Algorithms

ROBERT SEDGEWICK | KEVIN WAYNE



 \checkmark

Robert Sedgewick | Kevin Wayne

http://algs4.cs.princeton.edu

4.1 UNDIRECTED GRAPHS

- introduction
- graph API
- depth-first search
- breadth-first search
- connected components
- challenges

4.1 UNDIRECTED GRAPHS

introduction

graph APt

challenges

depth-first search

breadth-first search

connected components

Algorithms

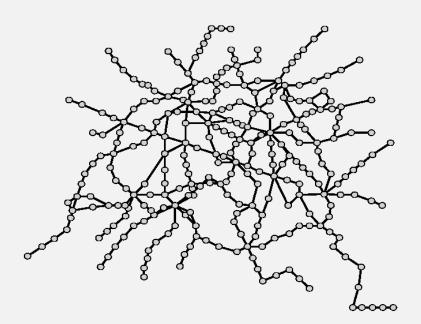
Robert Sedgewick | Kevin Wayne

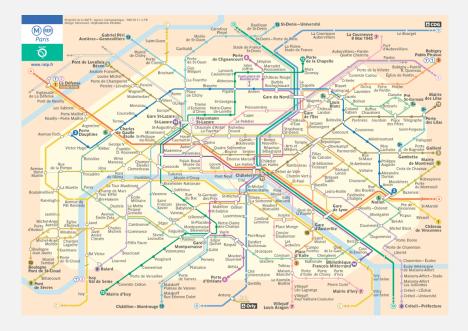
http://algs4.cs.princeton.edu

Graph. Set of vertices connected pairwise by edges.

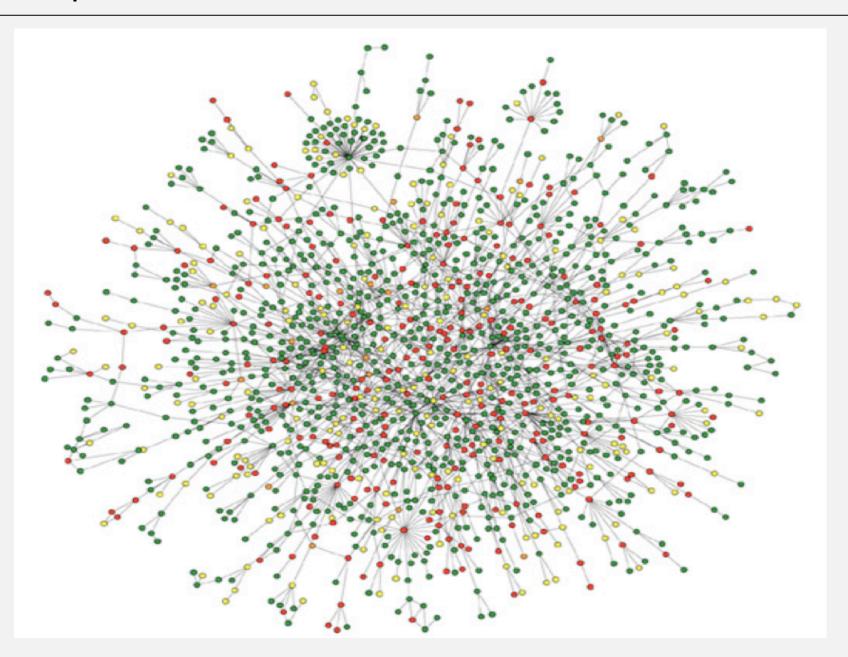
Why study graph algorithms?

- Thousands of practical applications.
- Hundreds of graph algorithms known.
- Interesting and broadly useful abstraction.
- Challenging branch of computer science and discrete math.



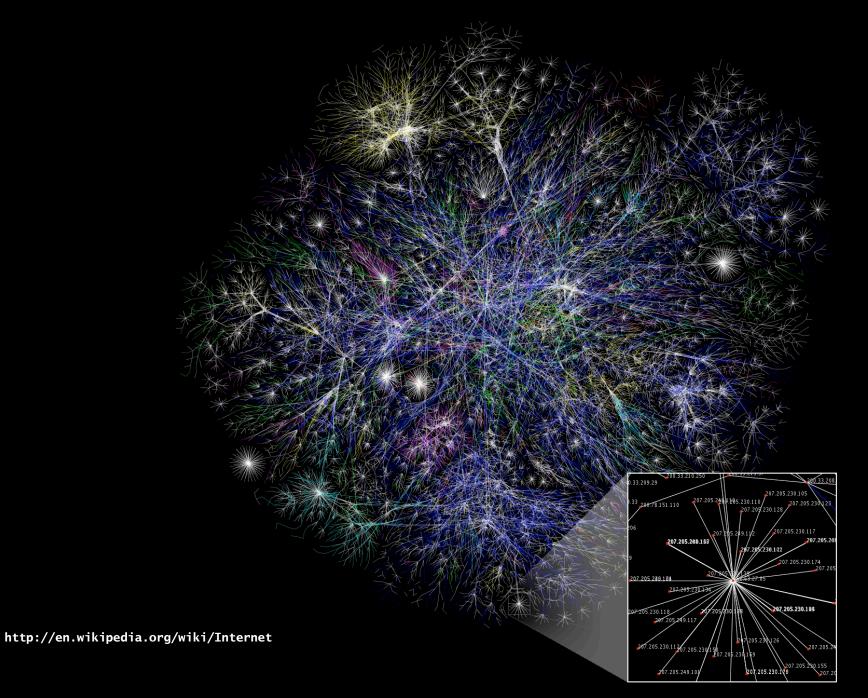


Protein-protein interaction network

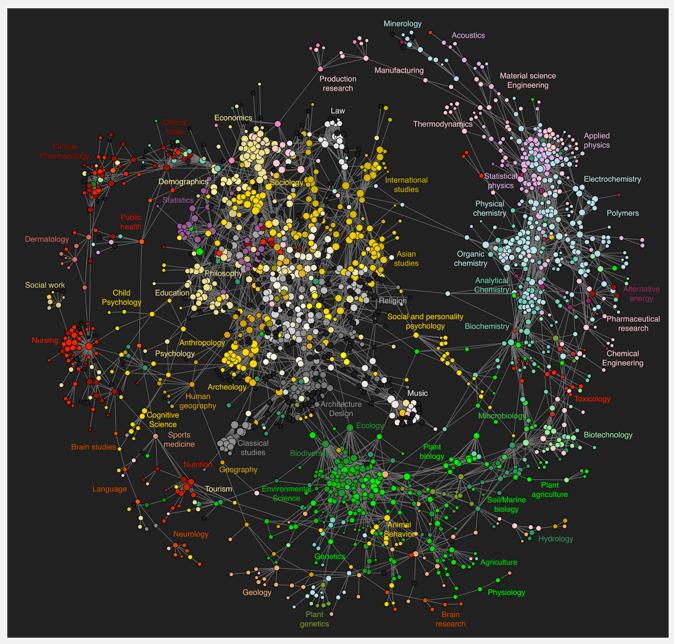


Reference: Jeong et al, Nature Review | Genetics

The Internet as mapped by the Opte Project



Map of science clickstreams



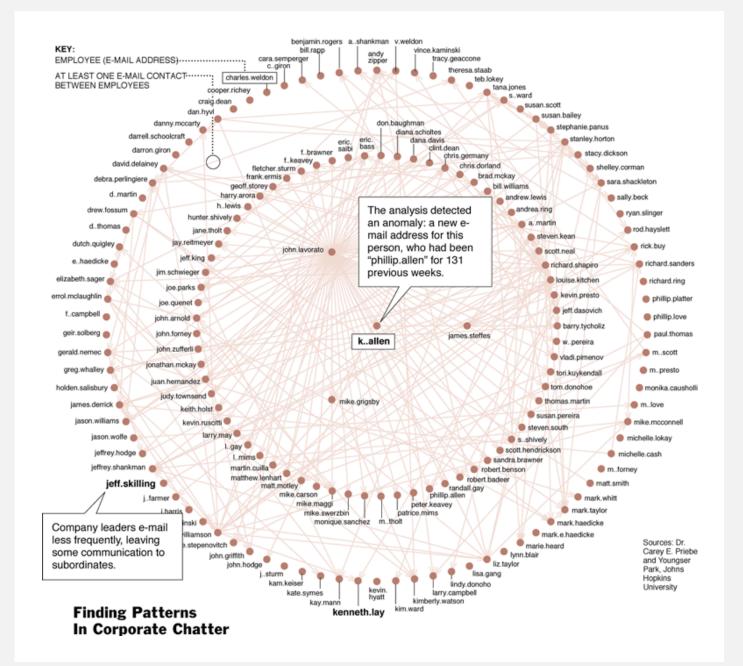
http://www.plosone.org/article/info:doi/10.1371/journal.pone.0004803

