Frontiers of Network Science
Fall 2018

Class 7: Small World and BA Networks (Chapters 4 and 5 in Textbook)

Boleslaw Szymanski

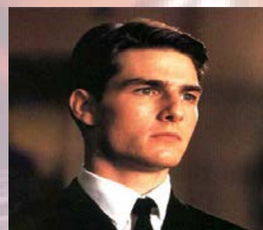
based on slides by
Albert-László Barabási
and Roberta Sinatra

Examples

ACTOR NETWORK

Nodes: actors
Links: cast jointly

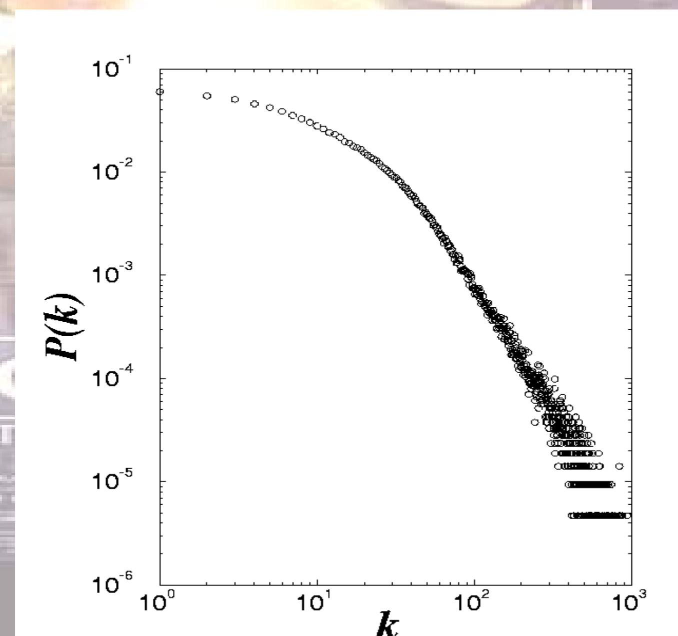
IMDb Internet Movie Database



Days of Thunder *(1990)*
Far and Away *(1992)*
Eyes Wide Shut *(1999)*

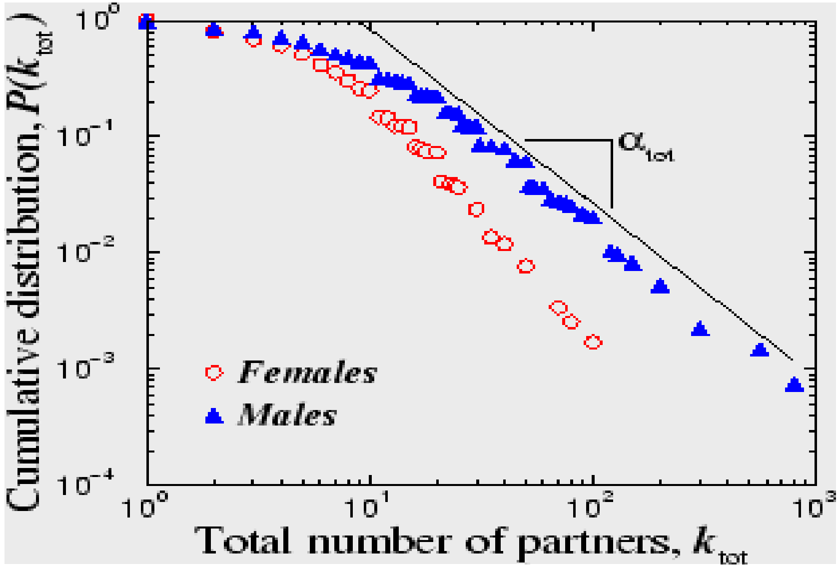
N = 212,250 actors
$\langle k \rangle = 28.78$

$P(k) \sim k^{-\gamma}$
$\gamma = 2.3$



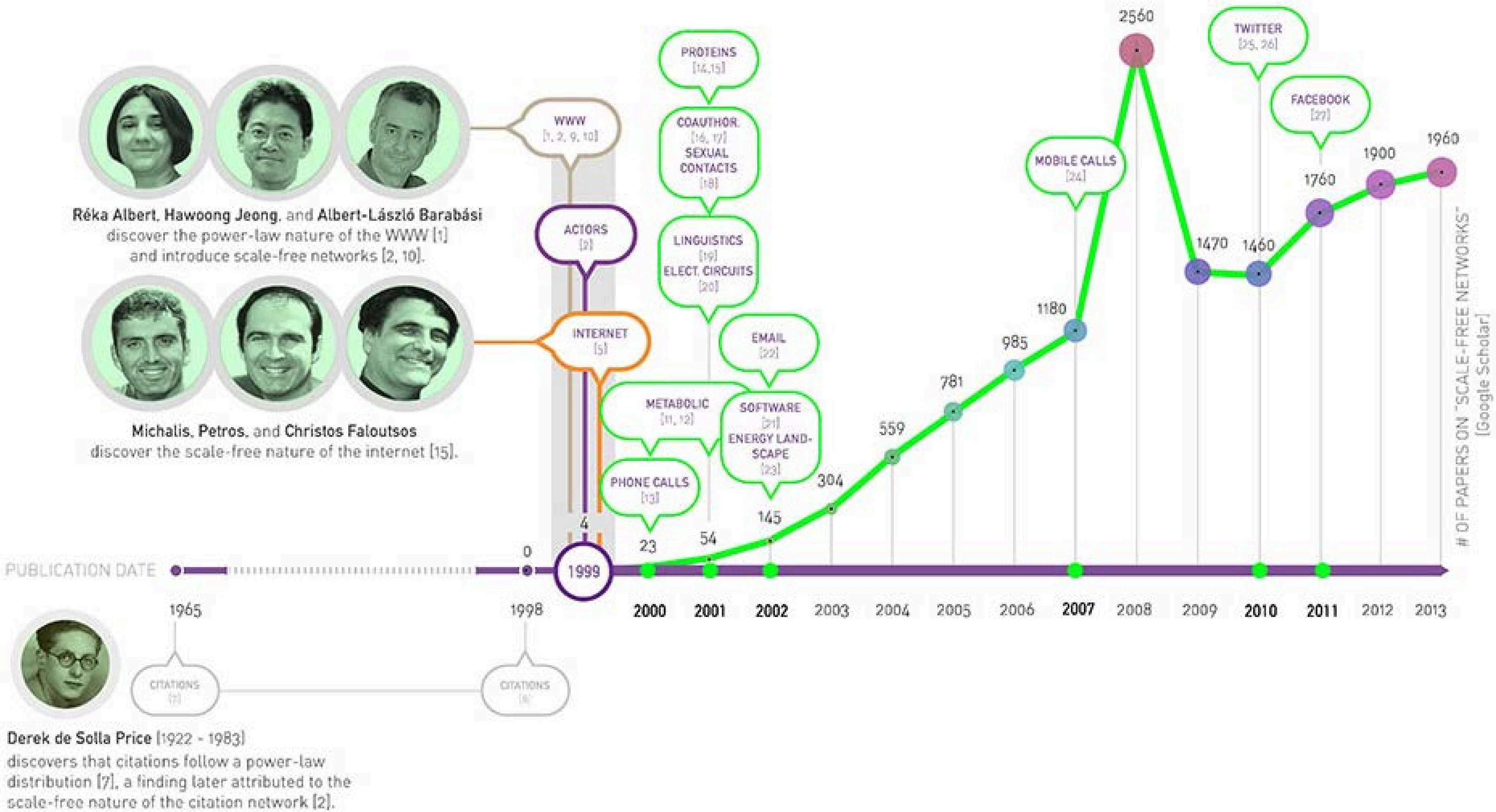
SWEDISH SE-WEB

Nodes: people (Females; Males)
Links: sexual relationships



4781 Swedes; 18-74;
59% response rate.

Liljeros et al. Nature 2001



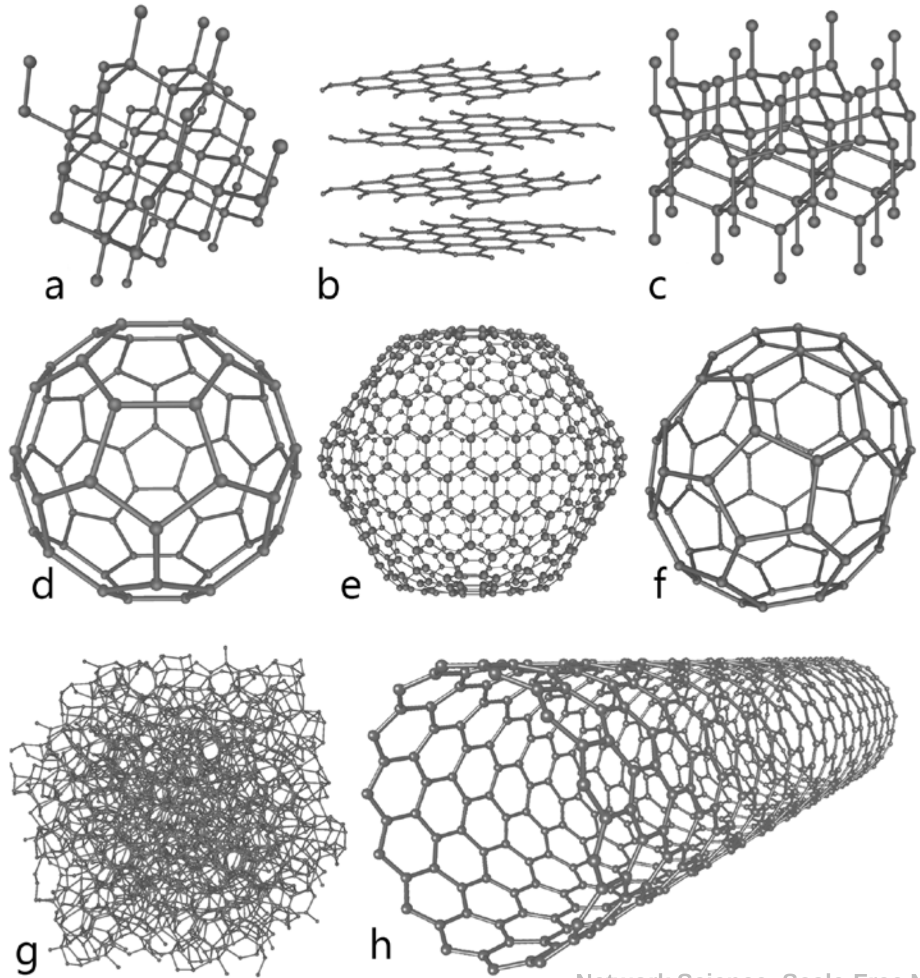
Réka Albert, Hawoong Jeong, and Albert-László Barabási
discover the power-law nature of the WWW [1]
and introduce scale-free networks [2, 10].
Michalis, Petros, and Christos Faloutsos
discover the scale-free nature of the internet [15].
Derek de Solla Price [1922 - 1983]
discovers that citations follow a power-law
distribution [7], a finding later attributed to the
scale-free nature of the citation network [2].
PUBLICATION DATE
1965
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
CITATIONS
WWW [1, 2, 9, 10]
ACTORS [2]
INTERNET [5]
PROTEINS [14,15]
COAUTHOR. [16, 17]
SEXUAL CONTACTS [18]
LINGUISTICS [19]
ELECT. CIRCUITS [20]
METABOLIC [11, 12]
PHONE CALLS [13]
EMAIL [22]
SOFTWARE [21]
ENERGY LAND-SCAPE [23]
MOBILE CALLS [24]
TWITTER [25, 26]
FACEBOOK [27]
0
4
23
54
145
304
559
781
985
1180
2560
1470
1460
1760
1900
1960
OF PAPERS ON "SCALE-FREE NETWORKS"
[Google Scholar]

