Fun with Path Compression

Raimund Seidel

Universität des Saarlandes

In 1975 Bob Tarjan published

Theorem:

Any sequence of m Union, Find operations in a universe of n elements that uses linking by rank and path compression takes time at most

$$O(\, m \cdot \alpha(m,n) + n \,)$$

With the introduction of $\alpha(\ )$ into algorithm analysis Bob became the

With the introduction of $\alpha(\ )$ into algorithm analysis Bob became the

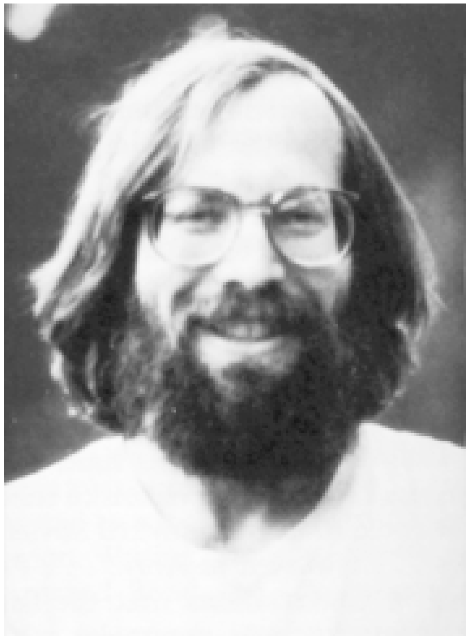
α – **male** of algorithms research.

With the introduction of $\alpha(\ )$ into algorithm analysis Bob became the

α – **male** of algorithms research.

With the introduction of α( ) into algorithm analysis Bob became the

α – **male** of algorithms research.



A two-parameter variation of the inverse Ackermann function can be defined as follows:

$$\alpha(m,n) = \min\{i \geq 1 : A(i, \lfloor m/n \rfloor) \geq \log_2 n\}.$$

This function arises in more precise analyses of the algorithms mentioned above, and gives a more refined time bound. In the disjoint-set data structure, *m* represents the number of operations while *n* represents the number of elements; in the minimum spanning tree algorithm, *m* represents the number of edges while *n* represents the number of vertices. Several slightly different definitions of α(*m*, *n*) exist; for example, $\log_2 n$ is sometimes replaced by *n*, and the floor function is sometimes replaced by a ceiling.

Ackermann function - Wikipedia, the free encyclopedia - Mozilla Firefox

Datei Bearbeiten Ansicht Gehe Lesezeichen Extras Hilfe

http://en.wikipedia.org/wiki/Ackerman's_function

LS FR Inf Uni AG Kurt MPI Talks DBLP Wikipedia

Definition and properties

[edit]

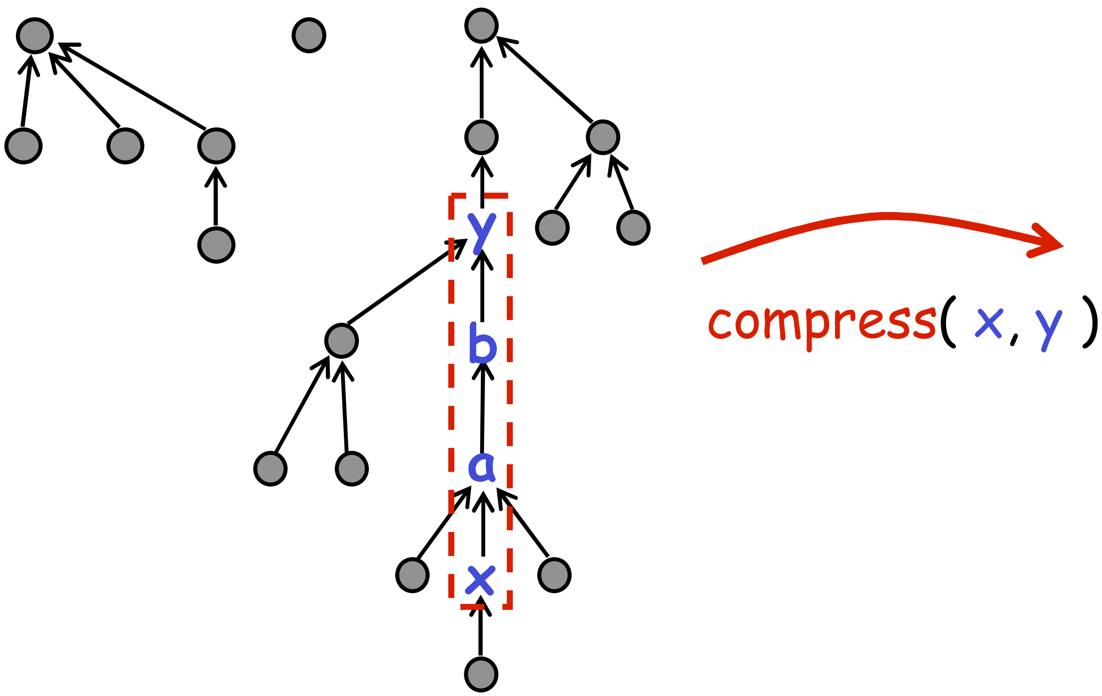
The Ackermann function is defined recursively for non-negative integers *m* and *n* as follows:

$$A(m,n) = \begin{cases} n+1 & \text{if } m = 0 \\ A(m-1, 1) & \text{if } m > 0 \text{ and } n = 0 \\ A(m-1, A(m, n-1)) & \text{if } m > 0 \text{ and } n > 0. \end{cases}$$

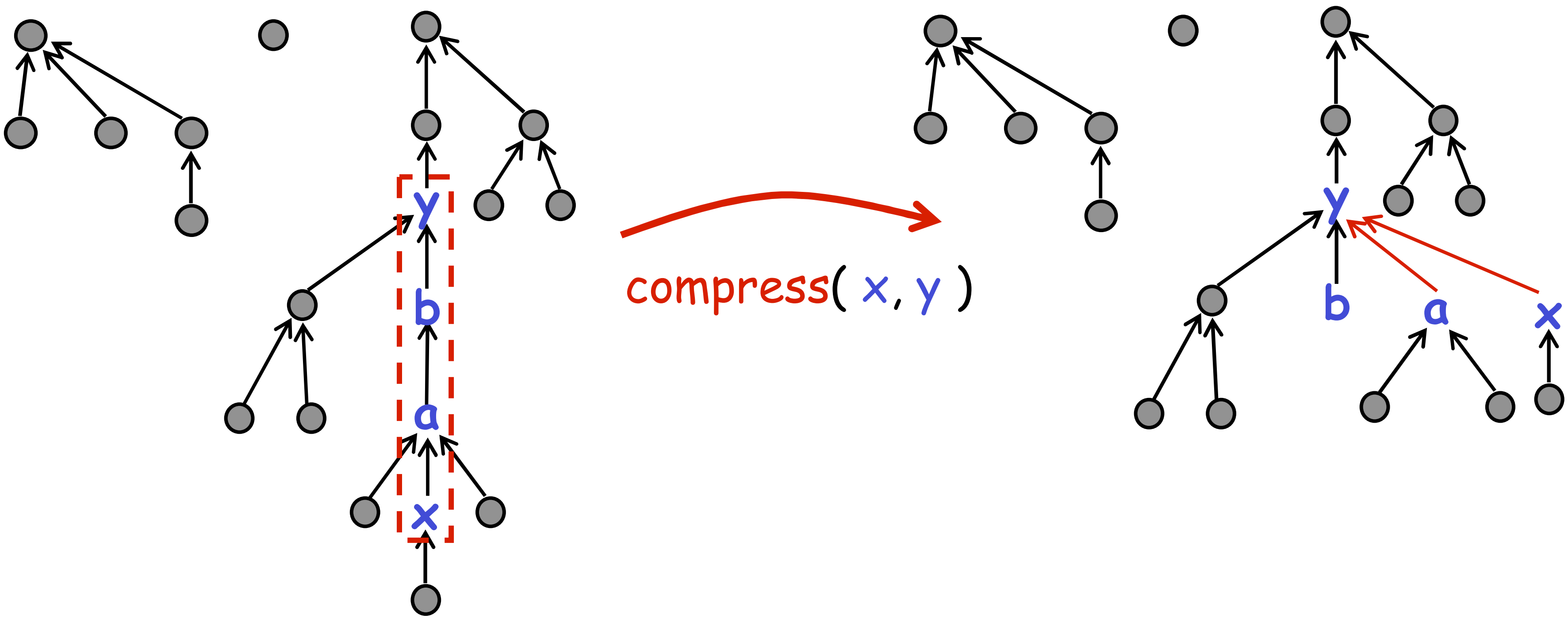
The Ackermann function can be calculated by a simple function based directly on the definition:

Fertig

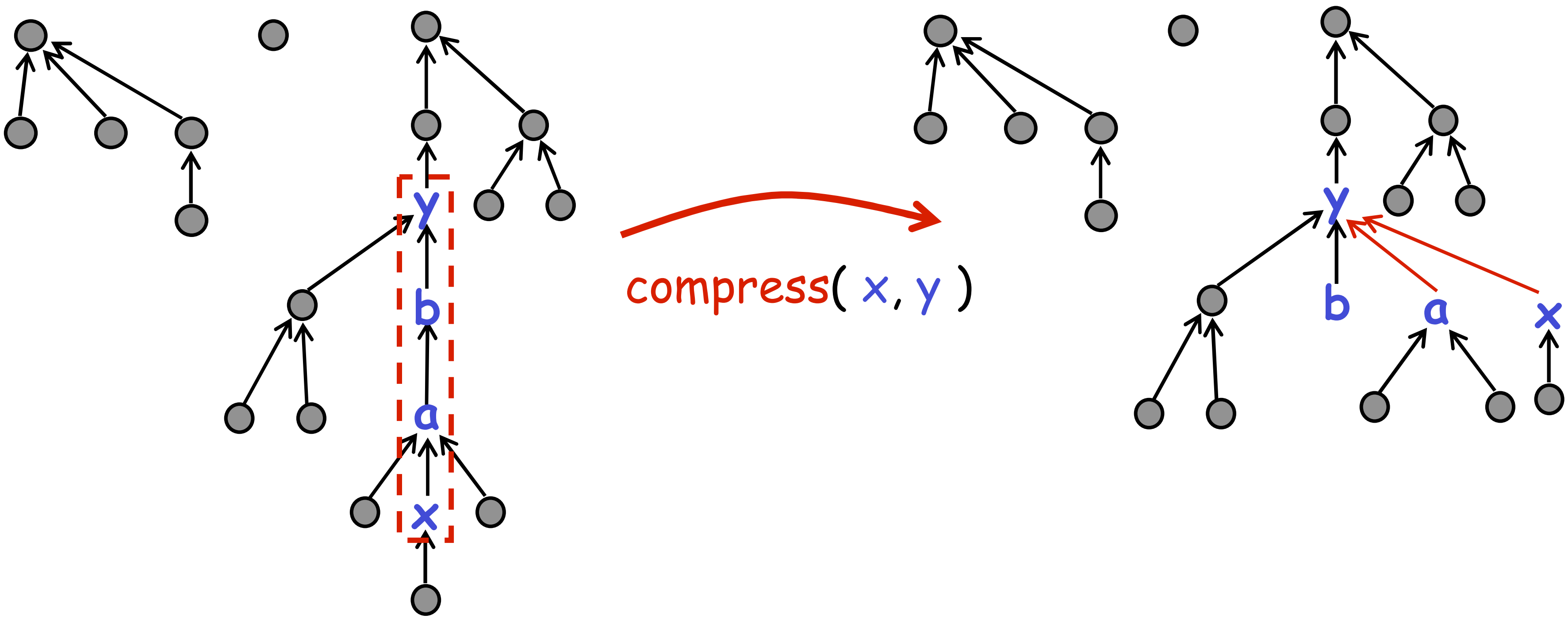
General path compression in forest $\mathcal{F}$



