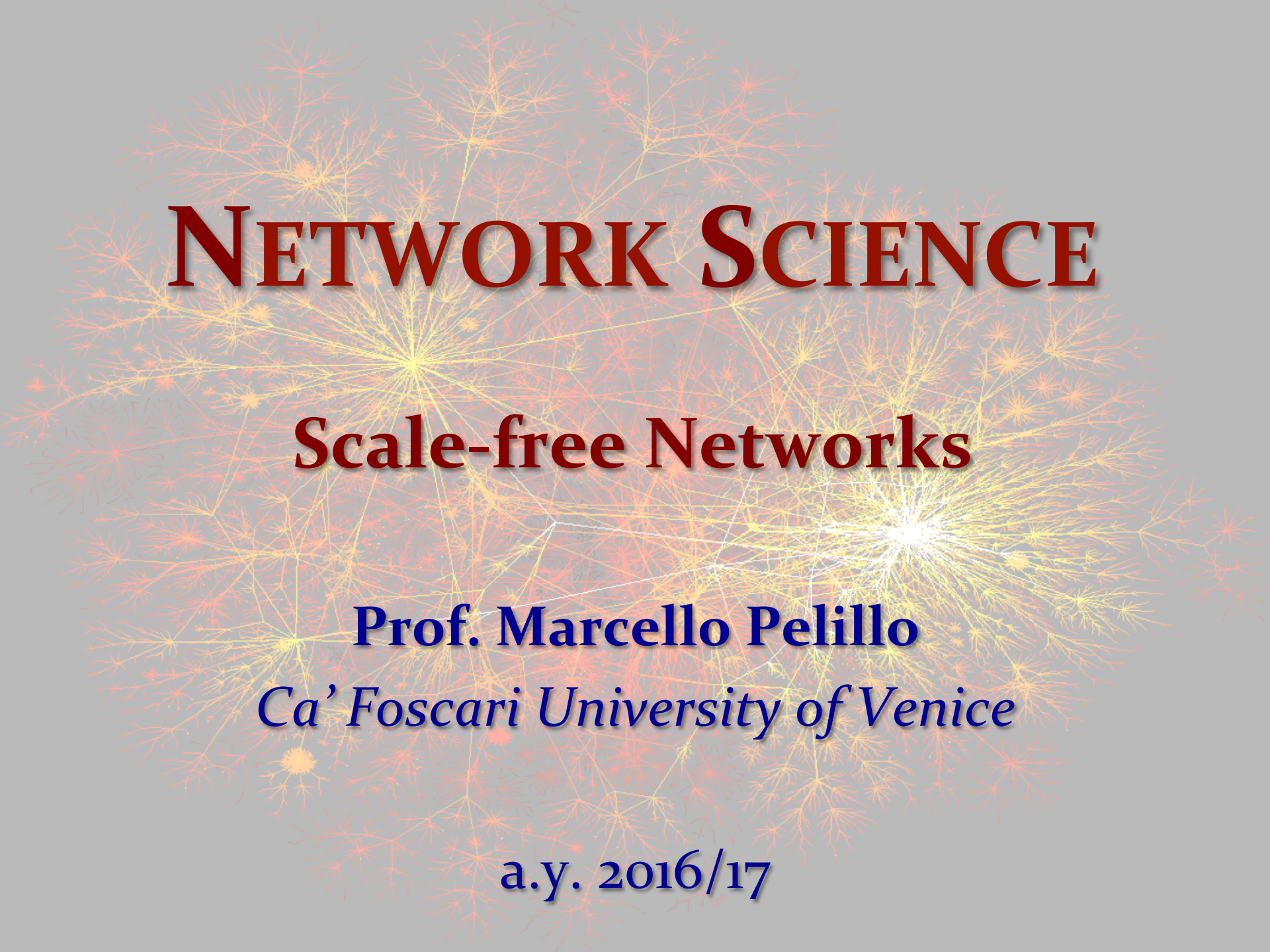


NETWORK SCIENCE

Scale-free Networks

Prof. Marcello Pelillo

Ca' Foscari University of Venice

a.y. 2016/17

The power law distribution: Discrete vs. Continuous formalism

Discrete Formalism

As node degrees are always positive integers, the discrete formalism captures the probability that a node has exactly k links:

$$p_k = Ck^{-\gamma}.$$

$$\sum_{k=1}^{\infty} p_k = 1.$$

$$C\sum_{k=1}^{\infty} k^{-\gamma} = 1 \qquad C = \frac{1}{\sum_{k=1}^{\infty} k^{-\gamma}} = \frac{1}{\zeta(\gamma)},$$

Riemann-Zeta function

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

INTERPRETATION: p_k

Continuous Formalism

In analytical calculations it is often convenient to assume that the degrees can take up any positive real value:

$$p(k) = Ck^{-\gamma}.$$

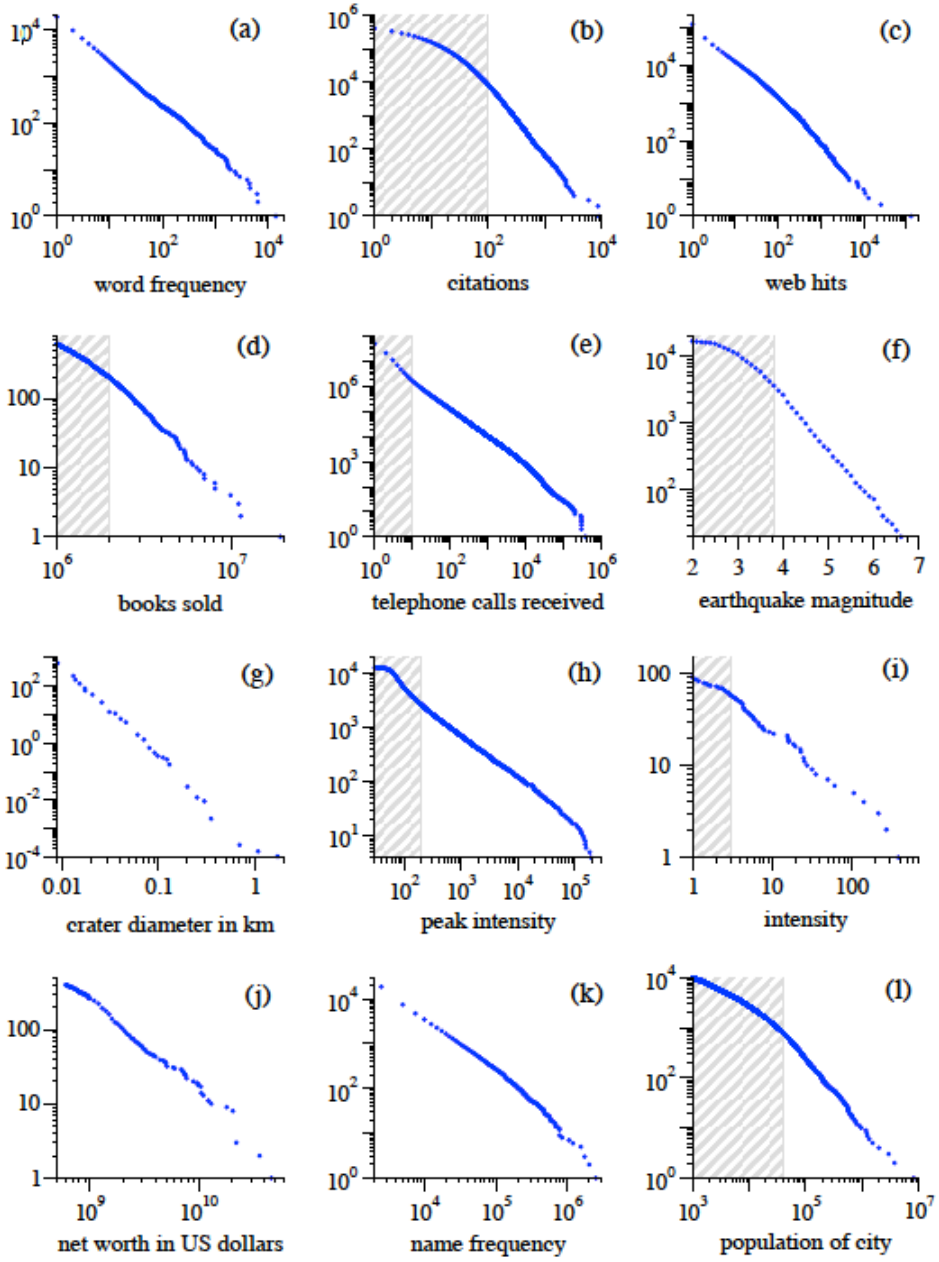
$$\int_{k_{\min}}^{\infty} p(k)dk = 1$$

$$C = \frac{1}{\int_{k_{\min}}^{\infty} k^{-\gamma}dk} = (\gamma - 1)k_{\min}^{\gamma-1}$$

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

$$\int_{k_1}^{k_2} p(k)dk$$

Power law



a) Numbers of occurrences of words in the novel Moby Dick by Hermann Melville.

b) Numbers of citations to scientic papers published in 1981, from time of publication until June 1997.

c) Numbers of hits on web sites by 60000 users of the America Online Internet service for the day of 1 December 1997.

d) Numbers of copies of bestselling books sold in the US between 1895 and 1965.

e) Number of calls received by AT&T telephone customers in the US for a single day.

f) Magnitude of earthquakes in California between January 1910 and May 1992. Magnitude is proportional to the logarithm of the maximum amplitude of the earthquake, and hence the distribution obeys a power law even though the horizontal axis is linear.

g) Diameter of craters on the moon. Vertical axis is measured per square kilometre.

h) Peak gamma-ray intensity of solar ares in counts per second, measured from Earth orbit between February 1980 and November 1989.

i) Intensity of wars from 1816 to 1980, measured as battle deaths per 10 000 of the population of the participating countries.

j) Aggregate net worth in dollars of the richest individuals in the US in October 2003.

k) Frequency of occurrence of family names in the US in the year 1990.

l) Populations of US cities in the year 2000.

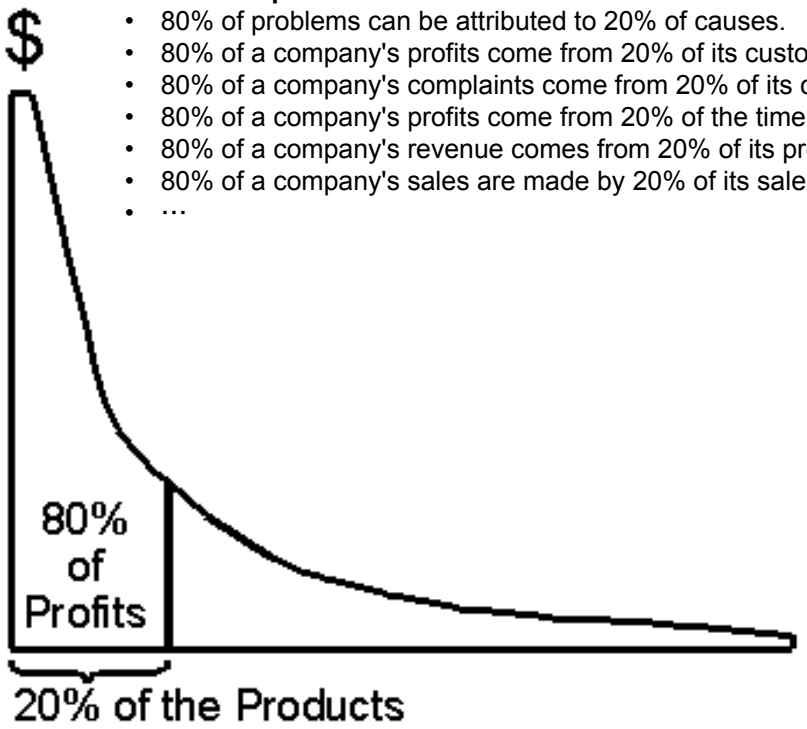
From: Newman 2006

The 80/20 rule

“80% of the wealthis in the hands of the richest 20% of people.”

