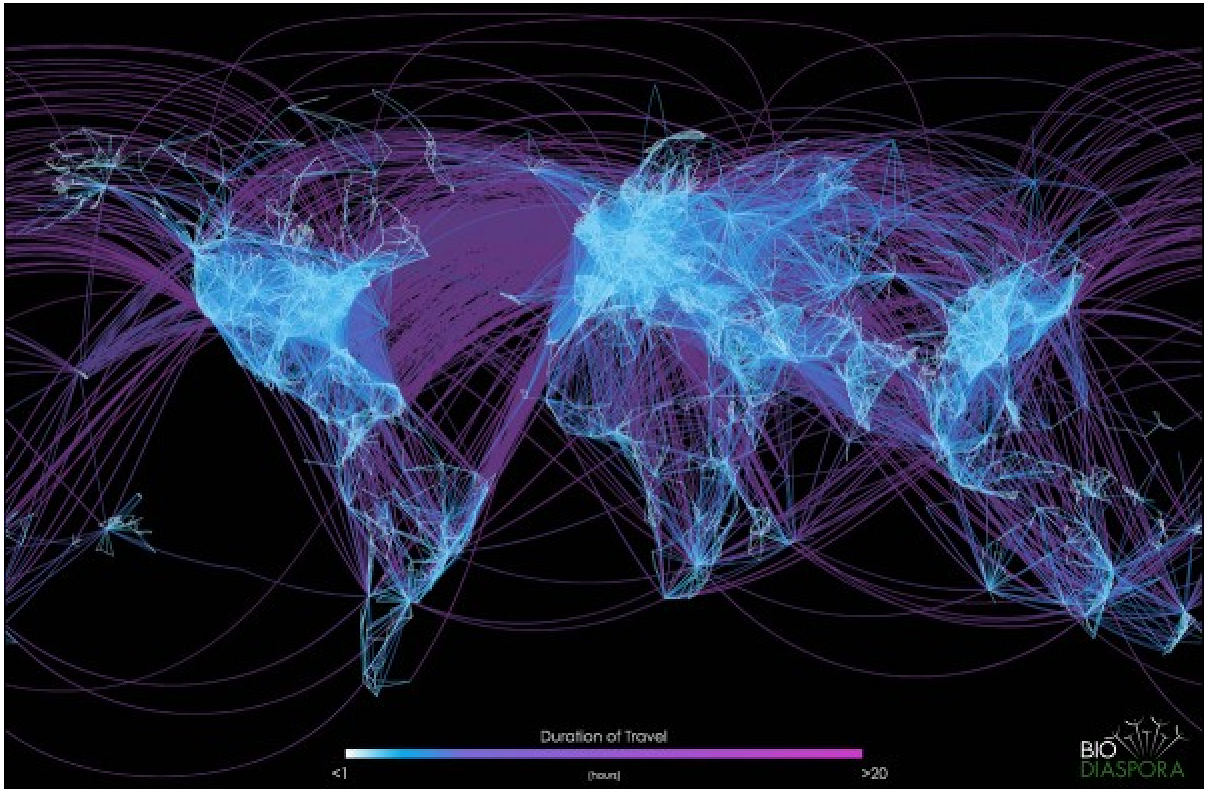
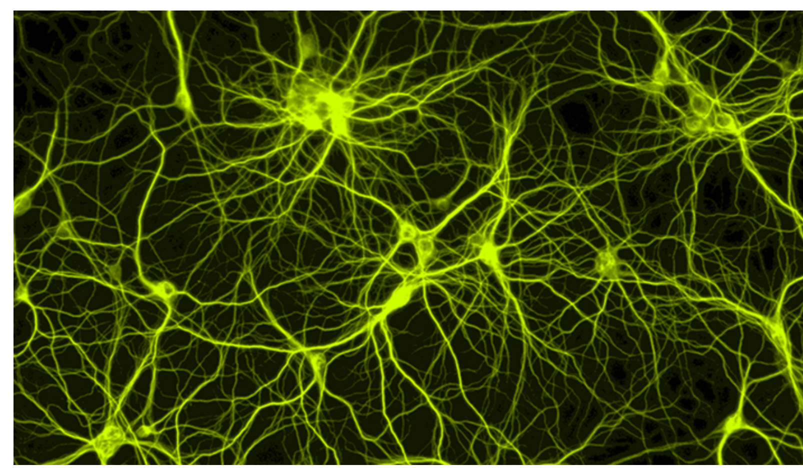
Introduction to Complex Networks

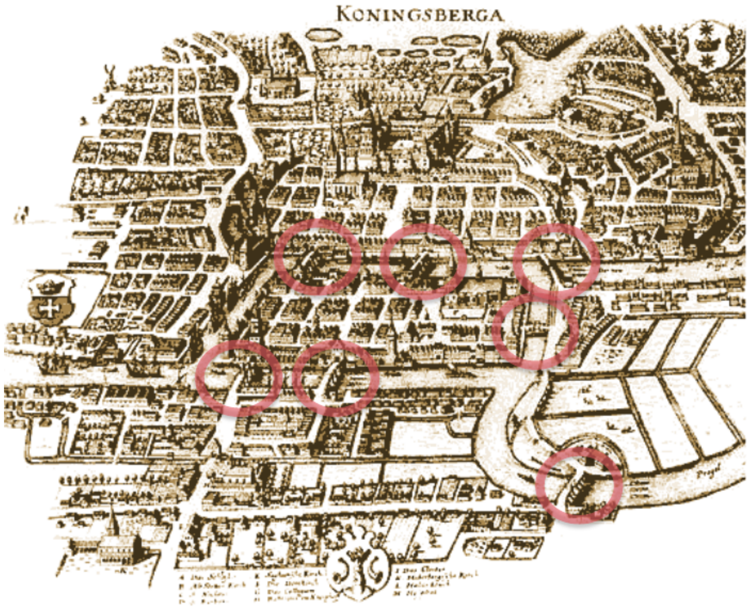


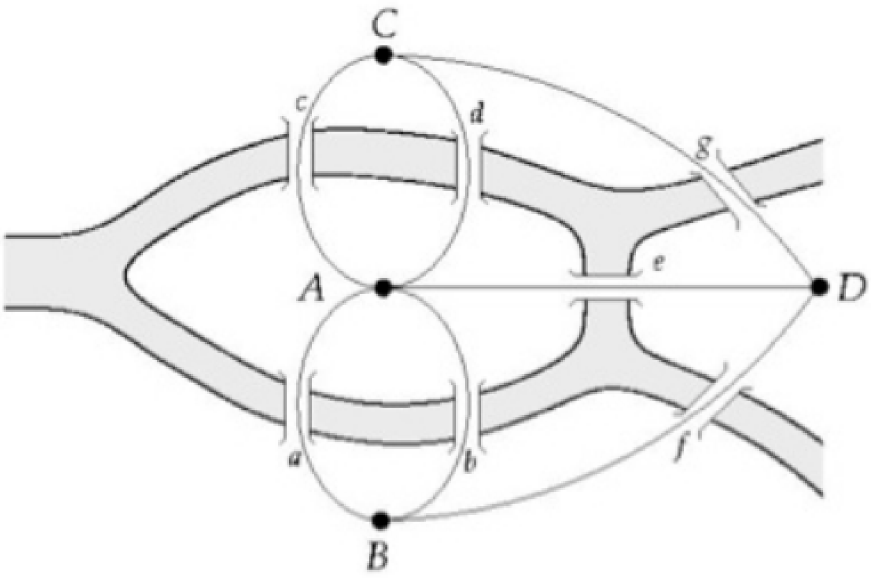
Complexity



“The whole is more than the sum of its parts.”

Graph theory



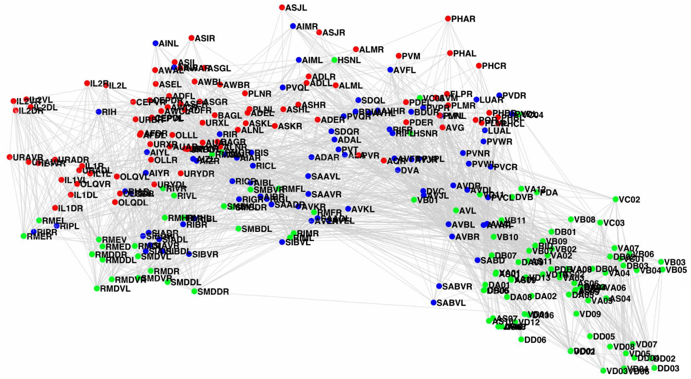


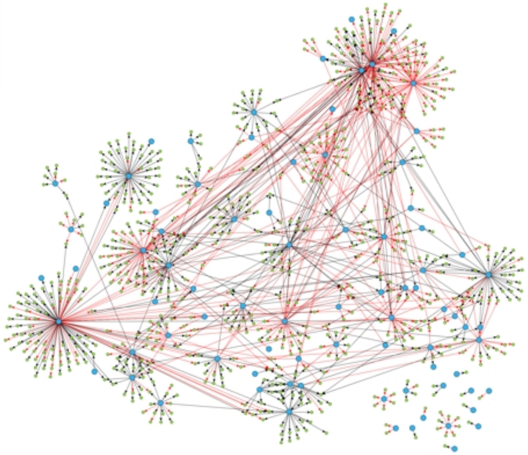
1735: EULER'S THEOREM:

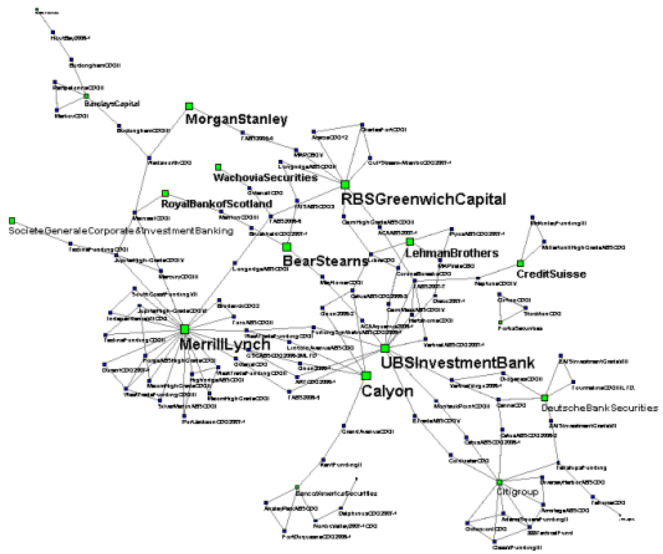
(A) IF A GRAPH HAS MORE THAN TWO NODES OF ODD DEGREE, THERE IS NO PATH.

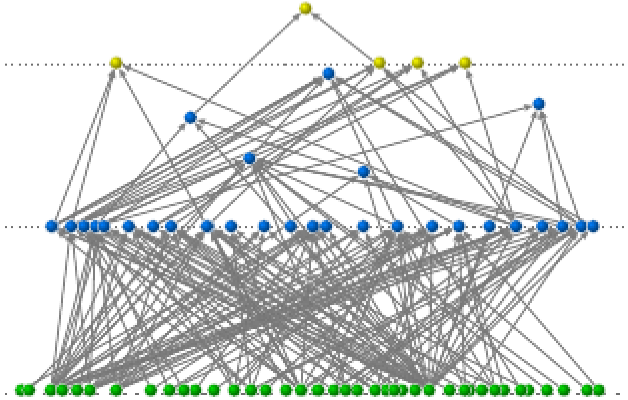
(B) IF A GRAPH IS CONNECTED AND HAS NO ODD DEGREE NODES, IT HAS AT LEAST ONE PATH.

