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

Sommersemester 2024
Semestralklausur
25. Juli 2024

# 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 ...........

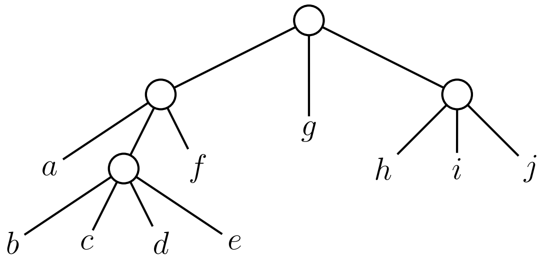
Vorzeitig abgegeben um ...........................

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

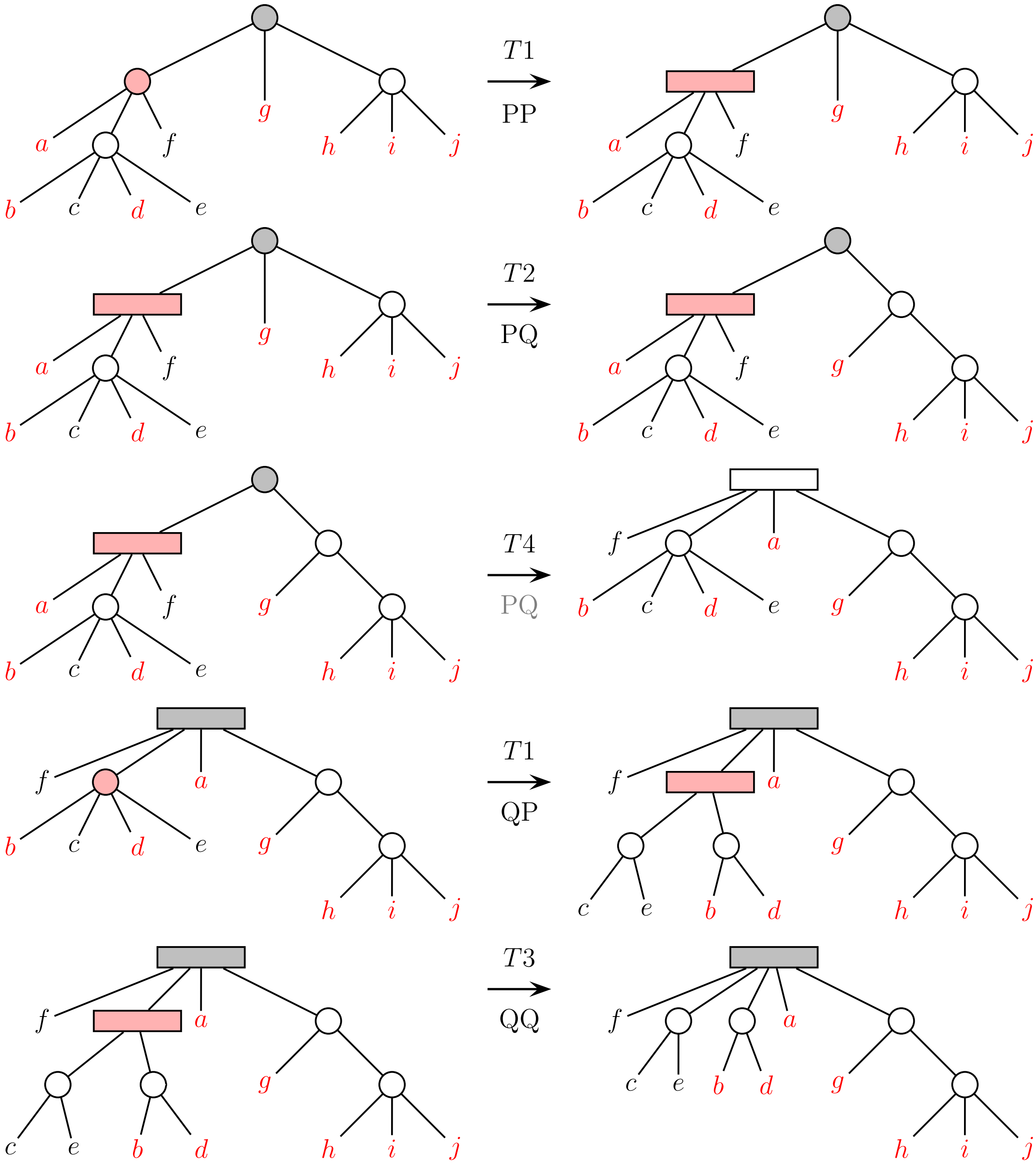
Vorname: ______________ Name: ______________ Matrikelnummer: ____________

## Aufgabe 1 (8 Punkte)

Betrachte den rechts abgebildeten PQR-Baum. Modifiziere diesen PQR-Baum gemäß den in der Vorlesung angegebenen Operationen, so dass der neue PQR-Baum zusätzlich noch die Restriktion $\{a, b, d, g, h, i, j\}$ erfüllt. Dabei sind alle Zwischenschritte anzugeben und zu dokumentieren.




