Ludwig-Maximilians-Universität München
Institut für Informatik
Prof. Dr. Volker Heun

Sommersemester 2023
Semestralklausur
26. Juli 2023

Algorithmische Bioinformatik: Bäume und Graphen

Vorname	Name	Matrikelnummer

Reihe	Platz	Unterschrift

Allgemeine Hinweise zur Semestralklausur

- Vor der Prüfung ist diese Seite mit Vornamen, Namen, Matrikelnummer, Reihe und Platz **leserlich mit Druckbuchstaben** zu versehen und zu unterschreiben.
- Bitte **nicht** in *roter* oder *grüner* Farbe bzw. **nicht** mit Bleistift schreiben.
- Der Studentenausweis und ein amtlicher Lichtbildausweis sind bereit zu halten.
- Die reine Bearbeitungszeit beträgt **120 Minuten**.
- Es sind insgesamt 40 Punkte zu erreichen, zum Bestehen sind 17 Punkte nötig.

Viel Erfolg!

Hörsaal verlassen von .... bis .... von .... bis ....

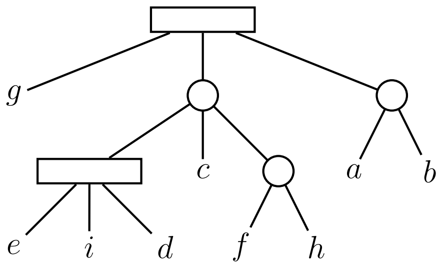
Vorzeitig abgegeben um ....

	Hz	A1	A2	A3	A4	A5	$\sum$
Erstkorrektur							
Nachkorrektur							
Zweitprüfer							

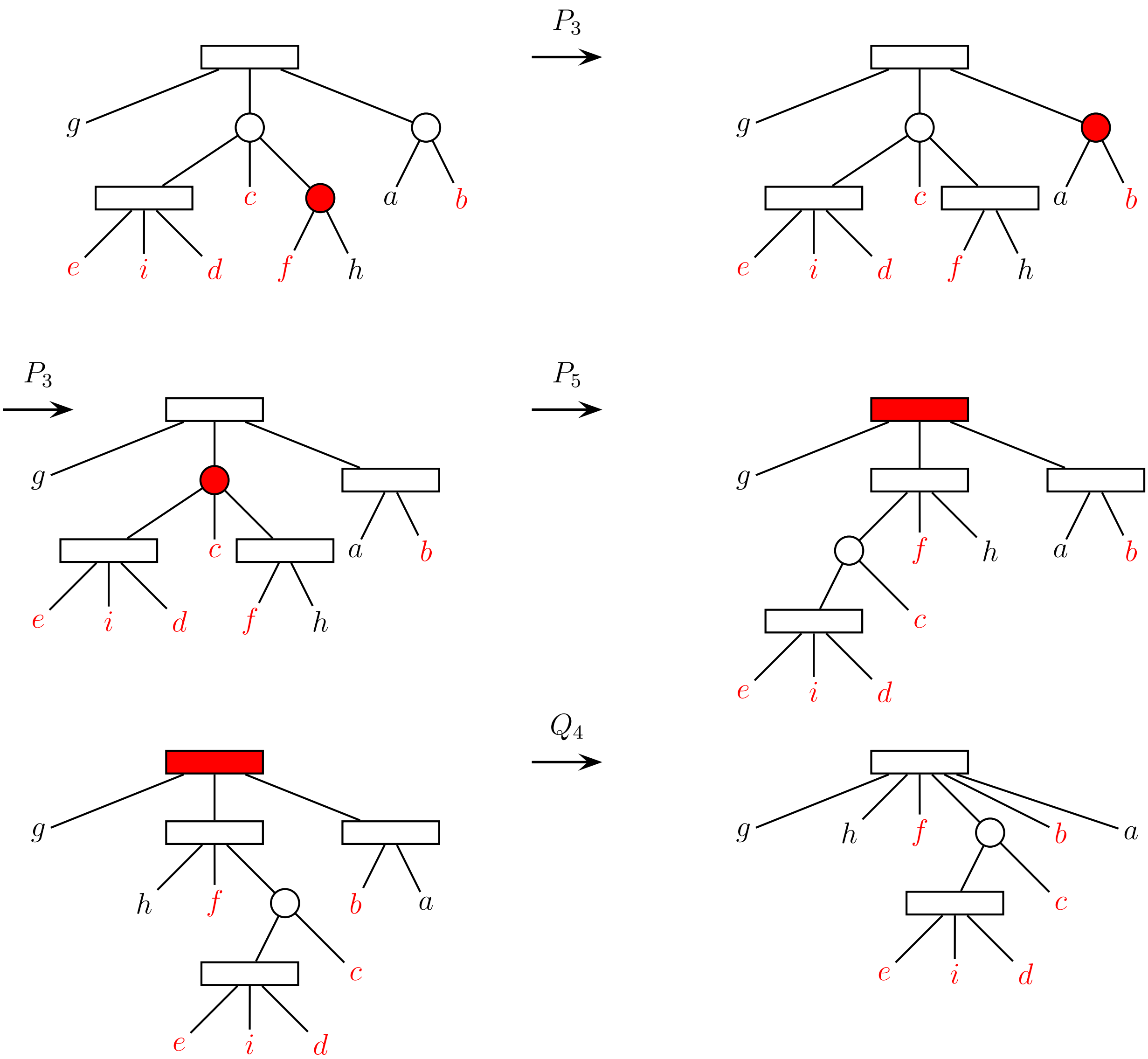
Vorname: ______________ Name: ______________ Matrikelnummer: ____________

