NETWORK MODELS

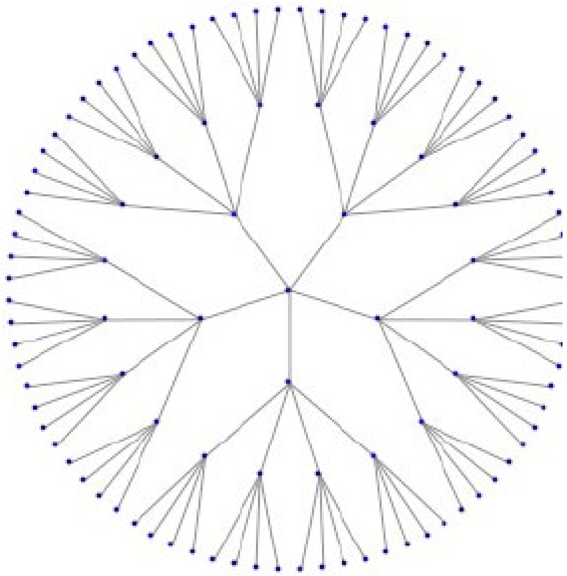
Carlo PICCARDI

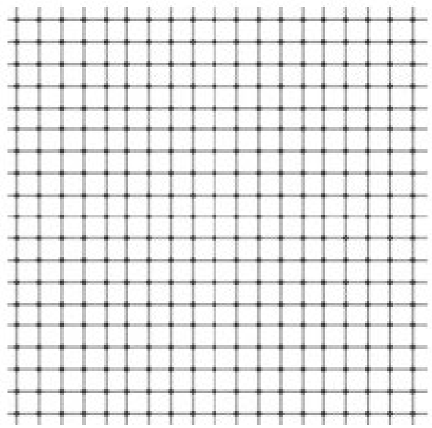
DEIB - Department of Electronics, Information and Bioengineering
Politecnico di Milano, Italy

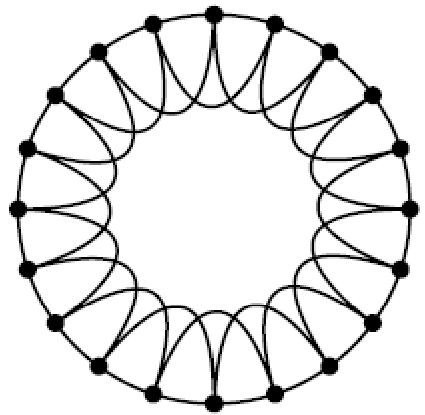
email carlo.piccardi@polimi.it
https://piccardi.faculty.polimi.it

"REGULAR" NETWORKS

Trees and lattices...





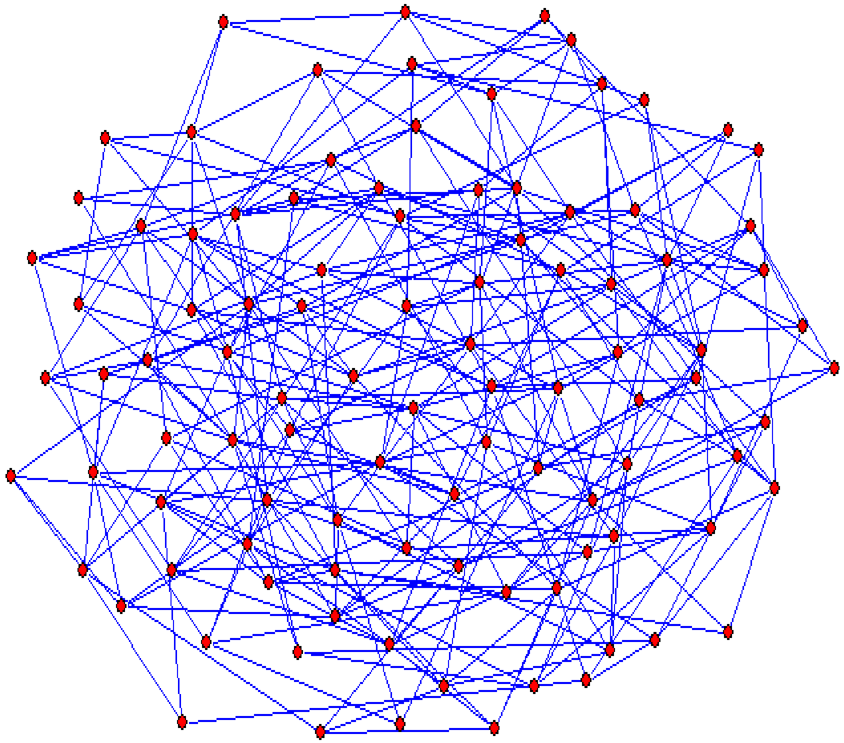


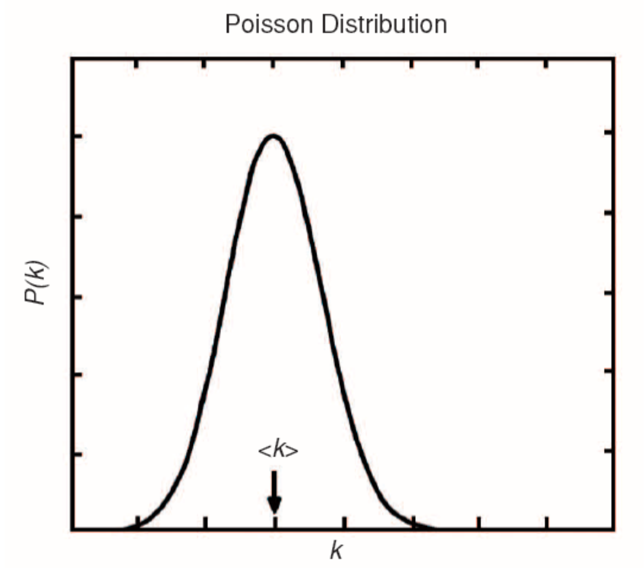
... are very rarely representative of real-world networks.

