Presentation for use with the textbook, Algorithm Design and Applications, by M. T. Goodrich and R. Tamassia, Wiley, 2015

Union-Find Structures

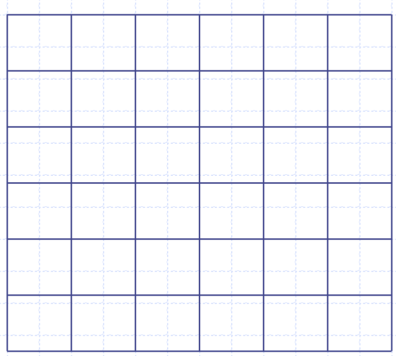
Merging galaxies, NGC 2207 and IC 2163. Combined image from NASA's Spitzer Space Telescope and Hubble Space Telescope. 2006. U.S. government image. NASA/JPL-Caltech/STSci/Vassar.

1

1

Maze Creation

- Build a random maze by erasing edges.

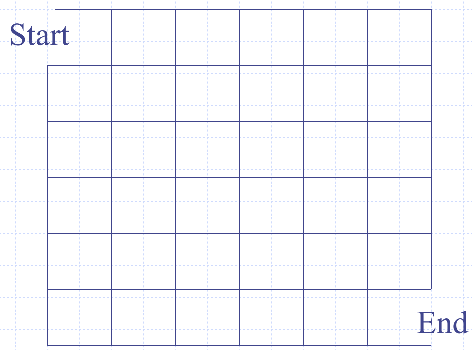


2

2

Maze Creation

- Pick Start and End

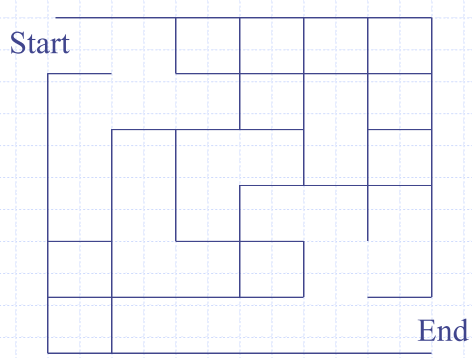


3

3

Maze Creation

- Repeatedly pick random edges to delete.



4

4

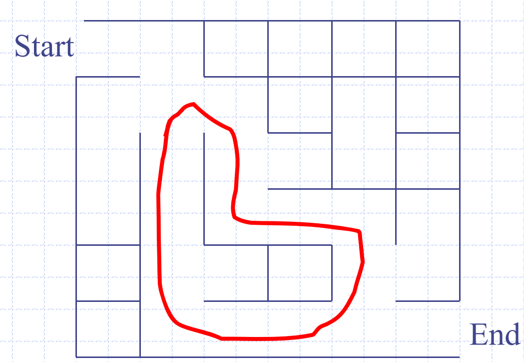
Desired Properties

- None of the boundary is deleted
- Every cell is reachable from every other cell.
- There are no cycles – no cell can reach itself by a path unless it retraces some part of the path.

5

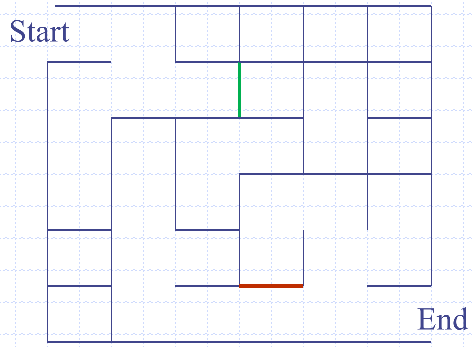
5

A Cycle, not allowed



6

Pick random edges to delete

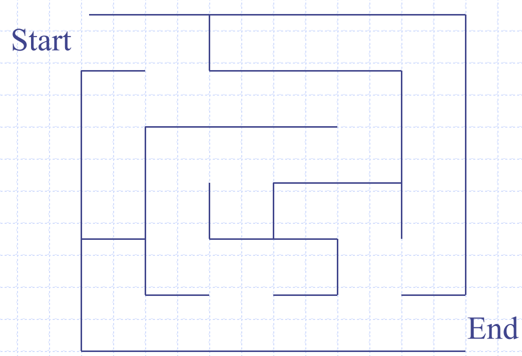


- Green edge can be deleted.
- Red edge cannot be deleted.

