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 $\alpha$-male of algorithms research.
With the introduction of $\alpha(\ )$ into algorithm analysis Bob became the $\alpha$-male of algorithms research.
With the introduction of $\alpha(\ )$ into algorithm analysis Bob became the \textit{\textbf{\textalpha{}-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 \( \alpha(m, n) \) exist; for example, \( \log_2 n \) is sometimes replaced by \( n \), and the floor function is sometimes replaced by a ceiling.
Definition and properties

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

compress($x, y$)
General path compression in forest $\mathcal{F}$

compress($x, y$)
General path compression in forest $\mathcal{F}$

Definition:

$\text{compress}(x, y) = \# \text{ of nodes that get a new parent}$

Diagram:

[Diagram showing the process of path compression with nodes and arrows indicating parent-child relationships.]
Problem formulation

\( \mathcal{F} \) forest on node set \( X \)

\( C \) sequence of compress operations on \( \mathcal{F} \)

|\( C \)| = \# of true compress operations in \( C \)

\[
\text{cost}(C) = \sum(\text{cost of individual operations})
\]

How large can \( \text{cost}(C) \) 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 $\Rightarrow$ $X_t, X_b$ is dissection for $\mathcal{F}'$
Main Lemma:

$C$ ... sequence of operations on $F$ with node set $X$

$X_t, X_b$ dissection for $F$ inducing subforests $F_t, F_b$
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

$$\text{cost}(C) \leq \text{cost}(C_b) + \text{cost}(C_t) + |X_b| + |C_t|$$
Proof: 1) How to get $C_b$ and $C_+$ from $C$: 
Proof: 1) How to get $C_b$ and $C_t$ from $C$:

compression paths from $C$

case 1: $\uparrow \downarrow$ into $C_t$
Proof: 1) How to get $C_b$ and $C_+$ from $C$:

compression paths from $C$

\begin{align*}
\text{case 1:} & & Y & & Y & & \text{into } C_+ \\
& & \downarrow & & \downarrow & & \\
& & X & & X & & \\
\text{case 2:} & & Y & & Y & & \text{into } C_b \\
& & \downarrow & & \downarrow & & \\
& & X & & X & & 
\end{align*}
Proof: 1) How to get $C_b$ and $C_t$ from $C$:

compression paths from $C$

case 1: $Y$ $\downarrow$ $X$ $\downarrow$ $X$ into $C_t$

case 2: $Y$ $\downarrow$ $X$ into $C_b$

case 3: $Y$ $\downarrow$ $X'$ into $C_t$

$\infty$ into $C_b$
“rootpath compress”

compress(\(x, \infty\))
“rootpath compress”

\[
\text{compress}(x, \infty) = \text{# of nodes that get a new parent}
\]

\[= 0\]
Proof:

\[ |C_b| + |C_t| \leq |C| \]

compression paths from \( C \)

case 1:

\[ \text{into } C_t \]

case 2:

\[ \text{into } C_b \]

case 3:

\[ \text{into } C_b \]
\[ \text{cost}(C) \leq \text{cost}(C_b) + \text{cost}(C_t) + |X_b| + |C_t| \]

- **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:** accounted by \( |X_b| \)
- **Brown node gets new green parent:** accounted by \( |C_t| \)

- **For the first time**
- **Again**
\[ \text{cost}(C) \leq \text{cost}(C_b) + \text{cost}(C_t) + |X_b| + |C_t| \]

green node gets new green parent:

brown node gets new brown parent:

brown node gets new green parent: for the first time

brown node gets new green parent: again

accounted by \text{cost}(C_t)

accounted by \text{cost}(C_b)

accounted by \ |X_b| - \#\text{roots}(F_b)

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

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: accounted by \(|X_b| - \#\text{roots}(\mathcal{F}_b)|

brown node gets new green parent: for the first time

brown node gets new green parent: again accounted by \(|C_t|\)
Main Lemma':

\( C \) ... sequence of operations on \( \mathcal{F} \) with node set \( X \)
\( X_t, X_b \) dissection for \( \mathcal{F} \) inducing subforests \( \mathcal{F}_t, \mathcal{F}_b \)

