



Genomanalyse

6. Übungsblatt WS 2009

Abgabe: Dienstag 24.11.2009 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 11: [6 Punkte] Eindeutiger minimaler spannender Baum

Beweisen Sie, gerne mit Mitteln aus der Vorlesung:

Lemma. Sei $G = (V, E)$ ein zusammenhängender Graph und $w : E \rightarrow \mathbb{R}$ eine injektive Gewichtsfunktion auf den Kanten, das heißt, $e \neq f \Rightarrow w(e) \neq w(f)$. Dann hat G einen eindeutig bestimmten minimalen spannenden Baum.

Hinweis (aber es gibt bestimmt auch andere Ansätze, die zum Ziel führen):

Angenommen, T und T' sind verschiedene MST. Sei e die schwerste Kante in der symmetrischen Differenz von T und T' , also $e := \arg \max\{w(e) | e \in (T \setminus T') \cup (T' \setminus T)\}$. Machen Sie etwas mit e , um einen der beiden MSTs zu verbessern (obwohl diese ja schon optimal sind) – Widerspruch.

Lösung: O.B.d.A. gilt $e \in T \setminus T'$. Dann hat $T - e$ genau zwei Komponenten. Andererseits gibt es in T' einen Pfad $P \subseteq E$, der die Endknoten von e miteinander verbindet. Nicht alle Kanten aus diesem Pfad können auch in T sein. Sei also f eine Kante aus $P \cap (T' \setminus T)$. (Tatsächlich gibt es genau eine Kante in $P \cap (T' \setminus T)$.) Nach Wahl von e gilt $w(f) < w(e)$. Da f beide Komponenten von $T - e$ verbindet, ist $T - e + f$ ein spannender Baum, und es gilt $w(T - e + f) = w(T) - w(e) + w(f) < w(T)$. Das ist ein Widerspruch zur Optimalität von T .

Aufgabe 12: [7 Punkte] Contig-Mate Graph

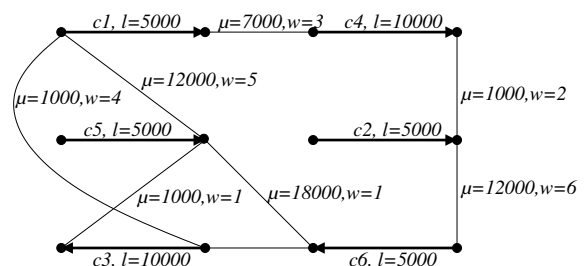
Gegeben sind 4 Contigs der Länge 5000, 3000, 2000 und 2000. Zudem seien 6 Mate-Pairs bekannt: $\{f_1, g_1\}, \{f_2, g_2\}, \dots$. Diese stammen aus zwei verschiedenen Libraries mit μ und σ wie angegeben. Für jeden Read steht in der folgenden Tabelle, auf welchem Contig er liegt, und die Anfangs- und 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

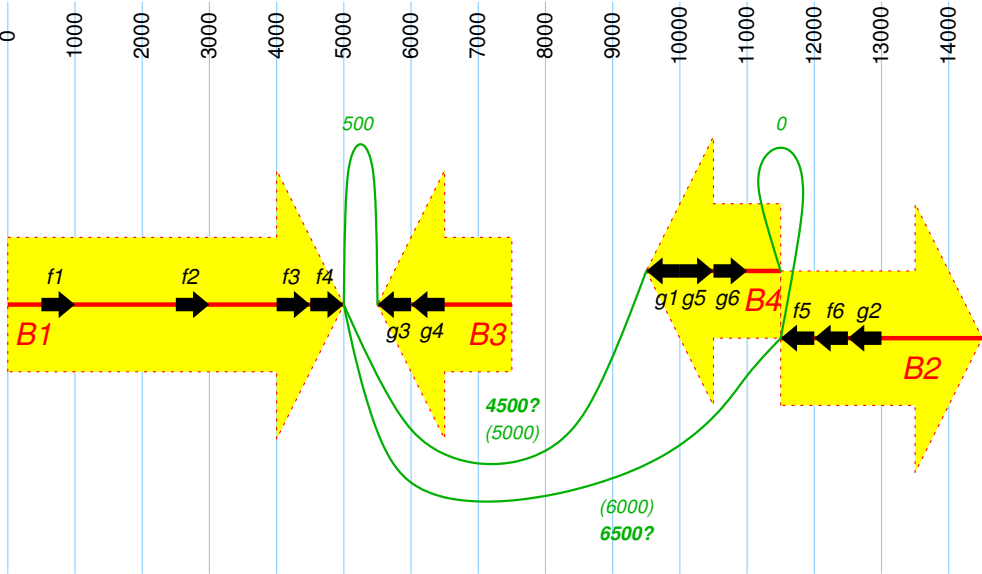
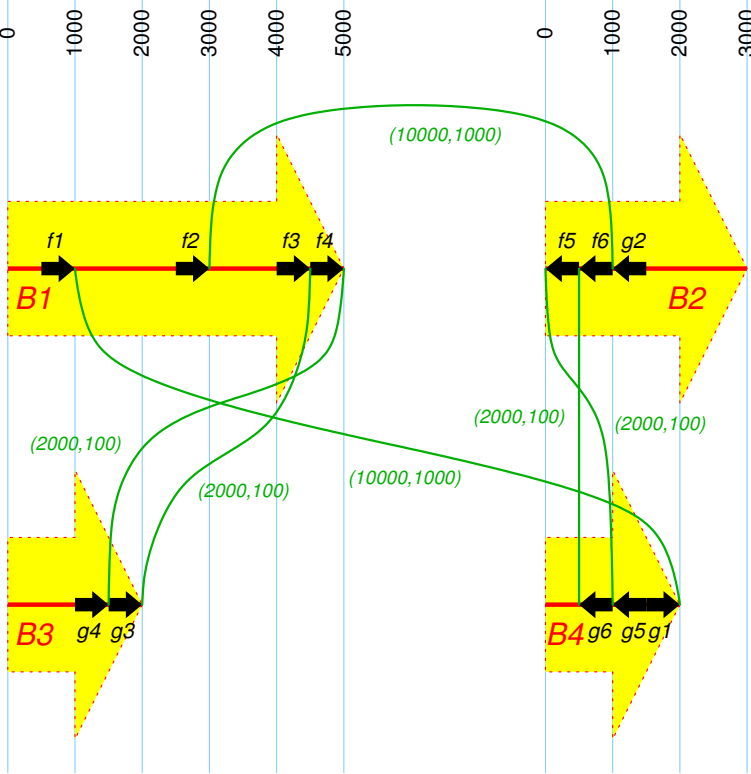
Konstruieren Sie den dazugehörigen initialen Contig-Mate Graph. Edge bundling und transitive edge reduction sind nicht erforderlich.

Aufgabe 13: [8 Punkte] 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?



Zu Aufgabe 12: Nachfolgend die Contigs mit den darin eingetragenen Fragmenten, darunter dann der resultierende Contig-Mate-Graph. Der Contig-Mate-Graph enthält einen Kreis, der etwas „unter Spannung“ steht (aber innerhalb der Standardabweichung). Die fettgedruckten Kantenlängen wurden für das Scaffolding verwendet. Im vorliegenden Fall würde man überprüfen, ob B_2 und B_4 an ihren Enden ein overlap alignment zulassen.



Zu Aufgabe 13: Nachfolgend das Tafelbild vom 25.11.2009.