Some real networks









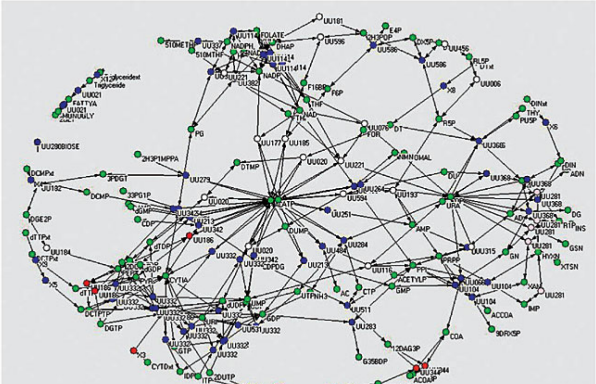
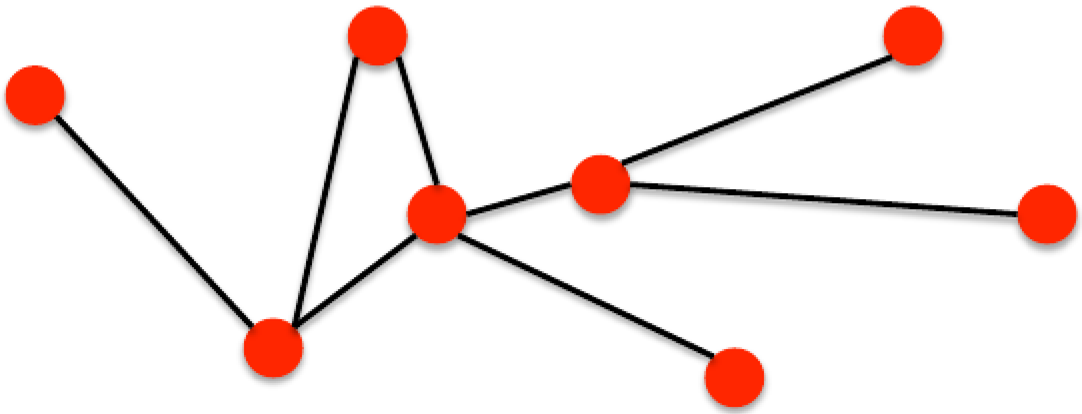


Figure 2. Bipartite graph of the metabolic network of *Ureaplasma urealyticum*. Dark gray and white nodes represent enzymes and light gray nodes represent metabolites (Lemke et al., 2004).

Nodes and links



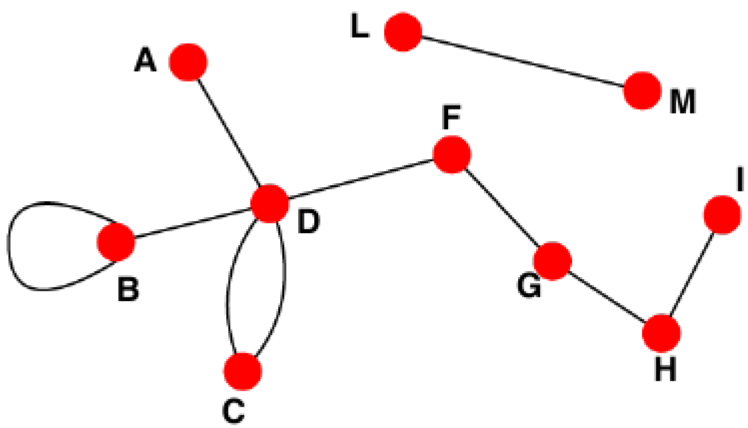
- **components**: nodes, vertices N
- **interactions**: links, edges L
- **system**: network, graph (N,L)

Nodes and links

Undirected

Links: undirected (*symmetrical*)

Graph:

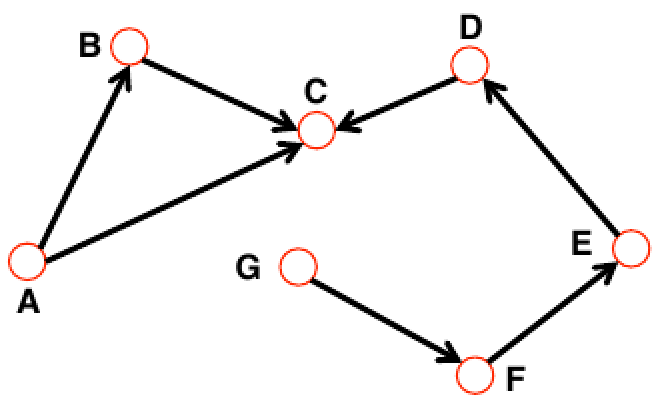


Undirected links :
coauthorship links
Actor network
protein interactions

Directed

Links: directed (*arcs*).

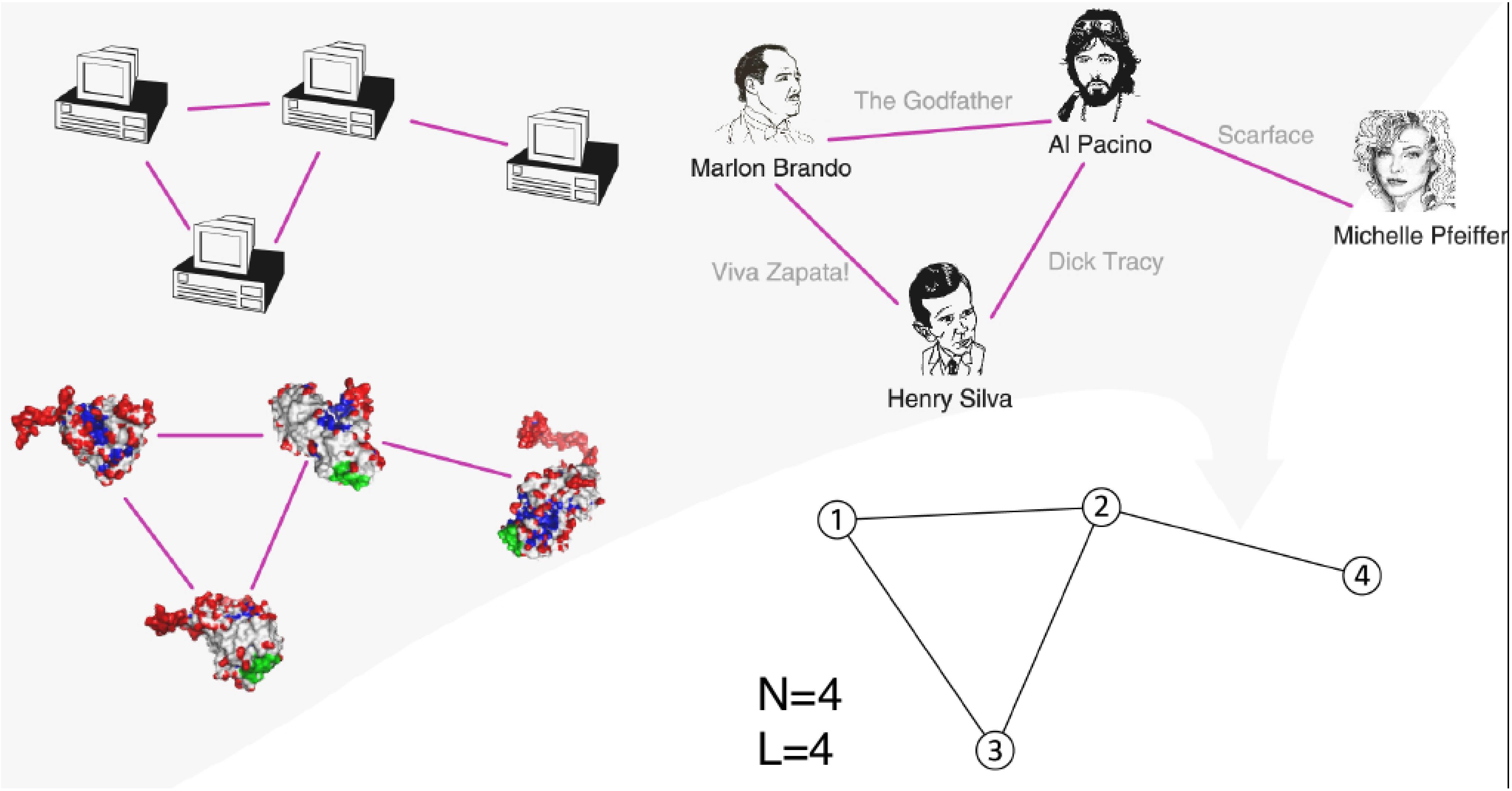
Digraph = directed graph:



An undirected link is the superposition of two opposite directed links.

Directed links :
URLs on the www
phone calls
metabolic reactions

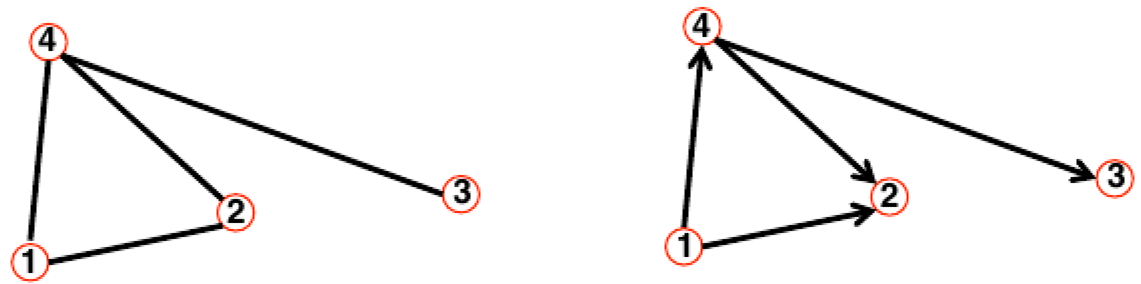
Nodes and links



Nodes and links

NETWORK	NODES	LINKS	DIRECTED UNDIRECTED	N	L
Internet	Routers	Internet connections	Undirected	192,244	609,066
WWW	Webpages	Links	Directed	325,729	1,497,134
Power Grid	Power plants, transformers	Cables	Undirected	4,941	6,594
Mobile Phone Calls	Subscribers	Calls	Directed	36,595	91,826
Email	Email addresses	Emails	Directed	57,194	103,731
Science Collaboration	Scientists	Co-authorship	Undirected	23,133	93,439
Actor Network	Actors	Co-acting	Undirected	702,388	29,397,908
Citation Network	Paper	Citations	Directed	449,673	4,689,479
E. Coli Metabolism	Metabolites	Chemical reactions	Directed	1,039	5,802
Protein Interactions	Proteins	Binding interactions	Undirected	2,018	2,930

Adjacency matrix



$A_{ij}=1$ if there is a link between node *i* and *j*

$A_{ij}=0$ if nodes *i* and *j* are not connected to each other.

$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \qquad A_{ij} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

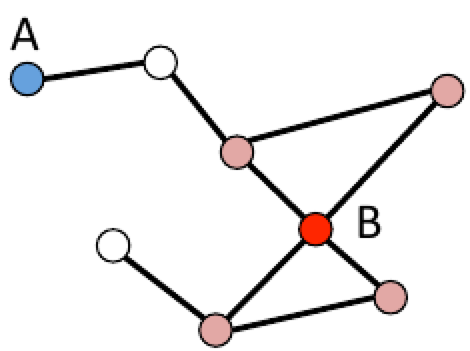
Note that for a directed graph (right) the matrix is not symmetric.

