CS 473: Fundamental Algorithms, Spring 2013

Greedy Algorithms for Minimum Spanning Trees

Lecture 12
March 1, 2013

Part I

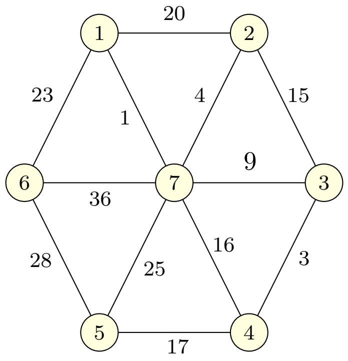
Greedy Algorithms: Minimum Spanning Tree

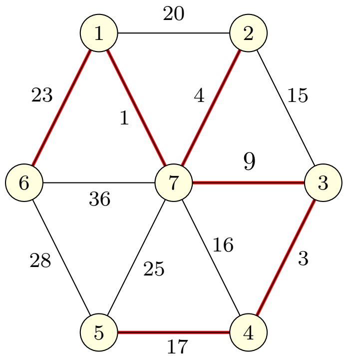
Minimum Spanning Tree

Input Connected graph $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ with edge costs

Goal Find $\mathbf{T} \subseteq \mathbf{E}$ such that $(\mathbf{V}, \mathbf{T})$ is connected and total cost of all edges in $\mathbf{T}$ is smallest

1. $\mathbf{T}$ is the minimum spanning tree (MST) of $\mathbf{G}$





Applications

1. Network Design
   1. Designing networks with minimum cost but maximum connectivity
2. Approximation algorithms
   1. Can be used to bound the optimality of algorithms to approximate Traveling Salesman Problem, Steiner Trees, etc.
3. Cluster Analysis

Greedy Template

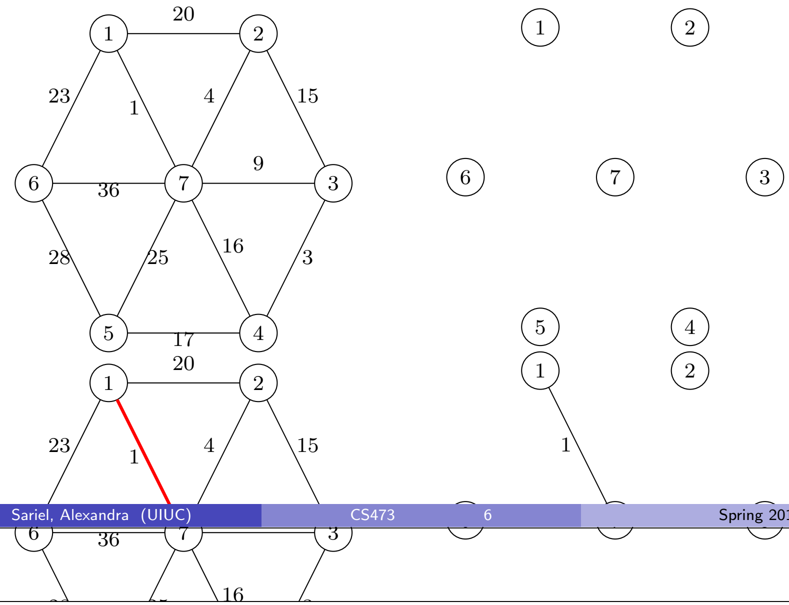
```
Initially E is the set of all edges in G
T is empty (* T will store edges of a MST *)
while E is not empty do
    choose i ∈ E
    if (i satisfies condition)
        add i to T
return the set T
```

Main Task: In what order should edges be processed? When should we add edge to spanning tree?

