Outline

1 What are Graphs?

2 Undirected Graphs

3 Depth-First Search (DFS)

4 Breadth-First Search (BFS)

5 Symbol Graphs

A graph is a set of V vertices connected pairwise by E edges



A graph is a set of V vertices connected pairwise by E edges



We use the names 0 through V-1 for the vertices in a V-vertex graph

A graph is a set of V vertices connected pairwise by E edges



We use the names 0 through V - 1 for the vertices in a V-vertex graph

We use the notation v-w to refer to an edge that connects vertices v and w

A graph is a set of V vertices connected pairwise by E edges



We use the names 0 through V-1 for the vertices in a V-vertex graph

We use the notation v-w to refer to an edge that connects vertices v and w

A self-loop is an edge that connects a vertex to itself

A graph is a set of V vertices connected pairwise by E edges



We use the names 0 through V-1 for the vertices in a V-vertex graph

We use the notation v-w to refer to an edge that connects vertices v and w

A self-loop is an edge that connects a vertex to itself

Parallel edges are edges that connect the same pair of vertices

The degree of a vertex is the number of vertices connected to it



The degree of a vertex is the number of vertices connected to it

A path is a sequence of vertices connected by edges



The degree of a vertex is the number of vertices connected to it

A path is a sequence of vertices connected by edges

A cycle is a path with at least one edge whose first and last vertices are the same



The degree of a vertex is the number of vertices connected to it

A path is a sequence of vertices connected by edges

A cycle is a path with at least one edge whose first and last vertices are the same $% \left({{{\rm{A}}_{\rm{B}}}} \right)$

The length of a path or a cycle is its number of edges



The degree of a vertex is the number of vertices connected to it

A path is a sequence of vertices connected by edges

A cycle is a path with at least one edge whose first and last vertices are the same $% \left({{{\rm{A}}_{\rm{B}}}} \right)$

The length of a path or a cycle is its number of edges

A graph is connected if there is a path from every vertex to every other vertex in the graph



The degree of a vertex is the number of vertices connected to it

A path is a sequence of vertices connected by edges

A cycle is a path with at least one edge whose first and last vertices are the same $% \label{eq:constraint}$

The length of a path or a cycle is its number of edges

A graph is connected if there is a path from every vertex to every other vertex in the graph

A graph that is not connected consists of a set of connected components, which are maximal connected subgraphs





An acyclic graph is a graph with no cycles

a bipartite graph



An acyclic graph is a graph with no cycles

A tree is an acyclic connected graph



An acyclic graph is a graph with no cycles

A tree is an acyclic connected graph

A bipartite graph is a graph whose vertices can be divided into two sets such that all edges connect a vertex in one set with a vertex in the other set



a bipartite graph



Graph applications

Graph	Vertex	Edge
communication	telephone, computer	fiber optic cable
circuit	gate, register, processor	wire
mechanical	joint	rod, beam, spring
financial	stock, currency	transactions
transportation	intersection	street
internet	class C network	connection
game	board position	legal move
social relationship	person	friendship
neural network	neuron	synapse
protein network	protein	protein-protein interaction
molecule	atom	bond

Example: Internet graph



Example: facebook graph



Example: c.elegans connectome graph



Example: coauthorship graph



Some graph-processing problems

Problem	Description
<i>s-t</i> path	is there a path between s and t?
shortest <i>s</i> - <i>t</i> path	what is the shortest path between s and t ?

🔳 Graph	
Graph(int V)	create a V-vertex graph with no edges
Graph(In in)	read a graph from input stream in
int V()	number of vertices
int E()	number of edges
<pre>void addEdge(int v, int w)</pre>	add edge <i>v-w</i> to this graph
Iterable <integer> adj(int v)</integer>	vertices adjacent to v
int degree(int v)	degree of v

🔳 Graph	
Graph(int V)	create a V-vertex graph with no edges
Graph(In in)	read a graph from input stream in
int V()	number of vertices
int E()	number of edges
<pre>void addEdge(int v, int w)</pre>	add edge <i>v-w</i> to this graph
<pre>Iterable<integer> adj(int v)</integer></pre>	vertices adjacent to v
int degree(int v)	degree of v

_ ~/workspace/dsa/programs

Graph input format

\$ more ../data/tinyG.txt
13 13
0 5 4 3 0 1 9 12 6 4 5 4 0 2
11 12 9 10 0 6 7 8 9 11 5 3



🔳 Graph	
Graph(int V)	create a V-vertex graph with no edges
Graph(In in)	read a graph from input stream in
int V()	number of vertices
int E()	number of edges
<pre>void addEdge(int v, int w)</pre>	add edge <i>v-w</i> to this graph
Iterable <integer> adj(int v)</integer>	vertices adjacent to v
int degree(int v)	degree of v