10 million Facebook friends

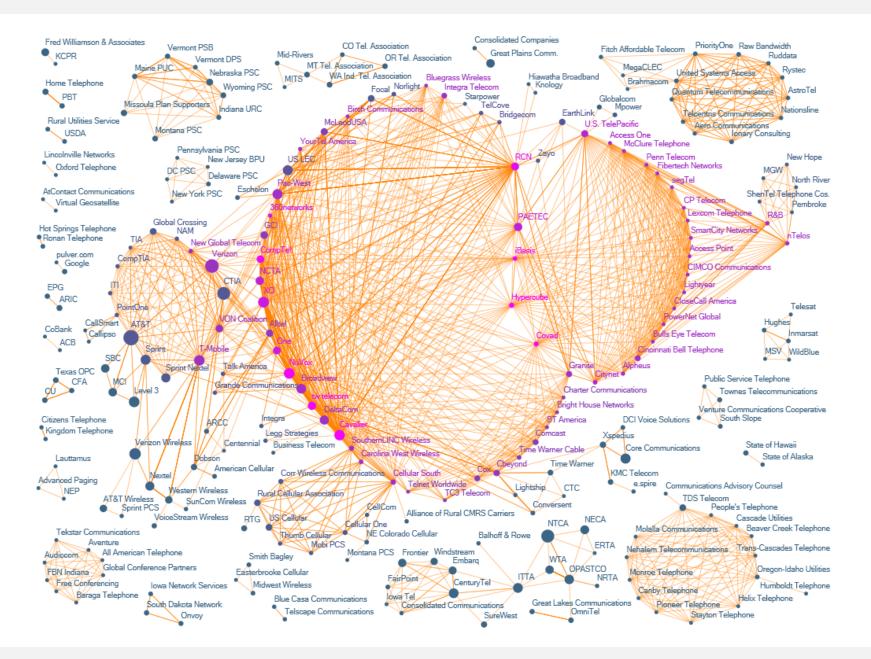


"Visualizing Friendships" by Paul Butler

One week of Enron emails

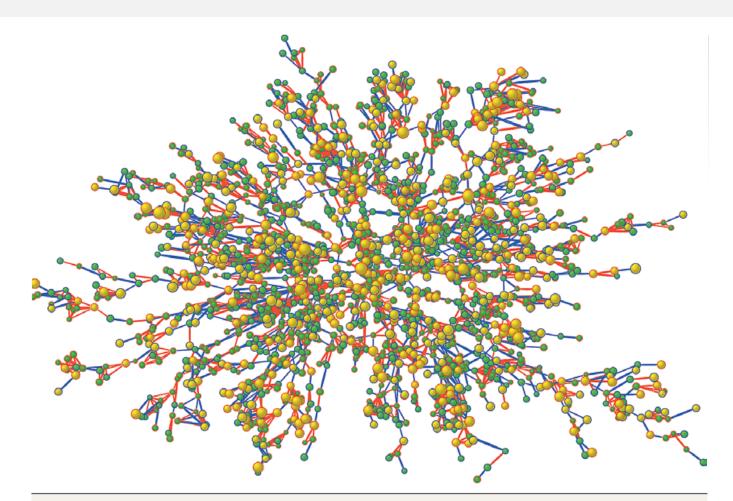


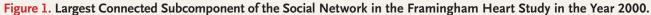
The evolution of FCC lobbying coalitions



"The Evolution of FCC Lobbying Coalitions" by Pierre de Vries in JoSS Visualization Symposium 2010

Framingham heart study





Each circle (node) represents one person in the data set. There are 2200 persons in this subcomponent of the social network. Circles with red borders denote women, and circles with blue borders denote men. The size of each circle is proportional to the person's body-mass index. The interior color of the circles indicates the person's obesity status: yellow denotes an obese person (body-mass index, \geq 30) and green denotes a nonobese person. The colors of the ties between the nodes indicate the relationship between them: purple denotes a friendship or marital tie and orange denotes a familial tie.

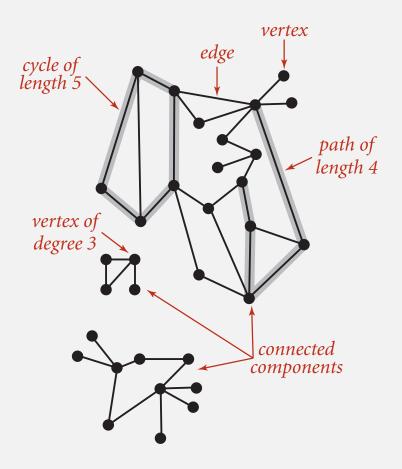
graph	vertex	edge
communication	telephone, computer	fiber optic cable
circuit	gate, register, processor	wire
mechanical	joint	rod, beam, spring
financial	stock, currency	transactions
transportation	street intersection, airport	highway, airway route
internet	class C network	connection
game	board position	legal move
social relationship	person, actor	friendship, movie cast
neural network	neuron	synapse
protein network	protein	protein-protein interaction
chemical compound	molecule	bond

Graph terminology

Path. Sequence of vertices connected by edges.

Cycle. Path whose first and last vertices are the same.

Two vertices are **connected** if there is a path between them.



Some graph-processing problems

Path. Is there a path between *s* and *t*?Shortest path. What is the shortest path between *s* and *t*?

Cycle. Is there a cycle in the graph? Euler tour. Is there a cycle that uses each edge exactly once? Hamilton tour. Is there a cycle that uses each vertex exactly once.

Connectivity. Is there a way to connect all of the vertices? MST. What is the best way to connect all of the vertices? Biconnectivity. Is there a vertex whose removal disconnects the graph?

Planarity. Can you draw the graph in the plane with no crossing edges Graph isomorphism. Do two adjacency lists represent the same graph?

Challenge. Which of these problems are easy? difficult? intractable?

4.1 UNDIRECTED GRAPHS

introduction

depth-first search

breadth-first search

connected components

graph API

challenges

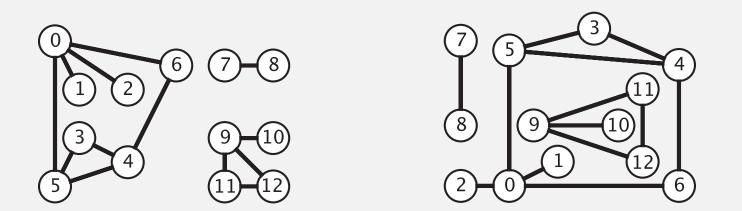
Algorithms

Robert Sedgewick | Kevin Wayne

http://algs4.cs.princeton.edu

Graph representation

Graph drawing. Provides intuition about the structure of the graph.



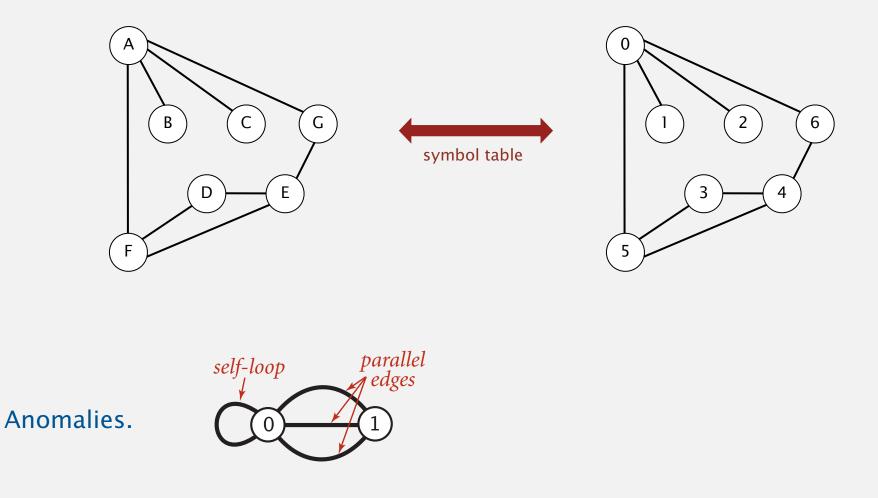
two drawings of the same graph

Caveat. Intuition can be misleading.

Graph representation

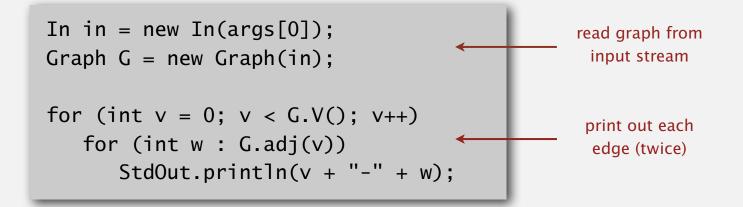
Vertex representation.

- This lecture: use integers between 0 and V-1.
- Applications: convert between names and integers with symbol table.