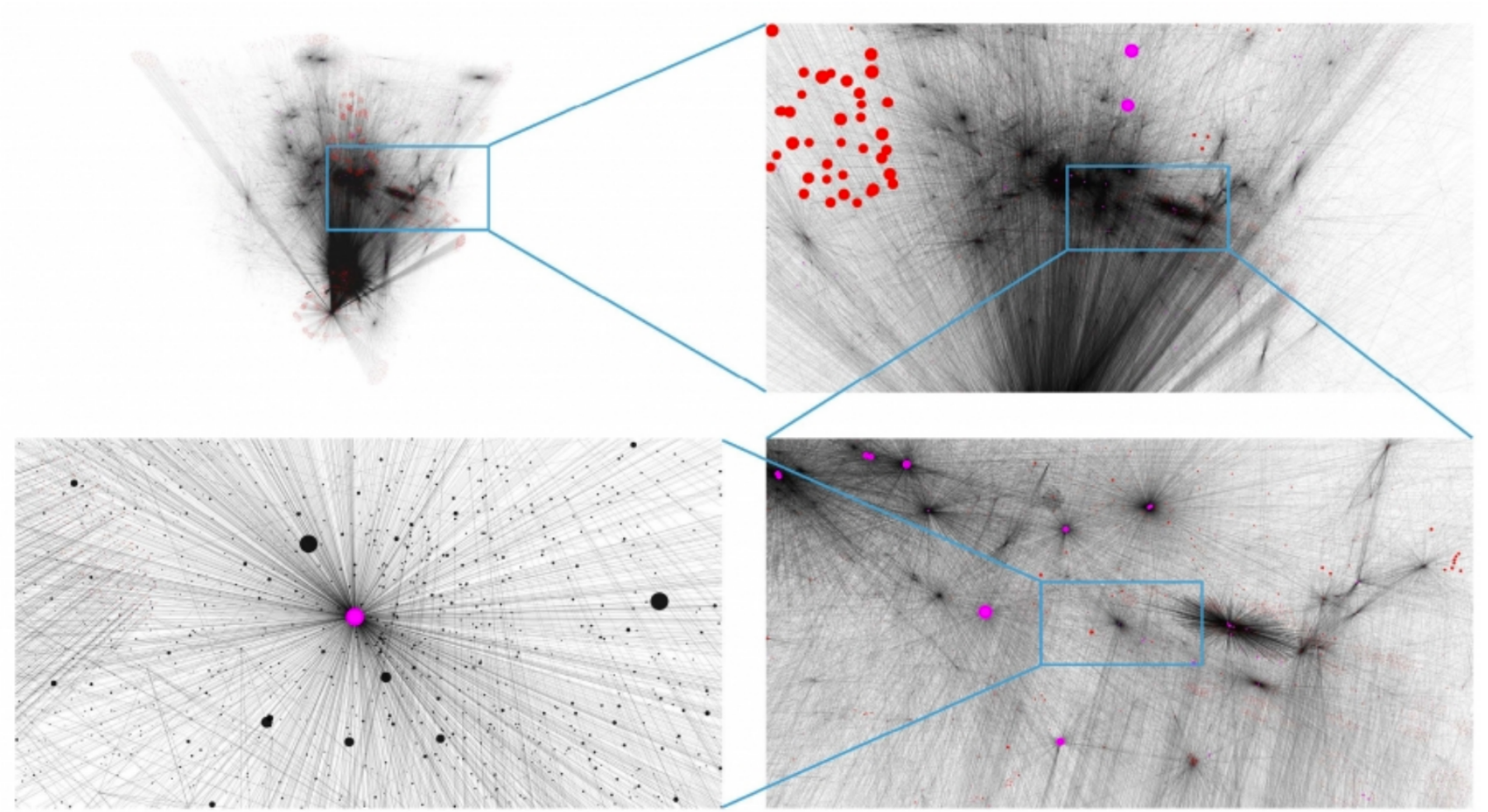
Other examples

- 80% of problems can be attributed to 20% of causes.
- 80% of a company's profits come from 20% of its customers
- 80% of a company's complaints come from 20% of its customers
- 80% of a company's profits come from 20% of the time its staff spent
- 80% of a company's revenue comes from 20% of its products
- 80% of a company's sales are made by 20% of its sales staff
- …



Vilfredo Pareto (1848 – 1923), Italian economist, political scientist and philosopher, who had important contributions to our understanding of income distribution and to the analysis of individuals choices. A number of fundamental principles are named after him, like Pareto efficiency, Pareto distribution (another name for a power-law distribution), the Pareto principle (or 80/20 law).

WORLD WIDE WEB



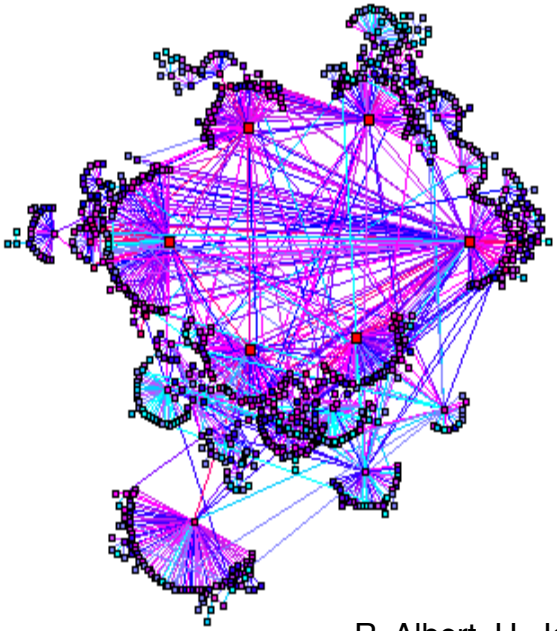
Snapshots of the World Wide Web sample mapped out by Hawoong Jeong in 1998 [1]. The sequence of images show an increasingly magnified local region of the network. The first panel displays all 325,729 nodes, offering a global view of the full dataset. Nodes with more than 50 links are shown in red and nodes with more than 500 links in purple. The closeups reveal the presence of a few highly connected nodes, called hubs, that accompany scale-free networks.

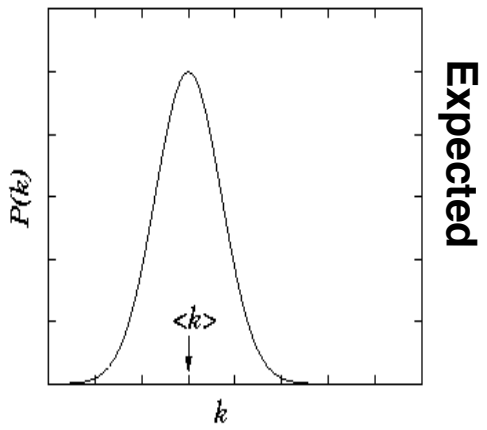
WORLD WIDE WEB

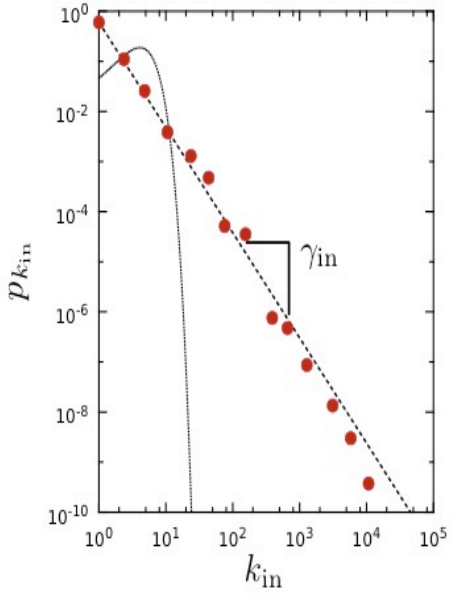
Nodes: **WWW documents**
Links: **URL links**

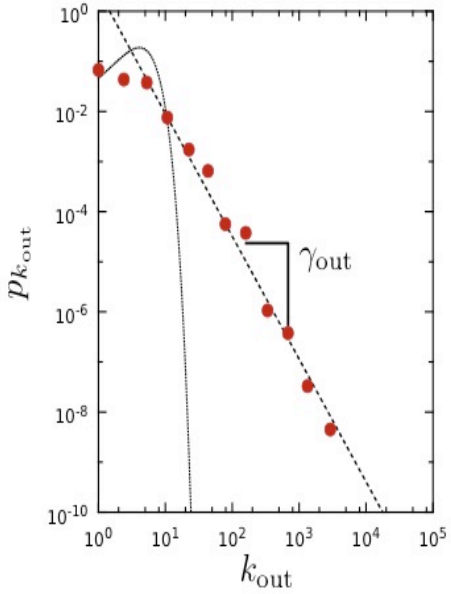
Over 3 billion documents

ROBOT: collects all URL's found in a document and follows them recursively





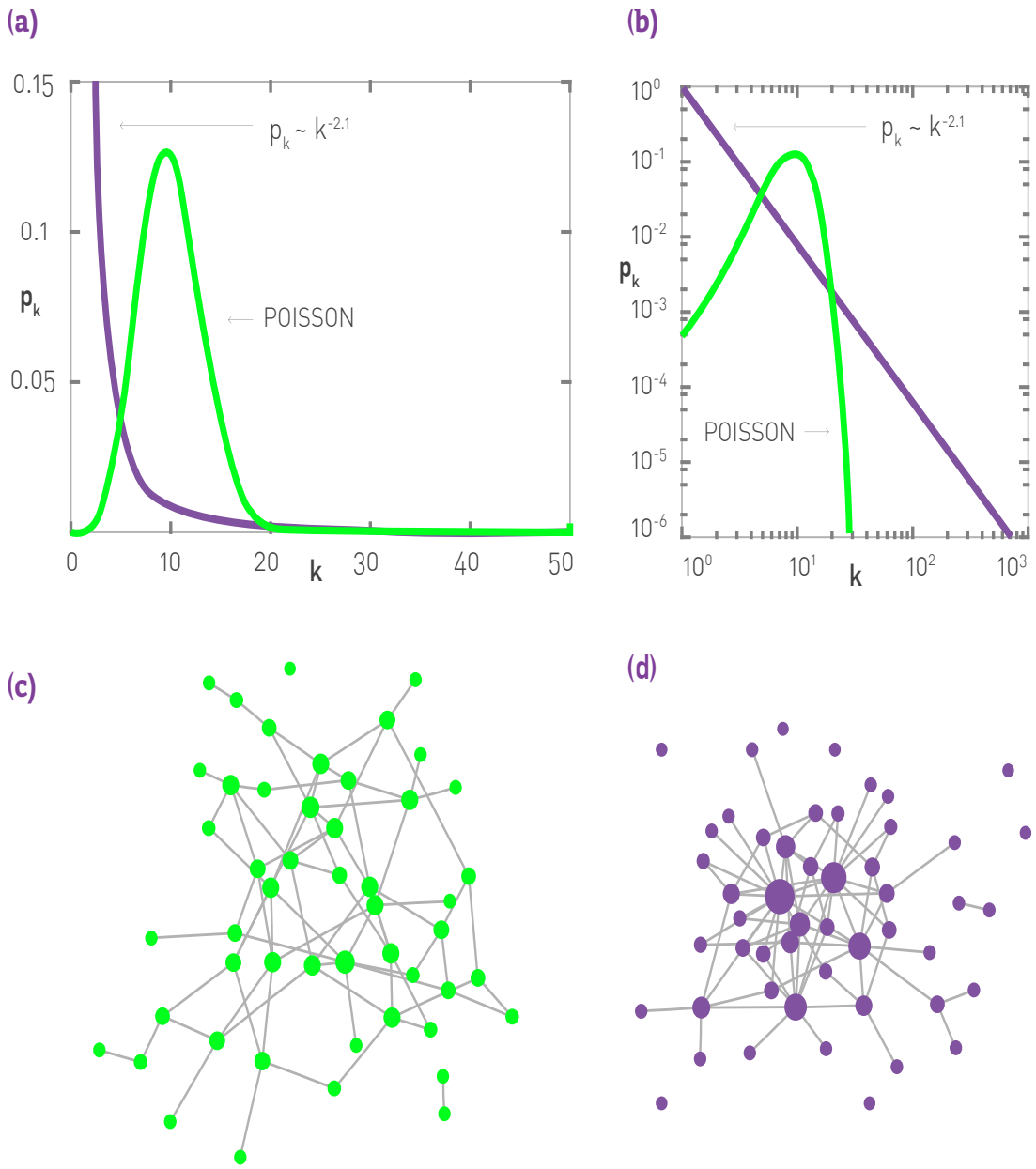




R. Albert, H. Jeong, A-L Barabasi, *Nature*, 401 130 (1999).

Hubs

The difference between a power law and an exponential distribution



The difference between a power law and an exponential distribution

Let us use the WWW to illustrate the properties of the high-*k regime.*
The probability to have a node with k~100 is

