NETWORK MODELS

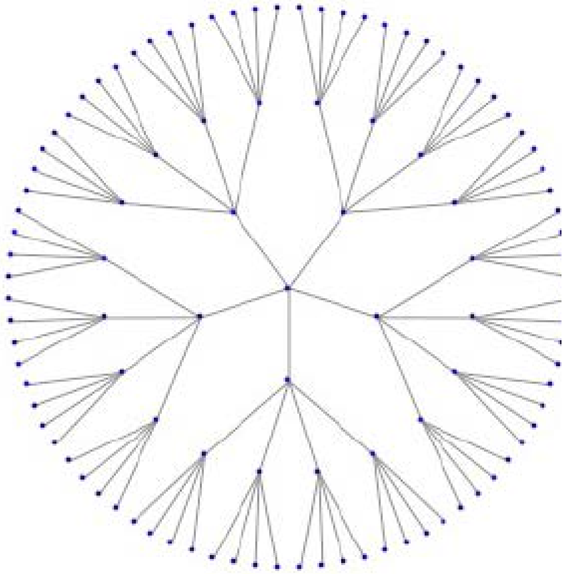
Carlo PICCARDI

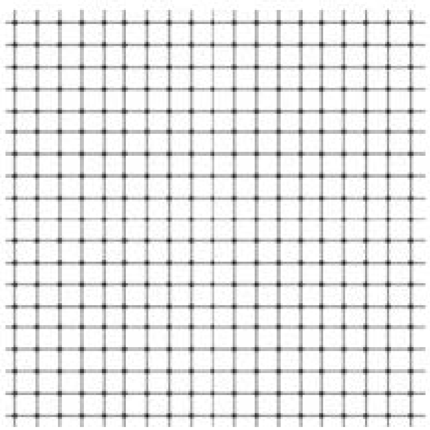
DEIB - Department of Electronics, Information and Bioengineering
Politecnico di Milano, Italy

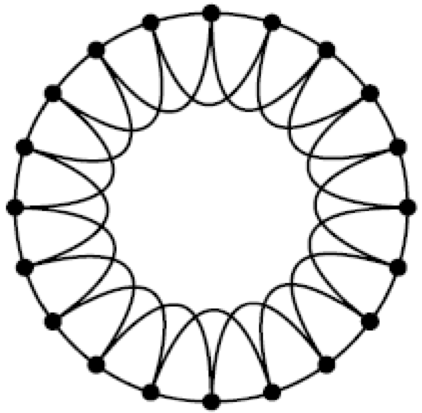
email carlo.piccardi@polimi.it
http://home.deib.polimi.it/piccardi

"REGULAR" NETWORKS

Trees and lattices...

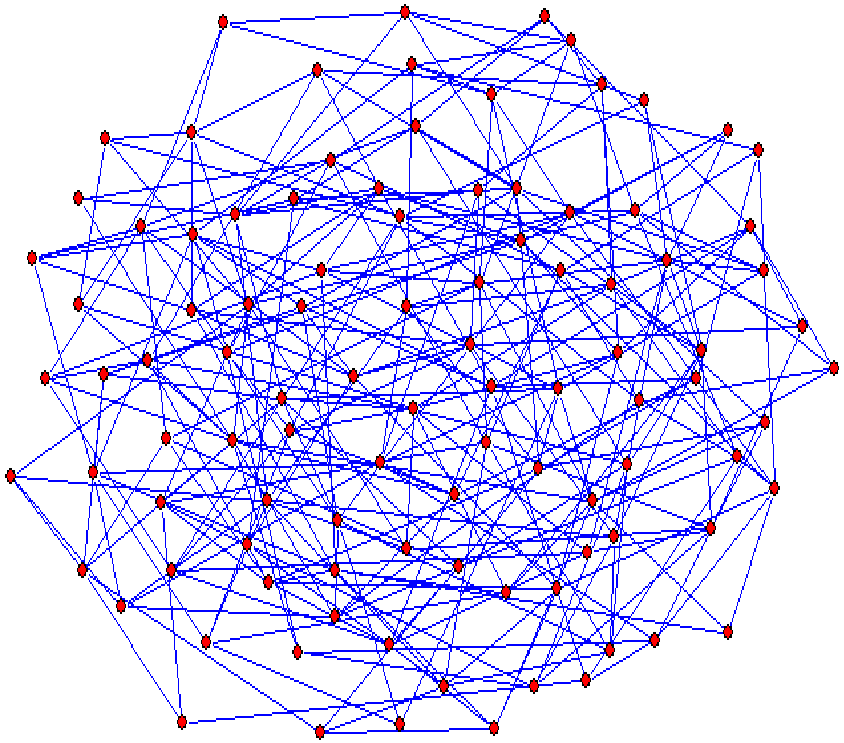




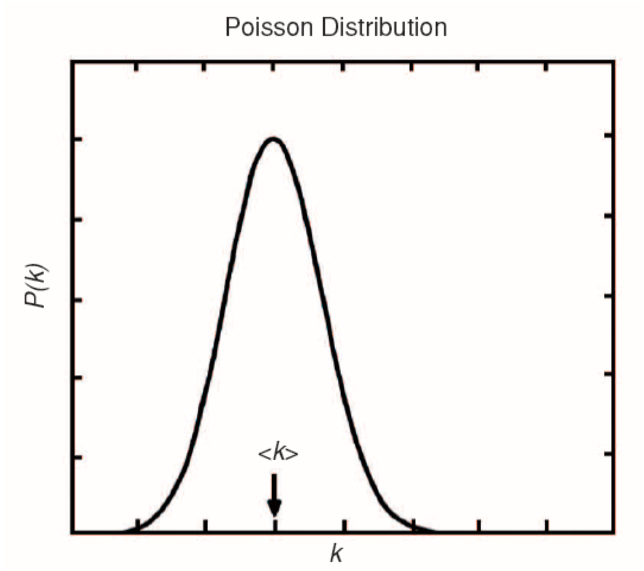


... are very rarely representative of real-world networks.

