Day 11	
Graphs	

# Graph Theory

Graph theory is the study of graphs, mathematical structures used to model pair wise relations between objects from a certain collection

Graph paper is not very useful

"Graphs" in this context are not to be confused with "graphs of functions" and other kinds of graphs or representations of data

A graph is a finite set of dots called vertices (or nodes) connected by links called edges (or arcs)





Applications			
Graph	Vertices	Edges	
Transportation	street intersections, airports	highways, air routes	
Scheduling	tasks	precedence constraints	
Software systems	functions	function calls	
Internet	web pages	hyperlinks	
Social Networks	people, actors, terrorists	friendships, movie casts, associations	



Graph	Vertices	Edges
Communication	telephones, computers	Fiber optic cable
Circuits	gates, registers, processors	wires
Mechanical	joints	rods, beams, springs
Hydraulic	reservoirs, pumping stations	pipelines
Electrical Power Grid	transmission stations	cable
Financial	stocks, currency	transactions



Graph	Vertices	Edges
Games	board pieces	legal moves
Protein interaction	proteins	protein interaction
Genetics regulatory networks	genes	regulatory interactions
Neural Networks	neurons	synapses
Chemical Compounds	molecules	chemical bonds
Infectious Disease	people	infections



## Seven Bridges of Königsberg

Königsberg once a capital of the German providence of East Prussia it is a sea port and Russian exclave between Poland and Lithuania on the Baltic Sea





The question is whether it is possible to walk with a route that crosses each bridge exactly once, and return to the starting point.

Leonhard Euler proved that it was not possible

### Leonhard Euler

The paper written by Leonhard Euler on the Seven Bridges of Königsberg and published in 1736 is regarded as the first paper in the history of graph theory



1707-1783





### Remarks

The Konigsberg bridge problem appeared in *Solutio problematis ad geometriam situs pertinentis,* Commetarii Academiae Scientiarum Imperialis Petropolitanae (1736) which may be the earliest paper on graph theory

For more on this see

http://mathdl.maa.org/images/upload\_library/22/P olya/hopkins.pdf

### Other Contributors

10

Thomas Pennington Kirkman (Manchester, England 1806-1895) British clergyman who studied combinatorics

William Rowan Hamilton (Dublin, Ireland 1805-1865) applied "quaternions" worked on optics, dynamics and analysis created the "icosian game" in 1857, a precursor of Hamiltonian cycles

Denes Konig (Budapest, Hungary 1844-1944) Interested in four-color problem and graph theory 1936: published *Theory of finite and infinite graphs,* the first textbook on graph theory



# Six Degrees of Separation

Six degrees of separation refers to the idea that human beings are connected through relationships with at most six other people

Several studies, such as Milgram's small world experiment have been conducted to empirically measure this connectedness

While the exact number of links between people differs depending on the population measured, it is generally found to be relatively small

Hence, six degrees of separation is somewhat synonymous with the idea of the "small world" phenomenon

### **Bacon Number**



14

15

13

See http://oracleofbacon.org/

The Bacon number is the number of degrees of separation (see Six degrees of separation) they have from actor Kevin Bacon, as defined by the game known as Six Degrees of Kevin Bacon

The higher the Bacon number, the farther away from Kevin Bacon the actor is

### Example

Adolf Hitler Adolf Hitler was in *Der Ewige Jude* (1940) with Curt Bois

and

Curt Bois was in *The Great Sinner* (1949) with Kenneth Tobey

and

Kenneth Tobey was in *Hero at Large* (1980) with Kevin Bacon



Bacon Number = 3







### Erdős Number

The Erdős number, honoring the late Hungarian mathematician Paul Erdős, one of the most prolific writers of mathematical papers, is a way of describing the "collaborative distance", in regard to mathematical papers, between an author and Erdős



1913-1996

17

To be assigned an Erdős number, an author must cowrite a mathematical paper with an author with a finite Erdős number

For more info see http://www.oakland.edu/enp/

# Erdős Number

An eBay auction offered an Erdős number of 2 for a prospective paper to be submitted for publication to *Chance* (a magazine of the American Statistical Association) about skill in the World Series of Poker and the World Poker Tour

It closed on 22 July 2004 with a winning bid of \$127.40



### Weighted Graph

A weighted graph associates a label (weight) with every edge in the graph

a weighted graph G(E,V) with a real-valued weight function  $f \colon E \to R$ 

### **Review - Directed Graph**

19

A directed graph or digraph *G* is an ordered pair G := (V, A) with

V, a set of vertices or nodes,

*A*, a set of ordered pairs of vertices, called directed edges, arcs, or arrows

An edge or arc e = (x, y) is considered to be directed from x to y, y is called the head and x is called the tail of the arc; y is said to be a direct successor of x, and x is said to be a direct predecessor of y