General path compression in forest $\mathcal{F}$



General path compression in forest $\mathcal{F}$



cost(compress(x,y)) = # of nodes that get a new parent

Problem formulation

$\mathcal{F}$ forest on node set X

C sequence of compress operations on $\mathcal{F}$

$|C|$ = # of true compress operations in C

cost(C) = $\sum$(cost of individual operations)

How large can cost(C) be at most,
in terms of |X| and $|C|$?

Dissection of a forest $\mathcal{F}$ with node set X :

partition of X into "top part" X_t
and "bottom part" X_b

so that top part X_t is "upwards closed",

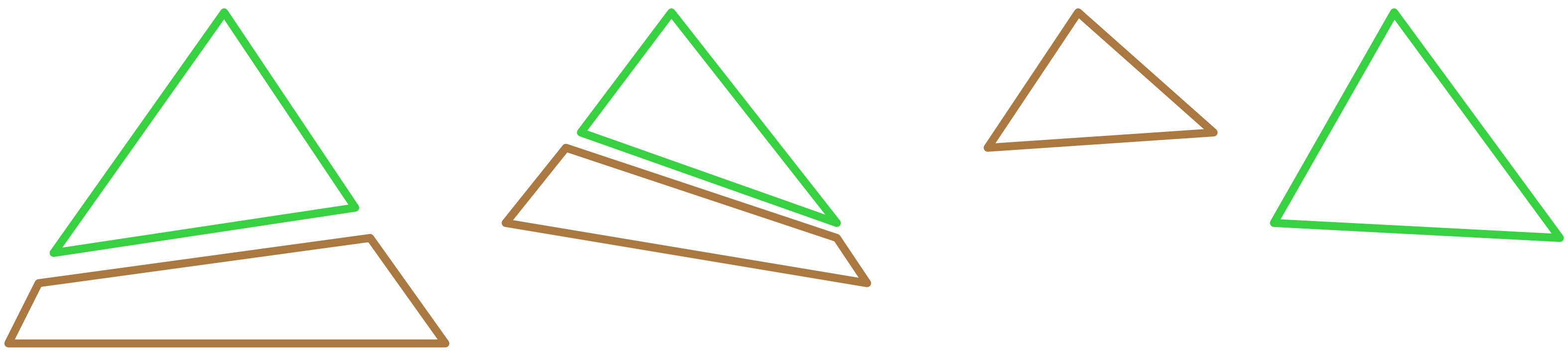
i.e. $x \in X_t \Rightarrow$ every ancestor of x is in X_t also

Dissection of a forest $\mathcal{F}$ with node set X :

partition of X into "top part" X_t
and "bottom part" X_b

so that top part X_t is "upwards closed",

i.e. $x \in X_t \Rightarrow$ every ancestor of x is in X_t also

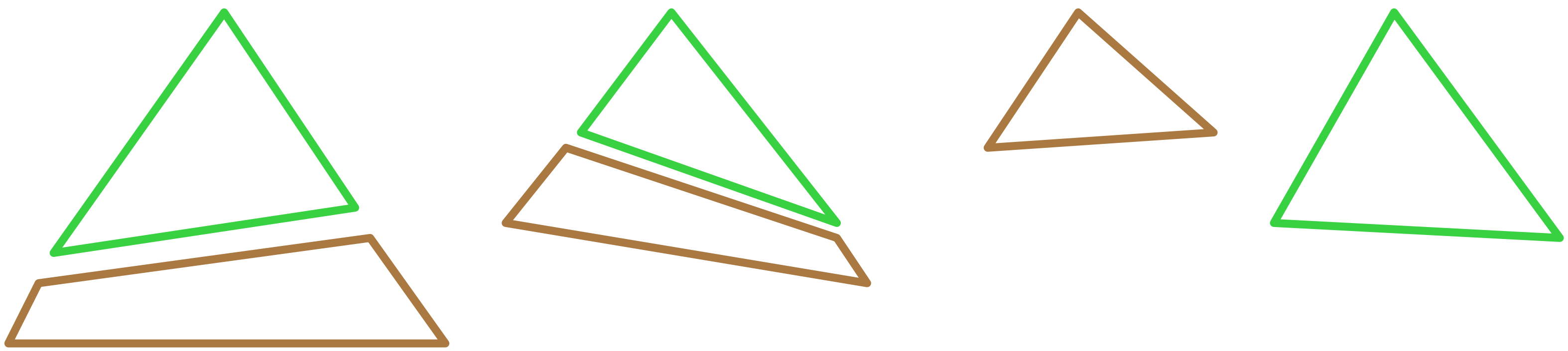


Dissection of a forest $\mathcal{F}$ with node set X :

partition of X into "top part" X_t
and "bottom part" X_b

so that top part X_t is "upwards closed",

i.e. $x \in X_t \Rightarrow$ every ancestor of x is in X_t also



Note: X_t, X_b dissection for $\mathcal{F}$
$\mathcal{F}'$ obtained from $\mathcal{F}$ by
sequence of path compressions $\Big\}$ $\Rightarrow$ X_t, X_b is dissection for $\mathcal{F}'$

Main Lemma:

$\mathcal{C}$ ... sequence of operations on $\mathcal{F}$ with node set X

X_t, X_b dissection for $\mathcal{F}$ inducing subforests $\mathcal{F}_t$, $\mathcal{F}_b$

Main Lemma:

C ... sequence of operations on $\mathcal{F}$ with node set X

X_t, X_b dissection for $\mathcal{F}$ inducing subforests $\mathcal{F}_t$, $\mathcal{F}_b$

$\Rightarrow$ $\exists$ compression sequences
C_b for $\mathcal{F}_b$ and C_t for $\mathcal{F}_t$
with

$$|C_b| + |C_t| \leq |C|$$

and

$$\mathrm{cost}(\,C\,) \leq \mathrm{cost}(\,C_b\,) + \mathrm{cost}(\,C_t\,) + |X_b| + |C_t|$$

Proof: 1) How to get C_b and C_t from C:

Proof: 1) How to get C_b and C_t from C:

compression paths from C

case 1: 

into C_t

Proof: 1) How to get C_b and C_t from C:

compression paths from C

case 1: 

into C_t

case 2: 

into C_b

Proof: 1) How to get C_b and C_t from C:

compression paths from C

case 1: 

into C_t

case 2: 

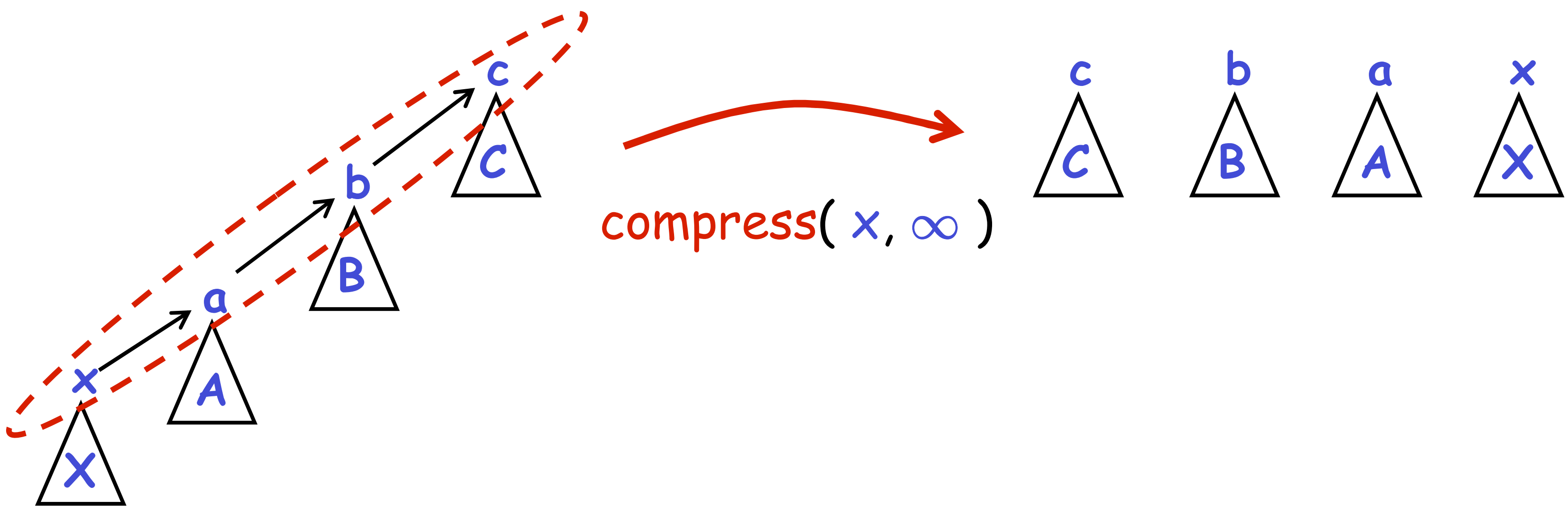
into C_b

case 3: 