"RANDOM" (Erdös-Rényi) NETWORKS



This is a random (Erdös-Rényi) network, obtained by letting $N = 100$ and connecting $L = 300$ randomly extracted pairs (hence $< k >= 2 \times 300 / 100 = 6$).



For large N, the degree is Poisson-distributed with $< k >= 2L / N$:

- the "typical" scale of node degree is $k_i =< k >$
- node degrees have small fluctuations around $< k >$
- the network is "almost homogeneous"

More on Erdös-Rényi networks...

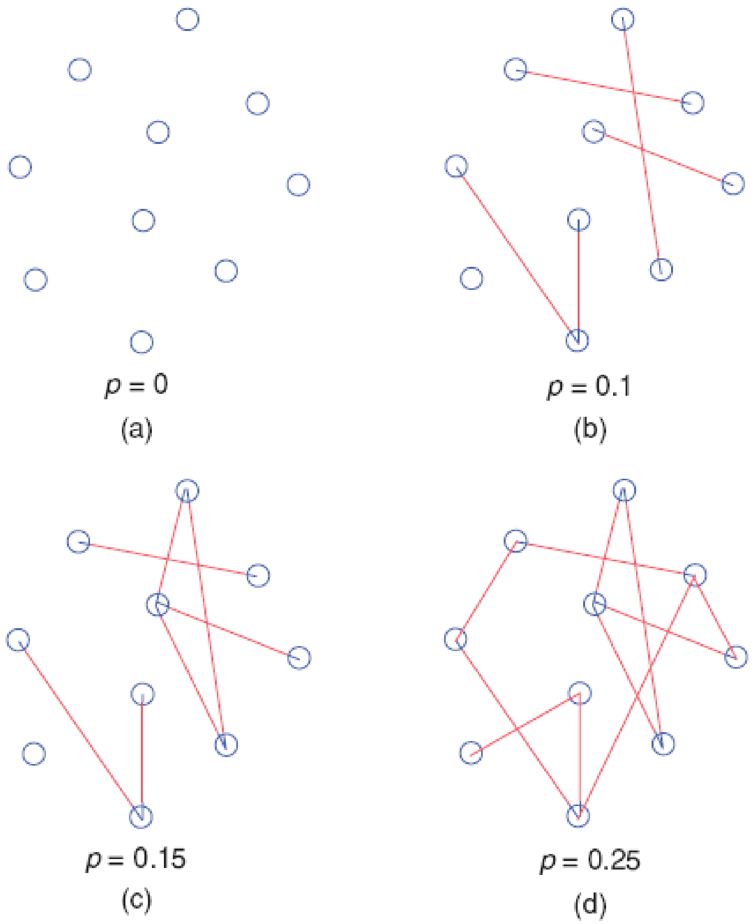
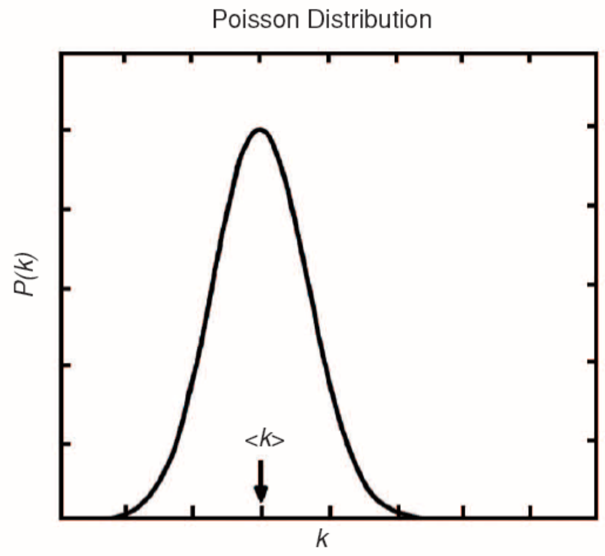


Figure 6. Evolution of a random graph. Given 10 isolated nodes in (a), one connects every pair of nodes with probability (b) $p = 0.1$, (c) $p = 0.15$ and (d) $p = 0.25$, respectively.

An alternative procedure:

Start from a graph with N nodes and no links, and connect each pair i, j with a given probability p.

Some properties (for $N \to \infty$):



- The degree is Poisson distributed, with $<k> = p(N-1)$:

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

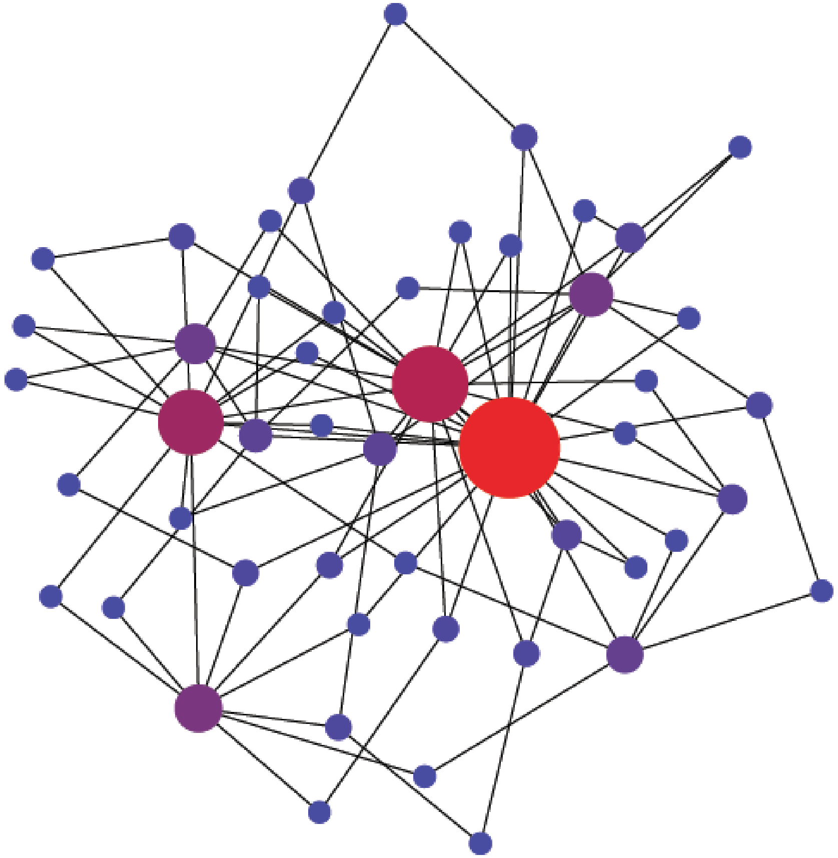
- The network has a giant component if $<k>$ is larger than 1 ($p > 1/N$).