KA PA RD

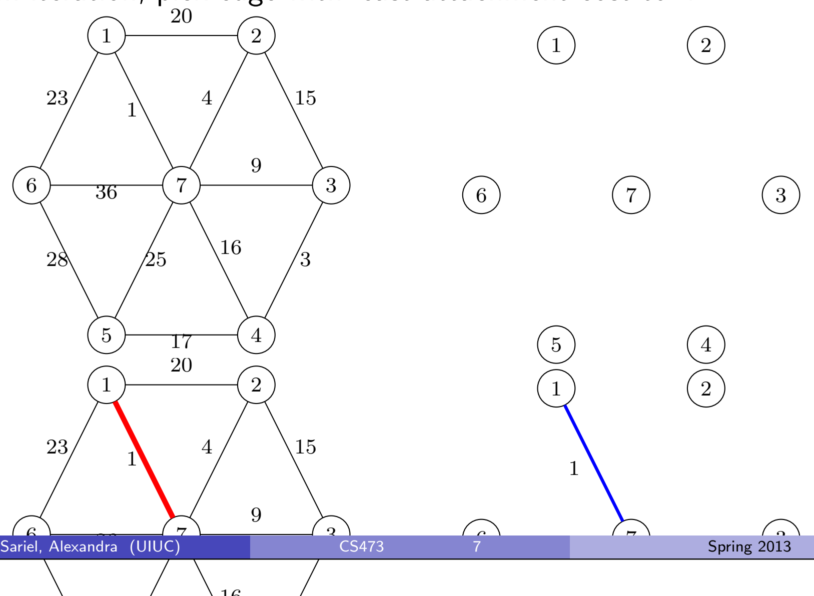
Kruskal's Algorithm

Process edges in the order of their costs (starting from the least) and add edges to **T** as long as they don't form a cycle.



Prim's Algorithm

T maintained by algorithm will be a tree. Start with a node in **T**. In each iteration, pick edge with least attachment cost to **T**.



Reverse Delete Algorithm

```
Initially E is the set of all edges in G
T is E (* T will store edges of a MST *)
while E is not empty do
    choose i ∈ E of largest cost
    if removing i does not disconnect T then
        remove i from T
return the set T
```

Returns a minimum spanning tree.

Back

Correctness of MST Algorithms

1. Many different MST algorithms
2. All of them rely on some basic properties of MSTs, in particular the **Cut Property** to be seen shortly.

Assumption

And for now ...

Assumption

Edge costs are distinct, that is no two edge costs are equal.

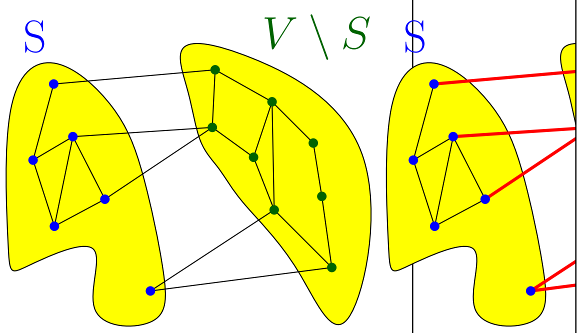
Cuts

Definition

Given a graph $\mathbf{G} = (\mathbf{V}, \mathbf{E})$, a **cut** is a partition of the vertices of the graph into two sets $(\mathbf{S}, \mathbf{V} \setminus \mathbf{S})$.

Edges having an endpoint on both sides are the **edges of the cut**.

A cut edge is **crossing** the cut.



Safe and Unsafe Edges

Definition

An edge $\mathbf{e} = (\mathbf{u}, \mathbf{v})$ is a safe edge if there is some partition of $\mathbf{V}$ into $\mathbf{S}$ and $\mathbf{V} \setminus \mathbf{S}$ and $\mathbf{e}$ is the unique minimum cost edge crossing $\mathbf{S}$ (one end in $\mathbf{S}$ and the other in $\mathbf{V} \setminus \mathbf{S}$).

Definition

An edge $\mathbf{e} = (\mathbf{u}, \mathbf{v})$ is an unsafe edge if there is some cycle $\mathbf{C}$ such that $\mathbf{e}$ is the unique maximum cost edge in $\mathbf{C}$.

Proposition

If edge costs are distinct then every edge is either safe or unsafe.

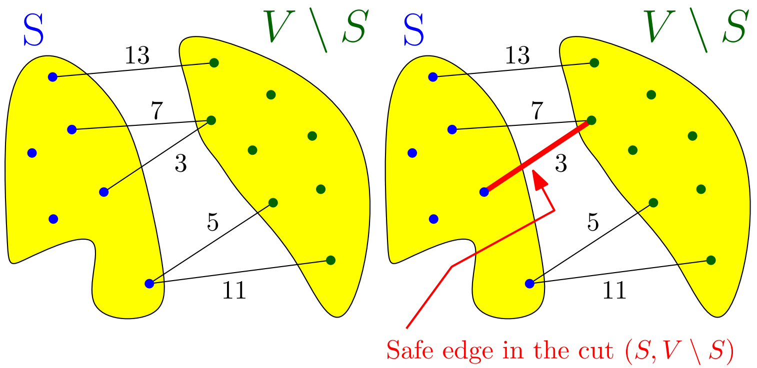
Proof.

Exercise. □

Safe edge

Example...

Every cut identifies one safe edge...



