

# Graphs

## Chapter Summary

- Graphs and Graph Models
- Graph Terminology and Special Types of Graphs
- Representing Graphs and Graph Isomorphism
- Connectivity
- Euler and Hamiltonian Graphs
- Shortest-Path Problems
- Planar Graphs
- Graph Coloring

# Graphs and Graph Models

3

## Section Summary

- Introduction to Graphs
- Graph Taxonomy
- Graph Models

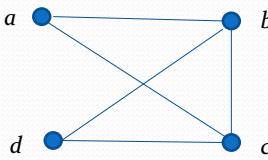
4

# Graphs

**Definition:** A *graph*  $G = (V, E)$  consists of a nonempty set  $V$  of *vertices* (or *nodes*) and a set  $E$  of *edges*. Each edge has either one or two vertices associated with it, called its *endpoints*. An edge is said to *connect* its endpoints.

**Example:**

This is a graph with four vertices and five edges.



**Remarks:**

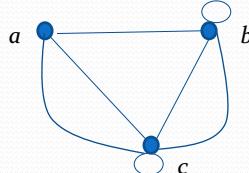
- The graphs we study here are unrelated to graphs of functions studied in Chapter 2.
- We have a lot of freedom when we draw a picture of a graph. All that matters is the connections made by the edges, not the particular geometry depicted. For example, the lengths of edges, whether edges cross, how vertices are depicted, and so on, do not matter.
- A graph with an infinite vertex set is called an *infinite graph*. A graph with a finite vertex set is called a *finite graph*. We (following the text) restrict our attention to finite graphs.

5

# Some Terminology

- In a *simple graph* each edge connects two different vertices and no two edges connect the same pair of vertices.
- Multigraphs* may have multiple edges connecting the same two vertices. When  $m$  different edges connect the vertices  $u$  and  $v$ , we say that  $\{u,v\}$  is an edge of *multiplicity*  $m$ .
- An edge that connects a vertex to itself is called a *loop*.
- A *pseudograph* may include loops, as well as multiple edges connecting the same pair of vertices.

**Example:**  
This pseudograph has both multiple edges and a loop.



**Remark:** There is no standard terminology for graph theory. So, it is crucial that you understand the terminology being used whenever you read material about graphs.

6

## Directed Graphs

**Definition:** A *directed graph* (or *digraph*)  $G = (V, E)$  consists of a nonempty set  $V$  of *vertices* (or *nodes*) and a set  $E$  of *directed edges* (or *arcs*).

Each edge is associated with an **ordered pair** of vertices. The directed edge associated with the ordered pair  $(u,v)$  is said to *start at u* and *end at v*.

**Remark:**

- Graphs where the end points of an edge are not ordered are said to be *undirected graphs*.

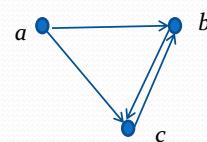
7

## Some Terminology (continued)

- A *simple directed graph* has no loops and no multiple edges.

**Example:**

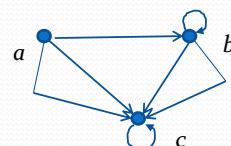
This is a directed graph with three vertices and four edges.



- A *directed multigraph* may have multiple directed edges. When there are  $m$  directed edges from the vertex  $u$  to the vertex  $v$ , we say that  $(u,v)$  is an edge of *multiplicity m*.

**Example:**

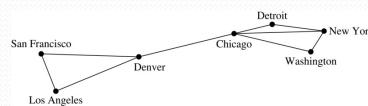
In this directed multigraph the multiplicity of  $(a,b)$  is 1 and the multiplicity of  $(b,c)$  is 2.



8

## Graph Models: Computer Networks

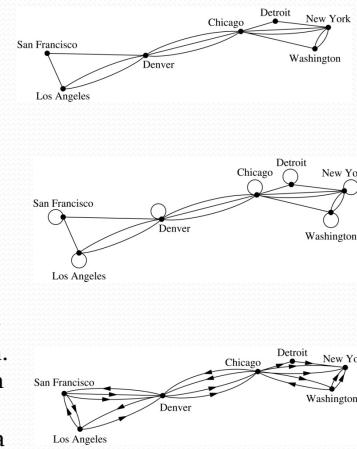
- When we build a graph model, we use the appropriate type of graph to capture the important features of the application.
- We illustrate this process using graph models of different types of computer networks. In all these graph models, the vertices represent data centers and the edges represent communication links.
- To model a computer network where we are only concerned whether two data centers are connected by a communications link, we use a simple graph. This is the appropriate type of graph when we only care whether two data centers are directly linked (and not how many links there may be) and all communications links work in both directions.



9

## Graph Models: Computer Networks (*continued*)

- To model a computer network where we care about the number of links between data centers, we use a multigraph.
- To model a computer network with diagnostic links at data centers, we use a pseudograph, as loops are needed.
- To model a network with multiple one-way links, we use a directed multigraph. Note that we could use a directed graph without multiple edges if we only care whether there is at least one link from a data center to another data center.



10

## Graph Terminology: Summary

- To understand the structure of a graph and to build a graph model, we ask these questions:
  - Are the edges of the graph undirected or directed (or both)?
  - If the edges are undirected, are multiple edges present that connect the same pair of vertices? If the edges are directed, are multiple directed edges present?
  - Are loops present?

**TABLE 1** Graph Terminology.

| Type                  | Edges                   | Multiple Edges Allowed? | Loops Allowed? |
|-----------------------|-------------------------|-------------------------|----------------|
| Simple graph          | Undirected              | No                      | No             |
| Multigraph            | Undirected              | Yes                     | No             |
| Pseudograph           | Undirected              | Yes                     | Yes            |
| Simple directed graph | Directed                | No                      | No             |
| Directed multigraph   | Directed                | Yes                     | Yes            |
| Mixed graph           | Directed and undirected | Yes                     | Yes            |

11

## Other Applications of Graphs

- We will illustrate how graph theory can be used in models of:
  - Social networks
  - Communications networks
  - Information networks
  - Software design
  - Transportation networks
  - Biological networks
- It's a challenge to find a subject to which graph theory has not yet been applied. Can you find an area without applications of graph theory?

12

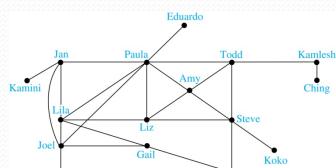
## Graph Models: Social Networks

- Graphs can be used to model social structures based on different kinds of relationships between people or groups.
- In a *social network*, vertices represent individuals or organizations and edges represent relationships between them.
- Useful graph models of social networks include:
  - *friendship graphs* - undirected graphs where two people are connected if they are friends (in the real world, on Facebook, or in a particular virtual world, and so on.)
  - *collaboration graphs* - undirected graphs where two people are connected if they collaborate in a specific way
  - *influence graphs* - directed graphs where there is an edge from one person to another if the first person can influence the second person

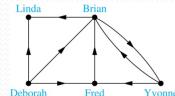
13

## Graph Models: Social Networks (continued)

**Example:** A friendship graph where two people are connected if they are Facebook friends.



**Example:** An influence graph



*Next Slide: Collaboration Graphs*

14

## Examples of Collaboration Graphs

- The *Hollywood graph* models the collaboration of actors in films.
  - We represent actors by vertices and we connect two vertices if the actors they represent have appeared in the same movie.
  - We will study the Hollywood Graph in Section 10.4 when we discuss Kevin Bacon numbers.
- An *academic collaboration graph* models the collaboration of researchers who have jointly written a paper in a particular subject.
  - We represent researchers in a particular academic discipline using vertices.
  - We connect the vertices representing two researchers in this discipline if they are coauthors of a paper.
  - We will study the academic collaboration graph for mathematicians when we discuss *Erdős numbers* in Section 10.4.

15

## Applications to Information Networks

- Graphs can be used to **model** different types of networks that link different types of information.
- In a *web graph*, web pages are represented by vertices and links are represented by directed edges.
  - A web graph models the web at a particular time.
  - We will explain how the web graph is used by search engines in Section 11.4.
- In a *citation network*:
  - Research papers in a particular discipline are represented by vertices.
  - When a paper cites a second paper as a reference, there is an edge from the vertex representing this paper to the vertex representing the second paper.

16

## Transportation Graphs

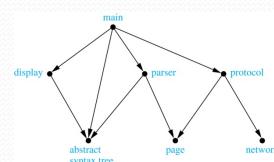
- Graph models are extensively used in the study of transportation networks.
- Airline networks can be modeled using directed multigraphs where
  - airports are represented by vertices
  - each flight is represented by a directed edge from the vertex representing the departure airport to the vertex representing the destination airport
- Road networks can be modeled using graphs where
  - vertices represent intersections and edges represent roads.
  - undirected edges represent two-way roads and directed edges represent one-way roads.

17

## Software Design Applications

- Graph models are extensively used in software design. We will introduce two such models here; one representing the dependency between the modules of a software application and the other representing restrictions in the execution of statements in computer programs.
- When a top-down approach is used to design software, the system is divided into modules, each performing a specific task.
- We use a *module dependency graph* to represent the dependency between these modules. These dependencies need to be understood before coding can be done.
  - In a module dependency graph vertices represent software modules and there is an edge from one module to another if the second module depends on the first.

**Example:** The dependencies between the seven modules in the design of a web browser are represented by this module dependency graph.



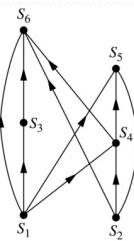
18

## Software Design Applications (continued)

- We can use a directed graph called a *precedence graph* to represent which statements must have already been executed before we execute each statement.
  - Vertices represent statements in a computer program
  - There is a directed edge from a vertex to a second vertex if the second vertex cannot be executed before the first

**Example:** This precedence graph shows which statements must already have been executed before we can execute each of the six statements in the program.

$S_1$      $a := 0$   
 $S_2$      $b := 1$   
 $S_3$      $c := a + 1$   
 $S_4$      $d := b + a$   
 $S_5$      $e := d + 1$   
 $S_6$      $e := c + d$

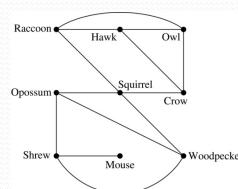


19

## Biological Applications

- Graph models are used extensively in many areas of the biological science. We will describe two such models, one to ecology and the other to molecular biology.
- *Niche overlap graphs* model competition between species in an ecosystem
  - Vertices represent species and an edge connects two vertices when they represent species who compete for food resources.

**Example:** This is the niche overlap graph for a forest ecosystem with nine species.

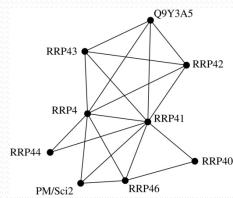


20

## Biological Applications (*continued*)

- We can model the interaction of proteins in a cell using a *protein interaction network*.
- In a *protein interaction graph*, vertices represent proteins and vertices are connected by an edge if the proteins they represent interact.
- Protein interaction graphs can be huge and can contain more than 100,000 vertices, each representing a different protein, and more than 1,000,000 edges, each representing an interaction between proteins
- Protein interaction graphs are often split into smaller graphs, called *modules*, which represent the interactions between proteins involved in a particular function.

**Example:** This is a module of the protein interaction graph of proteins that degrade RNA in a human cell.



21

## Graph Terminology and Special Types of Graphs

Section 10.2

22

## Section Summary

- Basic Terminology
- Some Special Types of Graphs
- Bipartite Graphs
- Bipartite Graphs and Matchings (*not currently included in overheads*)
- Some Applications of Special Types of Graphs (*not currently included in overheads*)
- New Graphs from Old

23

## Basic Terminology

**Definition 1.** Two vertices  $u, v$  in an undirected graph  $G$  are called *adjacent* (or *neighbors*) in  $G$  if there is an edge  $e$  between  $u$  and  $v$ . Such an edge  $e$  is called *incident with* the vertices  $u$  and  $v$  and  $e$  is said to *connect*  $u$  and  $v$ .

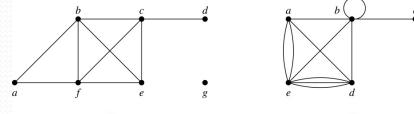
**Definition 2.** The set of all neighbors of a vertex  $v$  of  $G = (V, E)$ , denoted by  $N(v)$ , is called the *neighborhood* of  $v$ . If  $A$  is a subset of  $V$ , we denote by  $N(A)$  the set of all vertices in  $G$  that are adjacent to at least one vertex in  $A$ . So,  $N(A) = \bigcup_{v \in A} N(v)$ .