Aufgabe 1 (8 Punkte)

Betrachte den rechts abgebildeten PQ-Baum. Modifiziere den PQ-Baum gemäß der in der Vorlesung angegebenen Schablonen, so dass zusätzlich auch die Restriktion $\{b, c, d, e, f, i\}$ erfüllt wird. Dabei sind alle Zwischenschritte anzugeben und zu dokumentieren.



Lösungsskizze (nicht ausreichend für die volle Punktzahl)



Vorname: ______________ Name: ______________ Matrikelnummer: ____________

Aufgabe 2 (8 Punkte)

Entscheide mit Hilfe des in der Vorlesung angegebenen Algorithmus, ob die unten angegebene Merkmalsmatrix eine perfekte binäre Phylogenie besitzt oder nicht. Falls ja, konstruiere einen phylogenetischen Baum, ansonsten gib eine Begründung an, warum es keinen geben kann.

M	a	b	c	d	e	f
A	0	1	1	1	0	0
B	0	0	1	1	0	0
C	0	0	0	1	0	1
D	0	0	0	1	0	0
E	1	0	0	0	1	0
F	0	0	0	0	1	0

Lösungsskizze (nicht ausreichend für die volle Punktzahl)

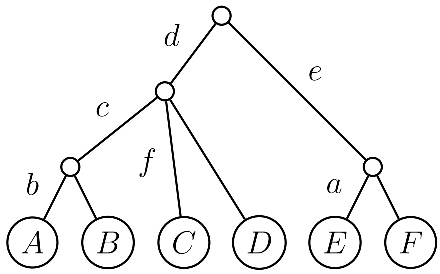
Wir sortieren erst die Spalten absteigend (interpretiert als Binärzahlen):

M'	d	c	b	f	e	a
A	1	1	1	0	0	0
B	1	1	0	0	0	0
C	1	0	0	1	0	0
D	1	0	0	0	0	0
E	0	0	0	0	1	1
F	0	0	0	0	1	0

Erstellen der Strings:

$$\begin{aligned} s_A &:= dcb\$ \\ s_B &:= dc\$ \\ s_C &:= df\$ \\ s_D &:= d\$ \\ s_E &:= ea\$ \\ s_F &:= e\$ \end{aligned}$$

Erstellen des Tries für $\{s_A, s_B, s_C, s_D, s_E, s_F\}$ (\$ bereits entfernt).



Alle Kantenlabels tauchen tatsächlich nur einmal auf.

Vorname: ______________ Name: ______________ Matrikelnummer: ______________

Aufgabe 3 (8 Punkte)

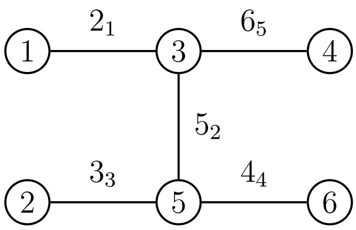
Gegeben seien die beiden folgenden 6×6-Distanzmatrizen D_ℓ und D_h. Entscheide mit Hilfe des in der Vorlesung angegebenen Algorithmus, ob es eine ultrametrische Matrix $D \in [D_\ell, D_h]$ gibt oder nicht. Gib ggf. die resultierende ultrametrische Matrix D an.