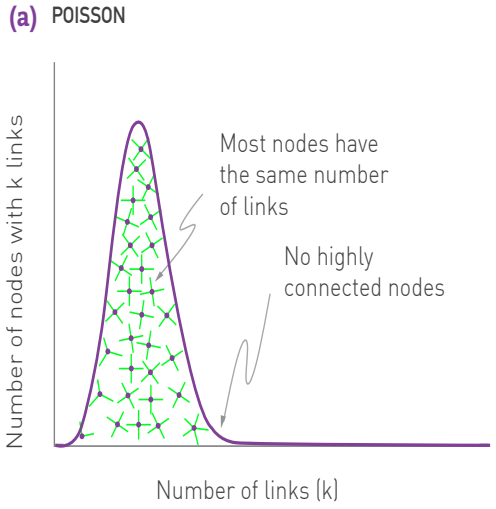
- *About* $p_{100} \simeq 10^{-30}$ *if* p_k *follows a Poisson distribution*

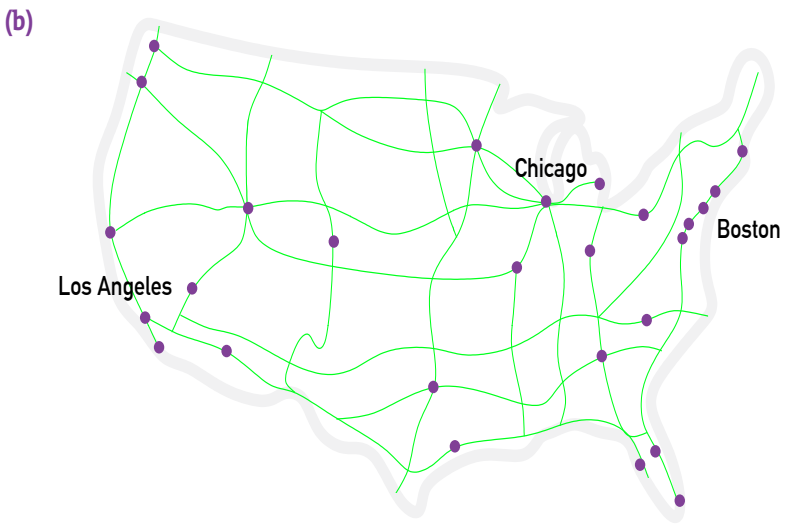
- *About* $p_{100} \simeq 10^{-4}$ *if* p_k *follows a power law.*

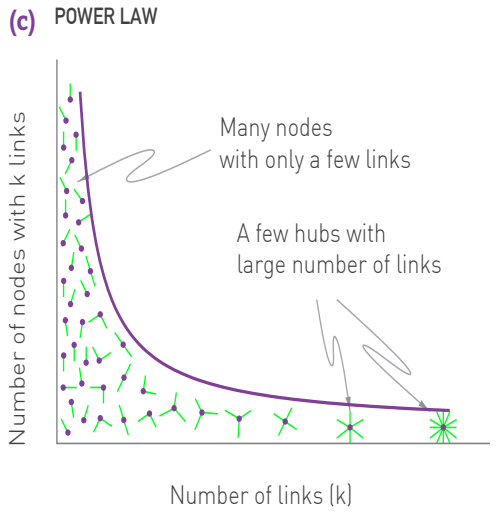
Consequently, if the WWW were to be a random network, according to the Poisson prediction we would expect 10^{-18} k>100 degree nodes, or none.

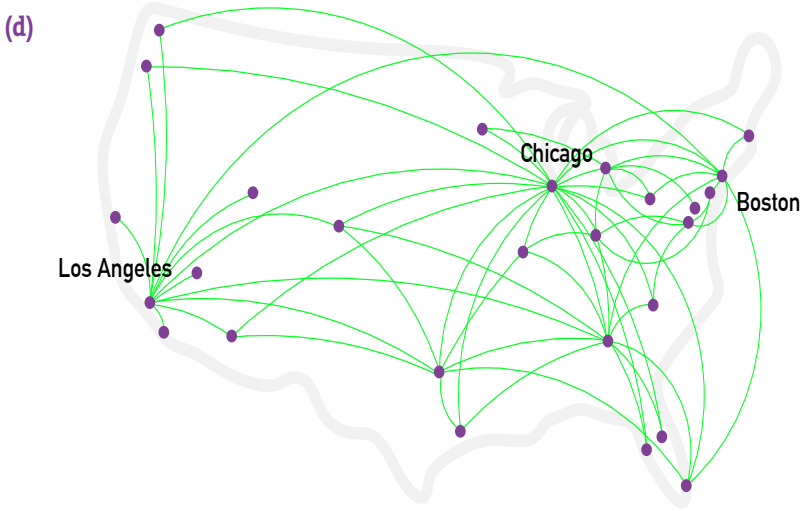
For a power law degree distribution, we expect about $N_{k>100} = 10^9$ k>100 degree nodes

The difference between a power law and an exponential distribution









The size of the largest hub

All real networks are finite → let us explore its consequences.
→ We have an expected maximum degree, k_{max}

Estimating k_{max}

$$\int_{k_{max}}^{\infty} P(k)dk \approx \frac{1}{N}$$

Why: we expect at most one node with degree > k_{max} (*natural upper cutoff*)

$$\int_{k_{max}}^{\infty} P(k)dk = (\gamma-1)k_{min}^{\gamma-1}\int_{k_{max}}^{\infty} k^{-\gamma}\,dk = \frac{(\gamma-1)}{(-\gamma+1)}k_{min}^{\gamma-1}\left[k^{-\gamma+1}\right]_{k_{max}}^{\infty} = \frac{k_{min}^{\gamma-1}}{k_{max}^{\gamma-1}} \approx \frac{1}{N}$$

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

The size of the largest hub

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

To illustrate the difference in the maximum degree of an exponential and a scale-free network let us return to the WWW sample, consisting of $N \approx 3 \times 10^5$ nodes.

As $k_{min} = 1$, if the degree distribution were to follow an exponential, (4.17) predicts that the maximum degree should be $k_{max} \approx 14$ for $\lambda=1$. In a scale-free network of similar size and $\gamma = 2.1$, (4.18) predicts $k_{max} \approx 95{,}000$, a remarkable difference.

Note that the largest in-degree of the WWW map of Image 4.1 is 10,721, which is comparable to kmax predicted by a scale-free network.

This reinforces our conclusion that in a random network hubs are effectivelly forbidden, while in scale-free networks they are naturally present.

The size of the largest hub

Expected maximum degree, k_{max}

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

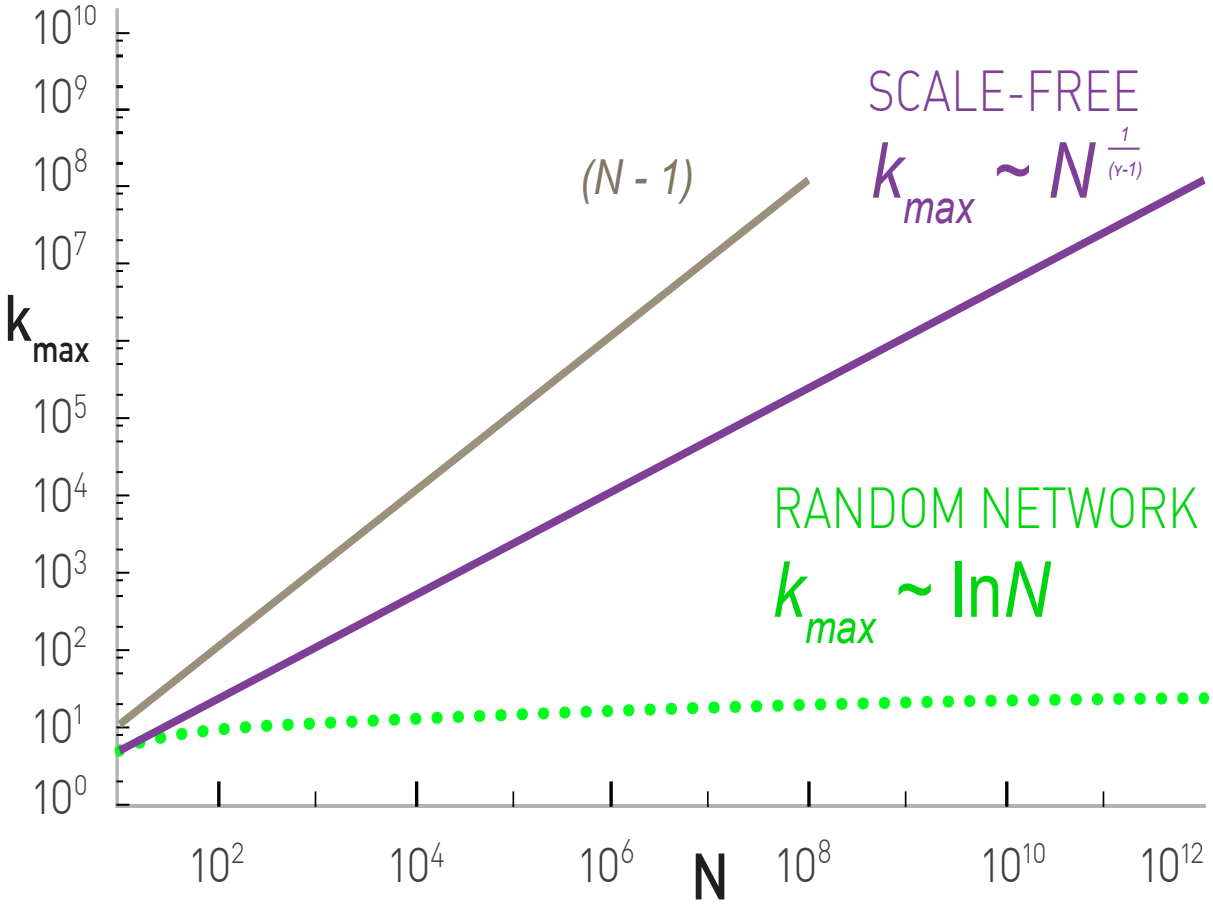
k_{max} increases with the size of the network
the larger a system is, the larger its biggest hub

$\gamma > 2$: k_{max} increases slower than N
The largest hub will contain a decreasing fraction of links as N increases.

