CSE 373
Data Structures & Algorithms

Lecture 17

Disjoint Sets (II)

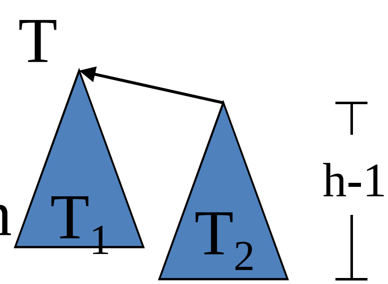
Brief Midterm Postmortem

- Heaps
- Hash tables
- Bubble sort
- Properties of Sorting Algorithms
- Merging

Analysis of Weighted Union

- With weighted union an up-tree of height h has weight at least 2^h.
- Proof by induction
  - Basis: $h = 0$. The up-tree has one node, $2^0 = 1$
  - Inductive step: Assume true for all $h' < h$.

Minimum weight up-tree of height h formed by weighted unions



$$W(T_1) \geq W(T_2) \geq 2^{h-1}$$

Weighted union

Induction hypothesis

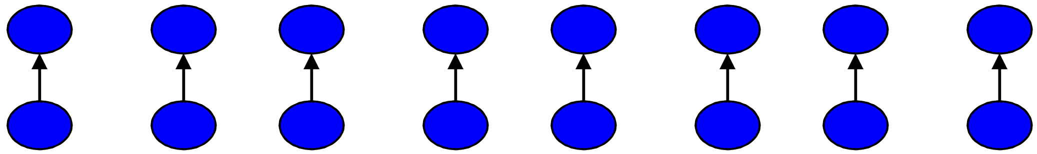
$$W(T) \geq 2^{h-1} + 2^{h-1} = 2^h$$

Analysis of Weighted Union

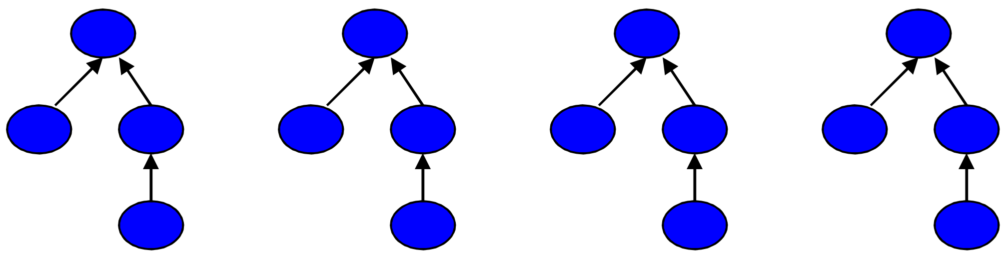
- Let T be an up-tree of weight n formed by weighted union. Let h be its height.
- $n \geq 2^h$
- $\log_2 n \geq h$
- Find(x) in tree T takes O(log n) time.
- Can we do better?