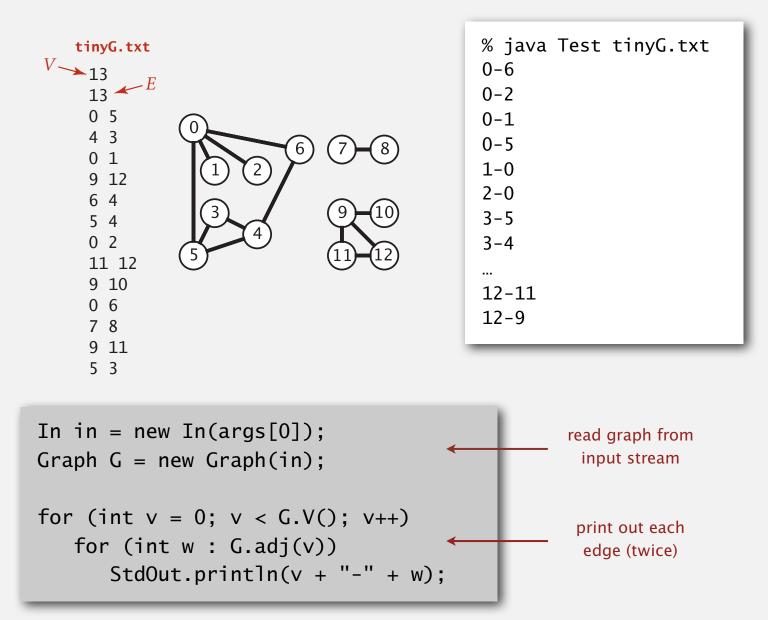


Graph API

public class	Graph	
	Graph(int V)	create an empty graph with V vertices
	Graph(In in)	create a graph from input stream
void	addEdge(int v, int w)	add an edge v-w
Iterable <integer></integer>	adj(int v)	vertices adjacent to v
int	V()	number of vertices
int	E()	number of edges



Graph input format.

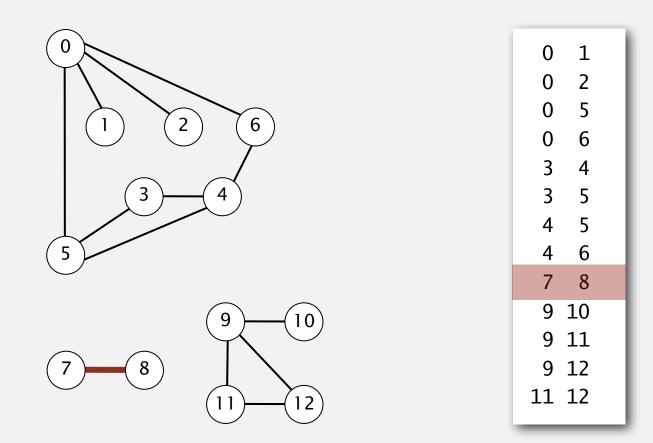


public class	Graph	
	Graph(int V)	create an empty graph with V vertices
	Graph(In in)	create a graph from input stream
void	addEdge(int v, int w)	add an edge v-w
Iterable <integer></integer>	adj(int v)	vertices adjacent to v
int	V()	number of vertices
int	E()	number of edges

```
// degree of vertex v in graph G
public static int degree(Graph G, int v)
{
    int degree = 0;
    for (int w : G.adj(v))
        degree++;
    return degree;
}
```

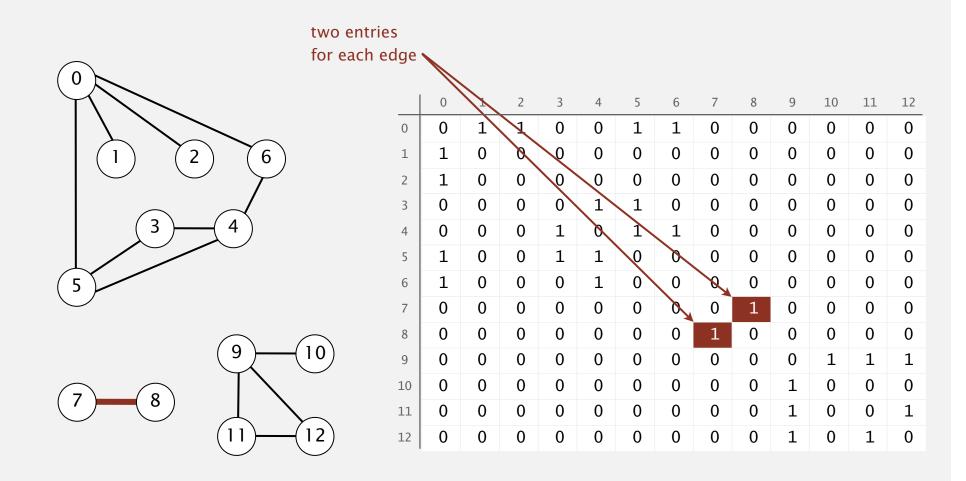
Set-of-edges graph representation

Maintain a list of the edges (linked list or array).



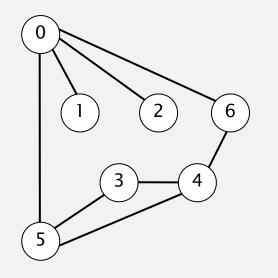
Adjacency-matrix graph representation

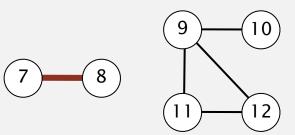
Maintain a two-dimensional *V*-by-*V* boolean array; for each edge v-w in graph: adj[v][w] = adj[w][v] = true.

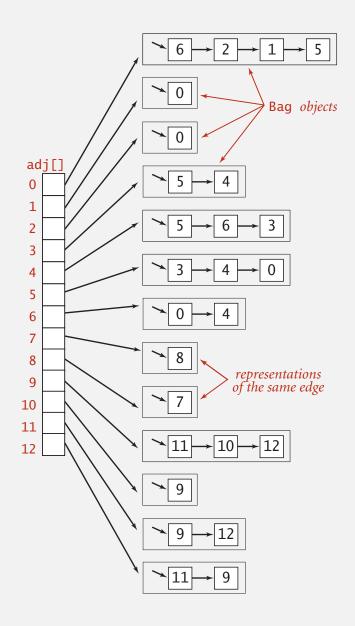


Adjacency-list graph representation

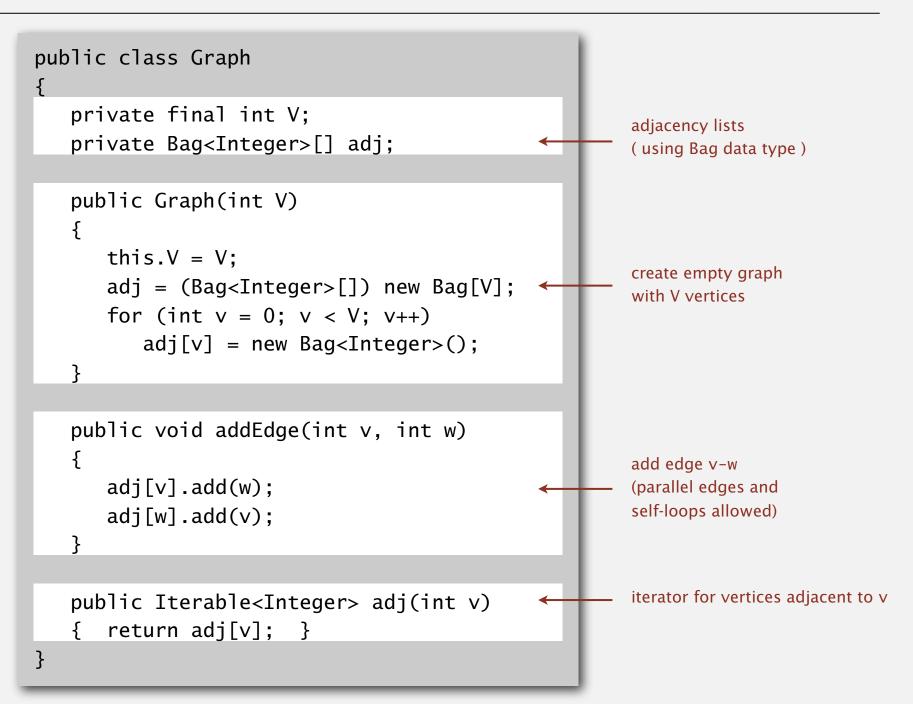
Maintain vertex-indexed array of lists.







Adjacency-list graph representation: Java implementation

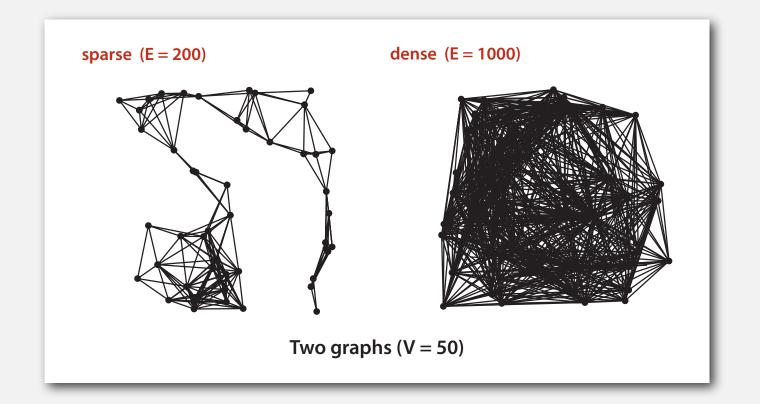


Graph representations

In practice. Use adjacency-lists representation.