7

7

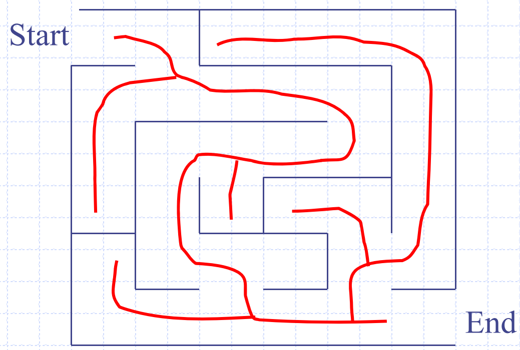
A Good Maze



8

8

A Hidden Tree



9

9

Maze Creation: Algorithm

Cost:
$n = s^2$

1. Given the dimension s of the maze, create an s by s matrix, give a name to each cell of the matrix, identify the beginning and ending cells. — O(n)
2. Collect all the possible edges between the cells, excluding the boundary edges, into E. — O(n)
3. If not all the cells are reachable from each other, randomly pick and remove an edge e from E; otherwise go to 5. — O(n)
4. If the two ends of edge e are already connected by a path, add e into M; otherwise, throw e away, and go to 3. — O(n)
5. Return the union of E and M as the edges of the maze. — O(1)

10

Number the Cells

We have disjoint sets S ={ {1}, {2}, {3}, {4},… {36} }
each cell is a singleton set.
We have all possible edges E ={ (1,2), (1,7), (2,8), (2,3), … }
60 edges total, representing the neighborhood relation.
Boundary edges are excluded.

Start

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36

End

For an s by s matrix, there are $n = s^2$ cells and $2s(s-1)$ edges. We need to delete $s^2 - 1$ edges. There are $(s-1)^2$ edges in the maze.

11

Group Cells into Disjoint Sets

- At any moment, two cells are in the same set if and only if they are connected by a path in the maze.
- An edge (x, y) is safe to delete, if x and y are not in the same set.
- After (x, y) is deleted, the two sets containing x and y, respectively, are joined together.

Disjoint sets are good data structure for implementing equivalence relations.

12

12

Equivalence Relation

Relation *R* on *S* is a subset of *SxS*.

- For every pair of elements *a, b* from a set *S*, *a R b* is either true or false.
- *a R b* is true iff (*a, b*) is in *R*. In this case, we say *a is related* to *b*.

An equivalence relation satisfies:

1. (Reflexive) a *R* a
2. (Symmetric) a *R* b iff b *R* a
3. (Transitive) a *R* b and b *R* c implies a *R* c

13

13

Equivalence Classes

- Given a set of things…
  { grapes, blackberries, plums, apples, oranges, peaches, raspberries, lemons, bananas }
- …define the equivalence relation
  All citrus fruit is related, all berries, all stone fruits, …
- …partition them into related subsets
  { grapes }, { blackberries, raspberries }, { oranges, lemons }, { plums, peaches }, { apples }, { bananas }

Everything belongs to a unique class.
Everything in an equivalence class is related to each other.

14

14

Determining equivalence classes

- Idea: give every equivalence class a name
  - { oranges, limes, lemons } = "like-ORANGES"
  - { peaches, plums } = "like-PEACHES"
  - Etc.
- To answer if two fruits are in the same class:
  - FIND the class name of one fruit.
  - FIND the class name of the other fruit.
  - Are they the same name?

15

15

Building Equivalence Classes

- Start with disjoint, singleton sets:
  - { apples }, { bananas }, { peaches }, …
- As you gain information about the equivalence relation, take UNION of sets that are now related:
  - { peaches, plums }, { limes, oranges, lemons }, { apples }, { bananas }, …
- E.g. if peaches R limes, then we get
  - { peaches, plums, limes, oranges, lemons }

16

16

Disjoint Union - Find

- Maintain a set of pairwise disjoint sets.
  - {3,5,7} , {4,2,8}, {9}, {1,6}
- Each set has a unique name, using one of its members as its name
  - {3,5,7} , {4,2,8}, {9}, {1,6}

17

17

Union

- Union(x, y) – return the union of two sets named by x and y
  - {3,5,7} , {4,2,8}, {9}, {1,6}

Union(5,1)

