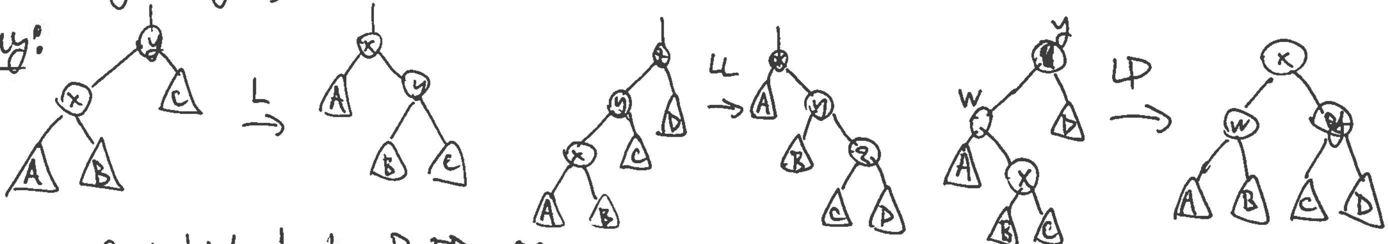
Splay stromy (Sleator & Tarjan, 1983)

Myšlenka: Kdykoli s nějakým prvkem pracujeme, vyrotujeme ho do kořene.
Pak by mohlo platit, že často používané prvky leží blízko kořene.
... ale aby to vyšlo, musíme rotovat šikovně:

Kroky:



... symetrický kroky P, PP a PL.

Operace Splay(x): Dokud x není kořen, provádím kroky, preferuji dvojité. (tedy L/P jen těsně pod kořenem)

Analýza: Zatím uvažujeme jen posloupnost operací Splay.

- Značení: $T(v)$:= podstrom zakořeněný ve v

  $m(v) := |T(v)|$ ... mohutnost vrcholu $m \le n$

  $r(v) := \log m(v)$ ... rank vrcholu (všechny log-y budou binární) $0 \le r \le \log n$

  $\Phi := \sum_v r(v)$ ... potenciál struktury $0 \le \Phi \le n \log n$

- Skutečná cena operace Splay := # rotací (každá trvá $\Theta(1)$ času)

Věta: Amortizovaná cena Splay(v) je nejvýše $3\cdot(r'(v) - r(v)) + 1$.] což je $O(\log n)$

vzhledem k potenciálu Φ — rank po — rank před

Lemma: Amort. cena kroků LL, PP, LP a PL je nejvýše $3(r'(v) - r(v))$
kroků L, P $3(r'(v) - r(v)) + 1$

Dk: zteleskopováním lemmat

Lemma: Pro $\forall \alpha, \beta \in \mathbb{R}^+$: $\log \frac{\alpha+\beta}{2} \ge \frac{\log\alpha + \log\beta}{2}$

Dk: z konkávnosti $\log(x)$. [nebo zlogaritmováním AG-nerovnosti]

Důsledek: $\log\frac{\alpha+\beta}{2} = \log(\alpha+\beta) - 1 \Rightarrow \log\alpha + \log\beta \le 2\log(\alpha+\beta) - 2$

Lemma LP: Amort. cena kroku LP $\le 3(r'(x) - r(x))$

Dk: Skutečná cena = 2

$$\Delta\Phi = (r'(w) - r(w)) + (r'(x) - r(x)) + (r'(y) - r(y))$$

Amort. cena $= 2 + \underbrace{r'(w)} + r'(x) + \underbrace{r'(y)} - \underbrace{r(w)}_{\ge r(x)} - r(x) - \underbrace{r(y)}_{\ge r(x)}$ ... chceme $\le 3r'(x) - 3r(y)$

~~$\log s'(w) + \log s'(y) \le \log(s'(w) + s'(y)) - 2$~~

$= \log m'(w) + \log m'(y) \le 2\log\underbrace{(m'(w) + m'(y))}_{\le m'(x)} - 2 \le 2r'(x) - 2$

∎

Lemma LL: Amort. cena kroku LL $\leq 3r'(x) - 3r(x)$.

Dk: Amort. cena $= 2 + r'(x) + r'(y) + r'(z) + r(x) - r(y) - r(z)$

Stejný trik na podstromy $T(x)$ a $T'(z)$:

$$r(x) + r'(z) = \log m(x) + \log m'(z) \leq 2\log(\underbrace{m(x) + m'(z)}_{\leq m'(x)}) - 2 \leq 2r'(x) - 2$$

$$\hookrightarrow r'(z) \leq 2r'(x) - r(x) - 2$$

Tudíž $A \leq 2 + 3r'(x) - 2r(x) - 2 + \underbrace{r'(y)}_{\leq r'(x)} - \underbrace{r(y)}_{\geq r(x)} - \underbrace{r(z)}_{= r'(x)} \leq 3r'(x) - 3r(x)$. ∎

Lemma L: Amort. cena kroku L $\leq 3r'(x) - 3r(x) + 1$.