**Definition 3.** The *degree of a vertex in an undirected graph* is the number of edges incident with it, except that, by convention (but this is not always the case), a loop at a vertex contributes two to the degree of that vertex. The degree of the vertex  $v$  is denoted by  $\deg(v)$ .

24

## Degrees and Neighborhoods of Vertices

**Example:** What are the degrees and neighborhoods of the vertices in the graphs  $G$  and  $H$ ?



**Solution:**

- $G$ :  $\deg(a) = 2$ ,  $\deg(b) = \deg(c) = \deg(f) = 4$ ,  $\deg(d) = 1$ ,  
 $\deg(e) = 3$ ,  $\deg(g) = 0$ .  
 $N(a) = \{b, f\}$ ,  $N(b) = \{a, c, e, f\}$ ,  $N(c) = \{b, d, e, f\}$ ,  $N(d) = \{c\}$ ,  
 $N(e) = \{b, c, f\}$ ,  $N(f) = \{a, b, c, e\}$ ,  $N(g) = \emptyset$ .  
 $H$ :  $\deg(a) = 4$ ,  $\deg(b) = \deg(e) = 6$ ,  $\deg(c) = 1$ ,  $\deg(d) = 5$ .  
 $N(a) = \{b, d, e\}$ ,  $N(b) = \{a, b, c, d, e\}$ ,  $N(c) = \{b\}$ ,  
 $N(d) = \{a, b, e\}$ ,  $N(e) = \{a, b, d\}$ .

25

## Degrees of Vertices

**Theorem 1 (Handshaking Theorem):** If  $G = (V, E)$  is an undirected graph with  $m$  edges, then

$$2m = \sum_{v \in V} \deg(v)$$

**Proof:**

Each edge contributes twice to the degree count of all vertices. Hence, both the left-hand and right-hand sides of this equation equal twice the number of edges. ◀

*Think about the graph where vertices represent the people at a party and an edge connects two people who have shaken hands.*

26

## Handshaking Theorem

We now give two examples illustrating the usefulness of the handshaking theorem.

**Example:** How many edges are there in a graph with 10 vertices of degree six?

**Solution:** Because the sum of the degrees of the vertices is  $6 \cdot 10 = 60$ , the handshaking theorem tells us that  $2m = 60$ . So the number of edges  $m = 30$ .

**Example:** If a graph has 5 vertices, can each vertex have degree 3?

**Solution:** This is not possible by the handshaking theorem, because the sum of the degrees of the vertices  $3 \cdot 5 = 15$  is odd.

27

## Degree of Vertices (*continued*)

**Theorem 2:** An undirected graph has an even number of vertices of odd degree.

**Proof:** Let  $V_1$  be the vertices of even degree and  $V_2$  be the vertices of odd degree in an undirected graph  $G = (V, E)$  with  $m$  edges. Then

$$\text{even} \rightarrow 2m = \sum_{v \in V} \deg(v) = \sum_{v \in V_1} \deg(v) + \sum_{v \in V_2} \deg(v).$$

must be even since  $\deg(v)$  is even for each  $v \in V_1$

This sum must be even because  $2m$  is even and the sum of the degrees of the vertices of even degrees is also even. Because this is the sum of the degrees of all vertices of odd degree in the graph, there must be an even number of such vertices.

28

## Directed Graphs

Recall the definition of a directed graph.

**Definition:** A *directed graph*  $G = (V, E)$  consists of  $V$ , a nonempty set of *vertices* (or *nodes*), and  $E$ , a set of *directed edges* or *arcs*. Each edge is an ordered pair of vertices. The directed edge  $(u,v)$  is said to start at  $u$  and end at  $v$ .

**Definition:** Let  $(u,v)$  be an edge in  $G$ . Then  $u$  is the *initial vertex* of this edge and is *adjacent to*  $v$  and  $v$  is the *terminal* (or *end*) vertex of this edge and is *adjacent from*  $u$ .

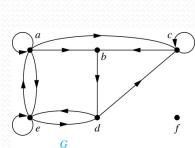
The initial and terminal vertices of a loop are the same.

29

## Directed Graphs (continued)

**Definition:** The *in-degree of a vertex*  $v$ , denoted  $\deg^-(v)$ , is the number of edges which terminate at  $v$ . The *out-degree of  $v$* , denoted  $\deg^+(v)$ , is the number of edges with  $v$  as their initial vertex. Note that a loop at a vertex contributes 1 to both the in-degree and the out-degree of the vertex.

**Example:** In the graph  $G$  we have



$$\begin{aligned} \deg^-(a) &= 2, \deg^-(b) = 2, \deg^-(c) = 3, \deg^-(d) = 2, \\ \deg^-(e) &= 3, \deg^-(f) = 0. \\ \deg^+(a) &= 4, \deg^+(b) = 1, \deg^+(c) = 2, \deg^+(d) = 2, \\ \deg^+(e) &= 3, \deg^+(f) = 0. \end{aligned}$$

30

## Directed Graphs (continued)

**Theorem 3:** Let  $G = (V, E)$  be a graph with directed edges. Then:

$$|E| = \sum_{v \in V} \deg^-(v) = \sum_{v \in V} \deg^+(v).$$

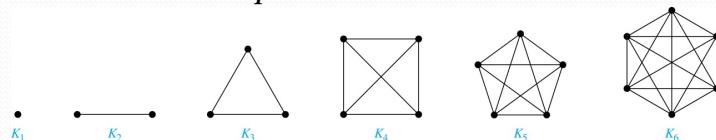
**Proof:** The first sum counts the number of outgoing edges over all vertices and the second sum counts the number of incoming edges over all vertices. It follows that both sums equal the number of edges in the graph.  $\blacktriangleleft$

31

## Special Types of Simple Graphs: Complete Graphs

A *complete graph on  $n$  vertices*, denoted by  $K_n$ , is the simple graph that contains exactly one edge between each pair of distinct vertices.

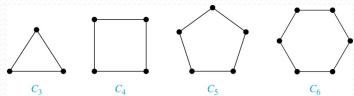
It is also called a *clique*



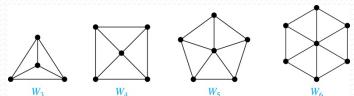
32

## Special Types of Simple Graphs: Cycles and Wheels

A *cycle*  $C_n$  for  $n \geq 3$  consists of  $n$  vertices  $v_1, v_2, \dots, v_n$ , and edges  $\{v_1, v_2\}, \{v_2, v_3\}, \dots, \{v_{n-1}, v_n\}, \{v_n, v_1\}$ .



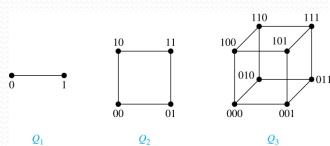
A *wheel*  $W_n$  is obtained by adding an additional vertex to a cycle  $C_n$  for  $n \geq 3$  and connecting this new vertex to each of the  $n$  vertices in  $C_n$  by new edges.



33

## Special Types of Simple Graphs: $n$ -Cubes

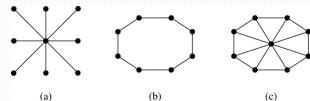
An  $n$ -dimensional hypercube, or  $n$ -cube,  $Q_n$ , is a graph with  $2^n$  vertices representing all bit strings of length  $n$ , where there is an edge between two vertices that differ in exactly one bit position.



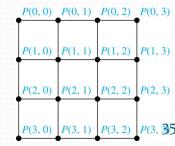
34

## Special Types of Graphs and Computer Network Architecture

Various special graphs play an important role in the design of computer networks.



- Some local area networks use a *star topology*, which is a complete bipartite graph  $K_{1,n}$ , as shown in (a). All devices are connected to a central control device.
- Other local networks are based on a *ring topology*, where each device is connected to exactly two others using  $C_n$ , as illustrated in (b). Messages may be sent around the ring.
- Others, as illustrated in (c), use a  $W_n$  – based topology, combining the features of a star topology and a ring topology.
- Various special graphs also play a role in parallel processing where processors need to be interconnected as one processor may need the output generated by another.
  - The *n-dimensional hypercube*, or *n-cube*,  $Q_n$ , is a common way to connect processors in parallel, e.g., Intel Hypercube.
  - Another common method is the *mesh network*, illustrated here for 16 processors.

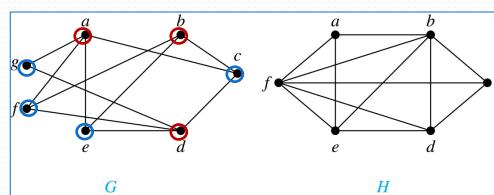


## Bipartite Graphs

**Definition:** A simple graph  $G$  is bipartite if  $V$  can be partitioned into two disjoint subsets  $V_1$  and  $V_2$  such that every edge connects a vertex in  $V_1$  and a vertex in  $V_2$ . In other words, there are no edges which connect two vertices in  $V_1$  or in  $V_2$ .

It is not hard to show that an equivalent definition of a bipartite graph is a graph where it is possible to color the vertices red or blue so that no two adjacent vertices are the same color.

$G$  is bipartite



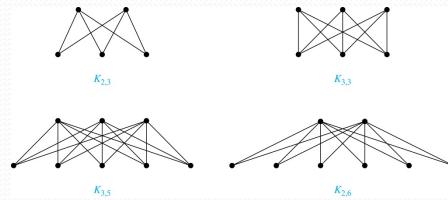
$H$  is not bipartite since if we color  $a$  red, then the adjacent vertices  $f$  and  $b$  must both be blue.

36

## Complete Bipartite Graphs

**Definition:** A *complete bipartite graph*  $K_{m,n}$  is a graph that has its vertex set partitioned into two subsets  $V_1$  of size  $m$  and  $V_2$  of size  $n$  such that there is an edge from every vertex in  $V_1$  to every vertex in  $V_2$ .

**Example:** We display four complete bipartite graphs here.

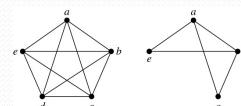


37

## New Graphs from Old

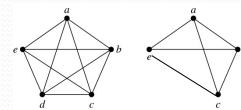
**Definition:** A proper *subgraph* of a graph  $G = (V, E)$  is a graph  $(W, F)$ , where  $W \subset V$  and  $F \subset E$ . Thus a subgraph  $H$  of  $G$  is a proper subgraph of  $G$  if  $H \neq G$ .

**Example:** Here we show  $K_5$  and one of its subgraphs.



**Definition:** Let  $G = (V, E)$  be a simple graph. The *subgraph induced* by a subset  $W$  of the vertex set  $V$  is the graph  $(W, F)$ , where the edge set  $F$  contains an edge in  $E$  if and only if both endpoints are in  $W$ .

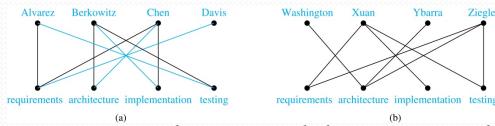
**Example:** Here we show  $K_5$  and the subgraph induced by  $W = \{a, b, c, e\}$ .



38

## Bipartite Graphs and Matchings

- Bipartite graphs are used to model applications that involve matching the elements of one set to elements in another, for example:
  - *Job assignments* - vertices represent the jobs and the employees, edges link employees with those jobs they have been trained to do. A common goal is to match jobs to employees so that the most jobs are done.



- *Marriage* - vertices represent the men and the women and edges link a man and a woman if they are an acceptable spouse. We may wish to find the largest number of possible marriages.

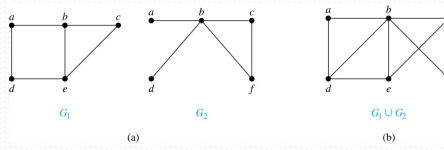
*See the text for more about matchings in bipartite graphs.*

39

## New Graphs from Old (*continued*)

**Definition:** The *union* of two simple graphs  $G_1 = (V_1, E_1)$  and  $G_2 = (V_2, E_2)$  is the simple graph with vertex set  $V_1 \cup V_2$  and edge set  $E_1 \cup E_2$ . The union of  $G_1$  and  $G_2$  is denoted by  $G_1 \cup G_2$ .

