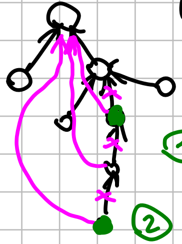
Union-Find



Union by Rank (UR)
$r(v) \le \log n$

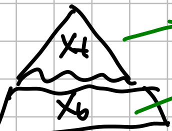
Path Compression (PC)

① $cost(\text{—}) :=$ # změn pointerů
Čas $= O(1 + cost)$

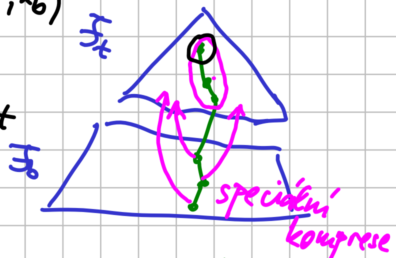
② nejdřív Uniony → les
pak komprese cest

Rozklad lesa $\mathcal{F}$ na $X \to (X_t, X_b)$

① X_t, X_b je rozklad X
② X_t je nahoru uzavřená:
pokud $v \in X_t$, pak $otec(v) \in X_t$



→ les $\mathcal{F}_t$
→ les $\mathcal{F}_b$



$\mathcal{F}$ les na X, $|X| = n$
C posl. kompresí na $\mathcal{F}$
$\|C\|$ # norm. kompresí v C ↳ m
$cost(C) \le f(m, n)$

Lemma: $\forall X, \mathcal{F}, C, \mathcal{F}_b, \mathcal{F}_t, X_b, X_t$
$\exists C_t, C_b$ posl. kompresí v $\mathcal{F}_t, \mathcal{F}_b$:
① $\|C_t\| + \|C_b\| \le \|C\|$
② $cost(C) \le cost(C_t) + cost(C_b) + |X_b| + \|C_t\|$ ~~− #roots($\mathcal{F}_b$)~~

Dk: T vrchol získá T otce → $cost(C_t)$
B vrchol získá B otce → $cost(C_b)$
B vrchol získá T otce — poprvé → $|X_b|$ − #roots($\mathcal{F}_b$)
— znovu → $\|C_t\|$

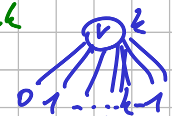
$f(m,n) :=$ max. $cost(C)$ pro C posl. komp. v n-vrcholovém lese, $\|C\| = m$

Věta: $f(m,n) \le (m+n)\log n$

Dk: Indukcí podle n … pro $n=1$: $f(m,n) = 0$
krok: $n \to n/2$ … X_t, X_b velikosti $n/2$
$n_t, n_b = n/2$ $m_t := \|C_t\|$ $m_b := \|C_b\|$
① $m_t + m_b \le m$
② $cost(C) \le (m_t + n/2)\log n/2 + (m_b + n/2)\cdot\log n/2 + n/2 + m_t$ ($n/2 \le n$, $m_t \le m$)
$\le m(\log n/2 + 1) + n(\log n/2 + 1)$ (každá závorka $= \log n$)
$\le (m+n)\log n.$

UR + PC

Rankový les $\mathcal{F}$ na X + $r: X \to \mathbb{N}$
① $r(v) :=$ výška $\mathcal{F}(v)$ měřená v hranách
② pokud $r(v) = k$, pak $\exists s_0 \dots s_k$ synové v t.ž. $r(s_i) = i$



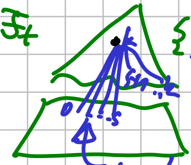
⇒ aspoň $r(v)+1$ synů

Pozorování: UR produkuje rankové lesy

$f(m,n,r)$ … max $cost(C)$ ($m = \|C\|$, $n = $ # vrcholů $|X|$, r = rank lesa)

Pozorování: $f(m,n,r) \le (r-1)n \le (r-1)m$ (k: $g(r) = r-1$)

Rozklad rankového lesa
parametr s $(0 \le s < r)$ ($r = r(\mathcal{F})$)



$\{v \mid r(v) > s\}$ rank. les ranku $\le r-s-1$
$\{v \mid r(v) \le s\}$ rankový les ranku $\le s$

Pro rank. rozklad z Lemmatu: ① $n_t + n_b = n$, $m_t + m_b \le m$
② $cost(C) \le f(m_t, n_t, r-s-1) + f(m_b, n_b, s) + (n - n_t) - (s+1)n_t + m_t$
$= n - (s+2)\cdot n_t$

Nechť $f(m,n,r) \le k\cdot m + n\cdot g(r)$
$cost(C) \le k\cdot m_t + n_t\cdot g(r-s-1) + f(m_b, n, s) + n + m_t - s\cdot n_t$
$\le (k+1)m_t + n_t\cdot g(r) + f(m_b, n, s) + n - s\cdot n_t$ ($s = g(r)$, oba členy se ruší)

$s := g(r)$

$f(m,n,r) \le (k+1)m_t + f(m_b, n, g(r)) + n$ $\big/ -(k+1)(m_t + m_b)$ ($\le m$)
$f(m,n,r) - (k+1)\cdot m \le f(m_b, n, g(r)) - (k+1)m_b + n$
$\varphi(m,n,r) \le \varphi(m_b, n, g(r)) + n$
$\varphi(m,n,r) \le n\cdot g^*(r)$
$f(m,n,r) \le n\cdot g^*(r) + (k+1)\cdot m$

Pokud
$f(m,n,r) \le km + n\cdot g(r)$
Pak
$f(m,n,r) \le (k+1)m + n\cdot g^*(r)$
$f(m,n,r) \le (k+i)m + n\cdot g^{*\cdots*}(r)$ (i hvězdiček)

dosadím tři, než
$f(m,n,r) \le n_t\cdot(r-s-2) + f(m_b, n_b, s) + n - (s+2)n_t + m_t$
$n_t\cdot(r-2s-4) + f(m_b, n_b, s) + n + m_t$ (< 0)

$s := r/2$
$f(m,n,r) \le f(m_b, n, r/2) + n + m_t$
$\boxed{f(m,n,r) \le n\cdot\log n + m}$

$f(m,n,r) \le (i+1)m + n\cdot\log^{*\cdots*}(r)$ (i hvězdiček)
$\alpha(n) := \min\{i \mid \log^{*\cdots*} r \le i\}$ (i hvězdiček)
$f(m,n,r) \le (\alpha(\log n)+1)m + n\cdot\alpha(\log n)$
$\le (\alpha(\log n)+1)(m+n)$

$\alpha(m,n,r) := \min\{i \mid \log^{*\cdots*} r \le m/n\}$ (i hvězdiček)
$f(m,n,r) \le (2 + \alpha(m,n,r))\,m$