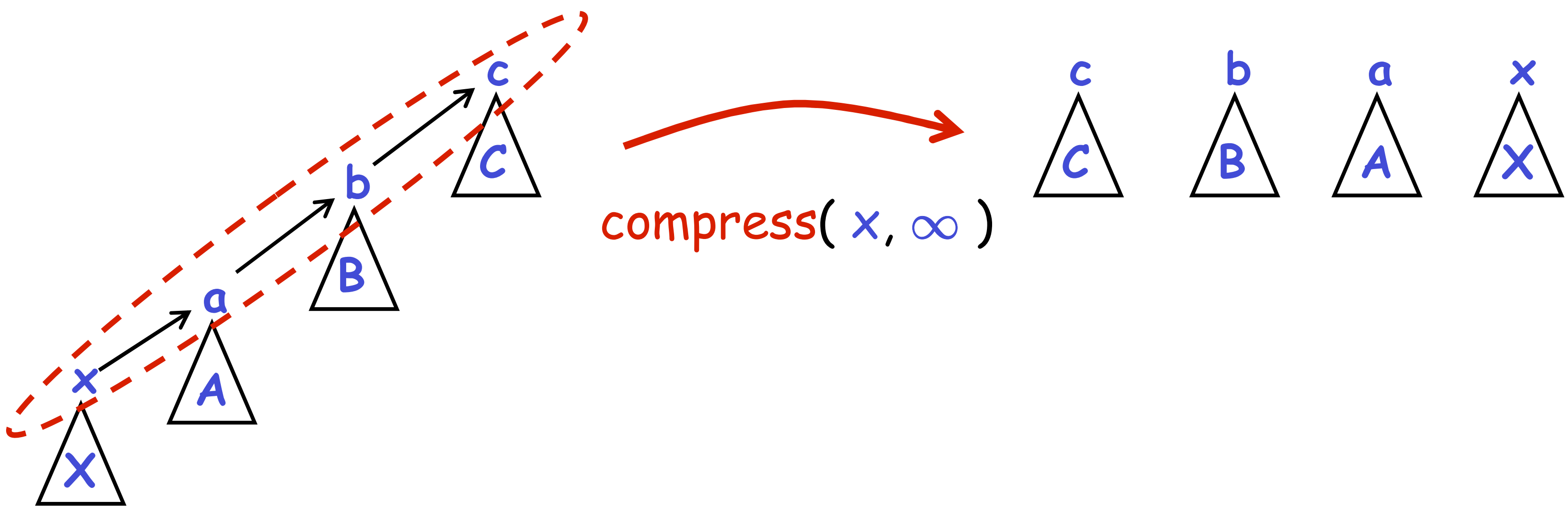
into C_t

into C_b

“rootpath compress”



"rootpath compress"



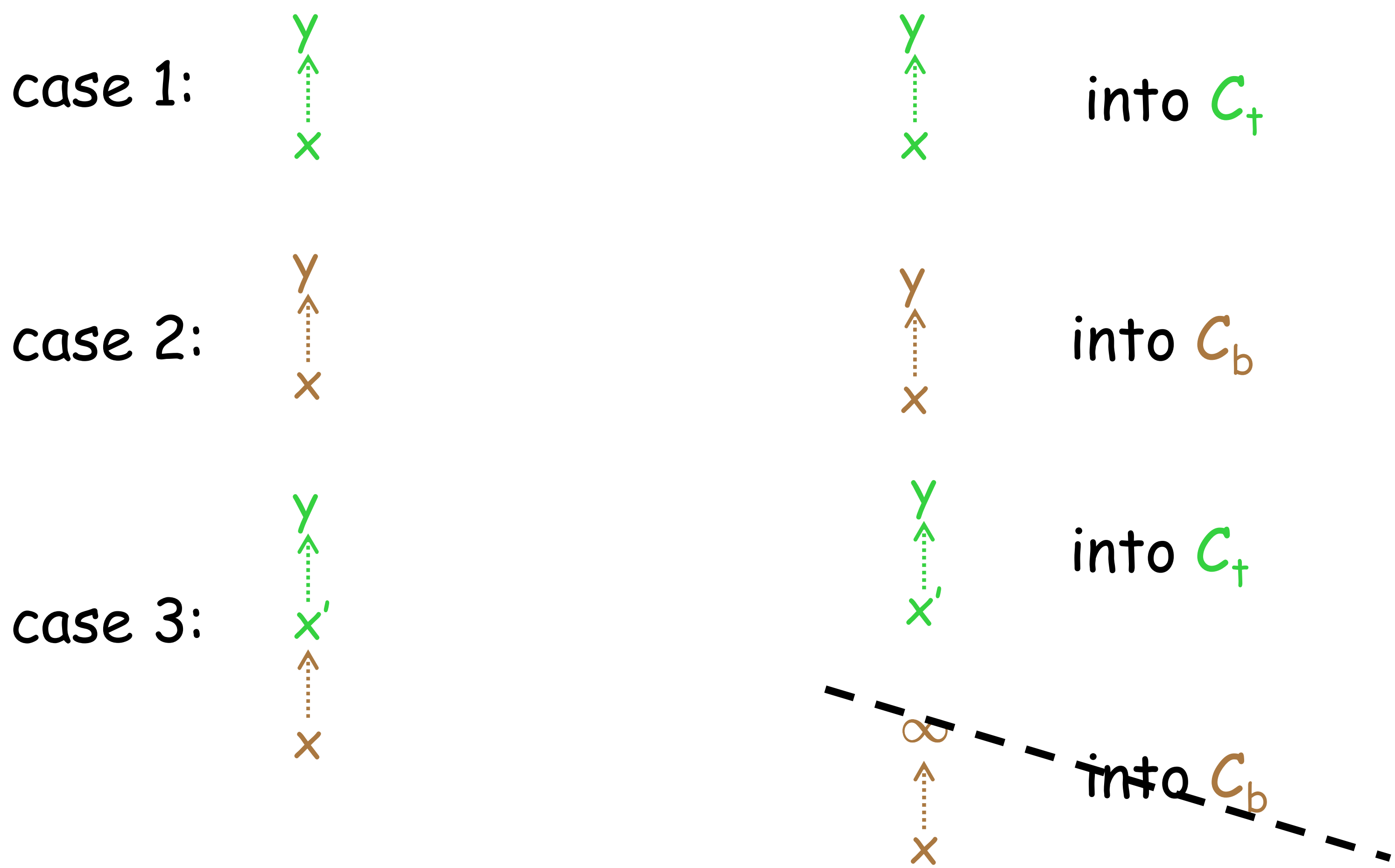
cost(compress(x, ∞)) = # of nodes that get a new parent

= 0

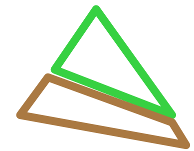
Proof:

$$|C_b| + |C_t| \leq |C|$$

compression paths from C

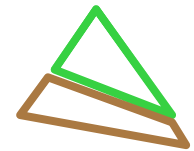


$$\text{cost}(C) \le \text{cost}(C_b) + \text{cost}(C_t) + |X_b| + |C_t|$$



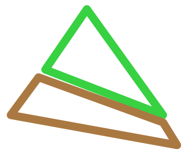
cost(C)	
green node gets new green parent:	accounted by $\text{cost}(C_t)$
brown node gets new brown parent:	accounted by $\text{cost}(C_b)$
brown node gets new green parent: for the first time	accounted by $\|X_b\|$
brown node gets new green parent: again	accounted by $\|C_t\|$

$$\text{cost}(C) \leq \text{cost}(C_b) + \text{cost}(C_t) + |X_b| + |C_t|$$



$\text{cost}(C)$	
green node gets new green parent:	accounted by $\text{cost}(C_t)$
brown node gets new brown parent:	accounted by $\text{cost}(C_b)$
brown node gets new green parent: for the first time	accounted by $|X_b|$ - #roots($\mathcal{F}_b$)
brown node gets new green parent: again	accounted by $|C_t|$

$$\text{cost}(C) \leq \text{cost}(C_b) + \text{cost}(C_t) + |X_b| - \#\text{roots}(\mathcal{F}_b) + |C_t|$$



cost(C)	
green node gets new green parent:	accounted by $\text{cost}(C_t)$
brown node gets new brown parent:	accounted by $\text{cost}(C_b)$
brown node gets new green parent: for the first time	accounted by $|X_b|$ $- \#\text{roots}(\mathcal{F}_b)$
brown node gets new green parent: again	accounted by $|C_t|$

Main Lemma':

$\mathcal{C}$ ... sequence of operations on $\mathcal{F}$ with node set X

X_t, X_b dissection for $\mathcal{F}$ inducing subforests $\mathcal{F}_t$, $\mathcal{F}_b$

$\Rightarrow$ $\exists$ compression sequences
$\mathcal{C}_b$ for $\mathcal{F}_b$ and $\mathcal{C}_t$ for $\mathcal{F}_t$
with

$$|\mathcal{C}_b| + |\mathcal{C}_t| \leq |\mathcal{C}|$$

and

$$\mathrm{cost}(\mathcal{C}) \leq \mathrm{cost}(\mathcal{C}_b) + \mathrm{cost}(\mathcal{C}_t) + |X_b| - \#\mathrm{roots}(\mathcal{F}_b) + |\mathcal{C}_t|$$

Path compression and union by rank

Path compression and union by rank

Def: $\mathcal{F}$ forest, x node in $\mathcal{F}$

$r(x)$ = height of subtree rooted at x

($r(\text{leaf}) = 0$)

$\mathcal{F}$ is a **rank forest**, if

for every node x

for every i with $0 \leq i < r(x)$,

there is a child y_i of x with $r(y_i) = i$.

Path compression and union by rank

Def: $\mathcal{F}$ forest, x node in $\mathcal{F}$

$r(x)$ = height of subtree rooted at x

($r(\text{leaf}) = 0$)

$\mathcal{F}$ is a **rank forest**, if

for every node x

for every i with $0 \leq i < r(x)$,

there is a child y_i of x with $r(y_i) = i$.

Note: Union by rank produces rank forests !

Path compression and union by rank

Def: $\mathcal{F}$ forest, x node in $\mathcal{F}$

$r(x)$ = height of subtree rooted at x

($r(\text{leaf}) = 0$)

$\mathcal{F}$ is a **rank forest**, if

for every node x

for every i with $0 \le i < r(x)$,

there is a child y_i of x with $r(y_i)=i$.

Note: Union by rank produces rank forests !

Lemma: $r(x)=r \Rightarrow x$ has at least r children.