Not All Networks Are Scale-free

- *Networks appearing in material science, like the network describing the bonds between the atoms in crystalline or amorphous materials, where each node has exactly the same degree.*

- *The neural network of the C.elegans worm.*

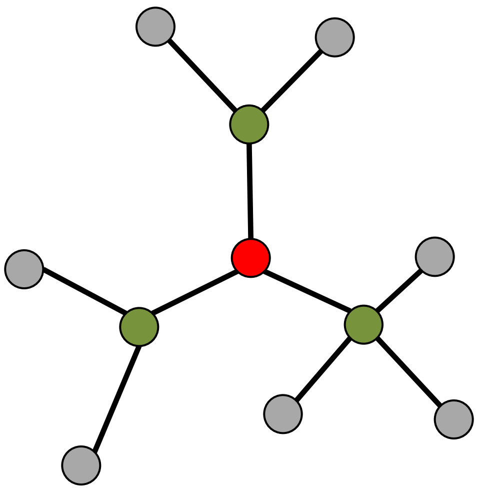
- *The power grid, consisting of generators and switches connected by transmission lines*



Ultra-small property

DISTANCES IN RANDOM GRAPHS

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



- number of first neighbors: $N_1 \cong \langle k \rangle$
- number of second neighbors: $N_2 \cong \langle k \rangle^2$
- number of neighbors at distance d: $N_d \cong \langle k \rangle^d$
- estimation of maximum distance:

$$1 + \sum_{l=1}^{l_{max}} \langle k \rangle^i = N \Rightarrow l_{max} = \frac{\log N}{\log \langle k \rangle}$$

SMALL WORLD BEHAVIOR IN SCALE-FREE NETWORKS

$$k_{\max} = k_{\min} N^{\frac{1}{\gamma-1}}$$

Ultra Small World

Small World

$$< l > \sim \begin{cases} const. & \gamma = 2 \\ \dfrac{\ln \ln N}{\ln(\gamma - 1)} & 2 < \gamma < 3 \\ \dfrac{\ln N}{\ln \ln N} & \gamma = 3 \\ \ln N & \gamma > 3 \end{cases}$$

$\gamma = 2$: Size of the biggest hub is of order O(N). Most nodes can be connected within two layers of it, thus the average path length will be independent of the system size.

$2 < \gamma < 3$: The average path length increases slower than logarithmically. In a random network all nodes have comparable degree, thus most paths will have comparable length. In a scale-free network the vast majority of the path go through the few high degree hubs, reducing the distances between nodes.

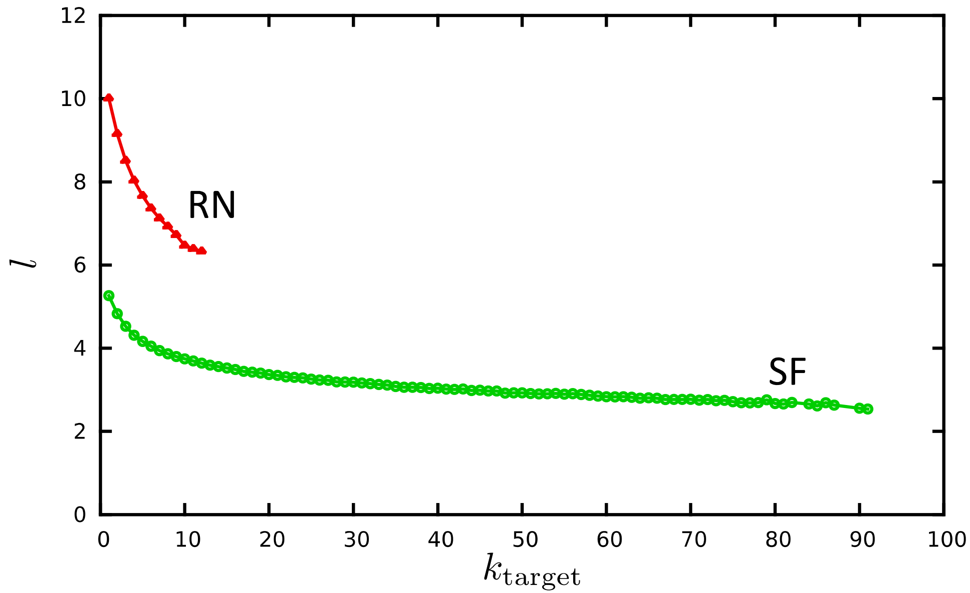
$\gamma = 3$: Some key models produce γ=3, so the result is of particular importance for them. This was first derived by Bollobas and collaborators for the network diameter in the context of a dynamical model, but it holds for the average path length as well.

$\gamma > 3$: The second moment of the distribution is finite, thus in many ways the network behaves as a random network. Hence the average path length follows the result that we derived for the random network model earlier.

Cohen, Havlin Phys. Rev. Lett. 90, 58701(2003); Cohen, Havlin and ben-Avraham, in Handbook of Graphs and Networks, Eds. Bornholdt and Shuster (Willy-VCH, NY, 2002) Chap. 4; Confirmed also by: Dorogovtsev et al (2002), Chung and Lu (2002); (Bollobas, Riordan, 2002; Bollobas, 1985; Newman, 2001

We Are Always Close to the Hubs

" it's always easier to find someone who knows a famous or popular figure than some run-the-mill, insignificant person."
(Frigyes Karinthy, 1929)

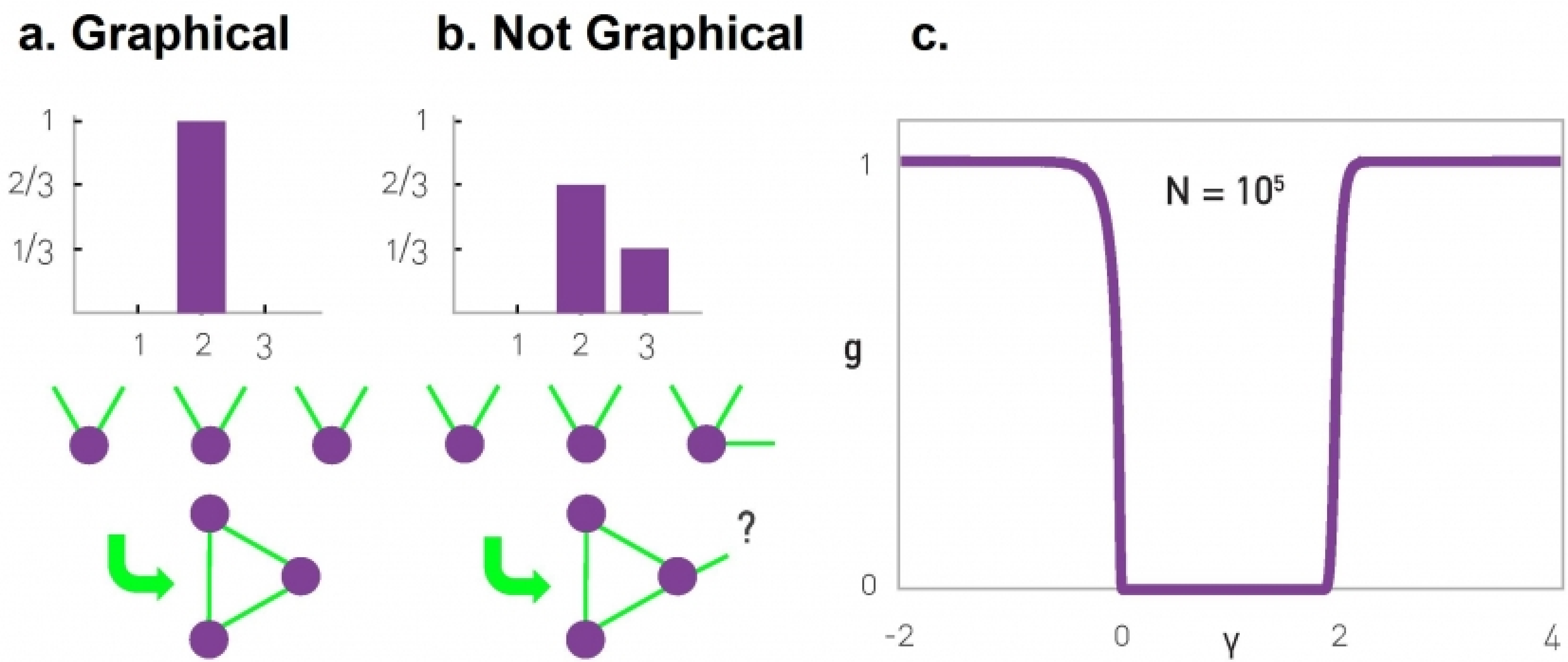


Average person is less popular then this person random friend!

The role of the degree exponent

SUMMARY OF THE BEHAVIOR OF SCALE-FREE NETWORKS

Graphicality: No Large Networks for γ<2



In scale-free networks: $k_{\max} = k_{\min} N^{\frac{1}{\gamma - 1}}$ For γ<2: 1/(γ-2)>1

Why Don't We See Networks with Exponents in the Range of γ=4,5,6, etc?