- Algorithms based on iterating over vertices adjacent to v.
- Real-world graphs tend to be sparse.

 huge number of vertices, small average vertex degree



Graph representations

In practice. Use adjacency-lists representation.

- Algorithms based on iterating over vertices adjacent to v.
- Real-world graphs tend to be sparse.

huge number of vertices, small average vertex degree

representation	space	add edge	edge between v and w?	iterate over vertices adjacent to v?
list of edges	E	1	E	E
adjacency matrix	V ²	1 *	1	V
adjacency lists	E + V	1	degree(v)	degree(v)

* disallows parallel edges

4.1 UNDIRECTED GRAPHS

introduction

sraph APt

challenges

depth-first search

breadth-first search

connected components

Algorithms

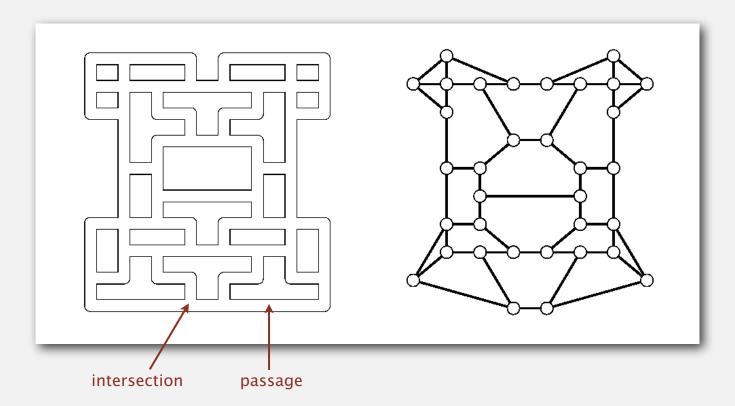
Robert Sedgewick | Kevin Wayne

http://algs4.cs.princeton.edu

Maze exploration

Maze graph.

- Vertex = intersection.
- Edge = passage.

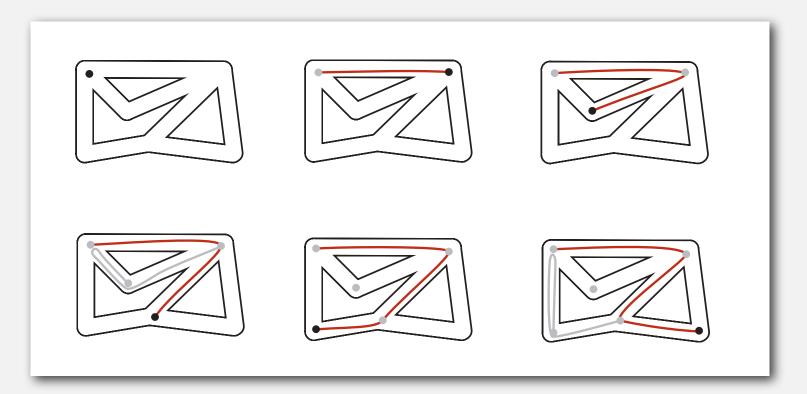


Goal. Explore every intersection in the maze.

Trémaux maze exploration

Algorithm.

- Unroll a ball of string behind you.
- Mark each visited intersection and each visited passage.
- Retrace steps when no unvisited options.



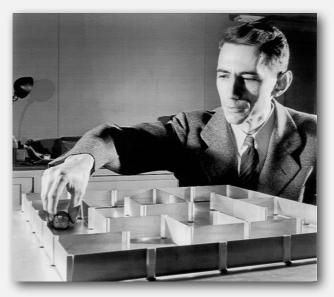
Trémaux maze exploration

Algorithm.

- Unroll a ball of string behind you.
- Mark each visited intersection and each visited passage.
- Retrace steps when no unvisited options.

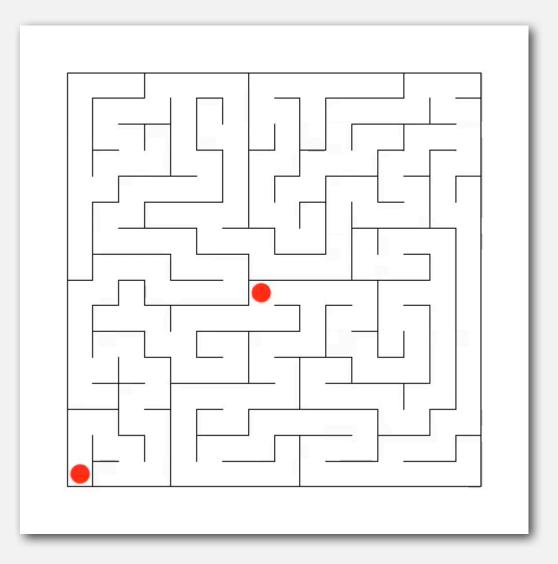
First use? Theseus entered Labyrinth to kill the monstrous Minotaur; Ariadne instructed Theseus to use a ball of string to find his way back out.



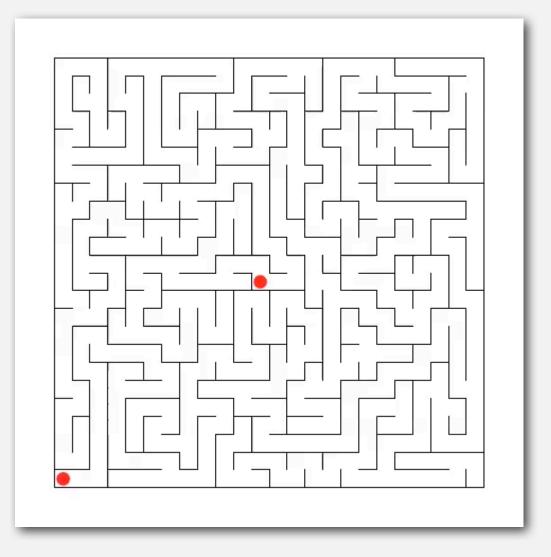


Claude Shannon (with Theseus mouse)

Maze exploration



Maze exploration



Depth-first search

- Goal. Systematically search through a graph.
- Idea. Mimic maze exploration.

DFS (to visit a vertex v)
Mark	v as visited.
Recur	sively visit all unmarked
	vertices w adjacent to v.

Typical applications.

- Find all vertices connected to a given source vertex.
- Find a path between two vertices.

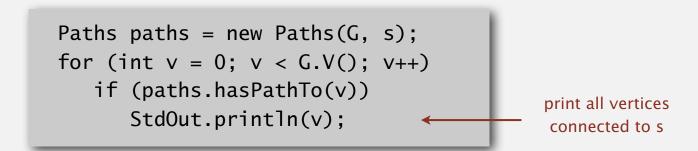
Design challenge. How to implement?

Design pattern for graph processing

Design pattern. Decouple graph data type from graph processing.

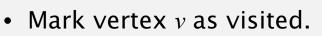
- Create a Graph object.
- Pass the Graph to a graph-processing routine.
- Query the graph-processing routine for information.

public class	Paths	
	Paths(Graph G, int s)	find paths in G from source s
boolean	hasPathTo(int v)	is there a path from s to v?
Iterable <integer></integer>	pathTo(int v)	path from s to v; null if no such path

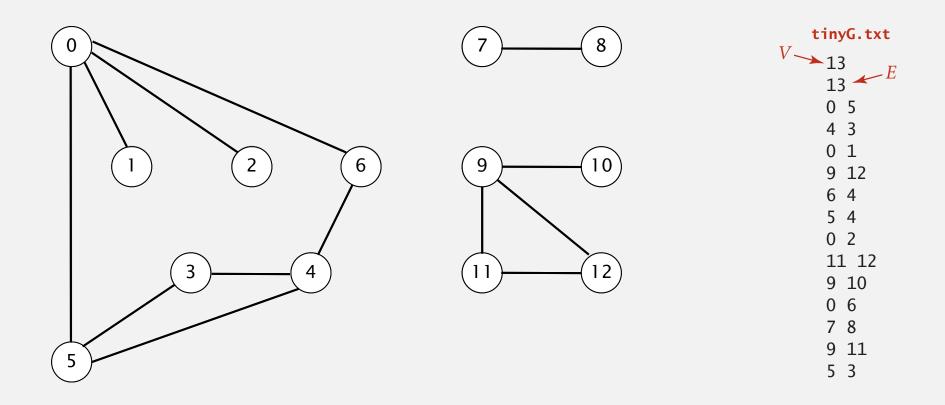


Depth-first search demo

To visit a vertex *v* :



• Recursively visit all unmarked vertices adjacent to v.