**Lösungsskizze** (nicht ausreichend für die volle Punktzahl)




Die Anpassung der Wurzel am Ende verändert den Baum nicht weiter.

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)

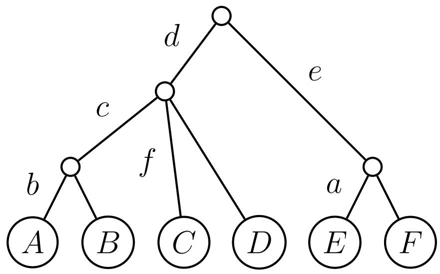
Wir sortieren erst die Spalten absteigend (interpretiert als Binärzahlen):

| $M'$ | $d$ | $c$ | $b$ | $f$ | $e$ | $a$ |
|---|---|---|---|---|---|---|
| $A$ | 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)

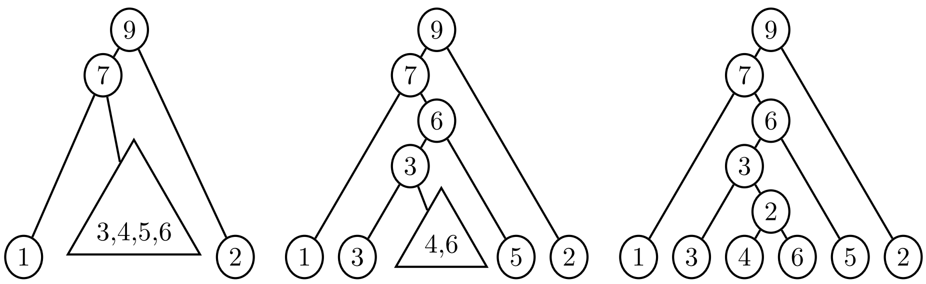
Gegeben sei die folgende $6 \times 6$-Distanzmatrix $D$. Entscheide mit Hilfe des in der Vorlesung angegebenen Algorithmus, ob $D$ additiv ist oder nicht. Falls ja, dann konstruiere einen additiven Baum für $D$, ansonsten gib eine Begründung an, warum es keinen geben kann.

| $D$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | 0 | 3 | 7 | 7 | 4 | 8 |
| 2 | | 0 | 8 | 8 | 5 | 9 |
| 3 | | | 0 | 4 | 7 | 5 |
| 4 | | | | 0 | 7 | 3 |
| 5 | | | | | 0 | 8 |
| 6 | | | | | | 0 |

**Lösungsskizze** (nicht ausreichend für die volle Punktzahl)

Das Maximum ist $D_{26} = 9$, also gilt $D'_{k\ell} = 9 - \frac{1}{2}(D_{2k} + D_{2\ell} - D_{k\ell})$:

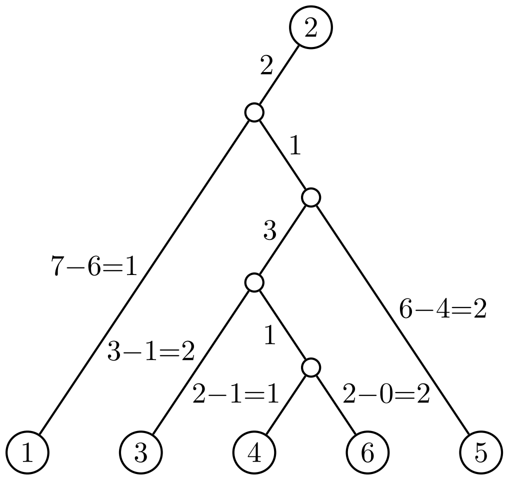
| $D'$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | 0 | 9 | 7 | 7 | 7 | 7 |
| 2 | | 0 | 9 | 9 | 9 | 9 |
| 3 | | | 0 | 3 | 6 | 3 |
| 4 | | | | 0 | 6 | 2 |
| 5 | | | | | 0 | 6 |
| 6 | | | | | | 0 |




Wie oben angegeben besitzt $D'$ einen (nach dem aus der Vorlesung berechnetet) ultrametrischen Baum.