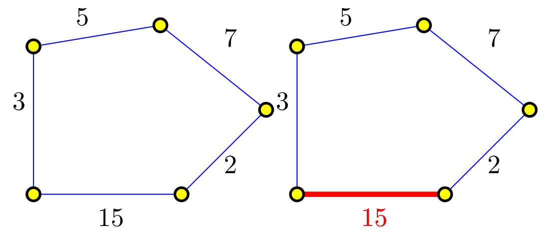
...the cheapest edge in the cut.

Note: An edge **e** may be a safe edge for *many* cuts!

Unsafe edge

Example...

Every cycle identifies one **unsafe** edge...



...the most expensive edge in the cycle.

Example

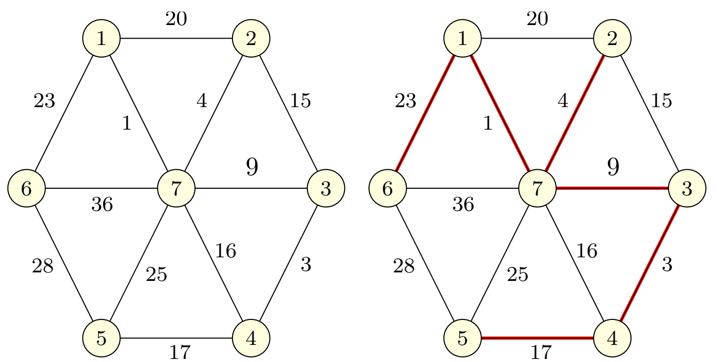


Figure: Graph with unique edge costs. Safe edges are red, rest are unsafe.

And all safe edges are in the MST in this case...

Key Observation: Cut Property

Lemma

If **e** *is a safe edge then every minimum spanning tree contains* **e**.

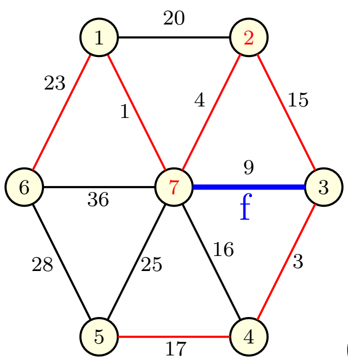
Proof.

1. Suppose (for contradiction) **e** is not in MST **T**.
2. Since **e** is safe there is an $\mathbf{S} \subset \mathbf{V}$ such that **e** is the unique min cost edge crossing **S**.
3. Since **T** is connected, there must be some edge **f** with one end in **S** and the other in $\mathbf{V} \setminus \mathbf{S}$
4. Since $\mathbf{c_f} > \mathbf{c_e}$, $\mathbf{T'} = (\mathbf{T} \setminus \{\mathbf{f}\}) \cup \{\mathbf{e}\}$ is a spanning tree of lower cost! Error: **T′** may not be a spanning tree!!

□

Error in Proof: Example

Problematic example. $\mathbf{S} = \{1, 2, 7\}$, $\mathbf{e} = (7, 3)$, $\mathbf{f} = (1, 6)$. $\mathbf{T} - \mathbf{f} + \mathbf{e}$ is not a spanning tree.



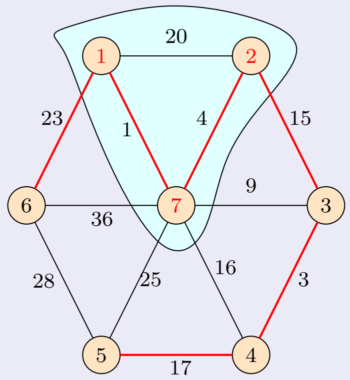
(A)

1. (A) Consider adding the edge $\mathbf{f}$.
2. (B) It is safe because it is the cheapest edge in the cut.
3. (C) Lets throw out the edge $\mathbf{e}$ currently in the spanning tree which is more expensive than $\mathbf{f}$ and is in the same cut. Put it $\mathbf{f}$ instead...
4. (D) New graph of selected edges is not a tree anymore. BUG.

(B)

Proof of Cut Property

Proof.



1. Suppose $\mathbf{e} = (\mathbf{v}, \mathbf{w})$ is not in MST $\mathbf{T}$ and $\mathbf{e}$ is min weight edge in cut $(\mathbf{S}, \mathbf{V} \setminus \mathbf{S})$. Assume $\mathbf{v} \in \mathbf{S}$.
2. $\mathbf{T}$ is spanning tree: there is a unique path $\mathbf{P}$ from $\mathbf{v}$ to $\mathbf{w}$ in $\mathbf{T}$
3. Let $\mathbf{w}'$ be the first vertex in $\mathbf{P}$ belonging to $\mathbf{V} \setminus \mathbf{S}$; let $\mathbf{v}'$ be the vertex just before it on $\mathbf{P}$, and let $\mathbf{e}' = (\mathbf{v}', \mathbf{w}')$
4. $\mathbf{T}' = (\mathbf{T} \setminus \{\mathbf{e}'\}) \cup \{\mathbf{e}\}$ is spanning tree of lower cost. (Why?) □

Proof of Cut Property (contd)

Observation

$\mathbf{T}' = (\mathbf{T} \setminus \{\mathbf{e}'\}) \cup \{\mathbf{e}\}$ *is a spanning tree.*

Proof.