$A_{ij} = 1$ if there is a link pointing from node *j* and *i*

$A_{ij} = 0$ if there is no link pointing from *j* to *i*.

Node degree

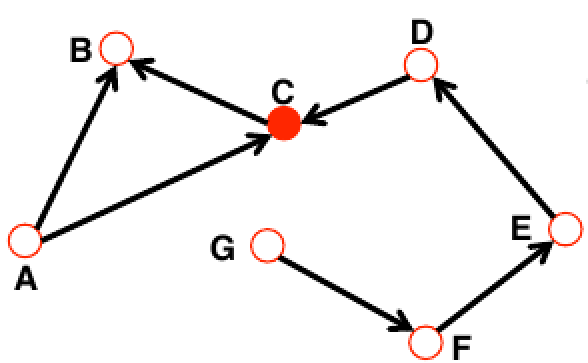
Undirected



Node degree: the number of links connected to the node.

$$k_A = 1 \qquad k_B = 4$$

Directed



In *directed networks* we can define an in-degree and out-degree.

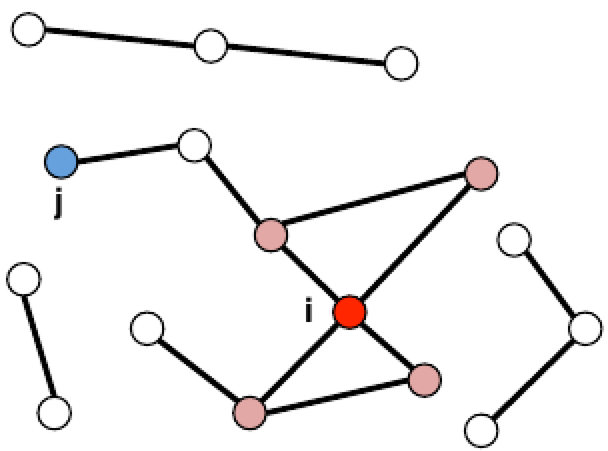
The (total) degree is the sum of in- and out-degree.

$$k_C^{in} = 2 \qquad k_C^{out} = 1 \qquad k_C = 3$$

Source: a node with $k^{in} = 0$; Sink: a node with $k^{out} = 0$.

Node degree

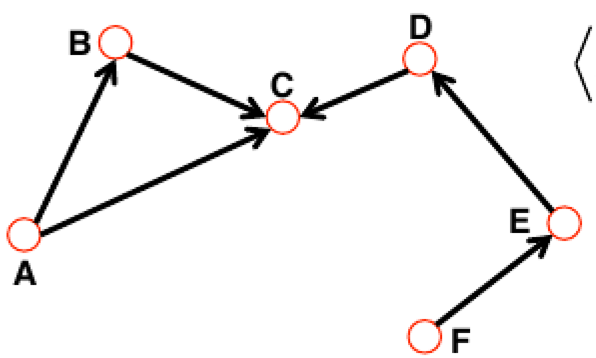
Undirected



$$\langle k \rangle \equiv \frac{1}{N}\sum_{i=1}^{N} k_i \qquad \langle k \rangle \equiv \frac{2L}{N}$$

N – the number of nodes in the graph

Directed



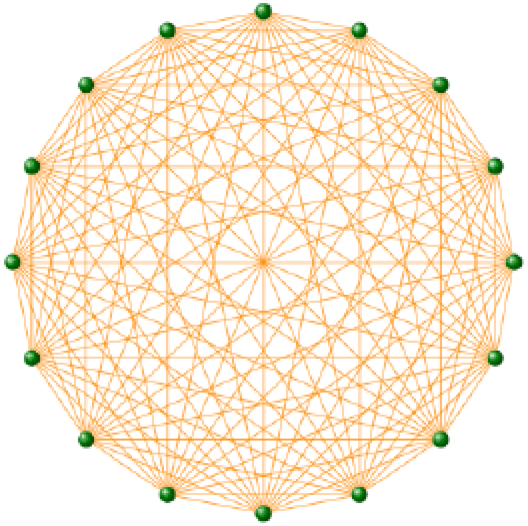
$$\langle k^{in} \rangle \equiv \frac{1}{N}\sum_{i=1}^{N} k_i^{in}, \quad \langle k^{out} \rangle \equiv \frac{1}{N}\sum_{i=1}^{N} k_i^{out}, \quad \langle k^{in} \rangle = \langle k^{out} \rangle$$

$$\langle k \rangle \equiv \frac{L}{N}$$

Node degree

NETWORK	NODES	LINKS	DIRECTED UNDIRECTED	N	L	⟨k⟩
Internet	Routers	Internet connections	Undirected	192,244	609,066	6.33
WWW	Webpages	Links	Directed	325,729	1,497,134	4.60
Power Grid	Power plants, transformers	Cables	Undirected	4,941	6,594	2.67
Mobile Phone Calls	Subscribers	Calls	Directed	36,595	91,826	2.51
Email	Email addresses	Emails	Directed	57,194	103,731	1.81
Science Collaboration	Scientists	Co-authorship	Undirected	23,133	93,439	8.08
Actor Network	Actors	Co-acting	Undirected	702,388	29,397,908	83.71
Citation Network	Paper	Citations	Directed	449,673	4,689,479	10.43
E. Coli Metabolism	Metabolites	Chemical reactions	Directed	1,039	5,802	5.58
Protein Interactions	Proteins	Binding interactions	Undirected	2,018	2,930	2.90

Node degree



The maximum number of links a network of N nodes can have is: $L_{\max} = \binom{N}{2} = \frac{N(N-1)}{2}$

A graph with degree $L=L_{max}$ is called a complete graph, and its average degree is **<k>=N-1**

Node degree

Most networks observed in real systems are sparse:

$$L \ll L_{max}$$

or

$$\langle k \rangle \ll N-1.$$

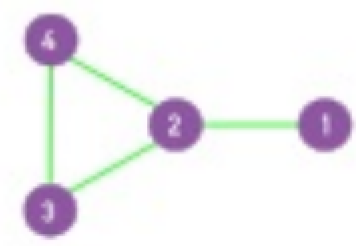
WWW (ND Sample):	N=325,729;	$L=1.4\ 10^6$	$L_{max}=10^{12}$	$\langle k \rangle=4.51$
Protein (*S. Cerevisiae*):	N= 1,870;	L=4,470	$L_{max}=10^7$	$\langle k \rangle=2.39$
Coauthorship (Math):	N= 70,975;	$L=2\ 10^5$	$L_{max}=3\ 10^{10}$	$\langle k \rangle=3.9$
Movie Actors:	N=212,250;	$L=6\ 10^6$	$L_{max}=1.8\ 10^{13}$	$\langle k \rangle=28.78$

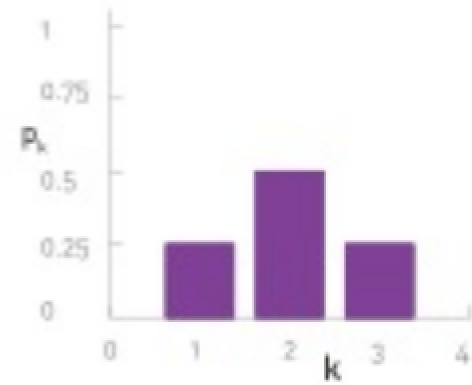
(Source: Albert, Barabasi, RMP2002)

Degree distribution

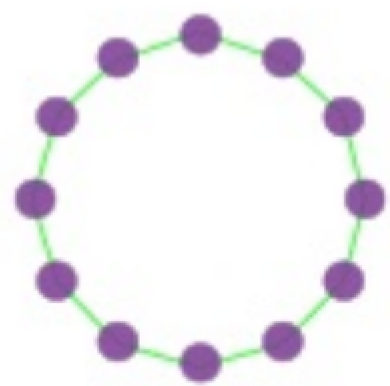
Degree distribution

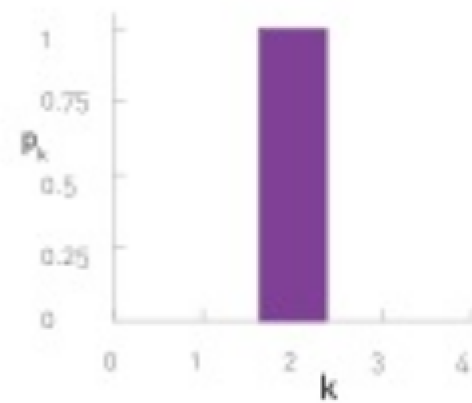
P(k): probability that a randomly chosen node has degree *k*





N_k = # nodes with degree k





$P(k) = N_k / N$ → plot

Degree distribution

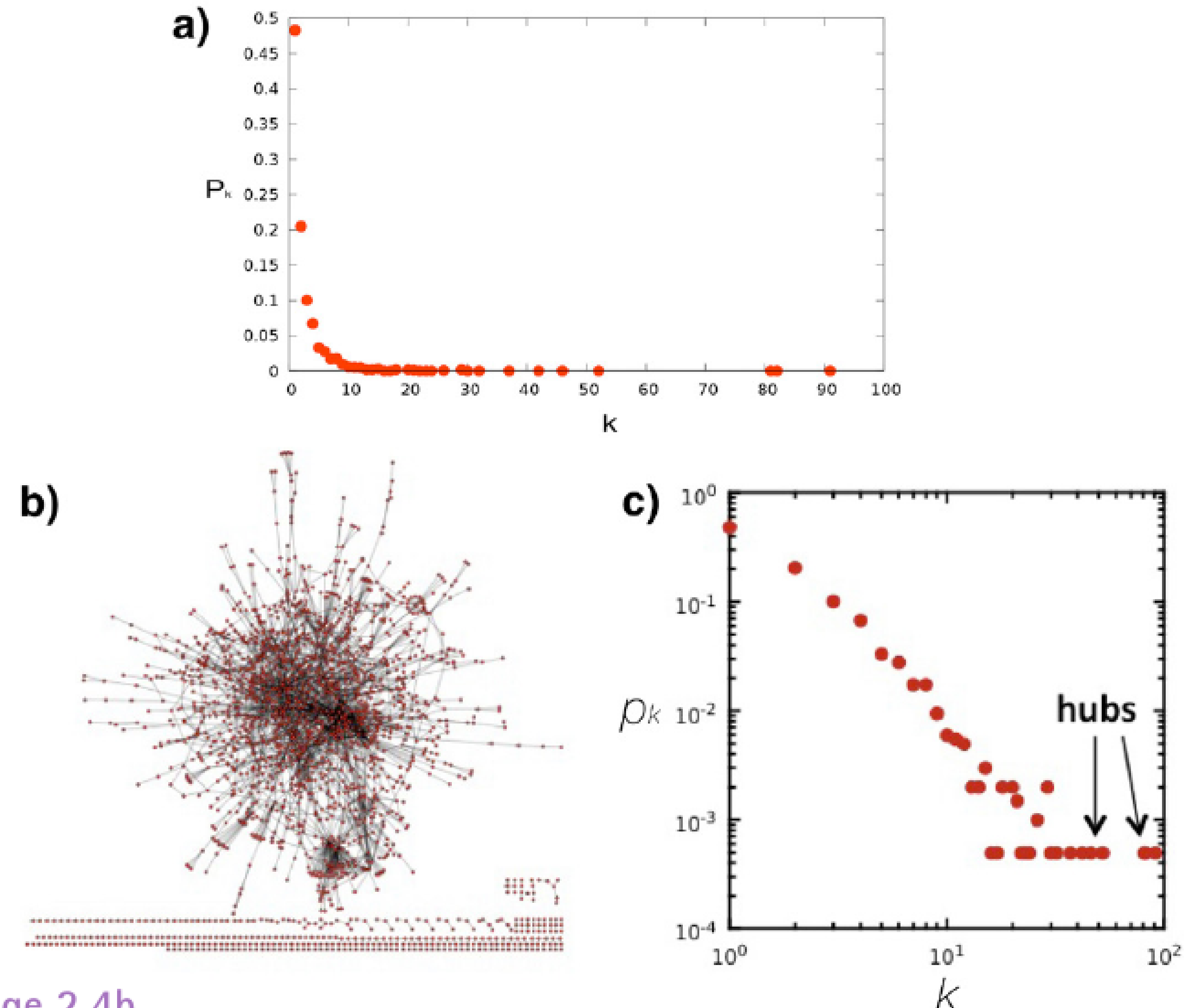
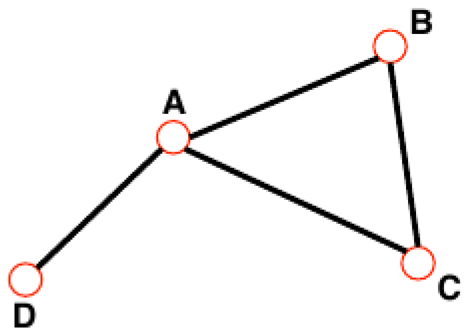