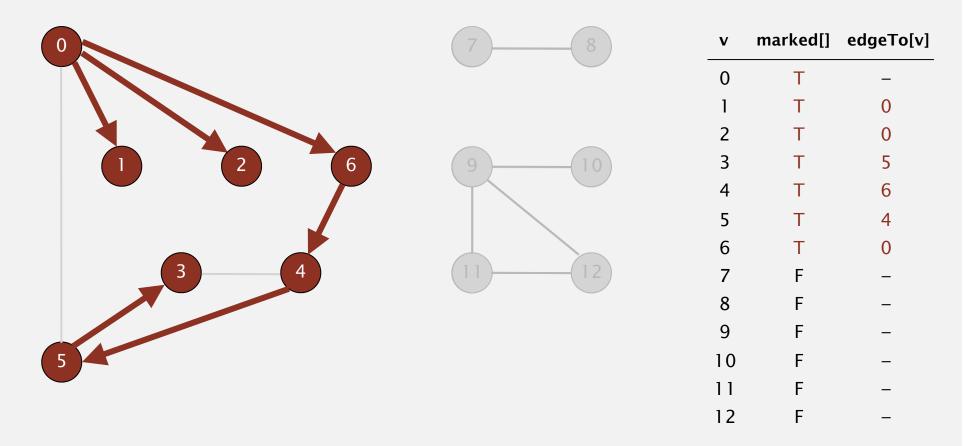


graph G

Depth-first search demo

To visit a vertex *v* :

- Mark vertex *v* as visited.
- Recursively visit all unmarked vertices adjacent to v.



vertices reachable from 0

Depth-first search

Goal. Find all vertices connected to *s* (and a corresponding path). Idea. Mimic maze exploration.

Algorithm.

- Use recursion (ball of string).
- Mark each visited vertex (and keep track of edge taken to visit it).
- Return (retrace steps) when no unvisited options.

Data structures.

- boolean[] marked to mark visited vertices.
- int[] edgeTo to keep tree of paths.

(edgeTo[w] == v) means that edge v-w taken to visit w for first time

Depth-first search



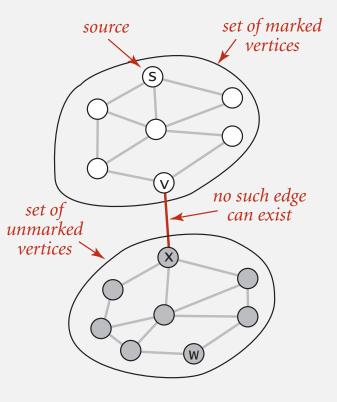
Proposition. DFS marks all vertices connected to *s* in time proportional to the sum of their degrees.

Pf. [correctness]

- If *w* marked, then *w* connected to *s* (why?)
- If *w* connected to *s*, then *w* marked.
 (if *w* unmarked, then consider last edge on a path from *s* to *w* that goes from a marked vertex to an unmarked one).

Pf. [running time]

Each vertex connected to *s* is visited once.

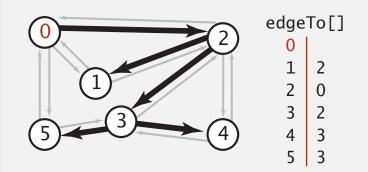


Depth-first search properties

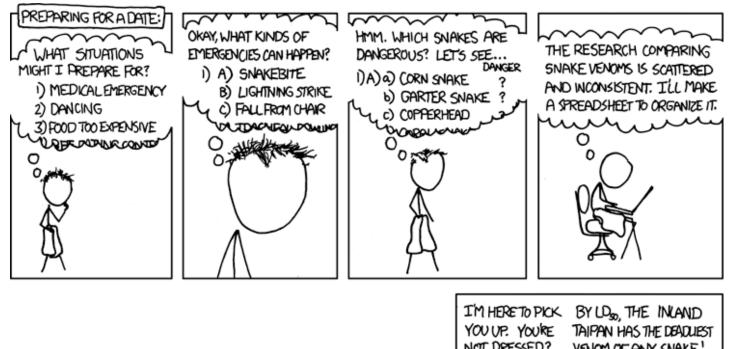
Proposition. After DFS, can find vertices connected to *s* in constant time and can find a path to *s* (if one exists) in time proportional to its length.

Pf. edgeTo[] is parent-link representation of a tree rooted at s.

```
public boolean hasPathTo(int v)
{ return marked[v]; }
public Iterable<Integer> pathTo(int v)
{
    if (!hasPathTo(v)) return null;
    Stack<Integer> path = new Stack<Integer>();
    for (int x = v; x != s; x = edgeTo[x])
        path.push(x);
    path.push(s);
    return path;
}
```



Depth-first search application: preparing for a date







I REALLY NEED TO STOP USING DEPTH-FIRST SEARCHES.

Depth-first search application: flood fill

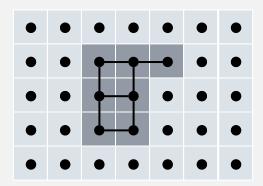
Challenge. Flood fill (Photoshop magic wand). Assumptions. Picture has millions to billions of pixels.





Solution. Build a grid graph.

- Vertex: pixel.
- Edge: between two adjacent gray pixels.
- Blob: all pixels connected to given pixel.



4.1 UNDIRECTED GRAPHS

introduction

graph APt

challenges

depth-first search

breadth-first search

connected components

Algorithms

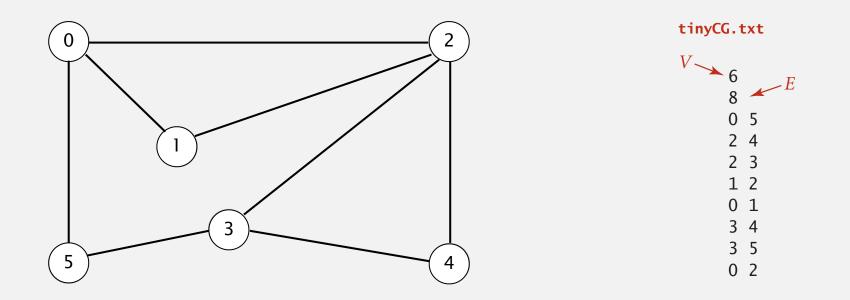
Robert Sedgewick | Kevin Wayne

http://algs4.cs.princeton.edu

Breadth-first search demo

Repeat until queue is empty:

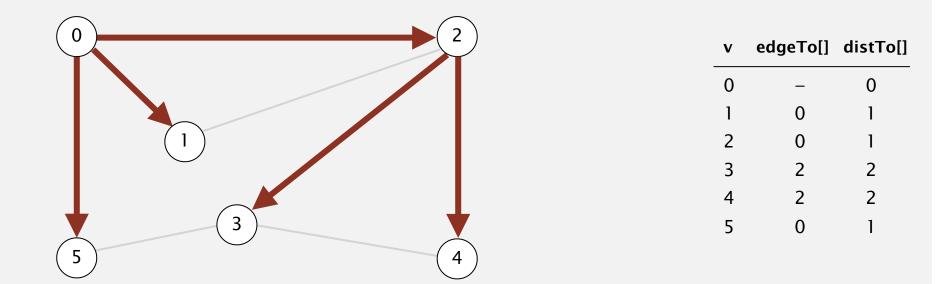
- Remove vertex *v* from queue.
- Add to queue all unmarked vertices adjacent to v and mark them.



Breadth-first search demo

Repeat until queue is empty:

- Remove vertex *v* from queue.
- Add to queue all unmarked vertices adjacent to v and mark them.



done

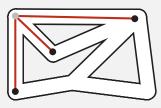
Depth-first search. Put unvisited vertices on a stack. Breadth-first search. Put unvisited vertices on a queue.

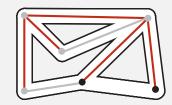
Shortest path. Find path from *s* to *t* that uses fewest number of edges.

BFS (from source vertex s)

Put s onto a FIFO queue, and mark s as visited. Repeat until the queue is empty:

- remove the least recently added vertex v
- add each of v's unvisited neighbors to the queue, and mark them as visited.





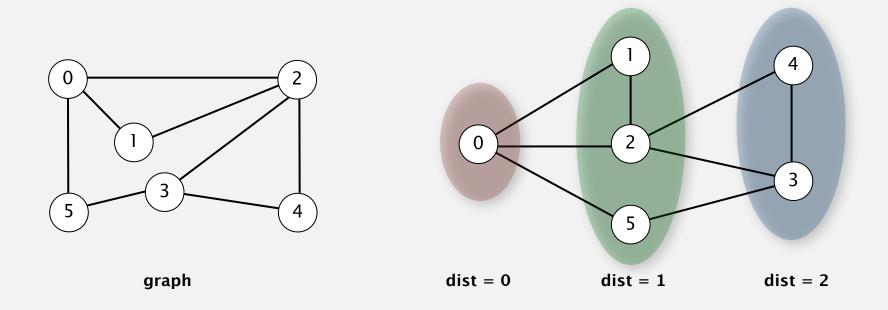


Intuition. BFS examines vertices in increasing distance from *s*.