_ ~/workspace/dsa/programs

Graph input format

\$ more ../data/tinyG.txt
13 13
0 5 4 3 0 1 9 12 6 4 5 4 0 2
11 12 9 10 0 6 7 8 9 11 5 3



Typical graph-processing code

```
public static int degree(Graph G, int v) {
    int degree = 0;
    for (int w : G.adj(v)) {
        degree++;
    }
    return degree;
}
```
Graph representations

- Edge list: maintain a list of the edges (linked list or array)
- Adjacency matrix: maintain a V-by-V matrix M, such that M[v][w] is 1 if there is an edge from v to w, and 0 otherwise
- Adjacency list: maintain a vertex-indexed array of lists

Graph representations

- Edge list: maintain a list of the edges (linked list or array)
- Adjacency matrix: maintain a V-by-V matrix M, such that M[v][w] is 1 if there is an edge from v to w, and 0 otherwise
- Adjacency list: maintain a vertex-indexed array of lists

Representation	Space	Add edge	ls v-w an edge?	Enumerate adj(v)
edge list	E	1	E	E
adjacency matrix	V^2	1	1	V
adjacency list	E + V	1	degree(v)	degree(v)

Performance characteristics

```
🕑 Graph.java
package dsa:
import stdlib.In;
import stdlib.StdOut:
public class Graph {
    private LinkedBag < Integer > [] adj;
    private int V:
    private int E;
    public Graph(int V) {
        adj = (LinkedBag<Integer>[]) new LinkedBag[V];
        for (int v = 0: v < V: v++) {
            adj[v] = new LinkedBag<Integer>();
        3
        this.V = V;
        this.E = 0:
    3
    public Graph(In in) {
        this(in readInt()):
        int E = in.readInt();
        for (int i = 0; i < E; i++) {
            int v = in.readInt();
            int w = in.readInt():
            addEdge(v, w);
        3
    3
    public int V() {
        return V;
    3
    public int E() {
        return E:
```

🕑 Graph.java 3 public void addEdge(int v, int w) { adi[v].add(w); adj[w].add(v); E++: 3 public Iterable <Integer > adj(int v) { return adj[v]; 3 public int degree(int v) { return adj[v].size(); public String toString() { StringBuilder sb = new StringBuilder(): sb.append(V + " vertices, " + E + " edges\n"); for (int v = 0; v < V; v++) { sb.append(v + ": "); for (int w : adj[v]) { sb.append(w + " "); 3 $sb.append("\n");$ return sb.toString().strip(); 3 public static void main(String[] args) { String filename = args[0]; In in = new In(filename); Graph G = new Graph(in); StdOut.println(G):

}

🕼 Graph.java
}

Goal: systematically traverse a graph













Goal: systematically traverse a graph

Idea: mimic maze exploration













Goal: systematically traverse a graph

Idea: mimic maze exploration

Typical applications

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













Goal: systematically traverse a graph

Idea: mimic maze exploration

Typical applications

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

To visit a vertex v

- Mark vertex v as visited
- Recursively visit all unmarked vertices adjacent to \boldsymbol{v}













Goal: systematically traverse a graph

Idea: mimic maze exploration

Typical applications

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

To visit a vertex v

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

Data structures

- Boolean array marked[] to mark visited vertices
- Integer array edgeTo[] to keep track of paths; edgeTo[w] = v means that edge v-w taken to visit w for first time













Design pattern for graph processing: decouple graph data type from graph processing

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

I Paths		
	boolean hasPathTo(int v)	is there a path from s to v ?
	<pre>Iterable<integer> pathTo(int v)</integer></pre>	path from s to v, or null

Design pattern for graph processing: decouple graph data type from graph processing

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

1	I Paths			
	boolean hasPathTo(int v)	is there a path from s to v ?		
	<pre>Iterable<integer> pathTo(int v)</integer></pre>	path from s to v , or null		

Typical graph-processing code

```
DFSPaths paths = new DFSPaths(G, s);
for (int v = 0; v < G.V(); v++) {
    if (paths.hasPathTo(v)) {
        StdOut.println(v);
    }
}
```

```
🕼 DFSPaths.java
package dsa:
import stdlib.In;
import stdlib.StdOut:
public class DFSPaths {
    private int s;
    private boolean[] marked:
    private int[] edgeTo;
    public DFSPaths(Graph G, int s) {
        this.s = s;
        marked = new boolean [G, V()]:
        edgeTo = new int[G.V()];
        dfs(G. s):
    public boolean hasPathTo(int v) {
        return marked[v]:
    3
    public Iterable <Integer > pathTo(int v) {
        if (lhasPathTo(v)) {
            return null:
        ι
        LinkedStack<Integer> path = new LinkedStack<Integer>():
        for (int x = v; x != s; x = edgeTo[x]) {
            path.push(x);
        3
        path.push(s);
        return path;
    3
    private void dfs(Graph G, int v) {
        marked[v] = true;
```