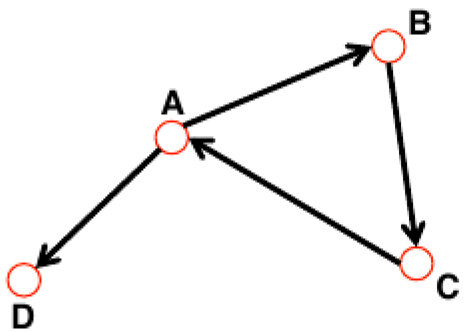
Image 2.4b

Paths



The *distance (shortest path, geodesic path)* between two nodes is defined as the number of edges along the shortest path connecting them.

*If the two nodes are disconnected, the distance is infinity.



In directed graphs each path needs to follow the direction of the arrows.

Thus in a digraph the distance from node A to B (on an AB path) is generally different from the distance from node B to A (on a BCA path).

Paths

N_{ij},number of paths between any two nodes *i* and *j*:

Length n=1: If there is a link between *i* and *j*, then A_{ij}=1 and A_{ij}=0 otherwise.

Length n=2: If there is a path of length two between *i* and *j*, then $A_{ik}A_{kj}$=1, and $A_{ik}A_{kj}$=0 otherwise.
The number of paths of length 2:

$$N_{ij}^{(2)} = \sum_{k=1}^{N} A_{ik}A_{kj} = [A^2]_{ij}$$

Length n: In general, if there is a path of length *n* between *i* and *j*, then $A_{ik}\ldots A_{lj}$=1 and $A_{ik}\ldots A_{lj}$=0 otherwise.
The number of paths of length *n* between *i* and *j* is*

$$N_{ij}^{(n)} = [A^n]_{ij}$$

* *holds for both directed and undirected networks.*

Cycles or loops: closed paths

Paths

Connectivity:

Undirected → Connected: there is a path between every pair of vertices.

Directed:
- Strongly connected: there is a directed path between every pair of vertices.
- Weakly connected: connected after replacing all directed edges with undirected edges.

Diameter: $\boldsymbol{d_{max}}$ the maximum distance between any pair of nodes in the graph.

Average path length/distance, <d>, for a connected graph:

where d_{ij} is the distance from node *i* to node j

$$\langle d \rangle \equiv \frac{1}{2L_{\max}} \sum_{i,j \neq i} d_{ij}$$

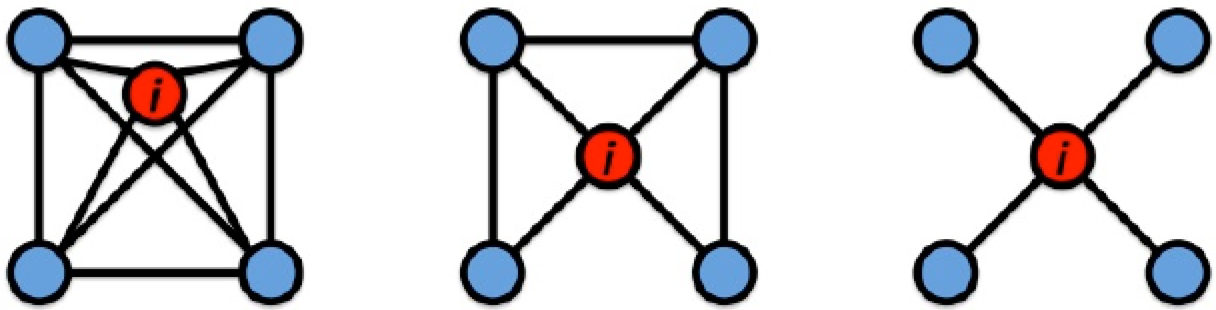
Clustering

- **Clustering coefficient:**

  what fraction of your neighbors are connected?

  - Node i with degree k_i
  - C_i in [0,1]

$$C_i = \frac{2e_i}{k_i(k_i - 1)}$$



$C_i = 1$ $C_i = 1/2$ $C_i = 0$

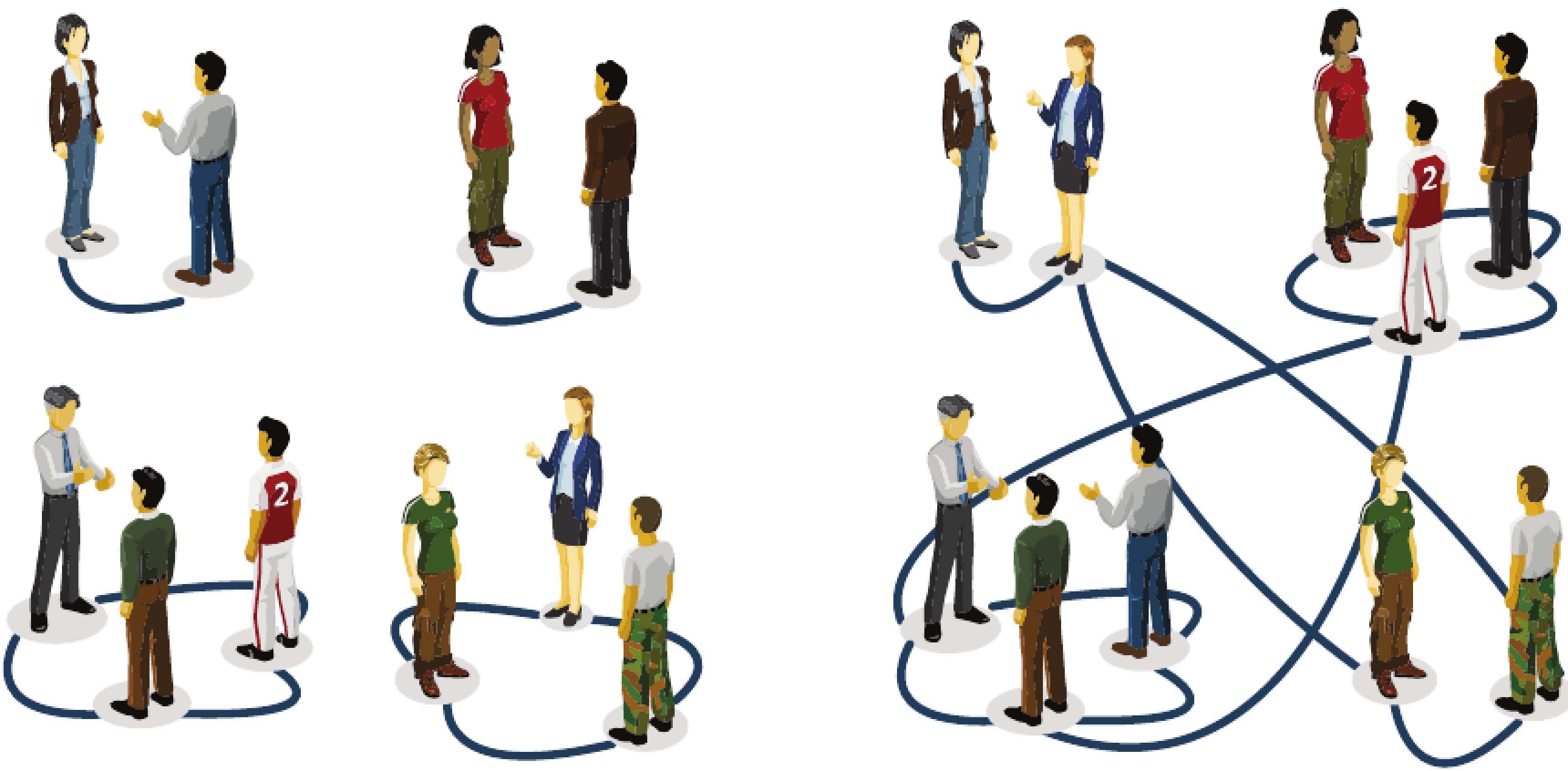
Key measures

Degree distribution: **P(k)**

Path length: **<*d*>**

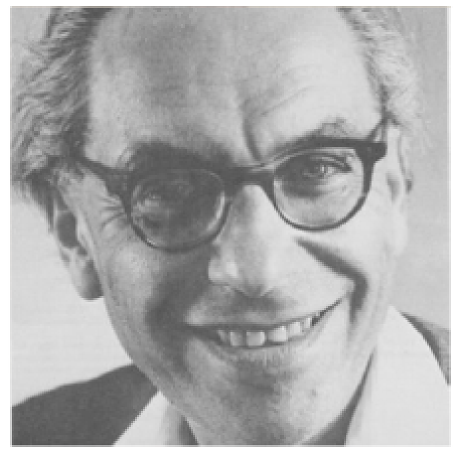
Clustering coefficient: $C_i = \frac{2e_i}{k_i(k_i - 1)}$

Random graphs



Random graphs

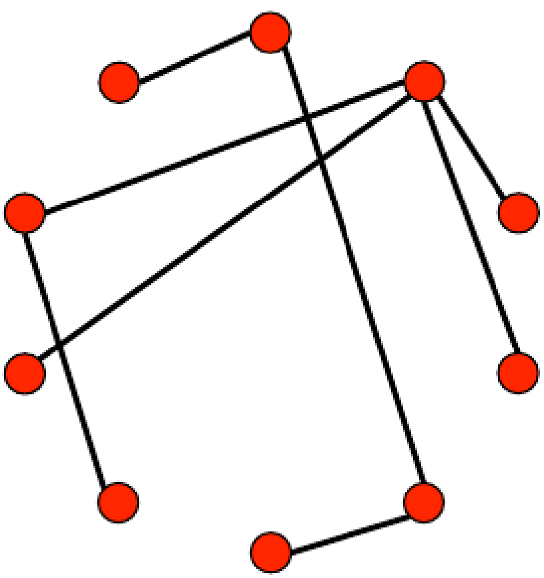
Pál Erdös
(1913-1996)



Alfréd Rényi
(1921-1970)

Erdös-Rényi model (1960)

Connect with probability p



p=1/6 N=10

<k> ~ 1.5

Random graphs

Two versions:

G(*n, M*) model: a graph is chosen uniformly at random from the collection of all graphs which have *n* nodes and *M* edges.

Erdős & Rényi (1959)

G(*n, p*) model: a graph is constructed by connecting nodes randomly. Each edge is included in the graph with probability *p* independent from every other edge.