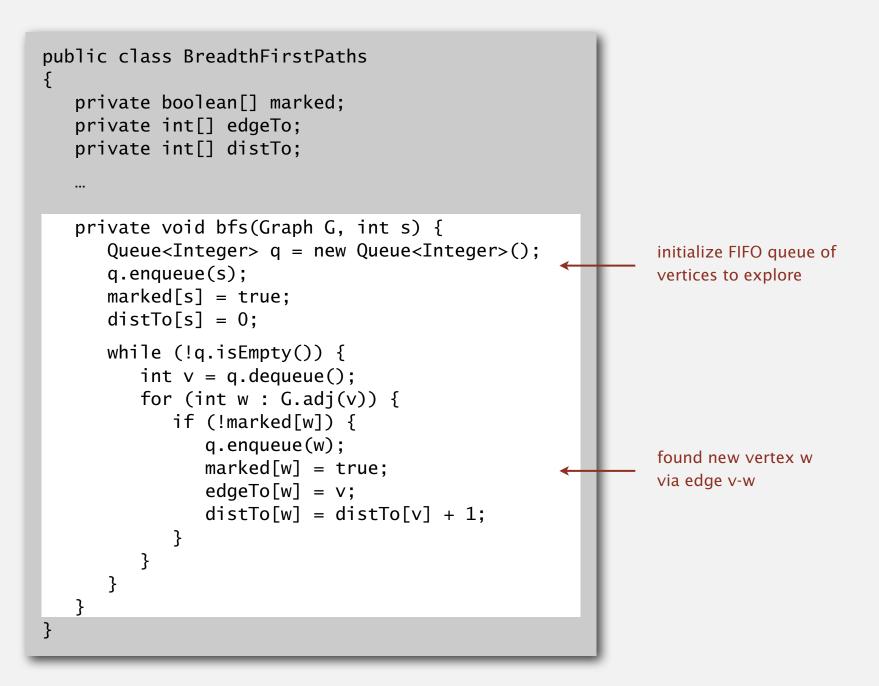
Proposition. BFS computes shortest paths (fewest number of edges) from *s* to all other vertices in a graph in time proportional to E + V.

Pf. [correctness] Queue always consists of zero or more vertices of distance k from s, followed by zero or more vertices of distance k + 1.

Pf. [running time] Each vertex connected to *s* is visited once.

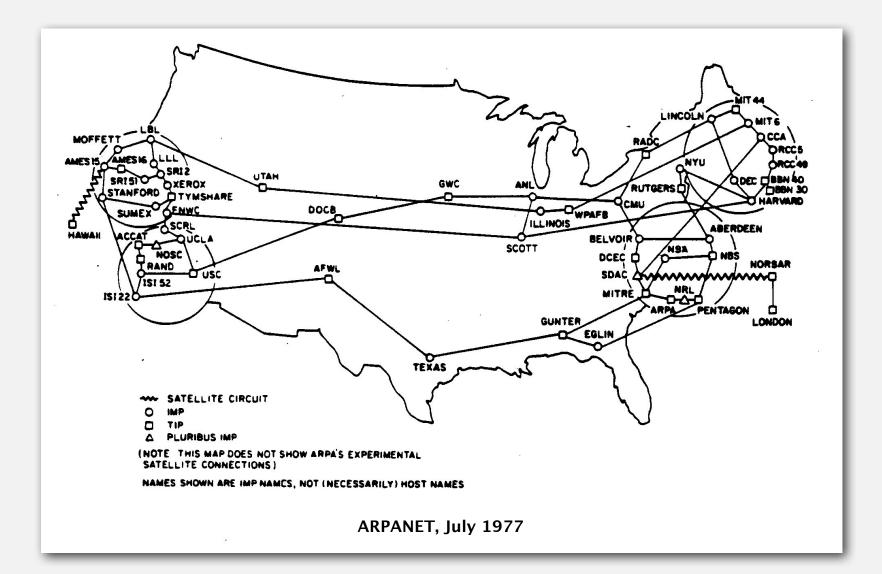


Breadth-first search



Breadth-first search application: routing

Fewest number of hops in a communication network.



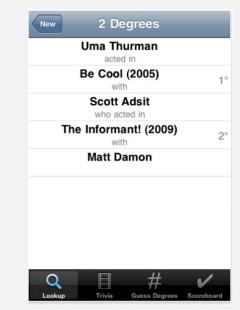
Breadth-first search application: Kevin Bacon numbers

Kevin Bacon numbers.





Endless Games board game

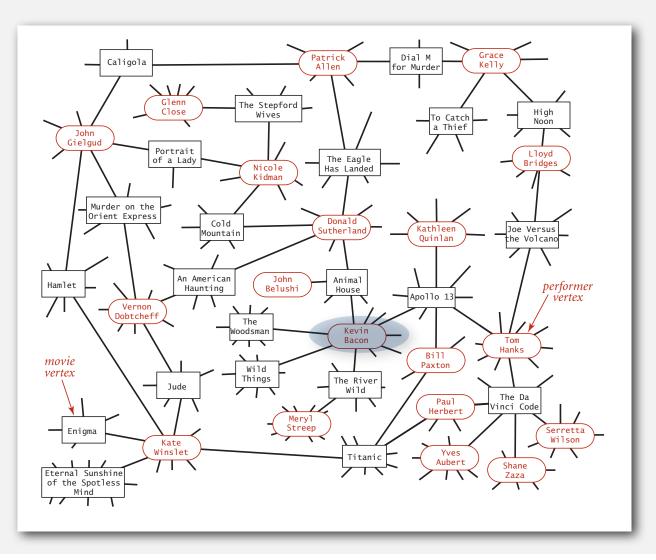


SixDegrees iPhone App

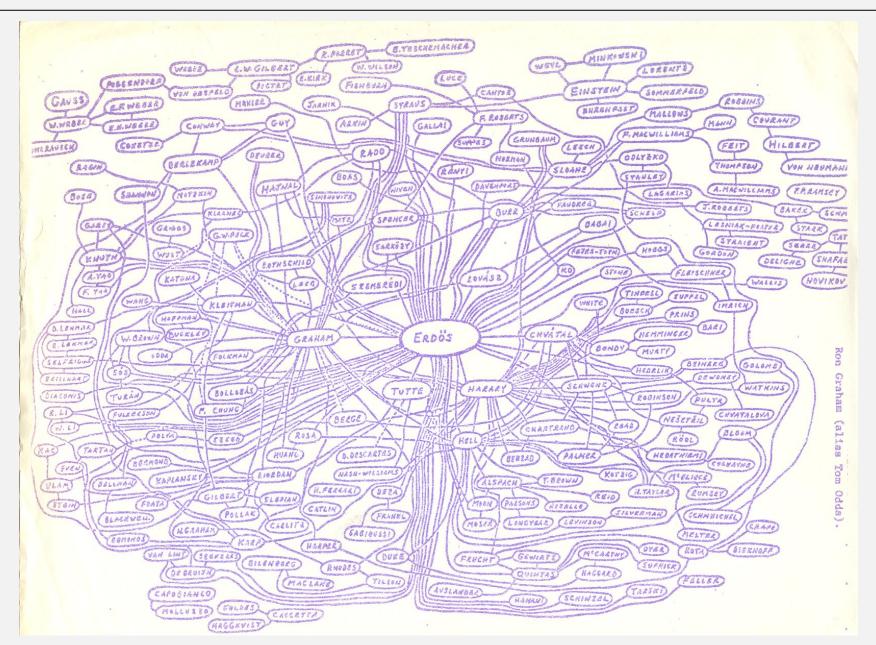
http://oracleofbacon.org

Kevin Bacon graph

- Include one vertex for each performer and one for each movie.
- Connect a movie to all performers that appear in that movie.
- Compute shortest path from *s* = Kevin Bacon.



Breadth-first search application: Erdös numbers



hand-drawing of part of the Erdös graph by Ron Graham

4.1 UNDIRECTED GRAPHS

Algorithms

connected components

depth-first search

breadth-first search

introduction

graph APt

challenges

Robert Sedgewick | Kevin Wayne

http://algs4.cs.princeton.edu

Def. Vertices *v* and *w* are **connected** if there is a path between them.