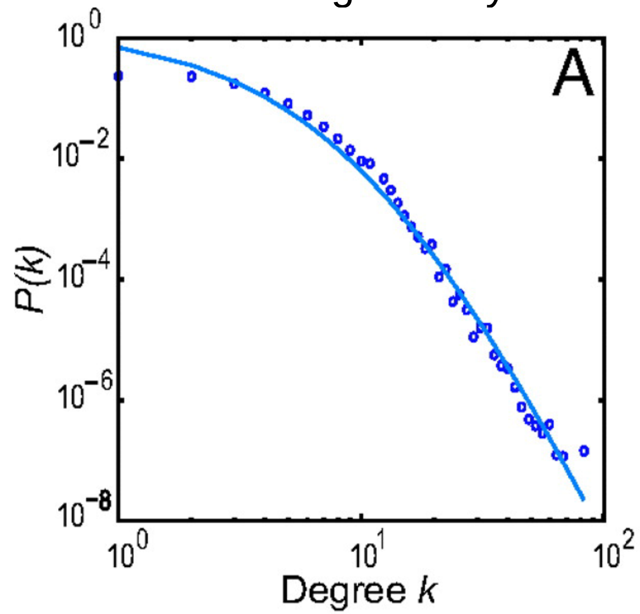
In order to document a scale-free networks, we need 2-3 orders of magnitude scaling. That is, $K_{max} \sim 10^3$

However, that constrains on the system size we require to document it.
For example, to measure an exponent $\gamma=5$,we need to maximum degree a system size of the order of

$$K_{\max} = K_{\min} N^{\frac{1}{\gamma-1}}$$

$$N = \left(\frac{K_{\max}}{K_{\min}} \right)^{\gamma-1} \approx 10^8$$

Mobile Call Network



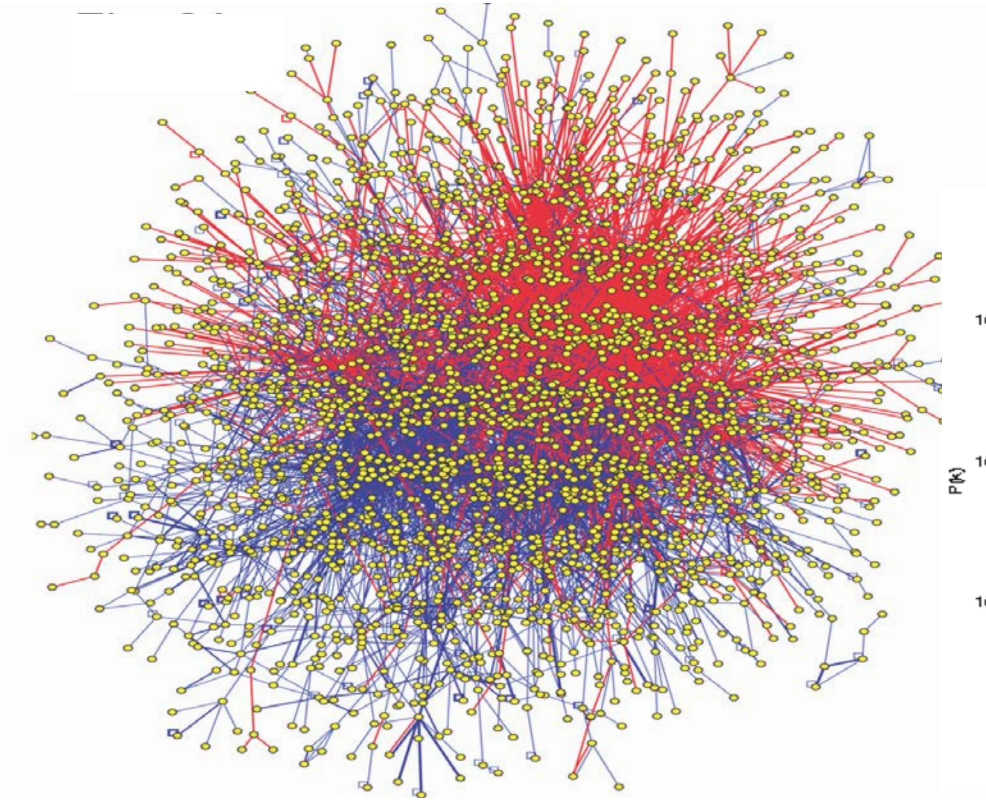
4.6x10[6]
3.4

Characterizing the large-scale structure and the tie strengths of the mobile call graph. Vertex degrees are shown

Onella et al. PNAS 2007

PLOTTING POWER LAWS

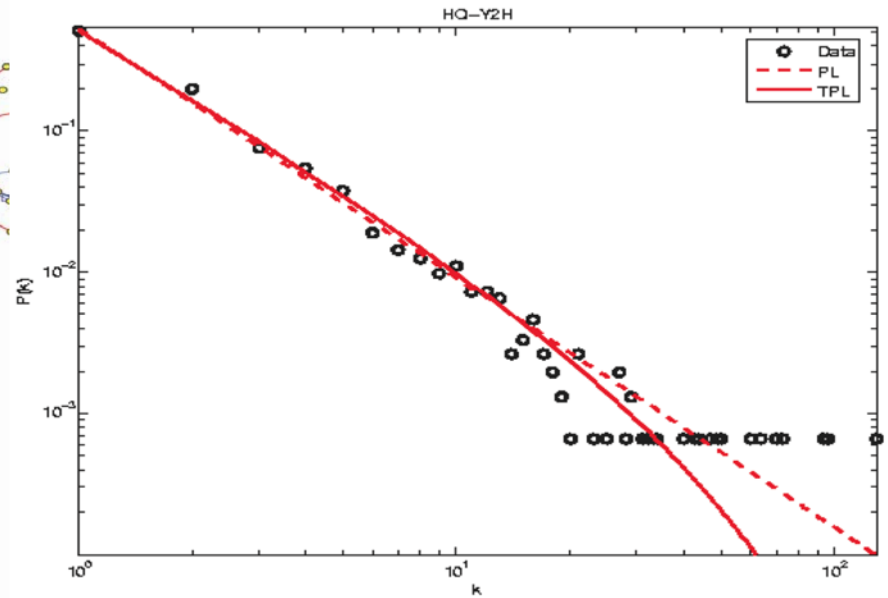
HUMAN INTERACTION NETWORK



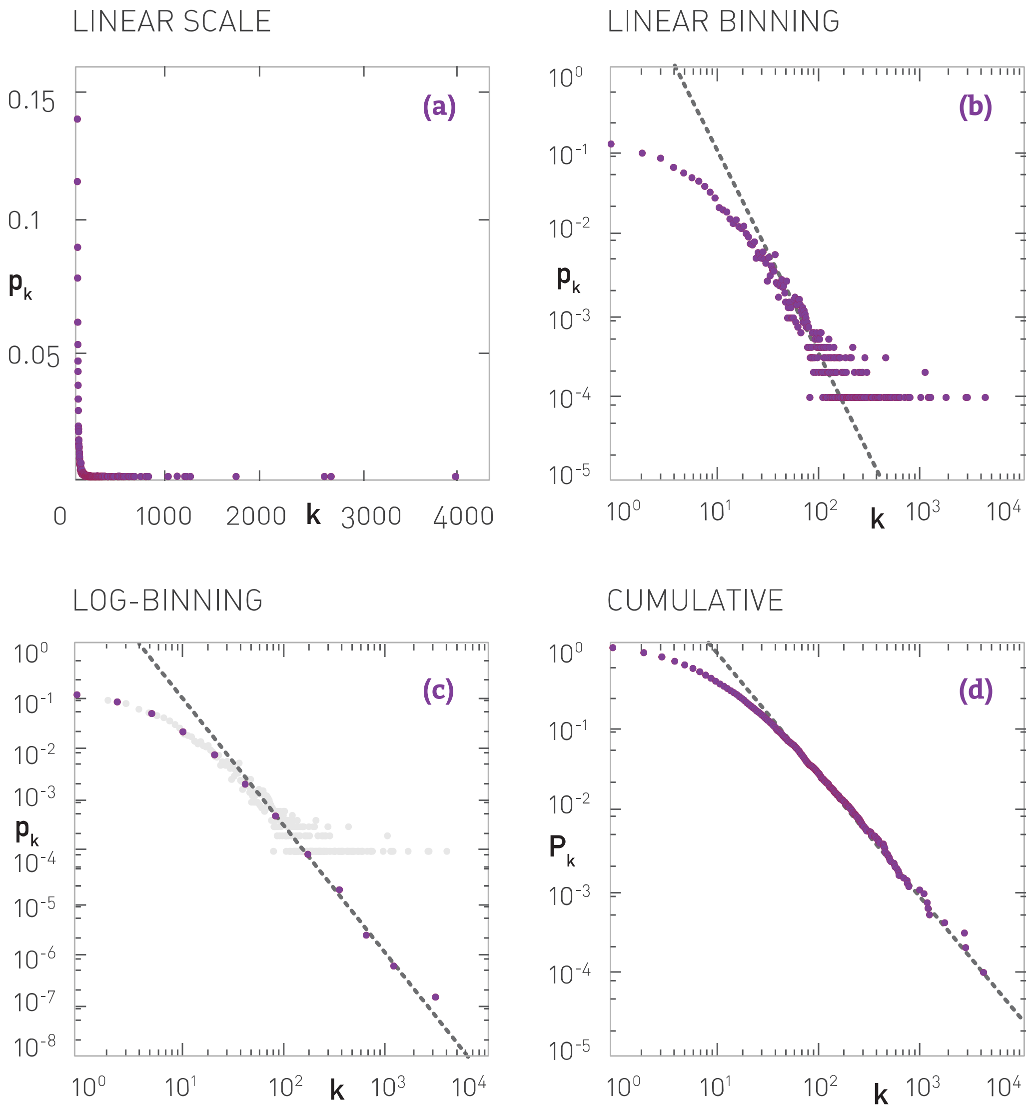
2,800 Y2H interactions

4,100 binary LC interactions

(HPRD, MINT, BIND, DIP, MIPS)



Rual et al. Nature 2005; Stelze et al. Cell 2005



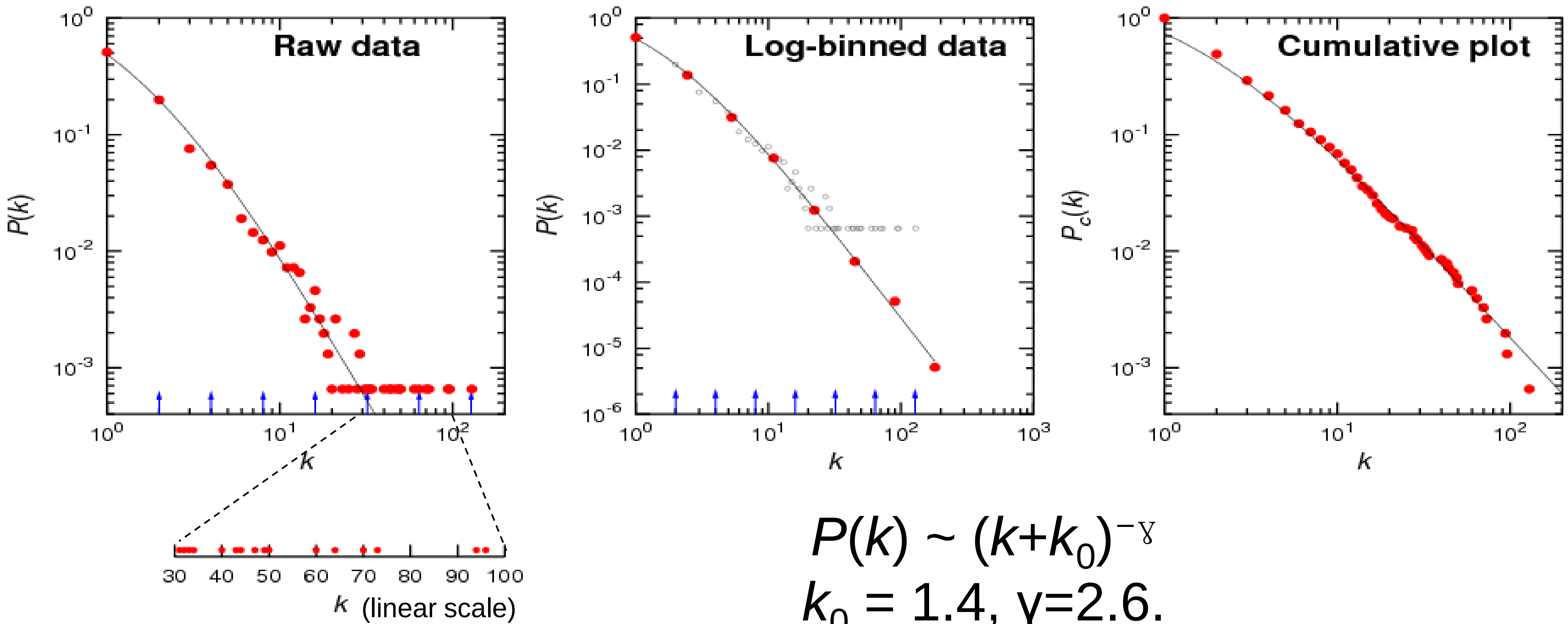
Use a Log-Log Plot

Avoid Linear Binning

Use Logarithmic Binning

Use Cumulative Distribution

HUMAN INTERACTION DATA BY RUAL ET AL.