  - {3,5,7,1,6}, {4,2,8}, {9},

18

18

Find

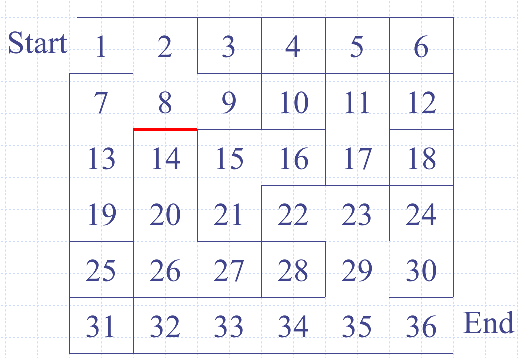
- Find(x) – return the name of the set containing x.
  - {3,5,7,1,6}, {4,2,8}, {9},
  - Find(1) = 5
  - Find(4) = 8

19

19

Example of Deletion

Pick edge (8,14)



S
{1,2,7,8,9,13,19}
{3}
{4}
{5}
{6}
{10}
{11,17}
{12}
{14,20,26,27}
{15,16,21}
.
.
{22,23,24,29,30,32
33,34,35,36}

20

20

Example: After Deletion

S
{1,2,7,8,9,13,19}
{3}
{4}
{5}
{6}
{10}
{11,17}
{12}
{14,20,26,27}
{15,16,21}
.
.
{22,23,24,29,39,32
33,34,35,36}

Find(8) = 7
Find(14) = 20

Union(7,20)

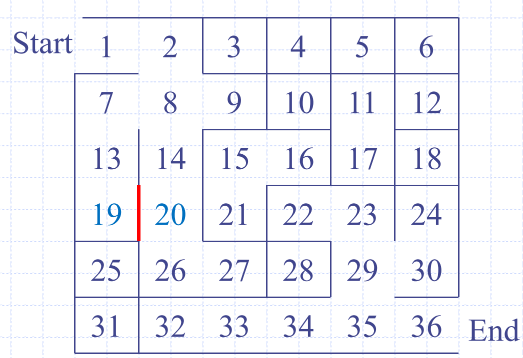
S
{1,2,7,8,9,13,19,14,20 26,27}
{3}
{4}
{5}
{6}
{10}
{11,17}
{12}
{15,16,21}
.
.
{22,23,24,29,39,32
33,34,35,36}

21

21

Example

Pick (19,20)



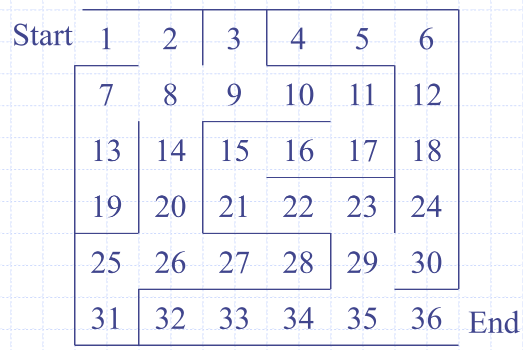
S
{1,2,7,8,9,13,19
14,20,26,27}
{3}
{4}
{5}
{6}
{10}
{11,17}
{12}
{15,16,21}
.
.
{22,23,24,29,39,32
33,34,35,36}

22

22

Example at the End

S
{1,2,3,4,5,6,7,… 36}



23

23

Maze Creation: Algorithm

- S = set of sets of connected cells
- Initially, S = { {1}, {2}, …, {s^2} }
- E = set of edges, representing the neighborhood of each cell.

```
Alg. CreateMaze (S, E) {
  while (|S| > 1) {
      pick a random, unused edge (x,y) from E;
      u = Find(x);
      v = Find(y);
      if (u ≠ v) { Union(u,v); remove (x, y) from E }
      else  mark (x, y) as "used"; /* move (x, y) into M */
  }
  return E;
}  // All remaining members of E form the maze.
```

24

24

Implementing Disjoint Sets

- *n* elements
  Total Cost of: *m* finds, at most *n* -1 unions
- Target complexity: total O(*m*+*n*) *i.e.* O(1) amortized per operation.