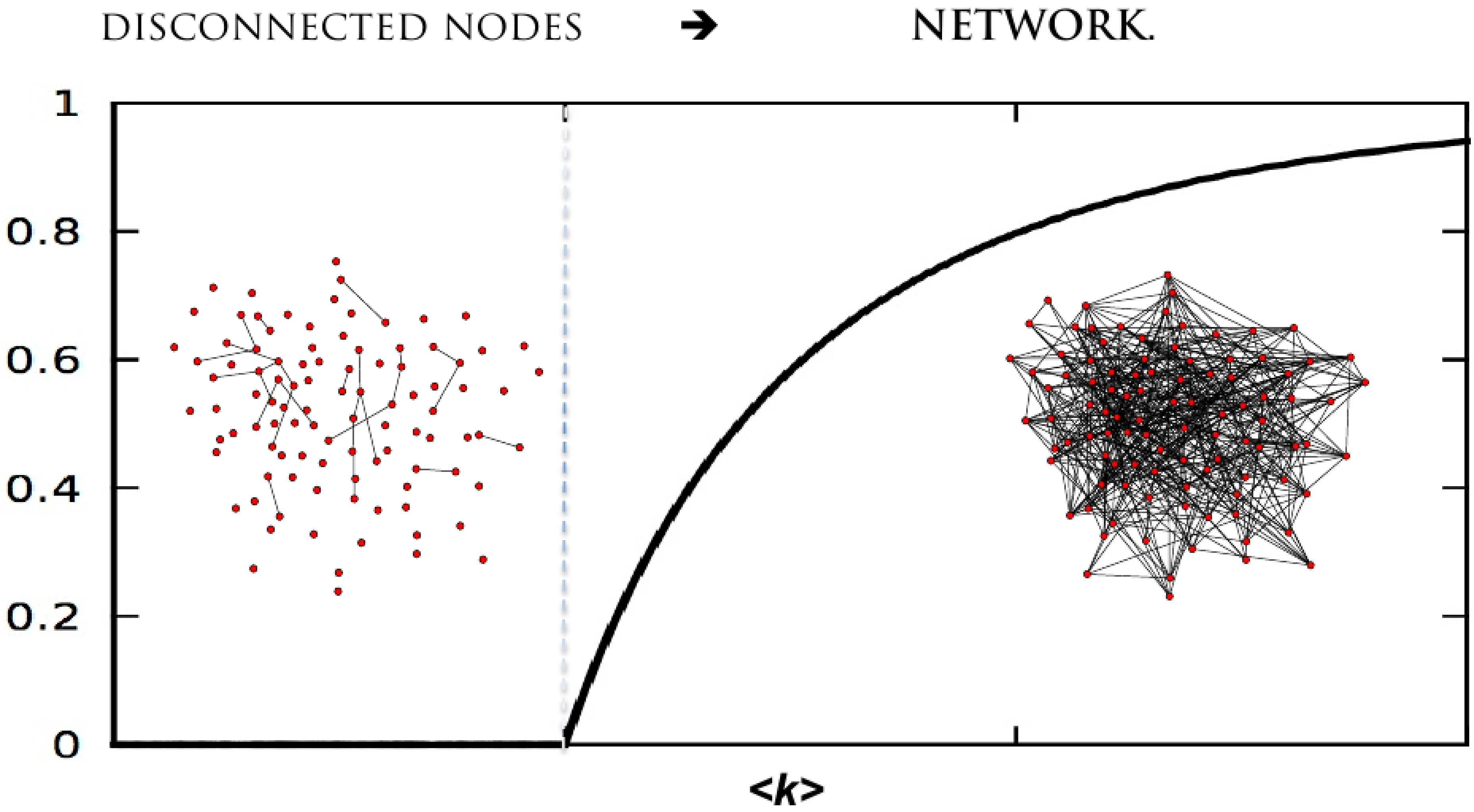
Gilbert (1959)

Microcanonical ensemble

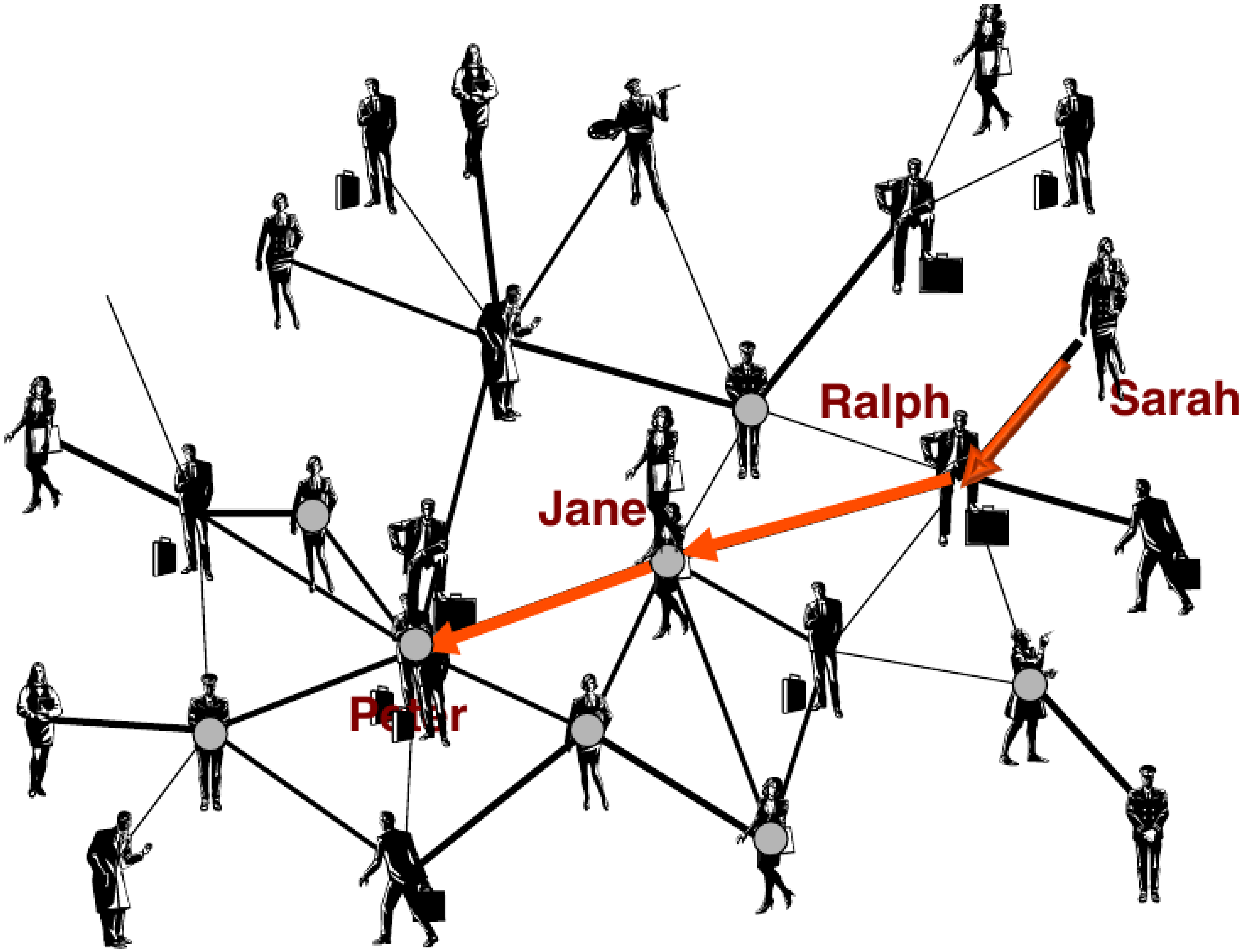
Canonical ensemble

$$P(k) = \binom{n-1}{k} p^k (1-p)^{n-1-k} \quad \rightarrow \frac{(np)^k \mathrm{e}^{-np}}{k!}$$

Random graphs



Six degrees of separation?



Frigyes Karinthy, 1929
Stanley Milgram, 1967

Six degrees of separation?

sport football opinion culture business lifestyle fashion environment tech

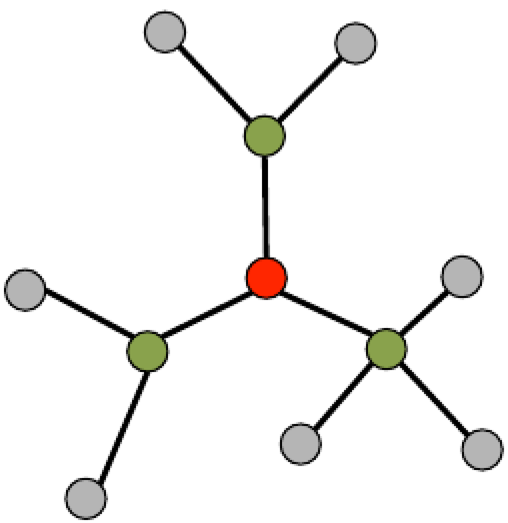
Facebook brings the world three-and-a-bit degrees of separation closer

The social media platform used its friend graph to calculate the degrees separating its 1.6 billion members and found it is as few as 3.57 people

Bringing the world together: Facebook says every person in the world is connected to every other person by an average of three and a half other people. Photograph: Alamy

Six degrees of separation?

Random graphs tend to have a tree-like topology with almost constant node degrees.



- nr. of first neighbors: $N_1 \cong \langle k \rangle$
- nr. of second neighbors: $N_2 \cong \langle k \rangle^2$
- nr. of neighbours at distance d: $N_d \cong \langle k \rangle^d$
- estimate maximum distance:

$$N = 1 + \langle k \rangle + \langle k \rangle^2 + \ldots + \langle k \rangle^{d_{\max}} = \frac{\langle k \rangle^{d_{\max}+1} - 1}{\langle k \rangle - 1} \approx \langle k \rangle^{d_{\max}} \quad \Rightarrow \quad d_{\max} = \frac{\log N}{\log \langle k \rangle}$$

Six degrees of separation?