<u>"RANDOM" (Erdös-Rényi) NETWORKS</u>

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

"$G(N, L)$ *model*"





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...

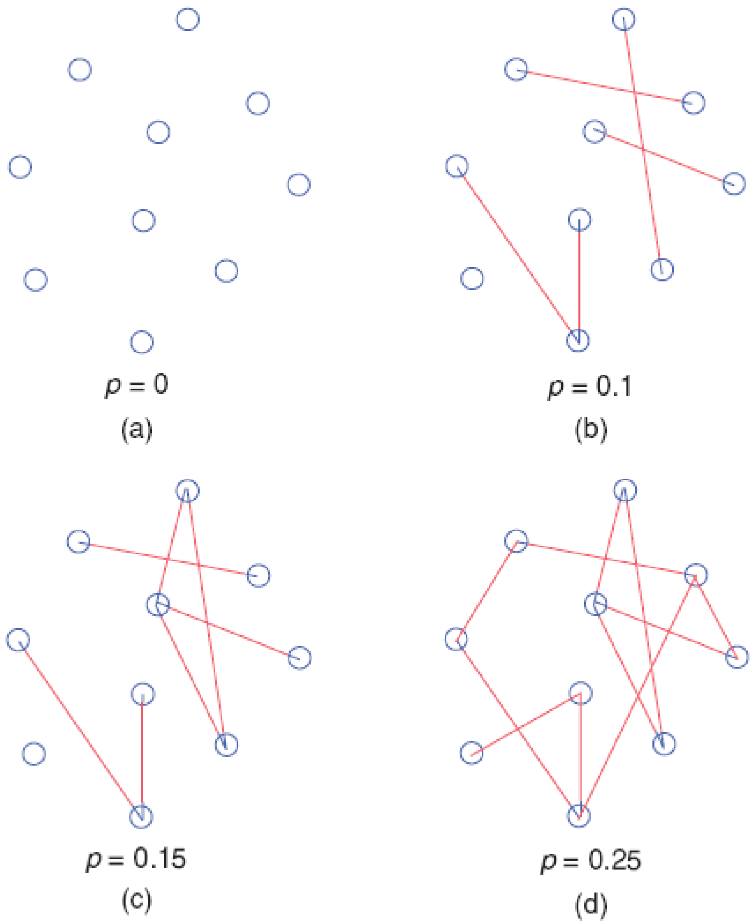


Figure 6. Evolution of a random graph. Given 10 isolated nodes in (a), one connects every pair of nodes with probability (b) $p = 0.1$,(c) $p = 0.15$ and (d) $p = 0.25$, respectively.

An alternative procedure ("$G(N, p)$ *model*"):

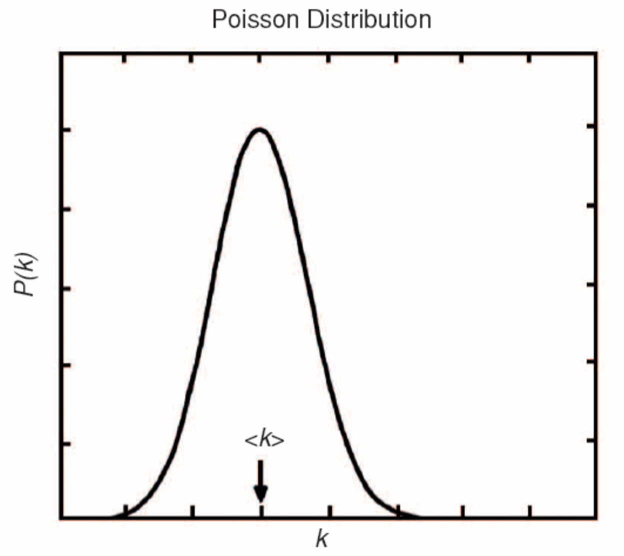
Start from a graph with N nodes and no links, and connect each pair i, j with a given probability p.

Then the degree distribution is binomial:

$$P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}$$

with $\langle k \rangle = p(N-1)$.

Some properties (for $N \to \infty$ and fixed $\langle k \rangle$):



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

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

- The network has a giant component $O(N)$ if $\langle k \rangle > 1$ ($p > 1/N$).

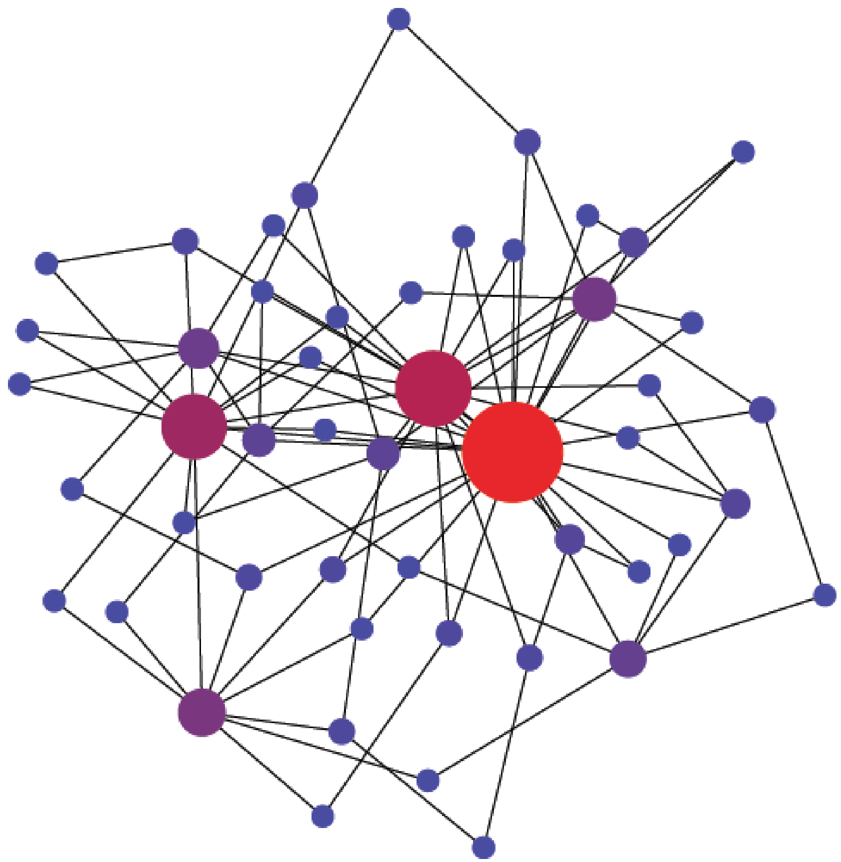
- The average distance $d \cong \log N / \log < k >$ grows "slowly" with N ("small-world" effect) $\rightarrow$ "Large" networks (=large N) have a relatively small average distance.