$\gamma = 2$: $k_{max} \sim N$
The size of the biggest hub is O(N)

$\gamma < 2$: k_{max} increases faster than N: condensation phenomena
The largest hub will grab an increasing fraction of links. Anomaly!

The size of the largest hub



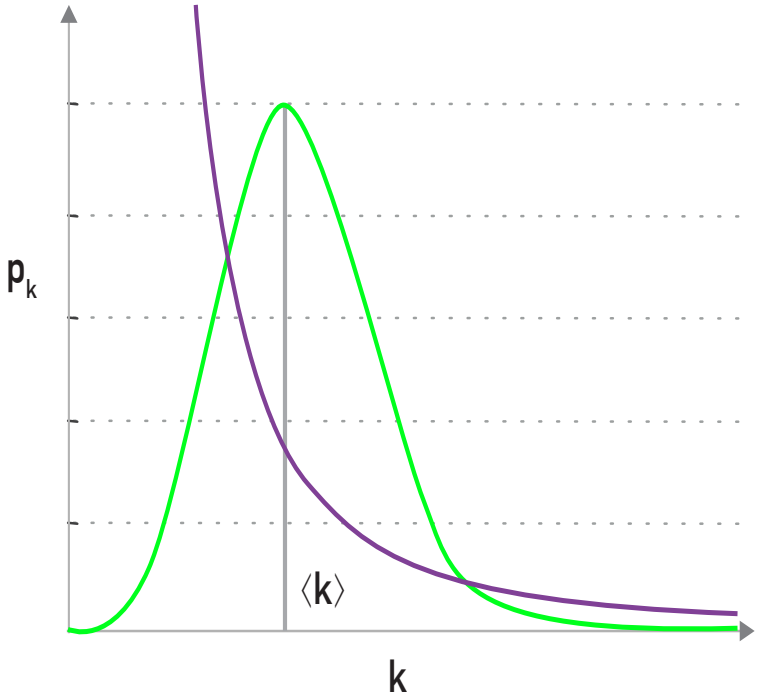
$$k_{\max} = k_{\min} N^{\frac{1}{\gamma-1}}$$

The estimated degree of the largest node in scale-free and random networks with the same average degree ⟨k⟩= 3.

For the scale-free network we chose γ = 2.5. For comparison, we also show the linear behavior, $k_{max} \sim N - 1$, expected for a complete network.

Overall, hubs in a scale-free network are several orders of magnitude larger than the biggest node in a random network with the same N and ⟨k⟩

The meaning of scale-free



Random Network

Randomly chosen node: $k = \langle k \rangle \pm \langle k \rangle^{1/2}$

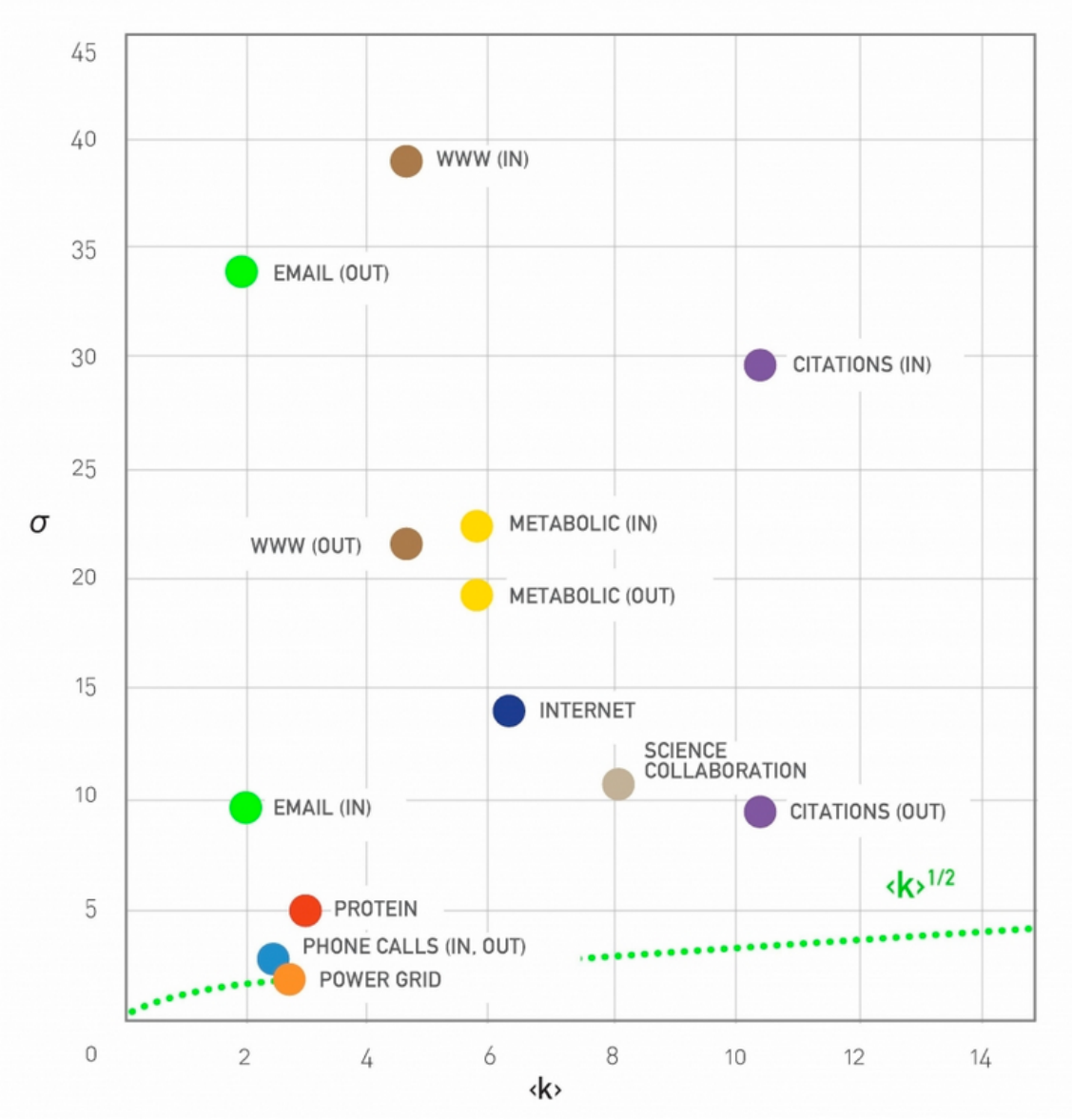
Scale: $\langle k \rangle$

Scale-Free Network

Randomly chosen node: $k = \langle k \rangle \pm \infty$

Scale: none

The meaning of scale-free



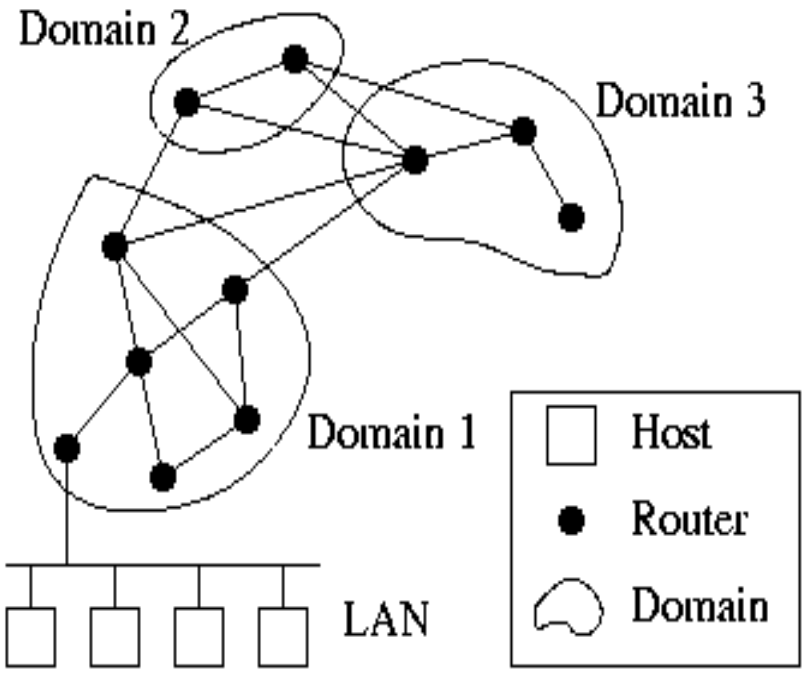
$$k = \langle k \rangle \pm \sigma_k$$

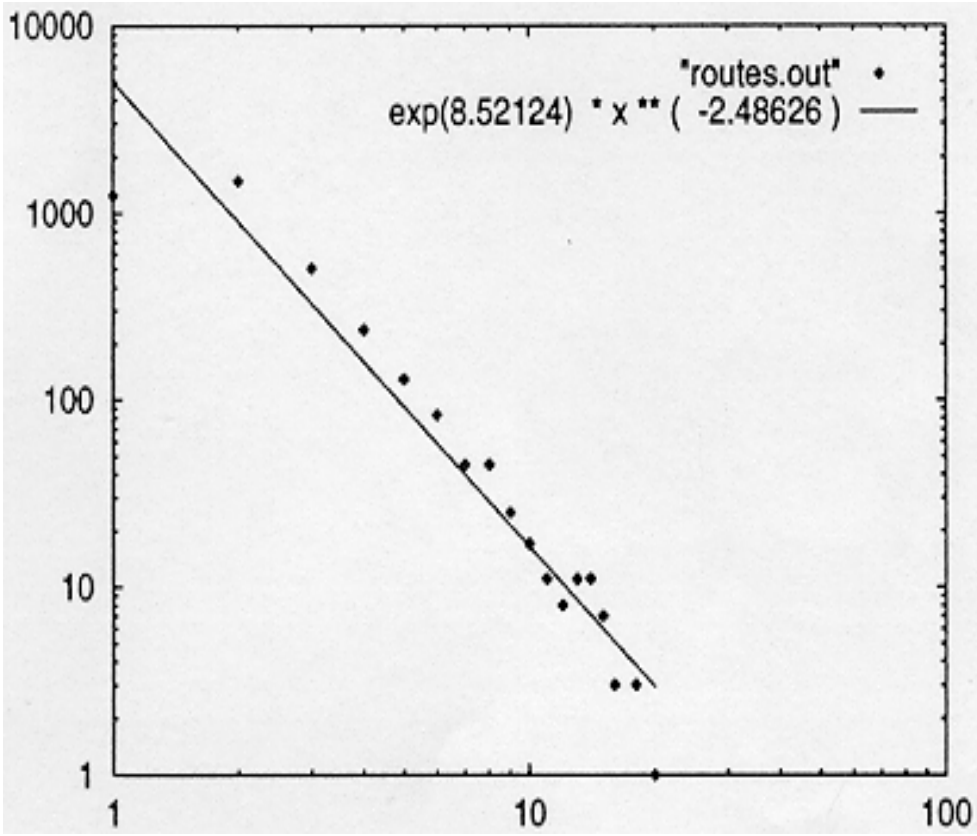
For a random network the standard deviation follows $\sigma = \langle k \rangle^{1/2}$ shown as a green dashed line on the figure. The symbols show σ for nine of the ten reference networks, calculated using the values shown in Table 4.1. The actor network has a very large ⟨k⟩ and σ, hence it omitted for clarity. For each network σ is larger than the value expected for a random network with the same ⟨k⟩. The only exception is the power grid, which is not scale-free. While the phone call network is scale-free, it has a large γ, hence it is well approximated by a random network.