Path compression and union by rank

Def: $\mathcal{F}$ forest, x node in $\mathcal{F}$

$r(x)$ = height of subtree rooted at x

($r(\text{leaf}) = 0$)

$\mathcal{F}$ is a **rank forest**, if

for every node x

for every i with $0 \leq i < r(x)$,

there is a child y_i of x with $r(y_i)=i$.

Note: Union by rank produces rank forests !

Lemma: $r(x)=r \Rightarrow x$ has at least r children and $\geq 2^r$ descendants.

Inheritance Lemma:

$\mathcal{F}$ rank forest with maximum rank r and node set X

s$\in\mathbb{N}$: $X_{>s} = \{ x \in X \mid r(x) > s \}$ $\mathcal{F}_{>s}$

$X_{\leq s} = \{ x \in X \mid r(x) \leq s \}$ $\mathcal{F}_{\leq s}$ induced forests

Inheritance Lemma:

$\mathcal{F}$ rank forest with maximum rank r and node set X

$s \in \mathbb{N}$: $X_{>s} = \{ x \in X \mid r(x) > s \}$ $\quad \mathcal{F}_{>s}$

$X_{\leq s} = \{ x \in X \mid r(x) \leq s \}$ $\quad \mathcal{F}_{\leq s}$ induced forests

i) $X_{\leq s}$, $X_{>s}$ is a dissection for $\mathcal{F}$

ii) $\mathcal{F}_{\leq s}$ is a rank forest with maximum rank $\leq s$

iii) $\mathcal{F}_{>s}$ is a rank forest with maximum rank $\leq r-s-1$

iv) $|X_{>s}| \leq |X| / 2^{s+1}$

Inheritance Lemma:

$\mathcal{F}$ rank forest with maximum rank r and node set X

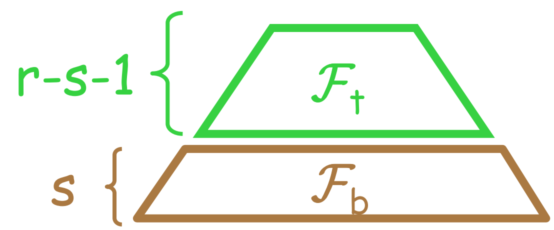
s$\in\mathbb{N}$: $X_{>s} = \{ x\in X \mid r(x)>s \}$ $\mathcal{F}_{>s}$ induced forests

$X_{\leq s} = \{ x\in X \mid r(x)\leq s \}$ $\mathcal{F}_{\leq s}$

i) $X_{\leq s}$, $X_{>s}$ is a dissection for $\mathcal{F}$

ii) $\mathcal{F}_{\leq s}$ is a rank forest with maximum rank $\leq$ s

iii) $\mathcal{F}_{>s}$ is a rank forest with maximum rank $\leq$ r-s-1



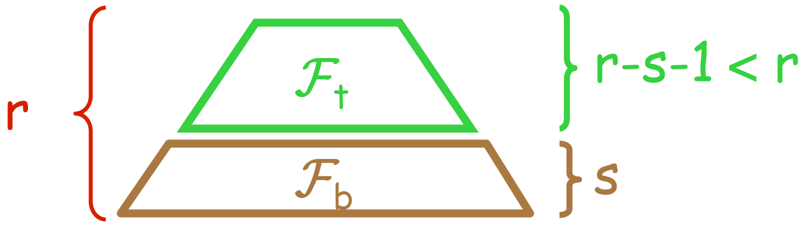
$f(m,n,r)$ = maximum cost of any compression sequence C, with $|C|=m$, in rank forest $\mathcal{F}$ with n nodes and maximum rank r.

$f(m,n,r)$ = maximum cost of any compression sequence C, with $|C|=m$, in rank forest $\mathcal{F}$ with n nodes and maximum rank r.

Trivial bounds:

$$f(m,n,r) \leq (r-1)\cdot n$$

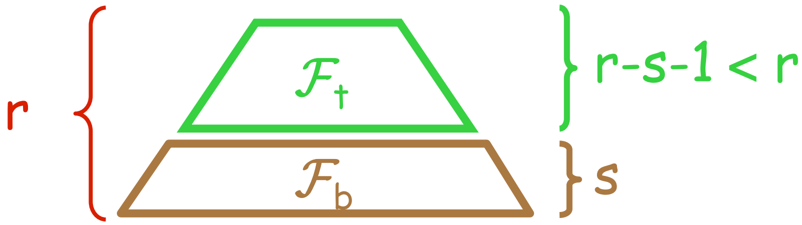
$$f(m,n,r) \leq (r-1)\cdot m$$



$|X_{>s}| = n_t$ $\qquad$ $|\mathcal{C}_t| = m_t$

$|X_{\leq s}| = n_b = n - n_t$ $\qquad$ $|\mathcal{C}_b| = m_b$

$$\text{cost}(\mathcal{C}) \leq \text{cost}(\mathcal{C}_t) + \text{cost}(\mathcal{C}_b) + |X_b| - \#\text{rts}(\mathcal{F}_b) + |\mathcal{C}_t|$$

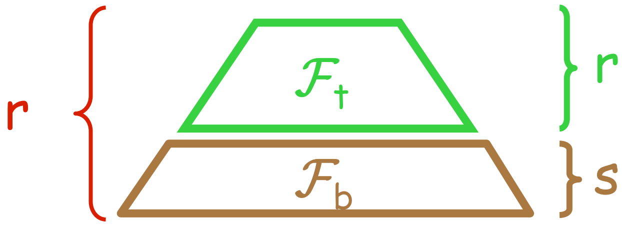


$|X_{>s}| = n_t$ $\quad |C_t| = m_t$

$|X_{\leq s}| = n_b = n - n_t$ $\quad |C_b| = m_b$

$$\text{cost}(C) \leq \text{cost}(C_t) + \text{cost}(C_b) + |X_b| - \#\text{rts}(\mathcal{F}_b) + |C_t|$$

$$\leq f(m_t, n_t, r-s-1) +$$

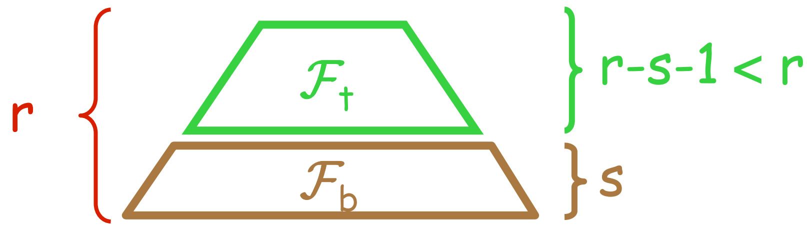


$|X_{>s}| = n_t$ $\quad |\mathcal{C}_t| = m_t$

$|X_{\leq s}| = n_b = n - n_t$ $\quad |\mathcal{C}_b| = m_b$

$$\text{cost}(\mathcal{C}) \leq \text{cost}(\mathcal{C}_t) + \text{cost}(\mathcal{C}_b) + |X_b| - \#\text{rts}(\mathcal{F}_b) + |\mathcal{C}_t|$$

$$\leq f(m_t, n_t, r-s-1) + f(m_b, n_b, s) +$$

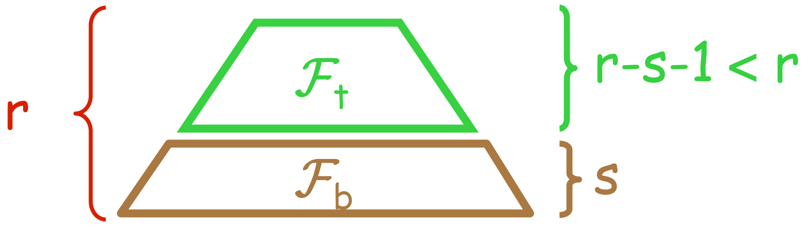


$|X_{>s}| = n_t$ $\quad |\mathcal{C}_t| = m_t$

$|X_{\leq s}| = n_b = n - n_t$ $\quad |\mathcal{C}_b| = m_b$

$$\text{cost}(\mathcal{C}) \leq \text{cost}(\mathcal{C}_t) + \text{cost}(\mathcal{C}_b) + |X_b| - \#\text{rts}(\mathcal{F}_b) + |\mathcal{C}_t|$$

$$\leq f(m_t, n_t, r-s-1) + f(m_b, n_b, s) + n - n_t -$$

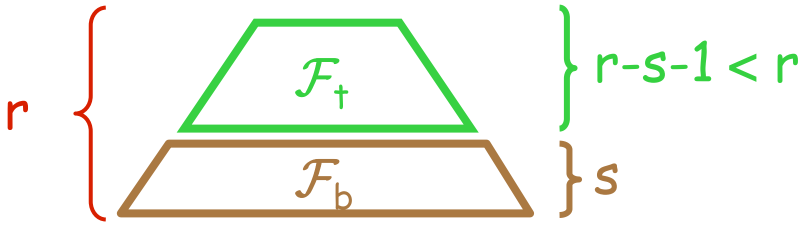


$|X_{>s}| = n_t$ $\quad$ $|C_t| = m_t$

$|X_{\leq s}| = n_b = n - n_t$ $\quad$ $|C_b| = m_b$

$$\text{cost}(\,C\,) \leq \text{cost}(\,C_t\,) + \text{cost}(\,C_b\,) + |X_b| - \#\text{rts}(\mathcal{F}_b) + |C_t|$$

$$\leq f(m_t, n_t, r-s-1) + f(m_b, n_b, s) + n - n_t - (s+1)\cdot n_t +$$



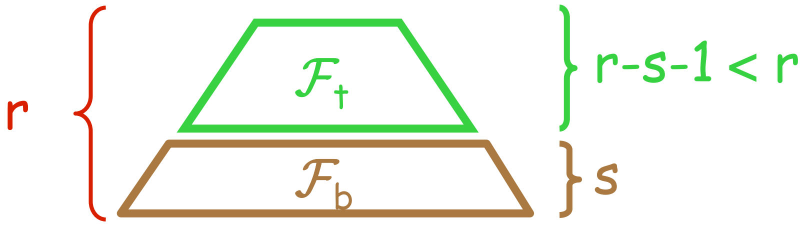
$|X_{>s}| = n_t$ $\quad$ $|C_t| = m_t$

$|X_{\leq s}| = n_b = n - n_t$ $\quad$ $|C_b| = m_b$

$$\text{cost}(\,C\,) \leq \text{cost}(\,C_t\,) + \text{cost}(\,C_b\,) + |X_b| - \#\text{rts}(\mathcal{F}_b) + |C_t|$$

$$\leq f(m_t, n_t, r-s-1) + f(m_b, n_b, s) + n - n_t - (s+1)\cdot n_t +$$

Each node in $\mathcal{F}_t$ has at least s+1 children in $\mathcal{F}_b$, and they must all be different roots of $\mathcal{F}_b$.