## Example:



40

# Representing Graphs and Graph Isomorphism

41

## Section Summary

- Adjacency Lists
- Adjacency Matrices
- Incidence Matrices
- Isomorphism of Graphs

42

## Representing Graphs: Adjacency Lists

**Definition:** An *adjacency list* can be used to represent a graph with no multiple edges by specifying the vertices that are adjacent to each vertex of the graph.

**Example:**

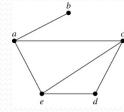


TABLE 1 An Adjacency List for a Simple Graph.

| Vertex | Adjacent Vertices |
|--------|-------------------|
| a      | b, c, e           |
| b      | a                 |
| c      | a, d, e           |
| d      | c, e              |
| e      | a, c, d           |

**Example:**

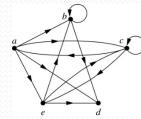


TABLE 2 An Adjacency List for a Directed Graph.

| Initial Vertex | Terminal Vertices |
|----------------|-------------------|
| a              | b, c, d, e        |
| b              | d                 |
| c              | a, c, e           |
| d              | e                 |
| e              | b, c, d           |

43

## Representation of Graphs: Adjacency Matrices

**Definition:** Suppose that  $G = (V, E)$  is a simple graph where  $|V| = n$ . Arbitrarily list the vertices of  $G$  as  $v_1, v_2, \dots, v_n$ .

- The *adjacency matrix*  $A_G$  of  $G$ , with respect to the listing of vertices, is the  $n \times n$  zero-one matrix with 1 as its  $(i, j)$ th entry when  $v_i$  and  $v_j$  are adjacent, and 0 as its  $(i, j)$ th entry when they are not adjacent.
- In other words, if the graphs adjacency matrix is

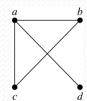
$$A_G = [a_{ij}], \text{ then}$$

$$a_{ij} = \begin{cases} 1 & \text{if } \{v_i, v_j\} \text{ is an edge of } G, \\ 0 & \text{otherwise.} \end{cases}$$

44

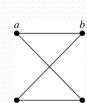
## Adjacency Matrices (continued)

**Example:**



$$\begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

The ordering of vertices is  $a, b, c, d$ .



$$\begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

The ordering of vertices is  $a, b, c, d$ .

When a graph is sparse, that is, it has few edges relatively to the total number of possible edges, it is much more efficient to represent the graph using an adjacency list than an adjacency matrix. But for a dense graph, which includes a high percentage of possible edges, an adjacency matrix is preferable.

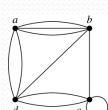
**Note:** The adjacency matrix of a simple graph is symmetric, i.e.,  $a_{ij} = a_{ji}$ .  
Also, since there are no loops, each diagonal entry  $a_{ii}$  for  $i = 1, 2, 3, \dots, n$ , is 0.

45

## Adjacency Matrices (continued)

- Adjacency matrices can also be used to represent graphs with loops and multiple edges.
- A loop at the vertex  $v_i$  is represented by a 1 (and not 2) at the  $(i, i)$ th position of the matrix.
- When multiple edges connect the same pair of vertices  $v_i$  and  $v_j$ , (or if multiple loops are present at the same vertex), the  $(i, j)$ th entry equals the number of edges connecting the pair of vertices.

**Example:** We give the adjacency matrix of the pseudograph shown here using the ordering of vertices  $a, b, c, d$ .



$$\begin{bmatrix} 0 & 3 & 0 & 2 \\ 3 & 0 & 1 & 1 \\ 0 & 1 & 1 & 2 \\ 2 & 1 & 2 & 0 \end{bmatrix}$$

46

## Adjacency Matrices (*continued*)

- Adjacency matrices can also be used to represent directed graphs. The matrix for a directed graph  $G = (V, E)$  has a 1 in its  $(i, j)$ th position if there is an edge from  $v_i$  to  $v_j$ , where  $v_1, v_2, \dots, v_n$  is a list of the vertices.
  - In other words, if the graphs adjacency matrix is  $A_G = [a_{ij}]$ , then
 
$$a_{ij} = \begin{cases} 1 & \text{if } \{v_i, v_j\} \text{ is an edge of } G, \\ 0 & \text{otherwise.} \end{cases}$$
  - The adjacency matrix for a directed graph does not have to be symmetric, because there may not be an edge from  $v_i$  to  $v_j$ , when there is an edge from  $v_j$  to  $v_i$ .
  - To represent directed multigraphs, the value of  $a_{ij}$  is the number of edges connecting  $v_i$  to  $v_j$ .

47

## Representation of Graphs: Incidence Matrices

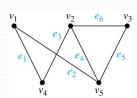
**Definition:** Let  $G = (V, E)$  be an undirected graph with vertices where  $v_1, v_2, \dots, v_n$  and edges  $e_1, e_2, \dots, e_m$ . The incidence matrix with respect to the ordering of  $V$  and  $E$  is the  $n \times m$  matrix  $M = [m_{ij}]$ , where

$$m_{ij} = \begin{cases} 1 & \text{when edge } e_j \text{ is incident with } v_i, \\ 0 & \text{otherwise.} \end{cases}$$

48

## Incidence Matrices (continued)

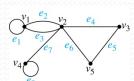
**Example:** Simple Graph and Incidence Matrix



$$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \end{bmatrix}$$

The rows going from top to bottom represent  $v_1$  through  $v_5$  and the columns going from left to right represent  $e_1$  through  $e_6$ .

**Example:** Pseudograph and Incidence Matrix



$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix}$$

The rows going from top to bottom represent  $v_1$  through  $v_5$  and the columns going from left to right represent  $e_1$  through  $e_8$ .

49

## Isomorphism of Graphs

**Definition:** The simple graphs  $G_1 = (V_1, E_1)$  and  $G_2 = (V_2, E_2)$  are *isomorphic* if there is a bijection (permutation)  $f$  from  $V_1$  to  $V_2$  with the property that  $a$  and  $b$  are adjacent in  $G_1$  if and only if  $f(a)$  and  $f(b)$  are adjacent in  $G_2$ , for all  $a$  and  $b$  in  $V_1$ .

Such a function  $f$  is called an *isomorphism*. Two simple graphs that are not isomorphic are called *nonisomorphic*.

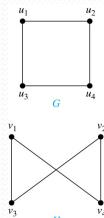
Isomorphisms preserve the structure of the graph.

50

## Isomorphism of Graphs (cont.)

**Example:** Show that the graphs  $G = (V, E)$  and  $H = (W, F)$  are isomorphic.

**Solution:** The function  $f$  with  $f(u_1) = v_1$ ,  $f(u_2) = v_4$ ,  $f(u_3) = v_3$ , and  $f(u_4) = v_2$  is a one-to-one correspondence between  $V$  and  $W$ . Note that adjacent vertices in  $G$  are  $u_1$  and  $u_2$ ,  $u_1$  and  $u_3$ ,  $u_2$  and  $u_4$ , and  $u_3$  and  $u_4$ . Each of the pairs  $f(u_1) = v_1$  and  $f(u_2) = v_4$ ,  $f(u_1) = v_1$  and  $f(u_3) = v_3$ ,  $f(u_2) = v_4$  and  $f(u_4) = v_2$ , and  $f(u_3) = v_3$  and  $f(u_4) = v_2$  consists of two adjacent vertices in  $H$ .



51

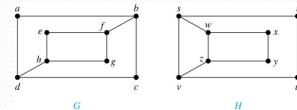
## Isomorphism of Graphs (cont.)

- It is difficult to determine whether two simple graphs are isomorphic using brute force because there are  $n!$  possible one-to-one correspondences between the vertex sets of two simple graphs with  $n$  vertices.
- The best algorithms for determining whether two graphs are isomorphic have exponential worst case complexity in terms of the number of vertices of the graphs.
- Sometimes it is not hard to show that two graphs are not isomorphic. We can do so by finding a property, preserved by isomorphism, that only one of the two graphs has. Such a property is called *graph invariant*.
- There are many different useful graph invariants that can be used to distinguish nonisomorphic graphs, such as the number of vertices, number of edges, and degree sequence (list of the degrees of the vertices in nonincreasing order). We will encounter others in later sections of this chapter.

52

## Isomorphism of Graphs (cont.)

**Example:** Determine whether these two graphs are isomorphic.



**Solution:** Both graphs have eight vertices and ten edges. They also both have four vertices of degree two and four of degree three.

However,  $G$  and  $H$  are not isomorphic. Note that since  $\deg(a) = 2$  in  $G$ ,  $a$  must correspond to  $t, u, x$ , or  $y$  in  $H$ , because these are the vertices of degree 2. But each of these vertices is adjacent to another vertex of degree two in  $H$ , which is not true for  $a$  in  $G$ .

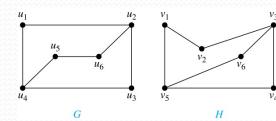
Alternatively, note that the subgraphs of  $G$  and  $H$  made up of vertices of degree three and the edges connecting them must be isomorphic. But the subgraphs, as shown at the right, are not isomorphic.



53

## Isomorphism of Graphs (cont.)

**Example:** Determine whether these two graphs are isomorphic.



**Solution:** Both graphs have six vertices and seven edges. They also both have four vertices of degree two and two of degree three. The subgraphs of  $G$  and  $H$  consisting of all the vertices of degree two and the edges connecting them are isomorphic. So, it is reasonable to try to find an isomorphism  $f$ .

We define an injection  $f$  from the vertices of  $G$  to the vertices of  $H$  that preserves the degree of vertices. We will determine whether it is an isomorphism.

The function  $f$  with  $f(u_1) = v_6, f(u_2) = v_3, f(u_3) = v_4, f(u_4) = v_5, f(u_5) = v_1$ , and  $f(u_6) = v_2$  is a one-to-one correspondence between  $G$  and  $H$ . Showing that this correspondence preserves edges is straightforward, so we will omit the details here. Because  $f$  is an isomorphism, it follows that  $G$  and  $H$  are isomorphic graphs.

*See the text for an illustration of how adjacency matrices can be used for this verification.*

54

## Algorithms for Graph Isomorphism

- The best algorithms known for determining whether two graphs are isomorphic have exponential worst-case time complexity (in the number of vertices of the graphs).
- However, there are algorithms with linear average-case time complexity.
- You can use a public domain program called NAUTY to determine in less than a second whether two graphs with as many as 100 vertices are isomorphic.
- Graph isomorphism is a problem of special interest because it is one of a few NP problems not known to be either tractable or NP-complete (see Section 3.3).

55

## Applications of Graph Isomorphism

- The question whether graphs are isomorphic plays an important role in applications of graph theory. For example,
  - chemists use molecular graphs to model chemical compounds. Vertices represent atoms and edges represent chemical bonds. When a new compound is synthesized, a database of molecular graphs is checked to determine whether the graph representing the new compound is isomorphic to the graph of a compound that this already known.
  - Electronic circuits are modeled as graphs in which the vertices represent components and the edges represent connections between them. Graph isomorphism is the basis for
    - the verification that a particular layout of a circuit corresponds to the design's original schematics.
    - determining whether a chip from one vendor includes the intellectual property of another vendor.

56

# Connectivity

57

## Section Summary

- Paths
- Connectedness in Undirected Graphs
- Vertex Connectivity and Edge Connectivity
- Connectedness in Directed Graphs
- Paths and Isomorphism
- Counting Paths between Vertices

58

## Paths

**Informal Definition:** A *path* is a sequence of edges that begins at a vertex of a graph and travels from vertex to vertex along edges of the graph. As the path travels along its edges, it visits the vertices along this path, that is, the endpoints of these.

**Applications:** Numerous problems can be modeled with paths formed by traveling along edges of graphs such as:

- determining whether a message can be sent between two computers.
- efficiently planning routes for mail delivery.

59

## Paths

**Definition:** Let  $n$  be a nonnegative integer and  $G$  an undirected graph. A *path of length  $n$*  from  $u$  to  $v$  in  $G$  is a sequence of  $n$  edges  $e_1, \dots, e_n$  of  $G$  for which there exists a sequence  $x_0 = u, x_1, \dots, x_{n-1}, x_n = v$  of vertices such that  $e_i$  has, for  $i = 1, \dots, n$ , the endpoints  $x_{i-1}$  and  $x_i$ .