- The average distance $d \cong \log N / \log <k>$ grows "slowly" with N (for fixed $<k>$ - "small-world" effect) $\to$ "Large" networks (=large N) have a relatively small average distance.

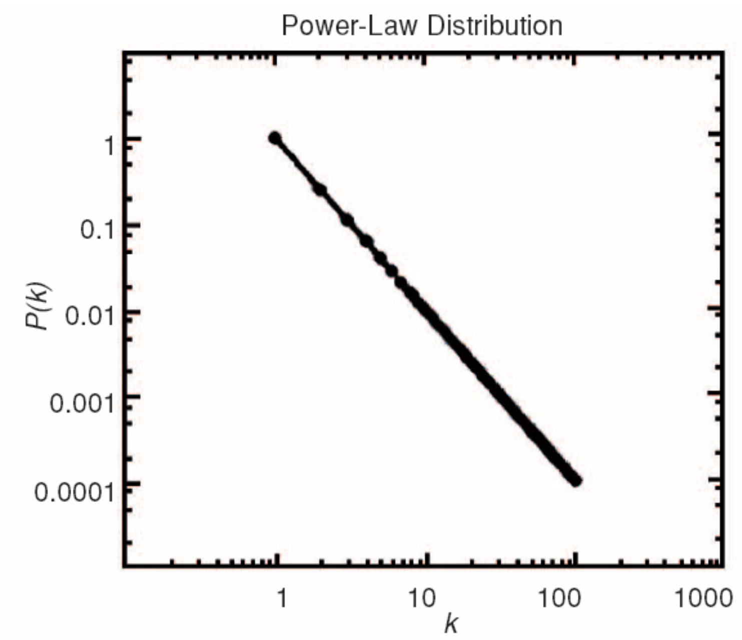
- The clustering coefficient $C = p \cong <k> / N$ tends to 0 as N grows (for fixed $<k>$) $\to$ "Large" networks have vanishing clustering.

SCALE-FREE (Barabási-Albert) NETWORKS

This is a scale-free network, obtained by adding one node at a time, and connecting it preferentially (=with higher probability) to nodes with higher degree (Barabási-Albert algorithm).



The network contains few very connected nodes ("hubs") and many scarcely connected nodes.

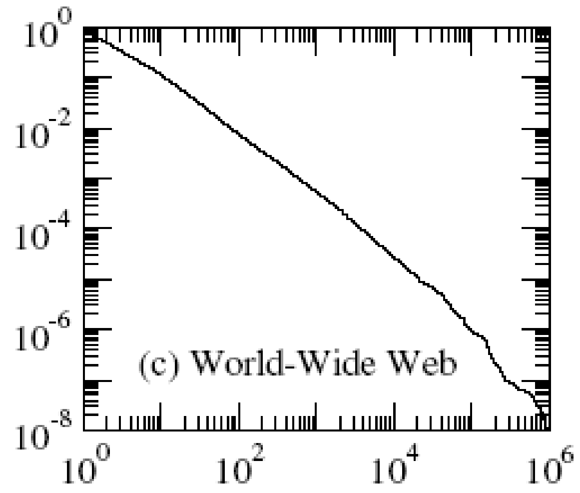


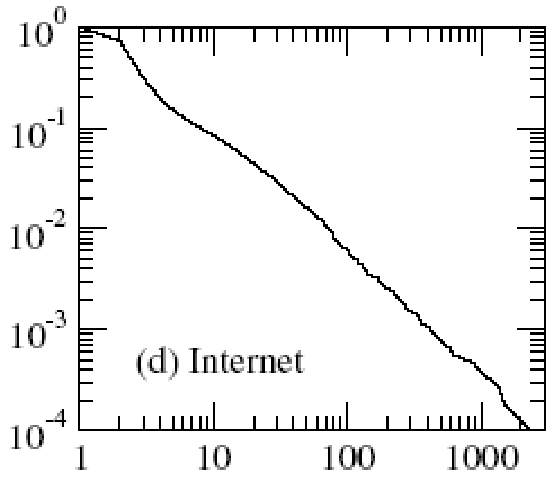
For large N, the degree distribution is a power-law function $P(k) \approx k^{-\alpha}$:

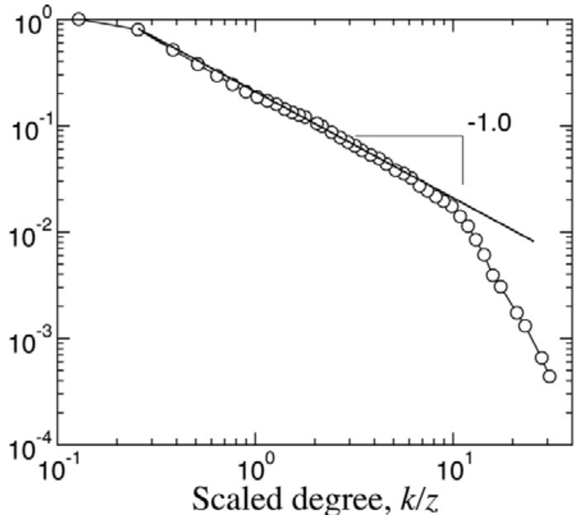
⇨ node degrees have large fluctuations around $< k >$: there is no "typical" scale of node degree