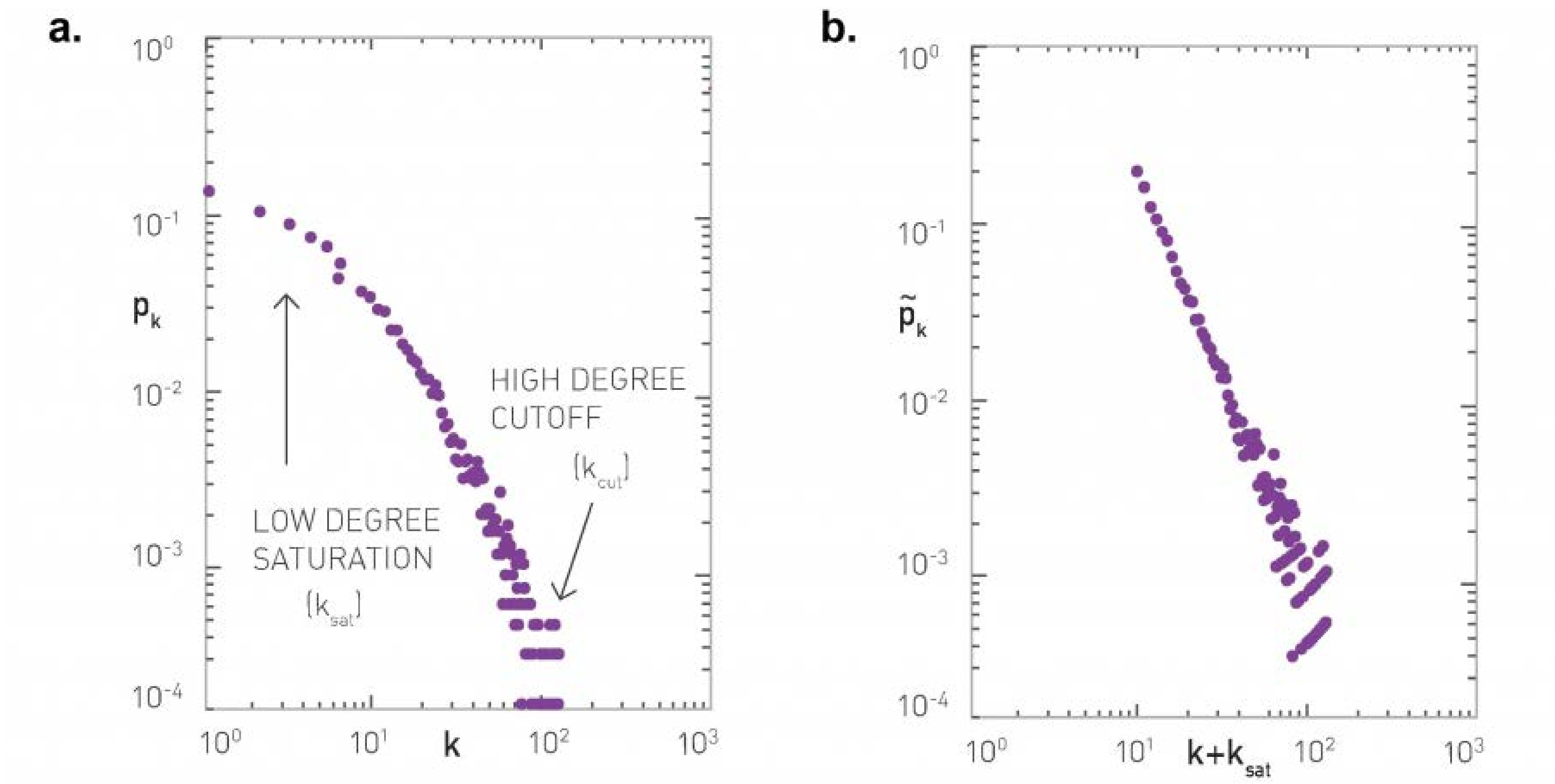


$$P(k) \sim (k+k_0)^{-\gamma}$$

$$k_0 = 1.4,\ \gamma=2.6.$$

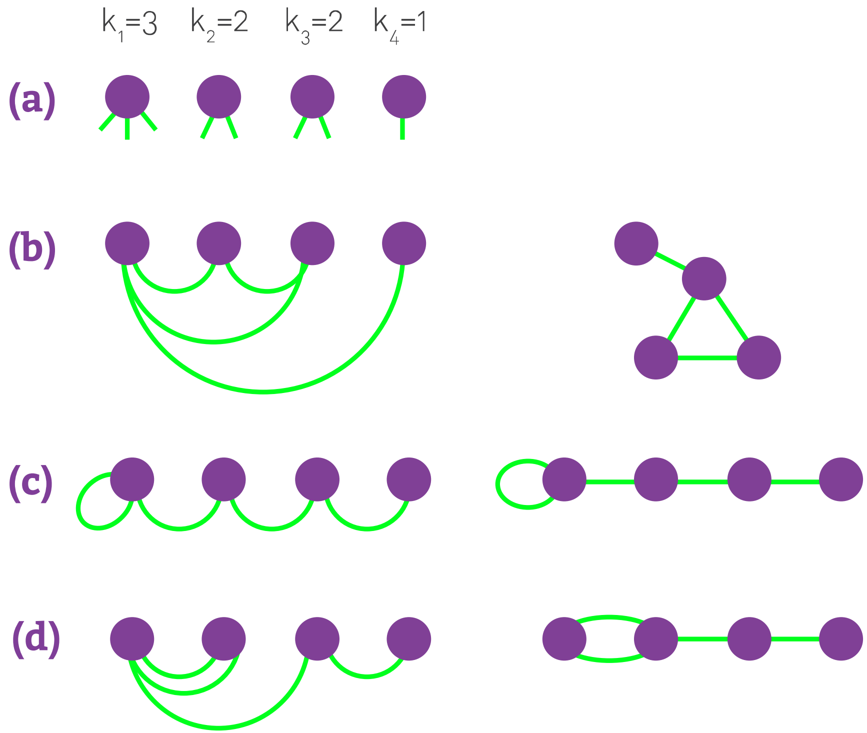
COMMON MISCONCEPTIONS

$$p_k = a(k + k_{sat})^{-\gamma} \exp\left(-\frac{k}{k_{CW}}\right) \quad \bar{p}_k = p_k \exp\left(\frac{k}{k_{CW}}\right) \quad \tilde{k} = k + k_{sat} \; p \sim \tilde{k}^{-\gamma}$$