$\mathbf{T}'$ is connected.

Removed $\mathbf{e}' = (\mathbf{v}', \mathbf{w}')$ from $\mathbf{T}$ but $\mathbf{v}'$ and $\mathbf{w}'$ are connected by the path $\mathbf{P} - \mathbf{f} + \mathbf{e}$ in $\mathbf{T}'$. Hence $\mathbf{T}'$ is connected if $\mathbf{T}$ is.

$\mathbf{T}'$ is a tree

$\mathbf{T}'$ is connected and has $\mathbf{n} - \mathbf{1}$ edges (since $\mathbf{T}$ had $\mathbf{n} - \mathbf{1}$ edges) and hence $\mathbf{T}'$ is a tree

□

Safe Edges form a Tree

Lemma

Let $\mathbf{G}$ *be a connected graph with distinct edge costs, then the set of safe edges form a connected graph.*

Proof.

1. Suppose not. Let $\mathbf{S}$ be a connected component in the graph induced by the safe edges.
2. Consider the edges crossing $\mathbf{S}$, there must be a safe edge among them since edge costs are distinct and so we must have picked it.

□

Safe Edges form an MST

Corollary
Let **G** *be a connected graph with distinct edge costs, then set of safe edges form the unique* MST *of* **G**.

Consequence: Every correct MST algorithm when **G** has unique edge costs includes exactly the safe edges.

Cycle Property

Lemma
If **e** *is an unsafe edge then no* MST *of* **G** *contains* **e**.

Proof.
Exercise. See text book. □

Note: Cut and Cycle properties hold even when edge costs are not distinct. Safe and unsafe definitions do not rely on distinct cost assumption.

Correctness of Prim's Algorithm

Prim's Algorithm
Pick edge with minimum attachment cost to current tree, and add to current tree.

Proof of correctness.
1. If **e** is added to tree, then **e** is safe and belongs to every MST.
   1. Let **S** be the vertices connected by edges in **T** when **e** is added.
   2. **e** is edge of lowest cost with one end in **S** and the other in $\mathbf{V} \setminus \mathbf{S}$ and hence **e** is safe.
2. Set of edges output is a spanning tree
   1. Set of edges output forms a connected graph: by induction, **S** is connected in each iteration and eventually $\mathbf{S} = \mathbf{V}$.
   2. Only safe edges added and they do not have a cycle □

Correctness of Kruskal's Algorithm

Kruskal's Algorithm
Pick edge of lowest cost and add if it does not form a cycle with existing edges.

Proof of correctness.
1. If $\mathbf{e} = (\mathbf{u}, \mathbf{v})$ is added to tree, then **e** is safe
   1. When algorithm adds **e** let **S** and **S**' be the connected components containing **u** and **v** respectively
   2. **e** is the lowest cost edge crossing **S** (and also **S**').
   3. If there is an edge $\mathbf{e}'$ crossing **S** and has lower cost than **e**, then $\mathbf{e}'$ would come before **e** in the sorted order and would be added by the algorithm to **T**
2. Set of edges output is a spanning tree : exercise

□

Correctness of Reverse Delete Algorithm

Reverse Delete Algorithm

Consider edges in decreasing cost and remove an edge if it does not disconnect the graph

Proof of correctness.

Argue that only unsafe edges are removed (see text book). □

When edge costs are not distinct

Heuristic argument: Make edge costs distinct by adding a small tiny and different cost to each edge

Formal argument: Order edges lexicographically to break ties

1. $\mathbf{e_i} \prec \mathbf{e_j}$ if either $\mathbf{c(e_i)} < \mathbf{c(e_j)}$ or $(\mathbf{c(e_i)} = \mathbf{c(e_j)}$ and $\mathbf{i} < \mathbf{j})$
2. Lexicographic ordering extends to sets of edges. If $\mathbf{A}, \mathbf{B} \subseteq \mathbf{E}$, $\mathbf{A} \neq \mathbf{B}$ then $\mathbf{A} \prec \mathbf{B}$ if either $\mathbf{c(A)} < \mathbf{c(B)}$ or $(\mathbf{c(A)} = \mathbf{c(B)}$ and $\mathbf{A} \setminus \mathbf{B}$ has a lower indexed edge than $\mathbf{B} \setminus \mathbf{A})$
3. Can order all spanning trees according to lexicographic order of their edge sets. Hence there is a unique MST.