⇨ the network is strongly heterogeneous

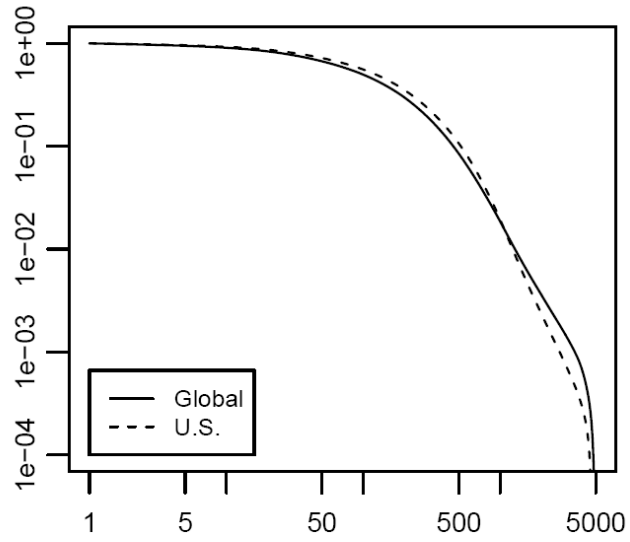
Some examples of (cumulative) degree distribution:







the air transportation network



Facebook (721 million nodes, May 2011)

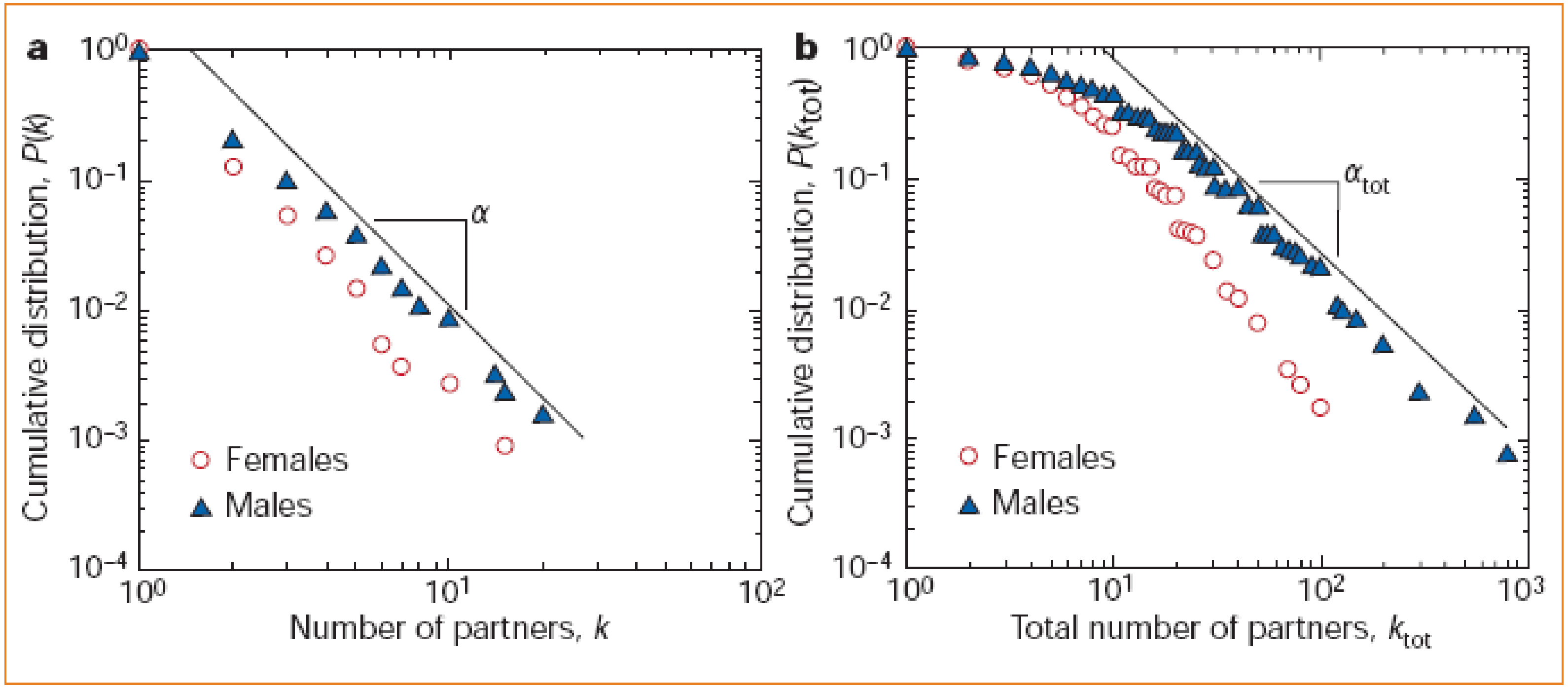
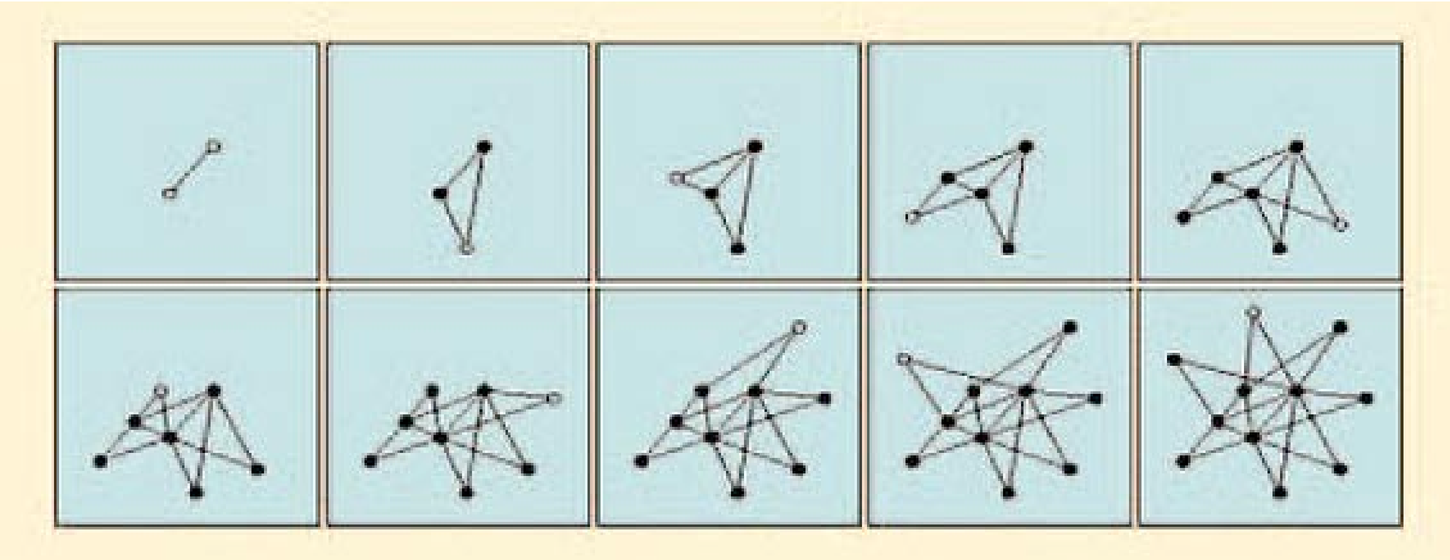