Worst Case for Weighted Union

n/2 Weighted Unions

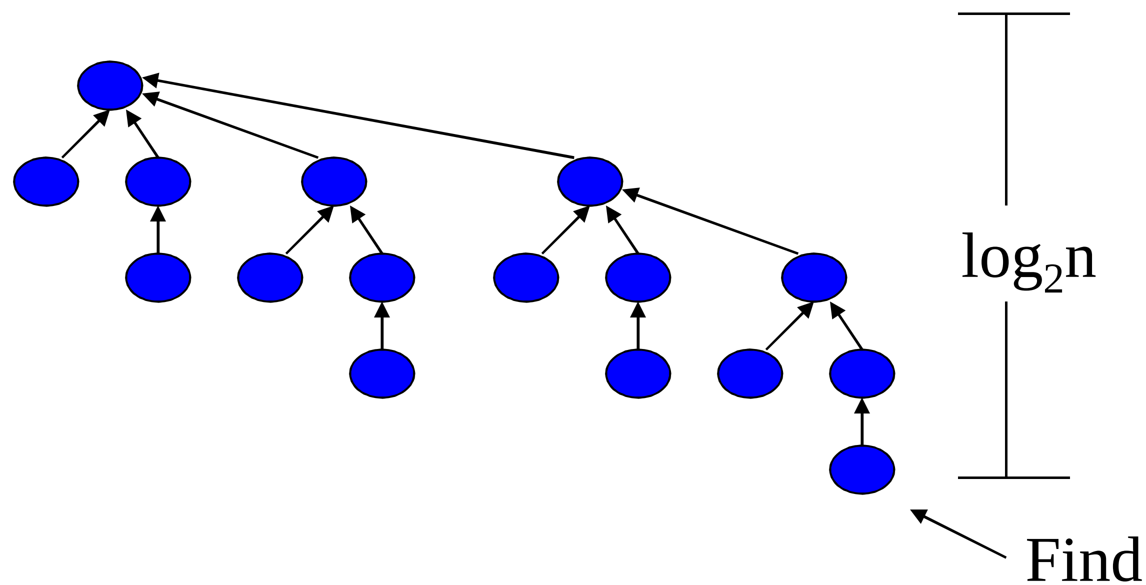


n/4 Weighted Unions



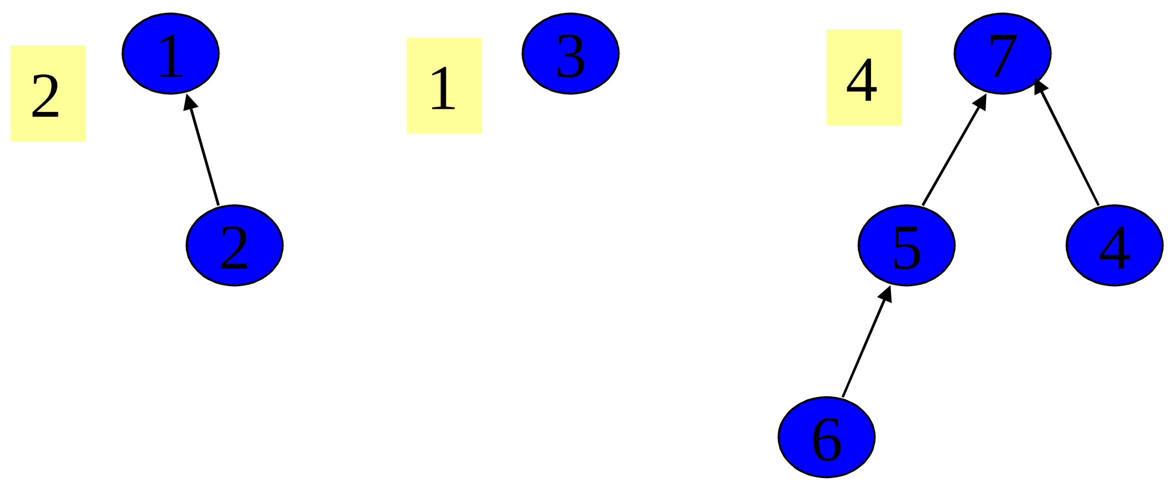
Example of Worst Cast (cont')

After n -1 = n/2 + n/4 + …+ 1 Weighted Unions



If there are $n = 2^k$ nodes then the longest path from leaf to root has length k.

Elegant Array Implementation



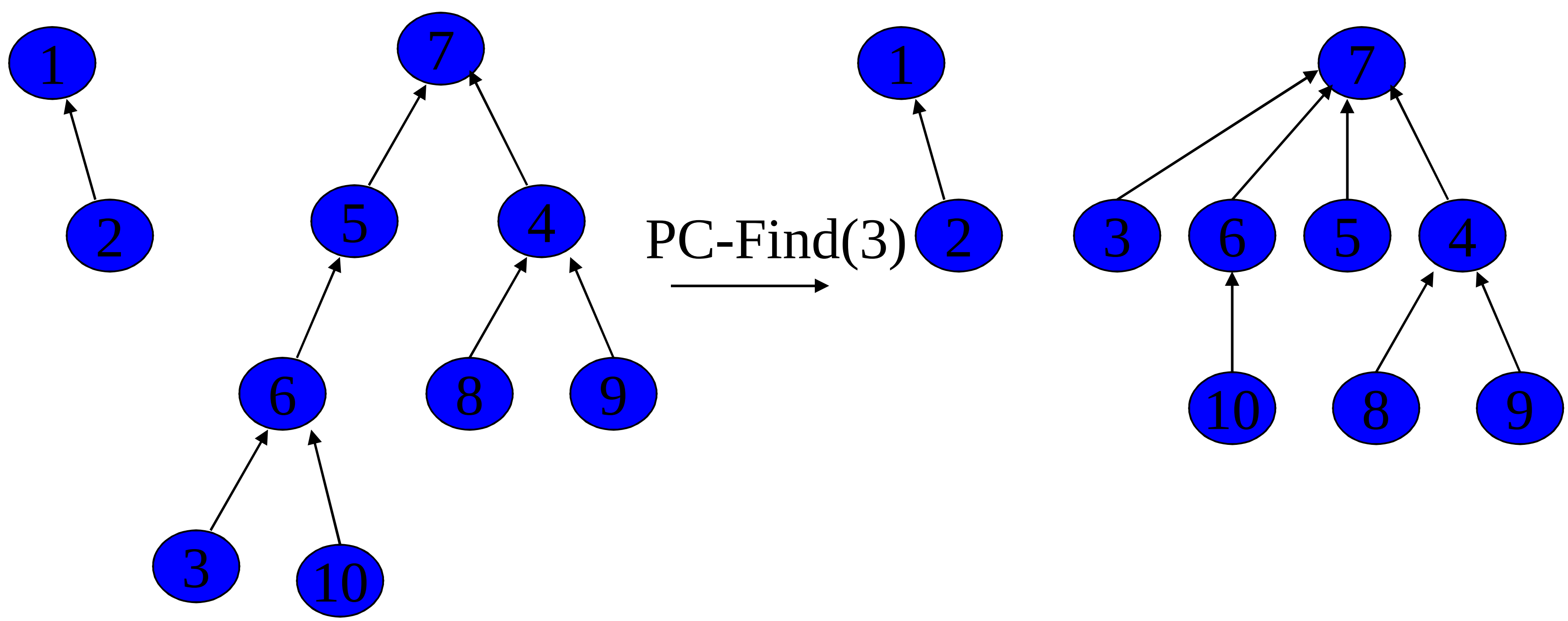
	1	2	3	4	5	6	7
up	0	1	0	7	7	5	0
weight	2		1				4