🕼 DFSPaths.java

```
for (int w : G.adi(v)) {
            if (!marked[w]) {
                edgeTo[w] = v;
                dfs(G, w);
            3
   public static void main(String[] args) {
        In in = new In(args[0]);
        int s = Integer.parseInt(args[1]);
        Graph G = \mathbf{new} Graph(in);
        DFSPaths dfs = new DFSPaths(G, s):
        for (int v = 0; v < G.V(); v++) {
            if (dfs.hasPathTo(v)) {
                StdOut.printf("%d to %d: ", s, v);
                for (int x : dfs.pathTo(v)) {
                    if (x == s) {
                        StdOut.print(x);
                    } else {
                        StdOut.print("-" + x);
                    3
                ι
                StdOut.println();
            } else {
                StdOut.printf("%d to %d: not connected\n", s, v);
            3
3
```

	marked[]	adj[]
đfs(0)	0 T 1 2 3 4 5	$\begin{array}{cccccccccccccccccccccccccccccccccccc$
dfs(2) check 0	0 T 1 2 T 3 4 5	$\begin{array}{cccccccccccccccccccccccccccccccccccc$
dfs(1) check 0 check 2 1 done	0 T 1 T 2 T 3 4 5	$\begin{array}{cccccccccccccccccccccccccccccccccccc$
dfs(3)	0 T 1 T 2 T 3 T 4 5	$\begin{array}{cccccccccccccccccccccccccccccccccccc$
dfs(5) check 3 check 0 5 done	0 T 1 T 2 T 3 T 4 5 T	$\begin{array}{cccccccccccccccccccccccccccccccccccc$
dfs(4) check 3 check 2 4 done check 2 3 done check 4 2 done	0 T 1 T 2 T 3 T 5 T	$\begin{array}{cccccccccccccccccccccccccccccccccccc$

Goal: given a graph and a source vertex $\boldsymbol{s},$ support queries of the form

- Is there a path from s to a given target vertex v?
- If so, find a shortest such path (one with minimal number of edges)







Goal: given a graph and a source vertex $\boldsymbol{s},$ support queries of the form

- Is there a path from s to a given target vertex v?
- If so, find a shortest such path (one with minimal number of edges)

Repeat until queue is empty

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







```
BreadthFirstPaths.java
package dsa:
import stdlib.In;
import stdlib.StdOut:
public class BFSPaths {
    private int s;
    private boolean[] marked:
    private int[] edgeTo:
    private int[] distTo;
    public BFSPaths(Graph G, int s) {
        this.s = s:
        marked = new boolean[G.V()];
        distTo = new int[G,V()]:
        for (int v = 0; v < G.V(); v++) {
            distTo[v] = Integer.MAX VALUE:
        3
        edgeTo = new int[G.V()]:
        bfs(G. s):
    public boolean hasPathTo(int v) {
        return marked[v]:
    3
    public Iterable < Integer > pathTo(int v) {
        if (lhasPathTo(v)) {
            return null:
        3
        LinkedStack<Integer> path = new LinkedStack<Integer>();
        for (int x = y; x \models s; x = edgeTo[x]) {
            path.push(x);
        3
        path.push(s);
```

```
🕑 BreadthFirstPaths.java
```

```
return path:
3
public int distTo(int v) {
    return distTo[v];
3
private void bfs(Graph G. int s) {
    LinkedQueue < Integer > q = new LinkedQueue < Integer > ();
    marked[s] = true;
    distTo[s] = 0:
    q.enqueue(s);
    while (!g.isEmptv()) {
        int v = q.dequeue();
        for (int w : G.adi(v)) {
            if (!marked[w]) {
                marked[w] = true:
                edgeTo[w] = v:
                distTo[w] = distTo[v] + 1;
                q.enqueue(w);
           3
        3
3
public static void main(String[] args) {
    In in = new In(args[0]);
    int s = Integer.parseInt(args[1]);
    Graph G = new Graph(in):
    BFSPaths bfs = new BFSPaths(G, s);
    for (int v = 0; v < G.V(); v++) {
        if (bfs,hasPathTo(v)) {
            StdOut.printf("%d to %d (%d): ", s, v, bfs,distTo(v));
            for (int x : bfs.pathTo(v)) {
                if(x = s)
```