universality

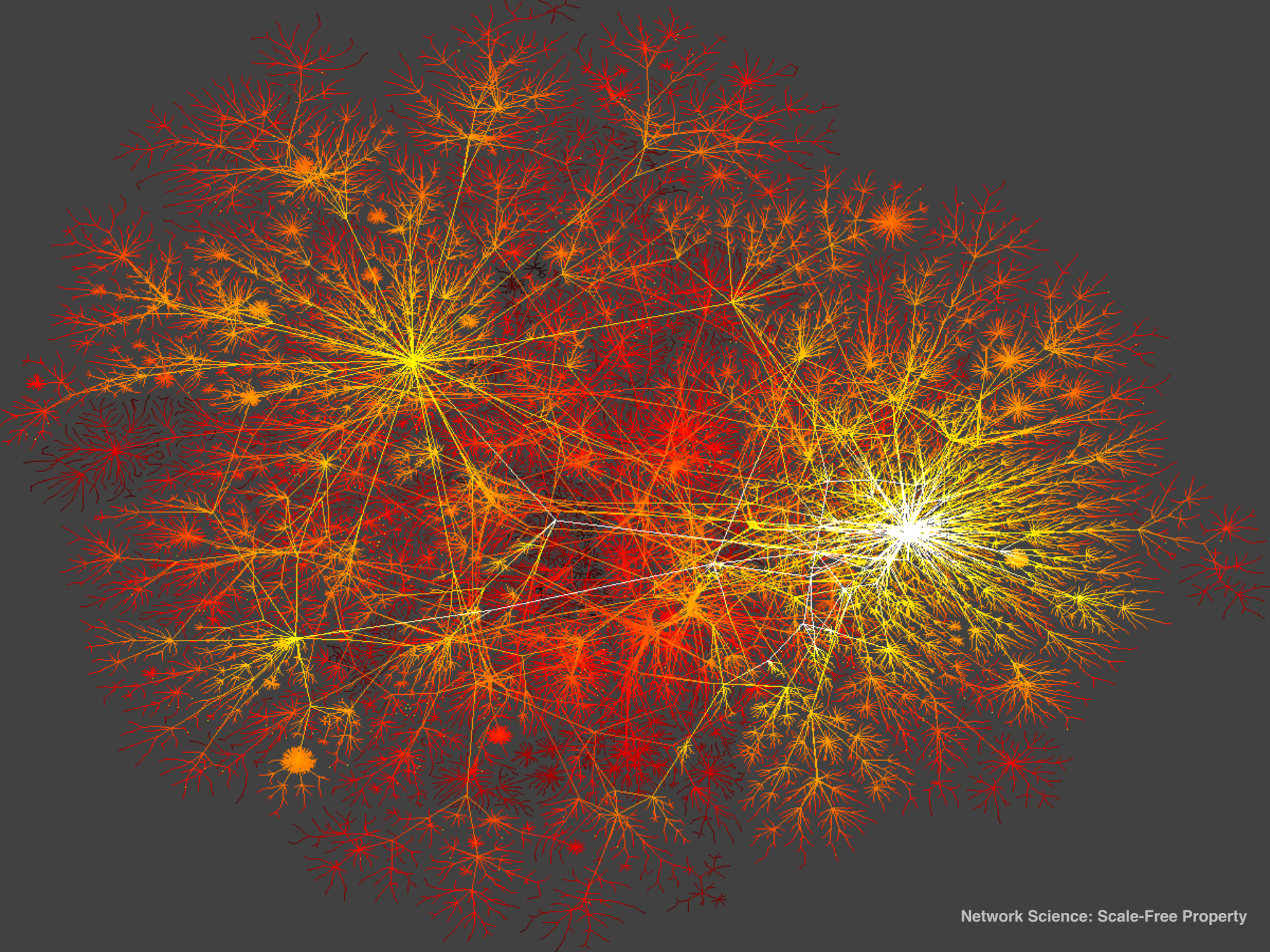
INTERNET BACKBONE

Nodes: computers, routers
Links: physical lines





(Faloutsos, Faloutsos and Faloutsos, 1999)



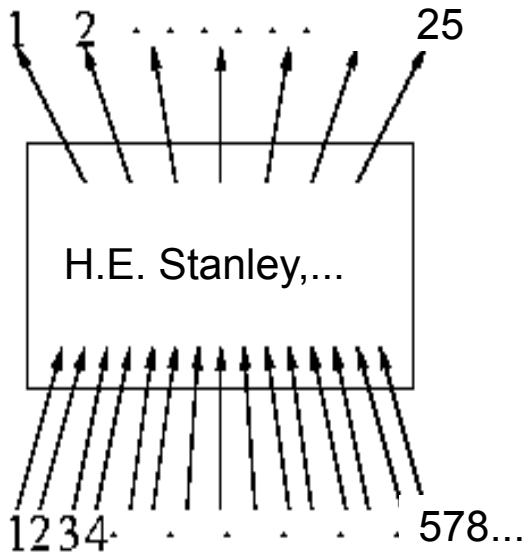
SCIENCE CITATION INDEX

Out of over 500,000 Examined
(see http://www.sst.nrel.gov)

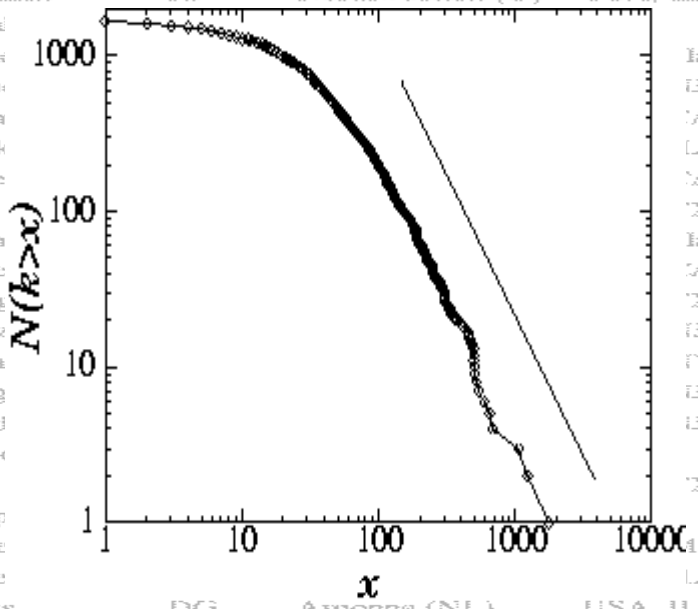
Author name		Institute	Country	Field	avg. cites	total art.	total cites	rank by total cit.
[illegible]	[illegible]	Princeton (U)	USA, NJ	High-energy (T)	168	138	23235	1
Gossard	AC	UCSB (U)	USA, CA	Semic				2
Cava	RJ	Bell Labs (I)	USA, NJ	Super				3
Batlogg	B	Bell Labs (I)	USA, NJ	Super				4
Ploog	K	Max-Planck (NL)	Germany	Semic				5
Ellis	J	Euro Nuclear Cent.	Switzerland	Astrop				6
Fisk	Z	Florida State (U)	USA, FL	Solid S				7
Cardona	M	Max Planck (NL)	Germany	Semic				8
Nanopoulos	DV	Texas A&M (U)	USA, TX	High-e				9
Heeger	AJ	UCSB (U)	USA, CA	Polym				10
Lee*	PA							11
Suzuki*	T							12
Anderson			NJ	Solid S				13
Suzuki*								14
Freeman	AJ	Northwestern (U)	USA, IL	Solid S				15
Tana								16
Mull			land	Super				17
Schm				Super				18
Chen			A	Optics				19
Mork			L	Semic				19
Mille			A	Semic				21
Chu			X	Superconductivity (E)	44	213	9453	22
Bedn			land	Supercond		85	9311	23
Cohe			A	Solid State (T)		284	9311	23
Meng			X	Superconductivity (E)	86	108	9300	25
Wasz				Superconductivity (E)	57	162	9170	26
Shira			Y	Superconductivity (E)	33	269	8841	27
Wieg				Semiconductors (E)	85	104	8822	28
Vand				Magnetism (E)	67	129	8686	29
Uchi					28	301	8520	30
Hor			X	Superconductivity (E)	72	119	8512	31
Murp				Astronomy (E)	111	76	8439	32
Birge			A	Superconductivity (E)	41	286	8375	33
Jorge			L	Superconductivity (E)	50	167	8298	34
Hinks	DG	Argonne (NL)	USA, IL	Superconductivity (E)	37	223	8263	35