Goal. Preprocess graph to answer queries of the form *is v connected to w?* in constant time.

public class	CC	
	CC(Graph G)	find connected components in G
boolean	<pre>connected(int v, int w)</pre>	are v and w connected?
int	count()	number of connected components
int	id(int v)	<i>component identifier for v</i> (<i>between</i> 0 <i>and</i> count() - 1)

Union-Find? Not quite. Depth-first search. Yes. [next few slides]

Connected components

The relation "is connected to" is an equivalence relation:

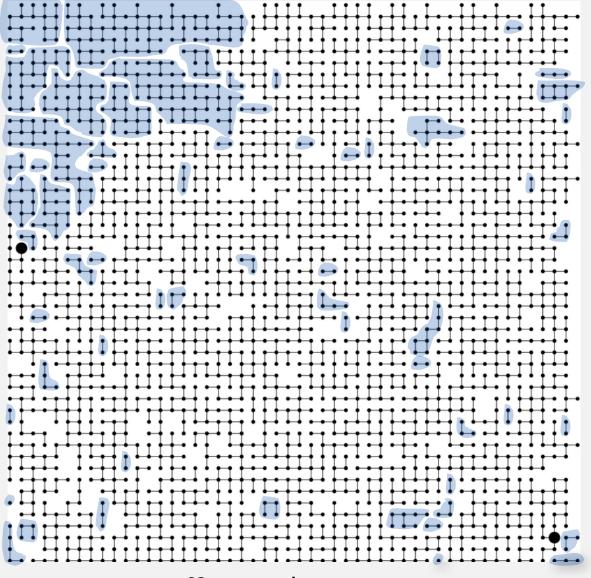
- Reflexive: *v* is connected to *v*.
- Symmetric: if v is connected to w, then w is connected to v.
- Transitive: if *v* connected to *w* and *w* connected to *x*, then *v* connected to *x*.

Def. A **connected component** is a maximal set of connected vertices.

	V	id[]
	0	0
	1	0
Y//	2	0
	3	0
$\begin{pmatrix} 1 \end{pmatrix}$ $\begin{pmatrix} 2 \end{pmatrix}$ $\begin{pmatrix} 6 \end{pmatrix}$	4	0
	5	0
	6	0
	7	1
	8	1
	9	2
	10	2
3 connected components	11	2
	12	2

Remark. Given connected components, can answer queries in constant time.

Def. A connected component is a maximal set of connected vertices.



63 connected components

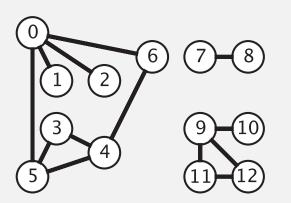
Connected components

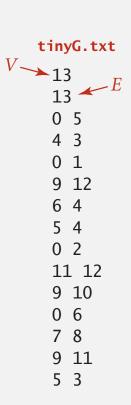
Goal. Partition vertices into connected components.



Initialize all vertices v as unmarked.

For each unmarked vertex v, run DFS to identify all vertices discovered as part of the same component.

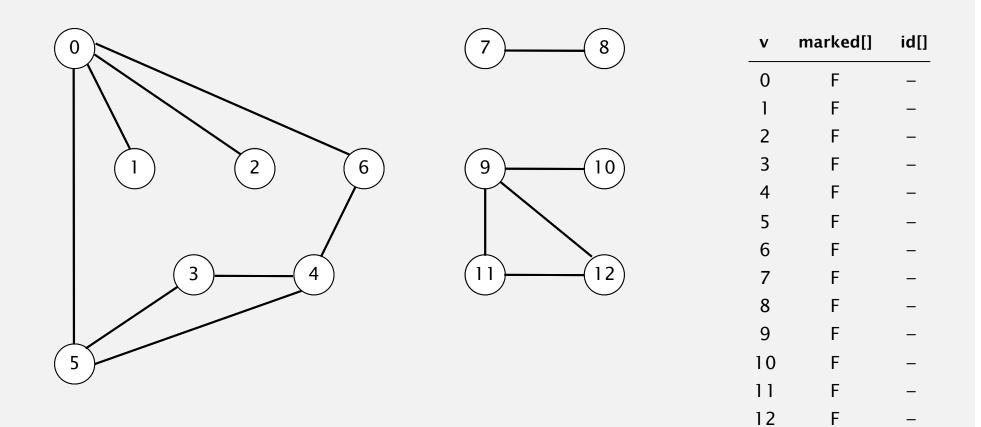




Connected components demo

To visit a vertex *v* :

- Mark vertex v as visited.
- Recursively visit all unmarked vertices adjacent to v.

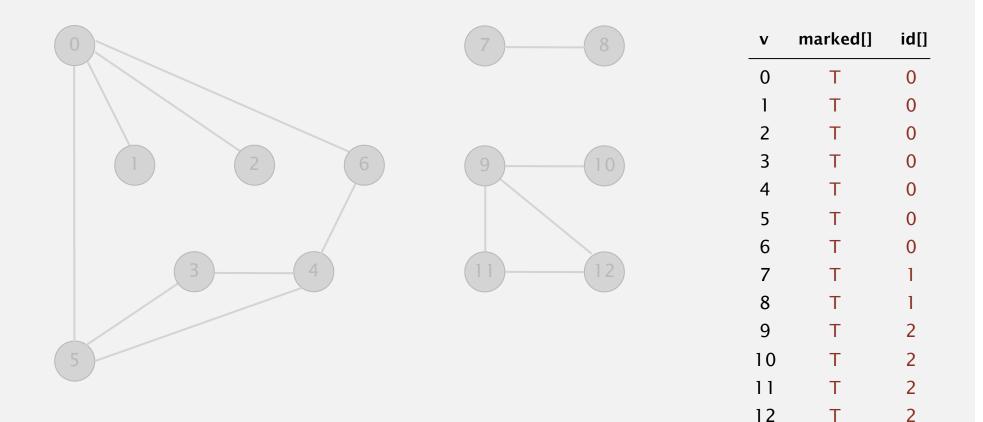


graph G

Connected components demo

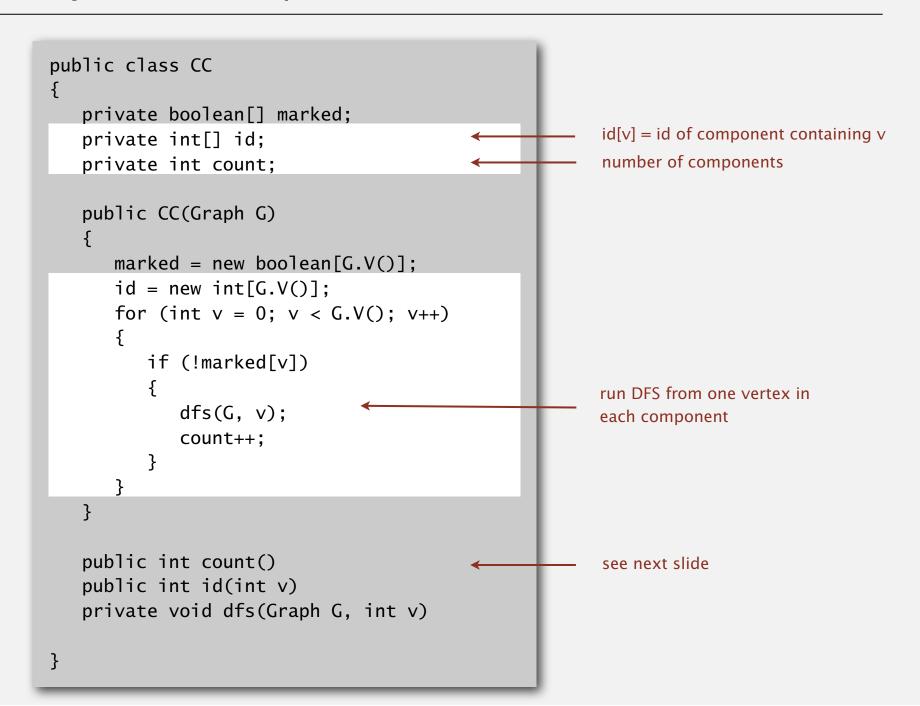
To visit a vertex *v* :

- Mark vertex v as visited.
- Recursively visit all unmarked vertices adjacent to v.