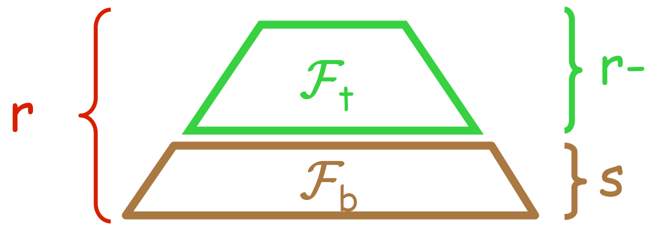
$|X_{>s}| = n_t$ $\quad |C_t| = m_t$

$|X_{\leq s}| = n_b = n-n_t$ $\quad |C_b| = m_b$

$$\text{cost}(C) \leq \text{cost}(C_t) + \text{cost}(C_b) + |X_b| - \#\text{rts}(\mathcal{F}_b) + |C_t|$$

$$\leq f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n-n_t - (s+1)\cdot n_t + m_t$$

Each node in $\mathcal{F}_t$ has at least $s+1$ children in $\mathcal{F}_b$, and they must all be different roots of $\mathcal{F}_b$.



$|X_{>s}| = n_t$ $\quad |C_t| = m_t$

$|X_{\leq s}| = n_b = n - n_t$ $\quad |C_b| = m_b$

$$\text{cost}(C) \leq \text{cost}(C_t) + \text{cost}(C_b) + |X_b| - \#\text{rts}(\mathcal{F}_b) + |C_t|$$

$$\leq f(m_t, n_t, r-s-1) + f(m_b, n_b, s) + n - n_t - (s+1)\cdot n_t + m_t$$

Each node in $\mathcal{F}_t$ has at least $s+1$ children in $\mathcal{F}_b$, and they must all be different roots of $\mathcal{F}_b$.

$$f(m,n,r) \leq f(m_t, n_t, r-s-1) + f(m_b, n_b, s) + n - (s+2)\cdot n_t + m_t$$

$$f(m,n,r) \leq f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n - (s+2)\cdot n_t + m_t$$

$$n_t + n_b = n$$

$$m_t + m_b \leq m$$

$$0 \leq s < r$$

$$f(m,n,r) \le f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n - (s+2)\cdot n_t + m_t$$

$$n_t + n_b = n$$
$$m_t + m_b \le m$$
$$0 \le s < r$$

Assume: $f(\ ,\nu,\rho) \le k\cdot \ + \nu\cdot g(\rho)$

$$\boxed{f(m,n,r) \le f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n - (s+2)\cdot n_t + m_t}$$

$$n_t + n_b = n \qquad 0 \le s < r$$

$$m_t + m_b \le m$$

Assume: $f(\ ,\nu,\rho) \le k\cdot\ \ + \nu\cdot g(\rho)$

$$f(m,n,r) \le k\cdot m_t + n_t\cdot g(r-s-1) + f(m_b,n_b,s) + n - (s+2)\cdot n_t + m_t$$

$$\le k\cdot m_t + n_t\cdot g(r) + f(m_b,n_b,s) + n - s\cdot n_t + m_t$$

$$\boxed{f(m,n,r) \le f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n - (s+2)\cdot n_t + m_t}$$

$$n_t + n_b = n \qquad 0 \le s < r$$

$$m_t + m_b \le m$$

Assume: $f(\ ,\nu,\rho) \le k\cdot\ \ + \nu\cdot g(\rho)$

$$f(m,n,r) \le k\cdot m_t + n_t\cdot g(r-s-1) + f(m_b,n_b,s) + n - (s+2)\cdot n_t + m_t$$

$$\le k\cdot m_t + n_t\cdot g(r) + f(m_b,n_b,s) + n - s\cdot n_t + m_t$$

choose $s = g(r)$

$$f(m,n,r) \le f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n - (s+2)\cdot n_t + m_t$$

$$n_t + n_b = n \qquad 0 \le s < r$$

$$m_t + m_b \le m$$

Assume: $f(\ ,\nu,\rho) \le k\cdot \ + \nu\cdot g(\rho)$

$$f(m,n,r) \le k\cdot m_t + n_t\cdot g(r-s-1) + f(m_b,n_b,s) + n - (s+2)\cdot n_t + m_t$$

$$\le k\cdot m_t + n_t\cdot g(r) + f(m_b,n_b,s) + n - s\cdot n_t + m_t$$

choose $s = g(r)$

$$f(m,n,r) \le (k+1)\cdot m_t + f(m_b,n_b,s) + n$$

$$\le (k+1)\cdot m_t + f(m_b,n,s) + n$$

$$s = g(r)$$

$$f(m,n,r) \leq (k+1) \cdot m_t + f(m_b,n,s) + n$$

$$s = g(r)$$

$$f(m,n,r) \leq (k+1)\cdot m_t + f(m_b,n,s) + n \quad \Big| \quad -(k+1)\cdot(m_b+m_t)$$

$$s = g(r)$$

$$f(m,n,r) \leq (k+1)\cdot m_t + f(m_b,n,s) + n \quad \Bigg| \quad -(k+1)\cdot\overbrace{(m_b+m_t)}^{m}$$

$s = g(r)$

$$f(m,n,r) \leq (k+1)\cdot m_t + f(m_b,n,s) + n \quad \Big| \quad -(k+1)\cdot\overbrace{(m_b+m_t)}^{m}$$

$$f(m,n,r) - (k+1)\cdot m \leq f(m_b,n,s) - (k+1)\cdot m_b + n$$

$$s = g(r)$$

$$f(m,n,r) \leq (k+1)\cdot m_t + f(m_b,n,s) + n \quad \Bigg| \quad -(k+1)\cdot\overbrace{(m_b+m_t)}^{m}$$

$$f(m,n,r) - (k+1)\cdot m \leq f(m_b,n,s) - (k+1)\cdot m_b + n$$

$$\phi(m,n,r) \leq \phi(m_b,n,g(r)) + n$$

$$s = g(r)$$

$$f(m,n,r) \le (k+1)\cdot m_t + f(m_b,n,s) + n \quad \Big| \quad -(k+1)\cdot\overbrace{(m_b+m_t)}^{m}$$

$$f(m,n,r) - (k+1)\cdot m \le f(m_b,n,s) - (k+1)\cdot m_b + n$$

$$\phi(m,n,r) \le \phi(m_b,n,g(r)) + n$$

$$\phi(m,n,r) \le n \cdot g^*(r)$$

$$s = g(r)$$

$$f(m,n,r) \le (k+1)\cdot m_t + f(m_b,n,s) + n \quad \Big| \quad -(k+1)\cdot\overbrace{(m_b+m_t)}^{m}$$

$$f(m,n,r) - (k+1)\cdot m \le f(m_b,n,s) - (k+1)\cdot m_b + n$$

$$\phi(m,n,r) \le \phi(m_b,n,g(r)) + n$$

$$\phi(m,n,r) \le n \cdot g^*(r)$$

$$f(m,n,r) \le (k+1)\cdot m + n \cdot g^*(r)$$

Shifting Lemma:

If $f(m,n,r) \leq k \cdot m + n \cdot g(r)$

then also $f(m,n,r) \leq (k+1) \cdot m + n \cdot g^*(r)$

Shifting Lemma:

If $f(m,n,r) \le k \cdot m + n \cdot g(r)$

then also $f(m,n,r) \le (k+1) \cdot m + n \cdot g^{*}(r)$

Shifting Corollary:

If $f(m,n,r) \le k \cdot m + n \cdot g(r)$

then also $f(m,n,r) \le (k+i) \cdot m + n \cdot g^{\overbrace{**\cdots*}^{i}}(r)$

for any $i \ge 0$

If $f(m,n,r) \leq k \cdot m + n \cdot g(r)$

then also $f(m,n,r) \leq (k+i) \cdot m + n \cdot g^{\overbrace{**\cdots*}^{i}}(r)$

for any $i \geq 0$

If $f(m,n,r) \le k \cdot m + n \cdot g(r)$

then also $f(m,n,r) \le (k+i) \cdot m + n \cdot g^{\overbrace{**\cdots*}^{i}}(r)$

for any $i \ge 0$

Trivial bound: $f(m,n,r) \le n \cdot (r-1)$

If $f(m,n,r) \leq k \cdot m + n \cdot g(r)$

then also $f(m,n,r) \leq (k+i) \cdot m + n \cdot g^{\overbrace{**\cdots*}^{i}}(r)$

for any $i \geq 0$

Trivial bound: $f(m,n,r) \leq n \cdot (r-1)$

$= 0 \cdot m + n \cdot (r-1)$

If $f(m,n,r) \leq k \cdot m + n \cdot g(r)$

then also $f(m,n,r) \leq (k+i) \cdot m + n \cdot g^{\overbrace{** \cdots *}^{i}}(r)$

for any $i \geq 0$

Trivial bound: $f(m,n,r) \leq n \cdot (r-1)$

$= 0 \cdot m + n \cdot (r-1)$

$g(r) = r-1$

$g^*(r) = r-1$

$$f(m,n,r) \leq f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n - (s+2)\cdot n_t + m_t$$

$$n_t + n_b = n$$

$$m_t + m_b \leq m$$

$$0 \leq s < r$$

$$f(m,n,r) \le f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n - (s+2)\cdot n_t + m_t$$

$$n_t + n_b = n \qquad 0 \le s < r$$

$$m_t + m_b \le m$$

Trivial bound: $f(\ ,\nu,\rho) \le \nu\cdot(\rho-1)$

$$f(m,n,r) \le f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n - (s+2)\cdot n_t + m_t$$

$$n_t + n_b = n \qquad 0 \le s < r$$
$$m_t + m_b \le m$$

Trivial bound: $f(\ ,\nu,\rho) \le \nu\cdot(\rho-1)$

$$f(m,n,r) \le n_t\cdot(r-s-2) + f(m_b,n_b,s) + n - (s+2)\cdot n_t + m_t$$

$$f(m,n,r) \leq f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n - (s+2)\cdot n_t + m_t$$

$$n_t + n_b = n \qquad 0 \leq s < r$$
$$m_t + m_b \leq m$$

Trivial bound: $f(\ ,\nu,\rho) \leq \nu\cdot(\rho-1)$

$$f(m,n,r) \leq n_t\cdot(r-s-2) + f(m_b,n_b,s) + n - (s+2)\cdot n_t + m_t$$
$$\leq n_t\cdot(r-2s-4) + f(m_b,n_b,s) + n + m_t$$