* citation total may be skewed because of multiple authors with the same name

Nodes: papers
Links: citations



1736 PRL papers (1988)



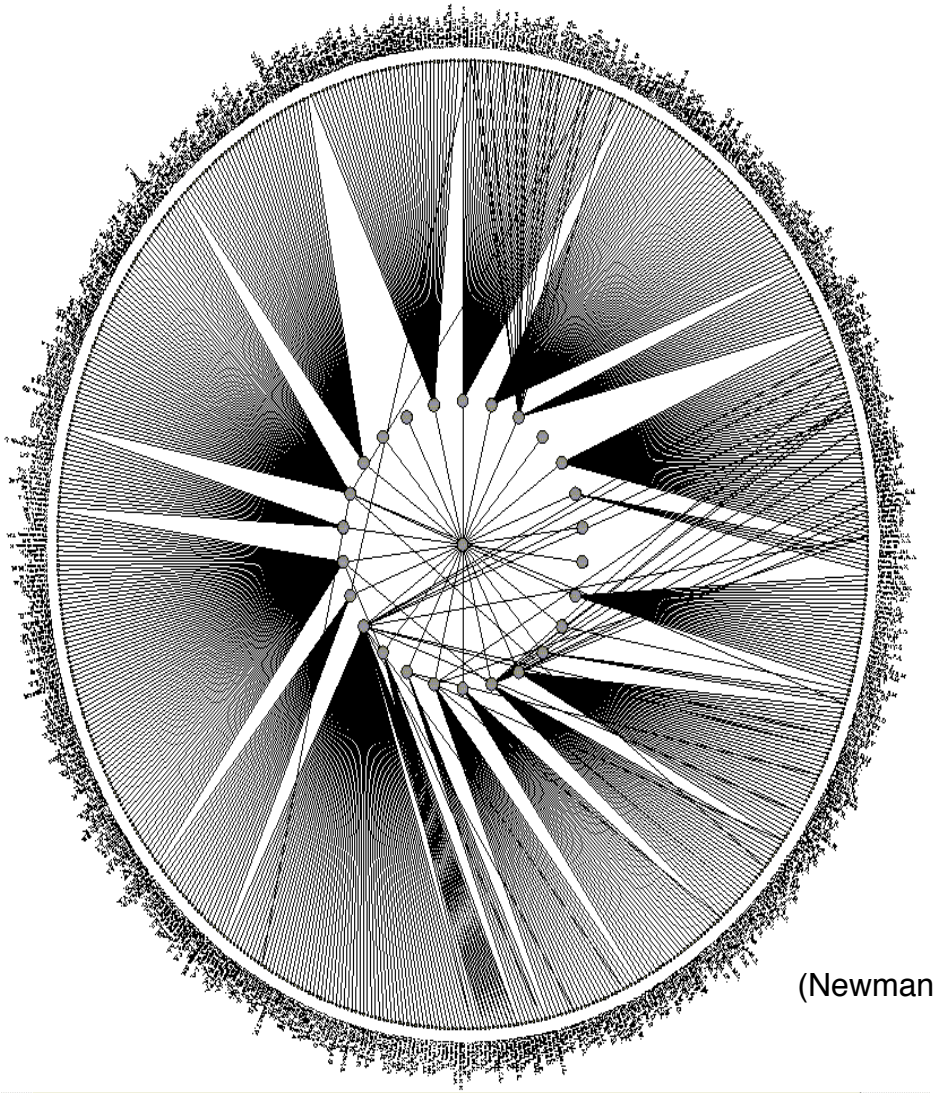
$P(k) \sim k^{-\gamma}$

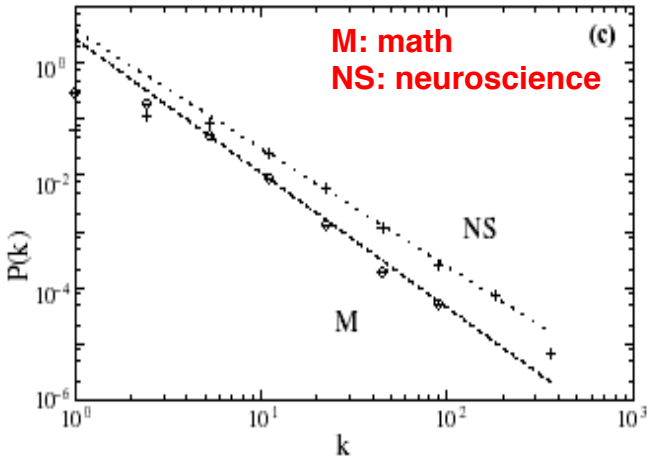
$(\gamma = 3)$

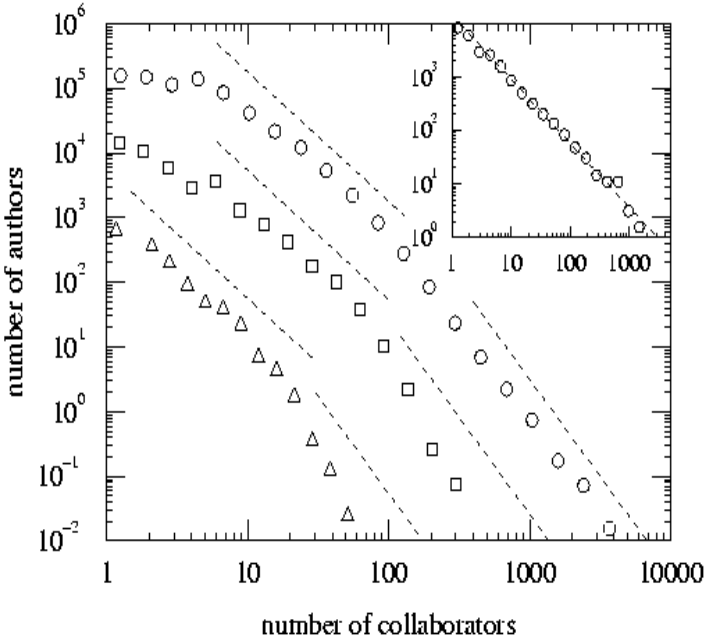
(S. Redner, 1998)

SCIENCE COAUTHORSHIP

Nodes: scientist (authors)
Links: joint publication





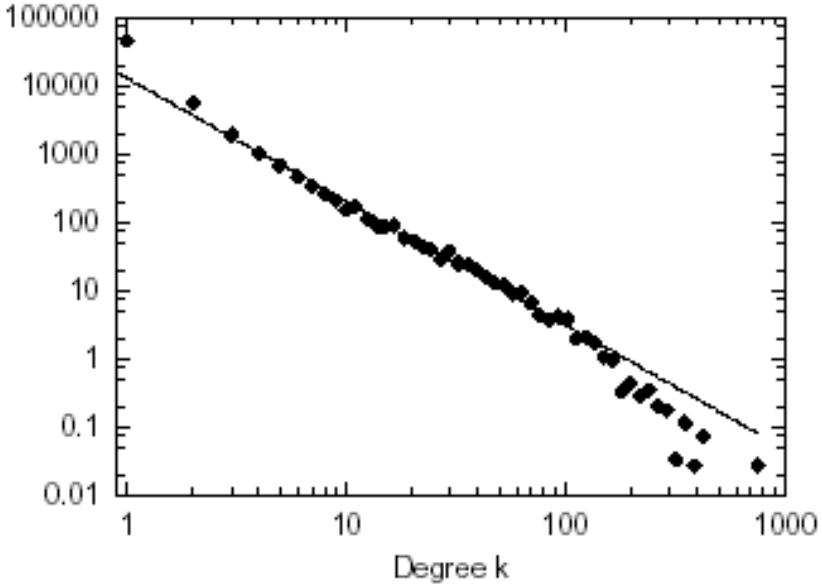


(Newman, 2000, Barabasi et al 2001)

ONLINE COMMUNITIES

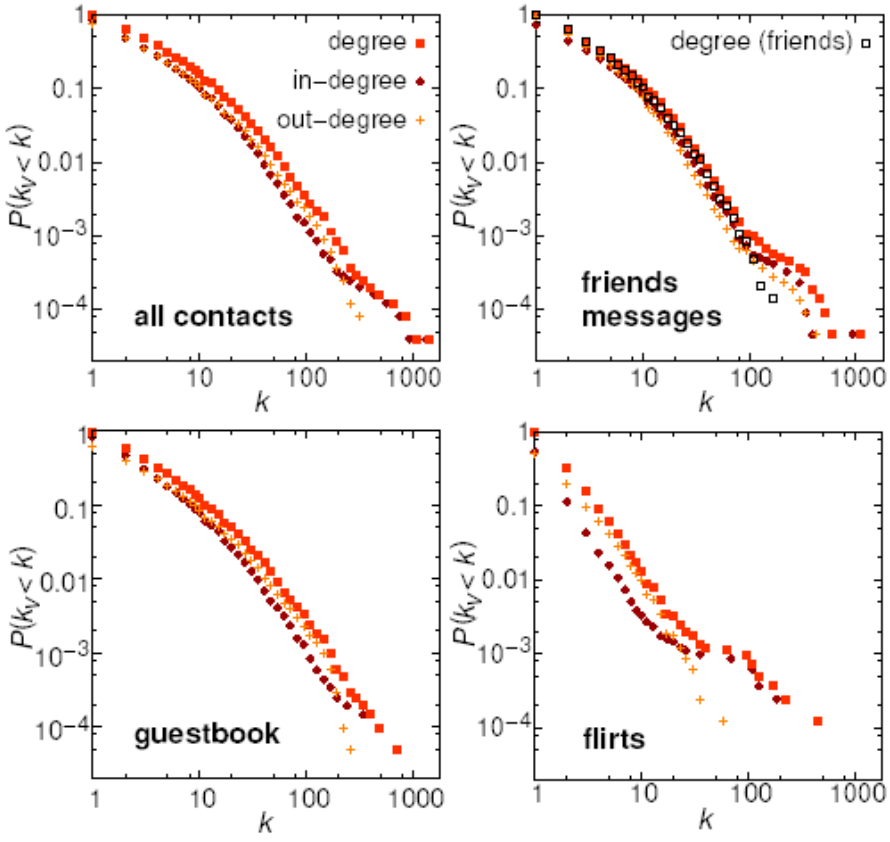
Nodes: online user
Links: email contact

Kiel University log files
112 days, N=59,912 nodes



Ebel, Mielsch, Bornholdtz, PRE 2002.

Pussokram.com online community;
512 days, 25,000 users.



Holme, Edling, Liljeros, 2002.

Network	N	L	$\langle k \rangle$	$\langle k_{in}^2 \rangle$	$\langle k_{out}^2 \rangle$	$\langle k^2 \rangle$	γ_{in}	γ_{out}	γ
Internet	192,244	609,066	6.34	-	-	240.1	-	-	3.42*
WWW	325,729	1,497,134	4.60	1546.0	482.4	-	2.00	2.31	-
Power Grid	4,941	6,594	2.67	-	-	10.3	-	-	Exp.
Mobile-Phone Calls	36,595	91,826	2.51	12.0	11.7	-	4.69*	5.01*	-
Email	57,194	103,731	1.81	94.7	1163.9	-	3.43*	2.03*	-
Science Collaboration	23,133	93,437	8.08	-	-	178.2	-	-	3.35*
Actor Network	702,388	29,397,908	83.71	-	-	47,353.7	-	-	2.12*
Citation Network	449,673	4,689,479	10.43	971.5	198.8	-	3.03*	4.00*	-
E. Coli Metabolism	1,039	5,802	5.58	535.7	396.7	-	2.43*	2.90*	-
Protein Interactions	2,018	2,930	2.90	-	-	32.3	-	-	2.89*-

Table 4.1

Degree Fluctuations in Real Networks

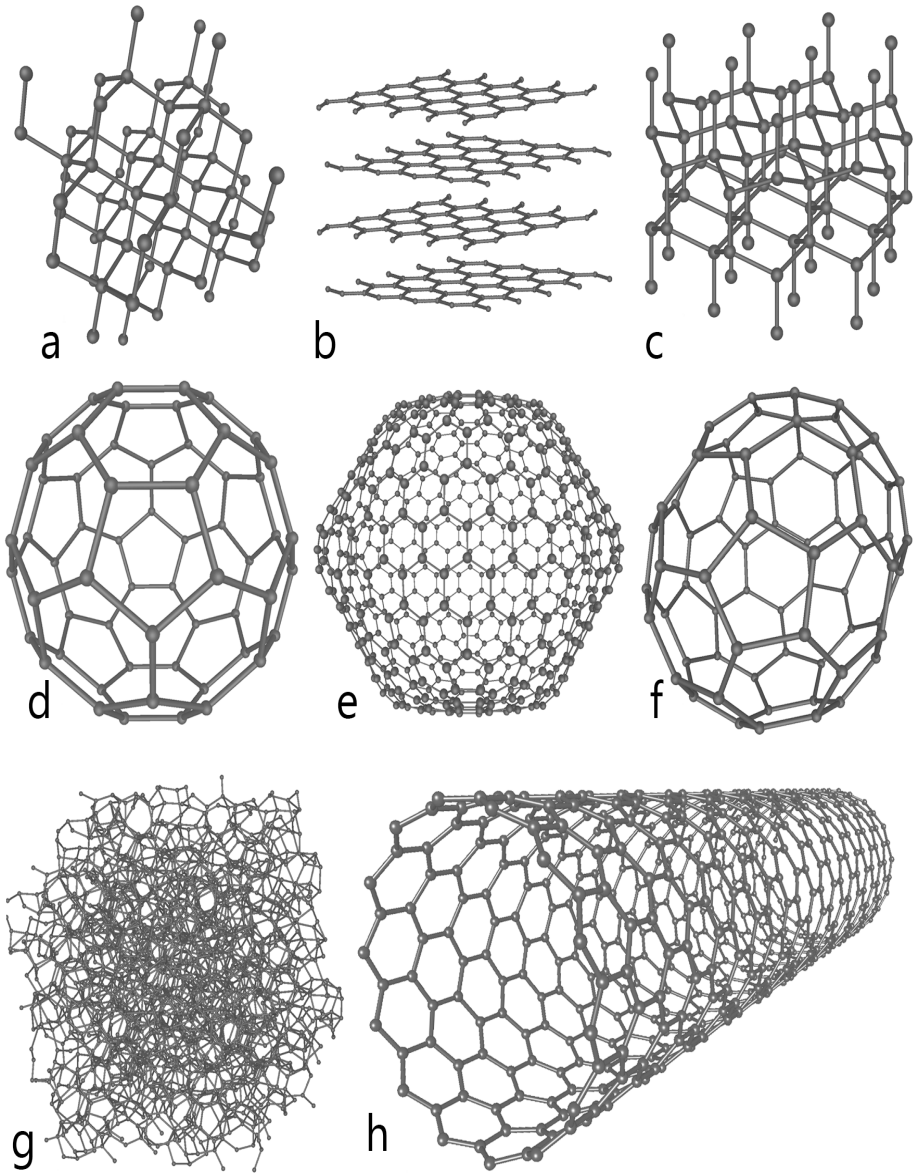
The table shows the first $\langle k \rangle$ and the second moment $\langle k^2 \rangle$ ($\langle k_{in}^2 \rangle$ and $\langle k_{out}^2 \rangle$ for directed networks) for ten reference networks. For directed networks we list $\langle k \rangle = \langle k_{in} \rangle = \langle k_{out} \rangle$. We also list the estimated degree exponent, γ, for each network, determined using the procedure discussed in ADVANCED TOPICS 4.A. The stars next to the reported values indicate the confidence of the fit to the degree distribution. That is, * means that the fit shows statistical confidence for a power-law ($k^{-\gamma}$); while ** marks statistical confidence for a fit (4.39) with an exponential cutoff. Note that the power grid is not scalefree. For this network a degree distribution of the form $e^{-\lambda k}$ offers a statistically significant fit, which is why we placed an "Exp" in the last column.