Figure 2 Scale-free distribution of the number of sexual partners for females and males. **a,** Distribution of number of partners, k, in the previous 12 months. Note the larger average number of partners for male respondents: this difference may be due to 'measurement bias' — social expectations may lead males to inflate their reported number of sexual partners. Note that the distributions are both linear, indicating scale-free power-law behaviour. Moreover, the two curves are roughly parallel, indicating similar scaling exponents. For females, $\alpha = 2.54 \pm 0.2$ in the range $k > 4$, and for males, $\alpha = 2.31 \pm 0.2$ in the range $k > 5$. **b,** Distribution of the total number of partners k_{tot} over respondents' entire lifetimes. For females, $\alpha_{tot} = 2.1 \pm 0.3$ in the range $k_{tot} > 20$, and for males, $\alpha_{tot} = 1.6 \pm 0.3$ in the range $20 < k_{tot} < 400$. Estimates for females and males agree within statistical uncertainty.

More on "scale-free" networks...

Barabási-Albert algorithm (1999) is inspired by the WWW growth:

- initialization: start with m_0 nodes (arbitrarily connected)
- growth: at each step, add a new node i with $m \le m_0$ new links connecting i to m existing nodes.
- preferential attachment: attach the new m links preferentially (=with higher probability) to nodes with high degree (*"rich get richer"*): that is, let the probability of connecting a new node i to an existing node j be $k_j / \Sigma_h k_h$.



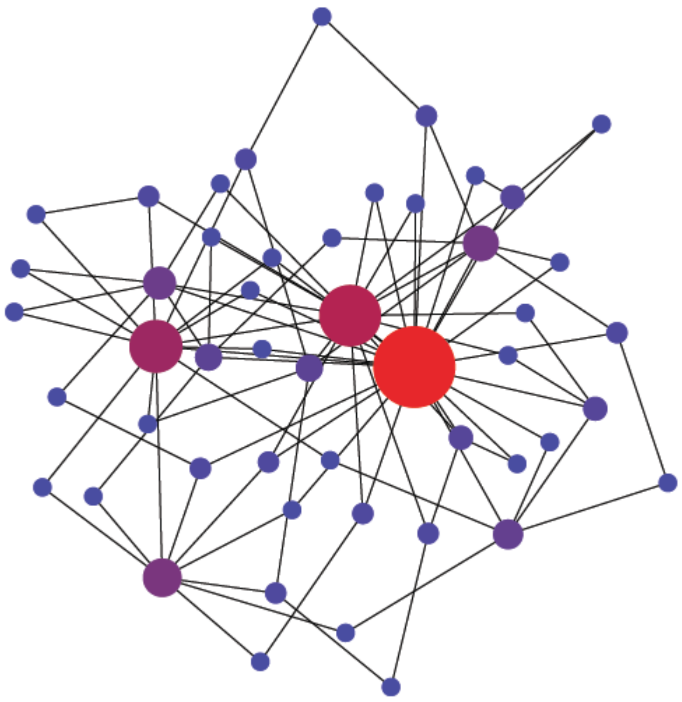
Then for $N \to \infty$:

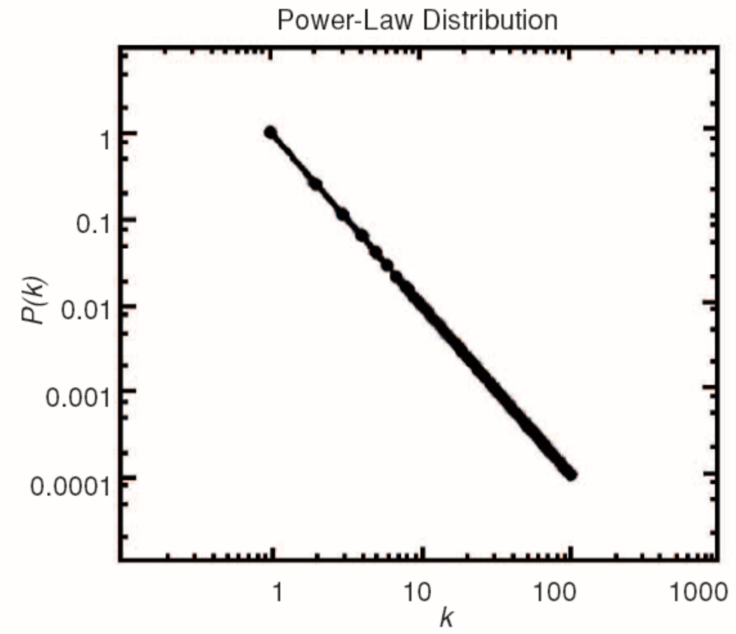
- the average degree tends to $<k>=2m$ and the degree distribution to the power-law $P(k) \approx k^{-3}$