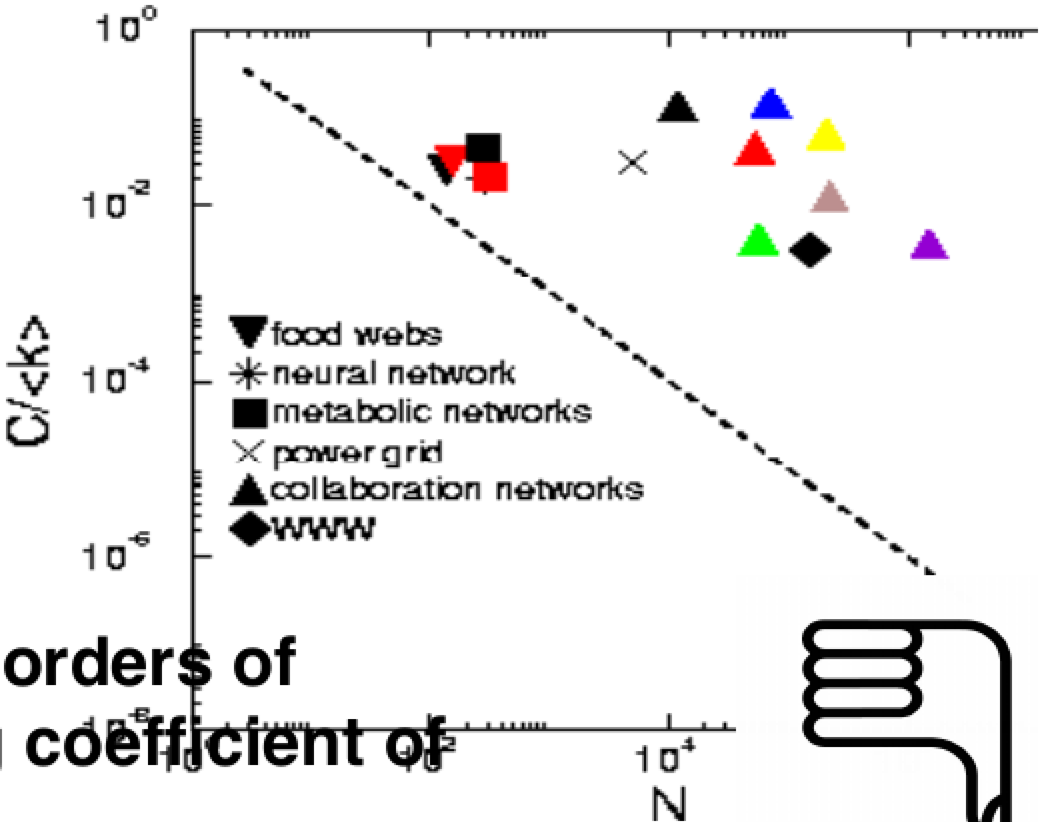
Network Name	N	L	$\langle k \rangle$	$\langle d \rangle$	d_{max}	$\frac{\log N}{\log \langle k \rangle}$
Internet	192,244	609,066	6.34	6.98	26	6.59
WWW	325,729	1,497,134	4.60	11.27	93	8.32
Power Grid	4,941	6,594	2.67	18.99	46	8.66
Mobile Phone Calls	36,595	91,826	2.51	11.72	39	11.42
Email	57,194	103,731	1.81	5.88	18	18.4
Science Collaboration	23,133	186,936	8.08	5.35	15	4.81
Actor Network	212,250	3,054,278	28.78	-	-	-
Citation Network	449,673	4,707,958	10.47	11.21	42	5.55
E Coli Metabolism	1,039	5,802	5.84	2.98	8	4.04
Yeast Protein Interactions	2,018	2,930	2.90	5.61	14	7.14

Clustering

Prediction:

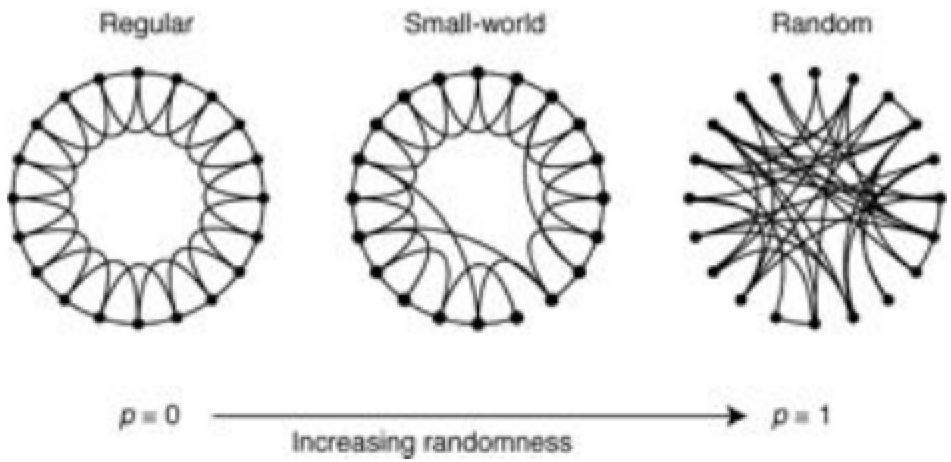
$$C_{rand} = \frac{\langle k \rangle}{N}$$

Data:



C_{rand} underestimates with orders of magnitudes the clustering coefficient of real networks.

Watts-Strogatz model



random network

Low clustering

Small-world property

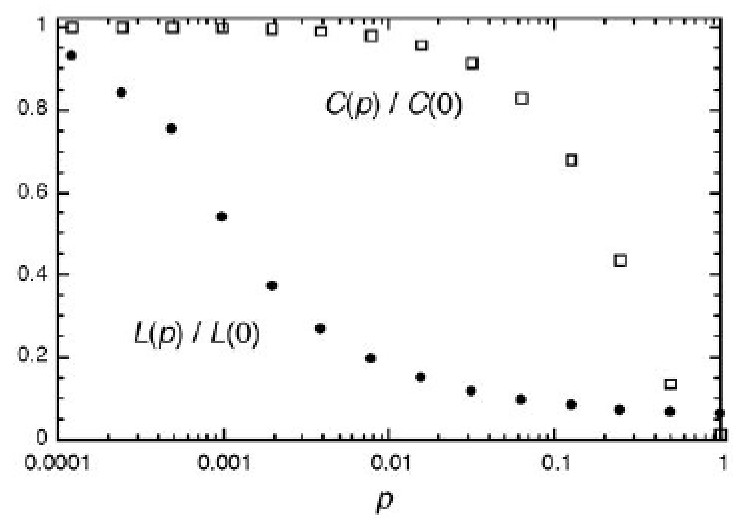
regular ring lattice

High clustering

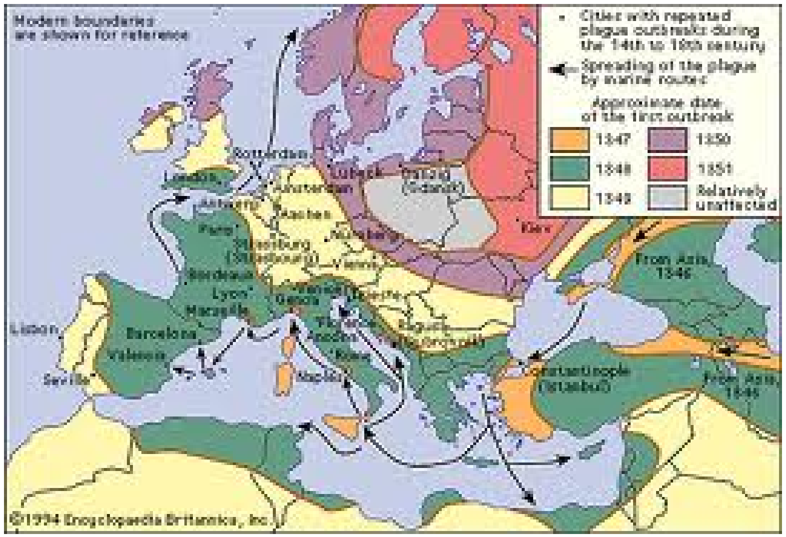
High average path length

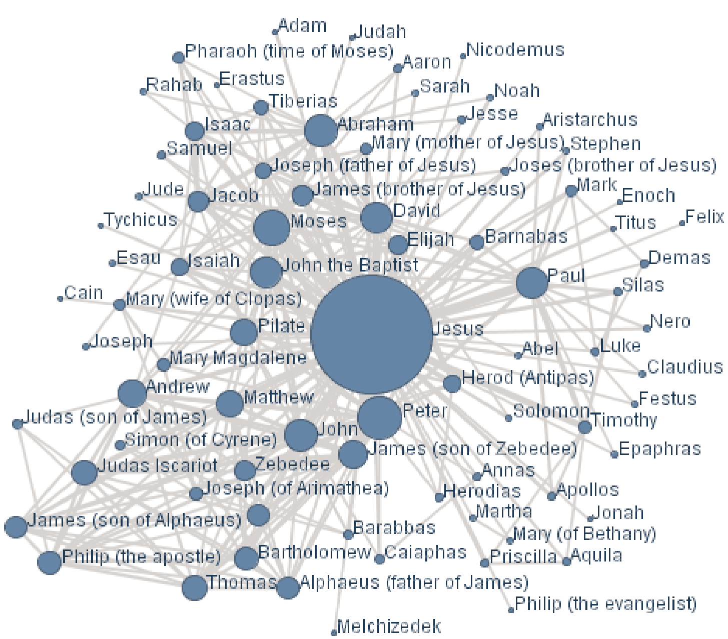
small worlds even with small concentrations of shortcuts

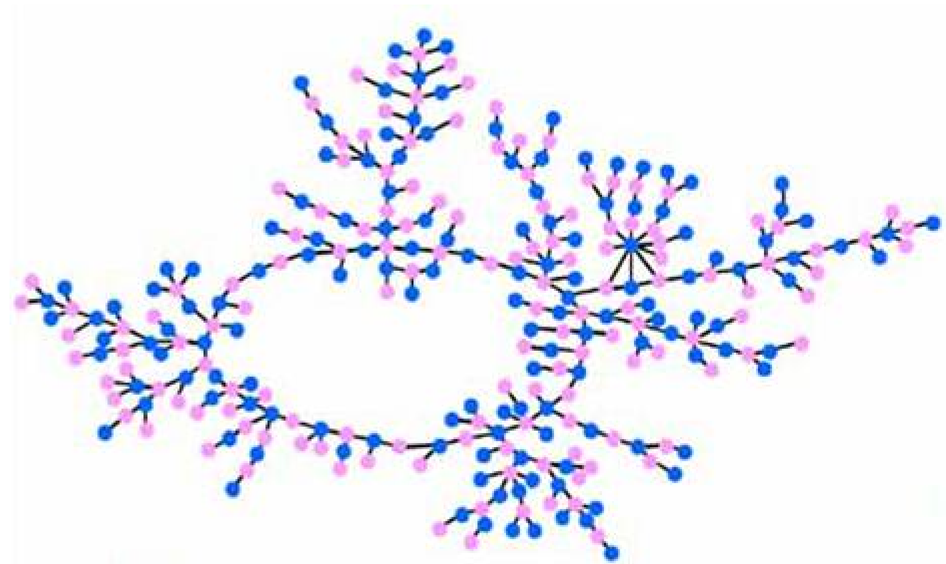
Collective dynamics of 'small-world' networks, D. J. Watts and S. H. Strogatz, Nature 393, 440-442 (1998)



