

Complex Networks

Lecture 1: Basic Concepts

EE 6605

Instructor: G Ron Chen



Some Background and Motivation

- Between two randomly selected persons in the world, how many friends are there connecting them together?
- When searching from one webpage to another through the World Wide Web (WWW), how many clicks are needed in average?
- How can computer viruses propagate so fast and so wide through the Internet?
- How can people be infected so easily by diseases such as AIDS, SARS, and bird flu all over the world?
- How do rumors spread out in human societies?
- Can traffic jams in metropolitan cities be regulated effectively?
- How does electric power blackout emerge from small local system failures through the huge power grid?

We are living in a networked world today

- The influence of various complex and dynamical networks is currently pervading all kinds of sciences, ranging from physical to biological, even to social sciences. Its impact on modern engineering and technology is prominent and will be far-reaching.
- On the one hand, networks bring us with convenience and benefits, improve our efficiency of work and quality of life, and create tremendous opportunities that we never had before.
- On the other hand, however, networks also generate harms and damages to humans and societies, typically with epidemic spreading, computer virus propagation, and power blackout, to name just a few.

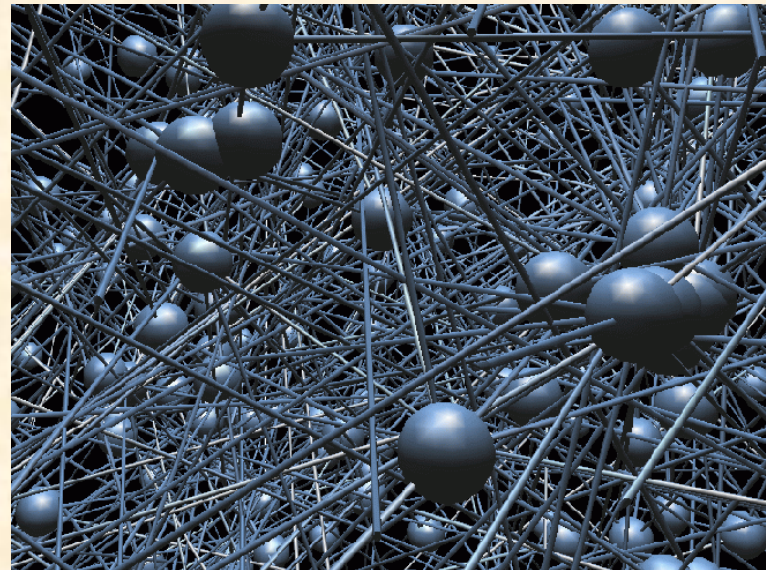
Development of network science studies

- For a long time in the history, studies of communication networks, power networks, biological networks, economic networks, social networks, etc., were carried out separately and independently.
- Recently, there were some rethinking of the general theory of complex dynamical networks towards a better understanding of the intrinsic relations, common properties and shared features of all kinds of networks in the real world.
- The new intention of studying the fundamental properties and dynamical behaviors of complex networks, both qualitatively and quantitatively, is important and timely, although very challenging technically.
- The current research along this line has been considered as “network science and engineering”, and has indeed become overwhelming.

Network Complexity (I)

Structural complexity:

- A network appears to be complicated in structure, which are seemingly messy and disordered.
- The network topology (i.e., structure) may vary in time.
- The connections (i.e., edges) among systems (i.e., nodes) may be weighted and/or directed.

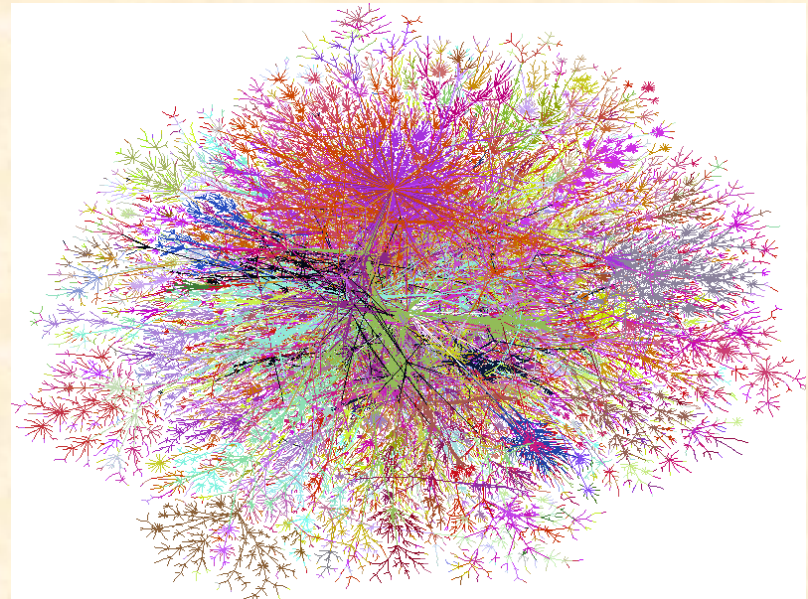


Graph of a social relationship structure in Canberra, Australia [A. S. Klov Dahl, ANU]

Network Complexity (II)

Node-dynamical complexity:

- A node in the network can be a dynamical system, which may have bifurcating and even chaotic behaviors.
- A network may have different kinds of nodes.

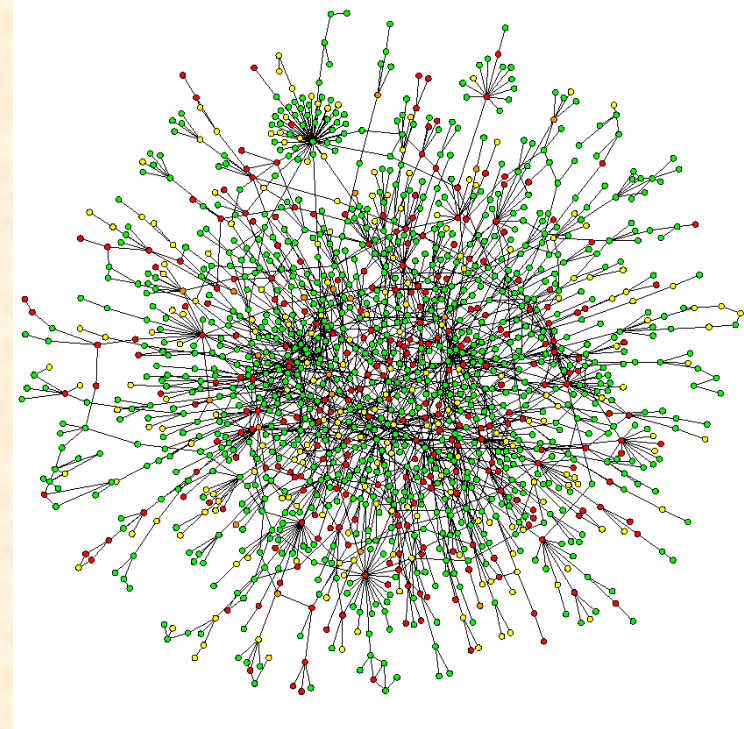


Graph of some IP addresses in the Internet
[W. R. Cheswick, Lumeta Corporation, USA]

Network Complexity (III)