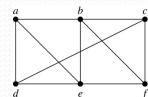
- When the graph is simple, we denote this path by its vertex sequence  $x_0, x_1, \dots, x_n$  (since listing the vertices uniquely determines the path).
- The path is a *circuit* if it begins and ends at the same vertex ( $u = v$ ) and has length greater than zero.
- The path or circuit is said to *pass through* the vertices  $x_1, x_2, \dots, x_{n-1}$  and *traverse* the edges  $e_1, \dots, e_n$ .
- A path or circuit is *simple* if it does not contain the *same edge* more than once.

This terminology is readily extended to directed graphs. (see text)

60

## Paths (*continued*)

**Example:** In the simple graph here:

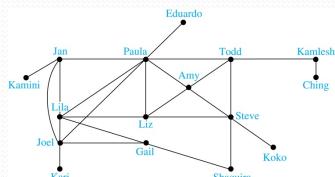


- $a, d, c, f, e$  is a simple path of length 4.
- $d, e, c, a$  is not a path because  $e$  is not connected to  $c$ .
- $b, c, f, e, b$  is a circuit of length 4.
- $a, b, e, d, a, b$  is a path of length 5, but it is not a simple path.

61

## Degrees of Separation

**Example: Paths in Acquaintance Graphs.** In an acquaintance graph there is a path between two people if there is a chain of people linking these people, where two people adjacent in the chain know one another. In this graph there is a chain of six people linking Kamini and Ching.



Some have speculated that almost every pair of people in the world are linked by a small chain of no more than six, or maybe even, five people. The play *Six Degrees of Separation* by John Guare is based on this notion.

62

## Erdős numbers



Paul Erdős

### Example: Erdős numbers.

In a collaboration graph, two people  $a$  and  $b$  are connected by a path when there is a sequence of people starting with  $a$  and ending with  $b$  such that the endpoints of each edge in the path are people who have collaborated.

| TABLE 1 The Number of Mathematicians with a Given Erdős Number (as of early 2006). |                  |
|--|------------------|
| Erdős Number   | Number of People |
| 0  | 1                |
| 1  | 504              |
| 2  | 6,593            |
| 3  | 33,605           |
| 4  | 83,642           |
| 5  | 87,760           |
| 6  | 40,014           |
| 7  | 11,591           |
| 8  | 3,146            |
| 9  | 819              |
| 10   | 244              |
| 11   | 68               |
| 12   | 23               |
| 13   | 5                |

- In the academic collaboration graph of people who have written papers in mathematics, the *Erdős number* of a person  $m$  is the length of the shortest path between  $m$  and the prolific mathematician Paul Erdős.
- To learn more about Erdős numbers, visit  
<http://www.ams.org/mathscinet/collaborationDistance.html>

63

## Bacon Numbers

- In the Hollywood graph, two actors  $a$  and  $b$  are linked when there is a chain of actors linking  $a$  and  $b$ , where every two actors adjacent in the chain have acted in the same movie.
- The *Bacon number* of an actor  $c$  is defined to be the length of the shortest path connecting  $c$  and the well-known actor Kevin Bacon. (Note that we can define a similar number by replacing Kevin Bacon by a different actor.)
- The *oracle of Bacon* web site  
<http://oracleofbacon.org/how.php> provides a tool for finding Bacon numbers.

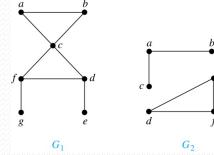
| TABLE 2 The Number of Actors with a Given Bacon Number (as of early 2011). |                  |
|--|------------------|
| Bacon Number   | Number of People |
| 0  | 1                |
| 1  | 2,367            |
| 2  | 242,407          |
| 3  | 785,389          |
| 4  | 200,602          |
| 5  | 14,048           |
| 6  | 1,277            |
| 7  | 114              |
| 8  | 16               |

64

## Connectedness in Undirected Graphs

**Definition:** An undirected graph is called *connected* if there is a path between every pair of vertices. An undirected graph that is not *connected* is called *disconnected*. We say that we *disconnect* a graph when we remove vertices or edges, or both, to produce a disconnected subgraph.

**Example:**  $G_1$  is connected because there is a path between any pair of its vertices, as can be easily seen. However  $G_2$  is not connected because there is no path between vertices  $a$  and  $f$ , for example.



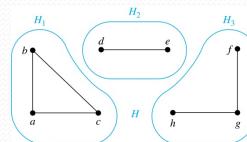
65

## Connected Components

**Definition:** A *connected component* of a graph  $G$  is a connected subgraph of  $G$  that is not a proper subgraph of another connected subgraph of  $G$  (it is the largest one).

A graph  $G$  that is not connected has two or more connected components that are disjoint and have  $G$  as their union.

**Example:** The graph  $H$  is the union of three disjoint subgraphs  $H_1$ ,  $H_2$ , and  $H_3$ , none of which are proper subgraphs of a larger connected subgraph of  $G$ . These three subgraphs are the connected components of  $H$ .



66

## Connectedness in Directed Graphs

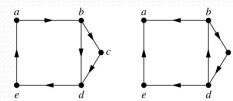
**Definition:** A directed graph is *strongly connected* if there is a path from  $a$  to  $b$  and a path from  $b$  to  $a$  whenever  $a$  and  $b$  are vertices in the graph.

**Definition:** A directed graph is *weakly connected* if there is a path between every two vertices in the underlying undirected graph, which is the undirected graph obtained by ignoring the directions of the edges of the directed graph.

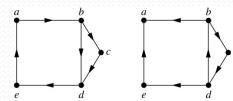
67

## Connectedness in Directed Graphs (continued)

**Example:**  $G$  is strongly connected because there is a path between any two vertices in the directed graph. Hence,  $G$  is also weakly connected.



The graph  $H$  is not strongly connected, since there is no directed path from  $a$  to  $b$ , but it is weakly connected.



**Definition:** The subgraphs of a directed graph  $G$  that are strongly connected but not contained in larger strongly connected subgraphs, that is, the maximal strongly connected subgraphs, are called the *strongly connected components* or *strong components* of  $G$ .

**Example (continued):** The graph  $H$  has three strongly connected components, consisting of the vertex  $a$ ; the vertex  $e$ ; and the subgraph consisting of the vertices  $b$ ,  $c$ ,  $d$  and edges  $(b,c)$ ,  $(c,d)$ , and  $(d,b)$ .

68

## The Connected Components of the Web Graph

- Recall that at any particular instant the web graph provides a snapshot of the web, where vertices represent web pages and edges represent links. According to a 1999 study, the Web graph at that time had over 200 million vertices and over 1.5 billion edges. (The numbers today are several orders of magnitude larger.)
- The underlying undirected graph of this Web graph has a connected component that includes approximately 90% of the vertices.
- There is a *giant strongly connected component (GSCC)* consisting of more than 53 million vertices. A Web page in this component can be reached by following links starting in any other page of the component. There are three other categories of pages with each having about 44 million vertices:
  - pages that can be reached from a page in the GSCC, but do not link back.
  - pages that link back to the GSCC, but can not be reached by following links from pages in the GSCC.
  - pages that cannot reach pages in the GSCC and can not be reached from pages in the GSCC.

69

## Counting Paths between Vertices

- We can use the adjacency matrix of a graph to find the number of paths between two vertices in the graph.

**Theorem:** Let  $G$  be a graph with adjacency matrix  $A$  with respect to the ordering  $v_1, \dots, v_n$  of vertices (with directed or undirected edges, multiple edges and loops allowed). The number of different paths of length  $r$  from  $v_i$  to  $v_j$ , where  $r > 0$  is a positive integer, equals the  $(i,j)$ th entry of  $A^r$ .

**Proof by mathematical induction:**

*Basis Step:* By definition of the adjacency matrix, the number of paths from  $v_i$  to  $v_j$  of length 1 is the  $(i,j)$ th entry of  $A$ .

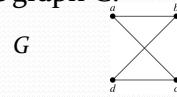
*Inductive Step:* For the inductive hypothesis, we assume that the  $(i,j)$ th entry of  $A^r$  is the number of different paths of length  $r$  from  $v_i$  to  $v_j$ .

- Because  $A^{r+1} = A^r A$ , the  $(i,j)$ th entry of  $A^{r+1}$  equals  $[A^{r+1}]_{ij} = b_{i1}a_{1j} + b_{i2}a_{2j} + \dots + b_{in}a_{nj}$ , where  $b_{ik}$  is the  $(i,k)$ th entry of  $A^r$  ( $B = A^r$ ). By the inductive hypothesis,  $b_{ik}$  is the number of paths of length  $r$  from  $v_i$  to  $v_k$ .
- A path of length  $r + 1$  from  $v_i$  to  $v_j$  is made up of a path of length  $r$  from  $v_i$  to some neighboring node  $v_k$ , and an edge from  $v_k$  to  $v_j$ . By the product rule for counting, the number of such paths is the product of the number of paths of length  $r$  from  $v_i$  to  $v_k$  (i.e.,  $b_{ik}$ ) and the number of edges from  $v_k$  to  $v_j$  (i.e.,  $a_{kj}$ ).
- The sum over all possible intermediate vertices  $v_k$  is  $b_{i1}a_{1j} + b_{i2}a_{2j} + \dots + b_{in}a_{nj}$ .
- But this is exactly  $[A^{r+1}]_{ij}$ .

70

## Counting Paths between Vertices (continued)

**Example:** How many paths of length four are there from  $a$  to  $d$  in the graph  $G$ .



$$A = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} \quad \text{adjacency matrix of } G$$

**Solution:** The adjacency matrix of  $G$  (ordering the vertices as  $a, b, c, d$ ) is given above. Hence the number of paths of length four from  $a$  to  $d$  is the  $(1, 4)$ th entry of  $A^4$ . The eight paths are as:

- |                 |                 |
|-----------------|-----------------|
| $a, b, a, b, d$ | $a, b, a, c, d$ |
| $a, b, d, b, d$ | $a, b, d, c, d$ |
| $a, c, a, b, d$ | $a, c, a, c, d$ |
| $a, c, d, b, d$ | $a, c, d, c, d$ |

71

## Euler and Hamiltonian Graphs

Section 10.5

72

## Section Summary

- Euler Paths and Circuits
- Hamilton Paths and Circuits
- Applications of Hamilton Circuits

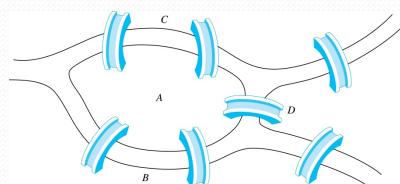
73



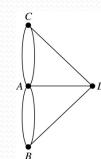
Leonard Euler  
(1707-1783)

## Euler Paths and Circuits

- The town of Königsberg, Prussia (now Kalingrad, Russia) was divided into four sections by the branches of the Pregel river. In the 18th century seven bridges connected these regions.
- People wondered whether it was possible to follow a path that crosses each bridge exactly once and returns to the starting point.
- The Swiss mathematician Leonard Euler proved that no such path exists. This result is often considered to be the first theorem ever proved in graph theory.



The 7 Bridges of Königsberg



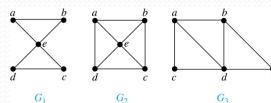
Multigraph  
Model of the  
Bridges of  
Königsberg

74

## Euler Paths and Circuits (continued)

**Definition:** An *Euler circuit* in a graph  $G$  is a simple circuit containing every edge of  $G$ . An *Euler path* in  $G$  is a simple path containing every edge of  $G$ .

**Example:** Which of the undirected graphs  $G_1$ ,  $G_2$ , and  $G_3$  has a Euler circuit? Of those that do not, which has an Euler path?



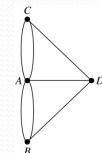
**Solution:** The graph  $G_1$  has an Euler circuit (e.g.,  $a, e, c, d, e, b, a$ ). But, as can easily be verified by inspection, neither  $G_2$  nor  $G_3$  has an Euler circuit. Note that  $G_3$  has an Euler path (e.g.,  $a, c, d, e, b, d, a, b$ ), but there is no Euler path in  $G_2$ , which can be verified by inspection.

75

## Necessary and Sufficient Conditions for Euler Circuits and Paths

**Theorem:** A connected multigraph with at least two vertices has an Euler circuit if and only if each of its vertices has an even degree and it has an Euler path if and only if it has exactly two vertices of odd degree.

