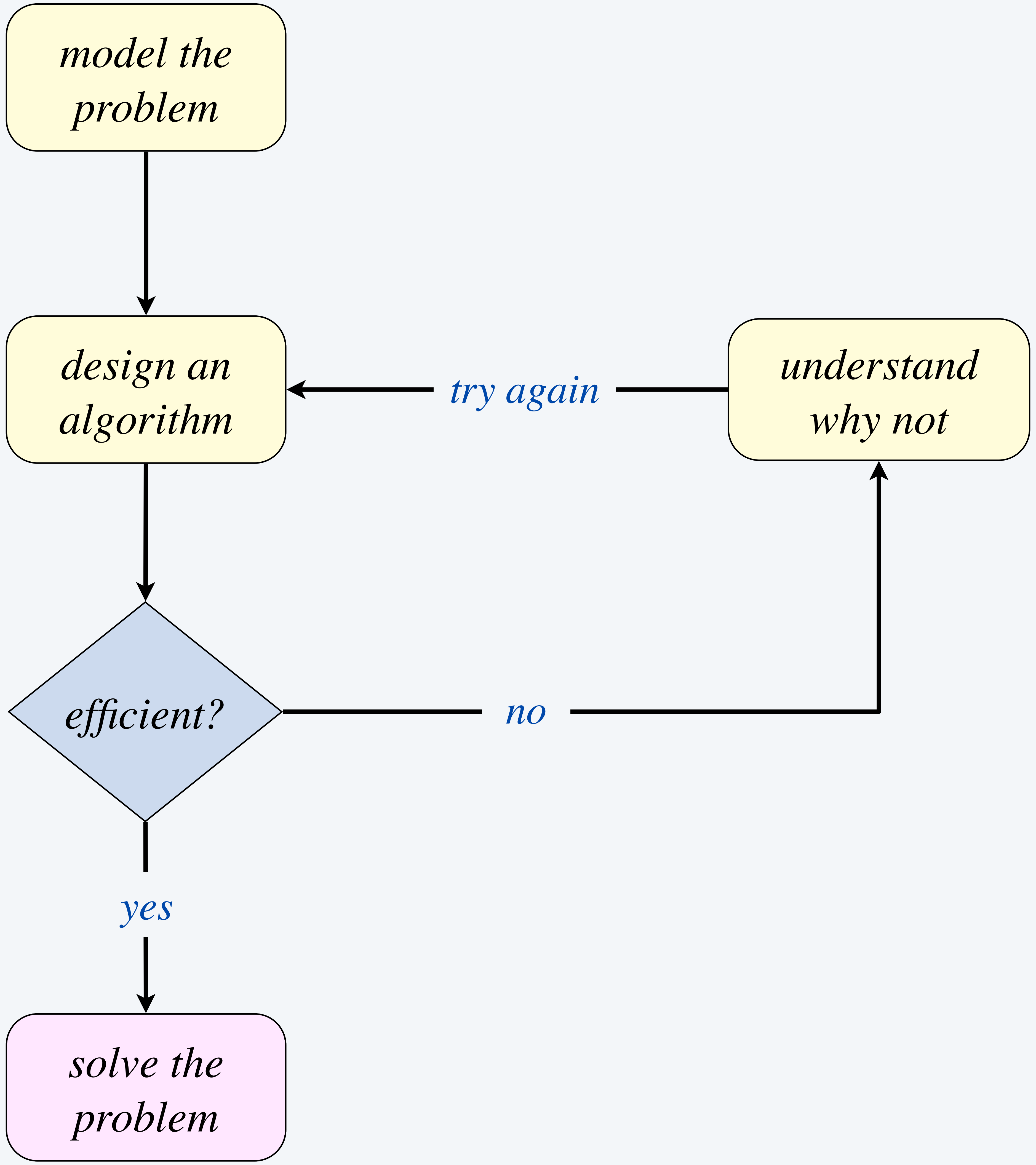
https://algs4.cs.princeton.edu

1.5 Union–Find

- *union–find data type*
- *quick-find*
- *quick-union*
- *weighted quick-union*

Subtext of today's lecture (and this course)

Steps to develop a usable algorithm to solve a computational problem.



https://algs4.cs.princeton.edu

1.5 Union–Find

- union–find data type
- quick-find
- quick-union
- weighted quick-union
- percolation

Union–find data type

Disjoint sets. A collection of sets containing n elements, with each element in exactly one set.

Leader. Each set designates one of its elements as leader (to uniquely identify it).

no restriction on which element is designated leader (but leader of a set can't change unless the set changes)

Find. Return the leader of the set containing element p. ← *main use case: are two elements in the same set ?*

Union. Merge the set containing element p with the set containing element q.

leader is 0   *leader is 4*   **find(1) = 4**, **find(4) = 4**, **find(5) = 4**   *leader is 6*

`{ 0 } { 1, 4, 5 } { 2, 3, 6, 7 }`

8 elements, 3 disjoint sets

union(2, 5) ⟹

leader is 0   *leader is 6*   **find(1) = 6**, **find(2) = 6**, **find(3) = 6**, **find(4) = 6**, **find(5) = 6**, **find(6) = 6**, **find(7) = 6**

`{ 0 } { 1, 2, 3, 4, 5, 6, 7 }`

2 disjoint sets

Union–find data type: API

Goal. Design an efficient union-find data type.

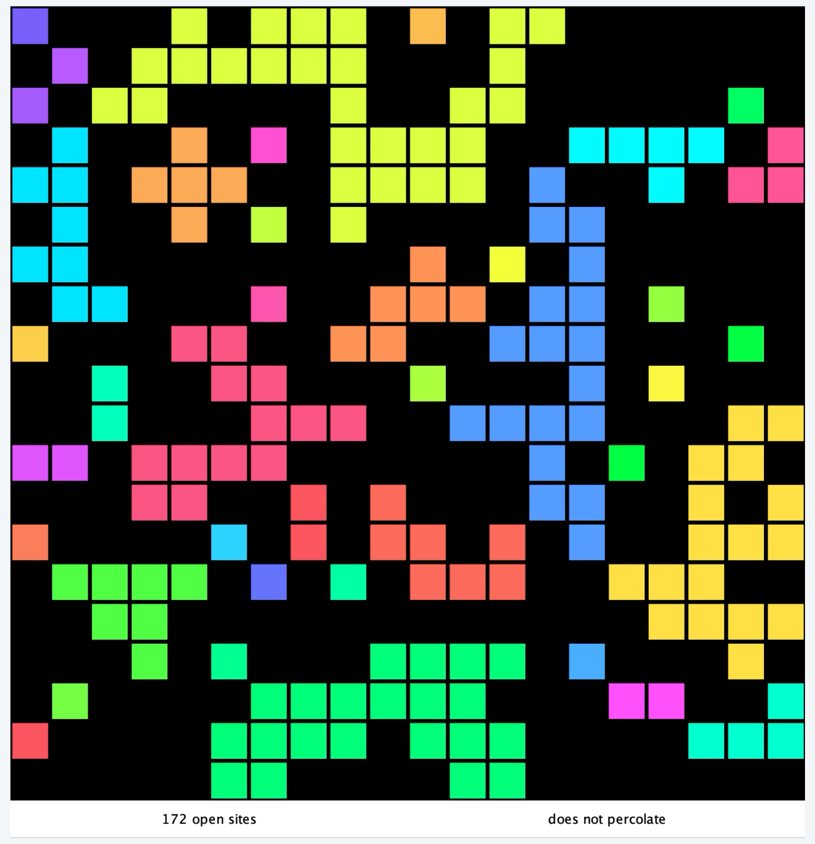
- Simplifying assumption: the n elements are named $0, 1, 2, \ldots, n - 1$.
- The `union()` and `find()` operations can be intermixed.
- Number of elements n can be huge.
- Number of operations m can be huge.