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

\[ |C_b| + |C_t| \leq |C| \]

and

\[ \text{cost}(C) \leq \text{cost}(C_b) + \text{cost}(C_t) + |X_b| - \#\text{roots}(\mathcal{F}_b) + |C_t| \]
Path compression and union by rank

Def: \( F \) forest, \( x \) node in \( F \)
  \( r(x) = \) height of subtree rooted at \( x \)
  \( (r(leaf) = 0) \)

\( F \) 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 $\mathcal{F}$

\[ r(x) = \text{height of subtree rooted at } x \]

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

$\mathcal{F}$ 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 \( \mathcal{F} \)
\[
\begin{align*}
r(x) &= \text{height of subtree rooted at } x \\
&= 0 \quad \text{(leaf)}
\end{align*}
\]

\( \mathcal{F} \) 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.
Path compression and union by rank

**Def:** \( \mathcal{F} \) forest, \( x \) node in \( \mathcal{F} \)

\[ r(x) = \text{height of subtree rooted at } x \]

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

\( \mathcal{F} \) 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 $X$

$s \in \mathbb{N}$:

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

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

induced forests
Inheritance Lemma:

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

\[
\begin{align*}
  s \in \mathbb{N}: & \quad X_{>s} = \{ x \in X \mid r(x) > s \} & \mathcal{F}_{>s} & \text{induced forests} \\
  X_{\leq s} = \{ x \in X \mid r(x) \leq s \} & \mathcal{F}_{\leq s}
\end{align*}
\]

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:

\( F \) rank forest with maximum rank \( r \) and node set \( X \)

\( s \in \mathbb{N} \):

\[ X_{>s} = \{ x \in X \mid r(x) > s \} \quad F_{>s} \quad \text{induced forests} \]

\[ X_{\leq s} = \{ x \in X \mid r(x) \leq s \} \quad F_{\leq s} \]

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

ii) \( F_{\leq s} \) is a rank forest with maximum rank \( \leq s \)

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

**Trivial bounds:**

\[ f(m, n, r) \leq (r-1) \cdot n \]
\[ f(m, n, r) \leq (r-1) \cdot m \]
\[
\text{cost}(C) \leq \text{cost}(C_t) + \text{cost}(C_b) + |X_b| - \nats(F_b) + |C_t|
\]
\[
\begin{align*}
\text{cost}( C ) & \leq \text{cost}( C_t ) + \text{cost}( C_b ) + |X_b| - \#\text{rts}( F_b ) + |C_t| \\
& \leq f(m_t, n_t, r-s-1) + 
\end{align*}
\]
\[
\begin{align*}
\text{cost}(C) & \leq \text{cost}(C_t) + \text{cost}(C_b) + |X_b| - \#\text{rts}(\mathcal{F}_b) + |C_t| \\
& \leq f(m_t, n_t, r-s-1) + f(m_b, n_b, s) + 
\end{align*}
\]
\[
\begin{aligned}
\text{cost}(C) & \leq \text{cost}(C_t) + \text{cost}(C_b) + |X_b| - \#\text{rts}(F_b) + |C_t| \\
& \leq f(m_t, n_t, r-s-1) + f(m_b, n_b, s) + n-n_t -
\end{aligned}
\]
\[
\text{cost}\left( C \right) \leq \text{cost}\left( C_t \right) + \text{cost}\left( C_b \right) + |X_b| - \#\text{rts}(F_b) + |C_t|
\]

\[
\leq f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n-n_t - (s+1)n_t +
\]
\[
\begin{align*}
\text{cost}(C) & \leq \text{cost}(C_t) + \text{cost}(C_b) + |X_b| - \#\text{rts}(F_b) + |C_t| \\
& \leq f(m_t, n_t, r-s-1) + f(m_b, n_b, s) + n-n_t - (s+1)n_t + \\
\end{align*}
\]