If a path leads from x to y, then y is said to be a successor of x, and x is said to be a predecessor of y







# Example

David travel throughout Europe every summer

He particularly enjoys visiting Hamburg, Salzburg, Rome, Vienna, and Madrid

He prefers Madrid over Salzburg over Hamburg; and Vienna and Rome over Madrid and Salzburg

Represent this using a directed graph









### **Traveling Salesperson Problem**

The first route (1-4-2-5-6-3-1), with a total length of 62 km, is a relevant selection but is not the best solution

The second route (1-2-5-4-6-3-1) represents a much better solution as the total distance, 48 km, is less than for the first route

This example assumes Euclidean distances and an isotropic space, but in reality the solution may be different considering the configuration of transport infrastructures

26

### Example

Whole pineapples are served in a restaurant in London To ensure freshness, the pineapples are purchased in Hawaii and air freighted from Honolulu to Heathrow in London

The following network diagram outlines the different routes that the pineapples could take

In this example, our weights are the freight associated freight costs

This is a "shortest-path" problem

The cost to freight a pineapple is known for each are:						
Honolulu Honolulu Chicago Chicago San Francisco San Francisco San Francisco Los Angeles Los Angeles Boston New York Atlanta	Chicago San Francisco Los Angeles Boston New York Boston New York Atlanta New York Atlanta Heathrow Heathrow Heathrow 7	105 75 68 45 56 71 48 63 44 57 88 65 6				
		28				















Problem: grouping objects into similarity classes based on various properties of the objects

Example:

Computer programs that implement the same algorithm have properties k = 1, 2 or 3 such as:

1. Number of lines in the program

- 2. Number of "return" statements
- 3. Number of function calls



32

Suppose five programs are compared and a table is made:

Program	# of lines	# of "return"	# of function calls
1	66	20	1
2	41	10	2
3	68	5	8
4	90	34	5
5	75	12	14







### **Dissimilarity Functions**

Define a dissimilarity function as follows: For each pair of vertices  $v = (p_1, p_2, p_3)$  and  $w = (q_1, q_2, q_3)$  let

$$s(v,w) = \sum_{1}^{3} |p_k - q_k|$$

$$k = 1$$

 $\hfill\square$  s(v,w) is a measure of dissimilarity between any two programs v and w

 $\square$  Fix a number N. Insert an edge between v and w if s(v,w) < N. Then:

□ We say that v and w are in the same class if v = w or if there is a path between v and w.



### Definition

A graph denoted G or G(V, E) consists of two parts

- (1) A set V = V(G) of vertices (points, nodes)
- (2) A collection of E = E(G) of unordered pairs of distinct vertices called edges

### **Drawing Graphs**

Graphs are represented graphically by drawing a dot for every vertex, and drawing an arc between two vertices if they are connected by an edge

If the graph is directed, the direction is indicated by drawing an arrow

# Remarks

A graph drawing should not be confused with the graph itself (the abstract, non-graphical structure) as there are several ways to structure the graph drawing

All that matters is which vertices are connected to which others by how many edges and not the exact layout

In practice it is often difficult to decide if two drawings represent the same graph

Depending on the problem domain some layouts may be better suited and easier to understand than others

39

37





Size refers to the number of edges

Order is the number of vertices

A trivial graph is a graph with a single vertex

### Terminology

When two vertices of a graph are connected by an edge, these vertices are said to be adjacent, and the edge is said to join them

A vertex and an edge that touch one another are said to be incident to one another

Suppose  $e=\{u,v\}$  is an edge, then the vertex u is said to be adjacent to v, and the edge e is said to be incident on u and on v

42









































# Multigraph

A Multigraph G = G(V, E) consists a set of vertices and a set of edges *E* may contain multiple edges,

i.e. edges connected to the same endpoint, and E may contain one or more loops

A loop is an edge whose endpoints are the same vertex





# Equivalent Graphs

56

Equivalent graphs have the same number of vertices and edges

They contain the G(V,E)

But they may be drawn differently









# Isomorphic Graphs

A Graph Isomorphism is a bijection (one-to-one and onto) mapping between the vertices of two graphs *G* and *H*  $f:V(G) \rightarrow V(H)$ 

with the property that any two vertices of uand v from G are adjacent if and only if f(u)and f(v) are adjacent in H.