🕑 BreadthFirstPaths.java

```
StdOut.print(x);
                   } else {
                        StdOut.print("-" + x);
                    3
                3
                StdOut.println();
           } else {
                StdOut.printf("%d to %d (-): not connected\n", s, v);
    3
}
```

Trace

q	marked[]	edgeTo[]	adj[]
0	0 T 1 2 3 4 5	0 1 2 3 4 5	$\begin{array}{cccccccccccccccccccccccccccccccccccc$
2 1 5	0 T 1 T 2 T 3 4 5 T	0 1 0 2 0 3 4 5 0	$\begin{array}{cccccccccccccccccccccccccccccccccccc$
1 5 3 4	0 T 1 T 2 T 3 T 4 T 5 T	0 1 0 2 0 3 2 4 2 5 0	0 2 1 5 1 0 2 2 0 1 3 4 3 5 4 2 4 3 2 5 3 0
5 3 4	0 T 1 T 2 T 3 T 4 T 5 T	0 1 0 2 0 3 2 4 2 5 0	0 2 1 5 1 0 2 2 0 1 3 4 3 5 4 2 4 3 2 5 3 0
3 4	0 T 1 T 2 T 3 T 4 T 5 T	0 1 0 2 0 3 2 4 2 5 0	0 2 1 5 1 0 2 2 0 1 3 4 3 5 4 2 4 3 2 5 3 0
4	0 T 1 T 2 T 3 T 4 T 5 T	0 1 0 2 0 3 2 4 2 5 0	$\begin{array}{cccccccccccccccccccccccccccccccccccc$

Typical applications involve processing graphs defined in files or on web pages, using strings, not integer indices, to define and refer to vertices

Typical applications involve processing graphs defined in files or on web pages, using strings, not integer indices, to define and refer to vertices

To accommodate such applications, we define an input format with these properties

- Vertex names are strings
- A specified delimiter separates vertex names (to allow for the possibility of spaces in names)
- Each line represents a set of edges, connecting the first vertex name on the line to each of the other vertices named on the line
- The number of vertices V and the number of edges E are both implicitly defined

Typical applications involve processing graphs defined in files or on web pages, using strings, not integer indices, to define and refer to vertices

To accommodate such applications, we define an input format with these properties

- Vertex names are strings
- A specified delimiter separates vertex names (to allow for the possibility of spaces in names)
- Each line represents a set of edges, connecting the first vertex name on the line to each of the other vertices named on the line
- The number of vertices V and the number of edges E are both implicitly defined

Example (routes.txt)

>_	"/workspace/dsa/programs
JF	K MCO
OR	D DEN
OR	D HOU
DF	W PHX
JF	K ATL
• •	

Typical applications involve processing graphs defined in files or on web pages, using strings, not integer indices, to define and refer to vertices

To accommodate such applications, we define an input format with these properties

- Vertex names are strings
- A specified delimiter separates vertex names (to allow for the possibility of spaces in names)
- Each line represents a set of edges, connecting the first vertex name on the line to each of the other vertices named on the line
- The number of vertices V and the number of edges E are both implicitly defined

Example (routes.txt)

>_	"/workspace/dsa/programs
JF	K MCD
OR	D DEN
OR	D HOU
DF	W PHX
JF	K ATL
• •	·

Example (movies.txt)

>_ ~/workspace/dsa/programs
'Breaker' Morant (1980)/Brown, Bryan (I)/Henderson, Dick (II)/ 'burbs, The (1989)/Jayne, Billy/Hovard, Rance/Ducommun, Rick/ 'Crocodile' Dundee II (1988)/Jbara, Gregory/Holt, Jim (I)/ *batteries not included (1987)/Aldredge, Tom/Boutsikaris, Dennis/ And Justice for All (1979)/Williams, Jonathan (XI)/

API for graphs with symbolic vertex names

I SymbolGraph			
SymbolGraph(String filename, String delim)	build graph specified in <i>filename</i> using <i>delim</i> to separate vertex names		
boolean contains(String key)	is key a vertex?		
int indexOf(String key)	index associated with key		
String nameOf(int v)	key associated with index v		
Graph G()	underlying graph as a Graph object		
>_ ~/workspace/dsa/programs			
--	--	--	
<pre>\$ java dsa.SymbolGraph/data/routes.txt " " Data models and the second s</pre>			
Done reading foutes.txt JFK			
ORD ATL			
NCO LAX			
LAS			
<pre>rnA <ctrl-d></ctrl-d></pre>			