Each node in \( F_t \) has at least \( s+1 \) children in \( F_b \), and they must all be different roots of \( F_b \).
\[ \begin{align*}
    r & \left\{ \begin{array}{l}
    \mathcal{F}_t \\
    \mathcal{F}_b \\
    \end{array} \right\} \begin{array}{l}
    r - s - 1 < r \\
    s \\
    \end{array} \quad |X_{> s}| = n_t \\
    |X_{\leq s}| = n_b = n - n_t \\
    |C_t| = m_t \\
    |C_b| = m_b
    \end{align*} \]

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

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

Each node in \( \mathcal{F}_t \) has at least \( s + 1 \) children in \( \mathcal{F}_b \), and they must all be different roots of \( \mathcal{F}_b \).
\[\begin{align*}
\text{cost}(C) & \leq \text{cost}(C_t) + \text{cost}(C_b) + |X_b| - \#\text{rts}(F_b) + |C_t| \\
& \leq f(m_t, n_t, r-s-1) + f(m_b, n_b, s) + n - n_t - (s+1) \cdot n_t + m_t
\end{align*}\]

Each node in \(F_t\) has at least \(s+1\) children in \(F_b\), and they must all be different roots of \(F_b\).

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

\[ n_t + n_b = n \]
\[ m_t + m_b \leq m \]
\[ 0 \leq s < r \]
\[ f(m,n,r) \leq f(m_+,n_+,r-s-1) + f(m_-,n_-,s) + n - (s+2) \cdot n_+ + m_+ \]

\[ n_+ + n_- = n \]
\[ m_+ + m_- \leq m \]
\[ 0 \leq s < r \]

Assume: \[ f(\mu,\nu,\rho) \leq k \cdot \mu + \nu \cdot g(\rho) \]
\[
\begin{align*}
f(m,n,r) &\leq f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n - (s+2) \cdot n_t + m_t \\
\text{Assume: } &\quad f(\mu,\nu,\rho) \leq k \cdot \mu + \nu \cdot g(\rho) \\
f(m,n,r) &\leq k \cdot m_t + n_t \cdot g(r-s-1) + f(m_b,n_b,s) + n - (s+2) \cdot n_t + m_t \\
&\leq k \cdot m_t + n_t \cdot g(r) + f(m_b,n_b,s) + n - s \cdot n_t + m_t
\end{align*}
\]
\[ f(m,n,r) \leq f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n - (s+2) \cdot n_t + m_t \]

\[
\begin{align*}
n_t + n_b &= n \\
m_t + m_b &\leq m \\
0 &\leq s < r
\end{align*}
\]

Assume: \( f(\mu,\nu,\rho) \leq k \cdot \mu + \nu \cdot g(\rho) \)

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

\[ \leq k \cdot m_t + n_t \cdot g(r) + f(m_b,n_b,s) + n - s \cdot n_t + m_t \]

choose \( s = g(r) \)
\[ f(m,n,r) \leq f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n - (s+2) \cdot n_t + m_t \]

\[ n_t + n_b = n \]
\[ m_t + m_b \leq m \]
\[ 0 \leq s < r \]

Assume: \( f(\mu,\nu,\rho) \leq k \cdot \mu + \nu \cdot g(\rho) \)

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

choose \( s = g(r) \)
\[ f(m,n,r) \leq (k+1) \cdot m_t + f(m_b,n_b,s) + n \]
\[ \leq (k+1) \cdot m_t + f(m_b,n,s) + n \]
\[ s = g(r) \]

\[ f(m, n, r) \leq (k+1) \cdot m + f(m_b, n, s) + n \]
\[ s = g(r) \]
\[ f(m, n, r) \leq (k+1) \cdot m_\perp + f(m_b, n, s) + n \quad \text{if} \quad -(k+1) \cdot (m_b + m_\perp) \]
\[ s = g(r) \]

\[ f(m, n, r) \leq (k+1) \cdot m_+ + f(m_b, n, s) + n \quad \text{for} \quad -(k+1) \cdot (m_b + m_+) \]
\[ s = g(r) \]

\[ f(m, n, r) \leq (k+1) \cdot m + f(m_b, n, s) + n \]

\[ -(k+1) \cdot (m_b + m) \]

\[ f(m, n, r) - (k+1) \cdot m \leq f(m_b, n, s) - (k+1) \cdot m_b + n \]
\[ s = g(r) \]

\[ f(m,n,r) \leq (k+1) \cdot m + f(m_b,n,s) + n \quad \text{-(k+1)} \cdot (m_b + m_t) \]

