Beweis:

Annahme: $T(n) \leq c \cdot n$, wobei $c = c(m)$ konstant ist.

Die Annahme ist ok, falls $T(n) \leq$

$$T\left(\left\lceil \frac{n}{m} \right\rceil\right) + T\left(\left\lfloor \frac{3}{4}n \right\rfloor\right) + \left\lceil \frac{n}{m} \right\rceil C_m + \left\lfloor \frac{n}{2} \right\rfloor \leq cn; \text{ dies gilt, falls}$$

$$\left\lceil \frac{n}{m} \right\rceil c + \left\lfloor \frac{3}{4}n \right\rfloor c + \left\lceil \frac{n}{m} \right\rceil C_m + \left\lfloor \frac{n}{2} \right\rfloor \leq cn \qquad \Big| \cdot \frac{1}{n} \quad \text{(IA)}$$

$$\Leftrightarrow \text{(bis auf } \lceil\rceil, \lfloor\rfloor\text{)} \ \frac{c}{m} + \frac{3}{4}c + \frac{C_m}{m} + \frac{1}{2} \leq c$$

$$\Leftrightarrow -\frac{c}{m} - \frac{3}{4}c + c \geq \frac{C_m}{m} + \frac{1}{2}$$

$$\Leftrightarrow c \geq \frac{\frac{C_m}{m} + \frac{1}{2}}{1 - \frac{3}{4} - \frac{1}{m}}$$

$\square$

Bemerkung: $m = 11 \rightsquigarrow c = c(m) \approx 20.$

Literatur:

Vaughan R. Pratt, Frances F. Yao:
On lower bounds for computing the i-th largest element
Proc. 14th Ann. IEEE SWAT, pp. 70–81 (1973)

Manuel Blum, Robert W. Floyd, Vaughan R. Pratt, Ron L. Rivest, Robert E. Tarjan:
Time bounds for selection
J. Comput. Syst. Sci. **7**, pp. 448–461 (1973)

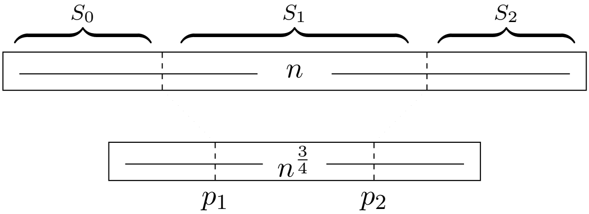
3. Randomisierter Median-Algorithmus

Problemstellung: Bestimme den Median von n Elementen

1. Wähle $n^{\frac{3}{4}}$ Elemente zufällig und gleichverteilt aus den n Elementen aus.
2. Sortiere diese $n^{\frac{3}{4}}$ Elemente mit einem (Standard-) $n \log n$-Algorithmus.
3. Setze

$$p_1 := \max\{\tfrac{n^{\frac{3}{4}}}{2} - \sqrt{n}, 1\}\text{-kleinstes Element der } n^{\frac{3}{4}} \text{ Elemente.}$$

$$p_2 := \min\{\tfrac{n^{\frac{3}{4}}}{2} + \sqrt{n}, n^{\frac{3}{4}}\}\text{-kleinstes Element der } n^{\frac{3}{4}} \text{ Elemente.}$$



4. Partitioniere die n Elemente in

   $S_0 := \{\text{Elemente} < p_1\}$
   $S_1 := \{p_1 \leq \text{Elemente} \leq p_2\}$
   $S_2 := \{p_2 < \text{Elemente}\}$

5. Falls $|S_0| \geq \lceil \frac{n}{2} \rceil$ oder $|S_2| \geq \lceil \frac{n}{2} \rceil$ oder $|S_1| \geq 4 \cdot n^{\frac{3}{4}}$, dann wiederhole den Algorithmus;
   ansonsten sortiere S_1 und liefere das $(\lceil \frac{n}{2} \rceil - |S_0|)$-kleinste Element davon ab.

Satz 80

Obiger randomisierter Algorithmus bestimmt den Median von n Elementen mit einer erwarteten Anzahl von $\frac{3}{2}n + o(n)$ Vergleichen.