Weighted Union

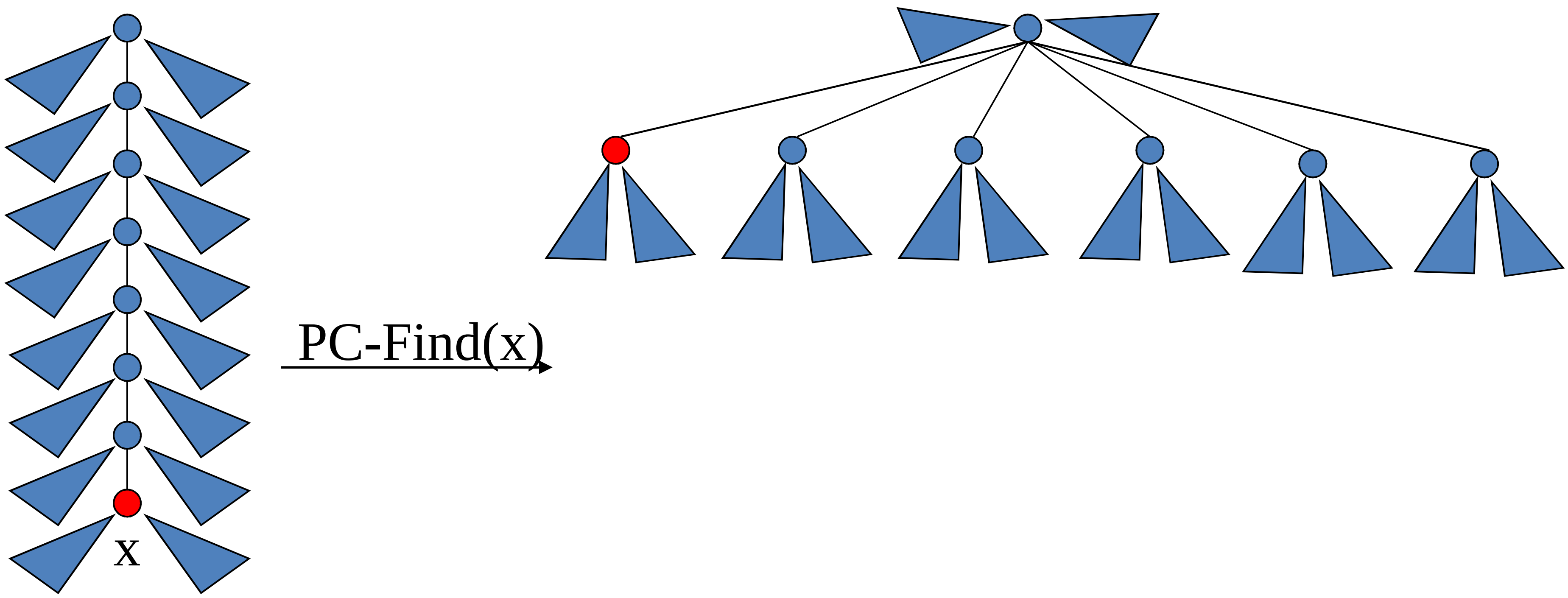
```
W-Union(i,j : index){
//i and j are roots//
  wi := weight[i];
  wj := weight[j];
  if wi < wj then
    up[i] := j;
    weight[j] := wi + wj;
  else
    up[j] :=i;
    weight[i] := wi + wj;
}
```

Path Compression

- On a Find operation point all the nodes on the search path directly to the root.