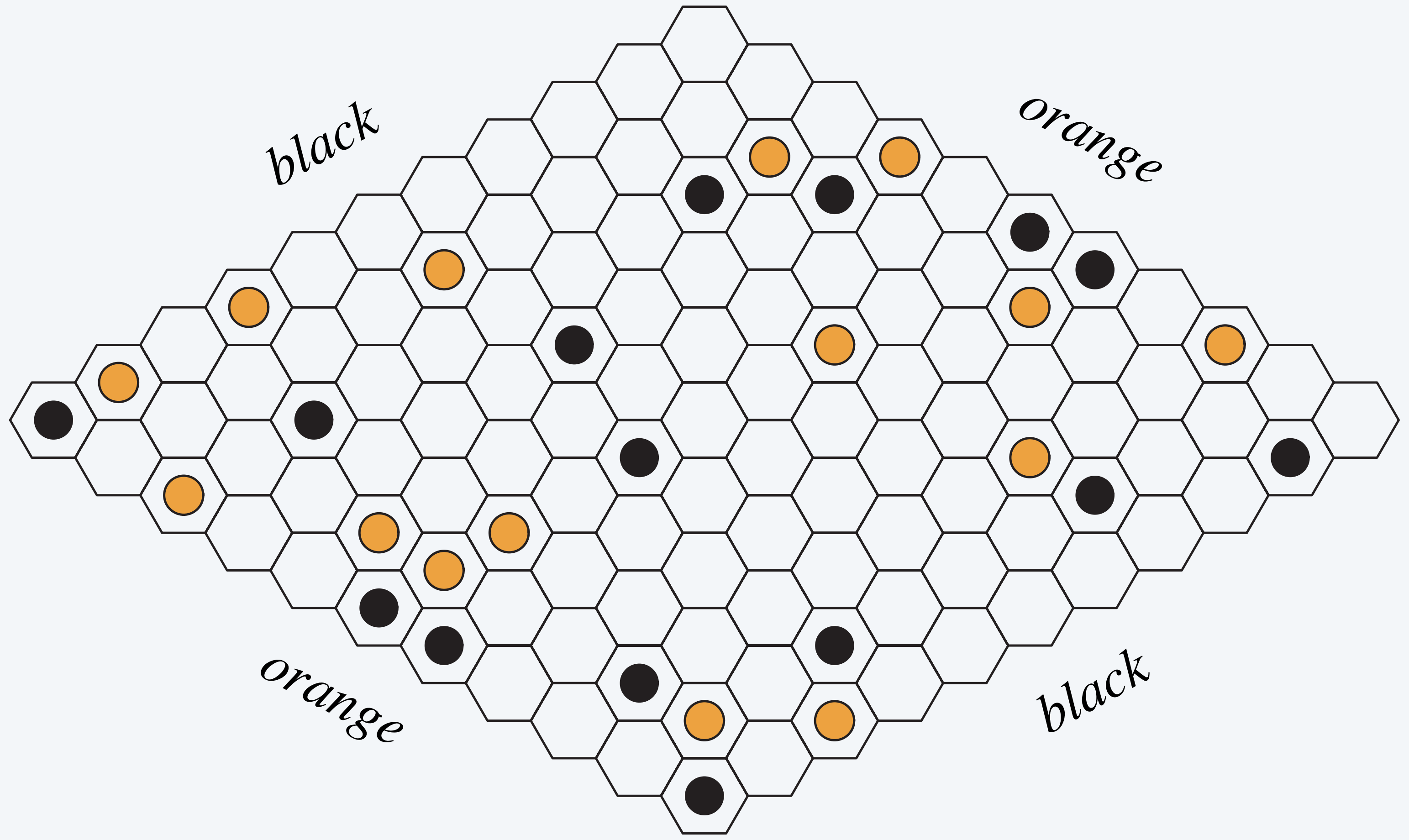
public class UF		**description**
	`UF(int n)`	*initialize with n singleton sets (0 to n – 1)*
`void`	`union(int p, int q)`	*merge sets containing elements p and q*
`int`	`find(int p)`	*return the leader of set containing element p*

Union–find data type: applications

Disjoint sets can represent:

- Clusters of conducting sites in a composite system. ⟵ *see Assignment 1 (Percolation)*
- Connected components in a graph. ⟵ *see Kruskal's algorithm (MST lecture)*
- Interlinked friends in a social network.
- Interconnected devices in a mobile network.
- Equivalent variable names in a Fortran program.
- Adjoining stones of the same color in the game of Hex.
- Contiguous pixels corresponding to same feature in a digital image.





https://algs4.cs.princeton.edu

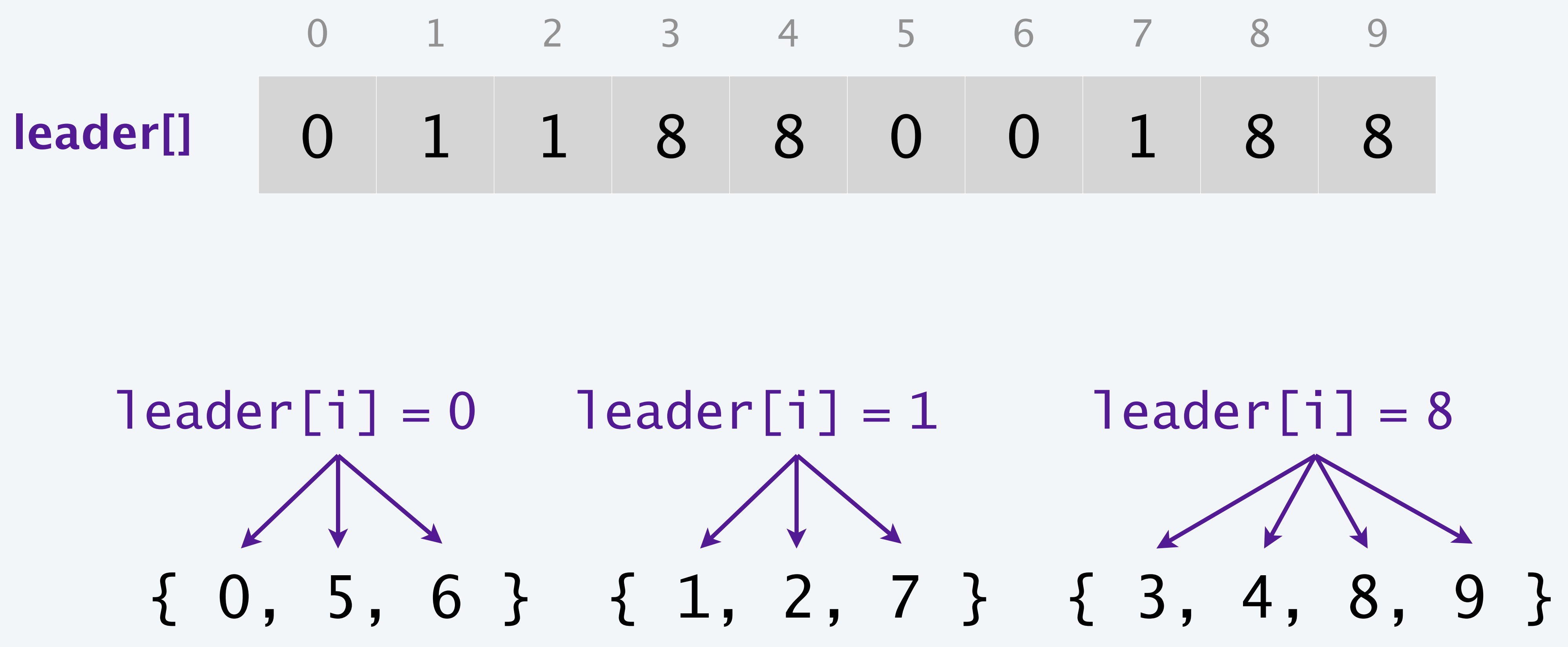
1.5 Union–Find

- *union–find data type*
- ***quick-find***
- *quick-union*
- *weighted quick-union*

Quick-find

Data structure.

- Integer array `leader[]` of length `n`.
- Interpretation: `leader[i]` is the leader of the set containing element `i`.