Generating networks with a pre-defined p_k

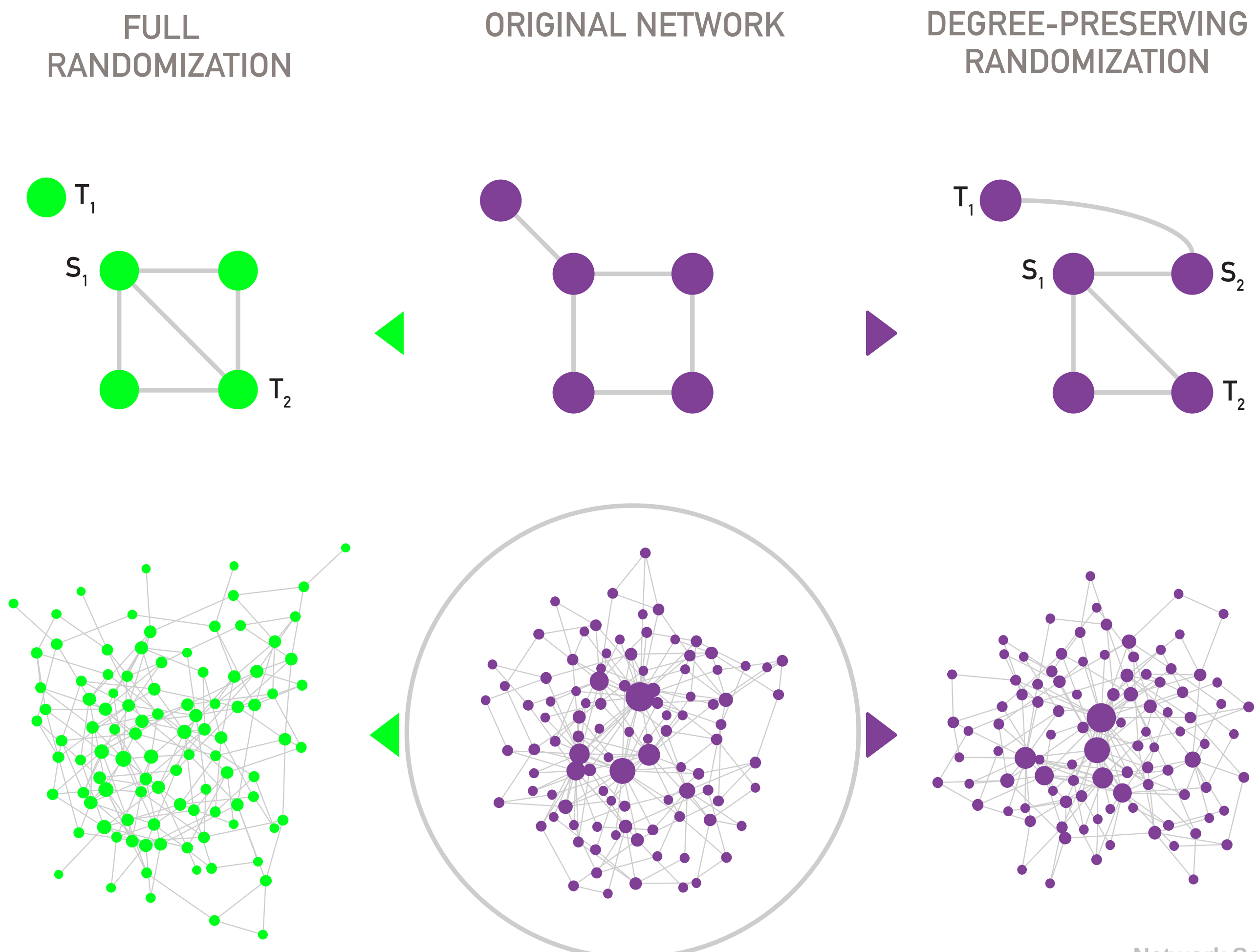
Configuration model



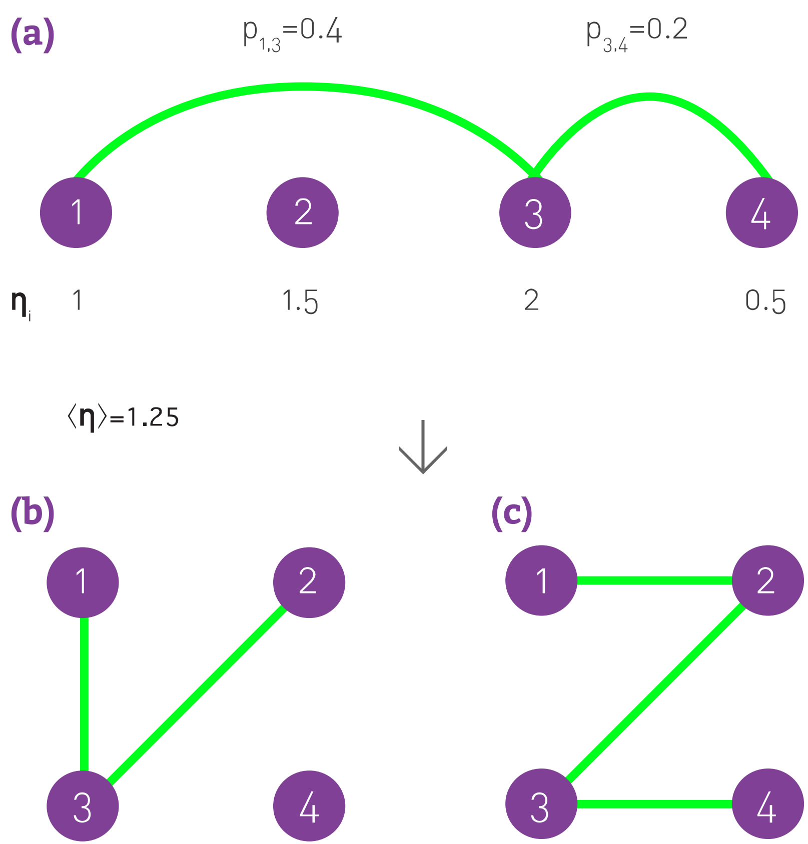
(1) ***Degree sequence****: Assign a degree to each node, represented as stubs or half-links. The degree sequence is either generated analytically from a preselected distribution (Box 4.5), or it is extracted from the adjacency matrix of a real network. We must start from an even number of stubs, otherwise we will be left with unpaired stubs. (2)* ***Network assembly****: Randomly select a stub pair and connect them. Then randomly choose another pair from the remaining stubs and connect them. This procedure is repeated until all stubs are paired up. Depending on the order in which the stubs were chosen, we obtain different networks. Some networks include cycles (2a), others self-edges (2b) or multi-edges (2c). Yet, the expected number of self- and multi-edges goes to zero in the limit.*

$$p_{ij} = \frac{k_i k_j}{2L - 1}$$

Degree Preserving randomization



Hidden parameter model



$$p(\eta_i, \eta_j) = \frac{\eta_i \eta_j}{\langle \eta \rangle N}$$

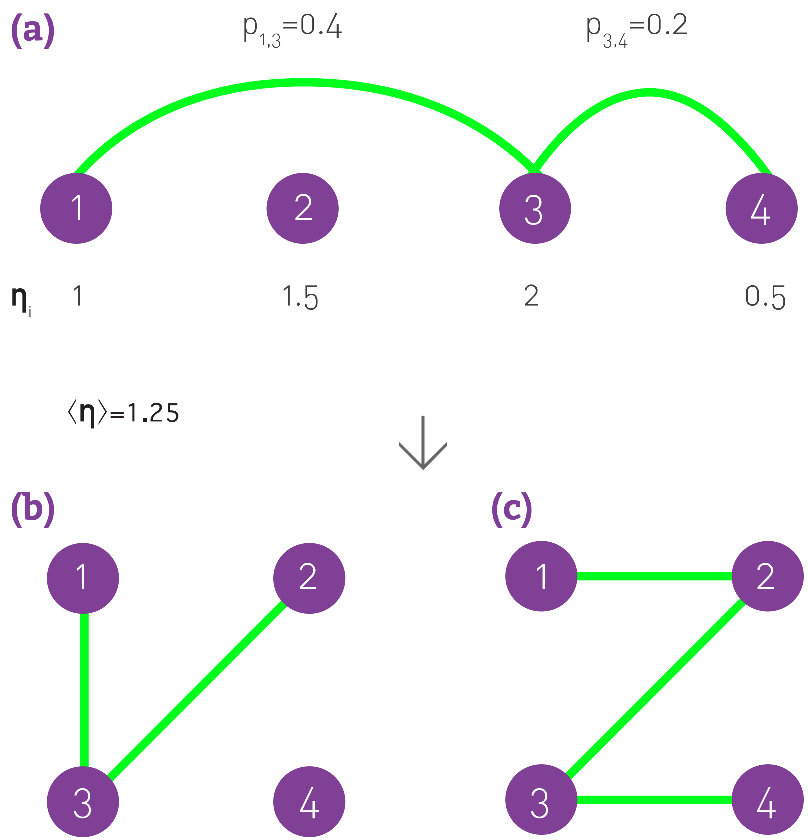
$$p_k = \int \frac{e^{-\eta}\eta^k}{k!} p(\eta) d\eta.$$

$$\{\eta_1, \eta_2, \ldots, \eta_N\}.$$

$$p_k = \frac{1}{N} \sum_j \frac{e^{-\eta_j} \eta_j^{\,k}}{k!}.$$

$$\eta_j = \frac{c}{i^{\alpha}}, i = 1, \ldots, N \qquad p_k \sim k^{-(1+\frac{1}{\alpha})}$$

Hidden parameter model



Start with N isolated nodes and assign to each node a "hidden parameter" η , which can be randomly selected from a ρ(η) distribution. We next connect each node pair with probability

$$p(\eta_i, \eta_j) = \frac{\eta_i \eta_j}{\langle \eta \rangle N}$$

For example, the figure shows the probability to connect nodes (1,3) and (3,4). After connecting the nodes, we end up with

the networks shown in (b) or (c), representing two independent realizations generated by the same hidden parameter sequence (a). The expected number of links in the obtained network is

$$L = \frac{1}{2} \sum_{N}^{i,j} \frac{\eta_i \eta_j}{\langle \eta \rangle N} = \frac{1}{2} \langle \eta \rangle N$$

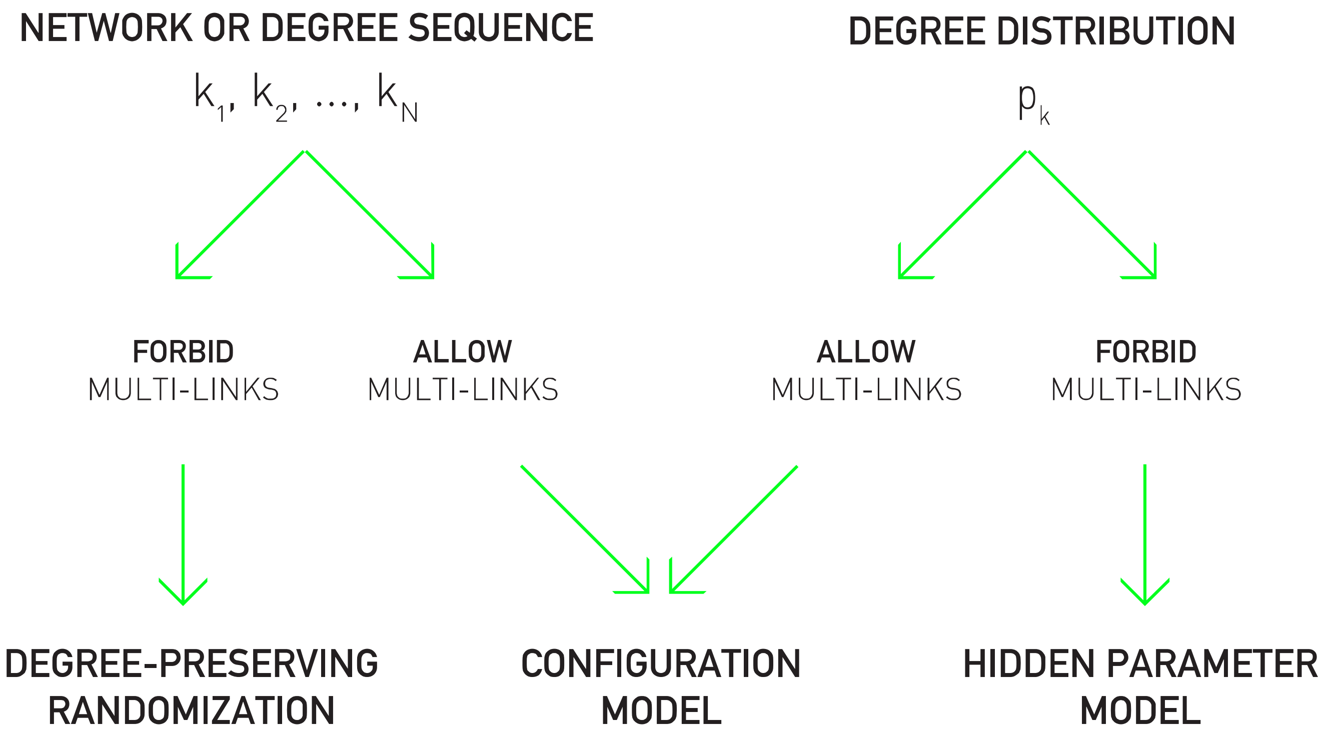
$$\eta_j = \frac{c}{i^{\alpha}}, i = 1, \ldots, N$$

$$p_k = \int \frac{e^{-\eta} \eta^k}{k!} p(\eta) d\eta.$$

$\{\eta_1, \eta_2, \ldots, \eta_N\}$. $p_k = \frac{1}{N} \sum_j \frac{e^{-\eta_j} \eta_j^{\,k}}{k!}.$

$$p_k \sim k^{-(1+\frac{1}{\alpha})}$$

Decision tree



summary

DEGREE DISTRIBUTION

Discrete form:

$$p_k = \frac{k^{-\gamma}}{\zeta(\gamma)}.$$

Continuous form:

$$p(k) = (\gamma - 1) k_{min}^{\gamma-1} k^{-\gamma}.$$

SIZE OF THE LARGEST HUB

$$k_{max} \sim k_{min} N^{\frac{1}{\gamma-1}}.$$

MOMENTS OF p_k *for* $N \to \infty$

$2 < \gamma < 3$: $\langle k \rangle$ finite, $\langle k^2 \rangle$ diverges.