- $<k^2>$ and thus the variance $\sigma^2 = <k^2> - <k>^2$ diverge ($P(k)$ has a "heavy tail")

- the average distance tends to $d \approx \log N / \log \log N$ ("small-world" effect)

- the clustering coefficient C vanishes with $c \approx (\log N)^2 / N \to 0$





Sketch of the proof (the "continuum approach"):

- After t steps, the network has $m_0 + t$ nodes and $\cong mt$ links.
- At each step t, the prob. for node i to be selected by one of new links is $k_i / \sum_j k_j$.
- Approximating the degree k_i with a continuous variable, its increase rate is
$$\frac{dk_i}{dt} = m \frac{k_i}{\sum_j k_j} = \frac{k_i}{2t}$$
because $(\sum_j k_j)/2 = mt$ is the number of links.
- Solving the differential equation for a node inserted at time t_i with $k_i(t_i) = m$:
$$k_i(t) = m \left(\frac{t}{t_i}\right)^{0.5}$$
Notice: All nodes evolve in the same way; older nodes (=smaller t_i) have larger degree.
- The cumulative degree distribution
$$\bar{P}(k) = prob(k_i \geq k) = prob(t_i \leq \frac{m^2 t}{k^2})$$
- Nodes are inserted uniformly in time, thus the fraction of nodes inserted before $m^2 t / k^2$ is
$$\bar{P}(k) = prob\left(t_i \leq \frac{m^2 t}{k^2}\right) = \frac{m^2 t}{k^2} / t = \frac{m^2}{k^2}$$
and the degree distribution is
$$P(k) = -\frac{d\bar{P}(k)}{dk} = 2m^2 k^{-3}$$

"SMALL-WORLD" (Watts-Strogatz) NETWORKS

In typical real-world networks, the average distance $d = < d_{ij} >$ turns out to be surprisingly small.

Empirically, it is observed that

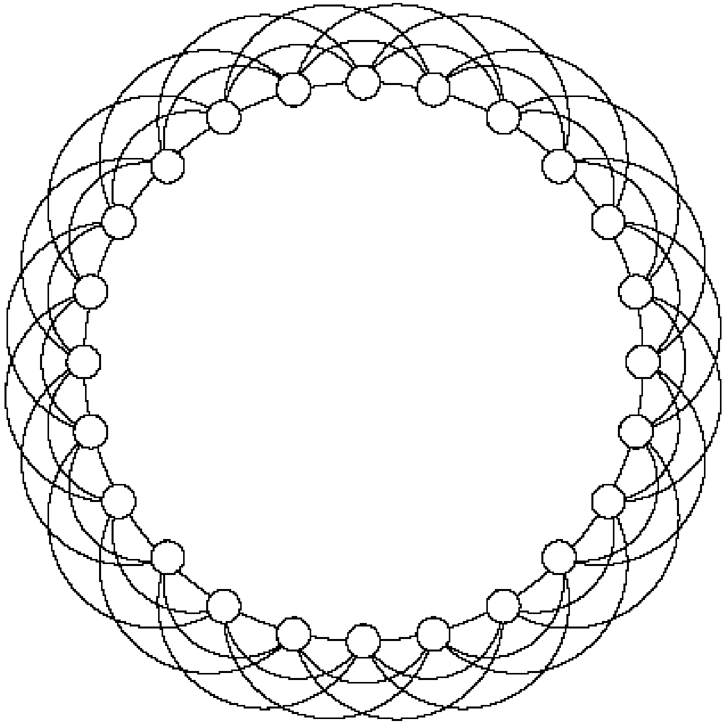
$$d \approx \log N$$

i.e., d increases "slowly" with N ("small-world" effect).

	Network	Type	n	m	z	ℓ
Social	film actors	undirected	449 913	25 516 482	113.43	3.48
	company directors	undirected	7 673	55 392	14.44	4.60
	math coauthorship	undirected	253 339	496 489	3.92	7.57
	physics coauthorship	undirected	52 909	245 300	9.27	6.19
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92
	telephone call graph	undirected	47 000 000	80 000 000	3.16	
	email messages	directed	59 912	86 300	1.44	4.95
	email address books	directed	16 881	57 029	3.38	5.22
	student relationships	undirected	573	477	1.66	16.01
	sexual contacts	undirected	2 810			
Information	WWW nd.edu	directed	269 504	1 497 135	5.55	11.27
	WWW Altavista	directed	203 549 046	2 130 000 000	10.46	16.18
	citation network	directed	783 339	6 716 198	8.57	
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87
	word co-occurrence	undirected	460 902	17 000 000	70.13	
Technological	Internet	undirected	10 697	31 992	5.98	3.31
	power grid	undirected	4 941	6 594	2.67	18.99
	train routes	undirected	587	19 603	66.79	2.16
	software packages	directed	1 439	1 723	1.20	2.42
	software classes	directed	1 377	2 213	1.61	1.51
	electronic circuits	undirected	24 097	53 248	4.34	11.05
	peer-to-peer network	undirected	880	1 296	1.47	4.28
Biological	metabolic network	undirected	765	3 686	9.64	2.56
	protein interactions	undirected	2 115	2 240	2.12	6.80
	marine food web	directed	135	598	4.43	2.05
	freshwater food web	directed	92	997	10.84	1.90
	neural network	directed	307	2 359	7.68	3.97

Watts and Strogatz (1998) demonstrated that adding a few long-distance connections to a regular network yields a dramatic decrease of d.

Start from a regular "ring" graph with N nodes, where each node is connected to the m right-neighbors and to the m left-neighbors (=each node has exactly degree $2m$).