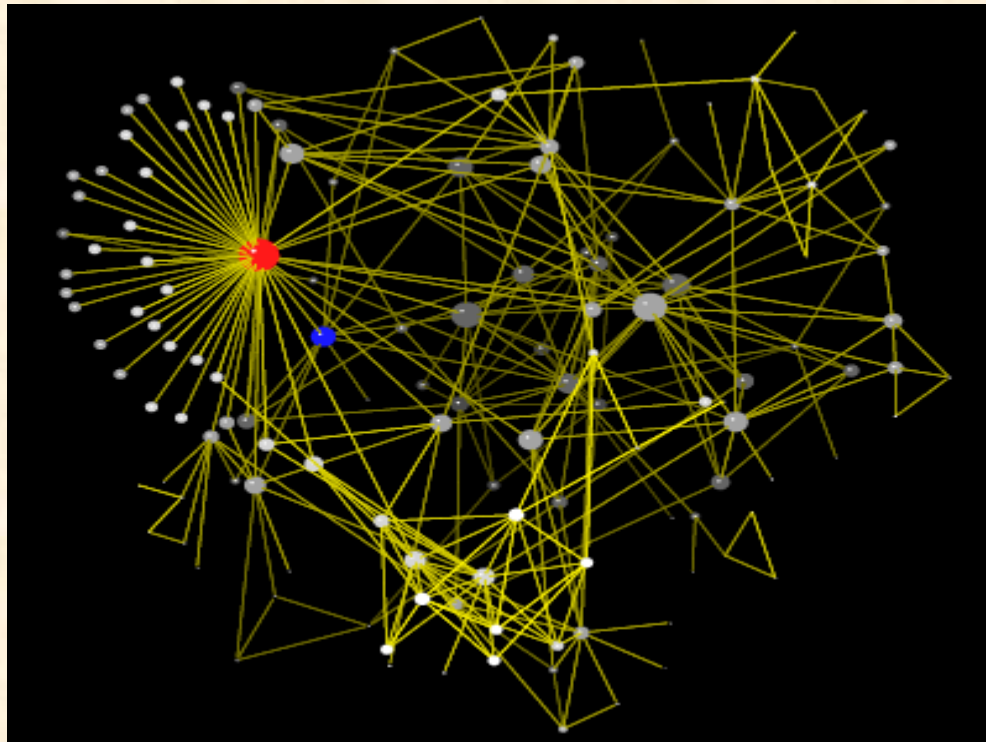
Mutual interactions among various complex factors:

- A real-world network is affected by many internal as well as external factors.
- The close relations between networks or sub-networks make the already-complicated behaviors of each of them become even more complex.



Graph of interactions among proteins
[H. Jeong, Korea Advanced Institute of Sci. Tech.]

Brief History of Network Research

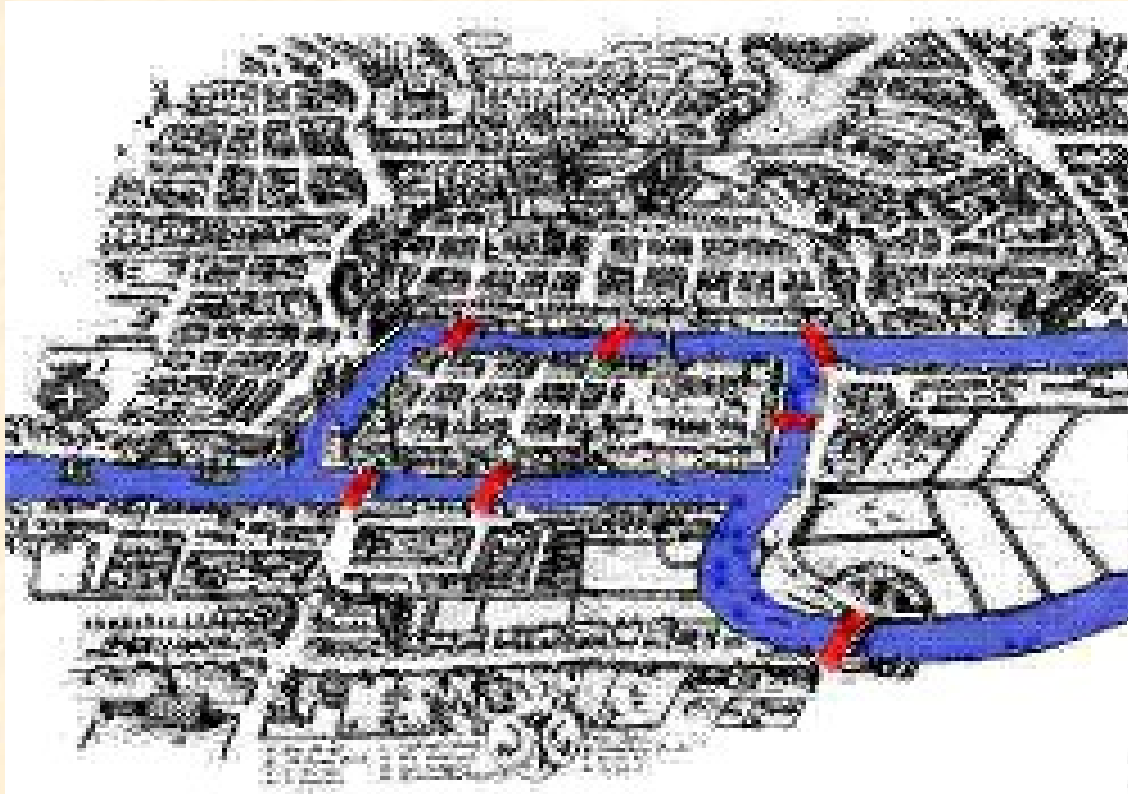


Seven-Bridges Problem of Königsberg



The town Königsburg and the seven bridges in Prussia (1525-1947)

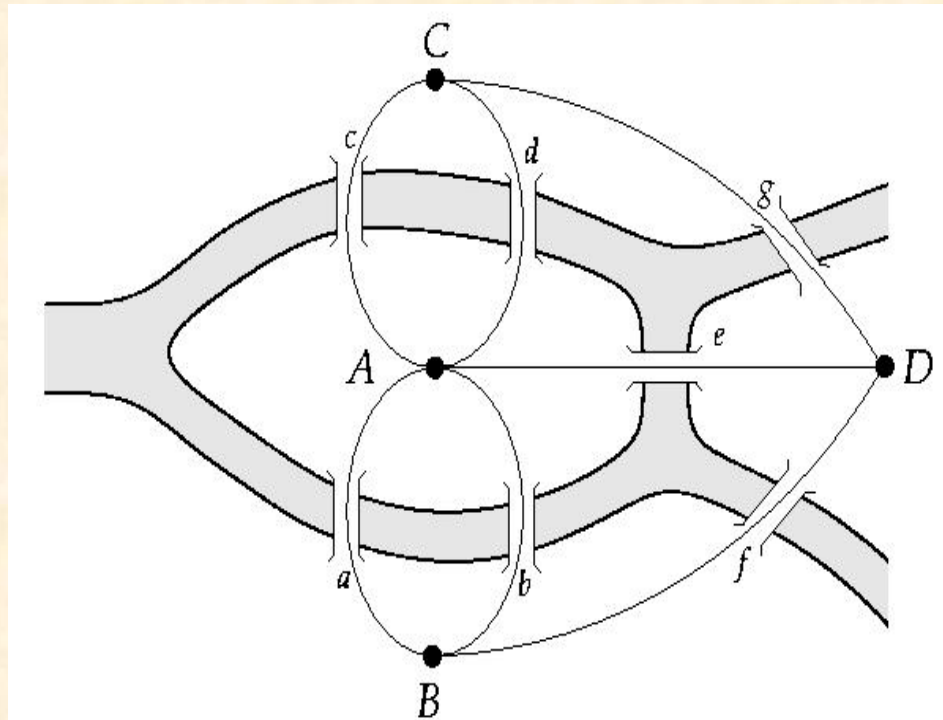
Seven-Bridges Problem of Königsberg



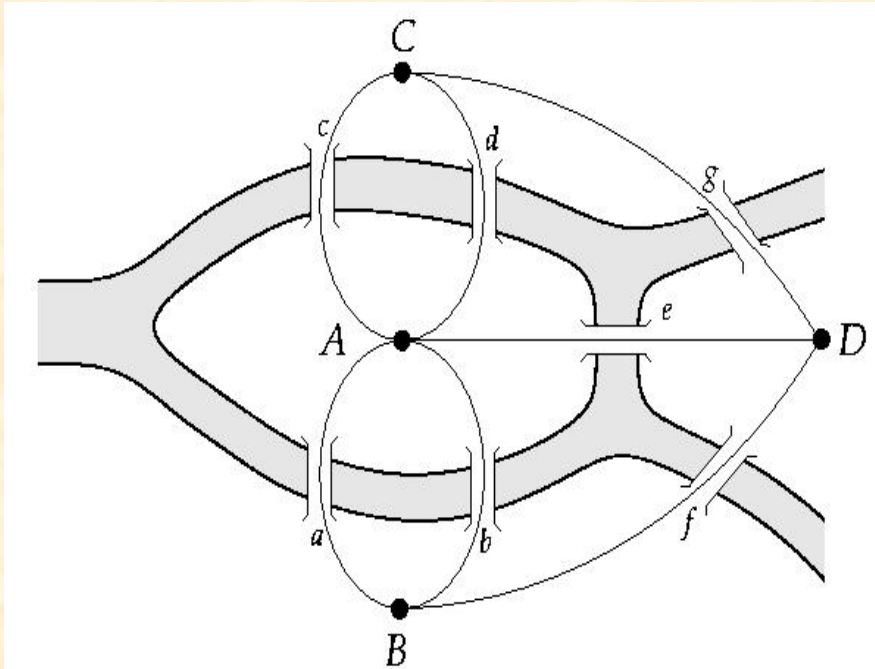
Q: Can one walk across all the seven bridges, once and once only on each bridge, and then return to the starting point ?