Dk: Amort. cena $= 1 + r'(x) + \underbrace{r'(y)}_{\leq r'(x)} - r(x) - \underbrace{r(y)}_{\geq r(x)} \leq 1 + 2r'(x) - 2r(x)$

… ~~a~~ stačí si uvědomit, že $r'(x) \geq r(x)$ [monotonie] ∎

Kroky PP, PL a P díky symetrii vyjdou stejně.

OPERACE

Věta: Posloupnost m operací Find na Splay stromu s n vrcholy trvá $O((m+n)\log n)$ [$n \log n$ za ~~počáteční potenciál~~ pokles potenciálu $\Phi_0 - \Phi_n$]

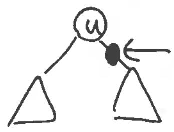
větu nejdřív vyslovit pro samotné Splayování

- **Find:** najdeme jako v obyčejném BVS, poté nalezený vrchol vysplayujeme
  ↳ nebo poslední navštívený

  … oproti samotnému Splay jsme skut. čas zvýšili konst.-krát ⇒ amortizovaný také
  (obecný princip: "účtujeme Splayi, ten pak amortizujeme")

- **Insert:** spustíme Find, najdeme u, který je buď Pred(x) nebo Succ(x), pak Splay(u)

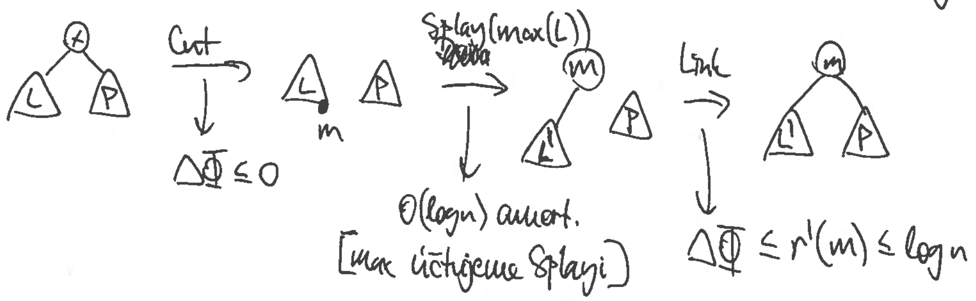
  DNO



← sem vložíme x

… to navíc stojí $O(1)$ skutečně $+ \Delta\Phi = r'(x) + r'(u) - r(u) \leq 2\log n$.

- **Delete:** Splay(x):



Celkem $O(\log n)$ amort.

Oproti jiným stromům umíme efektivní Link a Cut, takže je přirozené operace budovat z nich

Cv.: Insert pomocí Link a Cut

Cv: Nalezení následníka zadaného vrcholu.

Věta: Posloupnost m operací Find, Insert, Delete na zpočátku prázdném Splay stromu s nejvýše n vrcholy trvá $O(m \log n)$.

Statická optimalita

Uvažujme statický strom T, v němž hledáme prvek x.

Df: Cena přístupu $c_T(x) := \#$ vrcholů navštívených při hledání x v T.

Věta: Nechť T je statický BVS na n-prvkové množině X a $x_1 \ldots x_m \in X$ posloupnost hledaných prvků, obsahující všechny prvky X. Pak libovolný Splay strom na X provede při hledání $x_1 \ldots x_m$
$O(\cancel{n \log n +} \sum_{i=1}^{m} c_T(x_i))$ operací.

Odbočka: Vážená analýza Splay stromů

Vrcholům přiřadíme váhy $w(v) \in \mathbb{R}^+$ (algoritmus o nich neví!)

Předefinujeme mohutnost: $m(v) := \sum_{u \in T(v)} w(u)$ ← původní analýza tedy používá $w=1$ všude

Ponecháme definici ranku a potenciálu.

Už víme jenom $m(v) \geq 0$, $r(v) \in \mathbb{R}$, $\Phi \in \mathbb{R}$ (Φ může být záporný, pokud jsou váhy <1) !

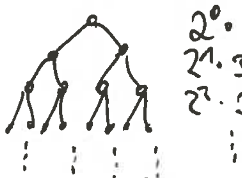
Ale pořád platí $T(u) \subseteq T(v) \Rightarrow m(u) \leq m(v) \Rightarrow r(u) \leq r(v)$.