$\gamma > 3$: $\langle k \rangle$ and $\langle k^2 \rangle$ finite.

DISTANCES

$$\langle d \rangle \sim \begin{cases} \text{const.} & \gamma = 2, \\ \dfrac{\ln \ln N}{\ln(\gamma - 1)} & 2 < \gamma < 3, \\ \dfrac{\ln N}{\ln \ln N} & \gamma = 3, \\ \ln N & \gamma > 3. \end{cases}$$

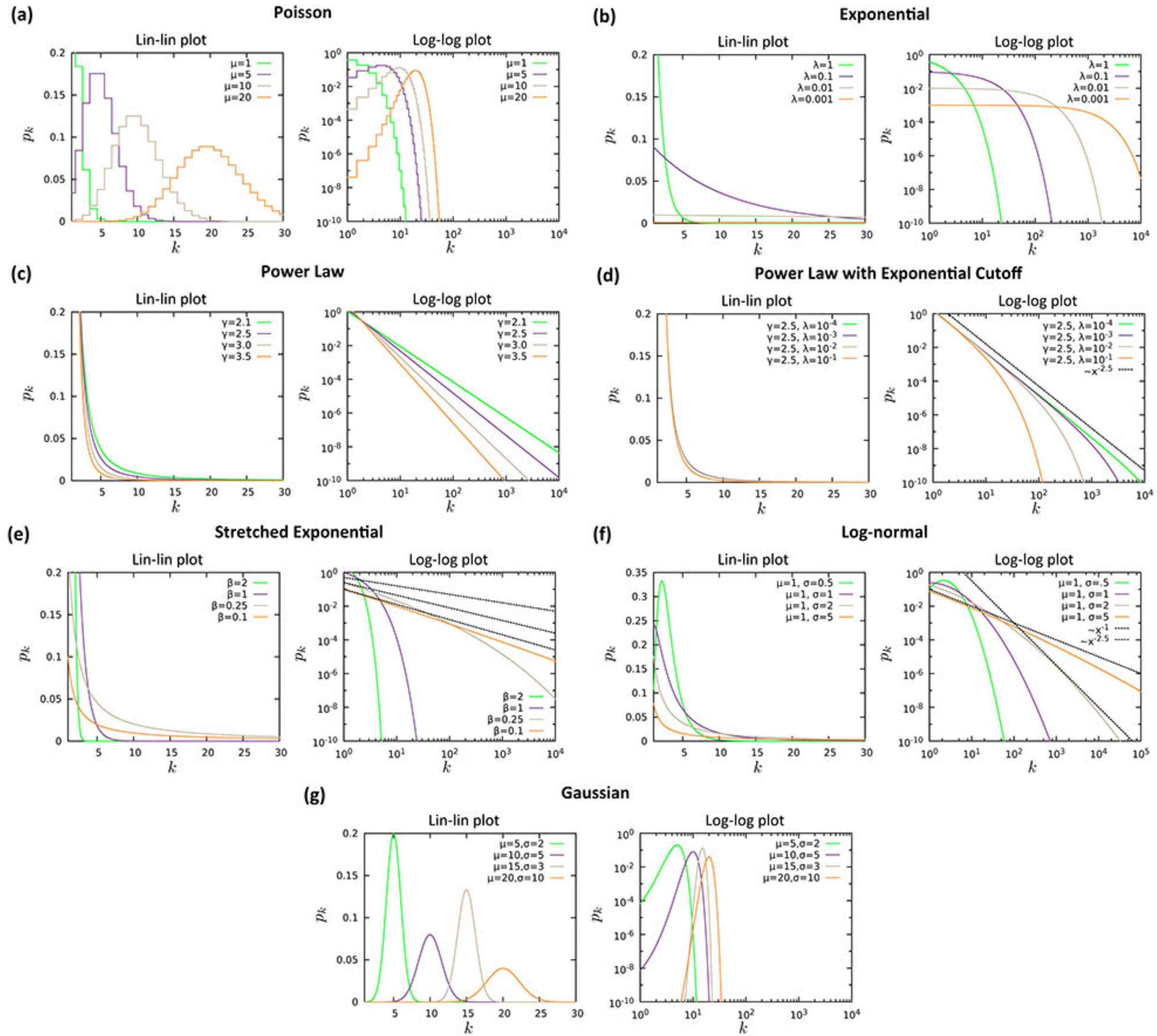
Bounded Networks

We call a network *bounded* if its degree distribution decrease exponentially or faster for high k. As a consequence $<k^2>$ is smaller than $<k>$, implying that we lack significant degree variations. Examples of p_k in this class include the Poisson, Gaussian, or the simple exponential distribution (Table 4.2). The Erdős-Rényi and the Watts-Strogatz networks are the best known network models belonging to this class. Bounded networks lack outliers, consequently most nodes have comparable degrees. Real networks in this class include highway networks and the power grid.

Unbounded Networks

We call a network *unbounded* if its degree distribution has a fat tail in the high-k region. As a consequence $<k^2>$ is much larger than $<k>$, resulting in considerable degree variations. Scale-free networks with a power-law degree distribution (4.1) offer the best known example of networks belonging to this class. Outliers, or exceptionally high-degree nodes, are not only allowed but are expected in these networks. Networks in this class include the WWW, the Internet, the protein interaction networks, and most social and online networks.

Random Distributions Important for Network Science



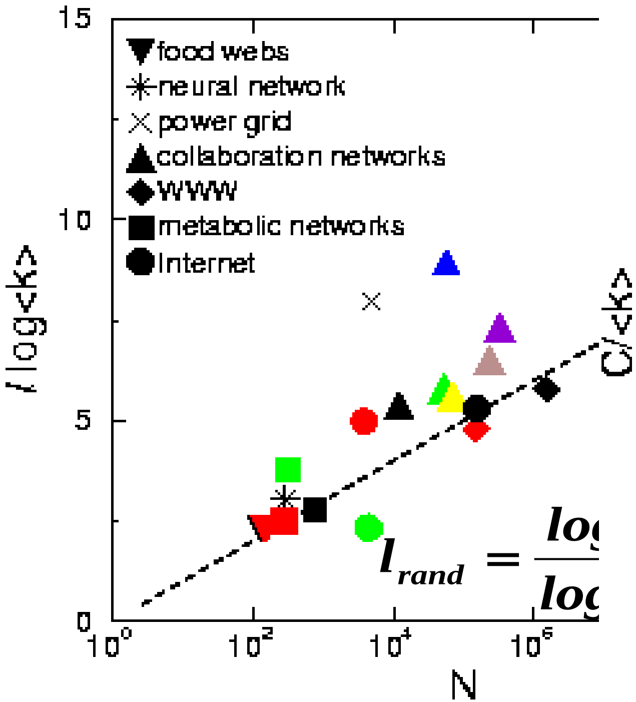
Introduction to BA Model

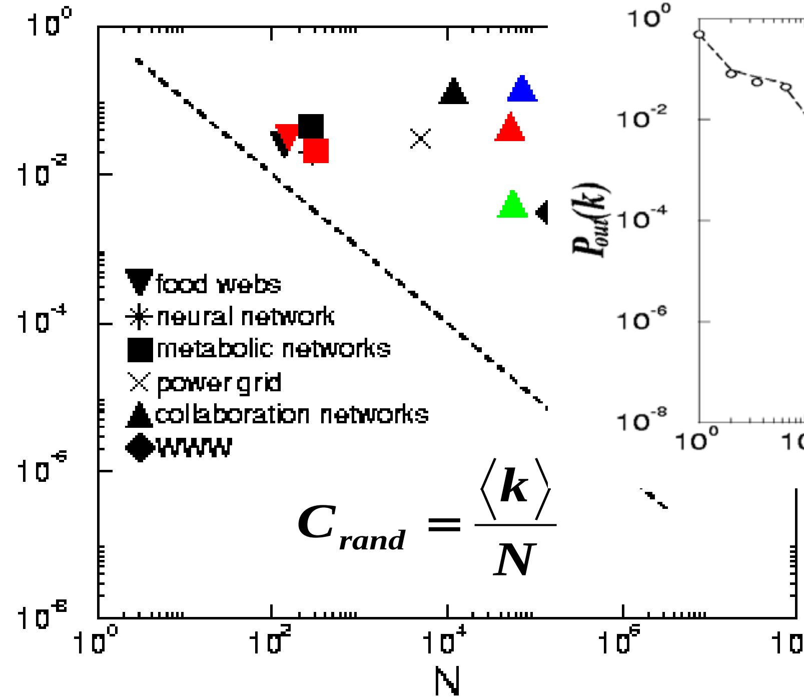
Hubs represent the most striking difference between a random and a scale-free network. Their emergence in many real systems raises several fundamental questions:

- Why does the random network model of Erdős and Rényi fail to reproduce the hubs and the power laws observed in many real networks?

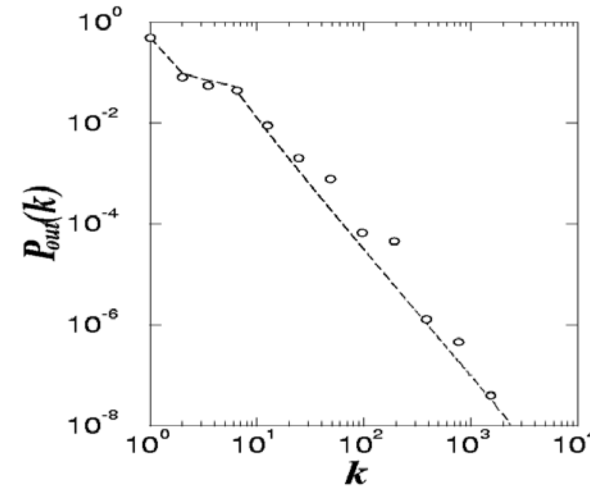
- Why do so different systems as the WWW or the cell converge to a similar scale-free architecture?

Empirical findings for real networks





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



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

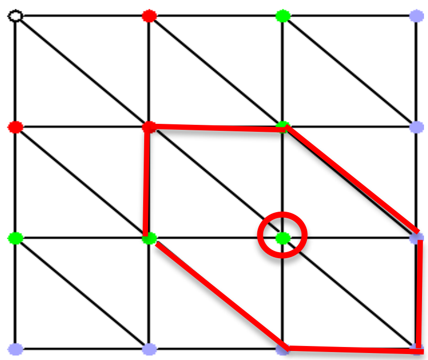
Small World:
distances scale logarithmically with the network size

Clustered:
clustering coefficient does not depend on network size.

Scale-free:
The degrees follow a power-laws distribution.