Leonhard Euler (1707-1783)

“Father of graph theory”



Viewed the problem
as a graph with 4
nodes and 7 edges



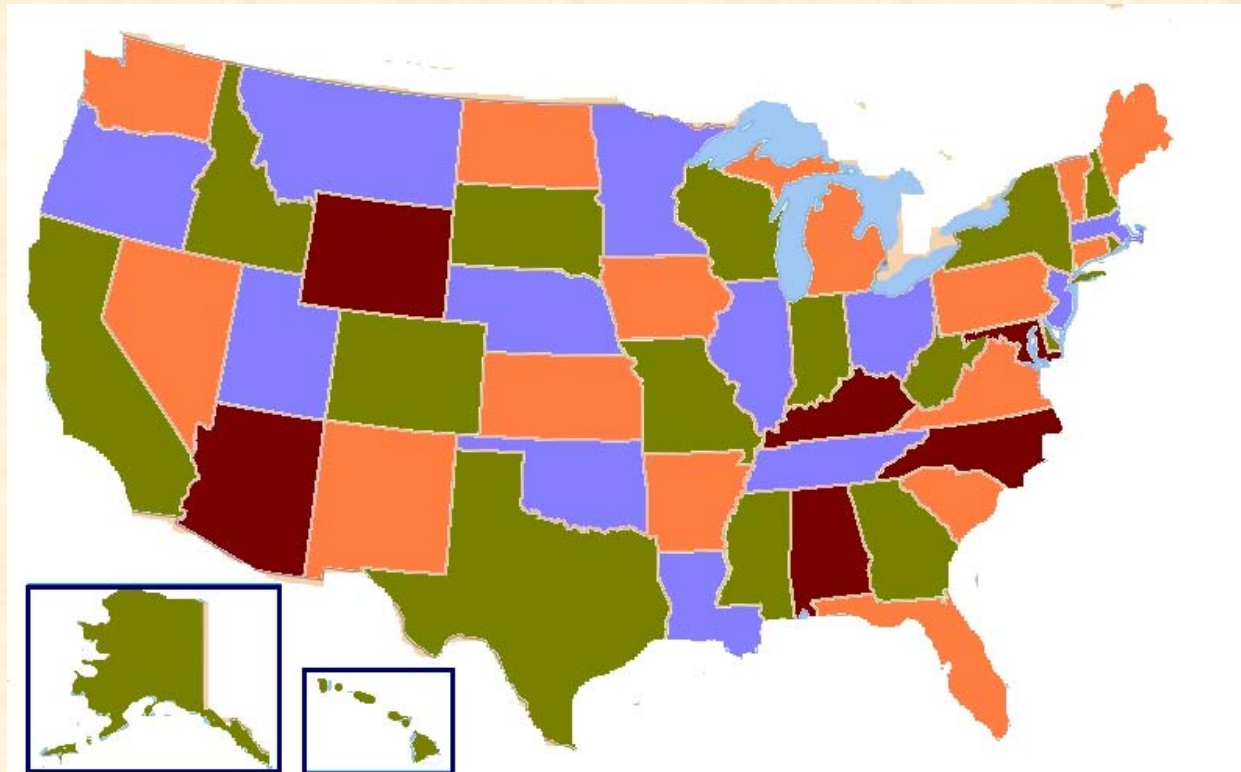
Terminology:

A **network** is a set of **nodes** (vertices) interconnected by **edges** (links)

- A path can only have one starting point and one end point → nodes with an odd number of edges must be either the starting or the end point → to have a solution, all points must have an even number of edges
- Thus, the Königsberg graph has no solution !
-- proved by Euler in 1736
- By the way, a new bridge was built between B and C in 1875

Four Color Problem

Q: Given any map, can the regions be colored by using at most **four** colors, so that no two adjacent regions have the same color?



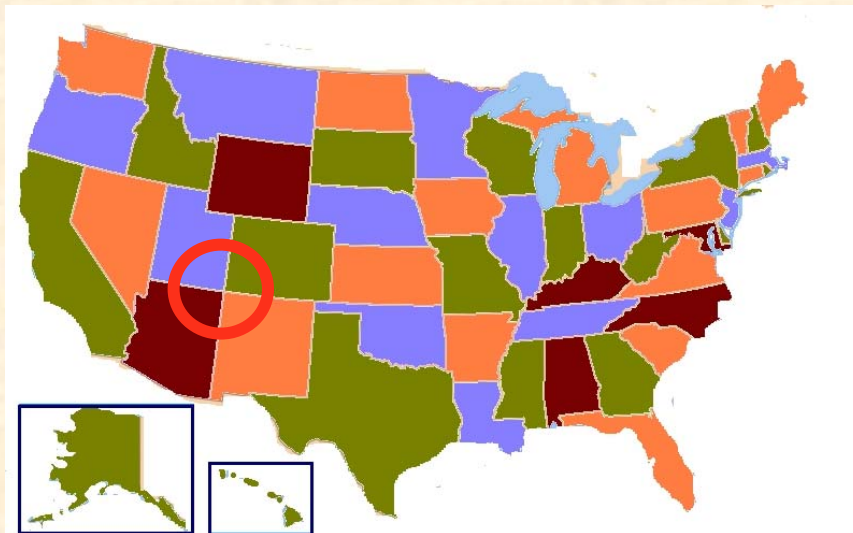
Terminology:

Two regions are called **adjacent** if they share a common boundary that is not a corner, where a **corner** is a point shared by more than two regions.

Convention:

All corners, and points that belong to more than two regions, are ignored. A region has to be a simply connected, or contiguous.

Bizarre maps (e.g. using regions of finite area but infinite perimeter) are ignored (which can require many different colors)



Q: Given any map, can the regions be colored by using at most four colors, so that no two adjacent regions have the same color?

Four Color Conjecture (Theorem):

Yes, four colors are sufficient

Brief History:

Heawood (1890): Five colors are sufficient (proved mathematically)

Appel and Haken (1976): Four colors are sufficient (proved by computer programming)

Robertson, Sanders, Seymour and Thomas (1997): Simplified the above computer-aided proof

Gonthier (2005): proved by using a general purpose theorem-proving software

Graph Theory provides a basic tool for studying complex networks

Some inferential historical markers in network science

Year	People	Event
1736	Euler	Seven-bridge problem
1959	Erdős and Rényi	Random graph theory
1967	Milgram	Small-world experiment
1973	Granovetter	Strength of weak ties
1998	Watts and Strogatz	Small-world network model
1999	Barabási and Albert	Scale-free network model

Inter-personal Ties

Johann Wolfgang von Goethe (歌德) (Germany, 1809) discussed the “marriage tie”

Anatol Rapoport (Russia, 1954): “the likely contacts of two individuals who are closely acquainted tend to be more overlapping than those of two arbitrarily selected individuals” (became a cornerstone of social network theory)

Mark Granovetter (USA 1973): “The Strength of Weak Ties” (one of the most influential sociology papers, cited by > 25,000 times)

Granovetter found that most jobs were found through “weak” acquaintances

Granovetter (1943 - , PhD from Harvard in 1970)

80-20 Rule

Vilfredo Pareto (Italy, 1906): Observed that about 80% of the land in Italy was owned by about 20% of the population

He then carried out some surveys on a number of other countries and found that all the distributions are quite similar

He further observed that 20% of the pea pods contained 80% of the peas in his garden

1992 United Nations Development Program Report: The richest 20% of population control 82.7% of the world's global income

Microsoft: By fixing the top 20% most reported bugs, 80% of the errors and crashes would be eliminated

Business: Roughly 80% of profits come from 20% of buyers; 80% of complaints come from 20% of customers; 80% of sales come from 20% of products; 80% of the sales come from 20% of the clients, ...