Nach dem Lemma müssen nun noch die zum Blatt $\ell$ inzidenten Kanten um $9 - D_{2\ell}$ erniedrigt werden, was machbar ist:

| $b$ | $9 - M_{b,2}$ | |
|---|---|---|
| 1 | $9 - M_{2,1}$ | $9 - 3 = 6$ |
| 2 | $9 - M_{2,2}$ | $9 - 0 = 9$ |
| 3 | $9 - M_{2,3}$ | $9 - 8 = 1$ |
| 4 | $9 - M_{2,4}$ | $9 - 8 = 1$ |
| 5 | $9 - M_{2,5}$ | $9 - 5 = 4$ |
| 6 | $9 - M_{2,6}$ | $9 - 9 = 0$ |



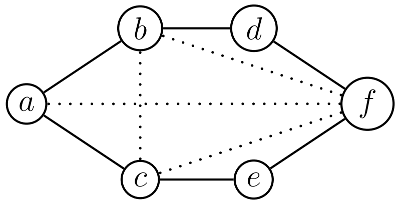

Kein Kantengewicht wird dabei negativ, also ist der konstruierte Baum und somit auch die gegebene Matrix additiv.

Vorname: ______________ Name: ______________ Matrikelnummer: ____________

## Aufgabe 4 (8 Punkte)

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

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

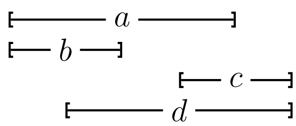



a) Weise anhand der Definition nach, dass $X = \{a, b, c, d\} \subsetneq V$ ein 4-zulässiger Kern ist.

b) Betrachte den 4-zulässigen Kern $X = \{a, b, c, d\} \subsetneq V$. Lässt sich dieser gemäß des Erweiterungslemmas zu einem Layout für $Y = X \cup \{e\}$ mit $d = 4$ 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) $\forall \{v, w\} \in M \cap \binom{X}{2} : I(v) \cap I(w) \neq \emptyset$:
   $M \cap \binom{X}{2} = \{\{a, b\}, \{a, c\}, \{b, d\}\}$ und die 3 zugehörigen Intervall-Paare überlappen sich.

2) $\forall \{v, w\} \in F \cap \binom{X}{2} : I(v) \cap I(w) = \emptyset$:
   $F \cap \binom{X}{2} = \{\{b, c\}\}$ und die 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 $r(I(b)) < r(I(a)) < r(I(c)) = r(I(d))$ und somit gilt die Bedingung.

4) $\forall v \in X \setminus \mathcal{A}(X) : I(v)$ schneidet höchstens $d = 4$ andere Intervalle:
   $X \setminus \mathcal{A}(X) = \{a, b\}$ und $I(a)$ bzw. $I(b)$ schneiden 3 bzw. 2 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. $|E(d, X)| = \{\{d, f\}\}$. Weiterhin schneidet $I(c)$ bzw. $I(d)$ höchstens $4 - 1 = 3$ andere Intervalle, nämlich 2 bzw. 3.

6) $|\mathcal{A}(X)| \leq d - 1$:
   Es gilt $|\mathcal{A}(X)| = |\{c, d\}| = 2 \leq 4 - 1 = d - 1$.

b) Wir weisen für den 4-zulässiger Kern $X$ nach, dass $Y = X \cup \{v\}$ mit $v = e$ eine Erweiterung zu einem 4-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 4-Layout $L(X)$. Wieder ist $\mathcal{A}(X) = \{c, d\}$ und es gilt $E(c, X) = \{\{c, e\}\}$ bzw. $E(d, X) = \{\{d, f\}\}$, wobei $\{c, e\} \in M$ bzw. $\{d, e\} \notin M$. Für $I(c)$ ist also nichts zu zeigen. Das Interval $I(e)$ schneidet 3 andere Intervalle, darf aber maximal $d - |E(u, X)| - 1 = 4 - 1 - 1 = 2$ andere Intervalle schneiden. Also ist eine Erweiterung von $X$ zu $Y$ **nicht** möglich.

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 $|\mathcal{A}(X)| = 2 \leq 4 - 1 = d - |E(e, Y)|$.

4) $|\mathcal{A}(Y)| \leq d - 1$:
Es gilt $\mathcal{A}(Y) = \{d, e\}$ und somit $\mathcal{A}(Y) = 2 \leq 4 - 1$.

Wegen Punkt 2 ist eine Erweiterung von $X$ zu $Y$ **nicht** möglich.