Beweis:

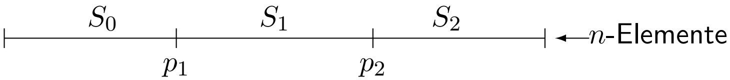
i) Korrektheit: klar.

Beweis (Forts.):

ii) Anzahl der Vergleiche in einer Iteration:

$$\mathcal{O}(n^{\frac{3}{4}} \log n^{\frac{3}{4}}) + \text{Kosten der Partitionierung}$$

Für die Partitionierung ist der naive Ansatz zu ungünstig, stattdessen:



Wähle zuerst jeweils mit Wahrscheinlichkeit $\frac{1}{2}$ aus, ob Element x mit p_1 oder p_2 verglichen wird, mache zweiten Vergleich nur, falls nötig.

Beweis (Forts.):

Die erwartete Anzahl von Vergleichen ist dann

$$
\begin{aligned}
&= \frac{n}{2}\left(\frac{|S_0|}{n}\cdot 1 + \frac{|S_1|+|S_2|}{n}\cdot 2\right) + \frac{n}{2}\left(\frac{|S_2|}{n}\cdot 1 + \frac{|S_0|+|S_1|}{n}\cdot 2\right)\\
&= \frac{n}{2}\left(\frac{|S_0|+|S_2|}{n} + 2\frac{\overbrace{|S_0|+|S_1|+|S_2|}^{n}+|S_1|}{n}\right)\\
&= \frac{n}{2}(3+\frac{|S_1|}{n}) = \frac{3}{2}n + o(n)
\end{aligned}
$$

Wir zeigen nun, dass der Algorithmus mit Wahrscheinlichkeit $\geq 1-\mathcal{O}(n^{-\frac{1}{4}})$ nur eine Iteration benötigt (daraus folgt dann, dass insgesamt die Anzahl der Vergleiche $\leq \frac{3}{2}n + o(n)$ ist).

Beweis (Forts.):

Dafür verwenden wir Hilfsmittel aus der Wahrscheinlichkeitstheorie/Stochastik:

- **Bernoulli-Zufallsvariable** (ZV): X, Werte $\in \{0, 1\}$ mit

$$X = \begin{cases} 1 \text{ mit WS } p \\ 0 \text{ mit WS } q = 1 - p \end{cases}$$

- **Erwartungswert** einer ZV:
  $\mathbb{E}[X] = \sum_{x \in \text{Wertebereich}} x \cdot \Pr[X = x]$
  (X ist diskret, d.h. der Wertebereich von X ist endlich)
- **Markov-Ungleichung**: $\Pr[X \geq t] \leq \frac{\mathbb{E}[X]}{t}$ für X nicht negativ
- **Chebyshev-Ungleichung**: $\Pr[|X - \mathbb{E}[X]| \geq t] \leq \frac{\mathrm{Var}(X)}{t^2}$

Beweis (Forts.):

- **Binomialverteilung**: Seien $X_1, \ldots, X_n$ unabhängige, identisch verteilte Bernoulli-Zufallsvariablen mit $\Pr[X_i = 1] = p$.

$$X := \sum_{i=1}^{n} X_i \,.$$

X ist binomial verteilt, mit Wertebereich $\{0, 1, \ldots, n\}$.

$$\begin{aligned}\Pr[X = k] &= \binom{n}{k} p^k (1-p)^{n-k}\\ \mathbb{E}[X] &= n \cdot p\\ \mathrm{Var}[X] &= n \cdot p \cdot (1-p) = n \cdot p \cdot q \,.\end{aligned}$$

In Zeichen: $X \sim \mathsf{B}(n, p)$

Beweis (Forts.):

Die Auswahl der $n^{\frac{3}{4}}$ Elemente wird wiederholt, falls $|S_0| \geq \frac{n}{2}$. Dies passiert gdw wir höchstens $\frac{1}{2}n^{\frac{3}{4}} - \sqrt{n}$ Elemente aus der Hälfte aller Elemente $\leq$ dem Median auswählen.