59

These are equivalent graphs



















































## Isomorphic Graphs

If an isomorphism can be constructed between two graphs, then we say those graphs are isomorphic

Determining whether two graphs are isomorphic is referred to as the graph isomorphism problem

Graphs G and H are isomorphic if and only if for some ordering they have the same adjacency matrix

Two graphs are isomorphic if they are they same graphs, drawn differently. Two graphs are isomorphic if you can label both graphs with the same labels so that every vertex has exactly the same neighbors in both graphs

### Terminology

Order of a graph: number of nodes in the graph

Degree: the number of edges at a node, without regard to whether the graph is directed or undirected

Connected graph: a graph in which all pairs of nodes are connected by a path. Informally, the graph is all in one piece

A graph that can be drawn in a plane or on a sphere so that its edges do not cross is said to be planar



Directed graph all the edges are directed e.g. route network

Undirected graph all the edges are undirected e.g. flight network

Mixed graph some edges are directed some edges are undirected

### Parallel Processing

76

Parallel processing is the simultaneous execution of the same task (split up and specially adapted) on multiple processors in order to obtain results faster

The idea is based on the fact that the process of solving a problem usually can be divided into smaller tasks, which may be carried out simultaneously with some coordination



OAK RIDGE, Tenn., Aug. 25, 2006 — An upgrade to the Cray XT3 supercomputer at Oak Ridge National Laboratory has increased the system's computing power to 54 teraflops, or 54 trillion mathematical calculations per second, making the Cray among the most powerful open scientific systems in the world

The Jaguar has more than 10,400 processing cores and 21 terabytes of memory. Probably fifth fastest computer today  $$_{78}$$ 

# Characteristics of Parallel Processors

Institution	Name	N	Topology	BW/Link (MB/s)	BW/Sys (MB/s)	year	
U. Illinois	Illiac IV	64	2D grid	40	2560	1972	
ICL	DAP1	4096	2D grid	0.6	2560	1980	
Goodyear	MPP	16384	2D grid	1.2	20,480	1982	t
Thinking Machines	CM-2	4096	12-cube	1	65,536	1987	t
nCube	nCube/ten	1024	10-cube	1.2	10,240	1987	Ī
Intel	iPSC/2	128	7-cube	2	896	1988	t
Maspar	MP1216	512	2D grid	3	23,000	1989	t
Intel	Delta	540	2D grid	40	21,600	1991	
Thinking Machines	CM-5	1024	Fat Tree	20	20,480	1991	t



# Hypercubes

Number of processors =  $2^n$ n dimensions (n = log N) A cube with n dimensions is made out of 2 cubes of dimension n-1

Also called an n-cube











A path is a sequence of edges that begins at an initial vertex and ends at a terminal vertex

A path that begins and terminates in the same vertex is called a cycle or circuit

A graph that contains no cycles is called an acyclic graph

A graph G = G(V,E) is finite if both V and E are finite

The empty graph has no vertices and no edges

A vertex is isolated if it does not belong to any edge

The degree of a vertex  $\boldsymbol{v}$  is equal to the number of edges which are incident on  $\boldsymbol{v}$ 

The vertex is said to be even or odd according to the degree









A path or walk is an alternating sequence of vertices and edges, beginning and ending with a vertex

A path is closed if its first and last vertices are the same, and open if they are different

A trail is a path in which all the edges are distinct

### Distance

The distance between u and v is written d(u,v) and is the shortest distance between u and v

### d(u,v)=0 **←→** u=v

 $d(\boldsymbol{u},\boldsymbol{v})$  is not defined if no path between  $\boldsymbol{u}$  and  $\boldsymbol{v}$  exists.

A path  $\alpha$  in G with origin  $v_o$  and end  $v_n$  is an alternating sequence of vertices and edges in the form  $v_0,e_1,v_1,e_2,v_2,...,e_{n-1},v_{n-1},e_n,v_n$  where each edge  $e_i$  is incident on vertices  $v_{i-1}$  and  $v_i$ . The number of edges is called the length of  $\alpha$ .