Vorname: ____________ Name: ____________ Matrikelnummer: ____________

## Aufgabe 5 (8 Punkte)

Sei $D$ eine $n \times n$-Distanzmatrix und sei $G(D) = (V, E, \gamma)$ der zu $D$ gehörige gewichtete ungerichtete Graph, wobei $V = [1 : n]$, $E := \binom{V}{2} := \{\{v, w\} : v \neq w \in V\}$ und $\gamma(v, w) = D(v, w)$.

Zeige, dass $D$ genau dann ultrametrisch ist, wenn es in jedem einfachen Kreis in $G(D)$ mindestens zwei Kanten mit maximalem Gewicht gibt.

*Hinweis:* In einem einfachen Kreis $C = (v_1, \ldots, v_\ell)$ der Länge $\ell$ haben genau dann mindestens zwei Kanten maximales Gewicht in $C$, wenn es $j \neq k \in [1 : \ell]$ gibt, so dass

$$\gamma(v_j, v_{j+1}) = \gamma(v_k, v_{k+1}) = \max\{\gamma(v_i, v_{i+1}) : i \in [1 : \ell]\},$$

wobei $v_{\ell+1} = v_1$ gilt.

**Lösungsskizze** (nicht ausreichend für die volle Punktzahl)

Wir müssen die Äquivalenz beweisen, also zwei Implikationen.

**$\Leftarrow$:** Zuerst halten wir fest, dass eine Distanzmatrix nach Definition definit und symmetrisch ist. Wenn es in jedem einfachen Kreis in $G(D)$ mindestens zwei Kanten mit demselben maximalem Gewicht gibt, dann auch in jedem Kreis der Länge genau 3. Dies ist die Dreipunktebedingung, von der wir in der Vorlesung nachgewiesen haben, dass daraus die ultrametrische Dreieckungleich für $D$ folgt.

**$\Rightarrow$:** Sei also jetzt $D$ ultrametrisch, Betrachte einen Kreis $C = (v_1, \ldots, v_\ell)$ der Länge $\ell$. Wir führen den Beweis mit Induktion über die Länge des Kreises, dass es in $C$ mindestens zwei Kanten maximalen Gewichts gibt.

**I.A. ($\boldsymbol{\ell = 3}$)** Nach der Dreipunktebedingung gilt für einen ultrametrische Matrix, dass in allen Kreisen der Länge 3 darin die beiden maximalen Distanzen gleich sind.

**I.S. ($\boldsymbol{\ell - 1 \to \ell}$)** Betrachte das Dreieck $(v_1, v_2, v_\ell)$. Wenn in diesem $\gamma(v_1, v_2) = \gamma(v_1, v_\ell)$ gilt, dann gilt dies natürlich auch im Kreis $C$ (eventuell nicht mit maximalem Gewicht in $C$).

Wenn dies nicht der Fall ist, muss wegen der Ultrametrik von $D$ und der Dreipunktebedingung gelten, dass $\gamma(v_1, v_2) = \gamma(v_2, v_\ell)$ oder $\gamma(v_1, v_\ell) = \gamma(v_2, v_\ell)$. O.B.d.A. gelte also $\gamma(v_1, v_2) = \gamma(v_2, v_\ell)$. Betrachte nun den Kreis $C' = (v_2, \ldots, v_\ell)$ der Länge $\ell - 1$. Nach Induktionsvoraussetzung haben in $C'$ zwei Kanten dasselbe maximale Gewicht. Sofern keine der beiden Kanten $(v_2, v_\ell)$ ist, haben diese beiden (oder eben $(v_1, v_2)$ und $(v_1, v_\ell)$ beide) auch im Kreis $C$ dasselbe maximale Gewicht (es muss dann mindestens $\gamma(v_2, v_\ell) \geq \max\{(v_1, v_2), (v_1, v_\ell)\}$ betragen).

Andernfalls ist $(v_2, v_\ell)$ eine Kante maximalen Gewichts und es muss nach Induktionsvoraussetzung eine weitere Kante $(v_i, v_j)$ in $C'$ geben. Da nun $(v_2, v_\ell)$ eine Kanten maximalen Gewichts in $C'$ und auch eine Kante maximalen Gewichts in $(v_1, v_2, v_\ell)$ gilt:

$$\gamma(v_1, v_2) = \gamma(v_2, v_\ell) = \gamma(v_i, v_j).$$

Damit ist $\gamma(v_1, v_2) = \gamma(v_i, v_j)$ und der Kreis $C$ besitzt zwei verschiedene Kanten $(v_1, v_2)$ und $(v_i, v_j)$, die maximales Gewicht in $C$ besitzen.