10 elements, 3 disjoint sets

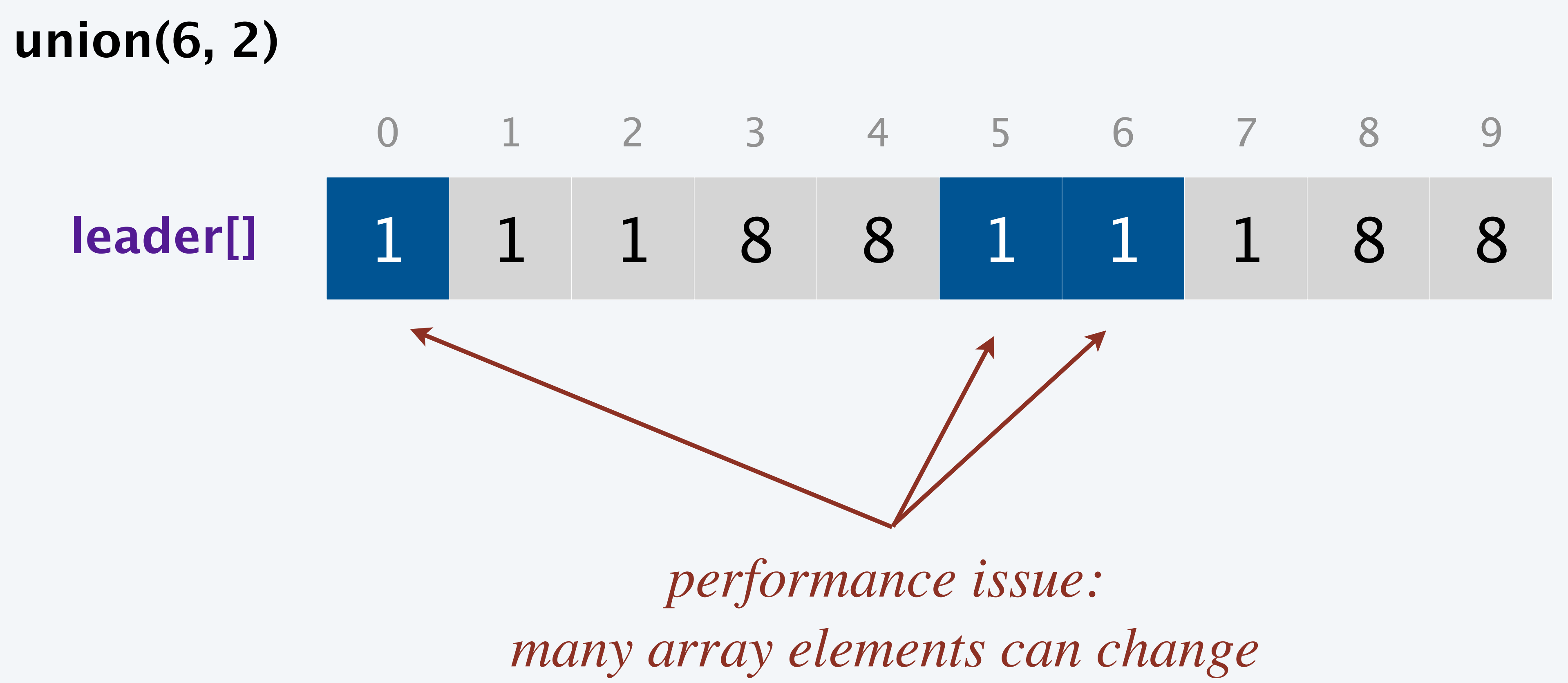
Q. How to implement `find(p)`?

A. Easy, just return `leader[p]`.

Quick-find

Data structure.

- Integer array `leader[]` of length `n`.
- Interpretation: `leader[i]` is the leader of the set containing element `i`.



Q. How to implement `union(p, q)`?

A. Change all array elements whose value is `leader[p]` to `leader[q]`. ← *or vice versa*

Quick-find: Java implementation

```
public class QuickFindUF {
   private int[] leader;

   public QuickFindUF(int n) {
      leader = new int[n];
      for (int i = 0; i < n; i++)
         leader[i] = i;
   }

   public int find(int p) {
      return leader[p];
   }

   public void union(int p, int q) {
      int leaderP = leader[p];
      int leaderQ = leader[q];
      for (int i = 0; i < leader.length; i++)
         if (leader[i] == leaderP)
            leader[i] = leaderQ;
   }

}
```

initialize leader of each element to itself (n array accesses)

return the leader of p (1 array access)

change all array elements whose value is `leader[p]` *to* `leader[q]` *(≥ n array accesses)*

https://algs4.cs.princeton.edu/15uf/QuickFindUF.java.html

Quick-find is too slow

Cost model. Number of array accesses (for read or write).

algorithm	initialize	union	find
quick-find	n	n	1

worst-case number of array accesses (ignoring leading coefficient)

Union is too expensive. Processing any sequence of m `union()` operations on n elements takes $\geq$ mn array accesses.

quadratic in input size!

Ex. Performing 10^9 `union()` operations on 10^9 elements might take 30 years.

https://algs4.cs.princeton.edu

1.5 Union–Find

- union–find data type
- quick-find
- **quick-union**
- weighted quick-union

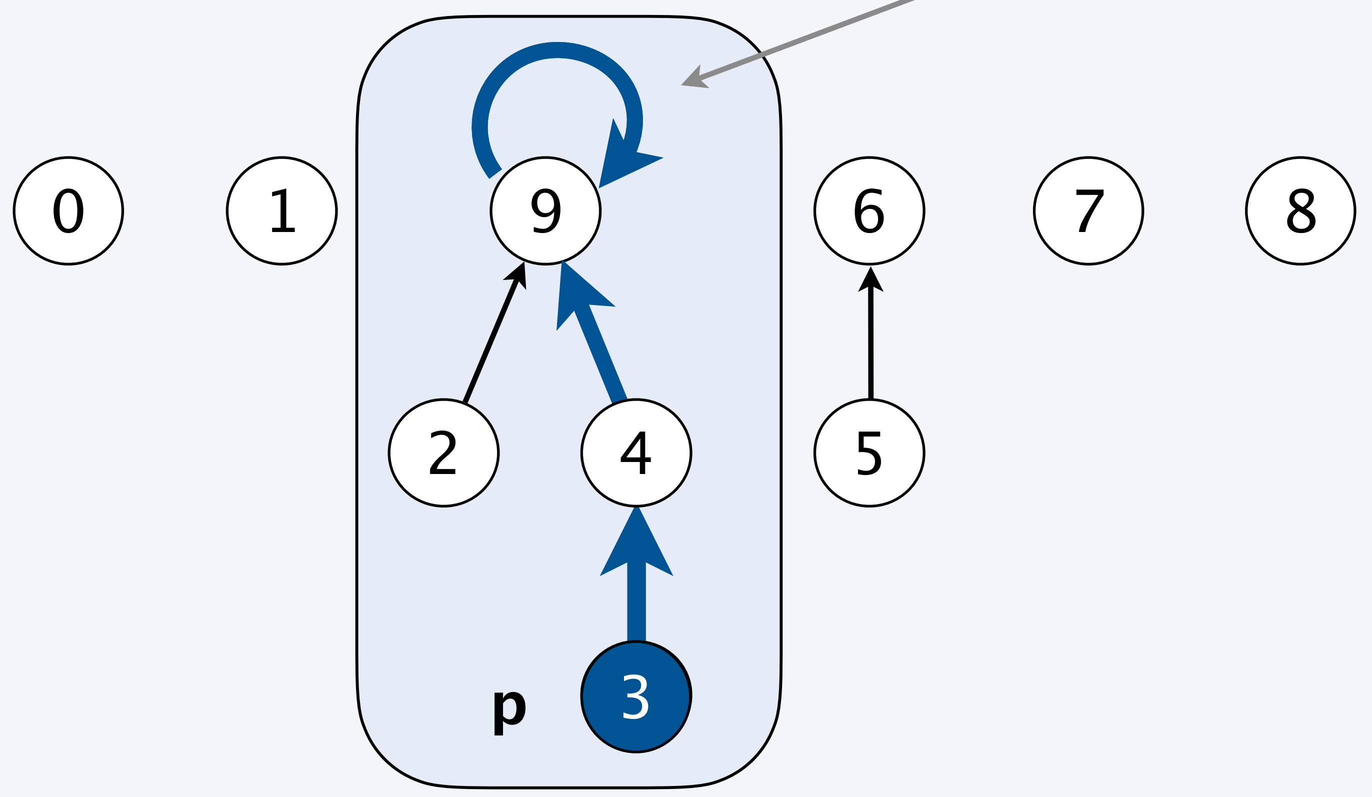
Quick-union

Data structure: Forest-of-trees.

- Interpretation: elements in one rooted tree correspond to one set.
- Integer array `parent[]` of length `n`, where `parent[i]` is parent of element `i` in tree.

by convention, we make root point to itself (but typically suppress from drawing)

	0	1	2	3	4	5	6	7	8	9
parent[]	0	1	9	4	9	6	6	7	8	9



`find(i)` = 9

{ 0 } { 1 } { 2, 3, 4, 9 } { 5, 6 } { 7 } { 8 }

10 elements, 6 disjoint sets (6 trees)