1% Rule (90-9-1 Rule)

1% rule (known also as 90-9-1 rule): hypothesizing that more people will lurk than will participate in a virtual community

Internet: 1% of people create content; 9% edit/modify the content; 90% only view the content without contributing anything to it

Blog: For every blog posted by 1 blogger, 9 people make some comments to follow; 99 other people only view the blog without posting comments

Note: According to a report in the 2012 BBC online briefing, about 83% of the population could properly be classified as lurkers, while 17% of the population could be classified as intense contributors

-- This is closer to the 80-20 rule instead

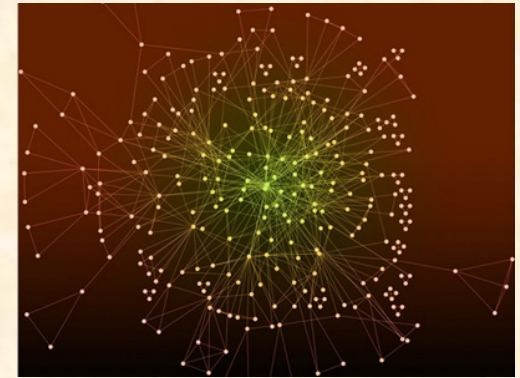
Present studies of complex networks

- ❖ **Discovering:** Trying to reveal the global statistical properties of a network and to develop measures for these properties.
- ❖ **Modeling:** Trying to establish a mathematical model of a given network, enabling better understanding of the network statistical properties and the causes of their appearance.
- ❖ **Analysis:** Trying to find out the basic characteristics and essential features of nodes, edges, and the whole network in a certain topology, to develop fundamental mathematical theories that can describe and predict the network dynamical behaviors.
- ❖ **Control:** Trying to develop effective methods and techniques that can be used to modify and improve network properties and performances, suggesting new and possibly optimal network designs and utilizations, particularly in the regards of network stability, synchronizability, and data-traffic management.
- ❖ **Applications:** Trying to apply and utilize some special and fundamental properties and characteristics of complex networks to facilitate the design and applications of network-related problems, such as data-flow congestion control on the Internet and traffic control for city transportations, optimal integrated circuit design for chip fabrication, better decision-making of policy and strategy for commercial trading and financial management, etc.

BREAK

10 minutes

Network Topology



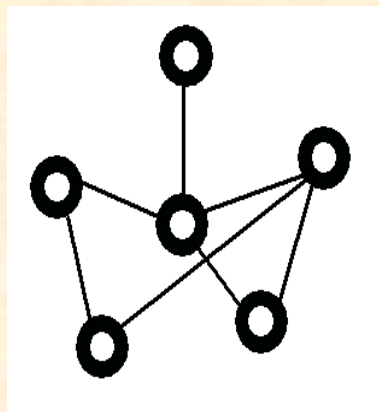
- ❖ A **network** is a set of **nodes** interconnected via **edges**

Examples:

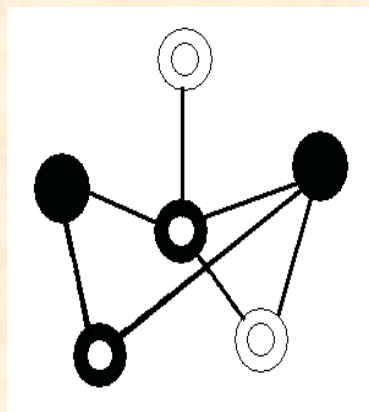
- ❖ Internet: Nodes – routers Edges – optical fibers
- ❖ WWW: Nodes – document files Edges – hyperlinks
- ❖ Scientific Citation Network:
Nodes – papers Edges – citation
- ❖ Social Networks:
Nodes – individuals Edges – relations

An Brief Introduction to Graphs Theory

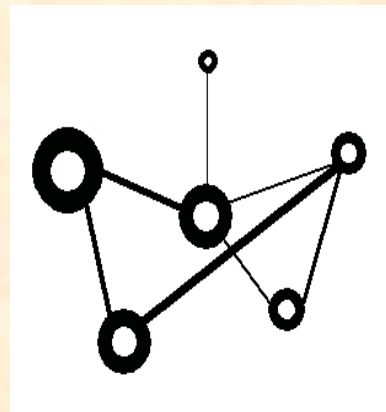
Some typical graphs:



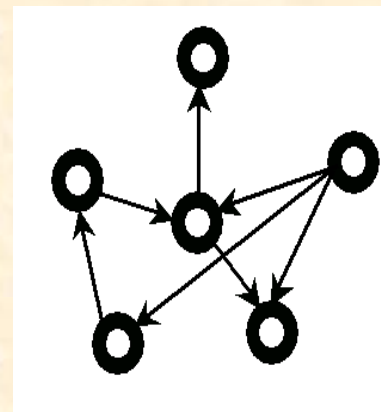
(a)



(b)



(c)



(d)

- (a) undirected and unweighted network with same type of nodes
- (b) undirected and unweighted network with different types of nodes and edges
- (c) undirected but weighted network with weights on both nodes and edges
- (d) directed but unweighted network with same type of nodes
- (e)

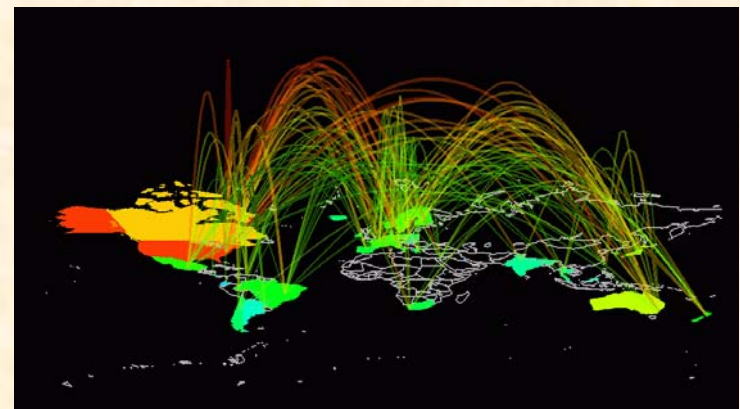
In this course: Nodes: all identical

Edges: no multi-connections and self-loops

Some Basic Concepts

- ◆ Degree and Degree Distribution
- ◆ Distance and Average Path Length
- ◆ Clustering Coefficient
- ◆ Coreness
- ◆ Betweenness
- ◆ Assortativeness

(Stephen G. Eick)



Degree and Degree Distribution

- **Degree** $k(i)$ **of node** i
= **total number of its edges**
- **Average Degree** $\langle k \rangle$ **over the network**

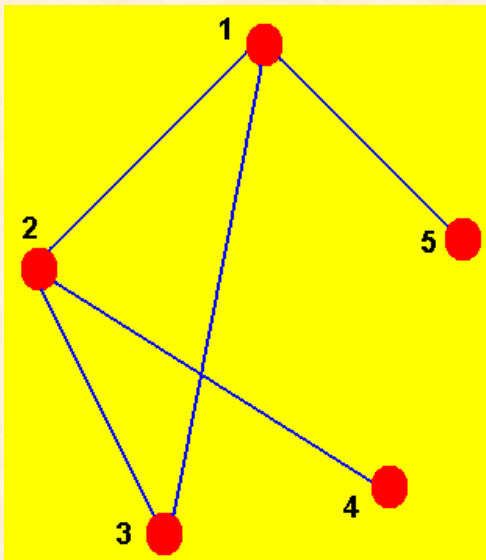
$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^N k(i)$$

