

Prof. Dr. Knut Reinert
 Dr. Roland Krause
 Matthias Winkelmann
 Patrick Pett

Institut für Informatik
 AG Algorithmische Bioinformatik

Algorithmische Bioinformatik

6. Übungsblatt WS 11/12

Abgabe 12.1.2012 !!!

Name:

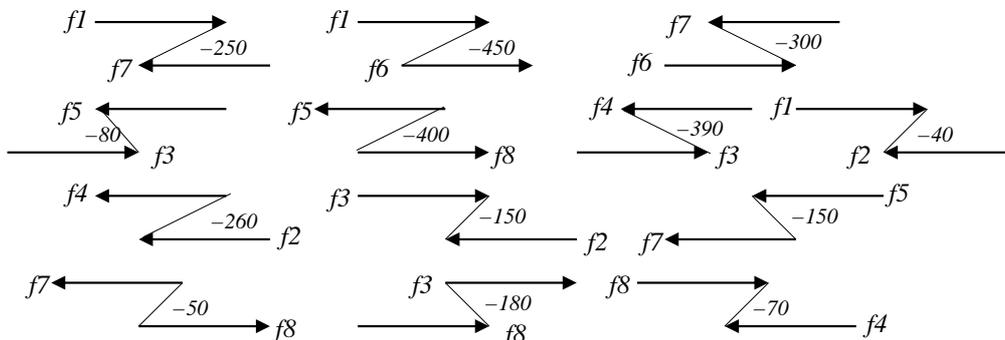
Matrikelnummer:

Übungsgruppe:

1 2 3

Aufgabe 1: 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, dass durch den MST 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 2: Der Contig-Mate Graph

Gegeben sind 4 Contigs der Länge 5000, 3000, 2000 und 2000. Zudem seien 6 Mate-Pairs bekannt: $\{f_1, g_1\}, \{f_2, g_2\}, \dots$ aus einer Library mit $\mu = 10k$ und standard deviation $\sigma = 1k$. Die anderen Paare kommen aus einer $\mu = 2k$ library. Für jeden Read steht in der folgenden Tabelle auf welchem Contig er liegt und die Anfangs- und Endpositionen

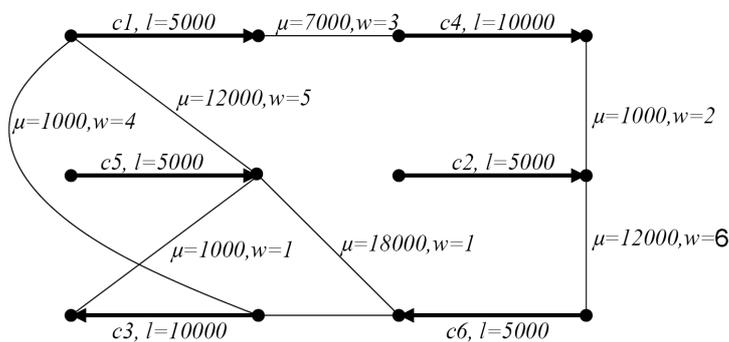
p_1 und p_2 innerhalb des Contigs an denen der Read beginnt bzw. endet. Alle Reads haben die Länge 500.

i	f_i		g_i		Library	
	contig	$p_1 - p_2$	contig	$p_1 - p_2$	μ	σ
1	B_1	500 – 1000	B_4	1500 – 2000	10k	1k
2	B_1	2500 – 3000	B_2	1500 – 1000	10k	1k
3	B_1	4000 – 4500	B_3	1500 – 2000	2k	100
4	B_1	4500 – 5000	B_3	1000 – 1500	2k	100
5	B_2	500 – 0	B_4	1500 – 1000	2k	100
6	B_2	1000 – 500	B_4	1000 – 500	2k	100

Konstruieren Sie den dazugehörigen initialen Contig-Mate Graph. Edge bundling und transitive edge reduction sind nicht erforderlich.

Aufgabe 3: Greedy Path-Merging Algorithmus

Gegeben sei folgender Contig-Mate Graph G . Zeigen Sie Schritt für Schritt wie der Greedy Path-Merging Algorithmus ein Layout der Contigs (Scaffolding) berechnet. Markieren Sie in jedem Schritt welche Mate-Pair Kante gerade aktiv ist und markieren Sie die selektierten Kanten. Welche impliziten Kanten werden eventuell eingefügt? Wie sieht das Scaffolding aus?



Aufgabe 4: Repeat Resolution

Seien die folgenden 5 reads mit 3 DNPs gegeben:

R1 ---A--T--C-----
R2 ---A--T--C-----
R3 ---A--C--G-----
R4 ---C--C--G-----
R5 ---C--C--A-----

- Geben Sie den gewichteten K_5 für das k -star problem an.
- Formulieren Sie das entsprechende ILP für $k = 2$.
- Geben Sie eine optimale Lösung an.