Epidemics

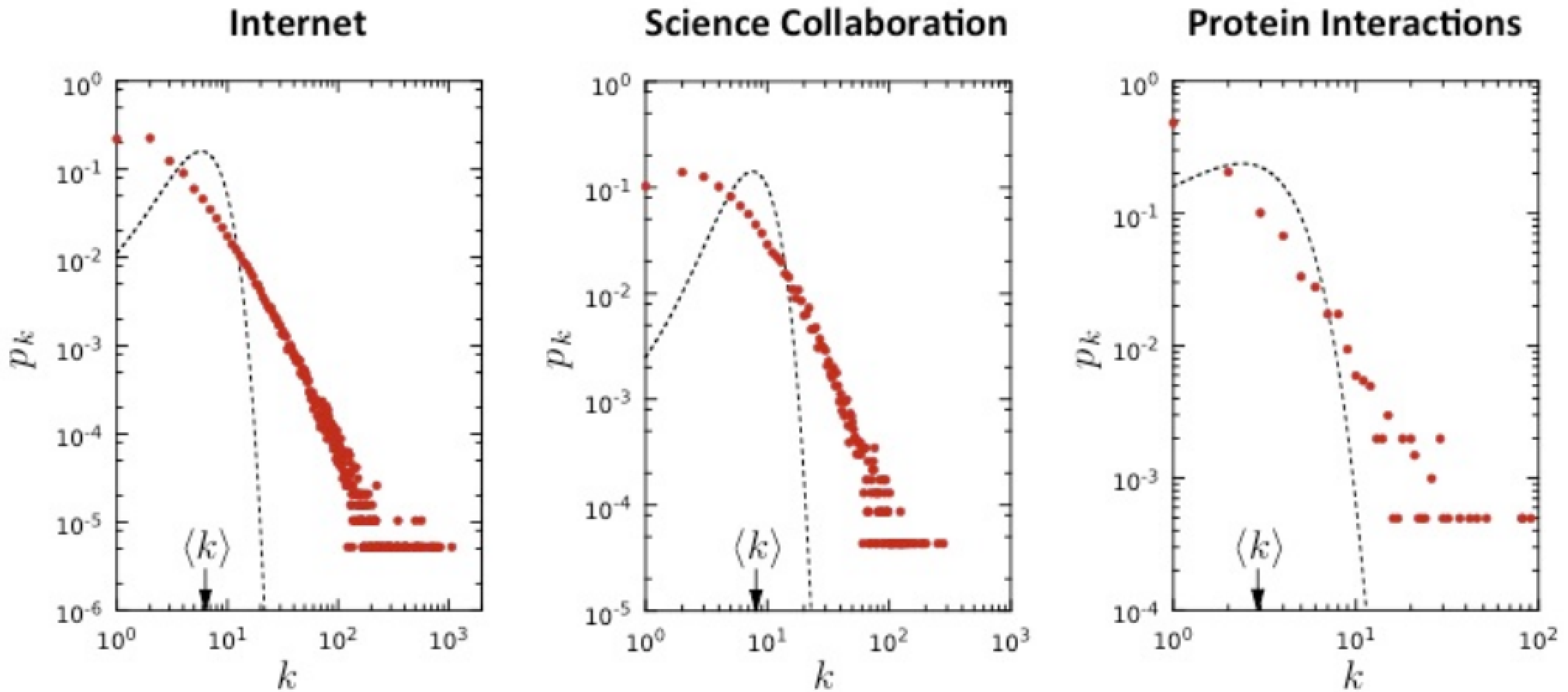



Adam
Judah
Pharaoh (time of Moses)
Aaron
Nicodemus
Rahab
Erastus
Sarah
Noah
Tiberias
Isaac
Abraham
Jesse
Aristarchus
Samuel
Mary (mother of Jesus)
Stephen
Joseph (father of Jesus)
Joses (brother of Jesus)
Jude
Jacob
James (brother of Jesus)
Mark
Enoch
Tychicus
Moses
David
Titus
Felix
Elijah
Barnabas
Esau
Isaiah
John the Baptist
Demas
Paul
Silas
Cain
Mary (wife of Clopas)
Pilate
Jesus
Nero
Joseph
Abel
Luke
Mary Magdalene
Claudius
Herod (Antipas)
Andrew
Matthew
Festus
Peter
Solomon
Timothy
Judas (son of James)
John
Simon (of Cyrene)
James (son of Zebedee)
Epaphras
Judas Iscariot
Zebedee
Annas
Joseph (of Arimathea)
Herodias
Apollos
Martha
Jonah
James (son of Alphaeus)
Barabbas
Mary (of Bethany)
Philip (the apostle)
Bartholomew
Caiaphas
Priscilla
Aquila
Thomas
Alphaeus (father of James)
Philip (the evangelist)
Melchizedek

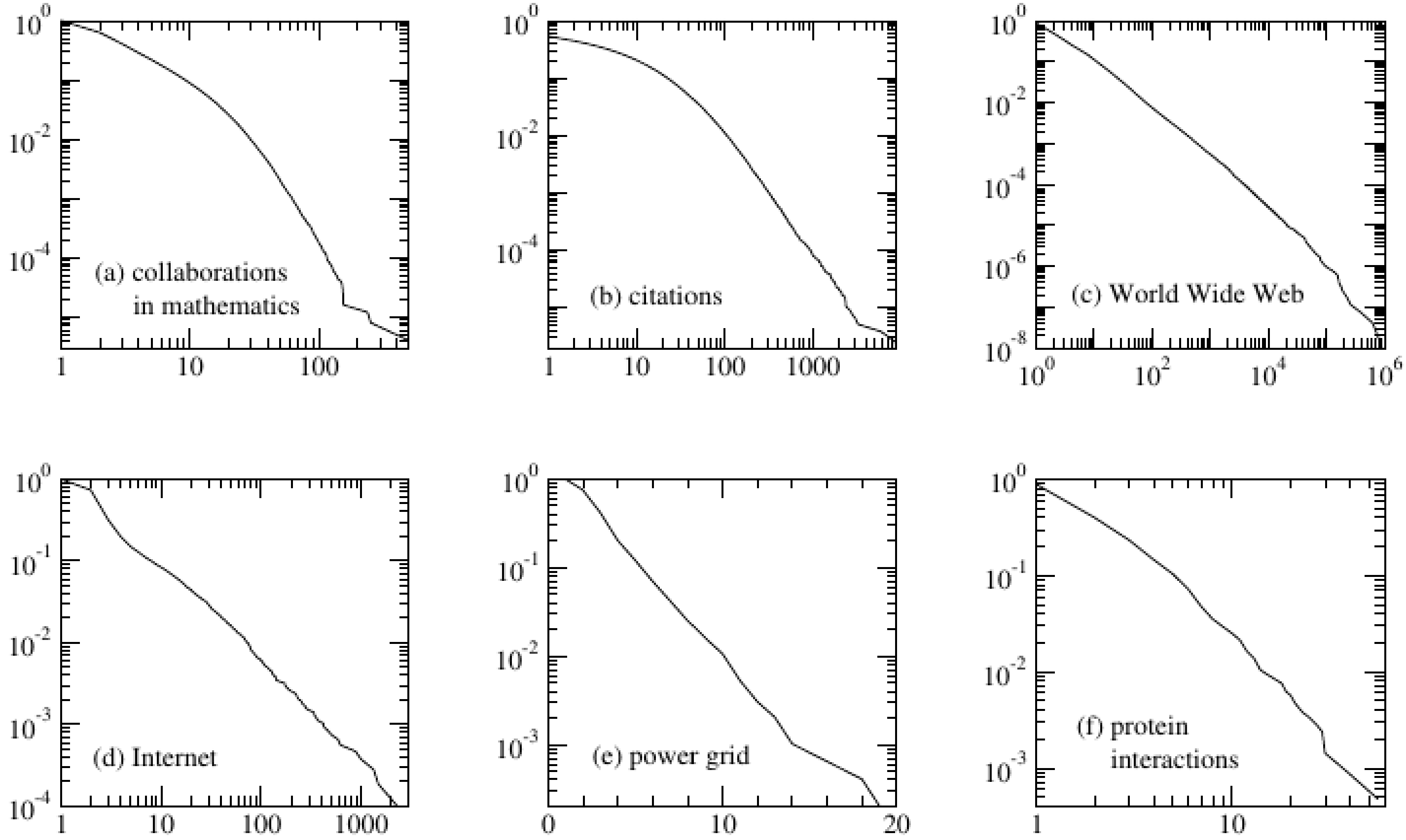


Scale-free networks

$$P(k) = e^{-<k>} \frac{<k>^k}{k!}$$



Scale-free networks

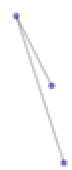


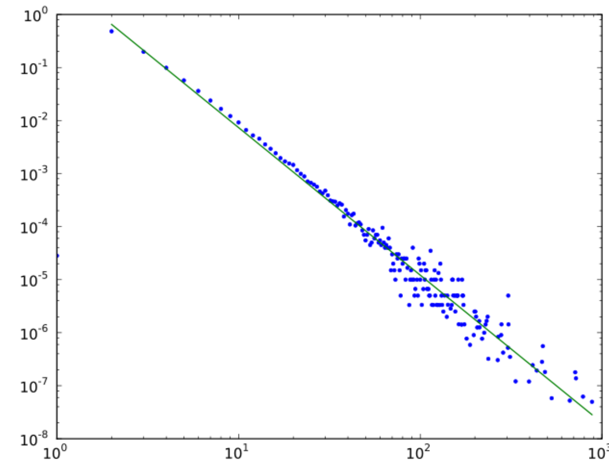
$$p(k) \sim k^{-\gamma}$$

Scale-free networks

Network	Size	$\langle k \rangle$	κ	γ_{out}	γ_{in}
WWW	325 729	4.51	900	2.45	2.1
WWW	4×10^7	7		2.38	2.1
WWW	2×10^8	7.5	4000	2.72	2.1
WWW, site	260 000				1.94
Internet, domain*	3015–4389	3.42–3.76	30–40	2.1–2.2	2.1–2.2
Internet, router*	3888	2.57	30	2.48	2.48
Internet, router*	150 000	2.66	60	2.4	2.4
Movie actors*	212 250	28.78	900	2.3	2.3
Co-authors, SPIRES*	56 627	173	1100	1.2	1.2
Co-authors, neuro.*	209 293	11.54	400	2.1	2.1
Co-authors, math.*	70 975	3.9	120	2.5	2.5
Sexual contacts*	2810			3.4	3.4
Metabolic, *E. coli*	778	7.4	110	2.2	2.2
Protein, *S. cerev.**	1870	2.39		2.4	2.4
Ythan estuary*	134	8.7	35	1.05	1.05
Silwood Park*	154	4.75	27	1.13	1.13
Citation	783 339	8.57			3
Phone call	53×10^6	3.16		2.1	2.1
Words, co-occurrence*	460 902	70.13		2.7	2.7
Words, synonyms*	22 311	13.48		2.8	2.8