- **The spread of node degrees over a network is characterized by a distribution function:**

$P(k)$ = **probability that a randomly selected node has exactly degree** k

$$\langle k \rangle = \sum_{k=0}^{\infty} kP(k)$$

Example:



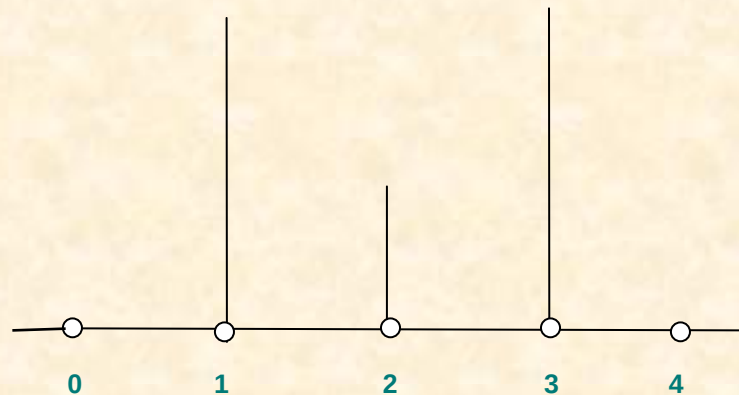
➤ Degree:

node1 = 3, node2 = 3, node3 = 2,
node4 = 1, node5 = 1

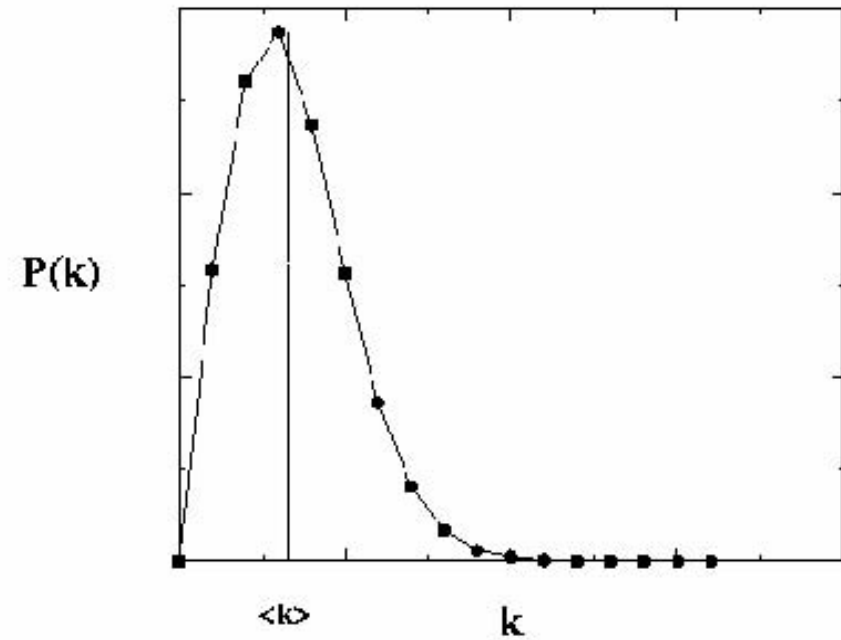
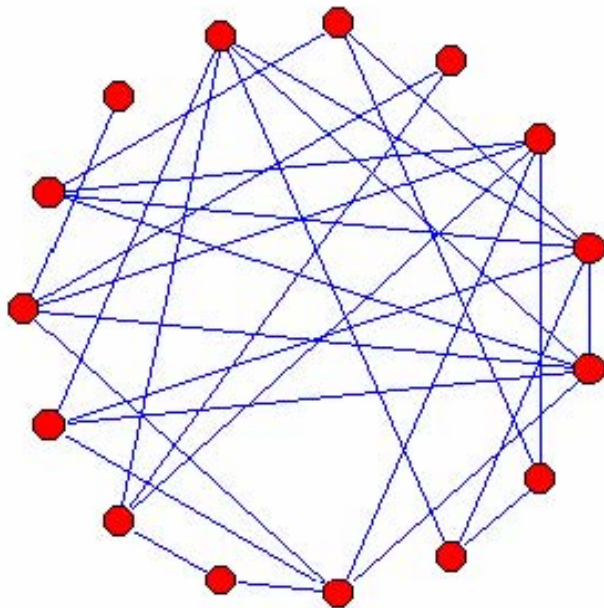
➤ Average Degree:

$$\langle k \rangle = (3 + 3 + 2 + 1 + 1) / 5 = 2$$

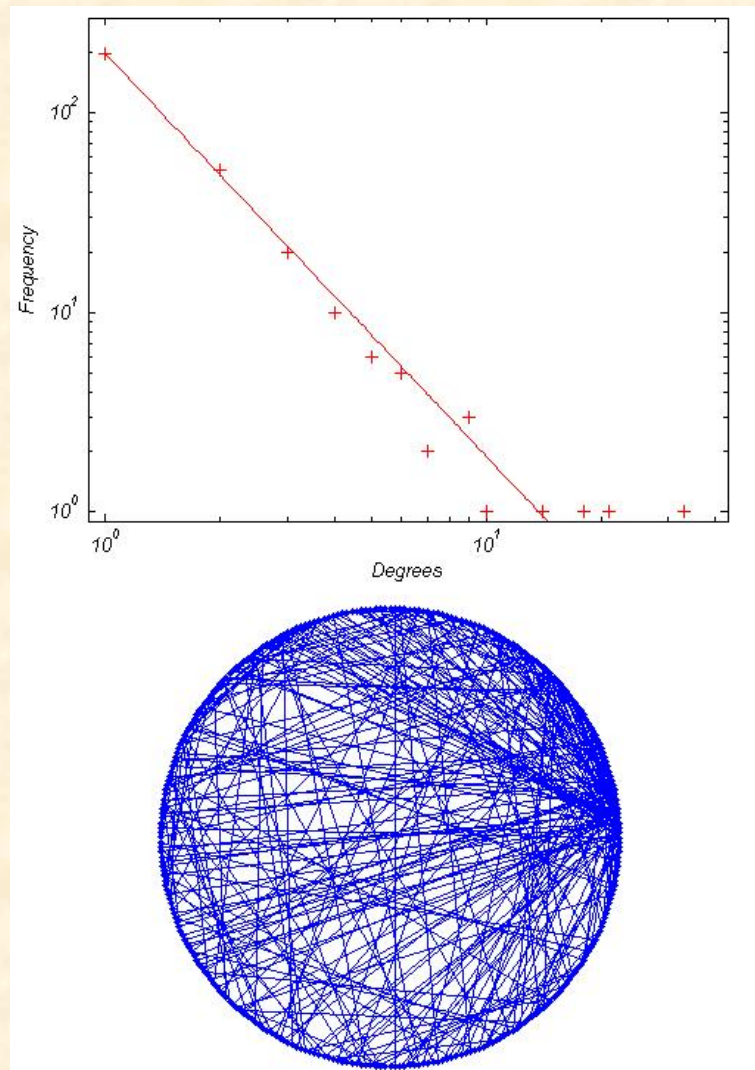
➤ Degree Distribution:



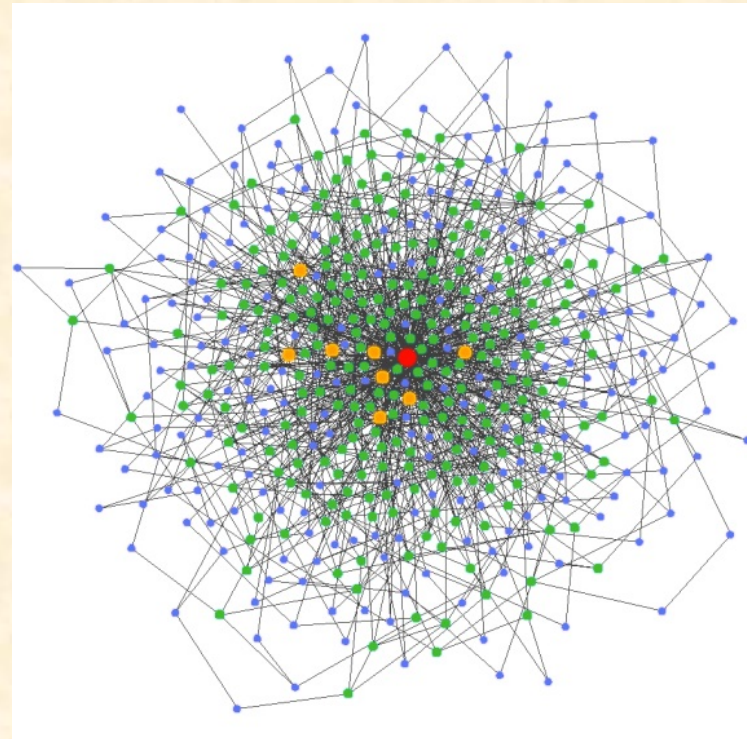
Example: Poisson Degree Distribution



Power-Law Degree Distribution



$$P(k) \sim k^{-\gamma}$$

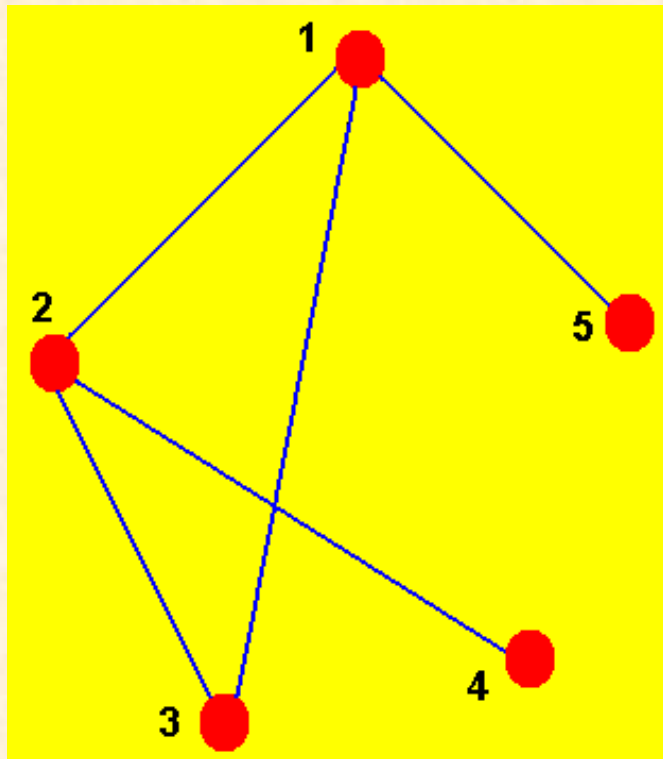


Distance and Average Path Length

- ❖ **Distance** $d_{i,j}$ **between two nodes** i **and** j
= the number of edges along the shortest path connecting them
- ❖ **Diameter** $D = \max\{d_{i,j}\}$
- ❖ **Average path length** L = average over all $d_{i,j}$
- ❖ For many large and complex networks:
small L is a small-world feature

Example:

A network having $N = 5$ nodes and 5 edges:



$$L = \frac{1}{\frac{1}{2}N(N-1)} \sum_{i < j} d_{ij}$$

$$d_{12} = 1$$

$$d_{13} = 1$$

$$d_{14} = 2$$

$$d_{15} = 1$$

$$d_{23} = 1$$

$$d_{24} = 1$$

$$d_{25} = 2$$

$$d_{34} = 2$$

$$d_{35} = 2$$

$$d_{45} = 3$$

Total = 16

Average: $L = 16 / 10 = 1.6$

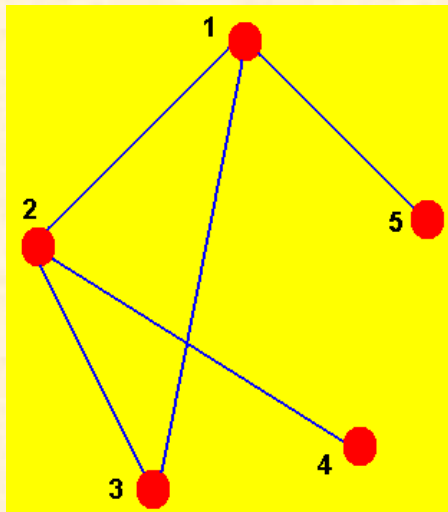
Clustering Coefficient

- ❖ **Clustering Coefficient C of a network:**
- ❖ **How many of your friends are also friends themselves?**
 - Consider a node, i
 - Suppose that node i has $k(i)$ edges (so it has $k(i)$ neighbors)
 - $E(i)$ = number of existing edges of these $k(i)$ neighbors actually have
 - $T(i)$ = number of possible edges of these $k(i)$ neighbors can have
= $k(i) [k(i)-1] / 2$
 - $C(i) = E(i) / T(i)$ -- Clustering coefficient of node i
 - C = average over all $C(i)$ -- Clustering coefficient of the network
- ◆ **Usually, $0 < C < 1$**
 - ❖ $C = 1$ **if all neighbors of a node are connected pair-wise**
 - ❖ $C = 0$ **if no neighbors of a node are connected pair-wise**
- ❖ **For many large and complex networks:**
large C is a small-world feature

Example:

How to compute the clustering coefficient ?

- Node i has $k(i)$ edges (so it has $k(i)$ neighbors)
- $E(i)$ = number of edges of these neighbors actually have
- $T(i)$ = number of edges of these neighbors can possibly have
 $= k(i) [k(i) - 1] / 2$
- $C(i) = E(i) / T(i) = 2E(i) / [k(i) (k(i) - 1)]$
- C = average over all $C(i)$



Node-1 has 3 neighbors, $E(1) = 1$, so $C(1) = 2 \times 1 / (3 \times 2) = 1/3$

Node-2 has 3 neighbors, $E(2) = 1$, so $C(2) = 2 \times 1 / (3 \times 2) = 1/3$

Node-3 has 2 neighbors, $E(3) = 1$, so $C(3) = 2 \times 1 / (2 \times 1) = 1$

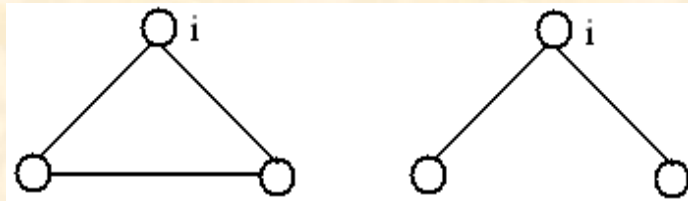
Node-4 has 1 neighbor, $E(4) = 0$, so $C(4) = 0$

Node-5 has 1 neighbor, $E(5) = 0$, so $C(5) = 0$

Average $C = (1/3 + 1/3 + 1 + 0 + 0) / 5 = 1/3$

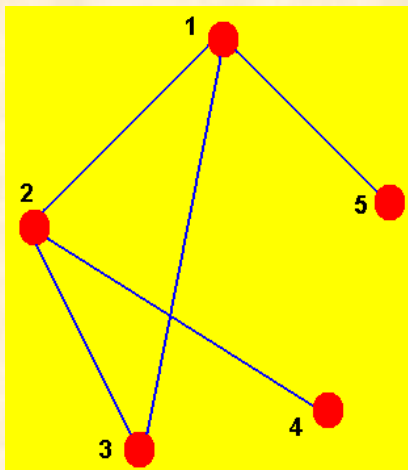
Another definition from the geometric viewpoint:

$$C(i) = \frac{\text{number of complete triangles with corner } i}{\text{number of all triangular graphs with corner } i}$$