Prim's, Kruskal, and Reverse Delete Algorithms are optimal with respect to lexicographic ordering.

Edge Costs: Positive and Negative

1. Algorithms and proofs don't assume that edge costs are non-negative! MST algorithms work for arbitrary edge costs.
2. Another way to see this: make edge costs non-negative by adding to each edge a large enough positive number. Why does this work for MSTs but not for shortest paths?
3. Can compute *maximum* weight spanning tree by negating edge costs and then computing an MST.

Part II

Data Structures for MST: Priority Queues and Union-Find

Implementing Prim's Algorithm

Implementing Prim's Algorithm

Prim_ComputeMST

```
E is the set of all edges in G
S = {1}
T is empty (* T will store edges of a MST *)
while S ≠ V do
    pick e = (v, w) ∈ E such that
        v ∈ S and w ∈ V − S
        e has minimum cost
    T = T ∪ e
    S = S ∪ w
return the set T
```

Analysis

1. Number of iterations $= O(n)$, where n is number of vertices
2. Picking e is $O(m)$ where m is the number of edges
3. Total time $O(nm)$

Implementing Prim's Algorithm

More Efficient Implementation

Prim_ComputeMST

```
E is the set of all edges in G
S = {1}
T is empty (* T will store edges of a MST *)
for v ∉ S, a(v) = min_{w∈S} c(w, v)
for v ∉ S, e(v) = w such that w ∈ S and c(w, v) is minimum
while S ≠ V do
    pick v with minimum a(v)
    T = T ∪ {(e(v), v)}
    S = S ∪ {v}
    update arrays a and e
return the set T
```

Maintain vertices in $V \setminus S$ in a priority queue with key $a(v)$.

Priority Queues

Data structure to store a set S of n elements where each element $v \in S$ has an associated real/integer key $k(v)$ such that the following operations

1. **makeQ**: create an empty queue
2. **findMin**: find the minimum key in S
3. **extractMin**: Remove $v \in S$ with smallest key and return it
4. **add**(v, $k(v)$): Add new element v with key $k(v)$ to S
5. **Delete**(v): Remove element v from S
6. **decreaseKey** (v, $k'(v)$): *decrease* key of v from $k(v)$ (current key) to $k'(v)$ (new key). Assumption: $k'(v) \leq k(v)$
7. **meld**: merge two separate priority queues into one

Prim's using priority queues

```
E is the set of all edges in G
S = {1}
T is empty (* T will store edges of a MST *)
for v ∉ S, a(v) = min_{w∈S} c(w, v)
for v ∉ S, e(v) = w such that w ∈ S and c(w, v) is minimum
while S ≠ V do
    pick v with minimum a(v)
    T = T ∪ {(e(v), v)}
    S = S ∪ {v}
    update arrays a and e
return the set T
```

Maintain vertices in $V \setminus S$ in a priority queue with key $a(v)$

1. Requires $O(n)$ **extractMin** operations
2. Requires $O(m)$ **decreaseKey** operations

Running time of Prim's Algorithm

$O(n)$ **extractMin** operations and $O(m)$ **decreaseKey** operations

1. Using standard Heaps, **extractMin** and **decreaseKey** take $O(\log n)$ time. Total: $O((m+n)\log n)$
2. Using Fibonacci Heaps, $O(\log n)$ for **extractMin** and $O(1)$ (amortized) for **decreaseKey**. Total: $O(n \log n + m)$.

Prim's algorithm and Dijkstra's algorithms are similar. Where is the difference?

Kruskal's Algorithm

```
Kruskal_ComputeMST
    Initially E is the set of all edges in G
    T is empty (* T will store edges of a MST *)
    while E is not empty do
        choose e ∈ E of minimum cost
        if (T ∪ {e} does not have cycles)
            add e to T
    return the set T
```

1. Presort edges based on cost. Choosing minimum can be done in $O(1)$ time
2. Do **BFS**/**DFS** on $T \cup \{e\}$. Takes $O(n)$ time
3. Total time $O(m \log m) + O(mn) = O(mn)$

Implementing Kruskal's Algorithm Efficiently

```
Kruskal_ComputeMST
    Sort edges in E based on cost
    T is empty (* T will store edges of a MST *)
    each vertex u is placed in a set by itself
    while E is not empty do
        pick e = (u, v) ∈ E of minimum cost
        if u and v belong to different sets
            add e to T
            merge the sets containing u and v
    return the set T
```

Need a data structure to check if two elements belong to same set and to merge two sets.