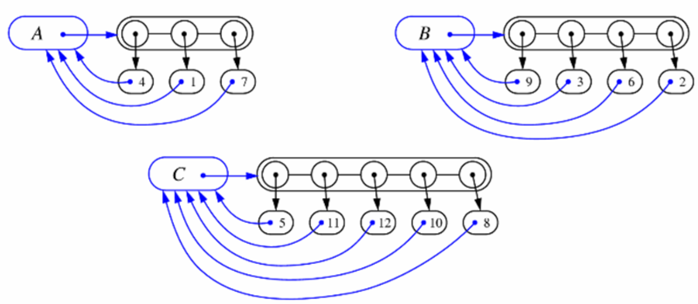
- O(1) worst-case for find as well as union would be great, but it cannot be done.
- *Known result*: find and union *can* be done practically in O(1) time.

25

25

List-based Implementation

- Each set is stored in a sequence represented with a linked-list
- Each node should store an object containing the element and a reference to the set name



26

26

Analysis of List-based Representation

- Worst case time for find is O(1).
- When doing a union, always move elements from the smaller set to the larger set
  - Each time an element is moved it goes to a set of size at least double its old set
  - Thus, an element can be moved at most O(log n) times
- Total time needed to do n – 1 unions and m finds is O(n log n + m).

27

27

Implementing Disjoint Sets

- Observation: *trees* let us find many elements given one root…
- Idea: if we *reverse* the pointers (make them point up from child to parent), we can find a single root from many element.
- Idea: Use one tree for each equivalence class. The name of the class is the tree root.

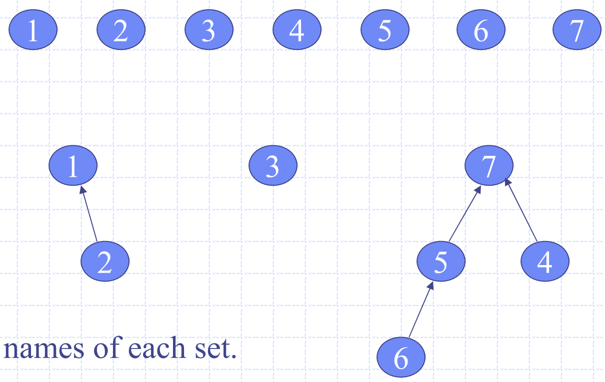
28

28

Up-Tree for Uinon/Find

Initial state

Intermediate state



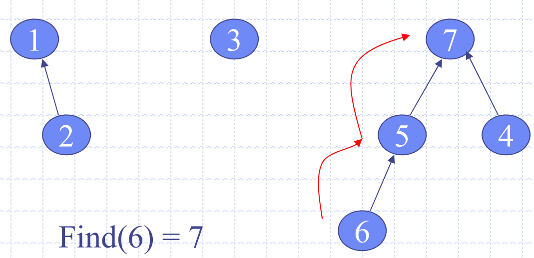
Roots are the names of each set.

29

29

Find Operation

- Find(x) follow x to the root and return the root.
- Cost: O(h), h: height of the tree



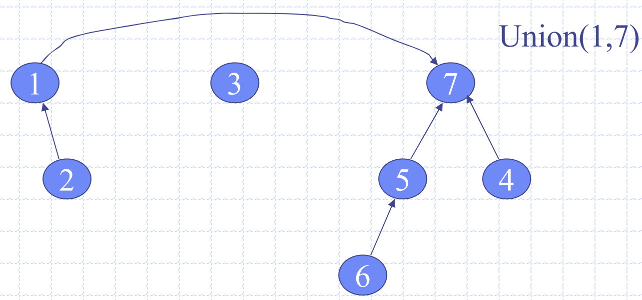
Find(6) = 7

30

30

Union Operation

- Union(i,j) - assuming i and j roots, point i to j.
- Cost: O(1)



31

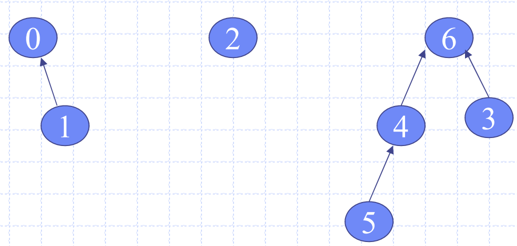
31

Array Implementation of Trees

- Array of indices

	0	1	2	3	4	5	6
Up	-	0	-	6	6	4	-

Up[x] = "-" or "-1", means x is a root.