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

\[ \phi(m,n,r) \leq \phi(m_b,n,g(r)) + n \]
\[ s = g(r) \]

\[
f(m,n,r) \leq (k+1) \cdot m_+ + f(m_b,n,s) + n \quad - (k+1) \cdot (m_b + m_+) \]

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

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

\[
\phi(m,n,r) \leq n \cdot g^*(r)
\]
\[ s = g(r) \]

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

\[ -(k+1) \cdot (m_b + m_t) \]

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

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

\[ \phi(m, n, r) \leq n \cdot g^*(r) \]

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

Shifting Corollary:

If \( f(m,n,r) \leq k \cdot m + n \cdot g(r) \)
then also \( f(m,n,r) \leq (k+i) \cdot m + n \cdot g^{**...*}(r) \)
for any \( i \geq 0 \)
If \( f(m,n,r) \leq k \cdot m + n \cdot g(r) \)
then also \( f(m,n,r) \leq (k+i) \cdot m + n \cdot g_{**...**}^{i}(r) \)
for any \( i \geq 0 \)
If \( f(m,n,r) \leq k \cdot m + n \cdot g(r) \)
then also \( f(m,n,r) \leq (k+i) \cdot m + n \cdot \underbrace{g ** ... *}_{i}(r) \)
for any \( i \geq 0 \)

Trivial bound: \( f(m,n,r) \leq n \cdot (r-1) \)
If \( f(m,n,r) \leq k \cdot m + n \cdot g(r) \)

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

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 \( f(m,n,r) \leq (k+i) \cdot m + n \cdot g^{**\ldots\ldots}(r) \)  
for any \( i \geq 0 \)

Trivial bound: \( f(m,n,r) \leq n \cdot (r-1) \)  
\[ = 0 \cdot m + n \cdot (r-1) \]

\( g(r) = r-1 \)  
\( g^{\ast}(r) = r-1 \)
\[ f(m,n,r) \leq f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n - (s+2) \cdot n_t + m_t \]

\[ n_t + n_b = n \]
\[ m_t + m_b \leq m \]
\[ 0 \leq s < r \]
\[ f(m,n,r) \leq f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n - (s+2) \cdot n_t + m_t \]

\[ \begin{align*}
    n_t + n_b &= n \\
    m_t + m_b &\leq m \\
    0 \leq s &< r
\end{align*} \]

Trivial bound: \( f(\mu,\nu,\rho) \leq \nu \cdot (\rho-1) \)
\[ f(m,n,r) \leq f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n - (s+2)\cdot n_t + m_t \]

\[ n_t + n_b = n \]
\[ m_t + m_b \leq m \]
\[ 0 \leq s < r \]

Trivial bound: \[ f(\mu,\nu,\rho) \leq \nu \cdot (\rho-1) \]

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

\[ \begin{align*}
    n_t + n_b &= n \\
    m_t + m_b &\leq m
\end{align*} \quad 0 \leq s < r \]

Trivial bound: \[ f(\mu,\nu,\rho) \leq \nu \cdot (\rho-1) \]

\[ f(m,n,r) \leq n_t \cdot (r-s-2) + f(m_b,n_b,s) + n - (s+2) \cdot n_t + m_t \]
\[ \leq n_t \cdot (r-2s-4) + f(m_b,n_b,s) + n + m_t \]
\[ f(m,n,r) \leq f(m_t,n_r,r-s-1) + f(m_b,n_b,s) + n - (s+2) \cdot n_t + m_t \]

\[ n_t + n_b = n \]
\[ m_t + m_b \leq m \]
\[ 0 \leq s < r \]