Left – complete triangle is counted in both numerator and denominator

Right – incomplete triangular graph is counted only in denominator



Node-1 has 1 complete triangle and 3 triangular graphs, so $C(1) = 1/3$

Node-2 has 1 complete triangle and 3 triangular graphs, so $C(2) = 1/3$

Node-3 has 1 complete triangle and 1 triangular graph, so $C(3) = 1$

Node-4 has 0 complete triangles, so $C(4) = 0$

Node-5 has 0 complete triangles, so $C(5) = 0$

Average $C = (1/3 + 1/3 + 1 + 0 + 0) / 5 = 1/3$

Coreness

Definition.:

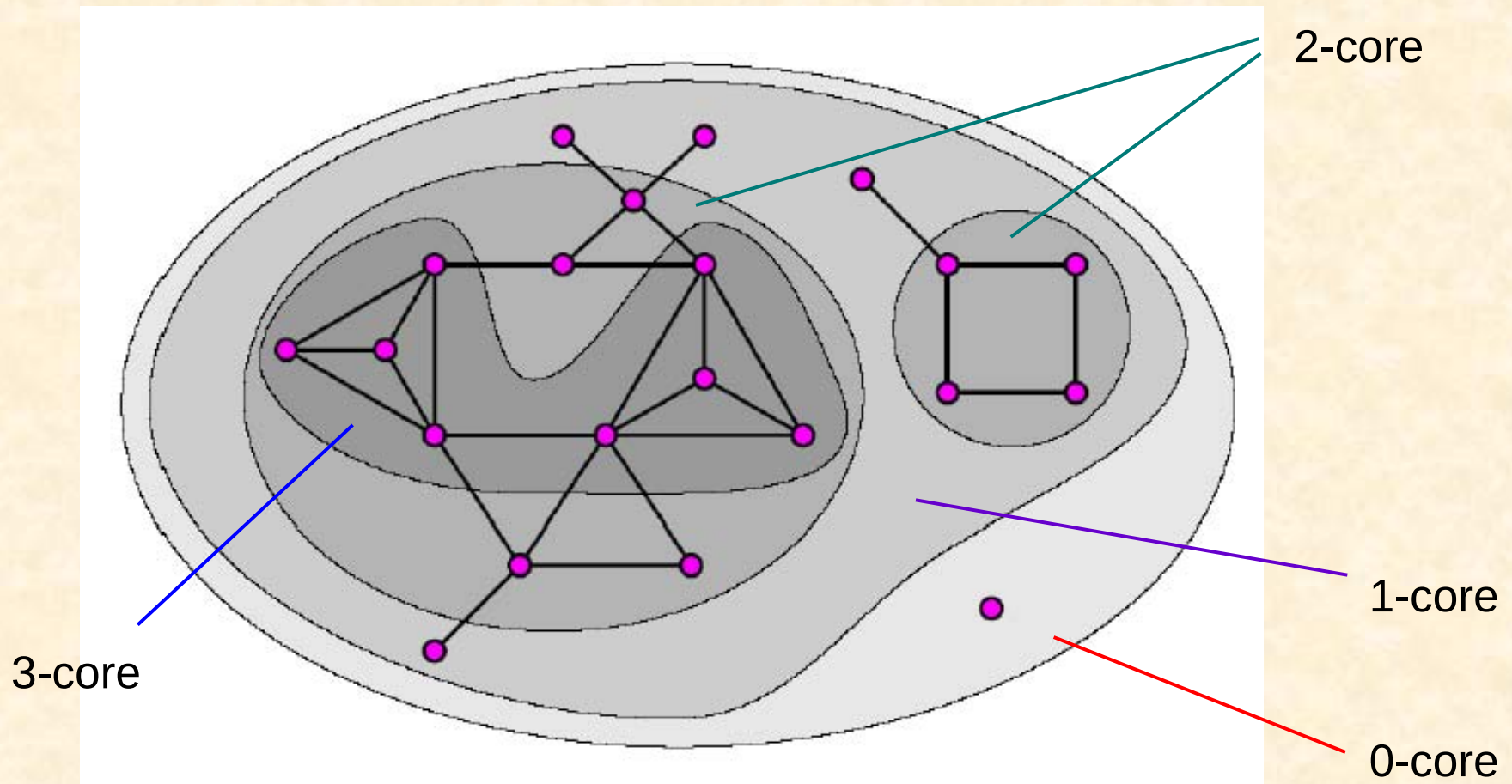
- ❖ The *k-core* in a graph is defined to be the remaining sub-graph after all the nodes with degrees $\leq k - 1$ have been removed successively, during which:
 - (i) when a node is removed, all its adjacent edges will also be removed;
 - (ii) after a node of degree $\leq k - 1$ is removed, in the remaining graph all the remaining nodes with a new degree $\leq k - 1$ also need to be removed.
- If a node belongs to a *k-core* of a graph, but it will be removed from the $(k+1)$ -core, then this node is said to have *coreness* (core value) *k*.
- The largest coreness in a graph is called the *coreness of the graph*.

Examples:

An isolated node has coreness $k = 0$

A fully-connected network of size N has coreness $k = N - 1$

Coreness



Coreness

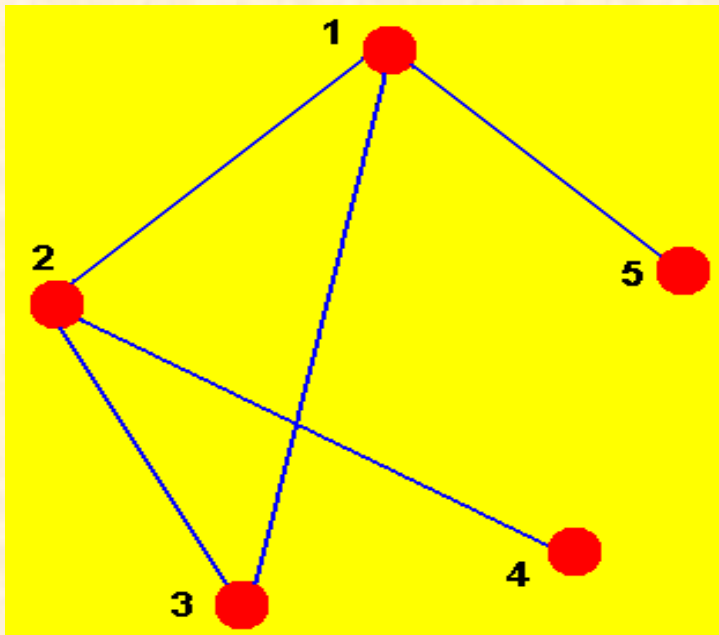
Star-shaped networks

- ❖ the 1-core of the network is the network itself
- ❖ all nodes, including the central node, have coreness 1
- ❖ the coreness of the network is 1

Ring networks

- ❖ the 1-core of the network is the network itself
- ❖ the 2-core of the network is the network itself
- ❖ all nodes have coreness 2
- ❖ the coreness of the network is 2

Coreness



Meaning here:

- Preserving 2-core
- = Cutting nodes of coreness 1
- = Removing tree-like subnets

Examples:

- 1-core is the network itself
- 2-core is Triangle 1-2-3
- Node 1, Node 2, Node 3 have coreness 2
- Node 4 and Node 5 have coreness 1
- Coreness of the graph is 2

Betweenness

Definition 1 (Node-betweenness) In a network of size N , the *node-betweenness* of node i is defined by

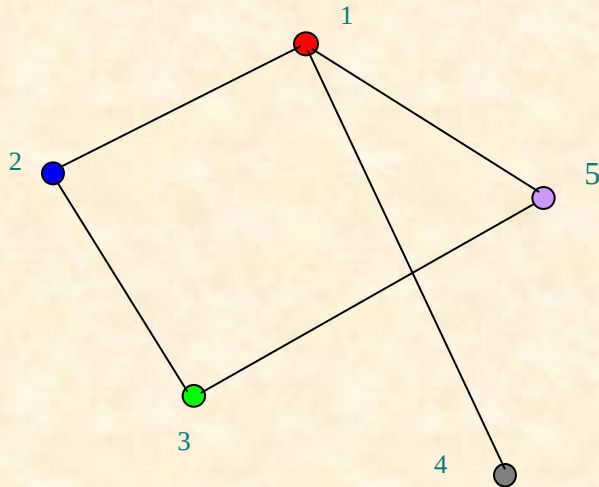
$$B(i) = \sum_{i \neq j \neq l} \frac{L_{jl}(i)}{L_{jl}}$$

where L_{jl} is the number of all existing shortest paths from node j to node l , and $L_{jl}(i)$ is the number of all shortest paths from node j to node l that pass through node i .