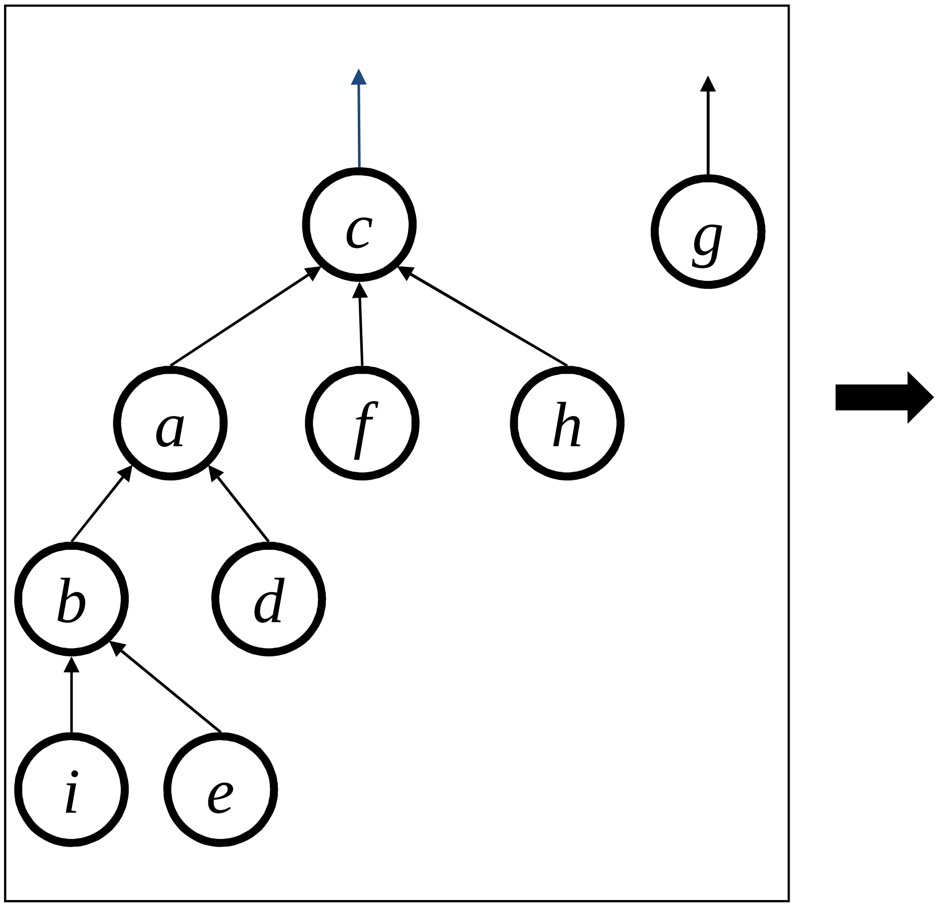


Self-Adjustment Works



Student Activity

Draw the result of Find(e):



Path Compression Find

```
PC-Find(i : index) {
  r := i;
  while up[r] ≠ 0 do //find root//
    r := up[r];
  if i ≠ r then  //compress path//
    k := up[i];
    while k ≠ r do
      up[i] := r;
      i := k;
      k := up[k]
  return(r)
}
```

Interlude: A Really Slow Function

Ackermann's function is a <u>really</u> big function $A(x, y)$ with inverse $\alpha(x, y)$ which is <u>really</u> small

How fast does $\alpha(x, y)$ grow?

$\alpha(x, y) = 4$ for x **far** larger than the number of atoms in the universe (2^{300})

α shows up in:

- Computation Geometry (surface complexity)
- Combinatorics of sequences

A More Comprehensible Slow Function

log* x = number of times you need to compute log to bring value down to at most 1

E.g. log* 2 = 1

log* 4 = log* 2^2 = 2

log* 16 = log* 2^{2^2} = 3 (log log log 16 = 1)

log* 65536 = log* $2^{2^{2^2}}$ = 4 (log log log log 65536 = 1)

log* 2^{65536} = …………… = 5

Take this: $\alpha(m,n)$ grows even slower than log* n !!

Disjoint Union / Find with Weighted Union and PC

- Worst case time complexity for a W-Union is O(1) and for a PC-Find is O(log n).
- Time complexity for $m \geq n$ operations on n elements is O(m log* n)
  - Log * n < 7 for all reasonable n. Essentially constant time per operation!
- Using "ranked union" gives an even better bound theoretically.

Amortized Complexity

- For disjoint union / find with weighted union and path compression.
  - average time per operation is essentially a constant.
  - worst case time for a PC-Find is O(log n).
- An individual operation can be costly, but over time the average cost per operation is not.

Find Solutions

Recursive

```
Find(up[] : integer array, x : integer) : integer {
//precondition: x is in the range 1 to size//
if up[x] = 0 then return x
else return Find(up,up[x]);
}
```

Iterative

```
Find(up[] : integer array, x : integer) : integer {
//precondition: x is in the range 1 to size//
while up[x] ≠ 0 do
  x := up[x];
return x;
}
```