parent of 3 is 4
parent of 4 is 9
parent of 9 is 9
root of 3 is 9

Q. How to implement `find(p)`?

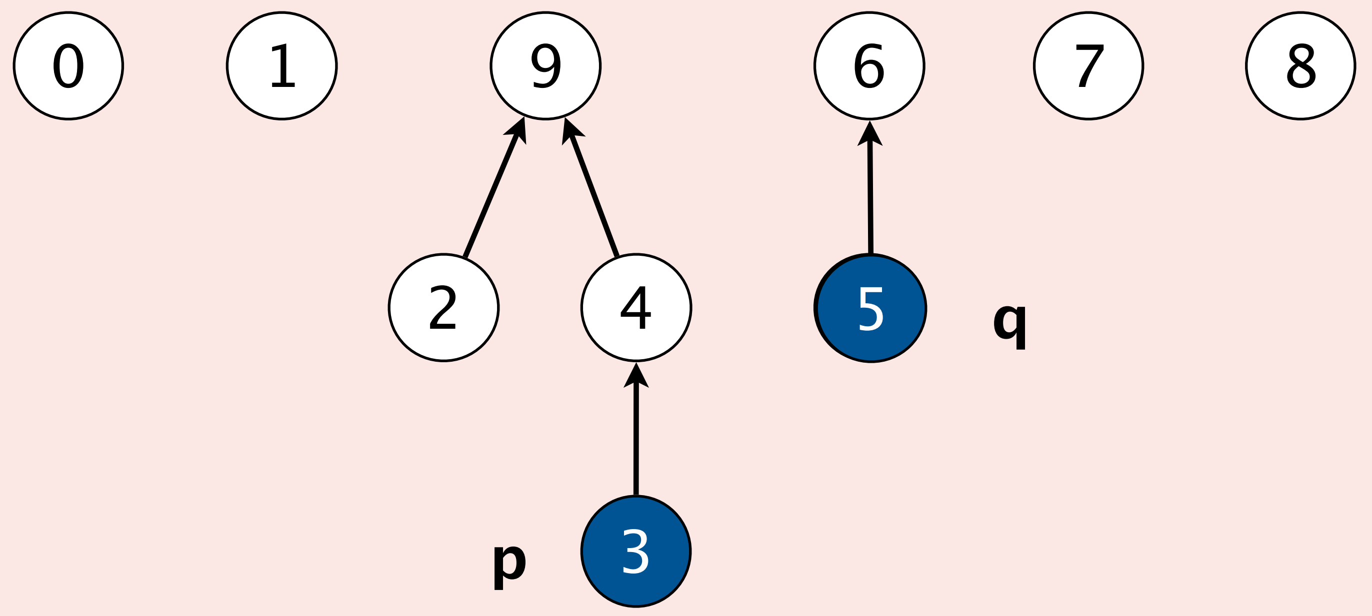
A. Use tree roots as leaders $\Longrightarrow$ return root of tree containing `p`.

Union-find: poll 1

Data structure: Forest-of-trees.

- Interpretation: elements in one rooted tree correspond to one set.
- Integer array `parent[]` of length `n`, where `parent[i]` is parent of element `i` in tree.

	0	1	2	3	4	5	6	7	8	9
parent[]	0	1	9	4	9	6	6	7	8	9



Which is not a valid way to implement `union(3, 5)` ?

A. Set `parent[6] = 9`.

B. Set `parent[9] = 6`.

C. Set `parent[3] = 5`.

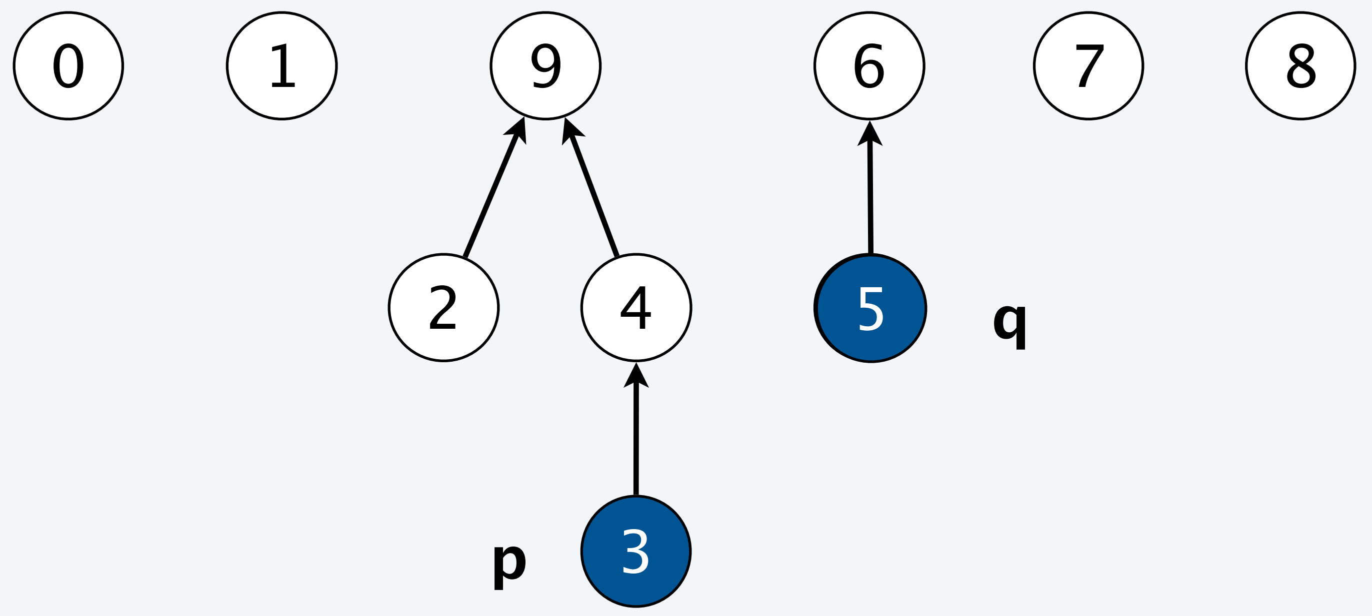
D. Set `parent[2] = parent[3] = parent[4] = parent[9] = 6`.

Quick-union

Data structure: Forest-of-trees.

- Interpretation: elements in one rooted tree correspond to one set.
- Integer array `parent[]` of length `n`, where `parent[i]` is parent of element `i` in tree.

	0	1	2	3	4	5	6	7	8	9
union(3, 5)	0	1	9	4	9	6	6	7	8	9



Q. How to implement `union(p, q)`?

A. Set `parent`[`p`'s root] = `q`'s root. ← *or vice versa*

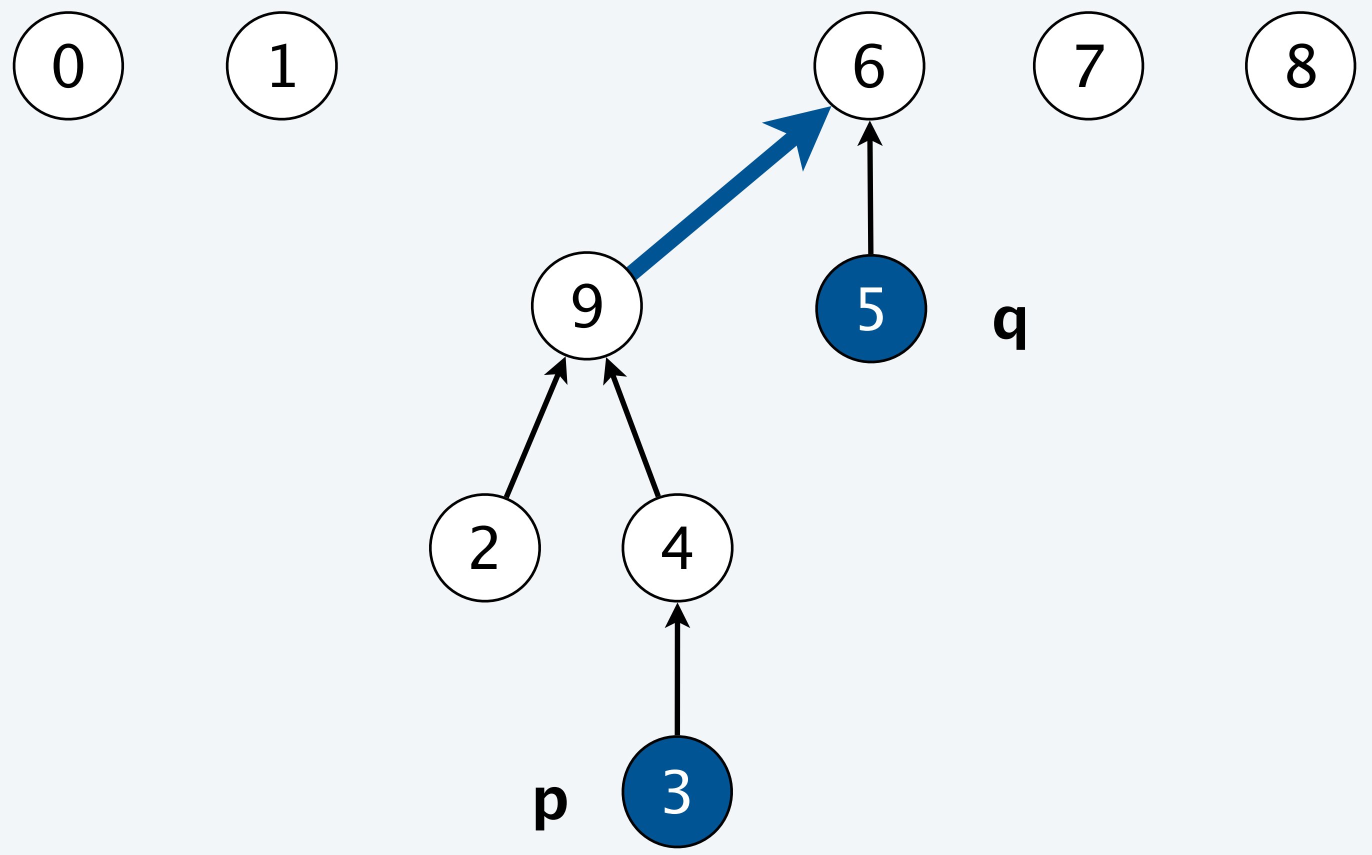
Quick-union

Data structure: Forest-of-trees.