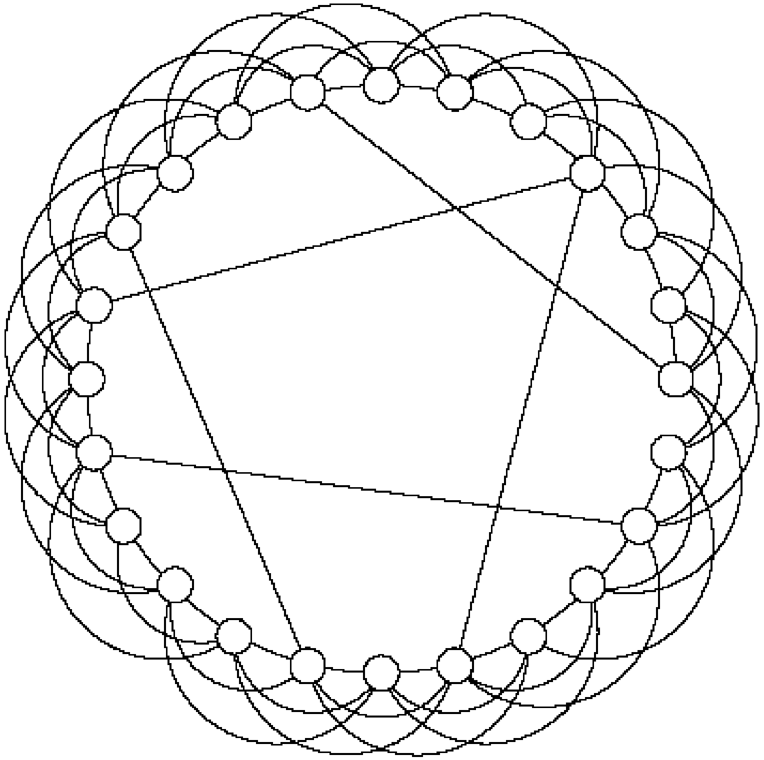
The network has large clustering coefficient (typical of "regular" networks)

$$C = \frac{3m-3}{4m-2}$$

and the average distance is also large (grows linearly with N)

$$d = \frac{N}{4m}$$

"Rewiring": Scan all nodes $i = 1,2,\ldots,N$. Consider all the links $i \leftrightarrow j$ connecting i to its right neighbors and, with probability p, break the connection to j and redirect it to a randomly selected node.



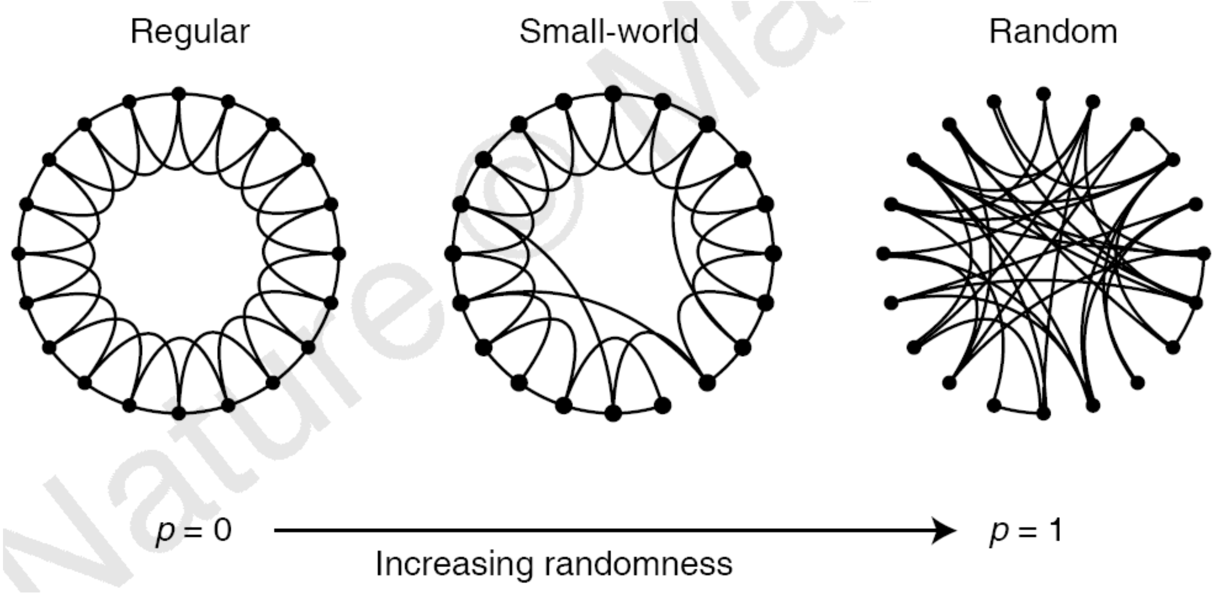
If p is small, the local properties are not significantly modified:

- the degree distribution remains concentrated around the average degree (unchanged!) $< k >= 2m$

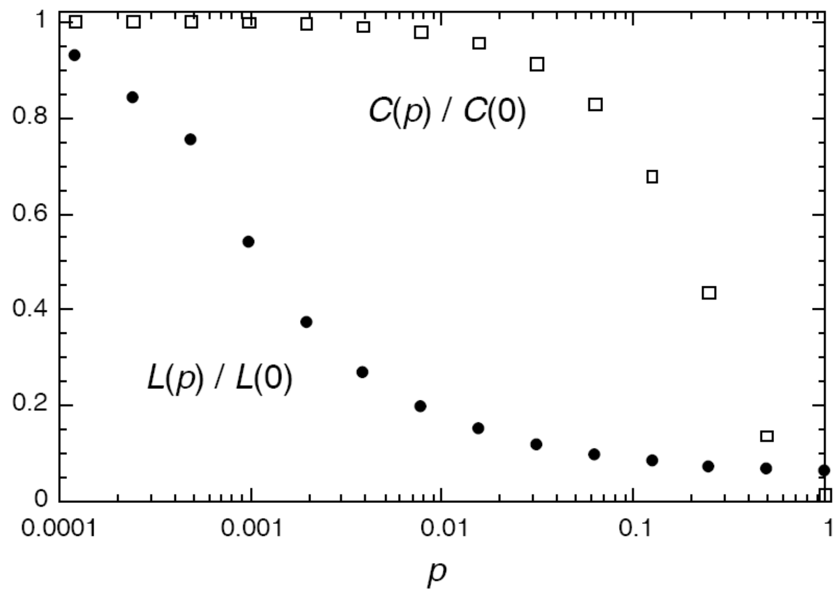
- the clustering coefficient C does not vary significantly

