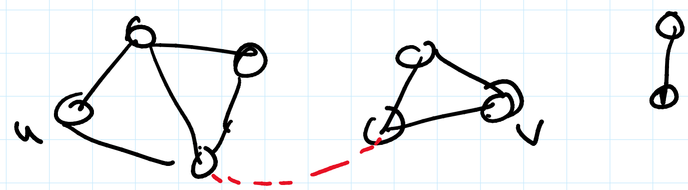
September 13, 2023 10:49 AM

Union-Find

Motivating Problem maintain undirected graph $G=(V,E)$
to support:
- connectivity queries: given $u, v \in V$, decide if u & v are connected
- insert an edge



appl: networks, Kruskal's MST alg'm

Sol'n: maintain connected components
- insert edge uv: union(find(u), find(v))
- query: find(u) == find(v)

Problem maintain collection of disjoint sets to support:
- union(A, B): union sets labeled A & B
- find (x): return label of set containing x
- makeset (x): create $\{x\}$.

Basic Method 1: Linked lists

e.g. $\underset{A}{\{0,2\}}$, $\underset{B}{\{1,3,4\}}$, $\underset{C}{\{5\}}$

Version 1.0:

A — 0 — 2

B — 1 — 3 — 4

union: $O(1)$
find: $O(n)$

B - 1 - 3 - 4

C - 5

find: $O(n)$

Version 1.1: with label field

0/A - 2/A

1/B - 3/B - 4/B

5/C

find: $O(1)$

union: $O(n)$

union(A,B): for each $x \in B$,
change x's label to A

Version 1.2: with "weighted union heuristic"

union(A,B): if size(A) > size(B)
as before
else for each $x \in A$,
change x's label to B.

Still worst-case $O(n)$

Claim $O(\log n)$ amortized time (find $O(1)$ worst case.)

i.e. any sequence of n makesets, $\leq n-1$ unions takes $O(n \log n)$ total time.

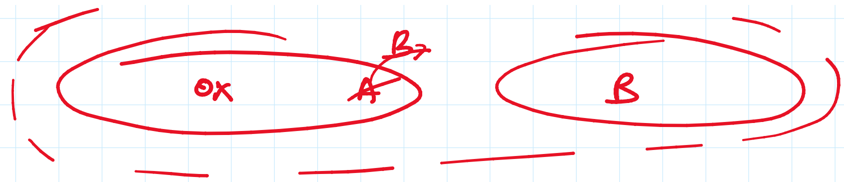
Pf:

total time $= O(n +$ total # of label changes$)$

fix an element x

let size(x) = size of set containing x

each time x changes label,



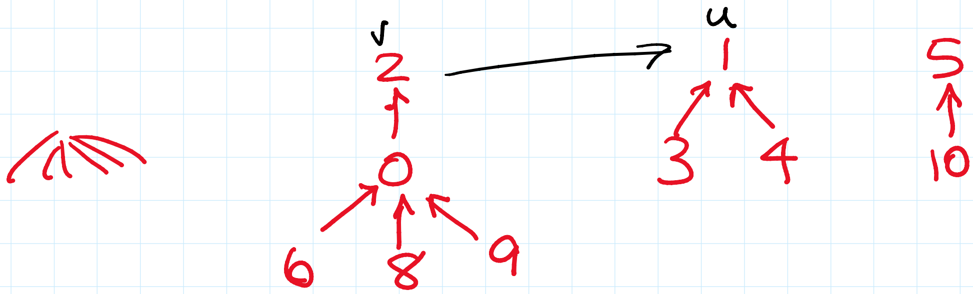
size(x) at least doubles

$\Rightarrow$ x changes label $\le \log n$ times

$\Rightarrow$ total $O(n \log n)$. $\square$

[$O(n \log s)$, let $s =$ max set size]

Basic Method 2: Trees

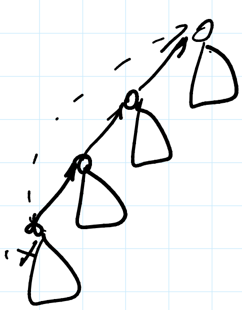


use roots as labels

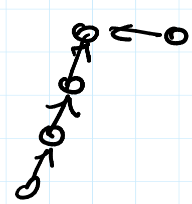
Version 2.1: union(u, v): $p(v) = u$ // link 2 trees

parent

find(x): if $p(x) == nil$
return x
else find(p(x)) // go up a path
return



union $O(1)$
find: $O(n)$ worst case

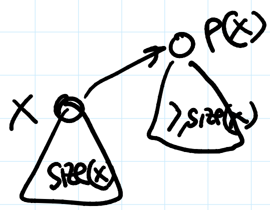


Version 2.2: with weighted union heuristic

union(u,v): if $size(u) > size(v)$
then $p(v) = u$, $size(u) \mathrel{+}= size(v)$
else $p(u) = v$, $size(v) \mathrel{+}= size(u)$

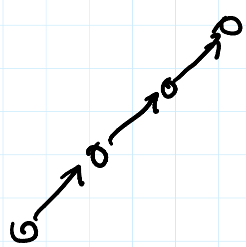
Claim find $O(\log n)$ worst case (union $O(1)$ worst case)

Pf: when we link x to $p(x)$,



$size(p(x)) \geq 2\, size(x)$

afterwards, $size(p(x)) \uparrow$
$size(x)$ stays same.



path length $\leq \log n$. $\square$

Q: better than $\log n$?
YES!

$O(\alpha(n))$ is possible!

Tarjan '75
la Poutre '90