- Interpretation: elements in one rooted tree correspond to one set.
- Integer array `parent[]` of length `n`, where `parent[i]` is parent of element `i` in tree.

	0	1	2	3	4	5	6	7	8	9
union(3, 5)	0	1	9	4	9	6	6	7	8	6

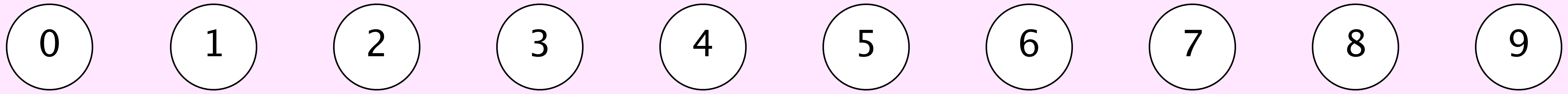
only one array element changes



Q. How to implement `union(p, q)`?

A. Set `parent`[`p`'s root] = `q`'s root. ← *or vice versa*

Quick-union demo



0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9

Quick-union: Java implementation

```
public class QuickUnionUF {
   private int[] parent;

   public QuickUnionUF(int n) {
      parent = new int[n];
      for (int i = 0; i < n; i++)
         parent[i] = i;
   }

   public int find(int p) {
      while (p != parent[p])
         p = parent[p];
      return p;
   }

   public void union(int p, int q) {
      int rootP = find(p);
      int rootQ = find(q);
      parent[rootP] = rootQ;
   }

}
```

set parent of each element to itself (to create forest of n singleton trees)

follow parent pointers until reach root; return resulting root

link root of p to root of q

https://algs4.cs.princeton.edu/15uf/QuickUnionUF.java.html

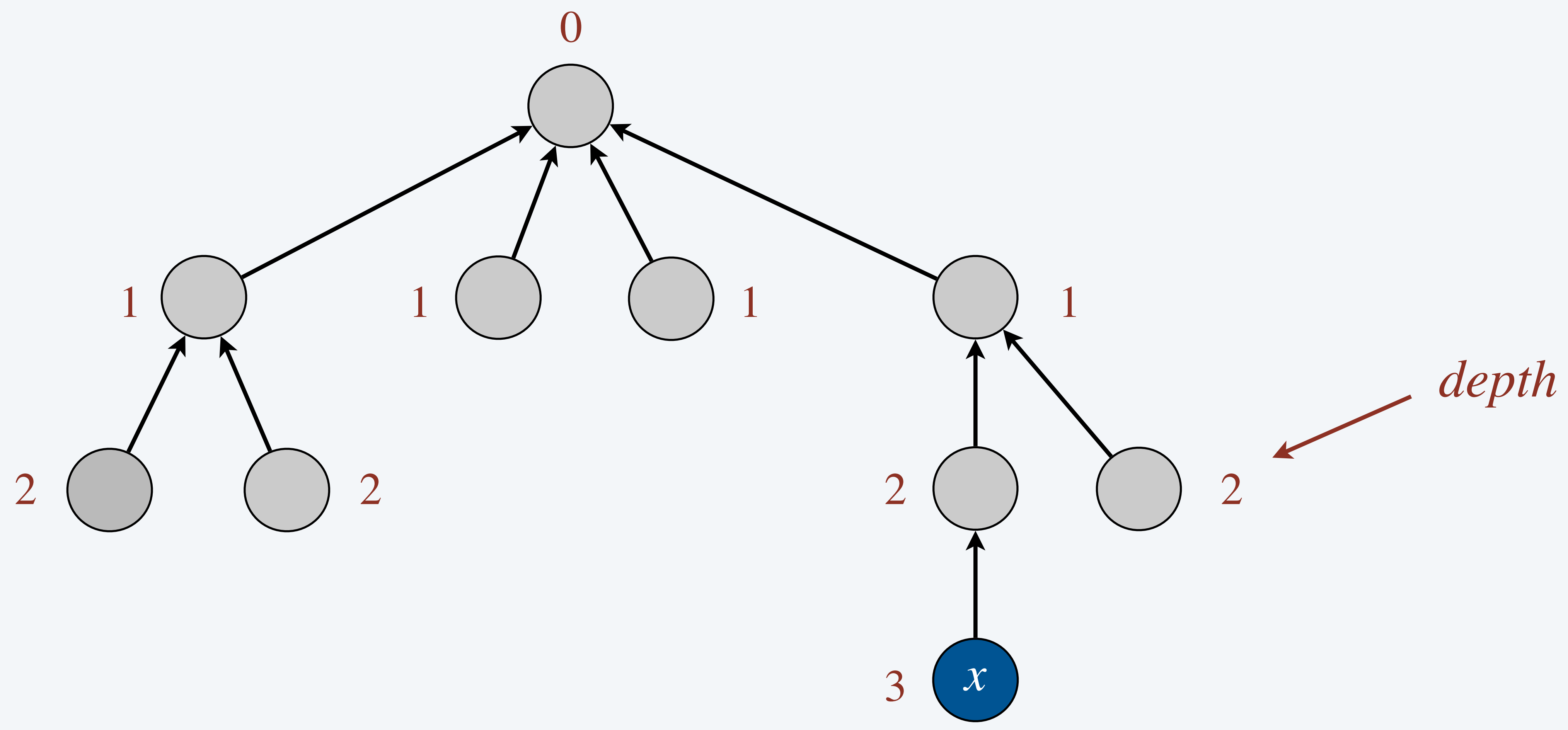
Quick-union analysis

Cost model. Number of array accesses (for read or write).

Running time.

- `union()` takes constant time, given two roots.
- `find()` takes time proportional to depth of node in tree.

 # links on path from node to root



depth(x) = 3

Quick-union analysis

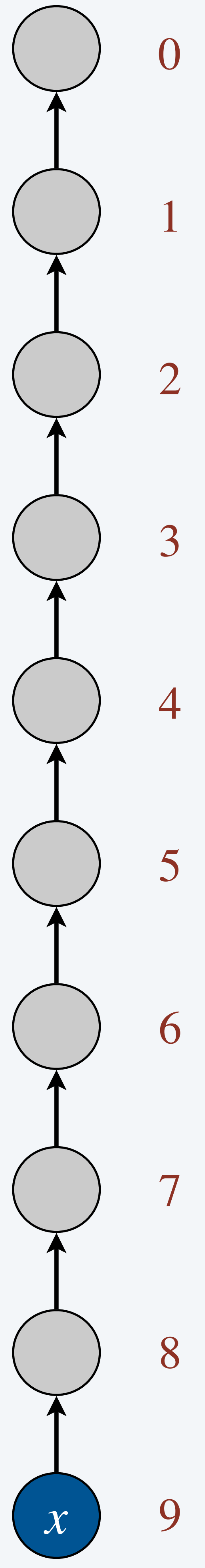
Cost model. Number of array accesses (for read or write).

Running time.