But the birth of few, "long distance" connections is sufficient to yield a dramatic decrease of the average distance, which passes from $d \approx N$ to

$$d \approx \log N$$



p = fraction of links rewired

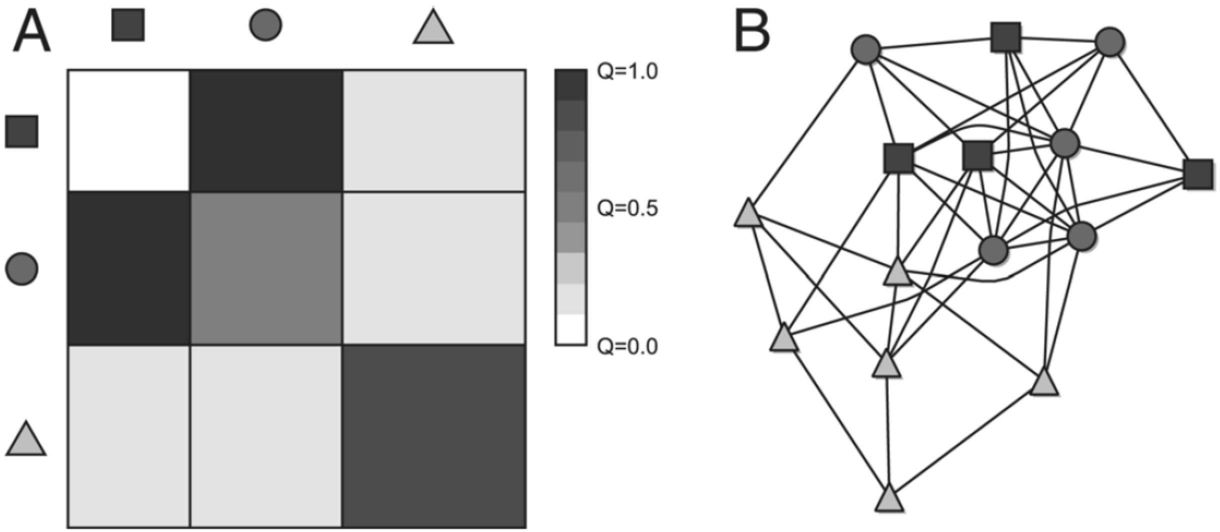


In a suitable p interval, the network mimics many typical real-world networks, i.e., at the same time:

- the clustering coefficient is large
- the average distance is small

STOCHASTIC BLOCK-MODEL

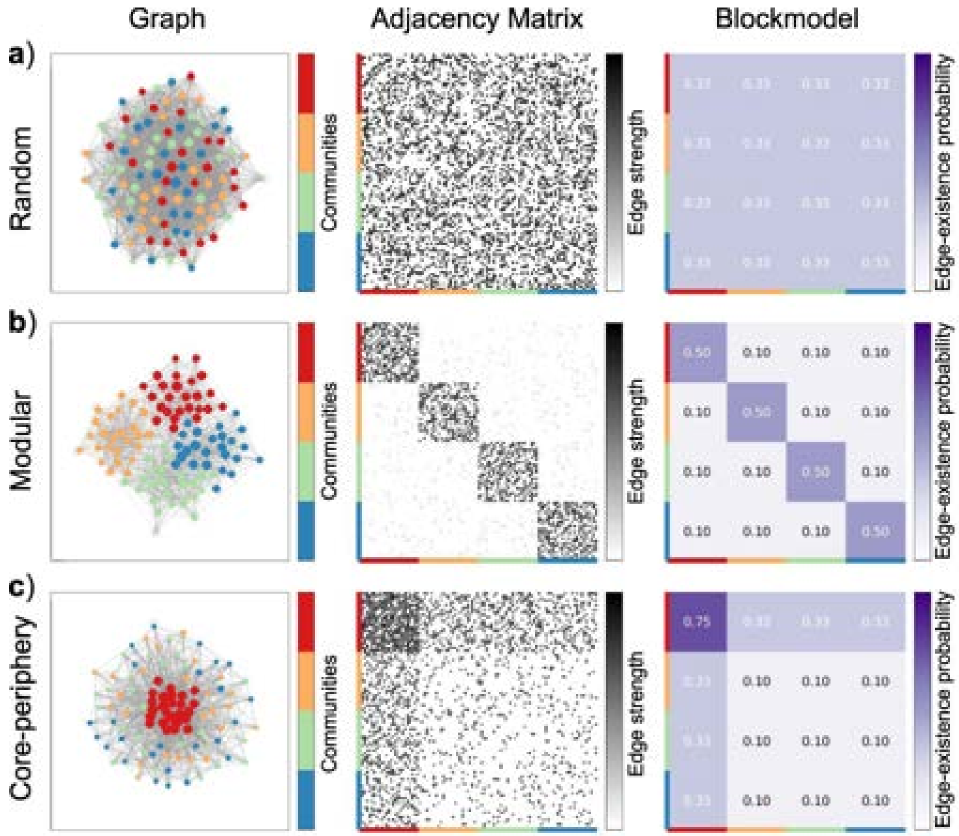
It is a “block” generalization of Erdös-Rényi networks.



The model is completely defined by:

- the number of nodes N and the number of groups (blocks) B
- a partition of the nodes, i.e., the group membership b_i of each node i
- the probabilities $p_{rs} = p_{sr}$ that a node in group r is linked to a node in group s (including $r = s$)

It is a general, versatile model for large-scale networks, suitable to parameter identification via statistical inference techniques.



Faskowitz et al, Sci.Rep 2018