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.

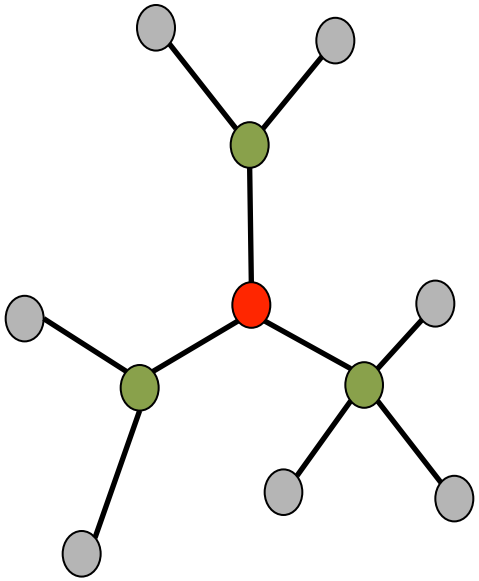
The power grid, consisting of generators and switches connected by transmission lines



Ultra-small property

DISTANCES IN RANDOM GRAPHS

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



- 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:

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

SMALL WORLD BEHAVIOR IN SCALE-FREE NETWORKS

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

Ultra Small World

Small World

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.

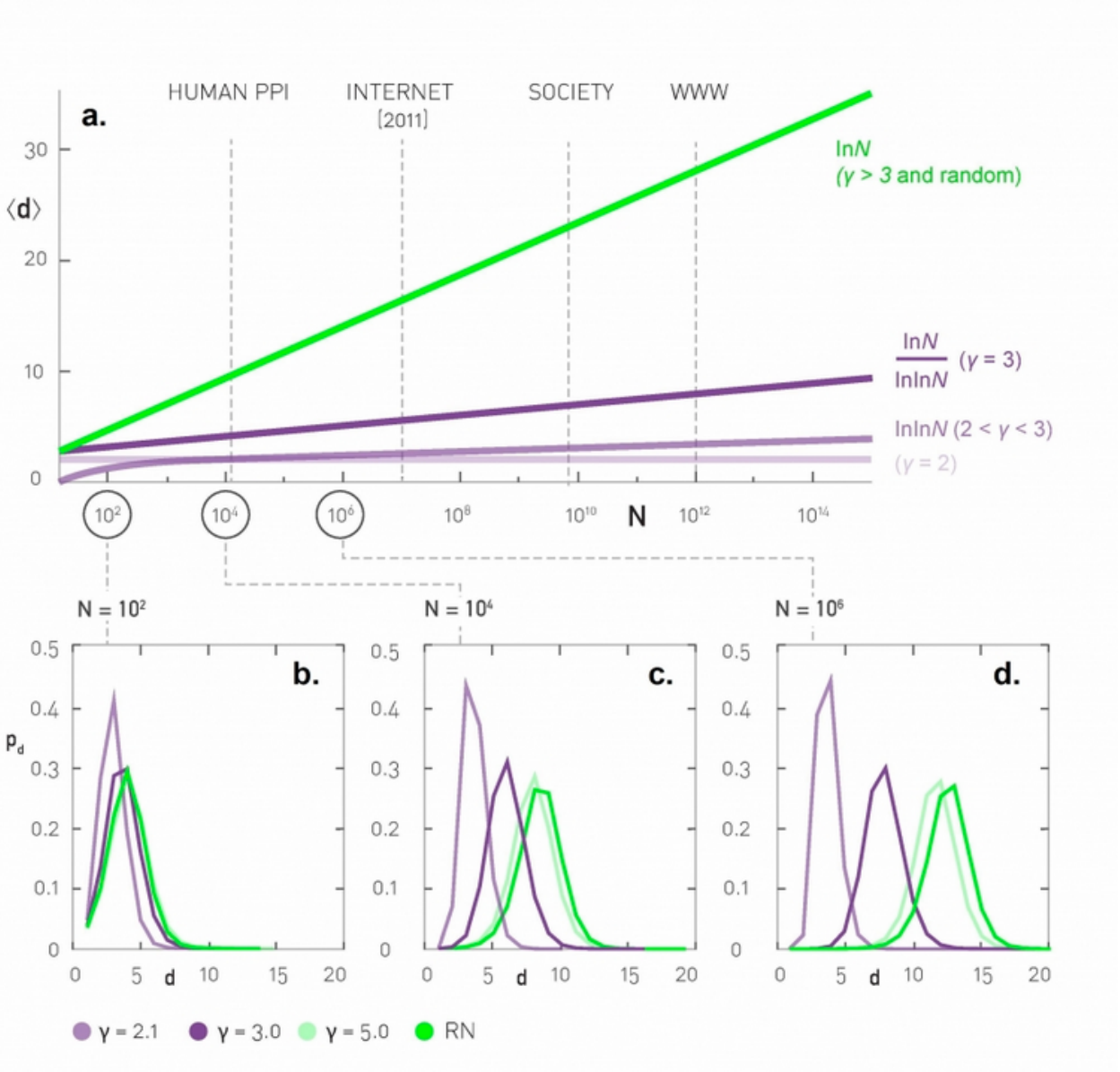
The average path length increases slower than logarithmically. In a random network all nodes have comparable degree, thus most paths will have comparable length. In a scale-free network the vast majority of the path go through the few high degree hubs, reducing the distances between nodes.

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.

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

SMALL WORLD BEHAVIOR IN SCALE-FREE NETWORKS

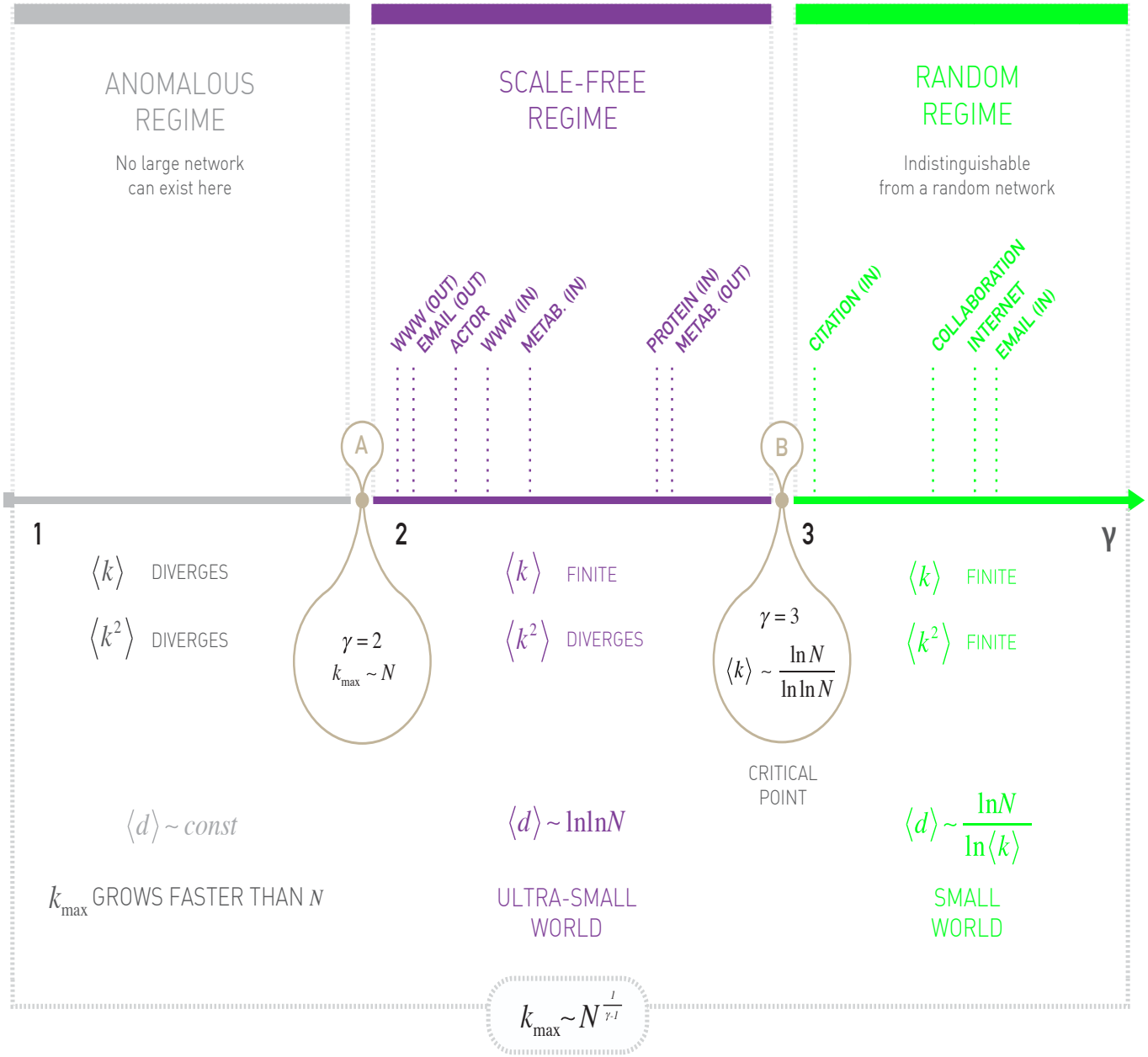


$$\langle d \rangle \sim \begin{cases} \text{const.} & \gamma = 2, \\ \dfrac{\ln\ln N}{\ln(\gamma - 1)} & 2 < \gamma < 3, \\ \dfrac{\ln N}{\ln\ln N} & \gamma = 3, \\ \ln N & \gamma > 3. \end{cases}$$