32

32

Union

```
void Union( int[] Up, int x, int y) {
   //precondition: x and y are roots
   Up[x] = y;
}
```

Constant Time!

33

33

FIND

- Design Find operator
  - Recursive version
  - Iterative version

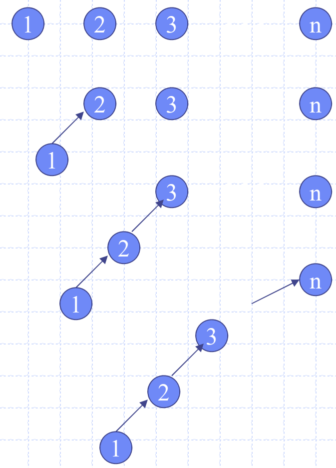
```
static int Find(int[] Up, int x) {
  //Pre: Up[0..(siz-1)] is the parent info;
  // x is in the range 0 to size-1
  if (Up[x] == “-1”) return x;
  return Find(Up[x]);
}
```

Complexity: Depth of x in the tree.

34

34

A Bad Case



Union(1,2)

Union(2,3)

:

:

Union(n-1,n)

Find(1) n steps!!

m finds: O(mn)

35

35

Now this doesn't look good ☹

Can we do better? *Yes!*

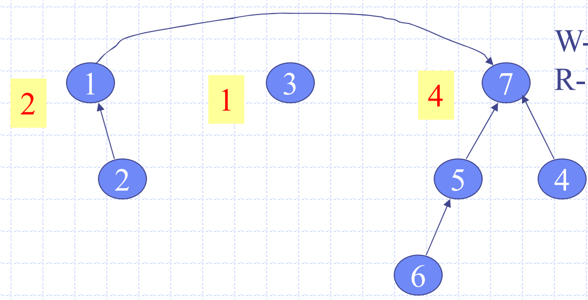
1. Improve union so that find only takes O(log *n*)
   - Union-by-size
   - Union-by-height (height)
   - The cost of m finds is $\Theta(m \log n)$
2. Improve find so that it becomes even better!
   - Path compression
   - Reduces complexity to <u>almost</u> O(1) per operation

36

36

Union by size/height

- Union by size (weight)
  - Always point the smaller tree to the root of the larger tree
- Union by height (rank)
  - Always point the shorter tree to the root of the higher tree

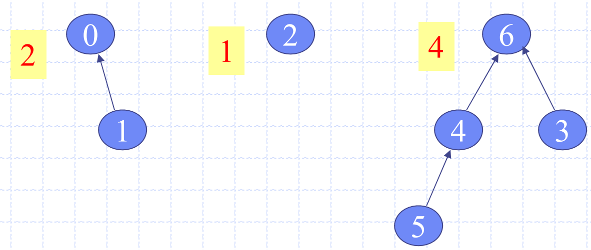


W-Union(1,7)
R-Union(1,7)

37

37

Array Implementation



	0	1	2	3	4	5	6
Up	-	0	-	6	6	4	-
size	2		1				4

	0	1	2	3	4	5	6
Up	-	0	-	6	6	4	-
height	1		0				2

38

38

Union by size

```
void W_Union(int i,j){
  //Pre: i and j are roots//
  int wi = size[i];
  int wj = size[j];
  if (wi < wj) {
    Up[i] = j;
    size[j] = wi + wj;
  } else {
    Up[j] = i;
    size[i] = wi + wj;
  }
}
```

Computing time?

39

39

Union by height

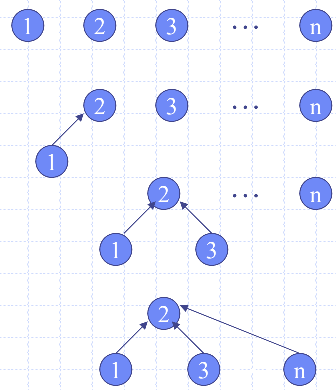
```
void R_Union(int i,j){
  //Pre: i and j are roots//
  int ri = height[i];
  int rj = height[j];
  if (ri < rj) {
    Up[i] = j;
  } if (ri > rj) {
    Up[j] = i;
  } else { // ri == rj
    height[j]++; Up[j] = i;
  }
}
```

Computing time?