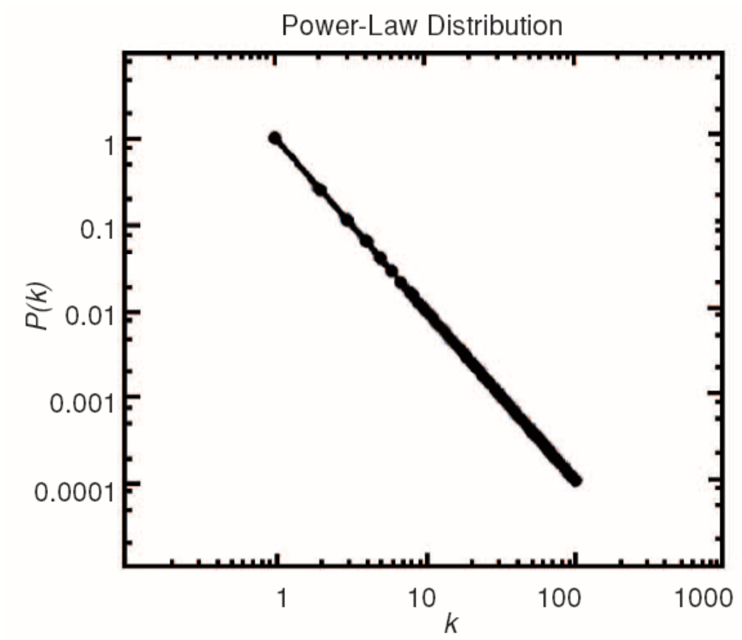
- The clustering coefficient $C = p \cong < k > / N$ tends to 0 as N grows $\rightarrow$ "Large" networks have vanishing clustering.

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.



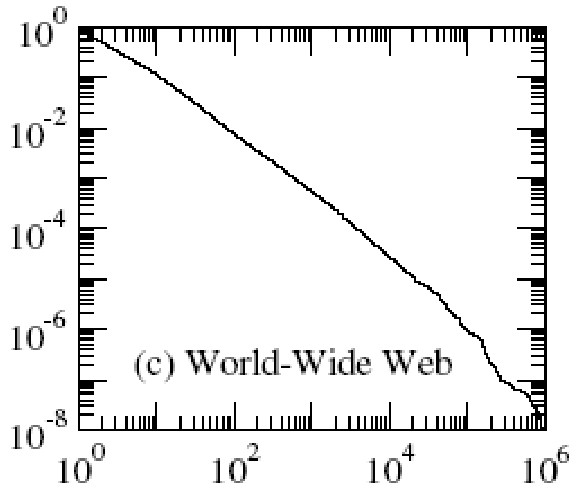


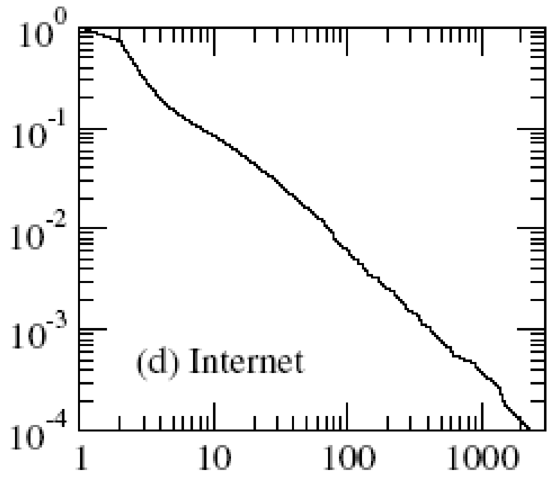
For large N, the degree distribution is a power-law function $P(k) \approx k^{-\alpha}$:

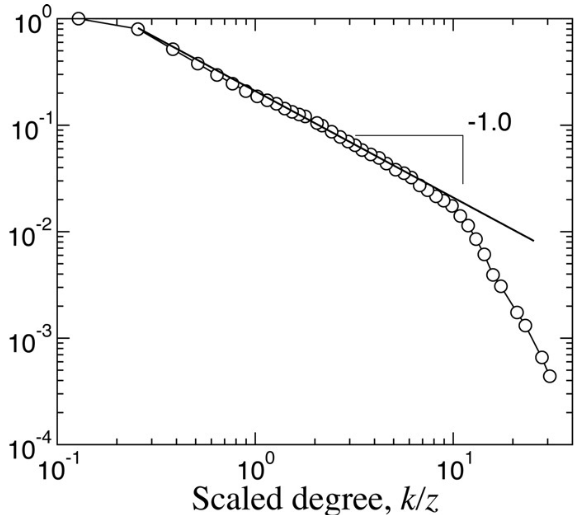
⇨ node degrees have large fluctuations around $< k >$: there is no "typical" scale of node degree