Union-Find Data Structure

Data Structure

Store disjoint sets of elements that supports the following operations

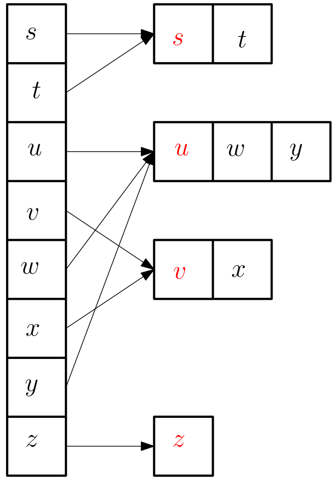
1. **makeUnionFind**(S) returns a data structure where each element of S is in a separate set
2. **find**(u) returns the *name* of set containing element u. Thus, u and v belong to the same set if and only if **find**(u) = **find**(v)
3. **union**(A, B) merges two sets A and B. Here A and B are the names of the sets. Typically the name of a set is some element in the set.

Implementing Union-Find using Arrays and Lists

Using lists

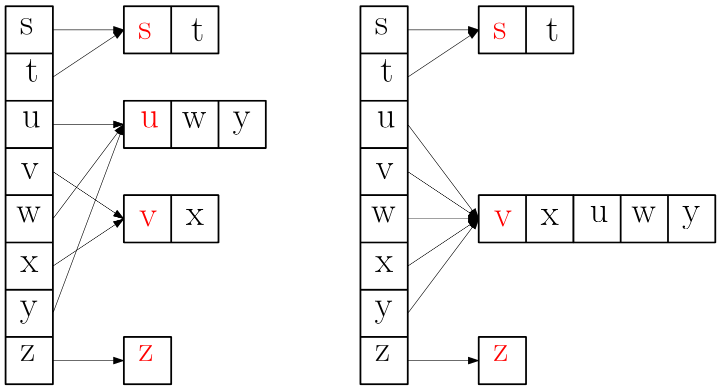
1. Each set stored as list with a name associated with the list.
2. For each element $u \in S$ a pointer to the its set. Array for pointers: `component`$[\mathbf{u}]$ is pointer for $\mathbf{u}$.
3. **makeUnionFind** ($\mathbf{S}$) takes $\mathbf{O(n)}$ time and space.

Example



Implementing Union-Find using Arrays and Lists

1. **find**($\mathbf{u}$) reads the entry `component`[$\mathbf{u}$]: $\mathbf{O(1)}$ time
2. **union**($\mathbf{A}$,$\mathbf{B}$) involves updating the entries `component`[$\mathbf{u}$] for all elements $\mathbf{u}$ in $\mathbf{A}$ and $\mathbf{B}$: $\mathbf{O(|A| + |B|)}$ which is $\mathbf{O(n)}$



Improving the List Implementation for Union

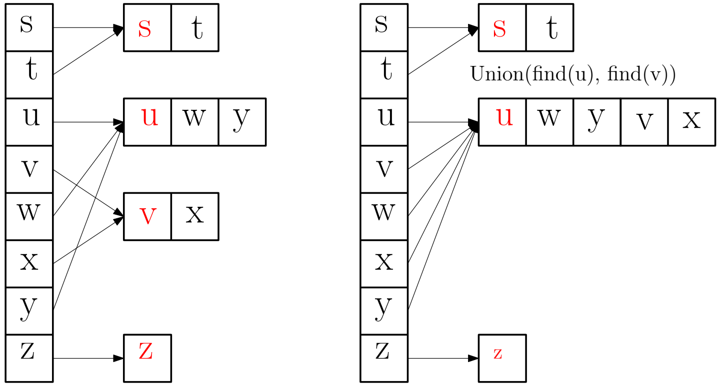
New Implementation

As before use `component`[$\mathbf{u}$] to store set of $\mathbf{u}$.
Change to **union**($\mathbf{A}$,$\mathbf{B}$):

1. with each set, keep track of its size
2. assume $|\mathbf{A}| \leq |\mathbf{B}|$ for now
3. Merge the list of $\mathbf{A}$ into that of $\mathbf{B}$: $\mathbf{O(1)}$ time (linked lists)
4. Update `component`$[\mathbf{u}]$ only for elements in the smaller set $\mathbf{A}$
5. Total $\mathbf{O(|A|)}$ time. Worst case is still $\mathbf{O(n)}$.

find still takes $\mathbf{O(1)}$ time

Example



The smaller set (list) is appended to the largest set (list)

Improving the List Implementation for Union

Question

Is the improved implementation provably better or is it simply a nice heuristic?