D_ℓ	1	2	3	4	5	6
1	0	4	1	3	5	6
2		0	5	5	2	3
3			0	3	4	5
4				0	4	4
5					0	3
6						0

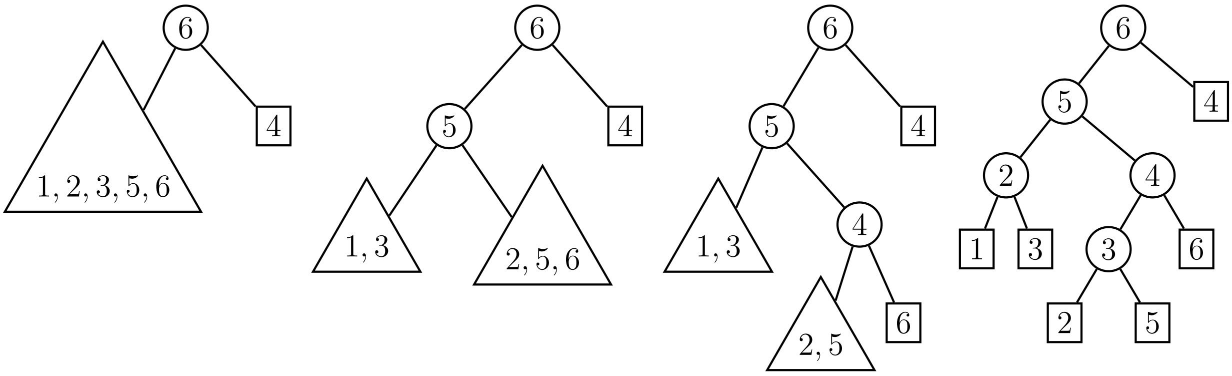
D_h	1	2	3	4	5	6
1	0	6	2	8	9	8
2		0	6	9	3	8
3			0	6	5	7
4				0	7	8
5					0	4
6						0

Lösungsskizze (nicht ausreichend für die volle Punktzahl)

Wir berechnen zuerst den minimalen Spannbaum für D_h nach Prim (die Indizes an den Kantengewichten geben die Reihenfolge an, in denen die Kanten hinzugefügt wurden).



Konstruktion des rekursiven Aufrufbaums (absteigend nach den Kantengewichten im Spannbaum, entspricht der Struktur des ultrametrischen Baums):



Damit ergibt sich die folgende ultrametrische Distanzmatrix D':

D'	1	2	3	4	5	6
1	0	5	2	6	5	5
2		0	5	6	3	4
3			0	6	5	5
4				0	6	6
5					0	4
6						0

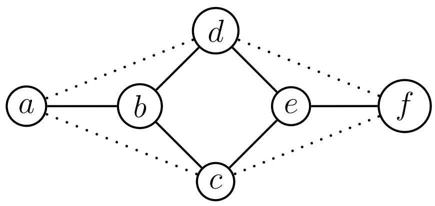
Man prüft leicht nach, dass $D_\ell(1,6) = 6 \not\leq 5 = D'(1,6)$ gilt. Also gibt es keine ultraetrische Matrix in $[D_\ell : D_h]$.

Vorname: ________________ Name: ________________ Matrikelnummer: ______________

Aufgabe 4 (8 Punkte)

Sei (V, M, F) eine Eingabe für das *Bounded Degree Interval Sandwich Problem*, wobei

$$\begin{aligned} V &= \{a, b, c, d, e, f\}, \\ M &= \Big\{\{a, b\}, \{b, c\}, \{b, d\}, \{c, e\}, \{d, e\}, \{e, f\}\Big\}, \\ F &= \Big\{\{a, c\}, \{a, d\}, \{c, f\}, \{d, f\}\Big\}. \end{aligned}$$

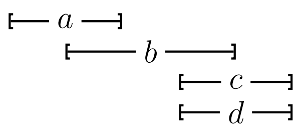


a) Weise anhand der Definition nach, dass $X = \{a, b, c, d\} \subsetneq V$ ein 3-zulässiger Kern ist.

b) Betrachte den 3-zulässigen Kern $X = \{a, b, c, d\} \subsetneq V$. Lässt sich dieser gemäß des Erweiterungslemmas zu einem Layout für $Y = X \cup \{e\}$ mit $d = 3$ erweitern?