BENCHMARK 1: Regular Lattices

Two-dimensional lattice:



Average path-length: $l \approx L \approx N^{1/2}$

Degree distribution: $P(k)=\delta(k-6)$ excluding corners and boundaries

Clustering coefficient: $C = \frac{2}{5}$ 6 neighbors, each with 2 edges = 12/30

D-dimensional lattice:

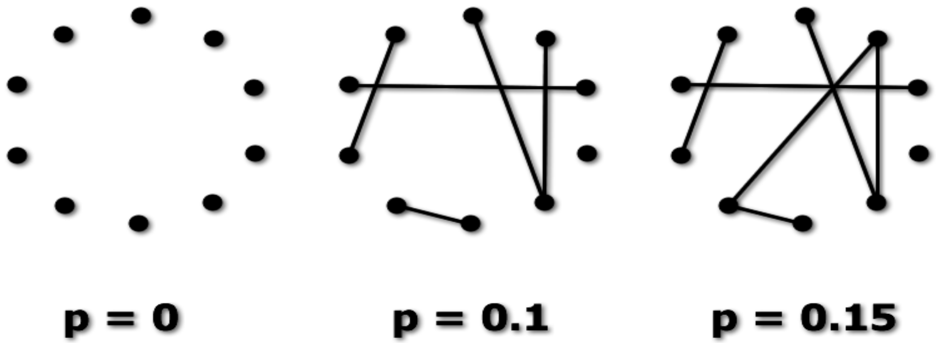
The average path-length varies as $l \approx N^{1/D}$

Constant degree $P(k)=\delta(k-k_d)$

Constant clustering coefficient $C=C_d$

BENCHMARK 2: Random Network Model

Erdös-Rényi Model- Publ. Math. Debrecen 6, 290 (1959)



- fixed node number N
- connecting pairs of nodes with probability p

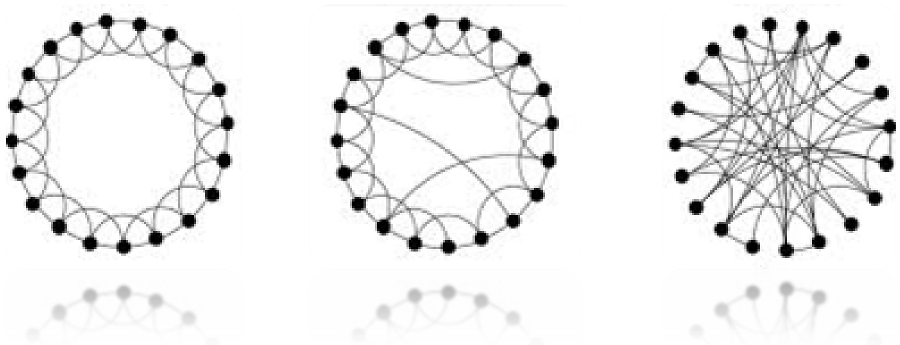
Degree distribution: $$P_{rand}(k) \cong C_{N-1}^{k} p^{k} (1-p)^{N-1-k}$$ 

Path length: $$l_{rand} \approx \frac{\log N}{\log\langle k \rangle}$$ 

Clustering coefficient: $$C_{rand} = p = \frac{\langle k \rangle}{N}$$ 

BENCHMARK 3: Small World Model

Watts-Strogatz algorithm – Nature 2008



• For fixed node number *N, first connect them into even number, k, degree ring in which k/2 nearest neighbors on each side of each node are connected to it*

• Then, with probability *p* re-wire ring edges of each node to nodes not currently connected to and different from it

Degree distribution: Exponential 

Path length: $$l_{rand} \approx \frac{\log N}{\log\langle k \rangle}$$ 

Clustering coefficient: $$C(\beta) = C(0)(1-\beta)$$ 

Growth and preferential attachment

Growth

ER, WS models: the number of nodes, N, is fixed (static models)

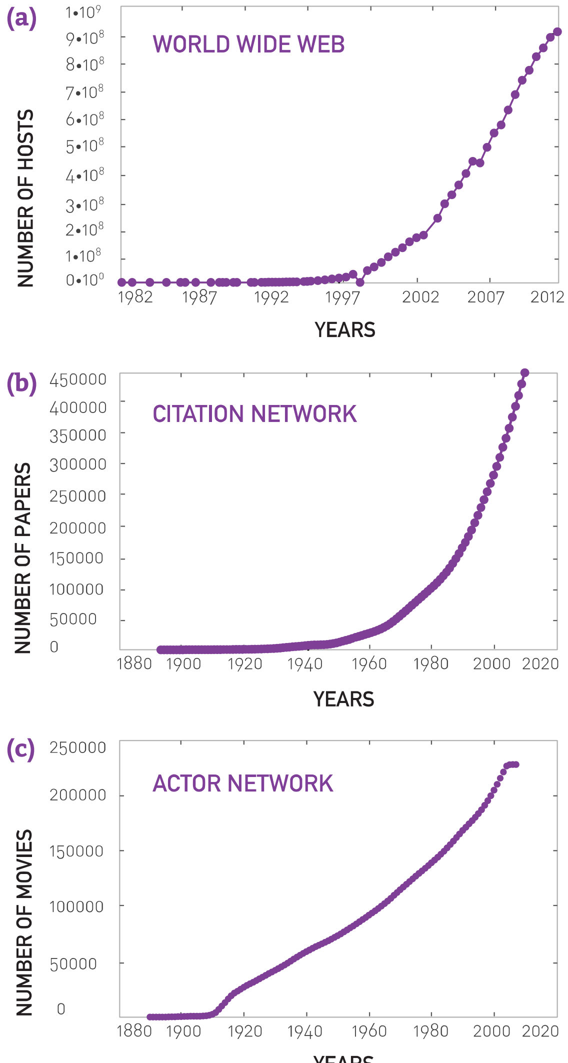
Real networks continuously expand by the addition of new nodes

Barabási & Albert, *Science* **286,** 509 (1999)

Growth

ER model:
the number of nodes, N, is fixed (static models)

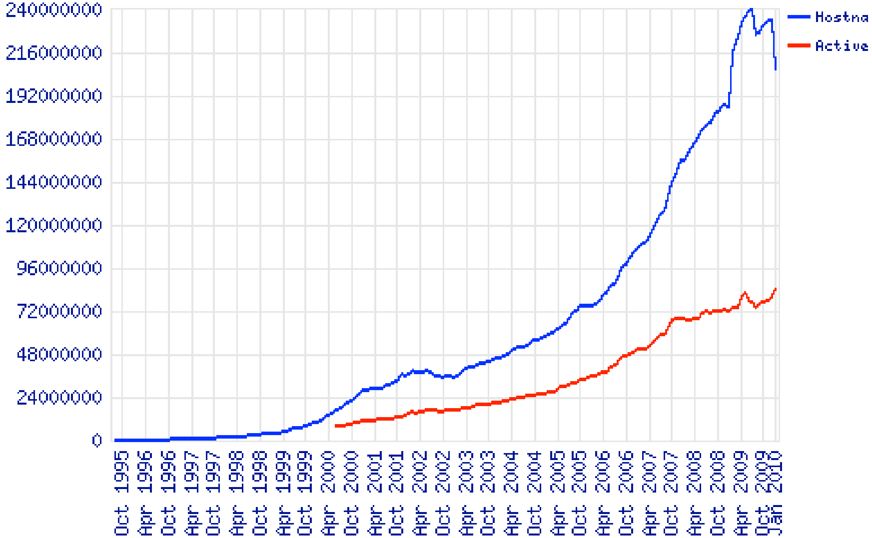
networks expand through the addition of new nodes



Barabási & Albert, *Science* **286,** 509 (1999)

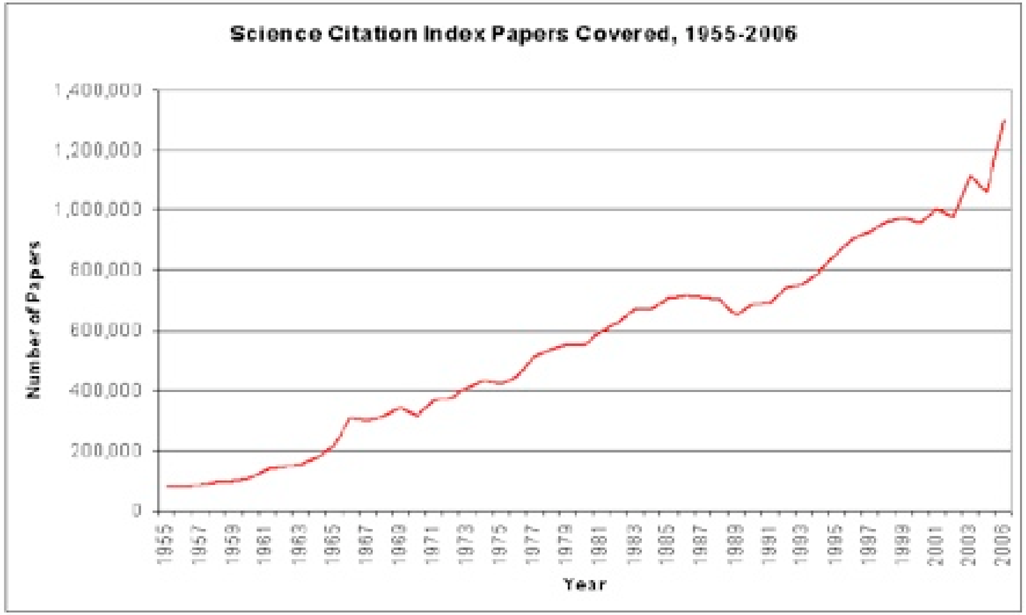
Growth (www/Pubs)