⇨ the network is strongly heterogeneous

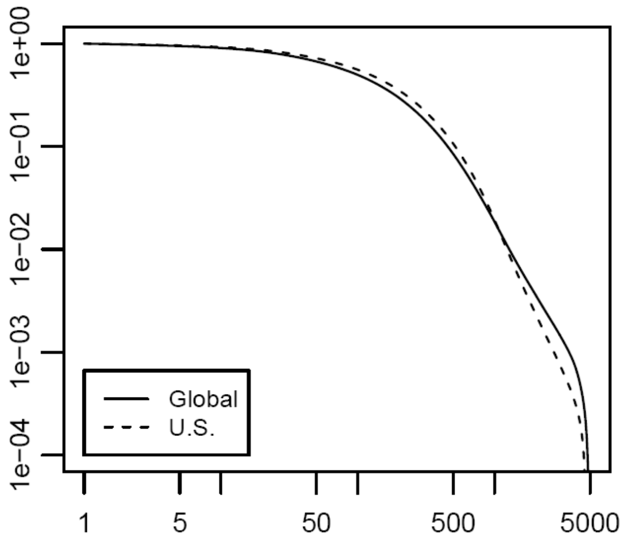
Some examples of (cumulative) degree distribution:







the air transportation network



Facebook (721 million nodes, May 2011)

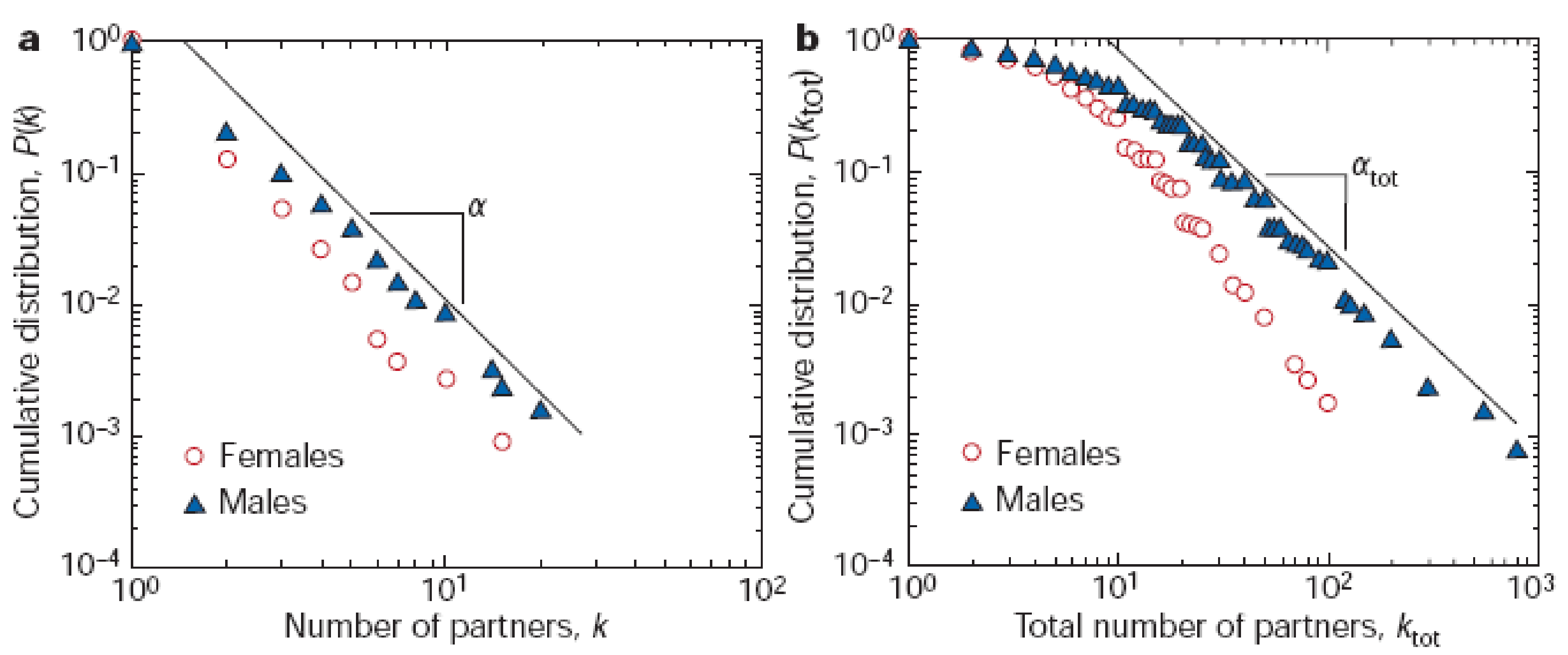
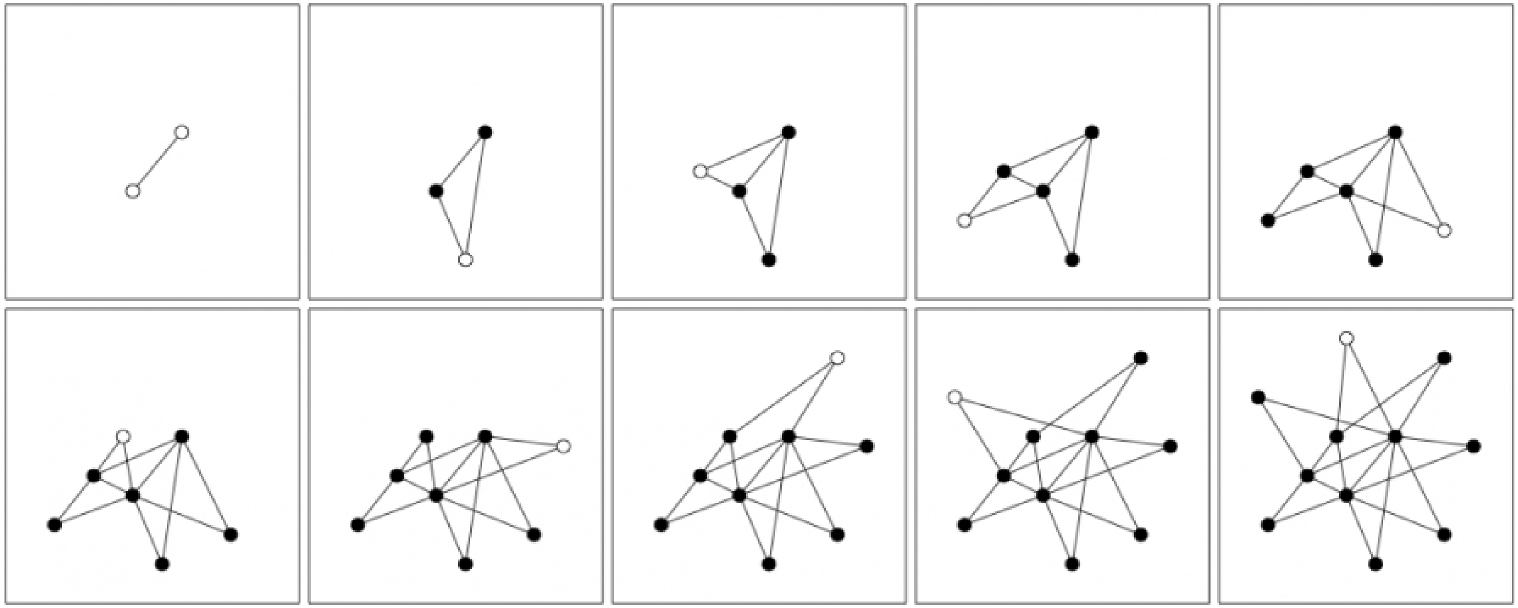