Wir bestimmen die Wahrscheinlichkeit dafür, dass keine neue Auswahl der $n^{\frac{3}{4}}$ Elemente stattfinden muss.
Setze Bernoulli-Zufallsvariable $X_1, \ldots, X_n$ mit:

$$X_i = \begin{cases} 1, \text{ falls Element } i < \text{Median ausgewählt wird} \\ 0, \text{ sonst} \end{cases}$$

$X := \sum X_i$ ist binomialverteilt mit Parametern n und $\frac{1}{2}n^{-\frac{1}{4}}$, und $\mathbb{E}[X] = \frac{1}{2}n^{\frac{3}{4}}$, $\mathrm{Var}[X] = n \cdot \frac{1}{2}n^{-\frac{1}{4}}(1 - \frac{1}{2}n^{-\frac{1}{4}}) = \frac{1}{2}n^{\frac{3}{4}}(1 - o(1))$.

Beweis (Forts.):

Die Wahrscheinlichkeit hierbei ist

$$\Pr[|S_0| \geq \frac{n}{2}] = \Pr[X \leq \frac{1}{2}n^{\frac{3}{4}} - \sqrt{n}] \leq \Pr[|X - \mathbb{E}[X]| \geq \sqrt{n}]$$
$$\leq \frac{\frac{1}{2}n^{\frac{3}{4}}(1 - o(1))}{n} \leq \frac{1}{2}n^{-\frac{1}{4}}(1 - o(1))$$

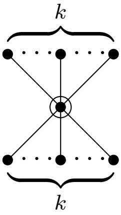
Die anderen beiden Wahrscheinlichkeitsbedingungen ($\Pr[|S_2| \geq \lceil \frac{n}{2} \rceil]$ und $\Pr[|S_1| \geq 4 \cdot n^{\frac{3}{4}}]$) ergeben analoge Abschätzungen.

Damit: Wiederholung mit WS $\leq \mathcal{O}(n^{-\frac{1}{4}})$. □

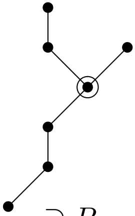
4. Schönhage/Paterson/Pippenger-Median-Algorithmus

Definition 81

Sei $k \in \mathbb{N}\setminus\{0\}$. P_k sei die folgende partielle Ordnung:



$2k+1$ Elemente

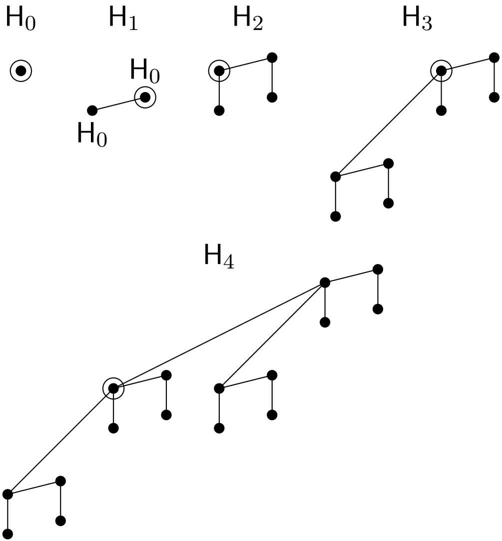


$\supseteq P_3$

Wir betrachten nun spezielle Binomialbäume mit „Zentrum“ (um Ordnungsinformation widerzuspiegeln).

Definition 82

1. Der Baum H_0 besteht aus einem Knoten, und dieser ist auch das Zentrum.
2. $H_{2h} (h > 0)$ besteht aus zwei H_{2h-1}, deren Zentren durch eine neue Kante verbunden sind. Das Zentrum des H_{2h} ist das kleinere der beiden Zentren der H_{2h-1}.
3. $H_{2h+1} (h \geq 0)$ besteht aus zwei H_{2h}, deren Zentren durch eine neue Kante verbunden sind, sein Zentrum ist das größere dieser beiden Zentren.