- `union()` takes constant time, given two roots.
- `find()` takes time proportional to depth of node in tree.

algorithm	initialize	union	find
quick-find	n	n	1
quick-union	n	n	n

worst-case number of array accesses (ignoring leading coefficient)



worst-case depth = n–1

Union and find are too expensive (if trees get tall). Processing some sequences of m `union()` and `find()` operations on n elements takes $\geq$ mn array accesses.

quadratic in input size !

https://algs4.cs.princeton.edu

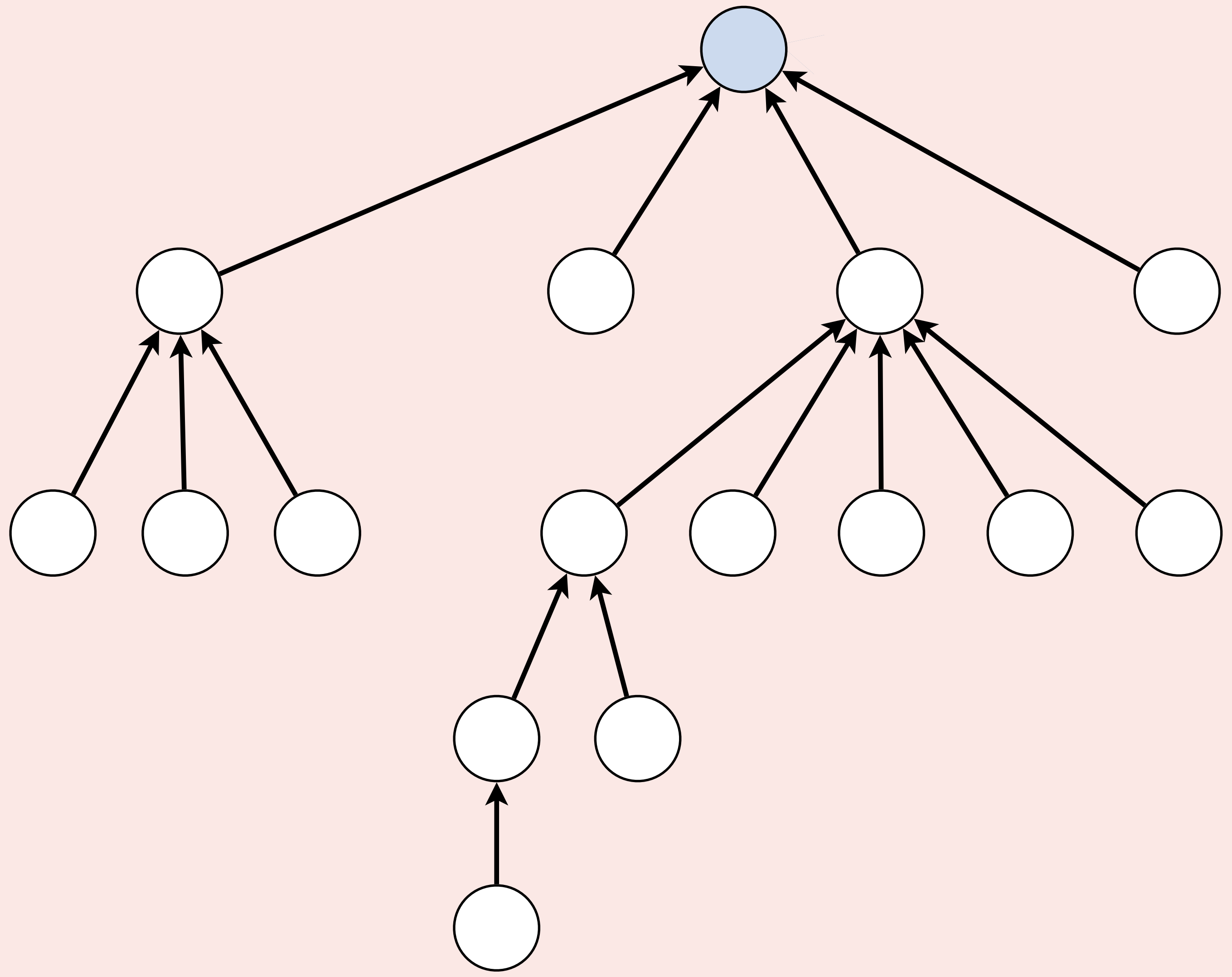
1.5 Union–Find

- union–find data type
- quick-find
- quick-union
- *weighted quick-union*

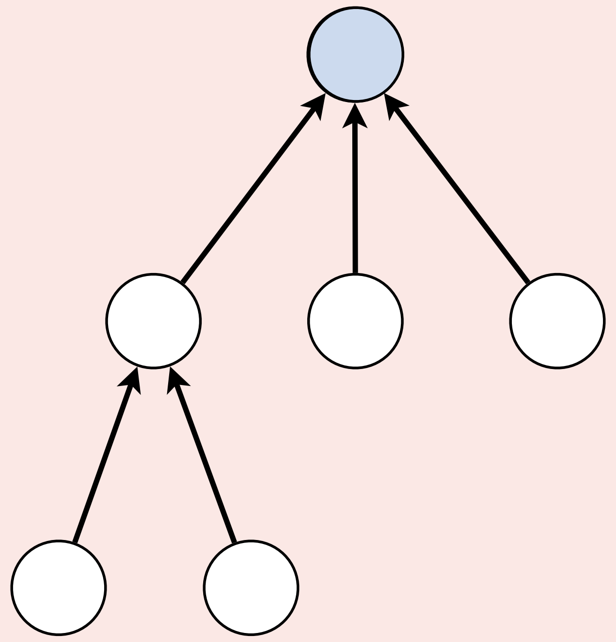
Union–find: poll 2

When linking two trees, which of these strategies is most effective?

- **A.** Link the root of the smaller tree to the root of the larger tree.
- **B.** Link the root of the larger tree to the root of the smaller tree.
- **C.** Flip a coin; randomly choose between A and B.
- **D.** All of the above.



larger tree
(size = 16, height = 4)