>_ ~/workspace/dsa/programs

\$ java dsa.SymbolGraph ../data/routes.txt " "
Done reading routes.txt
JFK
ORD
ATL
MCO
LAX
LAS
PHX
<<trl-d>

>_ ~/workspace/dsa/program

```
$ java dsa.SymbolGraph .../data/movies.txt "/"
Done reading movies.txt
Tin Men (1987)
Horshey, Barbara
Geppi, Cindy
...
Blumenfeld, Alan
DeBoy, David
Bacon, Kevin
Woodsman, The (2004)
Wild Things (1998)
...
Apollo 13 (1995)
Animal House (1978)
<<trl-d>
```

```
🕼 SymbolGraph.java
package dsa:
import stdlib.In;
import stdlib.StdIn:
import stdlib.StdOut;
public class SymbolGraph {
    private SeparateChainingHashST <String, Integer> st:
    private String[] kevs:
    private Graph G;
    public SymbolGraph(In in, String delim) {
        st = new SeparateChainingHashST <>();
        String[] lines = in.readAllLines();
        for (String line : lines) {
             String[] a = line.split(delim);
             for (int i = 0; i < a, length; i++) {
                if (!st.contains(a[i])) {
                     st.put(a[i].st.size()):
        kevs = new String[st.size()];
        for (String name : st.kevs()) {
             kevs[st.get(name)] = name;
        G = new Graph(st.size());
        for (String line : lines) {
             String[] a = line.split(delim);
            int v = st.get(a[0]);
            for (int i = 1; i < a.length; i++) {</pre>
                int w = st.get(a[i]);
                G.addEdge(v. w);
            3
        3
```

🕑 SymbolGraph.java

```
3
public boolean contains(String s) {
    return st.contains(s);
3
public int indexOf(String s) {
    return st.get(s);
3
public String nameOf(int v) {
    return keys[v];
public Graph graph() {
    return G;
public static void main(String[] args) {
    In in = new In(args[0]);
    String delim = args[1];
    SymbolGraph sg = new SymbolGraph(in, delim);
    Graph graph = sg.graph();
    while (!StdIn.isEmptv()) {
        String source = StdIn.readLine():
        if (sg.contains(source)) {
            int s = sg.indexOf(source);
            for (int v : graph.adj(s)) {
                StdOut.println(" " + sg.nameOf(v));
            ι
        } else {
            StdOut.println(source + " not in database"):
3
```

3 SymbolGraph.java
}

```
DegreesOfSeparation.java
import dsa.BFSPaths:
import dsa.Graph;
import dsa.SymbolGraph;
import stdlib.In:
import stdlib.StdIn;
import stdlib.StdOut;
public class DegreesOfSeparation {
    public static void main(String[] args) {
        String filename = args[0];
        String delim = args[1]:
        String source = args[2];
        In in = new In(filename);
        SymbolGraph sg = new SymbolGraph(in, delim);
        Graph G = sg.graph():
        if (!sg.contains(source)) {
            StdOut.println(source + " not in database"):
            return:
        int s = sg index\Omega f(source):
        BFSPaths bfs = new BFSPaths(G. s):
        while (!StdIn.isEmptv()) {
            String sink = StdIn.readLine();
            if (sg.contains(sink)) {
                int t = sg.indexOf(sink);
                if (bfs.hasPathTo(t)) {
                    for (int v : bfs.pathTo(t)) {
                         StdOut.println(" " + sg.nameOf(v));
                } else {
                    StdOut.println(source + " and " + sink + " are not connected");
                ι
            } else {
                StdOut.println(sink + " not in database");
            }
```

🕼 DegreesOfSeparation.java		
}		

>_ ~/workspace/dsa/programs
<pre>\$ java DegreesOfSeparation/data/routes.txt " " JFK Done reading routes.txt LAS JFK ORD PHX</pre>
LAS DFW JFK ORD DFW <ctrl-d></ctrl-d>

>_ ~/workspace/dsa/programs

```
$ java DegreesOfSeparation ../data/routes.txt " " JFK
Done reading routes.txt
LAS
JFK
ORD
PHX
LAS
DFW
JFK
ORD
DFW
Strl-d>
```

>_ ~/workspace/dsa/programs

```
$ java DegreesOfSeparation ../data/movies.txt "/" "Bacon, Kevin"
Kidman, Nicole
Bacon, Kevin
Woodsman, The (2004)
Grier, David Alan
Bewitched (2005)
Kidman, Nicole
Grant, Cary
Bacon, Kevin
Planes, Trains & Automobiles (1987)
Martin, Steve (I)
Dead Men Don't Wear Plaid (1982)
Grant, Cary
<ctrl-d>
```