$$\boxed{f(m,n,r) \leq f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n - (s+2)\cdot n_t + m_t}$$

$$n_t + n_b = n \qquad 0 \leq s < r$$
$$m_t + m_b \leq m$$

Trivial bound: $f(\ ,\nu,\rho) \leq \nu\cdot(\rho-1)$

$$f(m,n,r) \leq n_t\cdot(r-s-2) + f(m_b,n_b,s) + n - (s+2)\cdot n_t + m_t$$
$$\leq n_t\cdot(r-2s-4) + f(m_b,n_b,s) + n + m_t$$

set $s = \lfloor r/2 \rfloor$

$$f(m,n,r) \leq f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n - (s+2)\cdot n_t + m_t$$

$$n_t + n_b = n \qquad 0 \leq s < r$$

$$m_t + m_b \leq m$$

Trivial bound: $f(\ ,\nu,\rho) \leq \nu\cdot(\rho-1)$

$$f(m,n,r) \leq n_t\cdot(r-s-2) + f(m_b,n_b,s) + n - (s+2)\cdot n_t + m_t$$

$$\leq n_t\cdot(r-2s-4) + f(m_b,n_b,s) + n + m_t$$

set $s = \lfloor r/2 \rfloor$

$$f(m,n,r) \leq f(m_b,n_b,r/2) + n + m_t$$

$$f(m,n,r) \le f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n - (s+2)\cdot n_t + m_t$$

$$n_t + n_b = n \qquad 0 \le s < r$$
$$m_t + m_b \le m$$

Trivial bound: $f(\ ,\nu,\rho) \le \nu\cdot(\rho-1)$

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

set $s = \lfloor r/2 \rfloor$

$$f(m,n,r) \le f(m_b,n_b,r/2) + n + m_t$$

$$f(m,n,r) - m \le f(m_b,n_b,r/2) - m_b + n$$

$$f(m,n,r) \le f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n - (s+2)\cdot n_t + m_t$$

$$n_t + n_b = n \qquad 0 \le s < r$$
$$m_t + m_b \le m$$

Trivial bound: $f(\ ,\nu,\rho) \le \nu\cdot(\rho-1)$

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

set $s = \lfloor r/2 \rfloor$

$$f(m,n,r) \le f(m_b,n_b,r/2) + n + m_t$$

$$f(m,n,r) - m \le f(m_b,n_b,r/2) - m_b + n$$

$$\boxed{f(m,n,r) \le m + n\cdot\log r}$$

If $f(m,n,r) \leq k \cdot m + n \cdot g(r)$

then also $f(m,n,r) \leq (k+i) \cdot m + n \cdot g^{\overbrace{**\cdots*}^{i}}(r)$

for any $i \geq 0$

If $f(m,n,r) \le k \cdot m + n \cdot g(r)$

then also $f(m,n,r) \le (k+i) \cdot m + n \cdot g^{\overbrace{** \ldots *}^{i}}(r)$

for any $i \ge 0$

We know bound: $f(m,n,r) \le m + n \cdot \log r$

If $f(m,n,r) \le k \cdot m + n \cdot g(r)$

then also $f(m,n,r) \le (k+i) \cdot m + n \cdot g^{\overbrace{**\cdots*}^{i}}(r)$

for any $i \ge 0$

We know bound: $f(m,n,r) \le m + n \cdot \log r$

Therefore for any $i \ge 0$:

$$f(m,n,r) \le (i+1) \cdot m + n \cdot \log^{\overbrace{**\cdots*}^{i}}(r)$$

For any $i \geq 0$: $f(m,n,r) \leq (i+1) \cdot m + n \cdot \log^{\overbrace{** \cdots *}^{i}}(r)$

For any $i \geq 0$: $f(m,n,r) \leq (i+1) \cdot m + n \cdot \log^{\overbrace{** \ldots *}^{i}}(r)$

Choice of i :

For any $i \geq 0$: $f(m,n,r) \leq (i+1) \cdot m + n \cdot \log^{\overbrace{** \cdots *}^{i}}(r)$

Choice of i :

Define $\alpha(r) = \min\{\, i \mid \log^{\overbrace{** \cdots *}^{i}}(r) \leq i \,\}$

For any $i \geq 0$: $f(m,n,r) \leq (i+1)\cdot m + n\cdot \log^{\overbrace{**\cdots*}^{i}}(r)$

Choice of i :

Define $\alpha(r) = \min\{ i \mid \log^{\overbrace{**\cdots*}^{i}}(r) \leq i \}$

$$f(m,n,r) \leq (m+n)(1+\alpha(r))$$

For any $i \geq 0$: $f(m,n,r) \leq (i+1) \cdot m + n \cdot \log^{\overbrace{** \cdots *}^{i}}(r)$

Choice of i :

Define $\alpha(r) = \min\{ i \mid \log^{\overbrace{** \cdots *}^{i}}(r) \leq i \}$

$$f(m,n,r) \leq (m+n)(1+\alpha(r))$$

$$\leq (m+n)(1+\alpha(\log n))$$

For any $i \geq 0$: $f(m,n,r) \leq (i+1)\cdot m + n\cdot \log^{\overbrace{**\cdots*}^{i}}(r)$

Choice of i :

Define $\alpha(m,n,r) = \min\{\, i \mid \log^{\overbrace{**\cdots*}^{i}}(r) \leq m/n \,\}$

For any $i \geq 0$: $f(m,n,r) \leq (i+1)\cdot m + n\cdot \log^{\overbrace{**\cdots*}^{i}}(r)$

Choice of i :

Define $\alpha(m,n,r) = \min\{\, i \mid \log^{\overbrace{**\cdots*}^{i}}(r) \leq m/n \,\}$

$$f(m,n,r) \leq m(2+\alpha(m,n,r))$$

For any $i \geq 0$: $f(m,n,r) \leq (i+1)\cdot m + n\cdot \log^{\overbrace{**\cdots*}^{i}}(r)$

Choice of i :

Define $\alpha(m,n,r) = \min\{\, i \mid \log^{\overbrace{**\cdots*}^{i}}(r) \leq m/n \,\}$

$$f(m,n,r) \leq m(2+\alpha(m,n,r))$$

Define $\alpha(m,n) = \min\{\, i \mid \log^{\overbrace{**\cdots*}^{i}}(\log n) \leq m/n \,\}$

For any $i \geq 0$: $f(m,n,r) \leq (i+1)\cdot m + n\cdot \log^{\overbrace{**\cdots*}^{i}}(r)$

Choice of i :

Define $\alpha(m,n,r) = \min\{\, i \mid \log^{\overbrace{**\cdots*}^{i}}(r) \leq m/n \,\}$

$$f(m,n,r) \leq m(2+\alpha(m,n,r))$$

Define $\alpha(m,n) = \min\{\, i \mid \log^{\overbrace{**\cdots*}^{i}}(\log n) \leq m/n \,\}$

$$f(m,n,r) \leq m(2+\alpha(m,n))$$

Good bounds for $f(m,n,r)$ when r is small

$f(m,n,0) = 0$

$f(m,n,1) = 0$

$f(m,n,2) \leq m$

$$f(m,n,r) \leq f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n - (s+2)\cdot n_t + m_t$$

$$n_t + n_b = n$$

$$m_t + m_b \leq m$$

$$0 \leq s < r$$

$$f(m,n,r) \leq f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n - (s+2)\cdot n_t + m_t$$

$$n_t + n_b = n$$
$$m_t + m_b \leq m$$

$$0 \leq s < r$$

choose $s = 2$:

$$f(m,n,r) \leq f(m_t,n_t,r-3) + f(m_b,n_b,2) + n - 4\cdot n_t + m_t$$

$$f(m,n,r) \leq f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n - (s+2)\cdot n_t + m_t$$

$$n_t + n_b = n \qquad 0 \leq s < r$$
$$m_t + m_b \leq m$$

choose $s = 2$:

$$f(m,n,r) \leq \underbrace{f(m_t,n_t,r-3)}_{\leq n_t(r-4)} + \underbrace{f(m_b,n_b,2)}_{\leq m_b} + n - 4\cdot n_t + m_t$$

$$\boxed{f(m,n,r) \le f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n - (s+2)\cdot n_t + m_t}$$

$$n_t + n_b = n \qquad 0 \le s < r$$

$$m_t + m_b \le m$$

choose $s = 2$:

$$f(m,n,r) \le \underbrace{f(m_t,n_t,r-3)}_{\le n_t(r-4)} + \underbrace{f(m_b,n_b,2)}_{\le m_b} + n - 4\cdot n_t + m_t$$

$$f(m,n,r) \le m + n + (r-8)n_t$$

$$f(m,n,r) \leq f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n - (s+2)\cdot n_t + m_t$$

$$n_t + n_b = n \qquad 0 \leq s < r$$
$$m_t + m_b \leq m$$

choose $s = 2$:

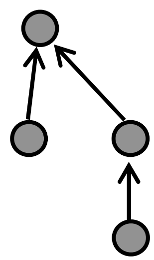
$$f(m,n,r) \leq \underbrace{f(m_t,n_t,r-3)}_{\leq n_t(r-4)} + \underbrace{f(m_b,n_b,2)}_{\leq m_b} + n - 4\cdot n_t + m_t$$

$$f(m,n,r) \leq m + n + (r-8)n_t$$

$f(m,n,r) \leq m + n$ for $r \leq 8$, i.e. for $n < 512$

$$f(m,n,r) \leq n(r-1)$$

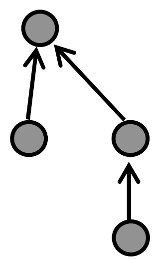
$f(m,n,r) \leq n(r-1)$



n=4, r=1: bound 4(2-1) = 4

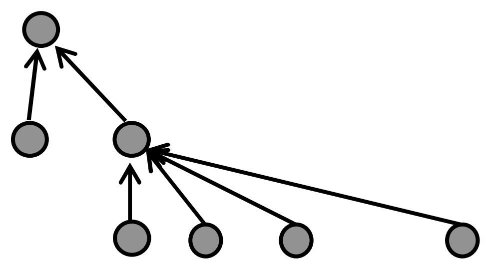
actual bound: 1

$f(m,n,r) \leq n(r-1)$



$n=4$, $r=1$: bound $4(2-1) = 4$

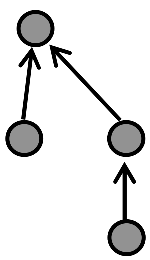
actual bound: 1



bound $n(2-1) = n$

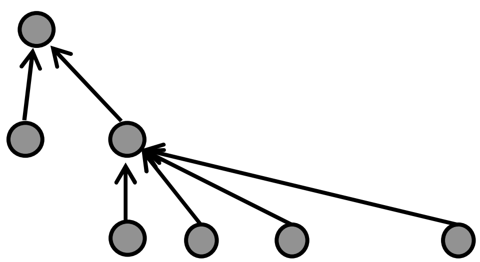
actual bound: $n-3$

$f(m,n,r) \leq n(r-1)$



$n=4,\ r=1$: bound $4(2-1) = 4$

actual bound: 1



bound $n(2-1) = n$

actual bound: $n-3$

$f(m,n,2) \leq n-3$
$f(m,n,3) \leq 2n-11$

$f(m,n,r) \leq (r-1)n-(r2^{r-1}-1)$

“trivial” bound for rank forest with n nodes and with max rank r, where there are R_i trees of rank i, for $0 \leq i \leq r$.