H_0
H_1
H_2
H_3
H_0
H_0
H_4

Lemma 83 (Zerlegungslemma)

a) *H_h hat 2^h Knoten, es werden $2^h - 1$ Vergleiche benötigt, um H_h zu konstruieren.*

b) *H_{2h} kann zerlegt werden in*

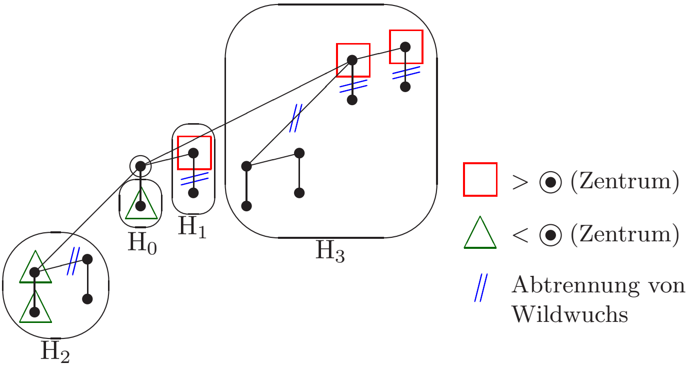
- *sein Zentrum*
- *eine Menge $\{H_1, H_3, \ldots, H_{2h-1}\}$ von disjunkten Teilbäumen, deren Zentren alle größer sind als das Zentrum von H_{2h}.*
- *eine Menge $\{H_0, H_2, H_4, \ldots, H_{2h-2}\}$ von disjunkten Teilbäumen mit Zentren kleiner als das von H_{2h}.*

Lemma 83 (Zerlegungslemma)

c) *H_{2h+1} kann so zerlegt werden, dass die Zusammenhangskomponente des Zentrums genau 2^h Knoten $\geq$ dem Zentrum enthält, indem höchstens $2^{h+1} - 1$ Kanten entfernt werden.*
H_{2h} kann so zerlegt werden, dass die Zusammenhangskomponente des Zentrums genau 2^h Knoten enthält, die alle $\leq$ dem Zentrum sind, indem höchstens $2^h - 1$ Kanten entfernt werden.

d) *Falls $k \leq 2^h - 1$, dann kann H_{2h} so zerlegt werden, dass die Zusammenhangskomponente des Zentrums genau $2k + 1$ Elemente enthält, von denen k größer und k kleiner als das Zentrum sind ($\Rightarrow P_k$).*
Dazu genügt es, höchstens $3k + 2h$ Kanten zu entfernen. Die restlichen Zusammenhangskomponenten sind wieder H_i's.

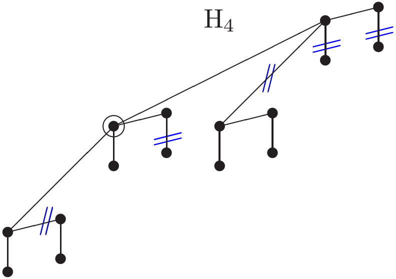
Zerlegungslemma



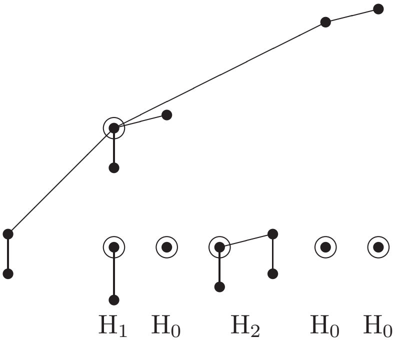
Bemerkung: Bei jedem Konstruktionsschritt wird ein Vergleich durchgeführt, um zu bestimmen, welcher der beiden Teilbäume das kleinere Zentrum hat. Im Algorithmus von Schönhage, Paterson und Pippenger werden aus Teilstücken H_r größere Bäume H_{r+1} zusammengebaut, wodurch schrittweise eine partielle Ordnung auf den Eingabewerten bestimmt wird. Wurde ein Baum H_{2h} hinreichender Größe hergestellt, so wird er durch Zerlegung in einen Baum umgewandelt, der nur noch sein altes Zentrum sowie k darüberliegende und k darunterliegende Elemente enthält, wobei $k \leq 2^h - 1$.

Beispiel 84

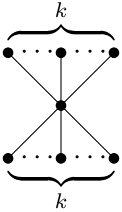
In diesem Beispiel wollen wir H_4 zerlegen und wählen $k = 3$:



Um einen H_4 derart zu zerlegen, müssen wir 5 Kanten aufbrechen. Dabei werden drei H_0, ein H_1 sowie ein H_2 abgespalten.