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

\[
\begin{align*}
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 \[ \leq n_t \cdot (r-2s-4) + f(m_b,n_b,s) + n + m_t \]
\end{align*}
\]

set \( s = \lfloor r/2 \rfloor \)
\[ f(m,n,r) \leq f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n - (s+2)n_t + m_t \]

\[ n_t + n_b = n \quad 0 \leq s < r \]
\[ m_t + m_b \leq m \]

Trivial bound: \[ f(\mu,\nu,\rho) \leq \nu(\rho-1) \]

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

set \( s = \lfloor r/2 \rfloor \)

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

\[
\begin{align*}
n_t + n_b &= n \\
m_t + m_b &\leq m \\
0 &\leq s < r
\end{align*}
\]

Trivial bound: \[ f(\mu,\nu,\rho) \leq \nu \cdot (\rho-1) \]

\[
\begin{align*}
f(m,n,r) &\leq n_t \cdot (r-s-2) + f(m_b,n_b,s) + n - (s+2) \cdot n_t + m_t \\
&\leq n_t \cdot (r-2s-4) + f(m_b,n_b,s) + n + m_t
\end{align*}
\]

set \( s = \lfloor r/2 \rfloor \)

\[
\begin{align*}
f(m,n,r) &\leq f(m_b,n_b,r/2) + n + m_t \\
f(m,n,r) - m &\leq f(m_b,n_b,r/2) - m_b + n
\end{align*}
\]
\[ f(m,n,r) \leq f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n - (s+2) \cdot n_t + m_t \]

\[ n_t + n_b = n \]
\[ m_t + m_b \leq m \]
\[ 0 \leq s < r \]

Trivial bound: \[ f(\mu,\nu,\rho) \leq \nu \cdot (\rho-1) \]

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

set \[ s = \lfloor r/2 \rfloor \]

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

\[ f(m,n,r) - m \leq f(m_b,n_b,r/2) - m_b + n \]

\[ f(m,n,r) \leq m + n \cdot \log r \]
If \( f(m,n,r) \leq k \cdot m + n \cdot g(r) \)
then also \( f(m,n,r) \leq (k+i) \cdot m + n \cdot g^{**...*}(r) \)
for any \( i \geq 0 \)
If \( f(m,n,r) \leq k \cdot m + n \cdot g(r) \)
then also \( f(m,n,r) \leq (k+i) \cdot m + n \cdot g^{**\ldots*}(r) \)
for any \( i \geq 0 \)

We know bound: \( f(m,n,r) \leq m + n \cdot \log r \)
If \( f(m,n,r) \leq k \cdot m + n \cdot g(r) \)

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

for any \( i \geq 0 \)

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

Therefore for any \( i \geq 0 \):

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

Choice of $i$:
For any $i \geq 0$:

\[ f(m,n,r) \leq (i+1) \cdot m + n \cdot \log^{\ldots}(r) \]

Choice of $i$:

Define $\alpha(r) = \min\{ i \mid \log^{\ldots}(r) \leq i \}$
For any $i \geq 0$:  
$$f(m, n, r) \leq (i+1) \cdot m + n \cdot \log^{**\cdots*}(r)$$

Choice of $i$:

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

$$f(m, n, r) \leq (m+n)(1+\alpha(r))$$
For any $i \geq 0$:

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

Choice of $i$:

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

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

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

Choice of \( i \):

Define \( \alpha(m,n,r) = \min\{ i \mid \log^{**\ldots\ast}(r) \leq m/n \} \)
For any $i \geq 0$:
\[ f(m,n,r) \leq (i+1) \cdot m + n \cdot \log^{**\ldots\ast}(r) \]

**Choice of $i$**:

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

\[ f(m,n,r) \leq m(2 + \alpha(m,n,r)) \]
For any $i \geq 0$:

$$f(m,n,r) \leq (i+1) \cdot m + n \cdot \log^{**\ldots\ast}(r)$$

Choice of $i$:

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

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

Define $\alpha(m,n) = \min\{i \mid \log^{**\ldots\ast}(\log n) \leq m/n\}$
For any $i \geq 0$:

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

**Choice of $i$:**

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

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

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

$$f(m,n,r) \leq m(2+\alpha(m,n))$$
Good bounds for \( f(m,n,r) \) when \( r \) is small

\[
\begin{align*}
  f(m,n,0) &= 0 \\
  f(m,n,1) &= 0 \\
  f(m,n,2) &\leq m
\end{align*}
\]
\[ f(m,n,r) \leq f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n - (s+2) \cdot n_t + m_t \]

\[ n_t + n_b = n \]
\[ m_t + m_b \leq m \]
\[ 0 \leq s < r \]
\[ f(m,n,r) \leq f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n - (s+2) \cdot n_t + m_t \]

\[ n_t + n_b = n \]
\[ m_t + m_b \leq m \]
\[ 0 \leq s < r \]

**choose** \( s = 2 \):

\[ f(m,n,r) \leq f(m_t,n_t,r-3) + f(m_b,n_b,2) + n - 4 \cdot n_t + m_t \]
$$f(m,n,r) \leq f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n - (s+2)n_t + m_t$$

\[
\begin{align*}
n_t + n_b &= n \\
m_t + m_b &\leq m \\
0 \leq s &< r
\end{align*}
\]

choose \(s = 2\):

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

\[
\leq n_t(r-4) \quad \leq m_b
\]
$$f(m,n,r) \leq f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n - (s+2)n_t + m_t$$

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

choose $s = 2$:

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

$$\leq n_t(r-4)$$

$$\leq m_b$$

$$f(m,n,r) \leq m + n + (r-8)n_t$$
\[ f(m,n,r) \leq f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n - (s+2)\cdot n_t + m_t \]

\[
\begin{align*}
    n_t + n_b &= n \\
    m_t + m_b &\leq m \\
    0 &\leq s < r
\end{align*}
\]

choose \( s = 2 \):

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

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

\[ f(m,n,r) \leq m + n \quad \text{for } r \leq 8, \text{ i.e. for } n < 512 \]
f(m, n, r) \leq n(r-1)
\[ f(m,n,r) \leq n(r-1) \]

\[
\begin{align*}
n=4, r=1: & \text{ bound } 4(2-1) = 4 \\
\text{actual bound: } & 1
\end{align*}
\]
\[ f(m,n,r) \leq n(r-1) \]

\[ n=4, r=1: \text{ bound } 4(2-1) = 4 \]

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 $0 \leq i \leq r$.

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

\[ C(r,i) = 2^{i-1}(2r-i)-1 \]
\[
f(m,n,r) \leq f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n - (s+2)\cdot n_t + m_t
\]

\[
n_t + n_b = n
m_t + m_b \leq m \quad 0 \leq s < r
\]

choose \( s = 8 \):

\[
f(m,n,r) \leq f(m_t,n_t,r-9) + f(m_b,n_b,8) + n - 10\cdot n_t + m_t
\]
\[ f(m,n,r) \leq f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n - (s+2)\cdot n_t + m_t \]

\[
\begin{align*}
n_t + n_b &= n \\
m_t + m_b &\leq m \\
0 \leq s < r
\end{align*}
\]

choose \( s = 8 \):

\[
\begin{align*}
f(m,n,r) &\leq f(m_t,n_t,r-3) + f(m_b,n_b,8) + n - 10\cdot n_t + m_t \\
&\leq n_t(r-4) - \sum R_i \cdot C(r-3,i) \\
&\leq m_b + n_b
\end{align*}
\]
\[ f(m,n,r) \leq f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n - (s+2)\cdot n_t + m_t \]

\[
\begin{align*}
n_t + n_b &= n \\
m_t + m_b &\leq m
\end{align*}
\]

\[ 0 \leq s < r \]