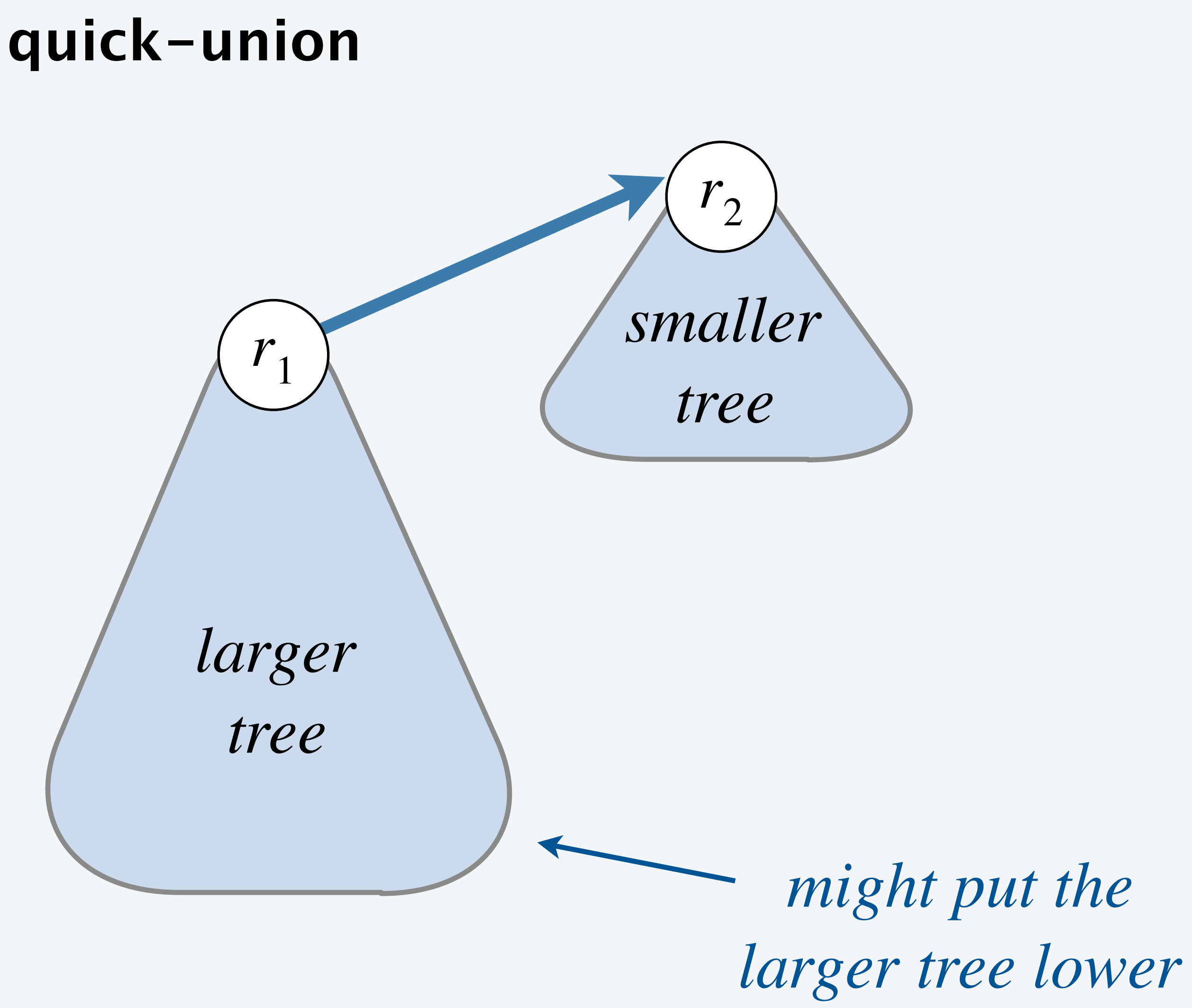


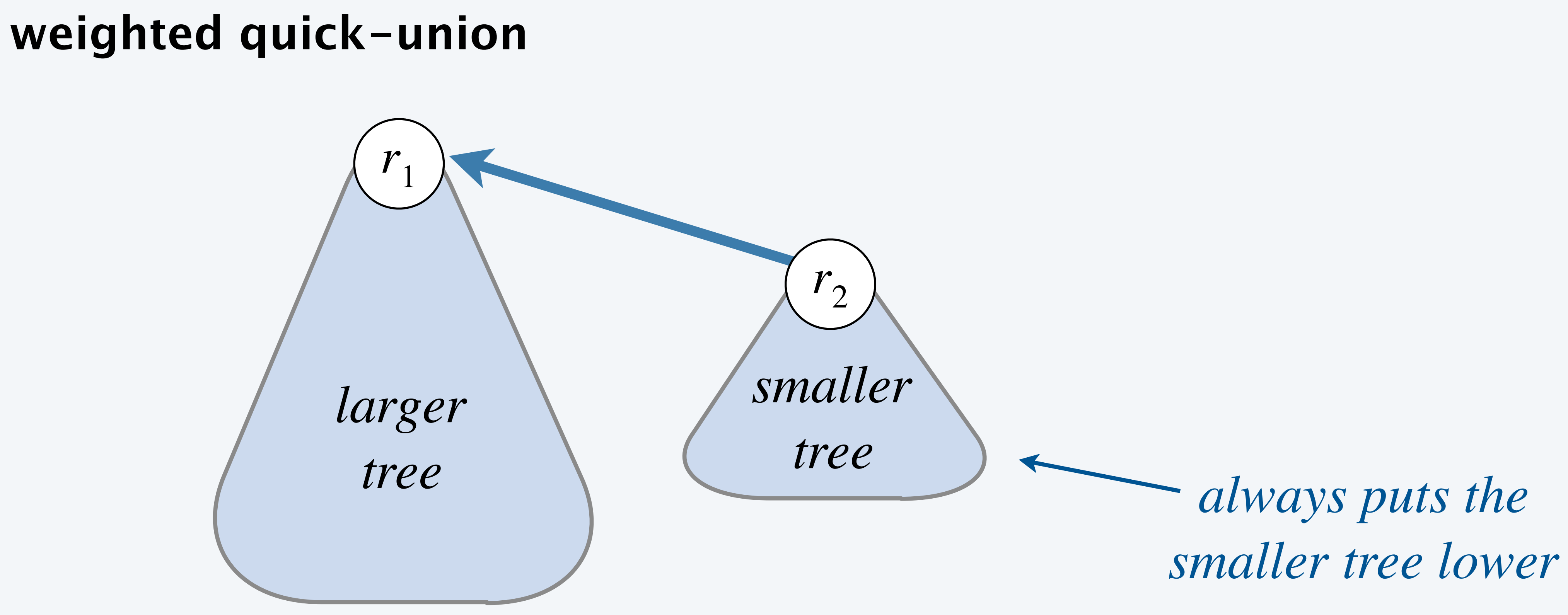
smaller tree
(size = 6, height = 2)

Weighted quick-union (link-by-size)

Link-by-size. Modify quick-union to avoid tall trees.

- Keep track of size of each tree = number of elements.
- Always link root of smaller tree to root of larger tree. ← *fine alternative: link-by-height (minimize worst-case depth vs. average depth)*





Weighted quick-union: Java implementation

Data structure. Same as quick-union, but maintain extra array `size[i]` to count number of elements in the tree rooted at `i`, initially `1`.

- `find()`: identical to quick-union.
- `union()`: link root of smaller tree to root of larger tree; update `size[]`.

```
public void union(int p, int q) {
   int rootP = find(p);
   int rootQ = find(q);
   if (rootP == rootQ) return;

   if (size[rootP] < size[rootQ]) {
      parent[rootP] = rootQ;
      size[rootQ] += size[rootP];
   }
   else {
      parent[rootQ] = rootP;
      size[rootP] += size[rootQ];
   }

}
```

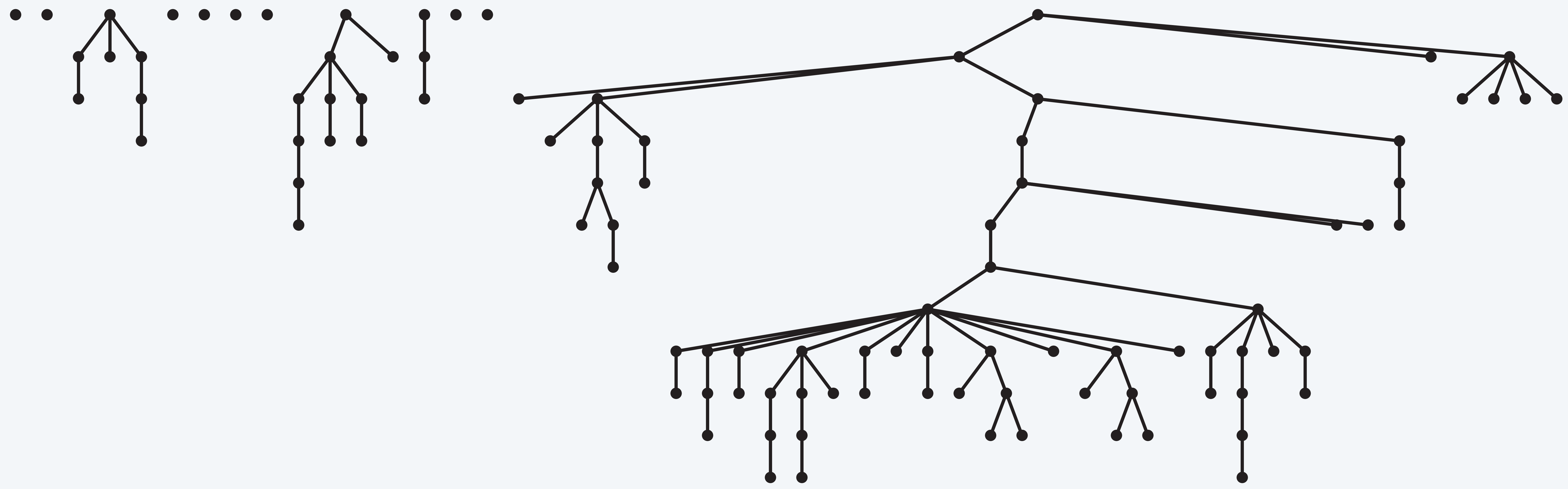
p and q already in the same set (annotation on `if (rootP == rootQ) return;`)

link root of smaller tree to root of larger tree (and update size) (annotation on `if (size[rootP] < size[rootQ])`)

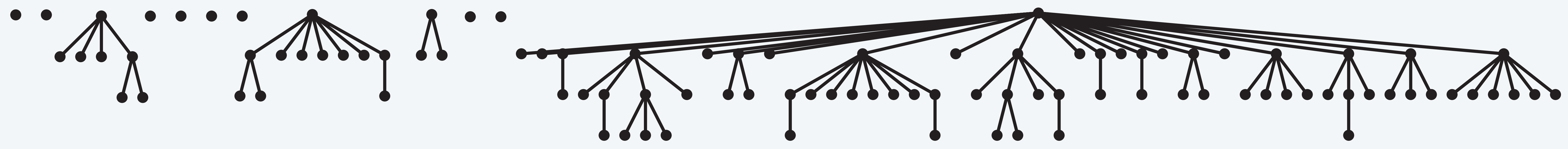
https://algs4.cs.princeton.edu/15uf/WeightedQuickUnionUF.java.html