Theorem

Any sequence of $\mathbf{k}$ **union** *operations, starting from* **makeUnionFind**$(\mathbf{S})$ *on set* $\mathbf{S}$ *of size* $\mathbf{n}$, *takes at most* $\mathbf{O(k \log k)}$.

Corollary

Kruskal's algorithm can be implemented in $\mathbf{O(m \log m)}$ *time.*

Sorting takes $\mathbf{O(m \log m)}$ time, $\mathbf{O(m)}$ finds take $\mathbf{O(m)}$ time and $\mathbf{O(n)}$ unions take $\mathbf{O(n \log n)}$ time.

Amortized Analysis

Why does theorem work?

Key Observation

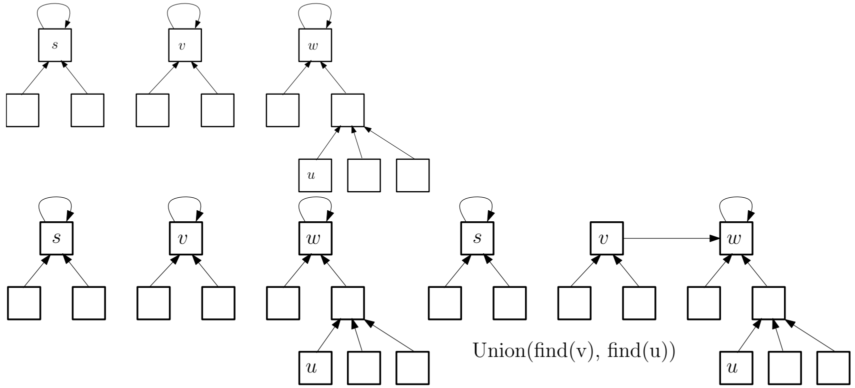
union$(\mathbf{A},\mathbf{B})$ takes $\mathbf{O(|A|)}$ time where $|\mathbf{A}| \leq |\mathbf{B}|$. Size of new set is $\geq \mathbf{2}|\mathbf{A}|$. Cannot double too many times.

Proof of Theorem

Proof.

1. Any union operation involves at most 2 of the original one-element sets; thus at least $\mathbf{n - 2k}$ elements have never been involved in a union
2. Also, maximum size of any set (after $\mathbf{k}$ unions) is $\mathbf{2k}$
3. **union**$(\mathbf{A},\mathbf{B})$ takes $\mathbf{O(|A|)}$ time where $|\mathbf{A}| \leq |\mathbf{B}|$.
4. *Charge* each element in $\mathbf{A}$ constant time to *pay* for $\mathbf{O(|A|)}$ time.
5. How much does any element get charged?
6. If `component`$[\mathbf{v}]$ is updated, set containing $\mathbf{v}$ *doubles* in size
7. `component`$[\mathbf{v}]$ is updated at most $\mathbf{\log 2k}$ times
8. Total number of updates is $\mathbf{2k \log 2k = O(k \log k)}$ $\square$

Improving Worst Case Time



Better data structure

Maintain elements in a forest of *in-trees*; all elements in one tree belong to a set with root's name.

1. **find(u)**: Traverse from **u** to the root
2. **union(A, B)**: Make root of **A** (smaller set) point to root of **B**. Takes $O(1)$ time.

Details of Implementation

Each element $u \in S$ has a pointer $\mathbf{parent(u)}$ to its ancestor.

```
makeUnionFind(S)
    for each u in S do
        parent(u) = u
```

```
find(u)
    while (parent(u) ≠ u) do
        u = parent(u)
    return u
```

```
union(component(u), component(v))
        (* parent(u) = u & parent(v) = v *)
    if (|component(u)| ≤ |component(v)|) then
        parent(u) = v
    else
        parent(v) = u
    set new component size to |component(u)| + |component(v)|
```


Analysis

Theorem

The forest based implementation for a set of size $\mathbf{n}$*, has the following complexity for the various operations:* **makeUnionFind** *takes* $O(n)$*,* **union** *takes* $O(1)$*, and* **find** *takes* $O(\log n)$*.*

Proof.