**Example:** Two of the vertices in the multigraph model of the Königsberg bridge problem have odd degree. Hence, there is no Euler circuit in this multigraph and it is impossible to start at a given point, cross each bridge exactly once, and return to the starting point.



76

## Necessary and Sufficient Conditions for Euler Circuits and Paths

**There exists an Euler circuit  $\rightarrow$  all degrees are even**

- An Euler circuit begins with a vertex  $a$  and continues with an edge incident with  $a$ , say  $\{a, b\}$ . The edge  $\{a, b\}$  contributes one to  $\deg(a)$ .
- Each time the circuit passes through a vertex it contributes two to the vertex's degree.
- Finally, the circuit terminates where it started, contributing one to  $\deg(a)$ . Therefore  $\deg(a)$  must be even.
- We conclude that the degree of every other vertex must also be even.
- By the same reasoning, we see that the initial vertex and the final vertex of an Euler path have odd degree, while every other vertex has even degree. So, a graph with an Euler path has exactly two vertices of odd degree.
- In the next slide we will show that these necessary conditions are also sufficient conditions.

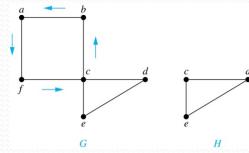
77

## Necessary and Sufficient Conditions for Euler Circuits and Paths

**All degrees are even  $\rightarrow$  there exists an Euler circuit**

Suppose that  $G$  is a connected multigraph with  $\geq 2$  vertices, all of even degree. Let  $x_0 = a$  be a vertex of even degree. Choose an edge  $\{x_0, x_1\}$  incident with  $a$  and proceed to build a simple path  $\{x_0, x_1\}, \{x_1, x_2\}, \dots, \{x_{n-1}, x_n\}$  by adding edges one by one until another edge can not be added.

We illustrate this idea in the graph  $G$  here.  
We begin at  $a$  and choose the edges  $\{a, f\}, \{f, c\}, \{c, b\}$ , and  $\{b, a\}$  in succession.

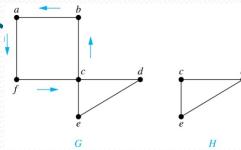


- The path begins at  $a$  with an edge of the form  $\{a, x\}$ ; we show that it must terminate at  $a$  with an edge of the form  $\{y, a\}$ . Since each vertex has an even degree, there must be an even number of edges incident with this vertex. Hence, every time we enter a vertex other than  $a$ , we can leave it. Therefore, the path can only end at  $a$ .
- If all of the edges have been used, an Euler circuit has been constructed. Otherwise, consider the subgraph  $H$  obtained from  $G$  by deleting the edges already used.

In the example  $H$  consists of the vertices  $c, d, e$ .

78

## Necessary and Sufficient Conditions for Euler Circuits and Paths (c)



- Because  $G$  is connected,  $H$  must have at least one vertex in common with the circuit that has been deleted.

In the example, the vertex is  $c$ .

- Every vertex in  $H$  must have even degree because all the vertices in  $G$  have even degree and for each vertex, pairs of edges incident with this vertex have been deleted. Beginning with the shared vertex construct a path ending in the same vertex (as was done before). Then splice this new circuit into the original circuit.

In the example, we end up with the circuit  $a, f, c, d, e, c, b, a$ .

- Continue this process until all edges have been used. This produces an Euler circuit since every edge is included and no edge is included more than once.
- Similar reasoning can be used to show that a graph with exactly two vertices of odd degree must have an Euler path connecting these two vertices of odd degree

79

## Algorithm for Constructing an Euler Circuits

In our proof we developed this algorithms for constructing a Euler circuit in a graph with no vertices of odd degree.

```

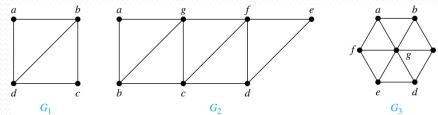
procedure Euler( $G$ : connected multigraph with all vertices of even degree)
   $circuit :=$  a circuit in  $G$  beginning at an arbitrarily chosen vertex with edges
    successively added to form a path that returns to this vertex.
   $H := G$  with the edges of this circuit removed
  while  $H$  has edges
     $subcircuit :=$  a circuit in  $H$  beginning at a vertex in  $H$  that also is
      an endpoint of an edge in  $circuit$ .
     $H := H$  with edges of  $subcircuit$  and all isolated vertices removed
     $circuit := circuit$  with  $subcircuit$  inserted at the appropriate vertex.
  return  $circuit$ { $circuit$  is an Euler circuit}

```

80

## Euler Circuits and Paths

**Example:**



$G_1$  contains exactly two vertices of odd degree ( $b$  and  $d$ ). Hence it has an Euler path, e.g.,  $d, a, b, c, d, b$ .

$G_2$  has exactly two vertices of odd degree ( $b$  and  $d$ ). Hence it has an Euler path, e.g.,  $b, a, g, f, e, d, c, g, b, c, f, d$ .

$G_3$  has six vertices of odd degree. Hence, it does not have an Euler path.

81

## Applications of Euler Paths and Circuits

- Euler paths and circuits can be used to solve many practical problems such as finding a path or circuit that traverses each
  - street in a neighborhood,
  - road in a transportation network,
  - connection in a utility grid,
  - link in a communications network.
- Other applications are found in the
  - layout of circuits,
  - network multicasting,
  - molecular biology, where Euler paths are used in the sequencing of DNA.

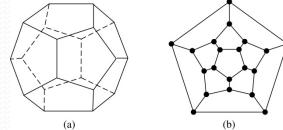
82



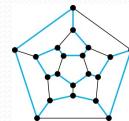
William Rowan  
Hamilton  
(1805- 1865)

## Hamilton Paths and Circuits

- Euler paths and circuits contained every edge only once. Now we look at paths and circuits that contain every vertex exactly once.
- William Hamilton invented the *Icosian puzzle* in 1857. It consisted of a wooden dodecahedron (with 12 regular pentagons as faces), illustrated in (a), with a peg at each vertex, labeled with the names of different cities. String was used to plot a circuit visiting 20 cities exactly once
- The graph form of the puzzle is given in (b).



- The solution (a Hamilton circuit) is given here.



83

## Hamilton Paths and Circuits

**Definition:** A simple path in a graph  $G$  that passes through every vertex exactly once is called a *Hamilton path*, and a simple circuit in a graph  $G$  that passes through every intermediate vertex exactly once is called a *Hamilton circuit*.

That is, a simple path  $x_0, x_1, \dots, x_{n-1}, x_n$  in the graph  $G = (V, E)$  is called a Hamilton path if  $V = \{x_0, x_1, \dots, x_{n-1}, x_n\}$  and  $x_i \neq x_j$  for  $0 \leq i < j \leq n$ , and the simple circuit  $x_0, x_1, \dots, x_{n-1}, x_n, x_0$  (with  $n > 0$ ) is a Hamilton circuit if  $x_0, x_1, \dots, x_{n-1}, x_n$  is a Hamilton path.

84

## Hamilton Paths and Circuits (continued)

**Example:** Which of these simple graphs has a Hamilton circuit or, if not, a Hamilton path?

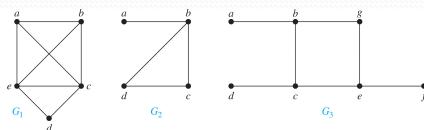


FIGURE 10 Three Simple Graphs.

**Solution:**  $G_1$  has a Hamilton circuit:  $a, b, c, d, e, a$ .

$G_2$  does not have a Hamilton circuit (Why?), but does at least have a Hamilton path :  $a, b, c, d$ .

$G_3$  does not have a Hamilton circuit, or a Hamilton path. Why?

85

## Necessary Conditions for Hamilton Circuits



Gabriel Andrew Dirac  
(1925-1984)

- Unlike for an Euler circuit, no simple necessary and sufficient conditions are known for the existence of a Hamilton circuit.
- However, there are some useful necessary conditions. We describe two of these now.

**Dirac's Theorem:** If  $G$  is a simple graph with  $n \geq 3$  vertices such that the degree of every vertex in  $G$  is  $\geq n/2$ , then  $G$  has a Hamilton circuit.

**Ore's Theorem:** If  $G$  is a simple graph with  $n \geq 3$  vertices such that  $\deg(u) + \deg(v) \geq n$  for every pair of nonadjacent vertices, then  $G$  has a Hamilton circuit.



Øystein Ore  
(1899-1968)

86

## Applications of Hamilton Paths and Circuits

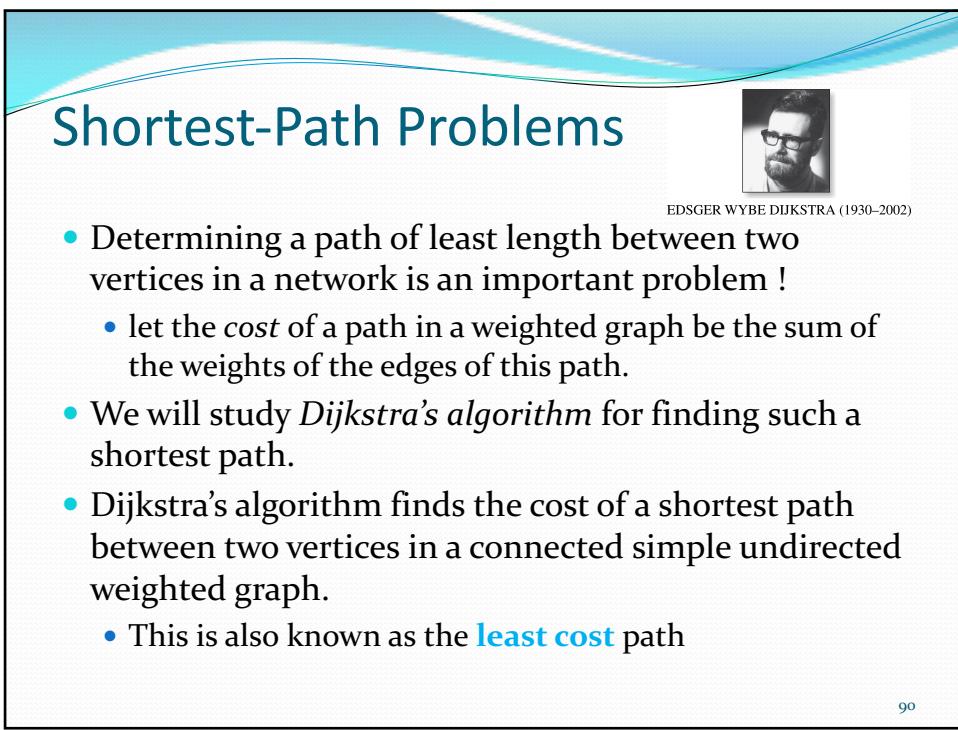
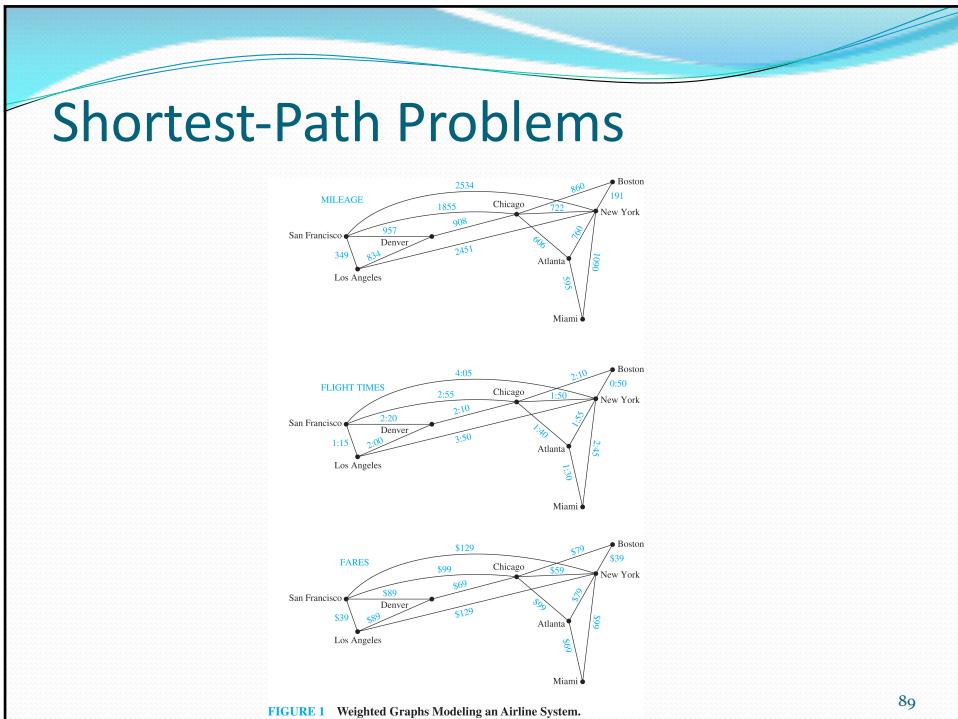
- Applications that ask for a path or a circuit that visits each intersection of a city, each place pipelines intersect in a utility grid, or each node in a communications network exactly once, can be solved by finding a Hamilton path in the appropriate graph.
- The famous *traveling salesperson problem (TSP)* asks for the shortest route a traveling salesperson should take to visit a set of cities. This problem reduces to finding a Hamilton circuit such that the total sum of the weights of its edges is as small as possible.
- A family of binary codes, known as *Gray codes*, which minimize the effect of transmission errors, correspond to Hamilton circuits in the  $n$ -cube  $Q_n$ . (*See the text for details.*)

87

## Shortest-Path Problems

- Many problems can be modeled using graphs with weights assigned to their edges.
- Graphs that have a number assigned to each edge are called **weighted graphs**.
  - The weight between  $u$  and  $v$  is denoted as  $w(u, v)$
  - Weights have to be **positive**,  $w(u, v) > 0$
- As an illustration, consider how an airline system can be modeled.