40

40

Example Again



Union(1,2)

Union(2,3)

:

:

Union(n-1,n)

Find(1) constant time

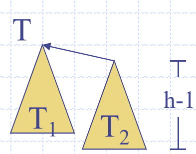
41

41

Analysis of Union by size/height

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

A tree T of height h must have a child T_2 of height h-1



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

Union by size/height

Induction hypothesis

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

42

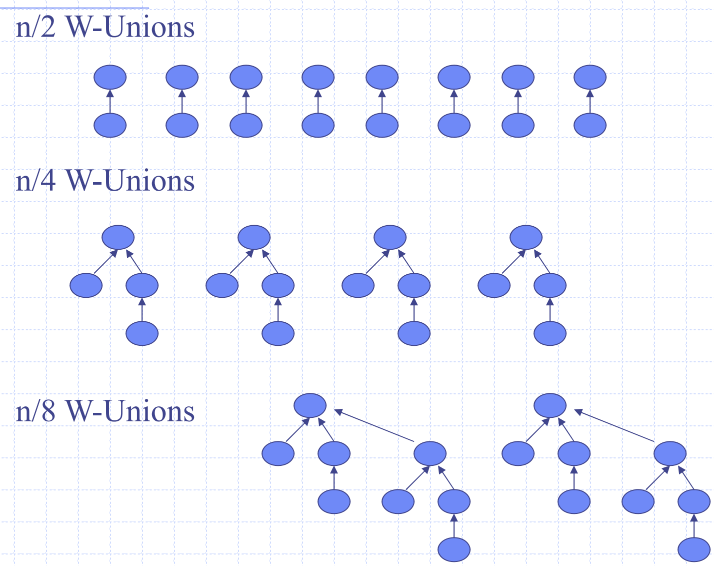
Analysis of Union by size/height

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

43

43

Worst Case for Union by size/height



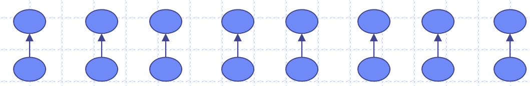
44

44

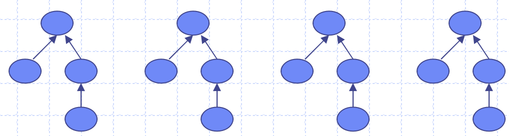
Binomial Trees

A single node is a binomial tree.
If two binomial trees have the same height (or size), the union of them is also a binomial tree.

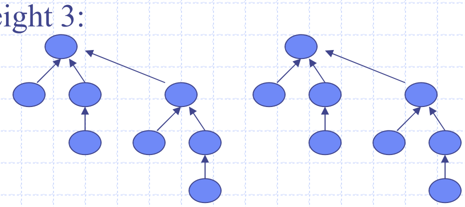
Binomial trees of height 1:



Binomial trees of height 2:



Binomial trees of height 3:



45

45

Binomial Trees

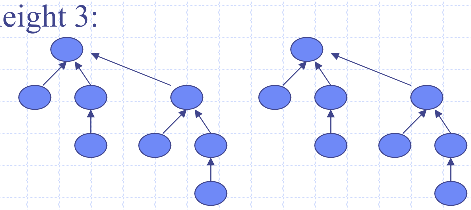
A single node is a binomial tree.
If two binomial trees have the same height (or size), the union of them is also a binomial tree.

Given a binomial tree T of height h:

How many nodes in T? 2^h

How many nodes at depth d in T? $C(h, d) = h!/(d!(h-d)!)$

Binomial trees of height 3:

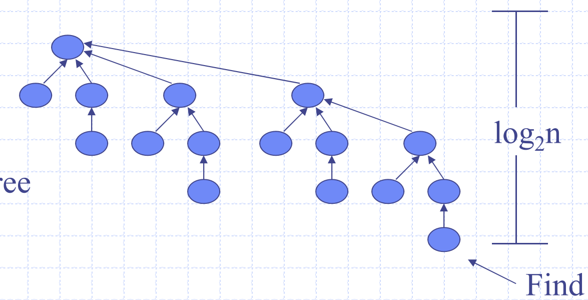


46

46

Example of Worst Cast (cont')