Übrig bleibt die gewünschte Struktur mit k Knoten über dem Zentrum und k unter dem Zentrum, wodurch eine partielle Ordnung auf $2k+1$ Eingabewerten bestimmt wurde:



$2k+1$ Elemente

Die bei der Zerlegung angefallenen Reststücke werden beim Aufbau weiterer Bäume benutzt. So geht das bereits angesammelte Wissen über die Ordnung der Elemente nicht verloren.

Beweis des Zerlegungslemmas:

Wir beginnen mit Teil a).

Lemma 85

H_r hat 2^r Knoten, es werden $2^r - 1$ Vergleiche benötigt, um H_r aufzubauen.

Beweis:

In jedem der r Konstruktionsschritte wird die Anzahl der Knoten verdoppelt. Da wir mit einem Knoten beginnen, hat H_r folglich 2^r Knoten. Die Anzahl der notwendigen Vergleiche C_r unterliegt folgender Rekursionsgleichung ($r \geq 1$):

$$C_r = 1 + 2C_{r-1} \text{ und } C_0 = 0\,.$$

Damit folgt sofort $C_r = 2^r - 1$. □

Beweis von b):

Lemma 86

H_r kann in folgende disjunkte Bereiche unterteilt werden:

- *sein Zentrum,*
- *eine Reihe $H_1, H_3, \ldots, H_{r-1}$ (falls r gerade) bzw. $H_1, H_3, \ldots, H_{r-2}$ (falls r ungerade) von Unterbäumen, deren Zentren alle über dem von H_r liegen,*
- *eine Reihe $H_0, H_2, \ldots, H_{r-2}$ (falls r gerade) bzw. $H_0, H_2, \ldots, H_{r-1}$ (falls r ungerade) von Unterbäumen, deren Zentren alle unter dem von H_r liegen.*

Beweis von b):

Beweis:

Durch Induktion über r.

Induktionsanfang: für H_0 gilt die Behauptung.

Induktionsannahme: die Behauptung gelte für H_{r-1}.

Beweis von b):

Beweis:

1. Sei $r = 2h, h > 0$.
   H_{2h} besteht aus zwei H_{2h-1}, wobei das kleinere der beiden alten Zentren das neue Zentrum z bildet. Wende auf den H_{2h-1}, der z enthält, die Induktionsannahme an. Wir können diesen Unterbaum also in z sowie $H_1, H_3, \ldots, H_{2h-3}$ (Zentren über z) und $H_0, H_2, \ldots, H_{2h-2}$ (Zentren unter z) partitionieren. Zusammen mit dem H_{2h-1}, dessen Zentrum über z liegt, ergibt sich die Induktionsbehauptung für $H_r = H_{2h}$.

Beweis von b):

Beweis:

2. Sei $r = 2h + 1$, $h \geq 0$. H_{2h+1} besteht aus zwei H_{2h}, wobei das größere der beiden alten Zentren das neue Zentrum z bildet. Wende auf den H_{2h}, der z enthält, die Induktionsannahme an. Wir können diesen Unterbaum also in z sowie $H_1, H_3, \ldots, H_{2h-1}$ (Zentren über z) und $H_0, H_2, \ldots, H_{2h-2}$ (Zentren unter z) partitionieren. Zusammen mit dem H_{2h}, dessen Zentrum unter z liegt, ergibt sich die Induktionsbehauptung für $H_r = H_{2h+1}$.

□

Beweis von c):

Wir bezeichnen im Folgenden mit H_{2h}^- den Baum, der entsteht, wenn wir H_{2h} so zerlegen, dass der verbleibende Rest nur mehr Elemente unterhalb des Zentrums (und das Zentrum selbst) enthält. Mit H_{2h+1}^+ bezeichnen wir den Baum, der entsteht, wenn wir H_{2h+1} so zerlegen, dass der verbleibende Rest nur mehr Elemente über dem Zentrum und dieses selbst enthält.

Lemma 87

H_{2h}^- und H_{2h+1}^+ haben jeweils 2^h Knoten. Bei der Herstellung aus H_{2h} bzw. H_{2h+1} werden $2^h - 1$ bzw. $2^{h+1} - 1$ Kanten aufgebrochen. Die wegfallenden Teile haben die Form H_s, $s < 2h$ bzw. $s < 2h + 1$.