WWW



http://website101.com/define-ecommerce-web-terms-definitions/

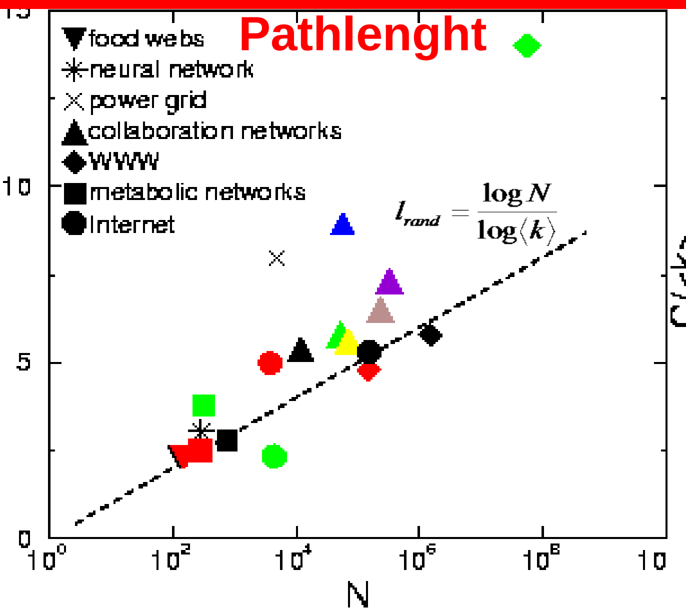
Scientific Publications

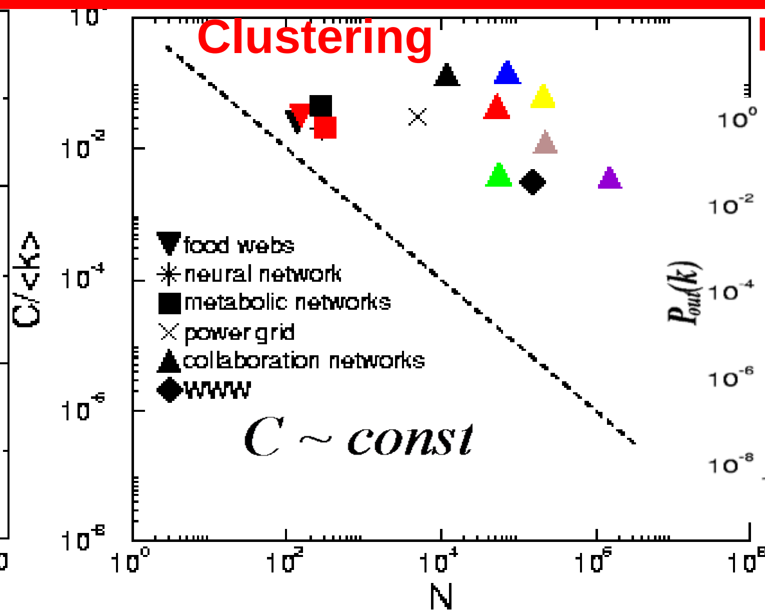


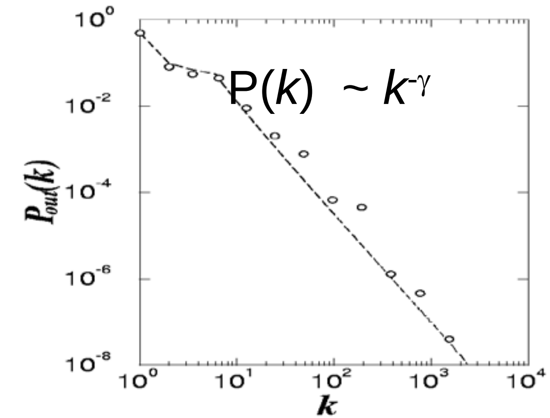
http://www.kk.org/thetechnium/archives/2008/10/the_expansion_o.php

Barabási & Albert, *Science* **286,** 509 (1999)

EMPIRICAL DATA FOR REAL NETWORKS







Regular network	$l \approx N^{1/D}$	$C \sim const$	P(k)=δ(k-k$_d$)
Erdos-Renyi	$l_{rand} \approx \frac{\log N}{\log\langle k\rangle}$	$C_{rand} = p = \frac{\langle k\rangle}{N}$	$P(k) = e^{-\langle k\rangle}\frac{\langle k\rangle^k}{k!}$
Watts-Strogatz	$l_{rand} \approx \frac{\log N}{\log\langle k\rangle}$	$C \sim const$	*Exponential*

Preferential attachment

ER model: links are added randomly to the network

New nodes prefer to connect to the more connected nodes

Barabási & Albert, *Science* **286,** 509 (1999)

The Barabási-Albert model

Section 2: Growth and Preferential Sttachment

The random network model differs from real networks in two important characteristics:

Growth: While the random network model assumes that the number of nodes is fixed (time invariant), real networks are the result of a growth process that continuously increases.

Preferential Attachment: While nodes in random networks randomly choose their interaction partner, in real networks new nodes prefer to link to the more connected nodes.

Barabási & Albert, *Science* **286,** 509 (1999)

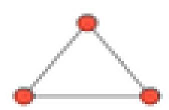
Origin of SF networks: Growth and preferential attachment

(1) Networks continuously expand by the addition of new nodes

WWW : addition of new documents

(2) New nodes prefer to link to highly connected nodes.

WWW : linking to well known sites



Barabási & Albert, *Science* **286,** 509 (1999)

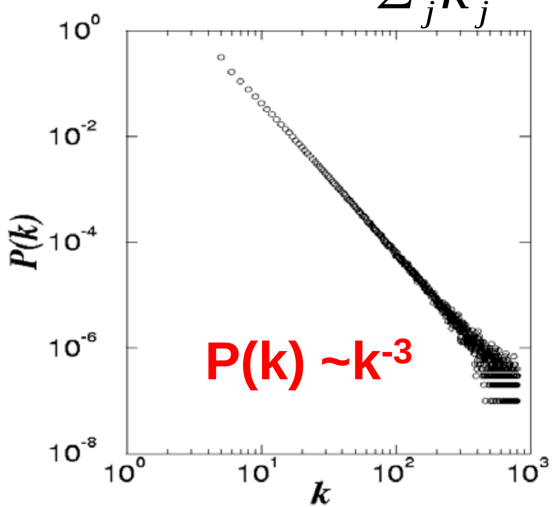
GROWTH:

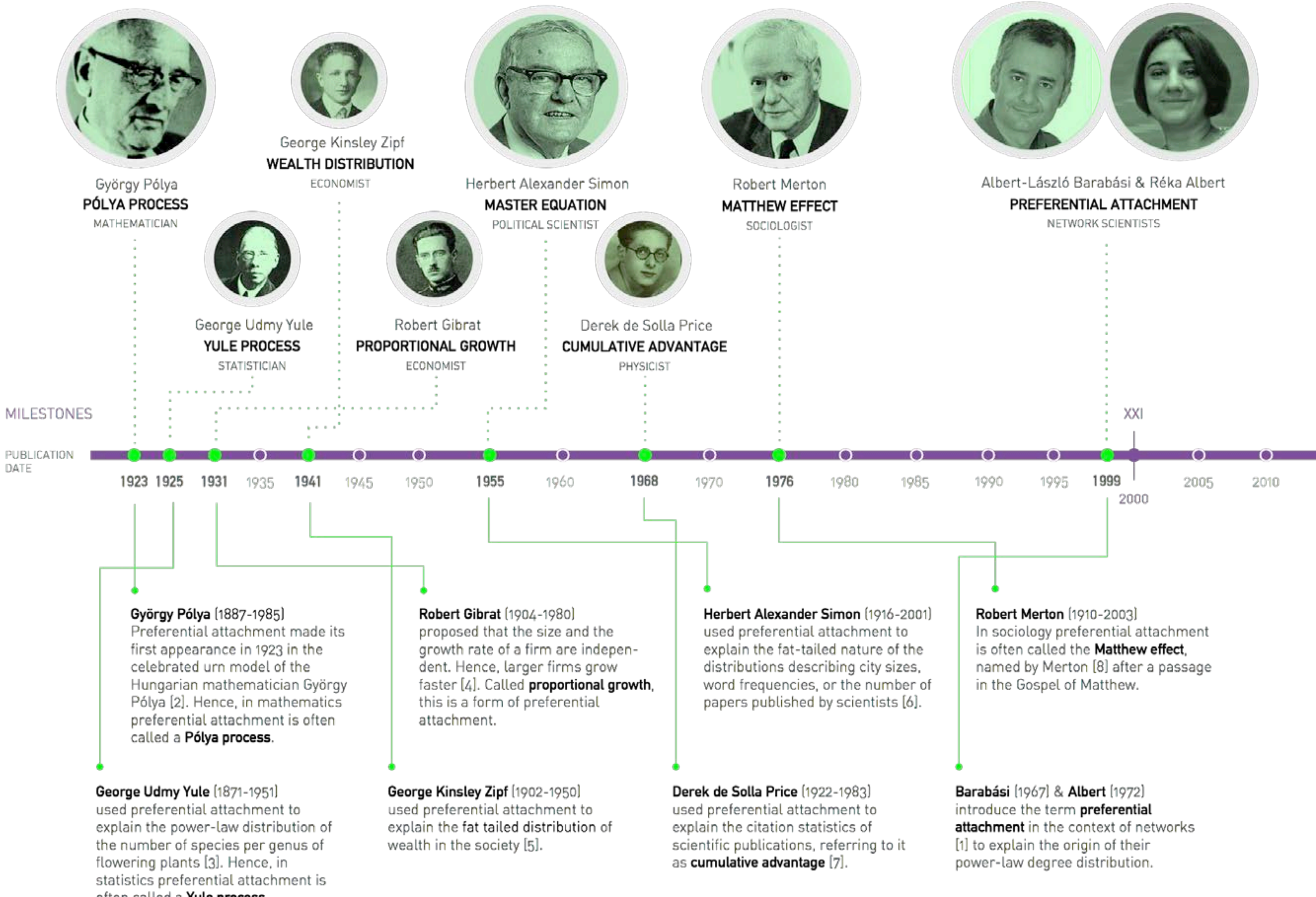
add a new node with m links

PREFERENTIAL ATTACHMENT:

the probability that a node connects to a node with k links is proportional to k.

$$\Pi(k_i) = \frac{k_i}{\Sigma_j k_j}$$





György Pólya
PÓLYA PROCESS
MATHEMATICIAN
George Kinsley Zipf
WEALTH DISTRIBUTION
ECONOMIST
George Udmy Yule
YULE PROCESS
STATISTICIAN
Robert Gibrat
PROPORTIONAL GROWTH
ECONOMIST
Herbert Alexander Simon
MASTER EQUATION
POLITICAL SCIENTIST
Derek de Solla Price
CUMULATIVE ADVANTAGE
PHYSICIST
Robert Merton
MATTHEW EFFECT
SOCIOLOGIST
Albert-László Barabási & Réka Albert
PREFERENTIAL ATTACHMENT
NETWORK SCIENTISTS
MILESTONES
XXI
PUBLICATION DATE
1923
1925
1931
1935
1941
1945
1950
1955
1960
1968
1970
1976
1980
1985
1990
1995
1999
2000
2005
2010
György Pólya (1887-1985) Preferential attachment made its first appearance in 1923 in the celebrated urn model of the Hungarian mathematician György Pólya [2]. Hence, in mathematics preferential attachment is often called a **Pólya process**.
Robert Gibrat (1904-1980) proposed that the size and the growth rate of a firm are independent. Hence, larger firms grow faster [4]. Called **proportional growth**, this is a form of preferential attachment.
Herbert Alexander Simon (1916-2001) used preferential attachment to explain the fat-tailed nature of the distributions describing city sizes, word frequencies, or the number of papers published by scientists [6].
Robert Merton (1910-2003) In sociology preferential attachment is often called the **Matthew effect**, named by Merton [8] after a passage in the Gospel of Matthew.
George Udmy Yule (1871-1951) used preferential attachment to explain the power-law distribution of the number of species per genus of flowering plants [3]. Hence, in statistics preferential attachment is often called a **Yule process**.
George Kinsley Zipf (1902-1950) used preferential attachment to explain the fat tailed distribution of wealth in the society [5].
Derek de Solla Price (1922-1983) used preferential attachment to explain the citation statistics of scientific publications, referring to it as **cumulative advantage** [7].
Barabási (1967) & **Albert** (1972) introduce the term **preferential attachment** in the context of networks [1] to explain the origin of their power-law degree distribution.

THE END