After n – 1 = n/2 + n/4 + …+ 1 Unions



A binomial tree

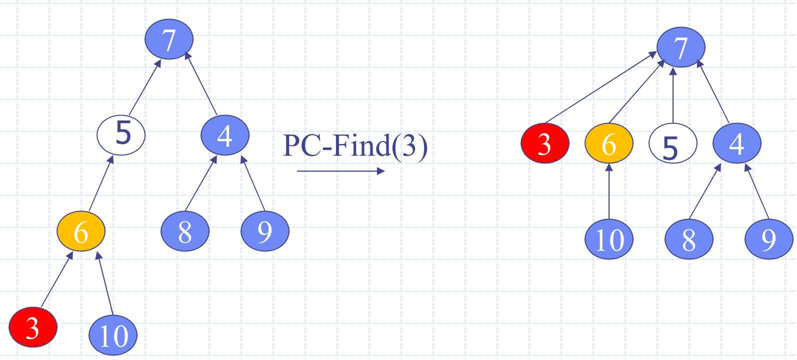
If there are $n = 2^k$ nodes then the longest path from leaf to root has length $k = \log_2(n)$.

47

47

Path Compression

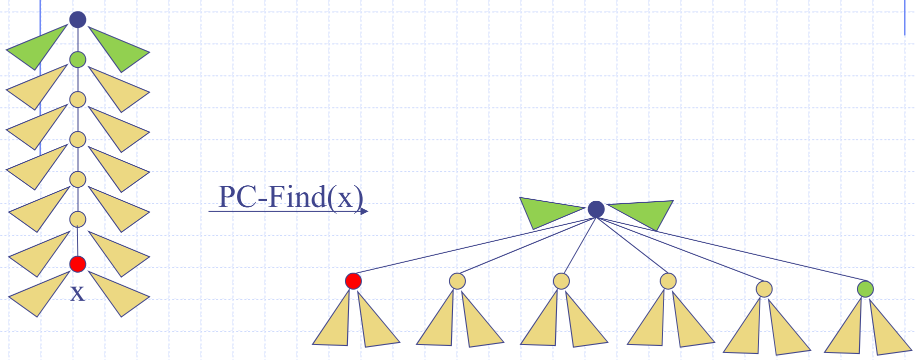
- On a Find operation point all the nodes on the search path directly to the root.



48

48

Self-Adjustment Works

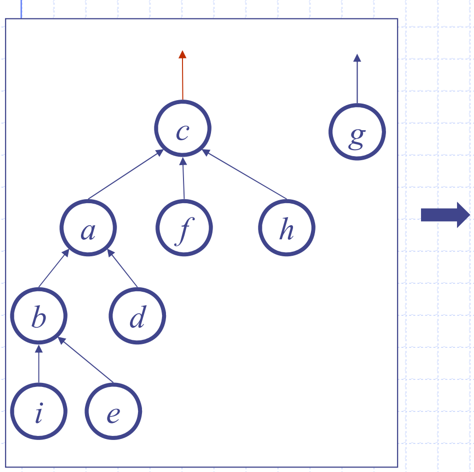


49

49

Possible Quiz Question:

Draw the result of Find(e)



50

50

Path Compression Find

```
int PC_Find(int i) {
  int r = i;
  while (Up[r] != -1) //find root
    r = Up[r];
  if (i != r) { //compress path//
    int k = Up[i];
    while (k != r) {
      Up[i] = r;
      i = k;
      k = Up[k];
    }
  }
  return r;
}
```

51

51

Function Definition

Ackermann's function was defined in 1920s by German mathematician and logician Wilhelm Ackermann (1896-1962).

A(m,n), m,n ∈ **N** such that**,**

A(0, n) = n + 1, n≥ 0;
A(m,0) = A(m-1, 1), m > 0;
A(m,n) = A(m-1, A(m, n-1)), m, n > 0;

52

52

$A(0, n) = n + 1, \quad n \geq 0;$
$A(m,0) = A(m-1, 1), \quad m > 0;$
$A(m,n) = A(m-1, A(m, n-1)), \quad m, n > 0;$

$A(1, 0) = A(0, 1) = 2$
$A(1, 1) = A(0, A(1, 0)) = A(0, 2) = 3$
$A(1, 2) = A(0, A(1, 1)) = A(0, 3) = 4$
$A(1, n) = n + 2$

$A(2, 0) = A(1, 1) = 3$
$A(2, 1) = A(1, A(2, 0)) = A(1, 3) = 5$
$A(2, 2) = A(1, A(2, 1)) = A(1, 5) = 7$
$A(2, n) = 2n + 3$

Simple addition and subtraction!!