Beweis:

Durch Induktion über r.

Induktionsanfang: für H_0 und H_1 gilt die Behauptung.

Induktionsannahme: die Behauptung gilt für alle H_p, $p < r$.

1. Sei $r = 2h$, $h > 0$. Wir betrachten die Partitionierung von H_{2h} mit Zentrum z wie in Lemma 86. Die Unterbäume H_1, H_3, ..., H_{2h-1} haben ihre Zentren oberhalb von z. Wir trennen sie von H_{2h}, indem wir h Kanten aufbrechen. Die abgetrennten Teile haben offensichtlich die Form H_s, $s < 2h$. Bei den Unterbäumen H_0, H_2, ..., H_{2h-2}, mit Zentren unterhalb von z, wenden wir jeweils die Induktionsannahme an, d.h. wir erzeugen H_0^-, H_2^-, ..., H_{2h-2}^-. Als Ergebnis erhalten wir H_{2h}^-.

Beweis (Forts.):

Damit gilt für die Zahl der aufzubrechenden Kanten $K^-(2h)$ zur Herstellung von H_{2h}^-:

$$K^-(2h) = h + \sum_{i=0}^{h-1} K^-(2i) \stackrel{I.A.}{=} h + \sum_{i=0}^{h-1} (2^i - 1) = \sum_{i=0}^{h-1} 2^i = 2^h - 1.$$

Für die Zahl $E^-(2h)$ der Elemente in H_{2h}^- gilt:

$$E^-(2h) = 1 + \sum_{i=0}^{h-1} E^-(2i) \stackrel{I.A.}{=} 1 + \sum_{i=0}^{h-1} 2^i = 1 + \underbrace{\sum_{i=1}^{h} 2^{i-1}}_{2^h-1} = 2^h.$$

Beweis (Forts.):

2. Sei $r = 2h + 1$, $h > 0$. Wir betrachten die Partitionierung von H_{2h+1} mit Zentrum z wie in Lemma 86. Die Unterbäume H_0, H_2, ..., H_{2h} haben ihre Zentren unterhalb von z. Wir trennen sie von H_{2h+1}, indem wir $h + 1$ Kanten aufbrechen. Die abgetrennten Teile haben offensichtlich die Form H_s, $s < 2h + 1$. Bei den Unterbäumen H_1, H_3, ..., H_{2h-1}, mit Zentren oberhalb von z, wenden wir jeweils die Induktionsannahme an, d.h. wir erzeugen H_1^+, H_3^+, ..., H_{2h-1}^+. Als Ergebnis erhalten wir H_{2h+1}^+. Damit gilt für die Zahl der aufzubrechenden Kanten $K^+(2h + 1)$ zur Herstellung von H_{2h+1}^+:

Beweis (Forts.):

$$\begin{aligned}
K^+(2h+1) &= h + 1 + \sum_{i=1}^{h} K^+(2(i-1)+1) \\
&\overset{I.A.}{=} h + 1 + \sum_{i=1}^{h} (2^i - 1) = 1 + \sum_{i=1}^{h} 2^i \\
&= 1 + \underbrace{\sum_{i=1}^{h+1} 2^{i-1}}_{2^{h+1}-1} - 1 = 2^{h+1} - 1.
\end{aligned}$$

Für die Zahl $E^+(2h+1)$ der Elemente in H^+_{2h+1} gilt:

$$E^+(2h+1) = 1 + \sum_{i=1}^{h} E^+(2(i-1)+1) \overset{I.A.}{=} 1 + \underbrace{\sum_{i=1}^{h} 2^{i-1}}_{2^h-1} = 2^h.$$

$\square$

Beweis von d):

Lemma 88

Falls $k \leq 2^h - 1$, dann kann H_{2h} so zerlegt werden, dass die Komponente des Zentrums genau $2k+1$ Elemente enthält, k davon über und k unter dem Zentrum. Dazu müssen $\leq 3k + 2h$ Kanten entfernt werden. Die entfernten Teile sind von der Form H_s, $s < 2h$.

Beweis:

Betrachte die Binärdarstellung von $k = k_0 2^0 + k_1 2^1 + \cdots + k_{h-1} 2^{h-1}$ und die Partitionierung von H_{2h} mit Zentrum z wie in Lemma 86.