### Definitions

A path  $\alpha$  is closed if  $v_0 = v_n$ 

The path  $\boldsymbol{\alpha}$  is simple if all the vertices are distinct

The path  $\boldsymbol{\alpha}$  is a trail if all the edges are distinct

The path  $\alpha$  is a cycle if it is closed and all vertices are distinct except v\_0=v\_f

A cycle of length k is called a k-cycle

A cycle in a graph must have a length of three or more. The diameter of graph written diam(G) is the maximum distance between any two of its vertices

90

88























### Subgraph

Let *G* be a graph. Then *H* is a subgraph of *G* if  $V(H) \subseteq V(G)$  and  $E(H) \subseteq E(G)$  i.e. the vertices of *H* are also vertices of *G* and the edges of *H* are also edges of *G* 

# Full Subgraph

97

98

Suppose H = H(V', E') is a subgraph of G = G(V, E)

Then H is a full subgraph of G if E' contains all the edges of E whose endpoints lie in V'

We say H is the subgraph generated by V'













# Cut Points

The subgraph G – v of G where v is a vertex in G

G-v is obtained by deleting the vertex v from the vertex set V(G) and deleting all edges in E(G) which are incident on v

 $\mathsf{G}-\mathsf{v}$  is the full subgraph of  $\mathsf{G}$  generated by the remaining vertices









# Bridges

G - e is obtained by deleting the edge e from the edge set E(G)

V(G - e) = V(G) and  $E(G - e) = E(G) / \{e\}$ 

An edge e is a bridge for G if G – e is disconnected









### Traversable Graph

A graph G is said to be traversable if it can be drawn without any breaks in the curve and without repeating any edge

Vertices may be repeated; edges may not be repeated

The path must include all vertices and all edges each exactly once

It has a path in which all may be traced exactly once without lifting the tracing instrument (without retracing)  $$_{\rm 109}$$ 





### Remarks

To be traversable, whenever our path enters a vertex v we must be a path leaving v, so we could expect our vertices to be even.

The exceptions of course are the first and last vertices which could be odd since we begin and stop at these vertices.

Thus to traversable G must have no more than two odd vertices, and we should start at one odd vertex and end at the other.

### Euler Path

If a graph is an Euler Path, that mean it has also can be traversed and has only two odd vertices

For a Euler path we start and stop on different odd nodes  $% \left( {{{\mathbf{r}}_{i}}} \right)$ 

We will now revisit Euler and the Seven Bridges of Königsberg problem

Present Day Satellite View of Königsberg Source: Google Earth





# Euler Cycle

An Euler (or Eulerian) cycle is path through a graph which starts and ends at the same vertex and includes every edge exactly once

Euler observed that a necessary condition for the existence of Euler cycles is that all vertices in the graph have an even degree, and that for an Euler path either all, or all but two, vertices have an even degree

Hence if G is a connected graph and every vertex of G has an even degree, then G has an Euler cycle

## William Rowan Hamilton

An Irish mathematician, physicist, and astronomer who made important contributions to the development of optics, dynamics, and algebra



His discovery of quaternions is perhaps his best known investigation. Hamilton's work in dynamics was later significant in the development of quantum mechanics

1805-1865

### Hamilton Cycle

A Hamilton cycle is a path through a graph that starts and ends at the same vertex and includes every other vertex exactly once

It is a closed path that includes every vertex exactly once

This differs from the Euler cycle which uses every edge exactly once but may repeat vertices

The Hamilton cycle uses each vertex exactly once (except for the first and last) and may skip edges









# Gray Code

Gray code after Frank Gray, is a binary numeral system where two successive values differ in only one digit

Gray codes are particularly useful in mechanical encoders since a slight change in position only affects one bit

Using a typical binary code, up to n bits could change, and slight misalignments between reading elements could cause wildly incorrect readings

An n-bit Gray code corresponds to a Hamiltonian cycle on an n-dimensional hypercube

### Gray Code

Example (N=3)

The binary coding of {0...7} is {000, 001, 010, 011, 100, 101, 110, 111},

while one Gray coding is {000, 001, 011, 010, 110, 111, 101, 100}

A Gray code takes a binary sequence and shuffles it to form some new sequence with the adjacency property

121

Grav Code Example	Decimal	Binary	Gray
	0	0000	0000
	1	0001	0001
For $n = 3$	2	0010	0011
The binary coding of {07} is	3	0011	0010
{000, 001, 010, 011, 100, 101, 110, 111}	4	0100	0110
	5	0101	0111
while one Gray coding is	6	0110	0101
{000, 001, 011, 010, 110, 111, 101, 100}	7	0111	0100
A Gray code takes a binary sequence and	8	1000	1100
shuffles it to form some new sequence	9	1001	1101
with the adjacency property	10	1010	1111
	11	1011	1110
The table on right is code $n = 4$	12	1100	1010
	13	1101	1011
	14	1110	1001
	15	1111	1000