Quick-union vs. weighted quick-union: larger example

quick-union

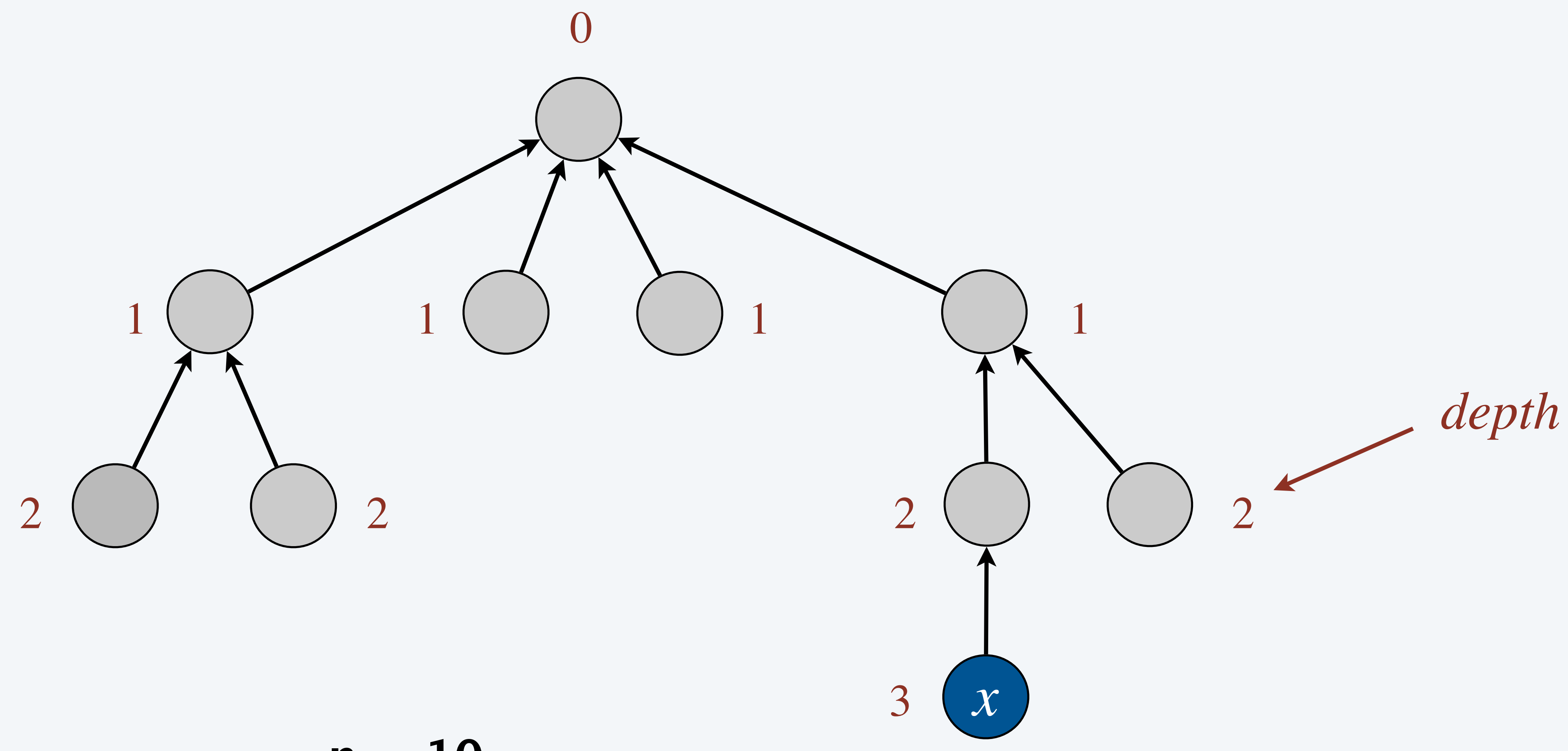


weighted



Weighted quick-union analysis

Proposition. Depth of any node $x \leq \log_2 n$.



n = 10
depth(x) = 3 ≤ $\log_2 n$

Weighted quick-union analysis

Proposition. Depth of any node $x \leq \log_2 n$.

Pf.

- Depth of x does not change unless root of tree T_1 containing x is linked to the root of a larger tree T_2, forming a new tree T_3.
- When this happens:
  - depth of x increases by exactly 1
  - size of tree containing x at least doubles because $size(T_3) = size(T_1) + size(T_2)$ $\geq 2 \times size(T_1)$.

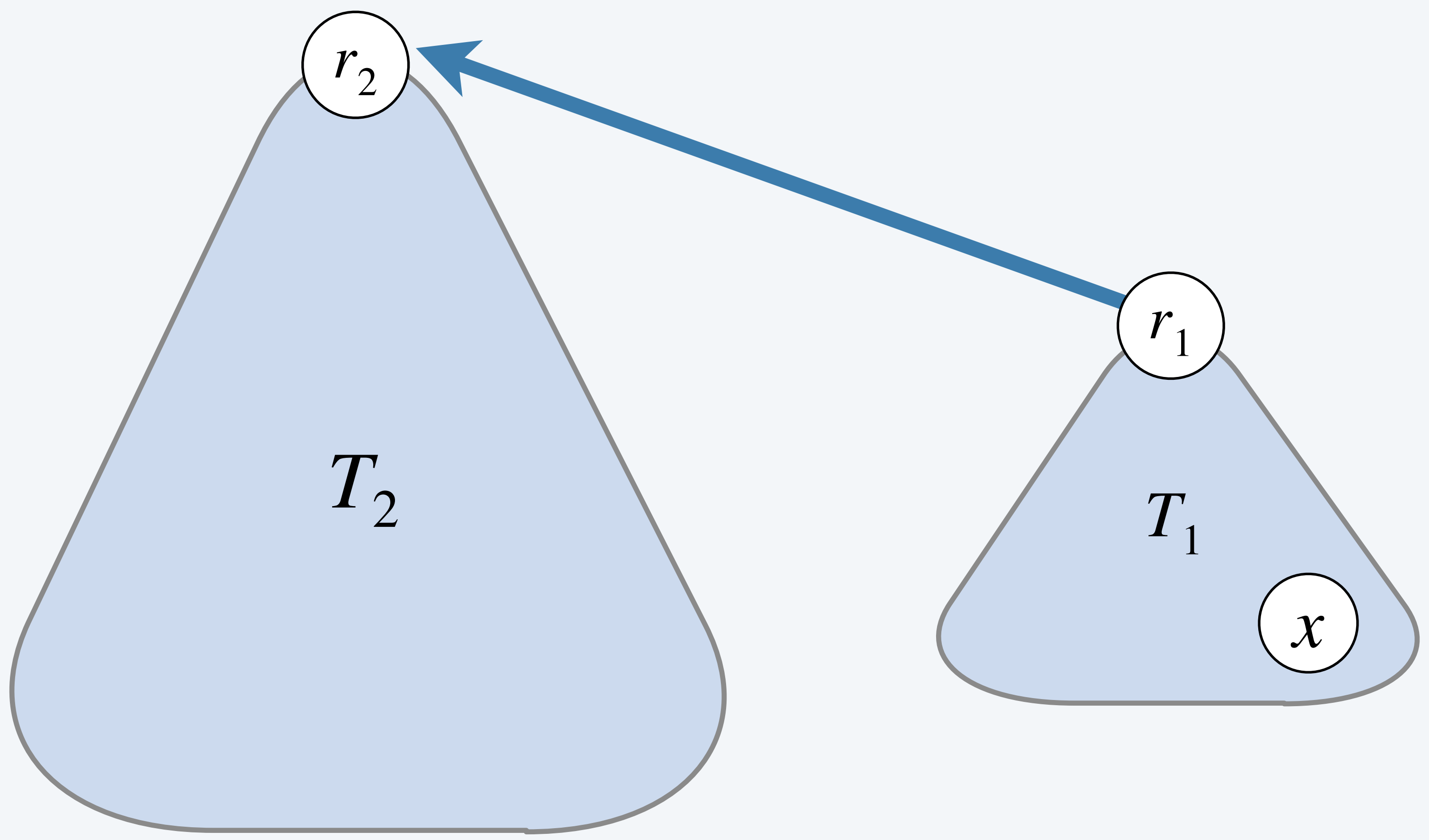
can happen at most $\log_2 n$ times. Why?

$1 \rightarrow 2 \rightarrow 4 \rightarrow 8 \rightarrow 16 \rightarrow \cdots \rightarrow n$

$\log_2 n$

size of tree containing x is initially 1

size of tree containing x can't exceed n



Weighted quick-union analysis

Proposition. Depth of any node $x \leq \log_2 n$.

Running time.

- `union()` takes constant time, given two roots.
- `find()` takes time proportional to depth of node in tree.

algorithm	initialize	union	find
quick-find	n	n	1
quick-union	n	n	n
weighted quick-union	n	$\log n$	$\log n$

worst-case number of array accesses (ignoring leading coefficient)

in this course, $\log$ *mean logarithm for some constant base*

Summary

Key point. Weighted quick-union empowers us to solve problems that could not otherwise be addressed.

algorithm	worst-case time
quick-find	$m\,n$
quick-union	$m\,n$
weighted quick-union	$m \log n$
quick-union + path compression	$m \log n$ ← *fastest for percolation?*
weighted quick-union + path compression	$m\, \alpha(m, n)$ ← *inverse Ackermann function (see COS 423)*

order of growth for m ≥ n union-find operations on a set of n elements

Ex. [10^9 union-find operations on 10^9 elements]

- Efficient algorithm reduces time from 30 years to 6 seconds.
- Supercomputer won't help much.

Credits

image	source
Game of Hex	Wolfram MathWorld
Cluster Labeling	Tiberiu Marita
Bob Tarjan	Princeton University
Computer and Supercomputer	New York Times

A final thought

" *The goal is to come up with algorithms that you can apply in practice that run fast, as well as being simple, beautiful, and analyzable.* " — Robert Tarjan