Figure 2 Scale-free distribution of the number of sexual partners for females and males. **a,** Distribution of number of partners, k, in the previous 12 months. Note the larger average number of partners for male respondents: this difference may be due to 'measurement bias' — social expectations may 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 (BA) 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 that a link of the new node i connects to the existing node j be $k_j / \Sigma_h k_h$.



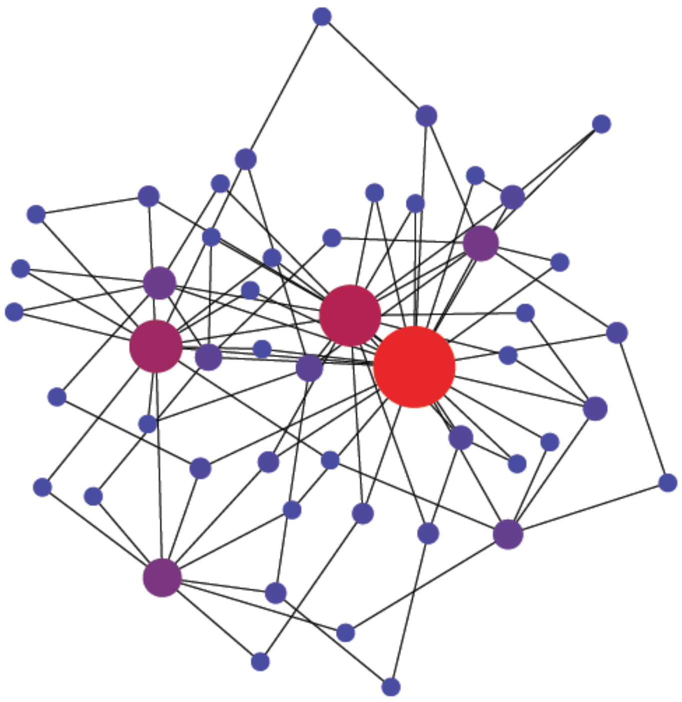
Then for $N \to \infty$:

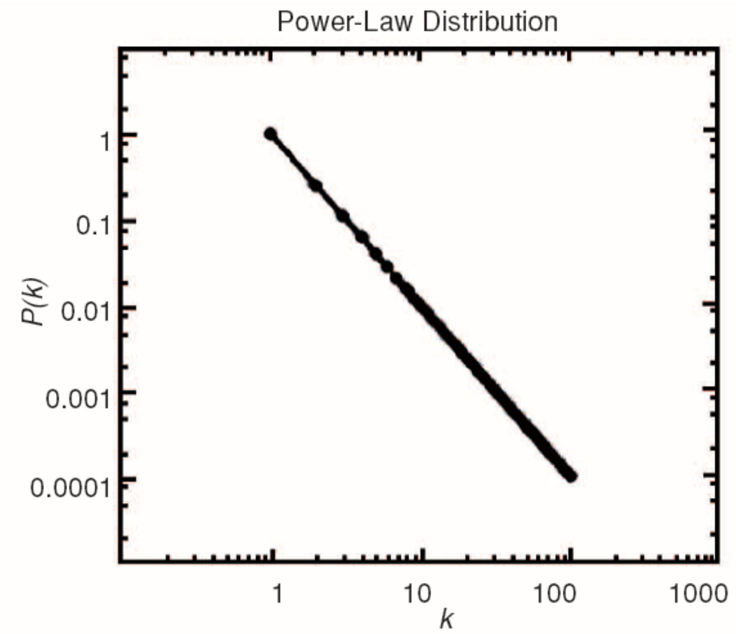
- the average degree tends to $<k> = 2m$ and the degree distribution to the power-law $P(k) \approx k^{-3}$

- $<k^2>$ 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 as $C \approx (\log N)^2 / N \to 0$