$$n(r-1) - \sum_{0 \leq i \leq r} R_i \cdot C(r,i)$$

$$C(r,i) = 2^{i-1}(2r-i)-1$$

$$f(m,n,r) \leq f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n - (s+2)\cdot n_t + m_t$$

$$n_t + n_b = n \qquad 0 \leq s < r$$

$$m_t + m_b \leq m$$

choose $s = 8$:

$$f(m,n,r) \leq f(m_t,n_t,r-9) + f(m_b,n_b,8) + n - 10\cdot n_t + m_t$$

$$\boxed{f(m,n,r) \le f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n - (s+2)\cdot n_t + m_t}$$

$$n_t + n_b = n \qquad 0 \le s < r$$

$$m_t + m_b \le m$$

choose $s = 8$:

$$f(m,n,r) \le \underbrace{f(m_t,n_t,r-3)}_{\le\; n_t(r-4) - \sum R_i \cdot C(r-3,i)} + \underbrace{f(m_b,n_b,8)}_{\le\; m_b+n_b} + n - 10\cdot n_t + m_t$$

$$\boxed{f(m,n,r) \le f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n - (s+2)\cdot n_t + m_t}$$

$$n_t + n_b = n \qquad 0 \le s < r$$
$$m_t + m_b \le m$$

choose $s = 8$:

$$f(m,n,r) \le \underbrace{f(m_t,n_t,r-3)}_{\le\, n_t(r-4) - \sum R_i \cdot C(r-3,i)} + \underbrace{f(m_b,n_b,8)}_{\le\, m_b+n_b} + n - 10\cdot n_t + m_t$$

$R_i \ge n_t$ for each i

$$f(m,n,r) \le f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n - (s+2)\cdot n_t + m_t$$

$$n_t + n_b = n \qquad 0 \le s < r$$
$$m_t + m_b \le m$$

choose $s = 8$:

$$f(m,n,r) \le \underbrace{f(m_t,n_t,r-3)}_{\le\; n_t(r-4) - \sum R_i \cdot C(r-3,i)} + \underbrace{f(m_b,n_b,8)}_{\le\; m_b+n_b} + n - 10\cdot n_t + m_t$$

$R_i \ge n_t$ for each i

$$f(m,n,r) \le m + 2n + n_t(r-202)$$

$$f(m,n,r) \leq f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n - (s+2)\cdot n_t + m_t$$

$$n_t + n_b = n \qquad 0 \leq s < r$$
$$m_t + m_b \leq m$$

choose $s = 8$:

$$f(m,n,r) \leq \underbrace{f(m_t,n_t,r-3)}_{\leq\ n_t(r-4) - \sum R_i \cdot C(r-3,i)} + \underbrace{f(m_b,n_b,8)}_{\leq\ m_b+n_b} + n - 10\cdot n_t + m_t$$

$R_i \geq n_t$ for each i

$$f(m,n,r) \leq m + 2n + n_t(r-202)$$

$$f(m,n,r) \leq m + 2n \quad \text{for } r \leq 202\text{, i.e. for } n < 2^{203}$$

This approach yields

$$f(m,n,r) \leq m + n(\log^* r - c)$$

Similar proof for $O(\, m \cdot \alpha(m,n) + n \,)$ bound also works for

linking by weight and path compression

Similar proof for $O(\, m \cdot \alpha(m,n) + n \,)$ bound also works for

linking by weight and path compression

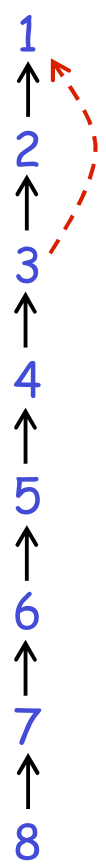
linking by rank and generalized path compaction

Heuristic 2': Path compaction

when performing a Find(x) operation make "all" nodes in the "findpath" child of some node further up.

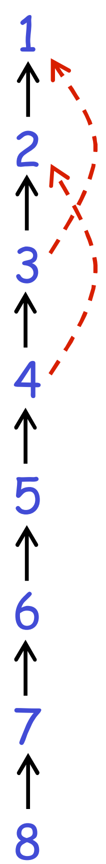
Heuristic 2': Path compaction

when performing a Find(x) operation make "all" nodes in the "findpath" child of some node further up.



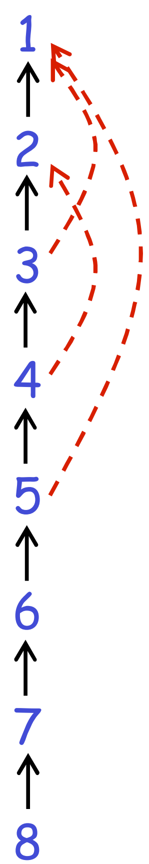
Heuristic 2': Path compaction

when performing a Find(x) operation make "all" nodes in the "findpath" child of some node further up.



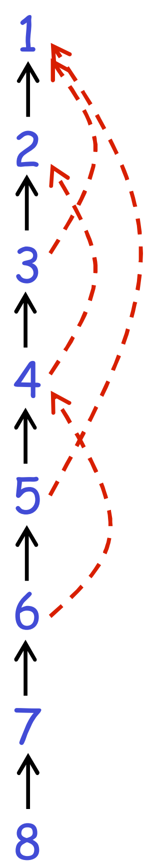
Heuristic 2': Path compaction

when performing a Find(x) operation make "all" nodes in the "findpath" child of some node further up.



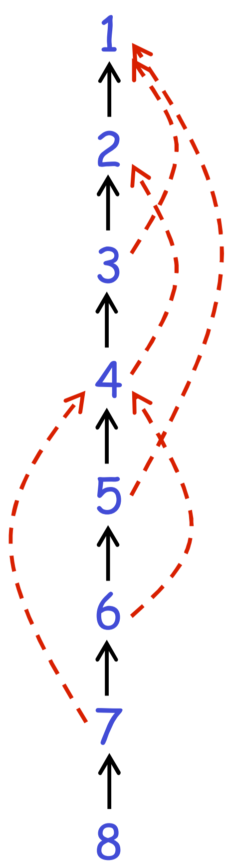
Heuristic 2': Path compaction

when performing a Find(x) operation make "all" nodes in the "findpath" child of some node further up.



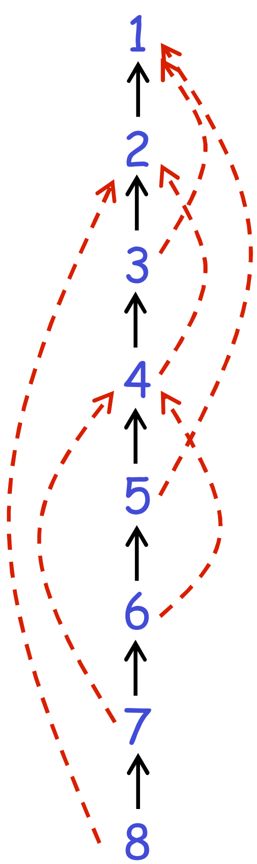
Heuristic 2': Path compaction

when performing a Find(x) operation make "all" nodes in the "findpath" child of some node further up.



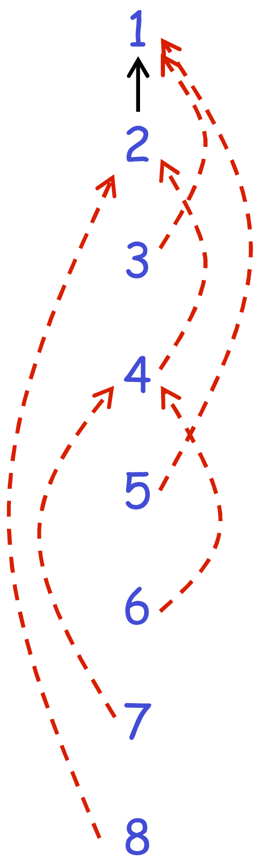
Heuristic 2': Path compaction

when performing a Find(x) operation make "all" nodes in the "findpath" child of some node further up.



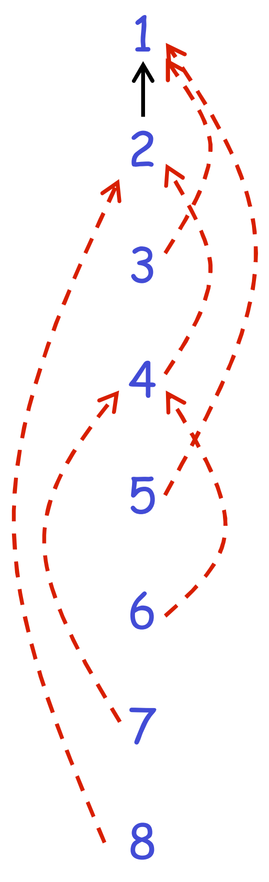
Heuristic 2': Path compaction

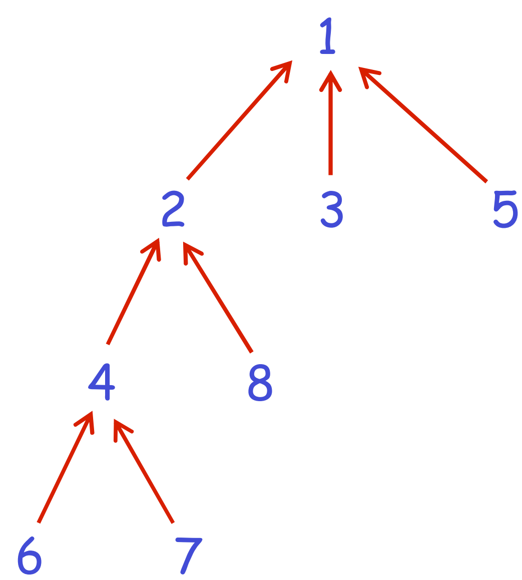
when performing a Find(x) operation make "all" nodes in the "findpath" child of some node further up.