Lösungsskizze (nicht ausreichend für die volle Punktzahl)

a) Wir weisen dazu die 6 Bedingungen der Definition für einen d-zulässigen Kern nach. Sei dazu das folgende Layout $L(X)$ für den Kern X in der Abbildung rechts gegeben.



1) $\forall \{v, w\} \in M \cap \binom{X}{2} \,:\, I(v) \cap I(w) \neq \emptyset$:
   $M \cap \binom{X}{2} = \{\{a, b\}, \{b, c\}, \{b, d\}\}$ und die 3 zugehörigen Intervall-Paare überlappen sich.

2) $\forall \{v, w\} \in F \cap \binom{X}{2} \,:\, I(v) \cap I(w) = \emptyset$:
   $F \cap \binom{X}{2} = \{\{a, c\}, \{a, d\}\}$ und die 2 zugehörigen Intervall-Paare überlappen sich nicht.

3) $\forall v \in \mathcal{A}(X) : r(I(v)) = \max\{r(I(w)) \,:\, w \in X\}$:
   $\mathcal{A}(X) = \{c, d\}$ und $r(I(a)) < r(I(b)) \leq r(I(c)) = r(I(d))$ und somit gilt die Bedingung.

4) $\forall v \in X \setminus \mathcal{A}(X) : I(v)$ schneidet höchstens $d = 3$ andere Intervalle:
   $X \setminus \mathcal{A}(X) = \{a, b\}$ und $I(a)$ bzw. $I(b)$ schneiden 1 bzw. 3 andere Intervalle. Somit gilt die Bedingung.

5) $\forall v \in \mathcal{A}(X) : I(v)$ schneidet höchstens $d - |E(v, X)|$ andere Intervalle, wobei $E(v, X) = \{\{v, w\} \in M \mid w \notin X\}$:
   $\mathcal{A}(X) = \{c, d\}$ und $|E(c, X)| = \{\{c, e\}\}$ bzw. $|E(d, X)| = \{\{d, e\}\}$. Weiterhin schneidet $I(c)$ bzw. $I(d)$ höchstens $3 - 1 = 2$ andere Intervalle, nämlich je 2.

6) $|\mathcal{A}(X)| \leq d - 1$:
   Es gilt $|\mathcal{A}(X)| = |\{c, d\}| = 2 \leq 3 - 1 = d - 1$.

b) Wir weisen für den 3-zulässiger Kern X nach, dass $Y = X \cup \{v\}$ mit $v = e$ eine Erweiterung zu einem 3-zulässigen Kern ist, indem wir die 4 Bedingungen des Lemmas nachweisen:

1) $\{v, w\} \notin F$ für alle $w \in \mathcal{A}(X)$:
$\mathcal{A}(X) = \{c, d\}$ und es gilt $\{c, e\} \notin F$ bzw. $\{d, e\} \notin F$.

2) X ein d-Layout L besitzt, so dass $I(u)$ für alle $u \in \mathcal{A}(X)$ mit $\{u, v\} \notin M$ höchstens $d - |E(u, X)| - 1$ andere Intervalle schneidet:
Wir betrachten das obige 3-Layout $L(X)$. Wieder ist $\mathcal{A}(X) = \{c, d\}$ und es gilt $E(c, X) = \{\{c, e\}\}$ bzw. $E(d, X) = \{\{d, e\}\}$, wobei $\{c, e\} \in M$ bzw. $\{d, e\} \in M$. Es ist also nichts zu zeigen.

3) $|\mathcal{A}(X)| \leq d - |E(v, Y)|$:
Wieder haben wir $\mathcal{A}(X) = \{c, d\}$ und $E(e, Y) = \{e, f\}$. Somit gilt offensichtlich $|\mathcal{A}(X)| = 2 \leq 3 - 1 = d - |E(e, Y)|$.

4) $|\mathcal{A}(Y)| \leq d - 1$:
Es gilt $\mathcal{A}(Y) = \{e\}$ und somit $\mathcal{A}(Y) = 1 \leq 3 - 1$.

Vorname: ______________ Name: ______________ Matrikelnummer: ______________

Aufgabe 5 (8 Punkte)

Sei $G = (V, E)$ ein zusammenhängender ungerichteter Graph und $T = (W, F)$ ein Baumzerlegung in Cliquen von G.