Barabási-Albert model





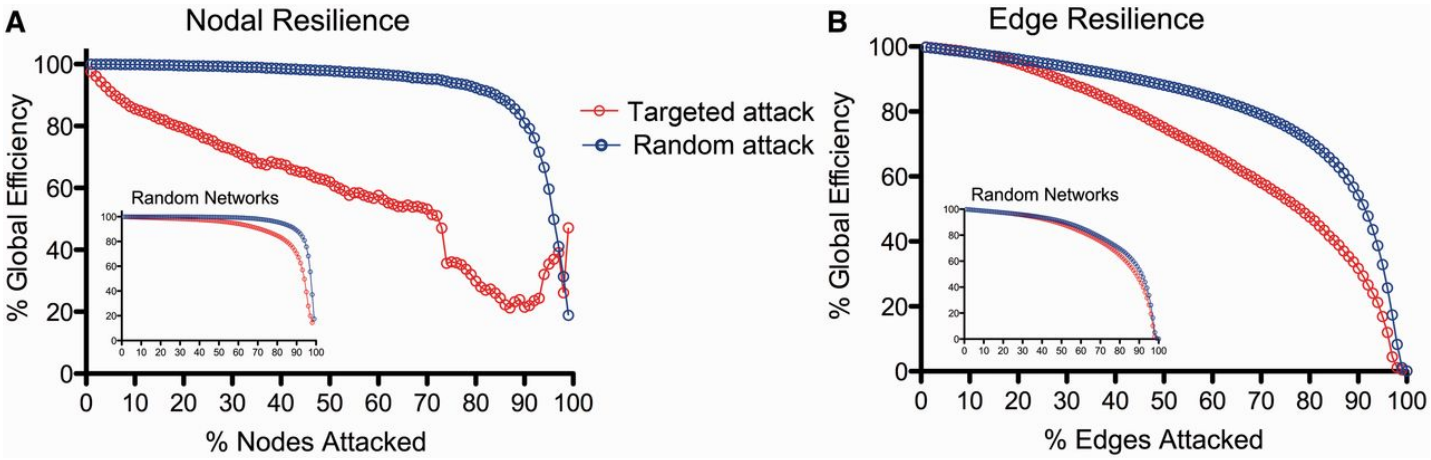
Emergence of Scaling in Random Networks

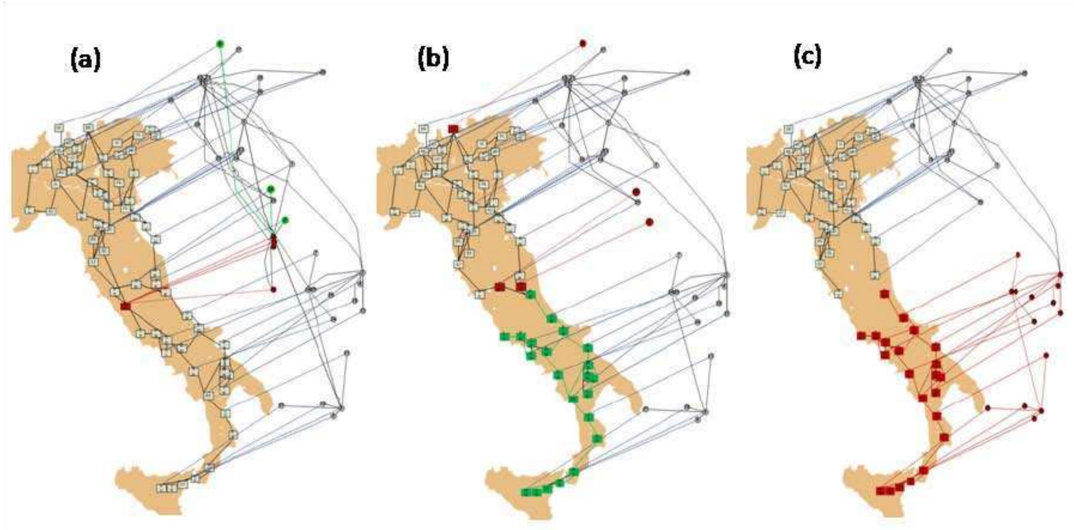
Albert-László Barabási* and Réka Albert

$$p(k) \sim k^{-3}$$

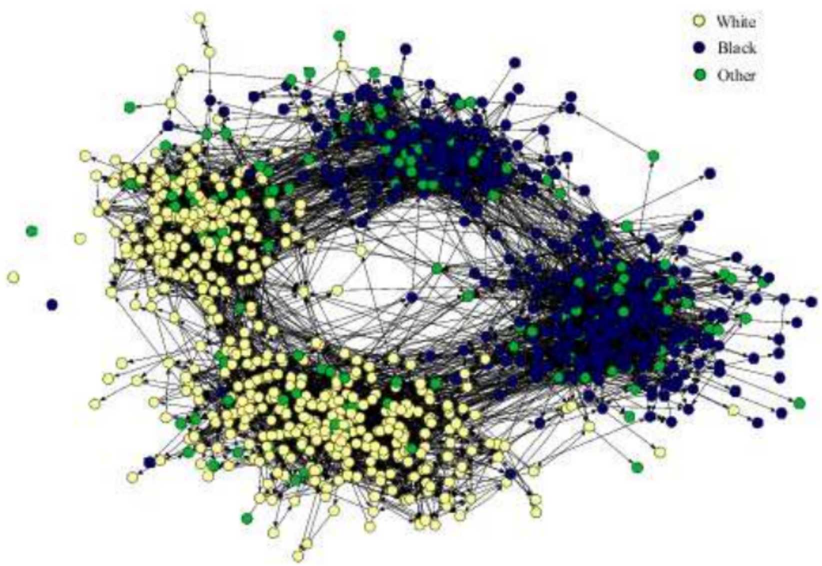
- m_0 initial nodes
- Each new node appears with $m \leq m$ new links
- Probability of attachment to i is $p_i = \frac{k_i}{\sum_j k_j}$

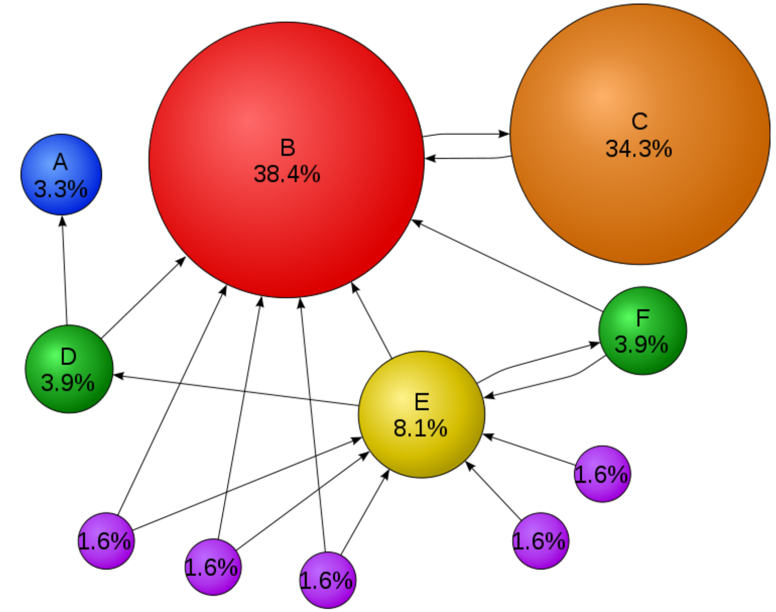
Network robustness

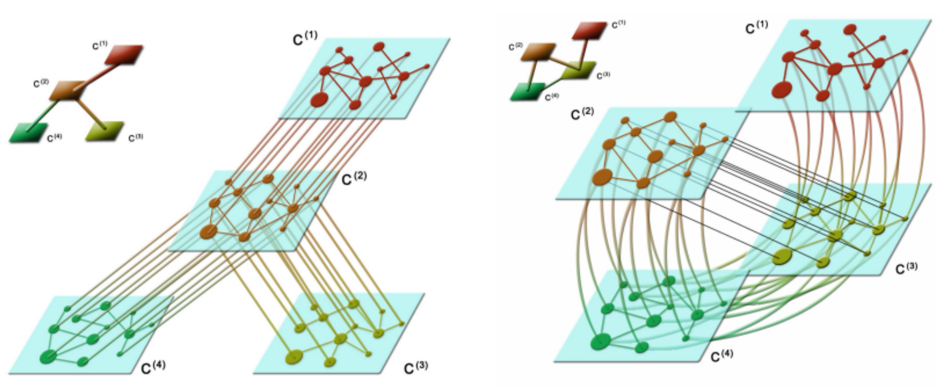




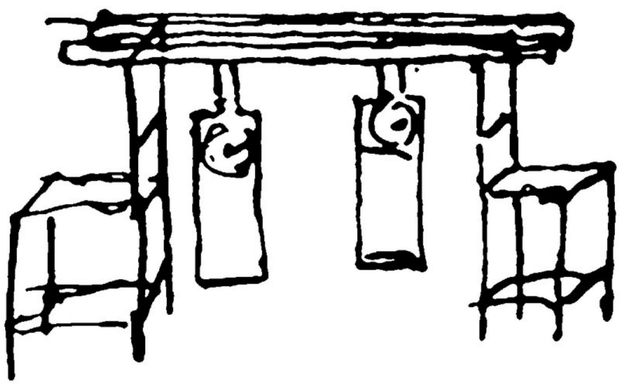
Structure of networks

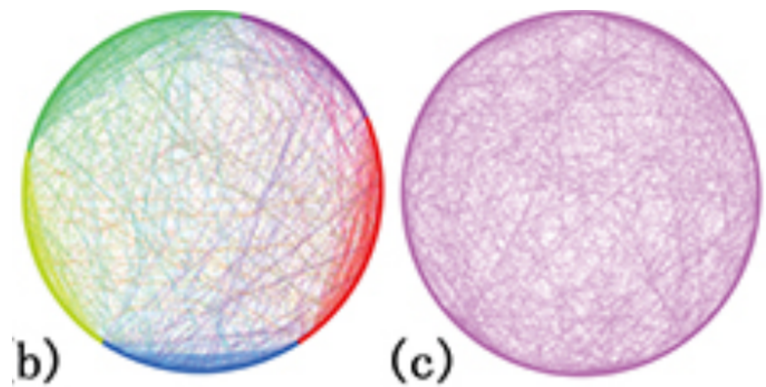


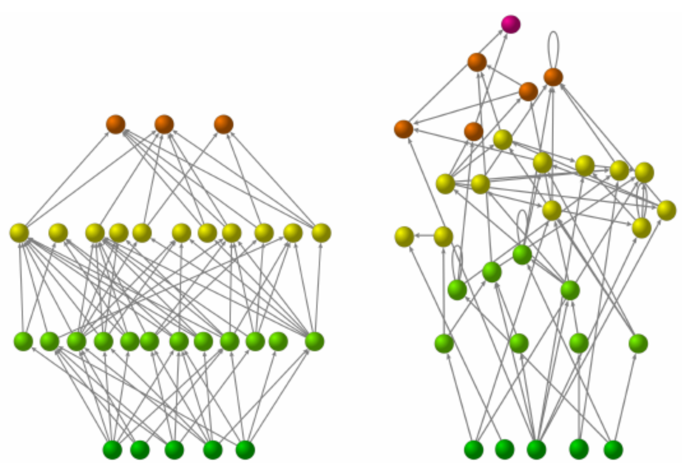


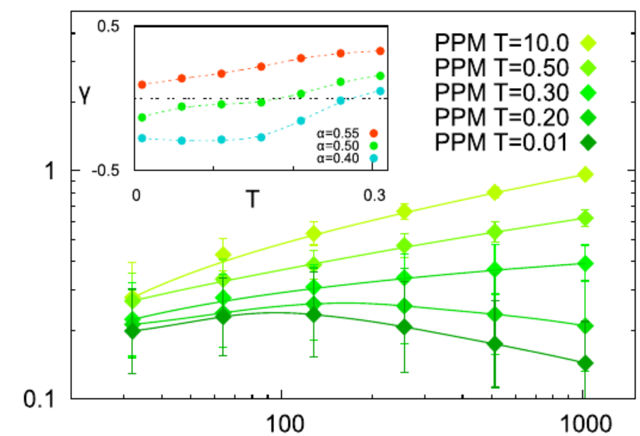


Dynamical processes on networks









Trophic coherence determines food-web stability

Samuel Johnson[a,1,2], Virginia Domínguez-García[b,1], Luca Donetti[c], and Miguel A. Muñoz[b]

Reviews

[R1] *Statistical mechanics of complex networks,* R. Albert, and A.-L. Barabási , *Rev. Mod. Phys.* **74**, 47-97 (2002)

[R2] *The structure and function of complex networks*, M. E. J. Newman, *SIAM Review* **45**, 167–256, 2003

[R3] *Critical phenomena in complex networks*, S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, *Rev. Mod. Phys.* **80**, 1275 (2008)

Books

[B1] *Evolution of Networks: From Biological Nets to the Internet and WWW*, S. N. Dorogovtsev and J. F. F. Mendes, Oxford University Press (2003)

[B2] *Networks: An Introduction*, M. E. J. Newman, Oxford University Press (2010)

[B3] *Dynamical Processes on Complex Networks*, A. Barrat, M. Barthélémy, and A. Vespignani, Cambridge University Press (2012)

Popular science

[P1] A.-L. Barabási. *Linked: How Everything Is Connected to Everything Else and What It Means*

[P2] D. J. Wats. *Small Worlds: The Dynamics of Networks between Order and Randomness*

[P3] R. Solé. *Redes Complejas. Del genoma a Internet*

Thank you for your attention!
…and to A.-L. Barabási for making material available online