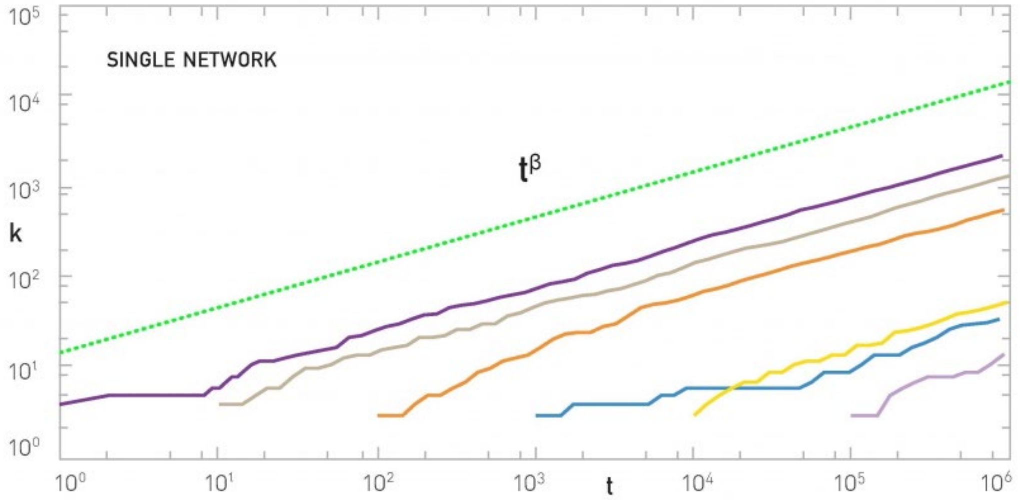
Computing the degree distribution (the "continuum approach"):

- 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: In the continuum approach, all nodes evolve exactly in the same way ($k_i \approx t^{0.5}$), but older nodes (=smaller t_i) have larger degree.

$\Longleftarrow$ *the degree of a few nodes in an actual (i.e. randomized) realization of a BA network*

- 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

$$P(k) = -\frac{d\bar{P}(k)}{dk} = 2m^2 k^{-3}$$

The continuum approach correctly predicts $P(k) \approx k^{-3}$. The exact degree distribution of a BA network is proved to be

$$P(k) = \frac{2m(m+1)}{k(k+1)(k+2)}$$

Two (of the many) generalizations of the Barabasi-Albert algorithm:

- Dorogovtsev-Mendes-Samukhin (DMS) model, to get a power law degree distribution $P(k) \approx k^{-\gamma}$ with arbitrary $\gamma \in (2, \infty)$.

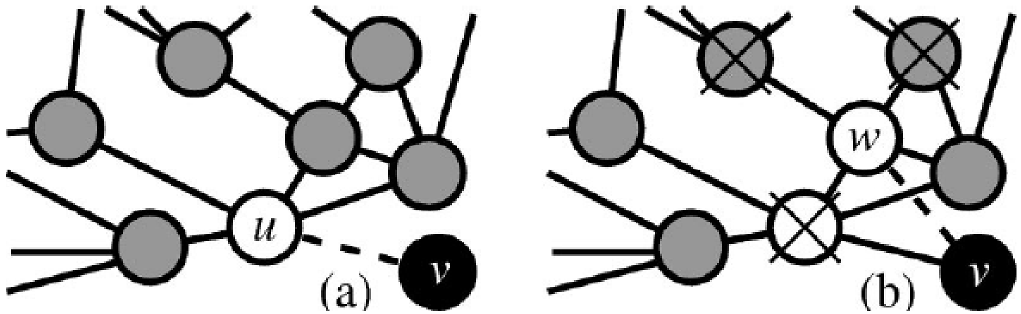
Modifies the preferential attachment probability that a link of the new node i connects to the existing node j

$$\text{from} \quad \frac{k_j}{\sum_h k_h} \qquad \text{to} \quad \frac{k_j + k_0}{\sum_h (k_h + k_0)}$$

By choosing a value $k_0 \in (-m, \infty)$, it is proved that $\gamma = 3 + k_0/m$.

- Holme-Kim (HK) model, to get a non-vanishing (large) clustering coefficient C.

Forces the creation of triangles by alternating (in a probabilistic fashion) preferential attachment steps and triad formation steps.



"SMALL-WORLD" (Watts-Strogatz) NETWORKS

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

Empirically, it is observed that

$$d \approx \log N$$

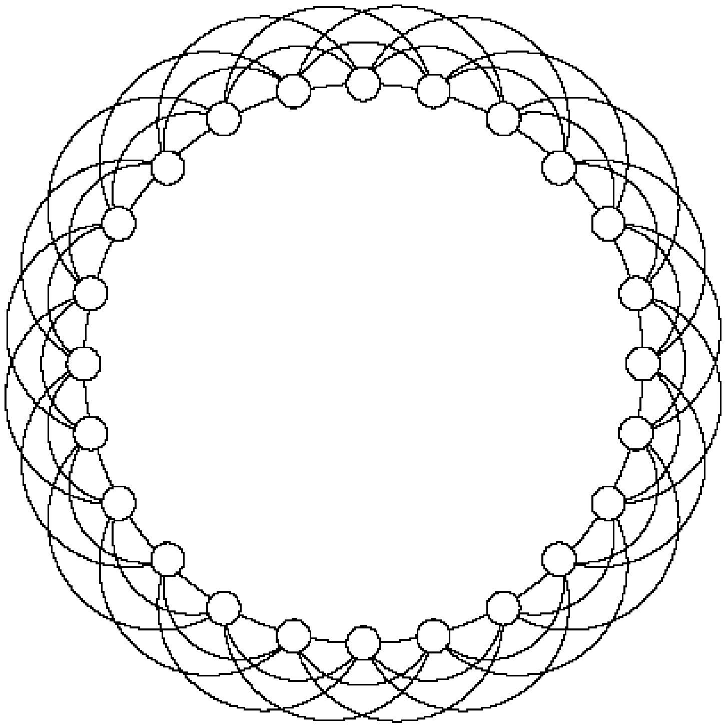
i.e. d increases "slowly" with N ("small-world" effect).

Network	N	L	‹k›	‹d›	d_{max}
Internet	192,244	609,066	6.34	6.98	26
WWW	325,729	1,497,134	4.60	11.27	93
Power Grid	4,941	6,594	2.67	18.99	46
Mobile-Phone Calls	36,595	91,826	2.51	11.72	39
Email	57,194	103,731	1.81	5.88	18
Science Collaboration	23,133	93,437	8.08	5.35	15
Actor Network	702,388	29,397,908	83.71	3.91	14
Citation Network	449,673	4,707,958	10.43	11.21	42
E. Coli Metabolism	1,039	5,802	5.58	2.98	8
Protein Interactions	2,018	2,930	2.90	5.61	14

Barabasi, 2016

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 $2m$).