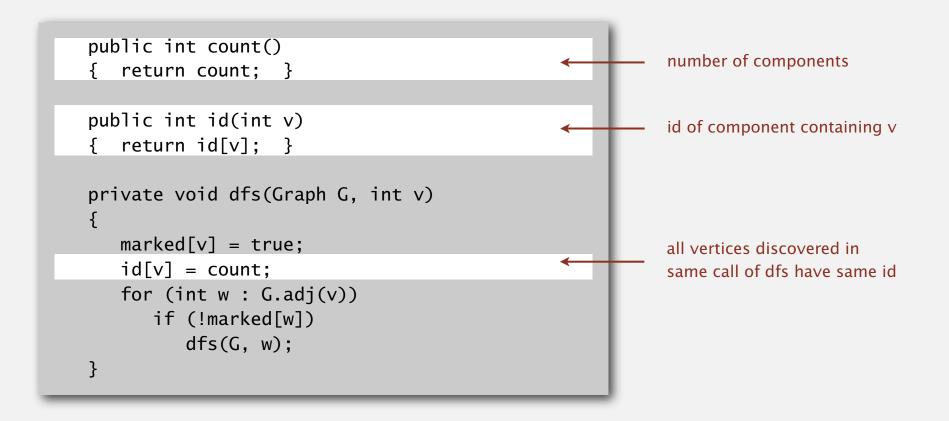


done

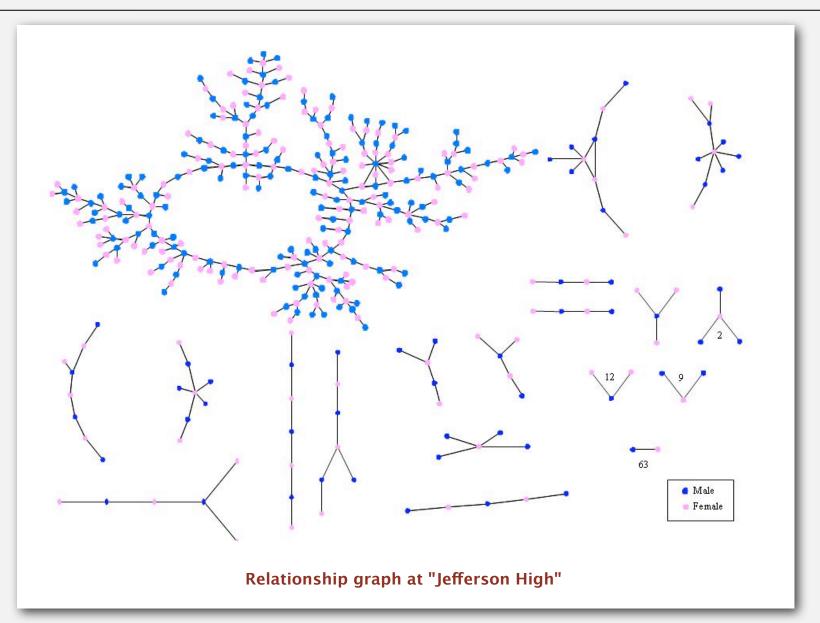
Finding connected components with DFS



Finding connected components with DFS (continued)



Connected components application: study spread of STDs

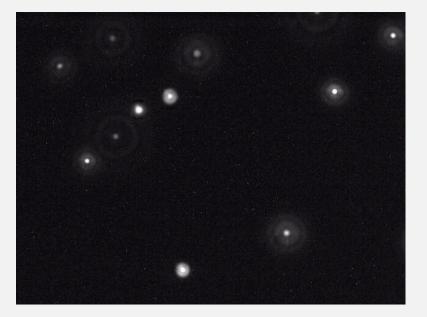


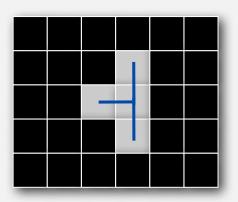
Peter Bearman, James Moody, and Katherine Stovel. Chains of affection: The structure of adolescent romantic and sexual networks. American Journal of Sociology, 110(1): 44-99, 2004.

Connected components application: particle detection

Particle detection. Given grayscale image of particles, identify "blobs."

- Vertex: pixel.
- Edge: between two adjacent pixels with grayscale value \ge 70.
- Blob: connected component of 20-30 pixels.





black = 0white = 255

Particle tracking. Track moving particles over time.

4.1 UNDIRECTED GRAPHS

Algorithms

Robert Sedgewick | Kevin Wayne

http://algs4.cs.princeton.edu

challenges

introduction

depth-first search

breadth-first search

connected components

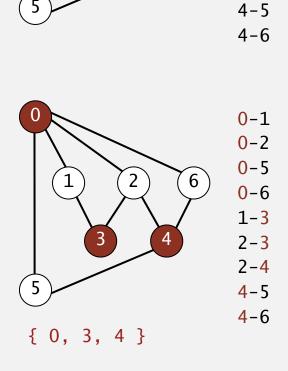
graph APt

Problem. Is a graph bipartite?

How difficult?

- Any programmer could do it.
- Typical diligent algorithms student could do it.
 - Hire an expert.
 - Intractable.
 - No one knows.
 - Impossible.

simple DFS-based solution (see textbook)



2

6

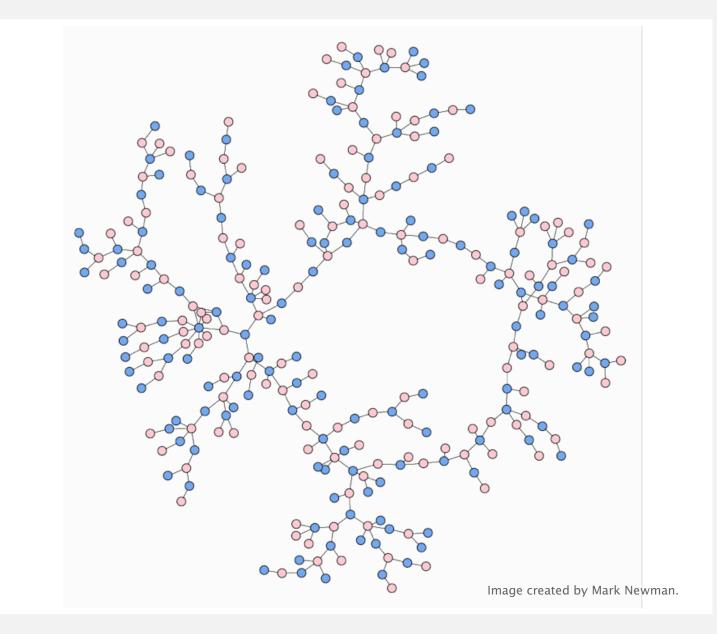
0-1 0-2 0-5

0-6 1-3

2 - 3

2 - 4

Bipartiteness application: is dating graph bipartite?

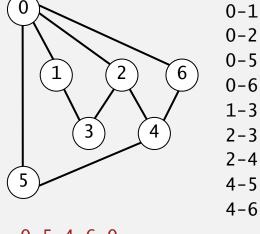


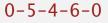
Problem. Find a cycle.

How difficult?

- Any programmer could do it.
- Typical diligent algorithms student could do it.
 - Hire an expert.
 - Intractable.
 - No one knows.
 - Impossible.

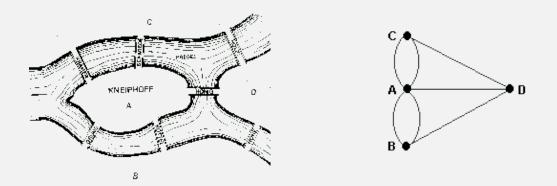
simple DFS-based solution (see textbook)





The Seven Bridges of Königsberg. [Leonhard Euler 1736]

"... in Königsberg in Prussia, there is an island A, called the Kneiphof; the river which surrounds it is divided into two branches ... and these branches are crossed by seven bridges. Concerning these bridges, it was asked whether anyone could arrange a route in such a way that he could cross each bridge once and only once."



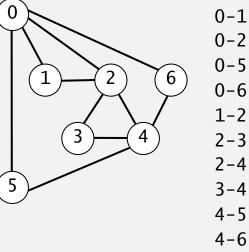
Euler tour. Is there a (general) cycle that uses each edge exactly once? Answer. Yes iff connected and all vertices have even degree.

Problem. Find a (general) cycle that uses every edge exactly once.

How difficult?

- Any programmer could do it.
- Typical diligent algorithms student could do it.
 - Hire an expert.
 - Intractable.
 - No one knows.
 - Impossible.

Eulerian tour (classic graph-processing problem)



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

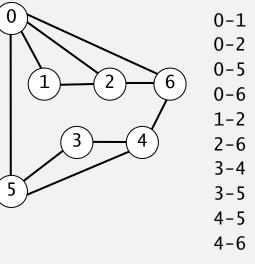
Problem. Find a cycle that visits every vertex exactly once.

How difficult?

- Any programmer could do it.
- Typical diligent algorithms student could do it.
- Hire an expert.
- 🗸 Intractable. 🔨
 - No one knows.

Hamiltonian cycle (classical NP-complete problem)

• Impossible.



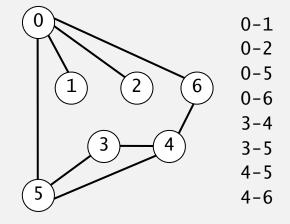
0-5-3-4-6-2-1-0

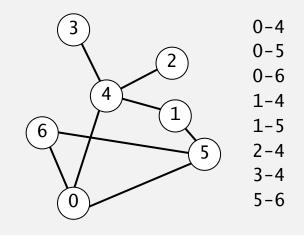
Problem. Are two graphs identical except for vertex names?

How difficult?

- Any programmer could do it.
- Typical diligent algorithms student could do it.
- Hire an expert.
- Intractable.
- ✓ No one knows.
 - Impossible.

right somorphism is longstanding open problem





 $0 \leftrightarrow 4$, $1 \leftrightarrow 3$, $2 \leftrightarrow 2$, $3 \leftrightarrow 6$, $4 \leftrightarrow 5$, $5 \leftrightarrow 0$, $6 \leftrightarrow 1$

Problem. Lay out a graph in the plane without crossing edges?

How difficult?

- Any programmer could do it.
- Typical diligent algorithms student could do it.
- ✓ Hire an expert.
 - Intractable.
 - No one knows.
 - Impossible.

linear-time DFS-based planarity algorithm discovered by Tarjan in 1970s (too complicated for most practitioners)