$A(3, 0) = A(2, 1) = 5$
$A(3, 1) = A(2, A(3, 0)) = A(2, 5) = 13$
$A(3, 2) = A(2, A(3, 1)) = A(2, 13) = 29$
$A(3, n) = 2^{n+3} – 3$

$A(4, 0) = A(3, 1) = 13$
$A(4, 1) = A(3, A(4, 0)) = A(3, 13) = 65533$
$A(4, 2) = A(3, A(4, 1)) = 2^{65536} – 3$
$A(4, 3) = A(3, A(4, 2)) = 2^{A(4,2)+3} – 3$
$A(4, n) = 2\uparrow\uparrow(n + 3) – 3$

$A(5, n) = 2\uparrow\uparrow\uparrow(n + 3) – 3$
$A(6, n) = 2\uparrow\uparrow\uparrow\uparrow(n + 3) – 3$

53

53

Equivalent Definition

$A(0, n) = n + 1$
$A(1, n) = 2 + (n + 3) - 3$
$A(2, n) = 2 \times (n + 3) - 3$
$A(3, n) = 2^{n + 3} – 3$
$A(4, n) = \underbrace{2^{2^{2^{\cdots 2}}}}_{(n + 3 \text{ terms})} – 3$

…

Terms of the form $2^{2^{2^{\cdots 2}}}$ are known as power towers.
It is a well defined total function that grows so fast.

54

54

Inverse of Ackermann's Function

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

$\alpha(x, y)$ is a really slowly growing function.

How slow 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

55

55

Disjoint Union / Find with Union by size/height and Path Compression

- Worst case time complexity for a W-Union/R-Union is O(1) and for a PC-Find is O(log n).
- The total time complexity for $m \geq n$ operations on n elements is O(m $\alpha(m, n)$)
  - $\alpha(m, n)$ <= 4 for all reasonable n. Essentially constant time per operation!

56

56

Amortized Complexity

- For disjoint union / find with union by size/height and path compression.
  - Amortized time per operation is essentially a constant.
  - Worst case time for a single union is O(1).
  - Worst case time for a single PC-Find is O(log n).
- An individual operation can be costly, but over time the average cost per operation is not.

57

57

Basic Algorithm

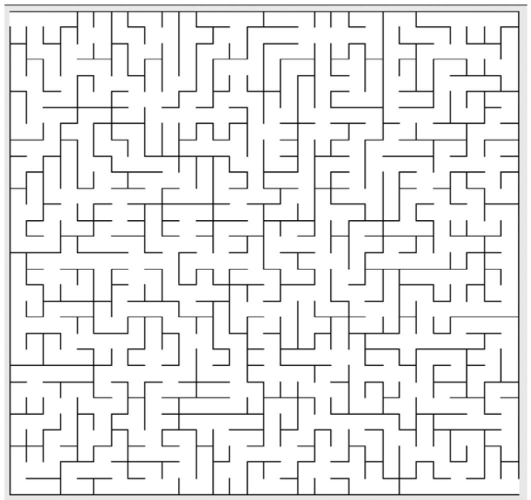
- S = set of sets of connected cells
- Initially, S = { {1}, {2}, …, {n^2} }
- E = set of edges, representing the neighborhood of each cell.

```
Alg. CreateMaze (S, E) {
  while (|S| > 1) {
        pick a random, unused edge (x,y) from E;
        u = Find(x);
        v = Find(y);
        if (u ≠ v) { Union(u,v); remove (x, y) from E }
        else  mark (x, y) as "used";
  }
  return E;
}  // All remaining members of E form the maze.
```

58

58

A larger size maze



59

59

A Maze Generator

Algorithm MazeGenerator(G, E):

Input: A grid, G, consisting of n cells and a set, E, of m "walls," each of which divides two cells, x and y, such that the walls in E initially separate and isolate all the cells in G

Output: A subset, R of E, such that removing the edges in R from E creates a maze defined on G by the remaining walls

while R has fewer than $n - 1$ edges **do**
  Choose an edge, (x, y), in E uniformly at random from among those previously unchosen
  if find$(x) \neq$ find(y) **then**
    union(find(x), find(y))
    Add the edge (x, y) to R
return R

60

60