Choose \( s = 8 \):

\[ f(m,n,r) \leq f(m_t,n_t,r-3) + f(m_b,n_b,8) + n - 10 \cdot n_t + m_t \]

\[
\leq n_t(r-4) - \sum R_i \cdot C(r-3,i)
\leq m_b + n_b
\]

\[ R_i \geq n_t \text{ for each } i \]
\[
f(m,n,r) \leq f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n - (s+2) \cdot n_t + m_t
\]

\[
\begin{align*}
n_t + n_b &= n \\
m_t + m_b &\leq m \\
0 &\leq s < r
\end{align*}
\]

choose \( s = 8 \):

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

\[
R_i \geq n_t \text{ for each } i
\]

\[
f(m,n,r) \leq m + 2n + n_t(r-202)
\]
\[ f(m,n,r) \leq f(m_t,n_t,r-s-1) + f(m_b,n_b,s) + n - (s+2) \cdot n_t + m_t \]

\[
\begin{align*}
    n_t + n_b &= n \\
    m_t + m_b &\leq m \\
    0 \leq s &< r
\end{align*}
\]

choose \( s = 8 \):

\begin{align*}
    f(m,n,r) &\leq f(m_t,n_t,r-3) + f(m_b,n_b,8) + n - 10 \cdot n_t + m_t \\
    &\leq n_t(r-4) - \sum R_i \cdot C(r-3,i) \\
    &\leq m_b + n_b
\end{align*}

\( R_i \geq n_t \) 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 \texttt{Find(x)} operation make
“all” nodes in the “findpath” child of some node
further up.
Heuristic 2': Path compaction

when performing a \texttt{Find( x )} operation make “all” nodes in the “findpath” child of some node further up.
Heuristic 2': Path compaction

when performing a \textbf{Find}(x) operation make “all” nodes in the “findpath” child of some node further up.
Heuristic 2': Path compaction

when performing a \texttt{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.

1
2
3
4
5
6
7
8
Heuristic 2': Path compaction

when performing a \texttt{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 $\text{Find}(x)$ operation make “all” nodes in the “findpath” child of some node further up.
Heuristic 2': Path compaction

when performing a $\textbf{Find}(x)$ 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

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

and

$$\text{cost}(C) \leq \text{cost}(C_b) + \text{cost}(C_t) + |X_b| + |C_t|$$
Main Lemma:
$C$ ... sequence of compaction operations on $F$ with node set $X$
$X_t, X_b$ dissection for $F$ inducing subforests $F_t, F_b$

$\Rightarrow \exists$ compaction sequences
$C_b$ for $F_b$ and $C_t$ for $F_t$
with

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

and

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

\[ |C_b| + |C_t| \leq |C| \]

compression paths from \( C \)

**case 1:**

\[ \downarrow \quad \downarrow \quad \text{into } C_t \]

**case 2:**

\[ \downarrow \quad \downarrow \quad \text{into } C_b \]

**case 3:**

\[ \downarrow \quad \downarrow \quad \text{into } C_t \]

\[ \downarrow \quad \downarrow \quad \text{into } C_b \]
Proof:

\[ |C_b| + |C_t| \leq |C| \]

compaction paths from \( C \)

- **case 1:**
  - \( Y \)
  - \( X \)

- **case 2:**
  - \( Y \)
  - \( X \)

- **case 3:**
  - \( Y \)
  - \( X' \)

\( \infty \) into \( C_b \)

\( \infty \) into \( C_t \)
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 $\text{height}(v)$
Main Lemma:

\( C \) ... 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

\( C_b \) for \( \mathcal{F}_b \) and \( C_t \) for \( \mathcal{F}_t \)

with

\[ |C_b| + |C_t| \leq |C| \]

and

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

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 \( \sum_{v \in X_b} \text{height}(v) \leq |X_b| \) 
\[ \leq \#	ext{ children of } v \]
\[ \text{cost}(C) \leq \text{cost}(C_b) + \text{cost}(C_t) + |X_b| + |C_t| \]

\[ + \sum_{v \in X_b} \text{height}(v) \]

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

\[ \leq \]

\# children of \( v \)

\[ \text{cost}(C) \leq \text{cost}(C_b) + \text{cost}(C_t) + 2 \cdot |X_b| + |C_t| \]
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