88



## Dijkstra's algorithm

- Assume we want to find the shortest path from  $a$  to  $z$ .
- It begins by labeling the starting vertex  $a$  with 0 and the other vertices with  $\infty$ . We use the notation  $L_0(a) = 0$  and  $L_0(v) = \infty$  for these initial labels.
  - Before any iterations have taken place (the subscript 0 stands for the “0th” iteration).
- Dijkstra's algorithm proceeds by forming a distinguished set of vertices. Let  $S_k$  denote this set after  $k$  iterations of the labeling procedure.
- We begin with  $S_0 = \emptyset$ . The set  $S_k$  is formed from  $S_{k-1}$  by adding a vertex  $u$  not in  $S_{k-1}$  with the *smallest label*.

91

## Dijkstra's algorithm

- Let  $v$  be a vertex not in  $S_k$ . To update the label of  $v$ , note that  $L_k(v)$  is the length of a shortest path from  $a$  to  $v$  containing only vertices in  $S_k$  as intermediate nodes (to be proved later).
- A shortest path from  $a$  to  $v$  containing only elements of  $S_k$  when  $u$  is added to  $S_{k-1}$  is either
  - a shortest path from  $a$  to  $v$  that contains only elements of  $S_{k-1}$  (that is, the distinguished vertices not including  $u$ ),
  - or it is a shortest path from  $a$  to  $u$  at the  $(k - 1)$ st stage with the edge  $\{u, v\}$  added.
- In this case,

$$L_k(v) = \min\{L_{k-1}(v), L_{k-1}(u) + w(u, v)\}$$

92

## Dijkstra's algorithm

- This procedure is iterated by successively adding vertices to the distinguished set  $S$  until  $z$  is added.
- When  $z$  is added to the distinguished set, we will show that its label is the length of an overall shortest path from  $a$  to  $z$ .

93

## Dijkstra's algorithm

### ALGORITHM 1 Dijkstra's Algorithm.

```

procedure Dijkstra( $G$ : weighted connected simple graph, with
all weights positive)
{ $G$  has vertices  $a = v_0, v_1, \dots, v_n = z$  and lengths  $w(v_i, v_j)$ 
where  $w(v_i, v_j) = \infty$  if  $\{v_i, v_j\}$  is not an edge in  $G$ }
for  $i := 1$  to  $n$ 
   $L(v_i) := \infty$ 
   $L(a) := 0$ 
   $S := \emptyset$ 
{the labels are now initialized so that the label of  $a$  is 0 and all
other labels are  $\infty$ , and  $S$  is the empty set}
while  $z \notin S$ 
   $u :=$  a vertex not in  $S$  with  $L(u)$  minimal
   $S := S \cup \{u\}$ 
  for all vertices  $v$  not in  $S$ 
    if  $L(u) + w(u, v) < L(v)$  then  $L(v) := L(u) + w(u, v)$ 
    {this adds a vertex to  $S$  with minimal label and updates the
    labels of vertices not in  $S$ }
return  $L(z)$  ( $L(z) =$  length of a shortest path from  $a$  to  $z$ )

```

94

## Dijkstra's algorithm

- **Theorem:** Any sub-path of a shortest path has to be itself a shortest path.
- This can be proved by contradiction.
  - Assume the full path is a shortest path.
  - Now, if a sub-path is not a shortest path, the length of the full path could be decreased by replacing this sub-path by a shorter sub-path.
  - Which contradicts the fact that the full path is a shortest path.
- Thus, any sub-path of a shortest path has to be itself a shortest path.

95

## Dijkstra's algorithm

- It follows that
  - On any shortest path from  $a$  to  $v$  with intermediate node  $i$ , the shortest path distance from  $a$  to  $v$  is equal to the shortest path distance from  $a$  to  $i$  plus the shortest path distance from  $i$  to  $v$ .

96

## Dijkstra's algorithm

- Moreover, a node cannot appear more than once on a shortest path.
- This can also be proved by contradiction.

97

## Dijkstra's algorithm

- Dijkstra's algorithm: proof by induction
- Theorem: at the end of any iteration  $k$ ,
  - H1:** for any  $u$  in  $S_k$ ,  $L_k(u)$  contains the length of the shortest path from  $a$  to  $u$ ,  $L^*(u)$ .
  - H2:** for any  $u$  not in  $S_k$ ,  $L_k(u)$  holds the length of the shortest path from  $a$  to  $u$  containing only intermediate nodes in  $S_k$  (thus, except the ending node  $u$  itself).
  - Sources: Rosen's book + Kevin Wayne's presentation on Dijkstra's algorithm and Bellman-Ford algorithm (Princeton University, [www.princeton.edu/~cos226](http://www.princeton.edu/~cos226)).
- **Initialization:** At  $k = 0$ ,  $S_0 = \emptyset$  with  $L_0(a) = 0$  and  $L_0(v) = \infty$  for the other labels  $v$ .
  - It therefore holds the length of a shortest path from  $a$  containing only intermediate vertices in  $S_0 = \emptyset$ .
  - The property 2 is therefore true for  $k = 0$ , the basis case.
  - Moreover, initially,  $a$  is chosen to be included in  $S$  and 0 is indeed the length of the overall shortest path from  $a$  to  $a$ . The property 1 is therefore true at  $k = 0$ .

98

## Dijkstra's algorithm

### Part 1 of the theorem (induction)

- Assume the hypothesis 1 and 2 are true at the end of step  $k$ .
- Let  $v$  be a vertex not in  $S_k$ . Note that  $L_k(v)$  is the length of a shortest path from  $a$  to  $v$  containing only intermediate nodes in  $S_k$ .
  - Assume  $v$  is the node not in  $S_k$  with lowest  $L_k$ .
  - This node will be added to  $S_k$  to form  $S_{k+1}$ .
  - We have to prove that  $L_k(v)$  is also the length of the overall shortest path from  $a$  to  $v$ .

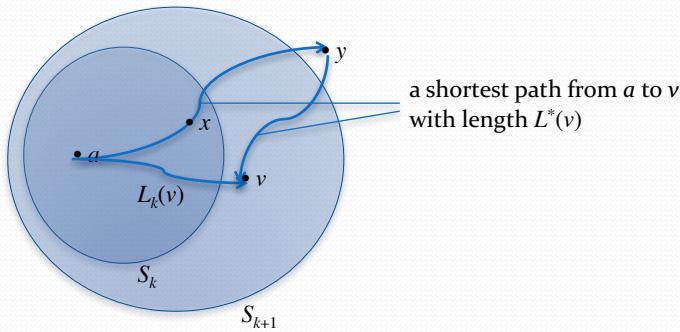
99

## Dijkstra's algorithm

- By **contradiction**: assume there is a *shorter path* from  $a$  to  $v$  with length  $L^*(v) < L_k(v)$ .
- By hypothesis 2, as  $L_k(v)$  holds the length of a shortest path from  $a$  to  $v$  containing only intermediate nodes in  $S_k$ , this shorter path must contain some nodes not in  $S_k$ .
  - Consider the node  $y$  which is *the first node not in  $S_k$*  on this shortest path from  $a$  to  $v$  (the first node leaving  $S_k$ ).
  - Moreover, node  $x$  is the node in  $S_k$  *immediately preceding*  $y$  on this shortest path.

100

## Dijkstra's algorithm



- We have  $L_k(v) \stackrel{(1)}{>} L^*(v) \stackrel{(2)}{\geq} L^*(y) = \stackrel{(3)}{L_k(y)} \stackrel{(4)}{\geq} L_k(v)$

101

## Dijkstra's algorithm

1. (1) follows from the fact that  $L^*(v)$  is the length of the shortest path to  $v$ .
2. (2) follows from the fact that any sub-path of a shortest path is itself a shortest sub-path whose length is lower or equal than the shortest-path length.
3. (3) follows from the induction hypothesis 2: all nodes preceding  $y$  are in  $S_k$
4. (4) follows from the fact that  $v$  is the node not in  $S_k$  with lowest  $L_k$ .

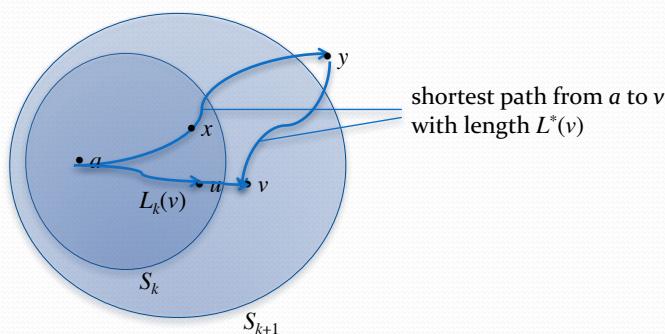
102

## Dijkstra's algorithm

- Here is an alternative proof of **part 1 of the theorem**.
- Assume hypotheses 1 and 2 are true at the end of step  $k$ .
- Let  $v$  be a vertex not in  $S_k$ . Note that, by induction hypothesis,  $L_k(v)$  is the length of a shortest path from  $a$  to  $v$  containing only nodes in  $S_k$ .
  - Assume  $v$  is the node not in  $S_k$  with the lowest  $L_k$ .
  - This node will be added to  $S_k$  to form  $S_{k+1}$ .
  - We thus have to show that  $L_k(v)$  must also be the overall shortest path from  $a$  to  $v$ :
- By contradiction, assume there is an overall **shortest path** from  $a$  to  $v$  of length  $L^*(v)$ . This shorter path must pass through at least one node not in  $S_k$ .
- Assume it is going through  $y$  (which is not in  $S_k$ ), which is the first node not in  $S_k$  on this shorter path.
- Since  $y$  is on the shortest path from  $a$  to  $v$  (a sub-path of a shortest path), we have  $L_k(y) \leq L^*(v) < L_k(v)$ , which breaks our assumption that  $v$  is the node not in  $S_k$  with the lowest  $L_k$ .
- Therefore, there is no such node  $y$ .