Heuristic 2': Path compaction

when performing a Find(x) operation make "all" nodes in the "findpath" child of some node further up.





Main Lemma:

$\mathcal{C}$ ... sequence of compress operations on $\mathcal{F}$ with node set X

X_t, X_b dissection for $\mathcal{F}$ inducing subforests $\mathcal{F}_t$, $\mathcal{F}_b$

$\Rightarrow$ $\exists$ compression sequences $\mathcal{C}_b$ for $\mathcal{F}_b$ and $\mathcal{C}_t$ for $\mathcal{F}_t$ with

$$|\mathcal{C}_b| + |\mathcal{C}_t| \leq |\mathcal{C}|$$

and

$$\text{cost}(\mathcal{C}) \leq \text{cost}(\mathcal{C}_b) + \text{cost}(\mathcal{C}_t) + |X_b| + |\mathcal{C}_t|$$

Main Lemma:

$\mathcal{C}$ ... sequence of compaction operations on $\mathcal{F}$ with node set X

X_t, X_b dissection for $\mathcal{F}$ inducing subforests $\mathcal{F}_t$, $\mathcal{F}_b$

$\Rightarrow$ $\exists$ compaction sequences $\mathcal{C}_b$ for $\mathcal{F}_b$ and $\mathcal{C}_t$ for $\mathcal{F}_t$ with

$$|\mathcal{C}_b| + |\mathcal{C}_t| \leq |\mathcal{C}|$$

and

$$\mathrm{cost}(\mathcal{C}) \leq \mathrm{cost}(\mathcal{C}_b) + \mathrm{cost}(\mathcal{C}_t) + |X_b| + |\mathcal{C}_t| + \sum_{v \in X_b} \mathrm{height}(v)$$

Proof:

$$|C_b| + |C_t| \le |C|$$

compression paths from C

case 1: $x \dashrightarrow y$ — $x \dashrightarrow y$ into C_t

case 2: $x \dashrightarrow y$ — $x \dashrightarrow y$ into C_b

case 3: $x \dashrightarrow x' \dashrightarrow y$ — $x' \dashrightarrow y$ into C_t; $x \dashrightarrow \infty$ into C_b

Proof:

$$|C_b| + |C_t| \leq |C|$$

compaction paths from C

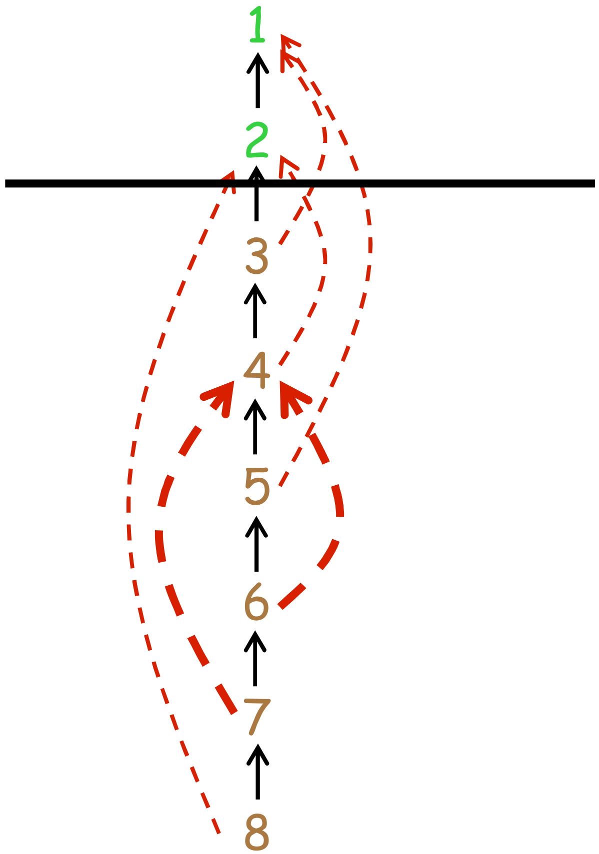
case 1: $x \to y$ — $x \to y$ into C_t

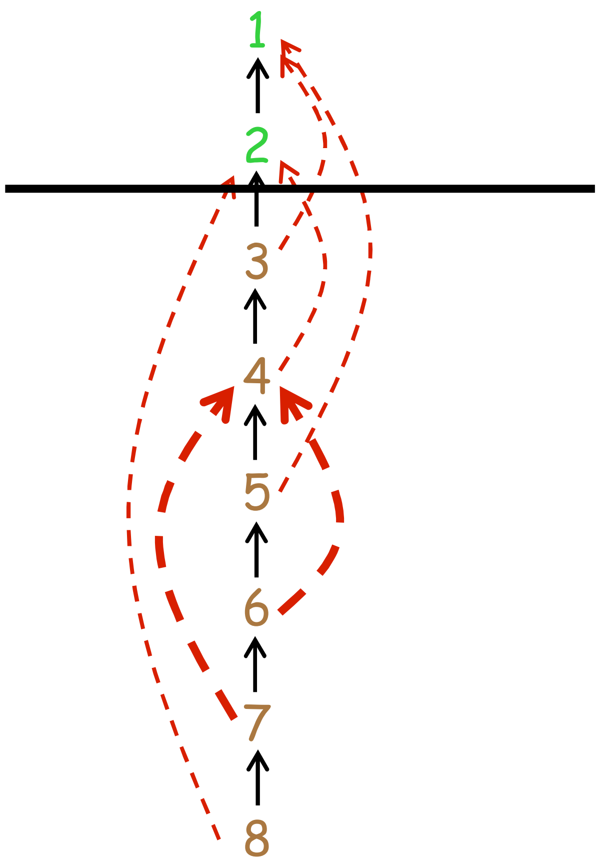
case 2: $x \to y$ — $x \to y$ into C_b

case 3: $x \to x' \to y$ — $x' \to y$ into C_t

$x \to \infty$ into C_b

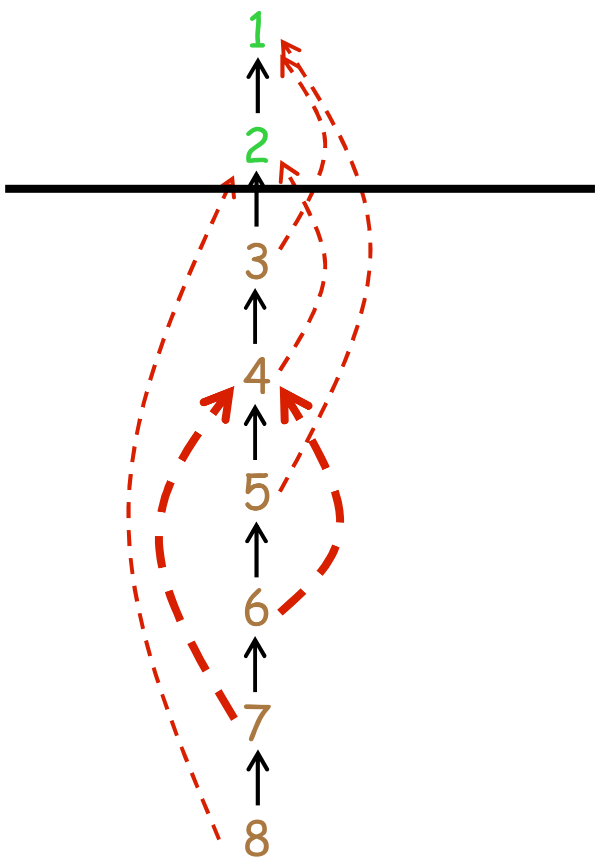
Compaction of path that crosses dissection boundary:





Compaction of path that crosses dissection boundary:

Charge getting a new brown parent to the topmost brown node v that gets a green parent for the first time.



Compaction of path that crosses dissection boundary:

Charge getting a new brown parent to the topmost brown node v that gets a green parent for the first time.

happens to v at most once;
v can be charged at most height(v)

Main Lemma:

$\mathcal{C}$ ... sequence of compaction operations on $\mathcal{F}$ with node set X

X_t, X_b dissection for $\mathcal{F}$ inducing subforests $\mathcal{F}_t$, $\mathcal{F}_b$

$\Rightarrow$ $\exists$ compaction sequences $\mathcal{C}_b$ for $\mathcal{F}_b$ and $\mathcal{C}_t$ for $\mathcal{F}_t$ with

$$|\mathcal{C}_b| + |\mathcal{C}_t| \leq |\mathcal{C}|$$

and

$$\text{cost}(\mathcal{C}) \leq \text{cost}(\mathcal{C}_b) + \text{cost}(\mathcal{C}_t) + |X_b| + |\mathcal{C}_t| + \sum_{v \in X_b} \text{height}(v)$$

$$\text{cost}(C) \leq \text{cost}(C_b) + \text{cost}(C_t) + |X_b| + |C_t| + \sum_{v \in X_b} \text{height}(v)$$

In a rank forest $\sum_{v \in X_b} \text{height}(v) \leq |X_b|$

$$\text{cost}(C) \leq \text{cost}(C_b) + \text{cost}(C_t) + |X_b| + |C_t| + \sum_{v \in X_b} \text{height}(v)$$

In a rank forest $\sum_{v \in X_b} \underbrace{\text{height}(v)}_{\leq \,\#\text{ children of } v} \leq |X_b|$

$$\text{cost}(C) \le \text{cost}(C_b) + \text{cost}(C_t) + |X_b| + |C_t| + \sum_{v \in X_b} \text{height}(v)$$

In a rank forest $\sum_{v \in X_b} \underbrace{\text{height}(v)}_{\le \text{ \# children of } v} \le |X_b|$

$$\text{cost}(C) \le \text{cost}(C_b) + \text{cost}(C_t) + 2 \cdot |X_b| + |C_t|$$

Corollary:

Any sequence of m Union, Find operations in a universe of n elements that uses linking by rank and path compaction takes time at most

$$O(\, m \cdot \alpha(m,n) + n \,)$$

Open problems:

- top-down approach for lower bounds