### Converting Gray Code to Binary

A. write down the number in gray code

B. the most significant bit of the binary number is the most significant bit of the gray code

C. add (using modulo 2) the next significant bit of the binary number to the next significant bit of the gray coded number to obtain the next binary bit

D. repeat step C till all bits of the gray coded number have been added modulo 2  $\ensuremath{\mathsf{2}}$ 

The resultant number is the binary equivalent of the gray number

# Binary $\rightarrow$ Gray

Let B[n:0] be the input array of bits in the usual binary representation, [0] being LSB

Let  $\mathsf{G}[n{:}0]$  be the output array of bits in Gray code

G[n] = B[n]

for i = n-1 downto 0

G[i] = B[i+1] XOR B[i]

# Gray $\rightarrow$ Binary

Let G[n:0] be the input array of bits in Gray code Let B[n:0] be the output array of bits in the usual binary representation B[n] = G[n]for i = n-1 downto 0 B[i] = B[i+1] XOR G[i]

125

124

### Complete

A graph G is complete if every vertex is connected to every other vertex

The complete graph with n vertices is denoted  $\boldsymbol{K}_n$ 









### Remarks

Let *m* be the number of edges in the complete graph  $K_n$ . Each pair of vertices determine an edge

Taking combinations of vertices two at a time we have

 $m = C \binom{n}{2} = \frac{n(n-1)}{2}$  ways if selecting two vertices from *n* vertices

Since each vertex is connected the diameter is one  $\text{diam}(K_{\text{n}}){=}1$ 





### Remarks

Every vertex is connected to every n-1 vertices; hence deg(v) = n - 1 for every v in  $K_n$ 

n odd  $\rightarrow$  deg(v) = n - 1 even; thus K<sub>n</sub> is traversable for n odd. Also K<sub>2</sub> is traversable since it has only one edge connecting the two vertices

n even  $\rightarrow$  deg(v) = n - 1 odd; so for n > 2 the complete graph will have n (more than 2) odd vertices, hence is not traversable

131

### Regular

A graph is regular of degree k or  $k\mbox{-regular}$  if every vertex has degree k

A graph is regular if every vertex has the same degree





### **Bipartite Graph**

In a bipartite graph, the vertices can be divided into two sets so that every edge has one vertex in each of the two sets

A graph is said to be bipartite if its vertices V can be partitioned into two subsets M and N such that each edge of G connect a vertex of M to a vertex of N

A bipartite graph is a special graph where the set of vertices can be divided into two disjoint sets M and N such that no edge has both end-points in the same set

134

### Complete Bipartite Graph

In a complete bipartite graph each vertex of M is connected to a vertex on N denoted  $K_{m,n}$  where m is the number of vertices in M and n is the number of vertices in N and for standardization  $m\leqslant n$ 

Since each of the m vertices in M is connected to each of the n vertices in N  $K_{\text{m,n}}$  has mn edges

























# 3 Utilities Puzzle







Is it possible to lay on water, gas, and electricity to three different houses without crossing any line over the other?













We will use Euler's formula to figure out how many faces there are:

 $\begin{array}{l} \mathsf{F} - \mathsf{E} + \mathsf{V} = 2 \\ \mathsf{F} = 2 + \mathsf{E} - \mathsf{V} \\ = 2 + 9 - 6 \\ = 5 \text{ faces} \end{array} \qquad \qquad \begin{array}{l} \text{So the number of edges in all the} \\ \text{faces is at least } 4 * 5 = 20 \text{ edges.} \end{array}$ 

This counts each edge twice, because every edge is a boundary for two faces. So, the smallest number of edges is 20 / 2 = 10 edges.

However, we know that there are only 9 edges! Since nothing can have nine edges and ten edges at the same time, drawing a solution to the three utilities problem must be impossible.

Practice











# Map Coloring

The goal of a map coloring problem is to color a map so that regions sharing a common border have different colors

Regions that meet only in a point may share a common color

### Remarks

A map coloring problem can be solved by first converting the map into a graph where each region is a vertex and an edge connects two vertices if and only if the corresponding regions share a border

Once a map is converted into a graph vertex coloring can be used to decide how the map should be colored

152

















# Remarks

Graph Information http://en.wikipedia.org/wiki/Graph\_theory

Graph Tutorial (includes glossary) http://www.cs.usask.ca/content/resources/tutori als/csconcepts/1999\_8/

Lots of information of graph theory http://math.fau.edu/Locke/GRAPHTHE.HTM