103

## Dijkstra's algorithm



104

## Dijkstra's algorithm

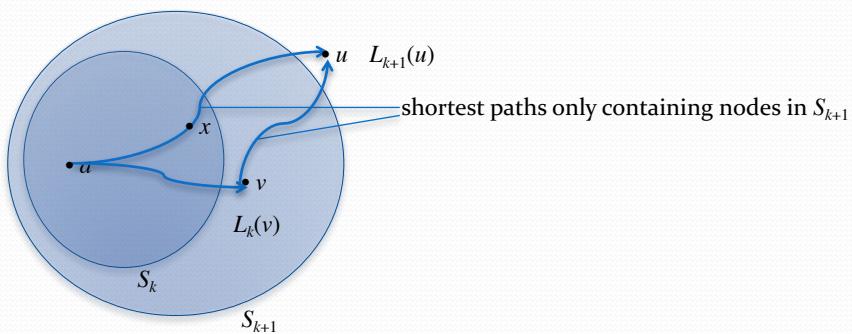
### Part 2 of the theorem (induction)

- Assume the hypothesis 1 and 2 are true at step  $k$ . Moreover, we just showed that hypothesis 1 is also true at step  $k+1$ .
- Let  $u$  be a vertex not in  $S_{k+1}$ . To update the label of  $u$ , note that, by induction hypothesis 2,  $L_k(u)$  is the length of a shortest path from  $a$  to  $u$  containing only vertices in  $S_k$ .
  - We now have to compute  $L_{k+1}(u)$  for all  $u$  not in  $S_{k+1}$ .
- A shortest path  $p$  from  $a$  to  $u$  containing only elements of  $S_{k+1}$  must either directly come from
  - a node of  $S_k$  (that is, the distinguished vertices not including  $v$ ), say node  $x$ . Note that, in this case, an overall shortest path from  $a$  to  $x$  contains only nodes in  $S_k$  and thus the path  $p$  from  $a$  to  $u$  is a shortest path only containing intermediate nodes in  $S_k$ . The new label of  $u$  is then unchanged,  $L_k(u)$ .
  - or the new node  $v$ , added at step  $k+1$ . In this case, the label should be  $L_k(v) + w(v, u)$ . since the sub-path of a shortest path must itself be a shortest path,
- The new label must therefore be the best among these two alternatives,

$$L_{k+1}(u) = \min\{L_k(u), L_k(v) + w(v, u)\}$$

105

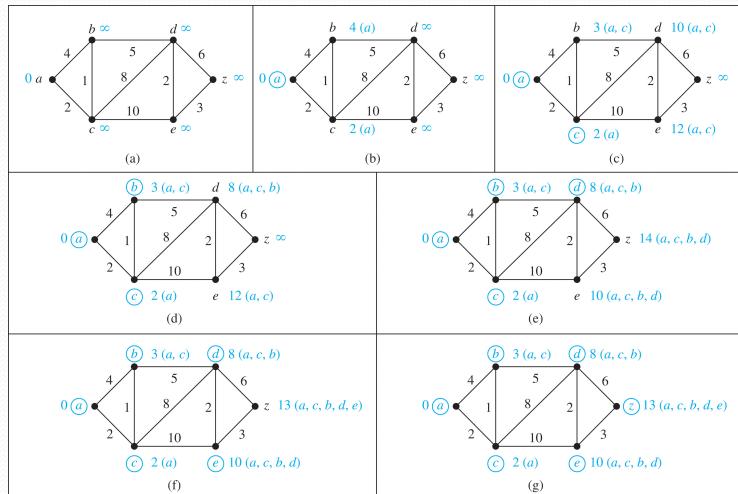
## Dijkstra's algorithm



The last node on the shortest path to  $u$  is either a node in  $S_k$  or the node  $v$  included in  $S_{k+1}$  at iteration  $k+1$ .

106

## Dijkstra's algorithm

FIGURE 4 Using Dijkstra's Algorithm to Find a Shortest Path from  $a$  to  $z$ .

107

## Dijkstra's algorithm

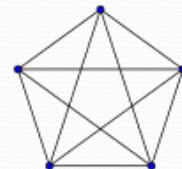
Solving the problem with a table:

|   | 0   | 1   | 2   | 3   | 4  | 5  | 6  |
|---|-----|-----|-----|-----|----|----|----|
| a | 0   | 0   | 0   | 0   | 0  | 0  | 0  |
| b | Inf | 4   | 3   | 3   | 3  | 3  | 3  |
| c | Inf | 2   | 2   | 2   | 2  | 2  | 2  |
| d | Inf | Inf | 10  | 8   | 8  | 8  | 8  |
| e | Inf | Inf | 12  | 12  | 10 | 10 | 10 |
| z | Inf | Inf | Inf | Inf | 14 | 13 | 13 |

108

## Planar graphs

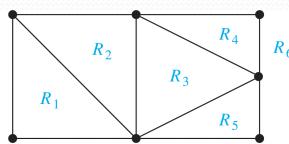
- A graph is called *planar* if it can be drawn in the plane without any edges crossing.
  - where a crossing of edges is the intersection of the lines or arcs representing them at a point other than their common endpoint.
- Such a drawing is called a *planar representation* of the graph.
- Wikipedia example of non-planar graph:



109

## Planar graphs

- Euler's formula
  - Let  $G$  be a connected planar simple graph with  $e$  edges and  $v$  vertices.
  - Let  $r$  be the number of regions in a planar representation of  $G$ .
- Then  $r = e - v + 2$ . In the example,  $6 = 11 - 7 + 2$ 
  - The proof is in the book.



**FIGURE 8** The Regions of the Planar Representation of a Graph. 110

## Graph coloring

- A *coloring* of a simple graph is the assignment of a color to each vertex of the graph
  - so that no two adjacent vertices are assigned the same color.
- The *chromatic number* of a graph is the least number of colors needed for a coloring of this graph.
  - The chromatic number of a graph  $G$  is denoted by  $\chi(G)$  (here  $\chi$  is the Greek letter *chi*).

111

## Graph coloring

- A simple example of translating a map into a graph.
  - Two regions are linked if they have a border.

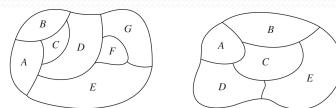


FIGURE 1 Two Maps.

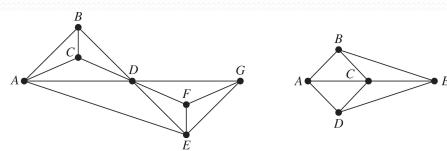


FIGURE 2 Dual Graphs of the Maps in Figure 1.

112

## Graph coloring

- **The four color problem.**
  - The chromatic number of a planar simple graph is no greater than four.
  - The proof is quite difficult.
  - It involved a computer enumerating many different cases.
- Many different heuristics have been developed for graph coloring of general graphs.

113

## Graph coloring

- Here is a simple example.

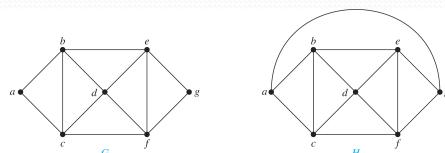


FIGURE 3 The Simple Graphs  $G$  and  $H$ .

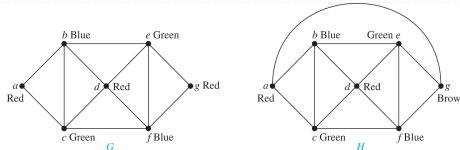


FIGURE 4 Colorings of the Graphs  $G$  and  $H$ .

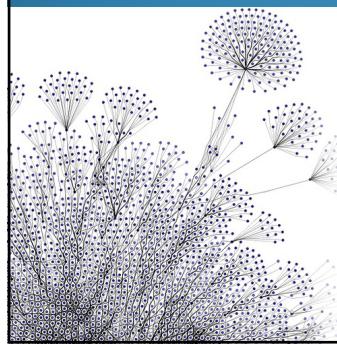
114

## Graph coloring

- An interesting application of graph coloring.
- How can the final exams at a university be scheduled so that no student has two exams at the same time?
- *Solution:* This scheduling problem can be solved using a graph model, with vertices representing courses and with an edge between two vertices if there is a common student in the courses they represent.
  - Each time slot for a final exam is represented by a different color.
  - A scheduling of the exams corresponds to a coloring of the associated graph.

115

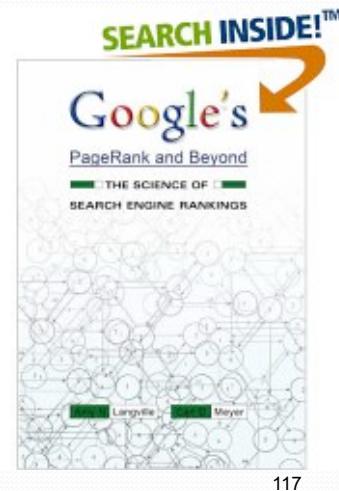
## The PageRank algorithm



116

## A nice book about PageRank

- A. Langville & C. Meyer (2006)
  - Google's PageRank and beyond
  - Princeton university press



117

## The basic PageRank algorithm

- Introduced by Page, Brin, Motwani & Winograd in 1998
- Partly implemented in Google
- Corresponds to a measure of « prestige » in a directed graph

118

## Web link analysis

- A set of techniques
  - Applied to: Hyperlink document repositories
  - Typically web pages
- Objective:
  - To exploit the link structure of the documents
  - In order to extract interesting information
  - Viewing the document repository as a graph where
    - Nodes are documents
    - Edges are directed links between documents
  - It does not exploit the content of the pages !!

119

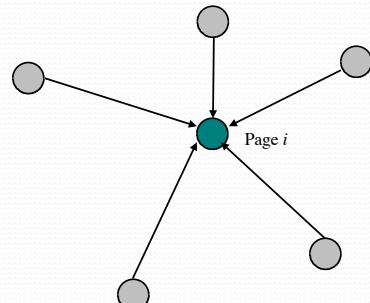
## Web link analysis

- Suppose we performed a search with a search engine
- **Objective:** to improve the (content-based) ranking of the search engine
  - Based on the graph structure of the web hyperlinks
  - PageRank is computed off-line

120

## The basic PageRank algorithm

- To each web page we associate a score,  $x_i$ 
  - The score of page  $i$ ,  $x_i$ , is proportional to the weighted averaged score of the pages pointing to page  $i$



121

## The basic PageRank algorithm

- Let  $w_{ij}$  be the non-negative weight (representing some **strength of affinity** between  $i$  and  $j$ ) of the link connecting page  $i$  to page  $j$ 
  - For the web, it is simply 0 or 1
  - Then,  $w_{ij} = 1$  if page  $i$  has a link to page  $j$ ;  $w_{ij} = 0$  otherwise
- Let  $\mathbf{W}$  be the matrix made of the elements  $w_{ij}$ 
  - Notice that this matrix is not symmetric (directed graph)
  - We suppose that the graph is strongly connected

122

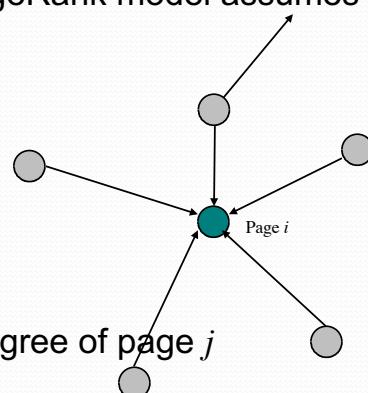
## The basic PageRank algorithm

- Indeed, the basic PageRank model assumes

$$x_i \propto \sum_{j=1}^n \frac{w_{ji} x_j}{w_{j.}}$$

$$w_{j.} = \sum_{i=1}^n w_{ji}$$

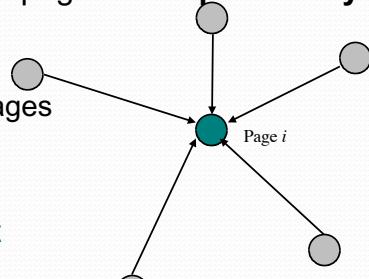
- where  $w_{j.}$  is the outdegree of page  $j$



123

## The basic PageRank algorithm

- In other words,
- A page with a **high score** is a page that **is pointed by**
  - Many pages
  - Having each a high score
  - and pointing to few other pages
- Thus a page is an **important** page if
  - It is pointed by **many, important**, pages
  - and these pages are citing few other pages



124

## The basic PageRank algorithm

- These equations can be updated **iteratively** until convergence
- In order to obtain the scores  $x_i$ 
  - Usually, we normalize the vector  $\mathbf{x}$  at each iteration
- The pages are then **ranked** according to their score
  - and presented to the user

125

## Random walk interpretation of PageRank

- This definition has a nice interpretation in terms of **random walk** or **surfing** on the graph
- If we define the probability of following the link from page  $j$  to page  $i$  as

$$P(\text{page}(k+1) = i | \text{page}(k) = j) = \frac{w_{ji}}{w_j}.$$

$$w_{j.} = \sum_{i=1}^n w_{ji}$$

126

## Random walk interpretation of PageRank

- We can write the updating equation as

$$\begin{aligned} x_i(k+1) &= P(\text{page}(k+1) = i) \\ &= \sum_{j=1}^n P(\text{page}(k+1) = i | \text{page}(k) = j) x_j(k) \end{aligned}$$