1. **find(u)** depends on the height of tree containing **u**.
2. Height of **u** increases by at most **1** only when the set containing **u** changes its name.
3. If height of **u** increases then size of the set containing **u** (at least) doubles.
4. Maximum set size is **n**; so height of any tree is at most $O(\log n)$. □

Further Improvements: Path Compression

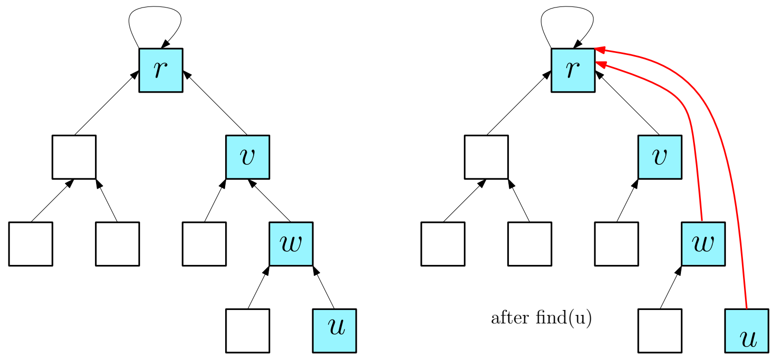
Observation

Consecutive calls of **find(u)** *take* $O(\log n)$ *time each, but they traverse the same sequence of pointers.*

Idea: Path Compression

Make all nodes encountered in the **find(u)** point to root.

Path Compression: Example



Path Compression

```
find(u):
    if (parent(u) ≠ u) then
        parent(u) = find(parent(u))
    return parent(u)
```

Question

Does Path Compression help?

Yes!

Theorem

With Path Compression, $\mathbf{k}$ *operations (***find** *and/or* **union***) take* $\mathbf{O(k\alpha(k, \min\{k, n\}))}$ *time where* α *is the* **inverse Ackermann function***.*

Ackermann and Inverse Ackermann Functions

Ackermann function $\mathbf{A(m, n)}$ defined for $\mathbf{m, n \geq 0}$ recursively

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

$\mathbf{A(3, n) = 2^{n+3} - 3}$
$\mathbf{A(4, 3) = 2^{65536} - 3}$

$\alpha\mathbf{(m, n)}$ is inverse Ackermann function defined as

$$\alpha\mathbf{(m, n) = \min\{i \mid A(i, \lfloor m/n \rfloor) \geq \log_2 n\}}$$

For **all practical** purposes $\alpha\mathbf{(m, n) \leq 5}$

Lower Bound for Union-Find Data Structure

Amazing result:

Theorem (Tarjan)

For **Union-Find***, any data structure in the pointer model requires* $\mathbf{\Omega(m\alpha(m, n))}$ *time for* $\mathbf{m}$ *operations.*

Running time of Kruskal's Algorithm

Using Union-Find data structure:

1. $\mathbf{O(m)}$ **find** operations (two for each edge)
2. $\mathbf{O(n)}$ **union** operations (one for each edge added to $\mathbf{T}$)
3. Total time: $\mathbf{O(m \log m)}$ for sorting plus $\mathbf{O(m\alpha(n))}$ for union-find operations. Thus $\mathbf{O(m \log m)}$ time despite the improved Union-Find data structure.

Best Known Asymptotic Running Times for MST

Prim's algorithm using Fibonacci heaps: $\mathbf{O(n \log n + m)}$.
If $\mathbf{m}$ is $\mathbf{O(n)}$ then running time is $\mathbf{\Omega(n \log n)}$.

Question

Is there a linear time ($\mathbf{O(m + n)}$ time) algorithm for MST?

1. $\mathbf{O(m \log^* m)}$ time **Fredman and Tarjan [1987]**.
2. $\mathbf{O(m + n)}$ time using bit operations in RAM model **Fredman and Willard [1994]**.
3. $\mathbf{O(m + n)}$ expected time (randomized algorithm) **Karger et al. [1995]**.
4. $\mathbf{O((n + m)\alpha(m, n))}$ time **Chazelle [2000]**.
5. Still open: Is there an $\mathbf{O(n + m)}$ time deterministic algorithm in the comparison model?

Chazelle, B. (2000). A minimum spanning tree algorithm with inverse-ackermann type complexity. *J. Assoc. Comput. Mach.*, 47(6):1028–1047.

Fredman, M. L. and Tarjan, R. E. (1987). Fibonacci heaps and their uses in improved network optimization algorithms. *J. Assoc. Comput. Mach.*, 34(3):596–615.

Fredman, M. L. and Willard, D. E. (1994). Trans-dichotomous algorithms for minimum spanning trees and shortest paths. *J. Comput. Sys. Sci.*, 48(3):533–551.

Karger, D. R., Klein, P. N., and Tarjan, R. E. (1995). A randomized linear-time algorithm to find minimum spanning trees. *J. Assoc. Comput. Mach.*, 42(2):321–328.