Amortizace Splaye stále funguje, dává $A \leq 3(r'(v) - r(v)) \in O(r'(v) - r(v))$.
↳ jelikož $T'(v)$ obsahuje vše, můžeme psát $O(1 + \log \frac{W}{m(v)})$, kde $W = \sum_u w(u)$ (Σ všech vah)

Amort. cena je invariantní vůči změně měřítka vah, dokonce i $\Phi_m - \Phi_0$ je. } změna měřítka mění ranky o konst., Φ o $n \cdot c$

Dk věty: Prvku $x \in X$ nastavíme váhu $3^{-c_T(x)}$.

Lemma: $W \leq 1$ **Dk:** Kdyby T byl nekonečný úplný bin. strom, vyjde:



$2^0 \cdot 3^{-1}$
$2^1 \cdot 3^{-2}$
$2^2 \cdot 3^{-3}$

$$W = \frac{1}{3} \sum_{i \geq 0} \left(\frac{2}{3}\right)^{i} = \frac{1}{3} \cdot \frac{1}{1-\frac{2}{3}} = 1.$$

Přístup k x tedy stojí amortizovaně $O\left(1 + \log \frac{1}{3^{-c_T(x)}}\right) = O(c_T(x))$.
↑ jelikož $c_T(x) \geq 1$

Zbývá započítat $\Phi_0 - \Phi_m$ ... stačí si uvědomit, že:

~~$\Phi = \sum_v r(v) = \sum_v \log m(v) \geq \sum$~~

$1 \geq W \geq m(v) \geq w(v) = 3^{-c_T(v)}$

$0 \geq r(v) \geq -c_T(v)$

$$0 \geq \Phi \geq -\sum_{v \in X} c_T(v) \Rightarrow \Phi_0 - \Phi_m \in O\left(\sum_{v \in X} c_T(v)\right) \subseteq O\left(\sum_{i=1}^{m} c_T(x_i)\right).$$

Další odhady

• zmínit, že projití v rostoucím pořadí stojí $O(n)$.

Věta (Static Finger Bound): Mějme Splay strom na množině $\{1 \dots n\}$, posloupnost přístupů $x_1 \dots x_m$ a pevné $f \in \{1 \dots n\}$. Potom celková cena přístupů činí $O(n \log n + \sum_i \log(|x_i - f| + 1))$.

Dk: Nastavíme váhy: $w(x) := \dfrac{1}{(|x-f|+1)^2}$ … $W \in O(1)$.

~~$w(x) \geq \frac{1}{n^2}$, $m(x) \leq \frac{1}{n}$~~

$\frac{1}{n^2} \leq w(x) \leq 1$

$\frac{1}{n^2} \leq m(x) \leq n$

$-2 \log n \leq r(x) \leq \log n$

Cena přístupu je $O\left(\log \frac{W}{w(x_i)}\right) = O(\log(|x_i - f| + 1))$.

$\Phi_0 - \Phi_m \in O(n \log n)$

Věta (Working Set Bound): Mějme S. s. na $\{1 \dots n\}$ a posl. přístupů $x_1 \dots x_m$. Nechť t_i je ~~čas předchozího~~ # různých prvků, ke kterým bylo přistoupeno od předchozího přístupu k x_i do kroku i. — working set
(předchozího přístupu: začátek výpočtu, pokud ∄ předchozí)

Potom celková cena přístupů činí $O(n \log n + m + \sum_i \log(t(i) + 1))$.

Dk: Opět váhy, tentokrát se budou průběžně měnit.

Df: Move-to-front posloupnost MTF: Na počátku obsahuje prvky v pořadí prvního přístupu, ty, k nimž se nepřistupuje, v lib. pořadí na konci. Přístup k x_i přesune x_i na začátek.
$P(x) :=$ aktuální pozice prvku x v MTF.

$w(x) := \dfrac{1}{(P(x))^2}$ … $W \leq \sum_{x \geq 1} \frac{1}{x^2} \in O(1)$.

↳ přesněji $W = \sum_{i=1}^{n} \frac{1}{i^2}$ stále (váhy v průběhu jen permutujeme)

$\frac{1}{n^2} \leq w(x) \leq 1$

$\frac{1}{n^2} \leq m(x) \leq n$

$-2 \log n \leq r(x) \leq \log n$

Přístup k x_i stojí $O\left(1 + \log \frac{W}{w(x_i)}\right) = O(1 + \log \underbrace{P(x_i)}_{= t(i)+1})$.

Pak změníme váhy: prvku x_i se váha změní na 1 → vzroste, prvkům, které byly v MTF před x_i, klesne, ostatním se nezmění } rank kořene se nemění (je to $\log W$ stále), ostatní ranky nerostou ⇓ Φ neroste

A nakonec $\Phi_0 - \Phi_m \in O(n \log n)$.

Hypotéza (Dynamická optimalita): $\exists c$: Splay strom provede nejvýše c-krát více operací ($+ n \log n$), než libovolný dynamický strom. ← chodí od kořene nahoru/dolů, může rotovat