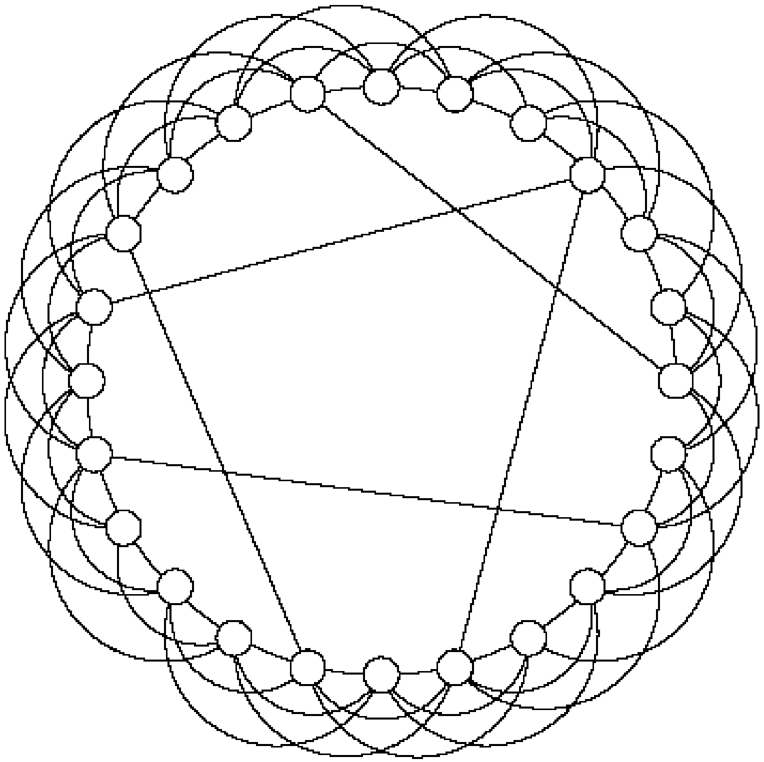
The network has large clustering coefficient (typical of "regular" networks)

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

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.



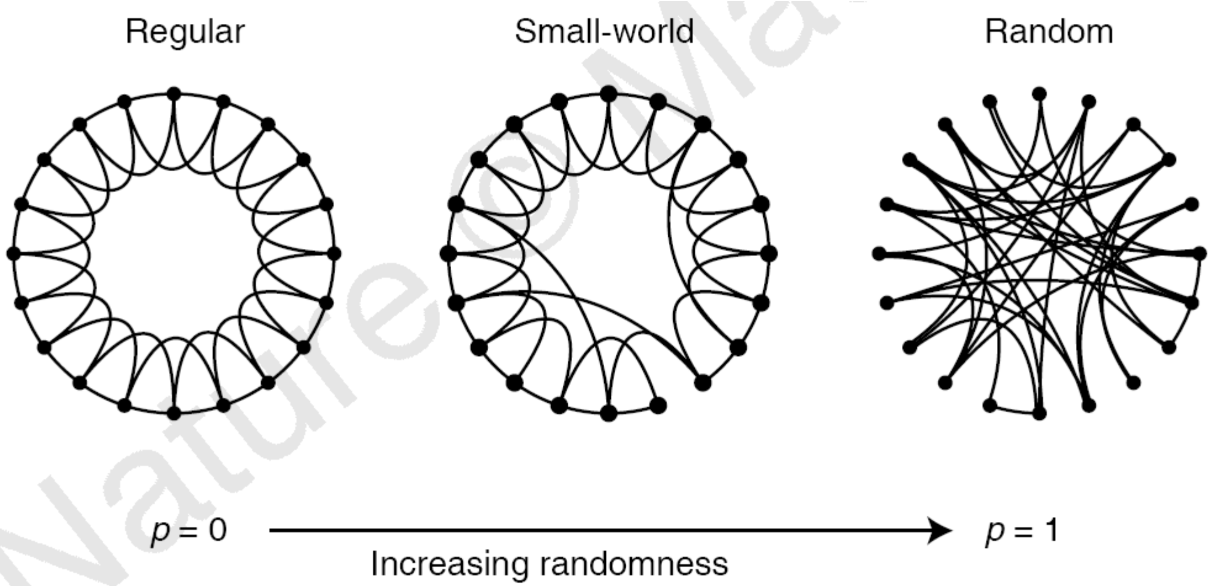
If p is small, the local properties are not significantly modified:

- 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

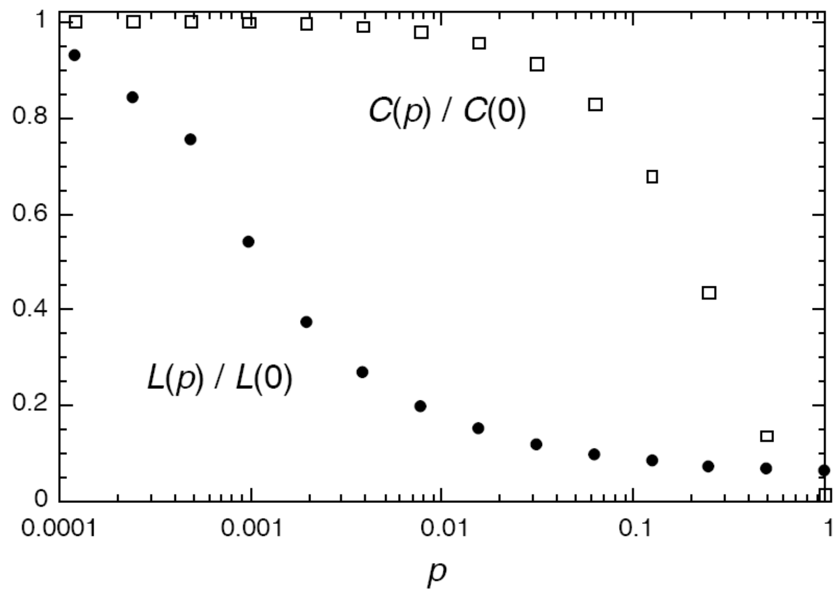
$$d \approx \log N$$



p = fraction of links rewired

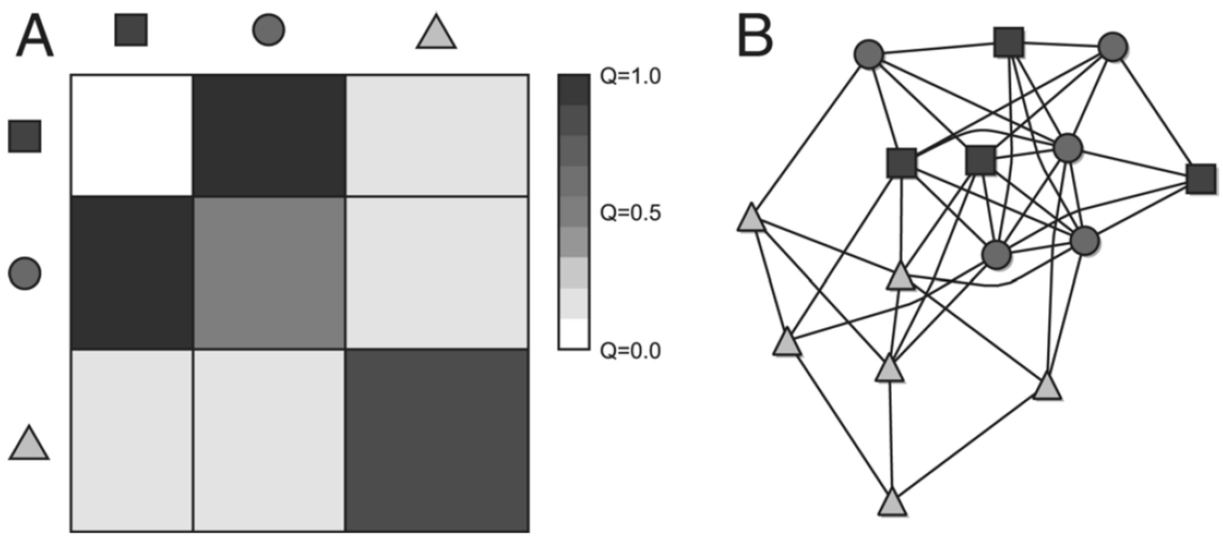
In a suitable p interval, the network mimics many typical real-world networks, i.e., at the same time:

- 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.

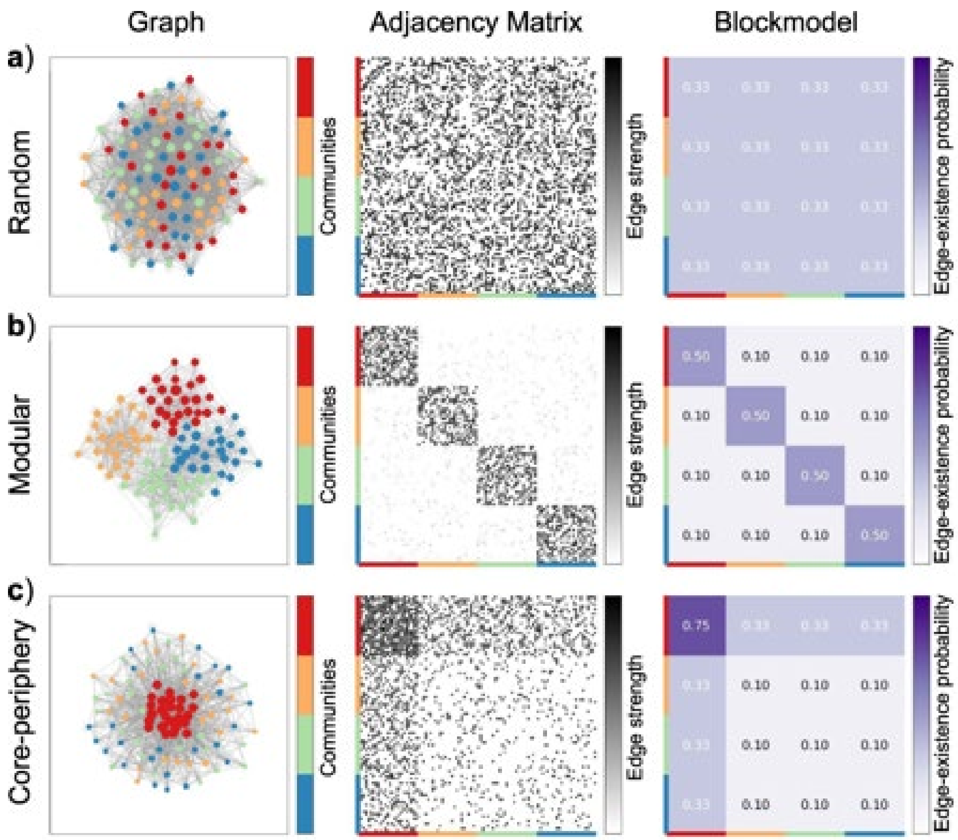
Special cases:

- Erdös-Rényi network, 1 block:
  $p_{rs} = p$ *for any node pair*

- Modular (community structure), q blocks:
  large intra-block connectivity p_{rr}
  small inter-block connectivity p_{rs} $(r \neq s)$

- Core-periphery structure, 2 blocks:
  large intra-core connectivity p_{11}
  small intra-periphery connectivity p_{22}
  intermediate core-periph. connectivity $p_{12} = p_{21}$

- Bipartite network, 2 blocks:
  null intra-block connectivity $p_{11} = p_{22} = 0$



Faskowitz et al, Sci.Rep 2018