- where  $x_i(k)$  is the probability of finding the random walker on page  $i$  at time step  $k$
- Thus we defined a **random surfer** following the links according to the transition probabilities

$$P(\text{page}(k+1) = i | \text{page}(k) = j) = \frac{w_{ji}}{w_j.}$$

127

## Random walk interpretation of PageRank

- This is the equation of a **Markov model** of random **surf** through the web
- It is very similar to the basic PageRank equation seen before, and recalled here:

$$x_i \propto \sum_{j=1}^n \frac{w_{ji} x_j}{w_j.}$$

$$w_{j.} = \sum_{i=1}^n w_{ji}$$

128

## Random walk interpretation of PageRank

- If we further denote element  $i, j$  of the transition probabilities matrix  $\mathbf{P}$  as  $p_{ij}$
- We then have

$$\begin{aligned} [\mathbf{P}]_{ij} &= p_{ij} = \text{P}(page(k+1) = j | page(k) = i) \\ &= \frac{w_{ij}}{w_i}. \end{aligned}$$

- And the **evolution equation** along time of  $x_i(k)$  can be rewritten as

$$x_i(k+1) = \sum_{j=1}^n p_{ji} x_j(k)$$

129

## Example of transition matrix

- Example (from Grinstead and Snell (2012)):
  - We take as states/nodes the kind of weather R (rain), N (nice), and S (snow)

$$\mathbf{P} = \begin{matrix} & \text{R} & \text{N} & \text{S} \\ \text{R} & \left( \begin{matrix} 1/2 & 1/4 & 1/4 \end{matrix} \right) \\ \text{N} & \left( \begin{matrix} 1/2 & 0 & 1/2 \end{matrix} \right) \\ \text{S} & \left( \begin{matrix} 1/4 & 1/4 & 1/2 \end{matrix} \right) \end{matrix}$$

130

## Random walk interpretation of PageRank

- In matrix form, if the vector  $\mathbf{x}$  has elements  $x_i$  with

$$\mathbf{x}(k+1) = \mathbf{P}^T \mathbf{x}(k)$$

- and therefore, by induction,

$$\mathbf{x}(k) = (\mathbf{P}^T)^k \mathbf{x}(0)$$

- The **stationary** or **equilibrium distribution** is given by setting  $\mathbf{x}(k+1) = \mathbf{x}(k) = \mathbf{x}$  (no change from time step  $k$  to time  $k+1$ ), and thus

$$\mathbf{x} = \mathbf{P}^T \mathbf{x}$$

131

## For the weather example

$$\mathbf{P}^1 = \begin{matrix} & \text{Rain} & \text{Nice} & \text{Snow} \\ \text{Rain} & .500 & .250 & .250 \\ \text{Nice} & .500 & .000 & .500 \\ \text{Snow} & .250 & .250 & .500 \end{matrix}$$

$$\mathbf{P}^2 = \begin{matrix} & \text{Rain} & \text{Nice} & \text{Snow} \\ \text{Rain} & .438 & .188 & .375 \\ \text{Nice} & .375 & .250 & .375 \\ \text{Snow} & .375 & .188 & .438 \end{matrix}$$

$$\mathbf{P}^3 = \begin{matrix} & \text{Rain} & \text{Nice} & \text{Snow} \\ \text{Rain} & .406 & .203 & .391 \\ \text{Nice} & .406 & .188 & .406 \\ \text{Snow} & .391 & .203 & .406 \end{matrix}$$

$$\mathbf{P}^4 = \begin{matrix} & \text{Rain} & \text{Nice} & \text{Snow} \\ \text{Rain} & .402 & .199 & .398 \\ \text{Nice} & .398 & .203 & .398 \\ \text{Snow} & .398 & .199 & .402 \end{matrix}$$

$$\mathbf{P}^5 = \begin{matrix} & \text{Rain} & \text{Nice} & \text{Snow} \\ \text{Rain} & .400 & .200 & .399 \\ \text{Nice} & .400 & .199 & .400 \\ \text{Snow} & .399 & .200 & .400 \end{matrix}$$

$$\mathbf{P}^6 = \begin{matrix} & \text{Rain} & \text{Nice} & \text{Snow} \\ \text{Rain} & .400 & .200 & .400 \\ \text{Nice} & .400 & .200 & .400 \\ \text{Snow} & .400 & .200 & .400 \end{matrix} = \text{equilibrium or stationary distribution}$$

132

## The basic PageRank algorithm

- The PageRank scores can therefore be obtained
  - By computing **the left eigenvector** of the matrix  $\mathbf{P}$  corresponding to **eigenvalue 1**
  - Which is also the right eigenvector of  $\mathbf{P}^T$ 
    - Where  $\mathbf{P}$  is **the transition probabilities matrix** of the Markov process (random walk/surf on the web)
    - Containing the transition probabilities
  - If the graph is undirected, the scores are simply the **indegrees** of the nodes !

133

## Random walk interpretation of PageRank

- $x_i$  can then be viewed as the probability of being at page  $i$ 
  - The solution to this equation is the **stationary distribution** of the random surf
  - Which is the probability of finding the surfer on page  $i$  on the **long-term behaviour**
- The « most probable page » is the best ranked

134

## The basic PageRank algorithm

- For computing this stationary distribution (right eigenvector of  $\mathbf{P}^T$ ), we can use the well-known **power method**
- The computation of the dominant right eigenvector  $\mathbf{x}$  of a matrix  $\mathbf{M}$  iterates

$$\begin{cases} \mathbf{x} \leftarrow \mathbf{Mx} \\ \mathbf{x} \leftarrow \frac{\mathbf{x}}{\|\mathbf{x}\|} \end{cases}$$

- And, in our case,  $\mathbf{M} = \mathbf{P}^T$
- The norm in the normalization step is the L1 norm ( $\mathbf{x}$  must sum to 1,  $\mathbf{e}^T \mathbf{x} = 1$  with  $\mathbf{e}$  being a column vector of 1s)
- Note that a **system of linear equations** ( $\mathbf{P}^T \mathbf{x} = \mathbf{x}$  with  $\mathbf{e}^T \mathbf{x} = 1$ ) can also be used to compute the basic PageRank scores (see a few slides later for the justification)

135

## Adjustments to the basic model

- However, there is a problem with
  - **Dangling nodes**
  - That is, nodes without any outgoing link
- In this case, the  $\mathbf{P}$  matrix is no more **stochastic**
  - That is, rows do not sum to one
- Moreover, the graph could have several **separate components**

136

## Adjustments to the basic model

- One potential solution is to allow to jump to any node of the graph
  - With some non-zero probability ([teleportation](#))
- Therefore, we define the new transition probabilities matrix

$$\mathbf{G} = \alpha \mathbf{P} + (1 - \alpha) \frac{\mathbf{e}\mathbf{e}^T}{n}$$

where  $\mathbf{G}$  is called the [Google matrix](#),  $\mathbf{e}$  is a column vector full of 1's and  $0 < \alpha < 1$

137

## Adjustments to the basic model

- Moreover, the dangling nodes are made absorbing
  - That is, if  $i$  is a dangling node,  $p_{ii} = 1$  and  $p_{ij} = 0$  for  $j \neq i$
- So that the row sum is equal to 1
- And since there is a transportation probability  $(1 - \alpha)$ , random walkers will be able to escape from these dangling nodes

138

## Adjustments to the basic model

- In this case,
  - The matrix is **stochastic** (each row sums to 1)
  - The matrix is **irreducible** (no separate component)
  - The matrix is **aperiodic** (not explained here)
- Then, the matrix is said to be **regular** and
  - it turns out that there is a unique eigenvector associated to eigenvalue 1
- However, **G** is no more **sparse** !

139

## Computing PageRank

- The problem is thus to find the **left eigenvector** of **G**
  - Corresponding to the eigenvalue 1
  - Instead of **P**

$$\mathbf{x}^T \mathbf{G} = \mathbf{x}^T$$

- with the normalization  $\|\mathbf{x}\|_1 = 1$  or  $\mathbf{x}^T \mathbf{e} = 1$
- One can use the standard **power method**

140

## Personalization in PageRank

- But how can we **favour** some pages in a natural way (advertising, etc) ?
- Rather than using  $\mathbf{e}\mathbf{e}^T/n$ ,
- use  $\mathbf{e}\mathbf{v}^T$  where
  - $\mathbf{v} > 0$  is a **probability vector** ( $\mathbf{v}^T \mathbf{e} = 1$ )
  - which is called the **personalization vector**
- It contains the **a priori probability** of jumping to any page

141

## Personalization in PageRank

- The Google matrix then becomes

$$\mathbf{G} = \alpha \mathbf{P} + (1 - \alpha) \mathbf{e}\mathbf{v}^T$$

- where  $\mathbf{v}$  is provided a priori, depending on the problem
- $\mathbf{v}$  is the probability of being teletransported to any node

142

## The PageRank problem as a sparse linear system

- Here is an alternative formulation of the PageRank problem

- As a usually sparse system of linear equations

$$\mathbf{x}^T \mathbf{G} = \mathbf{x}^T$$

$$\Rightarrow \mathbf{x}^T (\alpha \mathbf{P} + (1 - \alpha) \mathbf{e} \mathbf{v}^T) = \mathbf{x}^T$$

$$\Rightarrow \alpha \mathbf{x}^T \mathbf{P} + (1 - \alpha) \mathbf{v}^T = \mathbf{x}^T$$

$$\Rightarrow \mathbf{x}^T (\mathbf{I} - \alpha \mathbf{P}) = (1 - \alpha) \mathbf{v}^T$$

$$\Rightarrow (\mathbf{I} - \alpha \mathbf{P})^T \mathbf{x} = (1 - \alpha) \mathbf{v}$$

- In addition, if  $\alpha = 1$ , the equation  $\mathbf{e}^T \mathbf{x} = 1$  must be added to this matrix-vector equation (indeed,  $(\mathbf{I} - \mathbf{P})$  is rank-deficient)

- In order for  $\mathbf{x}$  to sum to 1 ( $\mathbf{x}$  represents a probability mass)

143

## The PageRank problem as a sparse linear system

- In other words, the problem is

Solve  $(\mathbf{I} - \alpha \mathbf{P})^T \mathbf{x} = (1 - \alpha) \mathbf{v}$  with  $\mathbf{x}^T \mathbf{e} = 1$

- Which has been shown (Del Corso, Gulli & Romani, 2005; Langville & Meyer, 2006) to be equivalent to

Solve  $(\mathbf{I} - \alpha \mathbf{P})^T \mathbf{x}' = \mathbf{v}$  and compute  $\mathbf{x} = \mathbf{x}' / \|\mathbf{x}'\|_1$

144

## The ItemRank model

- We now introduce an extension of PageRank that can be used for collaborative recommendation
- It is based on the « random walk with restart » idea (Faloutsos)
  - And was developed by Pucci and Gori in the framework of collaborative recommendation

145

## The ItemRank model

- It is called **ItemRank**
- Assume a graph has been defined
  - Nodes are defined for users and items
  - There is a link between an item and a user if the user bought the item
  - Weights quantify the number of times the item has been bought

146

## The ItemRank model

- Assume a **random walker** on the graph

$$\mathbf{x}_i(k+1) = \alpha \mathbf{P}^T \mathbf{x}_i(k) + (1 - \alpha) \mathbf{v}_i$$

- Where  $\mathbf{v}_i$  contains  $1/n_i$  for each item that user  $i$  bought and 0 everywhere else
- And where  $n_i$  is the number of bought items
- Where  $\mathbf{P}$  is the transition probabilities matrix derived from the graph

147

## The ItemRank model

- In other words, the random walker has a probability  $(1 - \alpha)\mathbf{v}_i$  of being teleported on a bought item node
- The steady-state solution will be the similarity score to the bought items, associated to each item
  - for user  $i$

$$\mathbf{x}_i = (1 - \alpha)(\mathbf{I} - \alpha \mathbf{P}^T)^{-1} \mathbf{v}_i$$

148

## The ItemRank model

- The solution can also be found by iterating the equation

$$\mathbf{x}_i(k+1) = \alpha \mathbf{P}^T \mathbf{x}_i(k) + (1 - \alpha) \mathbf{v}_i$$

- until convergence

149