Beweis (Forts.):

Für jedes i mit $k_i = 1$, betrachte H_{2i+1} aus der Sequenz H_1, H_3, $\ldots$, H_{2h-1} von Unterbäumen, deren Zentren oberhalb von z liegen, und schneide alle Elemente aus H_{2i+1}, die kleiner als sein Zentrum sind (bilde also H^+_{2i+1}). Dazu müssen höchstens $2k$ Kanten aufgebrochen werden, denn jedes $k_i = 1$ steht für 2^i in k, kostet aber nach Lemma 87 $K^+(2i+1) = 2^{i+1} - 1$ Kanten, also:

$$\sum_{i=0}^{h-1} k_i K^+(2i+1) \leq 2k\,.$$

Für jedes i mit $k_i = 0$, schneide H_{2i+1} ganz weg. Dabei werden $\leq h$ Kanten aufgebrochen. Genau k Elemente oberhalb z bleiben zurück, da jedes $k_i = 1$ für 2^i in k steht, und ein H^+_{2i+1} genau $E^+(2i+1) = 2^i$ Elemente enthält, also:

$$\sum_{i=0}^{h-1} k_i E^+(2i+1) = k\,.$$

Beweis (Forts.):

Für jedes i mit $k_i = 1$, betrachte H_{2i} aus der Sequenz H_0, H_2, $\ldots$, H_{2h-2} von Unterbäumen, deren Zentren unterhalb von z liegen, und schneide alle Elemente aus H_{2i}, die größer als sein Zentrum sind (bilde also H_{2i}^-). Dazu müssen höchstens $k-1$ Kanten aufgebrochen werden, denn jedes $k_i = 1$ steht für 2^i in k und kostet uns nach Lemma 87 $K^-(2i) = 2^i - 1$ Kanten, also:

$$\sum_{i=0}^{h-1} k_i(2^i - 1) \leq k - 1\,.$$

Für jedes i mit $k_i = 0$, schneide H_{2i} ganz weg. Dabei werden höchstens h Kanten aufgebrochen. Genau k Elemente unterhalb von z bleiben zurück, da jedes $k_i = 1$ für 2^i in k steht, und ein H_{2i}^- genau $E^-(2i) = 2^i$ Elemente enthält, also:

$$\sum_{i=0}^{h-1} k_i E^-(2i) = k\,.$$

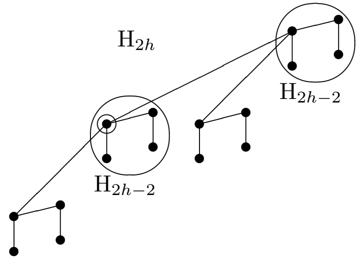
Beweis (Forts.):

Damit ergibt sich für die Gesamtanzahl aufzubrechender Kanten eine obere Schranke von $3k + 2h$. Lemma 87 liefert uns darüber hinaus die gewünschte Aussage über die Form der abgetrennten Teile. $\square$

Beweis von d):

Betrachte H_{2h}.

- „größer“: $H_{2h-1}, H_{2h-3}, \dots, H_1$
- „kleiner“: $H_{2h-2}, H_{2h-4}, \dots, H_0$



$U(h)$:= Anzahl der Elemente in $H_{2h} \geq$ Zentrum:
$U(h) = 2U(h-1) = 2^h;\ U(0) = 1$
$D(h)$:= Anzahl der Elemente in $H_{2h} \leq$ Zentrum:
$D(h) = 2D(h-1) = 2^h;\ D(0) = 1$

Beweis von d):

Anzahl der Kanten, die entfernt werden müssen:

$$\left.\begin{array}{ll} C_u(h) & \leq 2 + 2C_u(h-1) \\ & = 2 + 4 + 2^3 + \ldots + 2^h \\ & = 2^{h+1} - 2 \\ C_d(h) & \leq 1 + 2C_d(h-1) \\ & = 2^h - 1 \end{array}\right\} C(h) \leq 2^{h+1} - 2 + 2^h - 1 < 3 \cdot 2^h$$

Damit ist der Beweis des Zerlegungslemmas beendet. □