(since $i \neq j \neq l$ the node i itself is excluded)

Normalization: divided by total number of node-pairs not including node i as an end node: $(N-1)(N-2)/2$

Node-Betweenness



Example:
betweenness of node 1 is

$$\begin{aligned}
 B(1) &= \frac{(5,1,4)}{(5,1,4)} + \frac{0}{(5,3)} + \frac{(5,1,2)}{(5,1,2) + (5,3,2)} + \frac{(4,1,2,3) + (4,1,5,3)}{(4,1,2,3) + (4,1,5,3)} + \frac{(4,1,2)}{(4,1,2)} + \frac{0}{(3,2)} \\
 &= \frac{1}{1} + \frac{1}{2} + \frac{2}{2} + \frac{1}{1} = \frac{7}{2}
 \end{aligned}$$

Normalized: $B(1) = \frac{7/2}{(N-1)(N-2)/2} = \frac{7}{12}$

Betweenness

Definition 2 (Edge-betweenness)

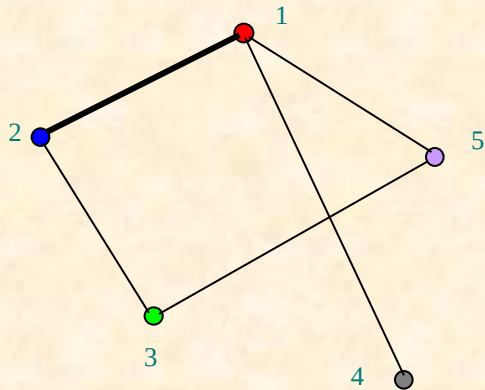
In a network of size N , the *edge-betweenness* of edge e_{ij} is defined by

$$B(e_{ij}) = \sum_{(l,q) \neq (i,j)} \frac{\tilde{L}_{lq}(e_{ij})}{\tilde{L}_{lq}}$$

where $\tilde{L}_{lq}$ is the number of all existing shortest paths from node l to node q , and $\tilde{L}_{lq}(e_{ij})$ is the number of all shortest paths from node l to node q that pass through edge e_{ij}

Normalization: divided by total number of edges not including e_{ij}
namely, divided by $N(N-1)/2 - 1$

Edge-Betweenness



Example:
betweenness of edge e_{12} is

$$\begin{aligned}
 B(e_{12}) &= \frac{0}{(5,1,4)} + \frac{0}{(5,3)} + \frac{(5,1,2)}{(5,1,2) + (5,3,2)} + \frac{(4,1,2,3)}{(4,1,2,3) + (4,1,5,3)} + \frac{(4,1,2)}{(4,1,2)} + \frac{0}{(3,2)} + \frac{(3,2,1)}{(3,2,1) + (3,5,1)} \\
 &= \frac{1}{2} + \frac{1}{2} + \frac{1}{1} + \frac{1}{2} = \frac{5}{2}
 \end{aligned}$$

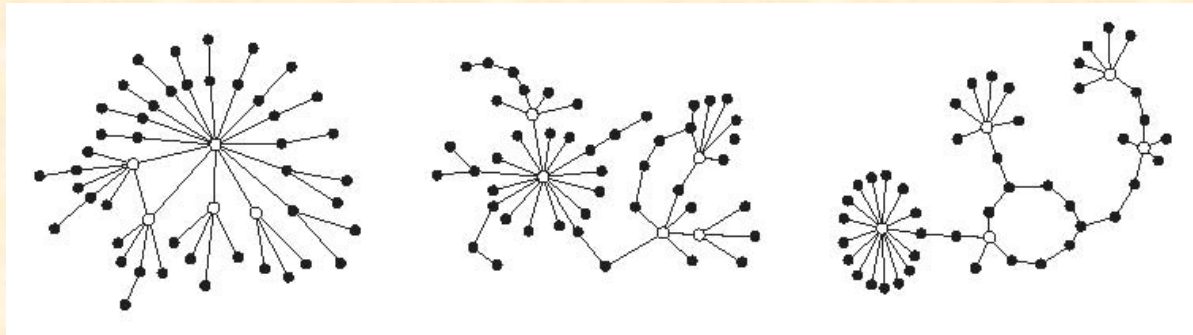
Normalized: $B(e_{12}) = \frac{5/2}{N(N-1)/2 - 1} = \frac{5}{18}$

Assortativity

In a network, if nodes with **large** degrees tend to connect to nodes also with **large** degrees on average, the network is said to be **assortative**

but if nodes with **large** degrees tend to connect to nodes with **small** degrees on average, the network is said to be **disassortative**

Note: networks having the same degree sequence (hence, the same **assortativity**) may have different topologies:



Statistical Properties of Some Real-World Complex Networks

	Network	Type	N	M	$\langle k \rangle$	L	γ	C
Social science	Film actors	undirected	449913	25516482	113	3.48	2.3	0.78
	Company directors	undirected	7673	55392	14.4	4.6	–	0.88
	Math coauthorship	undirected	253339	496489	3.92	7.57	–	0.34
	Physics coauthorship	undirected	52909	245300	9.27	6.19	–	0.56
	Biology coauthorship	undirected	1520251	11803064	15.5	4.92	–	0.6
	Telephone call graph	undirected	47000000	80000000	3.16			
	E-mail messages	undirected	59912	86300	1.44	4.95	1.5/2.0	0.16
	E-mail addresses books	undirected	16881	57029	3.38	5.22	–	0.13
	Student relationships	undirected	573	477	1.66	16	–	0
	Sexual contacts	undirected	2810				3.2	

N – number of nodes; M – number of edges; $\langle k \rangle$ - average degree;

L – average path length; r – power-law exponent; C – clustering coefficient

Statistical Properties of Some Real-World Complex Networks

	Network	Type	N	M	$\langle k \rangle$	L	γ	C
Information Science	WWW nd.edu	directed	269504	1497135	5.55	11.3	2.1/2.4	0.29
	WWW Altavista	directed	203549046	2.13E+09	10.5	16.2	2.1/2.7	
	Citation network	directed	783339	6716198	8.57		3.0/–	
	Roget's Thesaurus	directed	1022	5103	4.99	4.87	–	0.15
	Word co-occurrence	undirected	460902	1.7E+07	70.1		2.7	0.44
Technology	Internet (AS-level)	undirected	10697	31992	5.98	3.31	2.5	0.39
	Power grid	undirected	4941	6594	2.67	19	–	0.08
	Train routes	undirected	587	19603	66.8	2.16	–	0.69
	Software packages	directed	1439	1723	1.2	2.42	1.6/1.4	0.08
	Software classes	directed	1377	2213	1.61	1.51	–	0.01
	Electric circuits	undirected	24097	53248	4.34	11.1	3	0.03
	Peer-to-peer network	undirected	880	1296	1.47	4.28	2.1	0.01

Statistical Properties of Some Real-World Complex Networks

	Network	Type	N	M	$\langle k \rangle$	L	γ	C
Biology	Metabolic network	undirected	765	3686	9.64	2.56	2.2	0.67
	Protein network	undirected	2115	2240	2.12	6.8	2.4	0.07
	Marine food web	directed	135	598	4.43	2.05	–	0.23
	Freshwater food web	directed	92	997	10.8	1.9	–	0.09
	Neural network	directed	307	2359	7.68	3.97	–	0.28

Common Features:

- Small L and Small C → Random-Graph Networks
- Small L but Large C → Small-World Networks
- Power-Law $\sim k^{-\gamma}$ → Scale-Free Networks

End