Zeige, dass G genau dann ein Intervall-Graph ist, wenn T ein Pfad ist.

Lösungsskizze (nicht ausreichend für die volle Punktzahl)

$\Rightarrow$: Wir nehmen o.B.d.A. an, dass die Knoten aufsteigend nach ihrem linken Intervallende nummeriert sind, d.h. es gilt $\ell(I(1)) \leq \cdots \leq \ell(I(n))$, wobei $V = [1 : n]$. Wir beweisen die Aussage mittels Induktion über die Knotenanzahl, wobei wir zusätzlich annehmen, dass sich der Knoten n in einem Endpunkt des Pfades T befindet.

IA ($n \leq 2$): Für $|V| \leq 2$ ist die Aussage trivial, da der zusammenhängende Graph dann jeweils vollständig ist und somit eine einzige Clique bildet.

IS ($n - 1 \to n$): Sei also $|V| > 2$. Wir definieren zunächst

$$N \;:=\; \mathcal{N}(n) \;=\; \{v \in [1 : n-1] \;:\; I(n) \cap I(v) \neq \emptyset\}$$

Nach der Nummerierung der Knoten ist $C := N \cup \{n\}$ eine Clique in G und N ein vollständiger Teilgraph von G. Es gilt $N \neq \emptyset$, da G zusammenhängend ist. Da $\ell(I(n)) \geq \ell(I(i))$ für alle $i \in [1 : n]$ gilt, ist dann $G' = (V', E \cap \binom{V'}{2})$ mit $V' := V \setminus \{n\} = [1 : n-1]$ ebenfalls zusammenhängend. Nach Induktionsvoraussetzung besitzt G' mit $|V(G')| = n - 1$ eine Baumzerlegung in Cliquen $T' = (W', F')$, der ein Pfad ist, und in der sich der Knoten $n-1$ in einem Endpunkt $C' \in W'$ von T' befindet, d.h. $n - 1 \in C'$.

Sei zuerst $C' = N$. Dann erhalten wir die neue Pfad-Zerlegung in Cliquen für G, indem wir C' durch C ersetzen. Der Knoten n ist dann offensichtlich in einem Endknoten C des Pfades T enthalten.

Sei andererseits $C' \neq N$. Nach Nummerierung der Intervalle müssen alle Intervalle, die $I(n)$ schneiden, auch $I(n-1)$ schneiden. Nach der Nummerierung muss somit jede Clique, die den Knoten $n-1$ enthält, auch die Knoten aus N enthalten, d.h. $N \subsetneq C'$. Dann setzen wir $T = (W' \cup \{C\}, F' \cup \{\{C', C\}\})$.

In jedem Fall erhalten wir eine Baumzerlegung T in Cliquen für G, wobei T ein Pfad ist und ein Endpunkt dieses Pfades den Knoten n enthält.

$\Leftarrow$: Sei also $T = (W, F)$ ein Pfad und sei daher ohne Beschränkung der Allgemeinheit $W = \{C_i \;:\; i \in [1 : m]\}$ mit $F = \{\{C_{i-1}, C_i\} \;:\; i \in [2 : m]\}$.

Wir konstruieren nun eine Intervall-Darstellung $\mathcal{I} = \{I(v) \;:\; v \in V\}$ für V. Für $v \in V$ sei $\overline{v} = \max\{i \in [1 : m] \;:\; v \in C_i\}$ und $\underline{v} = \min\{i \in [1 : m] \;:\; v \in C_i\}$. Dann gilt wegen der Baumzerlegung in Cliquen für $i \in [\underline{v} : \overline{v}]$, dass $v \in C_i$. Wir setzen dann $I_v = [\underline{v} - 0.1, \overline{v} + 0.1] \subseteq \mathbb{R}$.

Wenn $I(v) \cap I(w) \neq \emptyset$, dann gibt es ein $i \in [1 : m] \cap (I_v \cap I_w)$ mit $v, w \in C_i$. Also gilt $\{v, w\} \in E$.

Umgekehrt gilt für $\{v, w\} \in E$, dass es ein $C_i \in F$ mit $v, w \in C_i$ geben muss. Dann ist aber $i \in I(v) \cap I(w)$ und somit $I(v) \cap I(w) \neq \emptyset$.

Also besitzt G eine Intervalldarstellung und ist somit ein Intervallgraph.