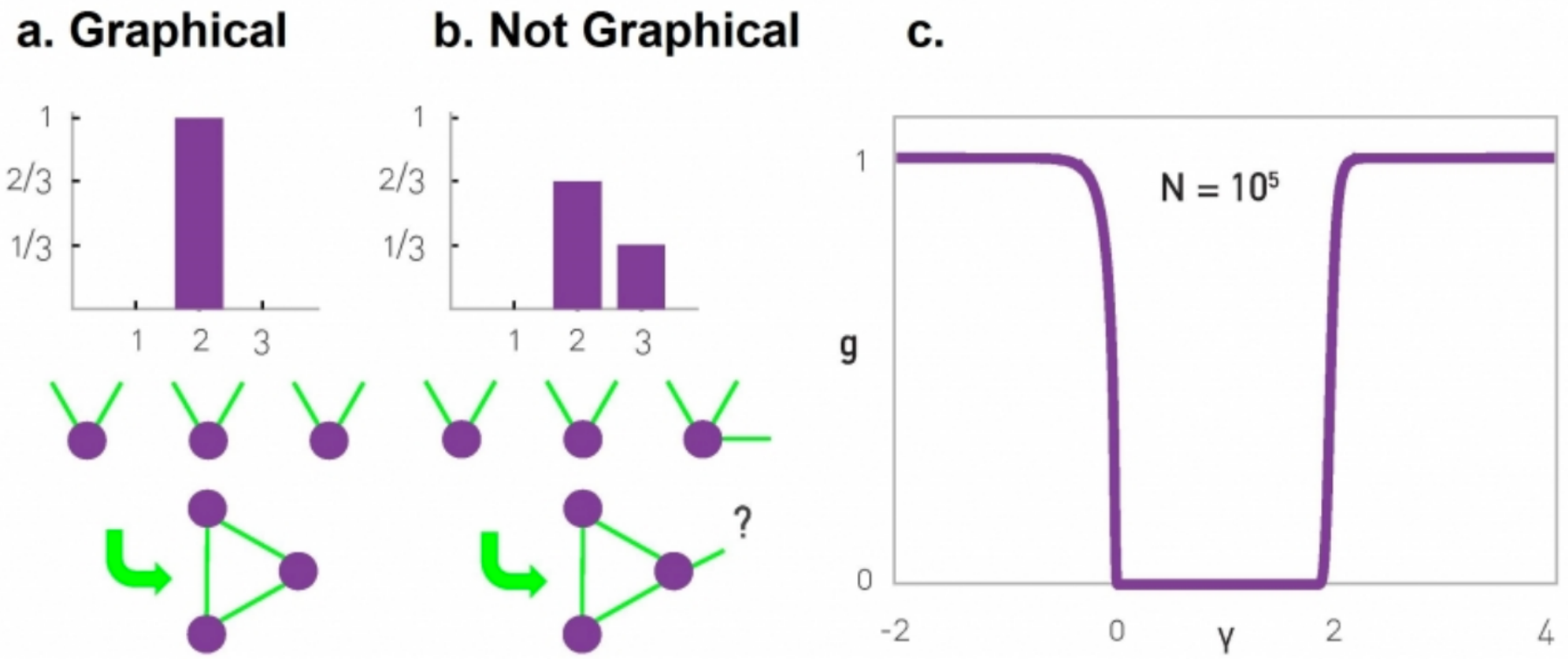
The scaling of the average path length in the four scaling regimes characterizing a cale-free network: constant (γ = 2), lnlnN (2 ‹ γ ‹ 3), lnN/ lnlnN (γ = 3), lnN (γ › 3 and random networks). The dotted lines mark the approximate size of several real networks. Given their modest size, in biological networks, like the human protein-protein interaction network (PPI), the differences in the node-to-node distances are relatively small in the four regimes. The differences in ⟨d⟩ is quite significant for networks of the size of the social network or the WWW. For these the small-world formula significantly underestimates the real ⟨d⟩.

(b)-(d) Distance distribution for networks of size N = 102, 104, 106, illustrating that while for small networks (N = 102) the distance distributions are not too sensitive to γ, for large networks (N = 106) pd and ⟨d⟩ change visibly with γ.

The role of the degree exponent



Graphicality: No large networks for γ<2

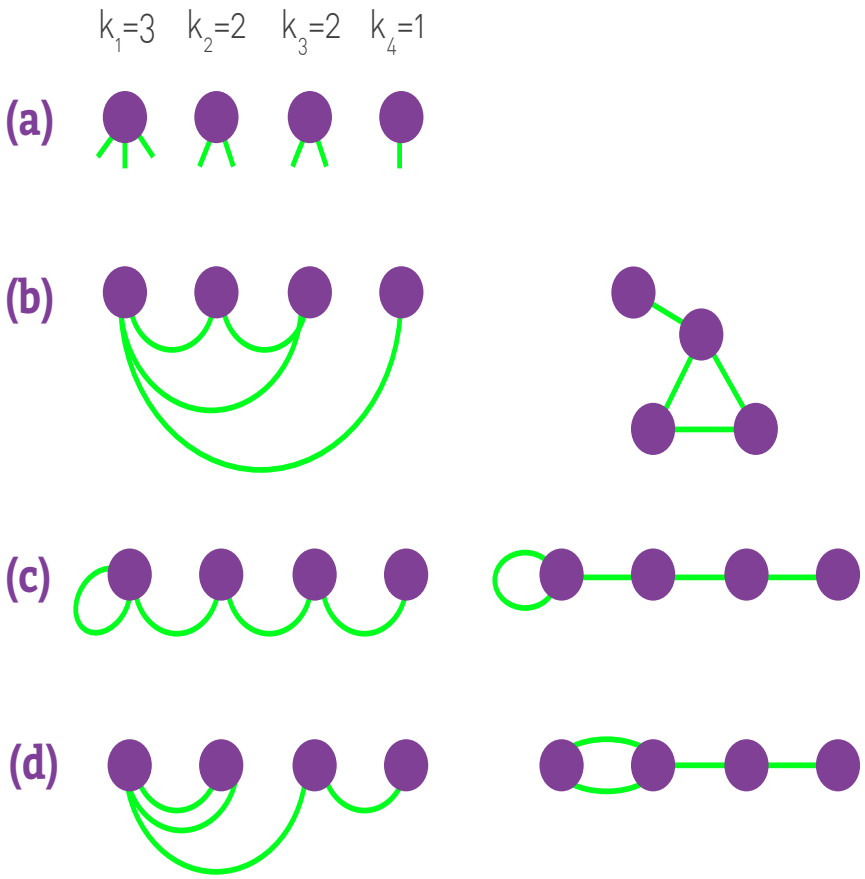


Networks With γ ‹ 2 are Not Graphical

Degree distributions and the corresponding degree sequences for two small networks. The difference between them is in the degree of a single node. While we can build a simple network using the degree distribution (a), it is impossible to build one using (b), as one stub always remains unmatched. Hence (a) is graphical, while (b) is not. Fraction of networks, g, for a given γ that are graphical. A large number of degree sequences with degree exponent γ and N = 105 were generated, testing the graphicality of each network. The figure indicates that while virtually all networks with γ › 2 are graphical, it is impossible to find graphical networks in the 0 ‹ γ ‹ 2 range

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

Generating networks with predefined degree seq: Configuration model



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

(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, 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 (2b), others self-loops (2c) or multi-edges (2d).

Yet, the expected number of self- and multi-edges goes to zero in the limit.

summary

The Barabási-Albert 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:

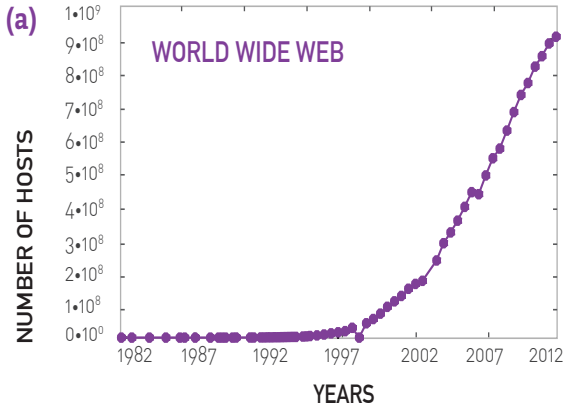
- *Why does the random network model of Erdős and Rényi fail to reproduce the hubs and the power laws observed in many real networks?*

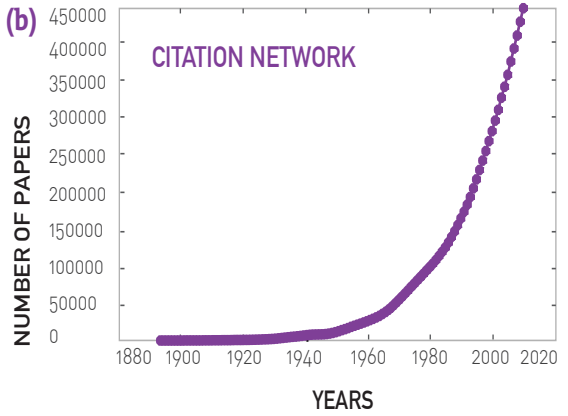
- *Why do so different systems as the WWW or the cell converge to a similar scale-free architecture?*

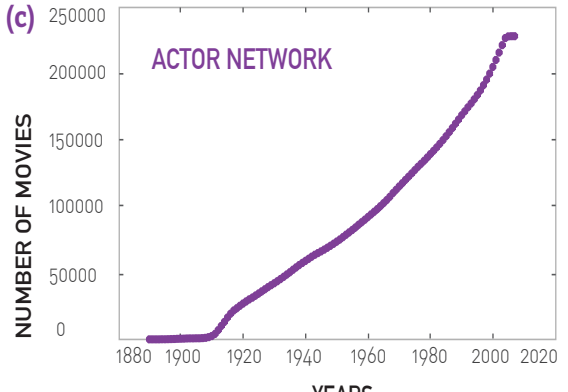
BA MODEL: Growth

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)

BA MODEL: 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)

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)

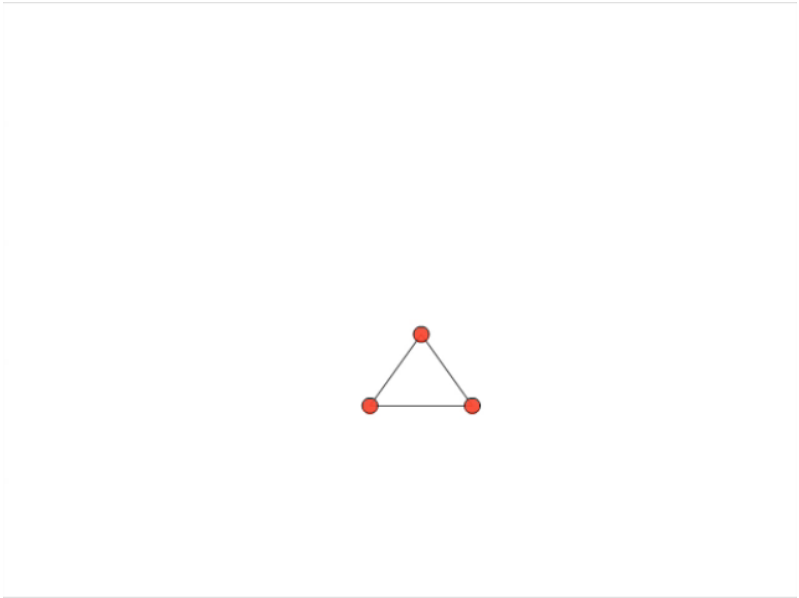
Origin of SF networks: Growth and preferential attachment

(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



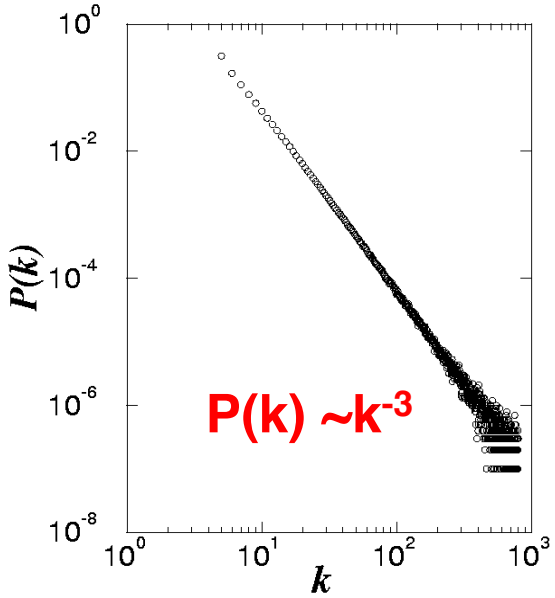
GROWTH:

add a new node with m links

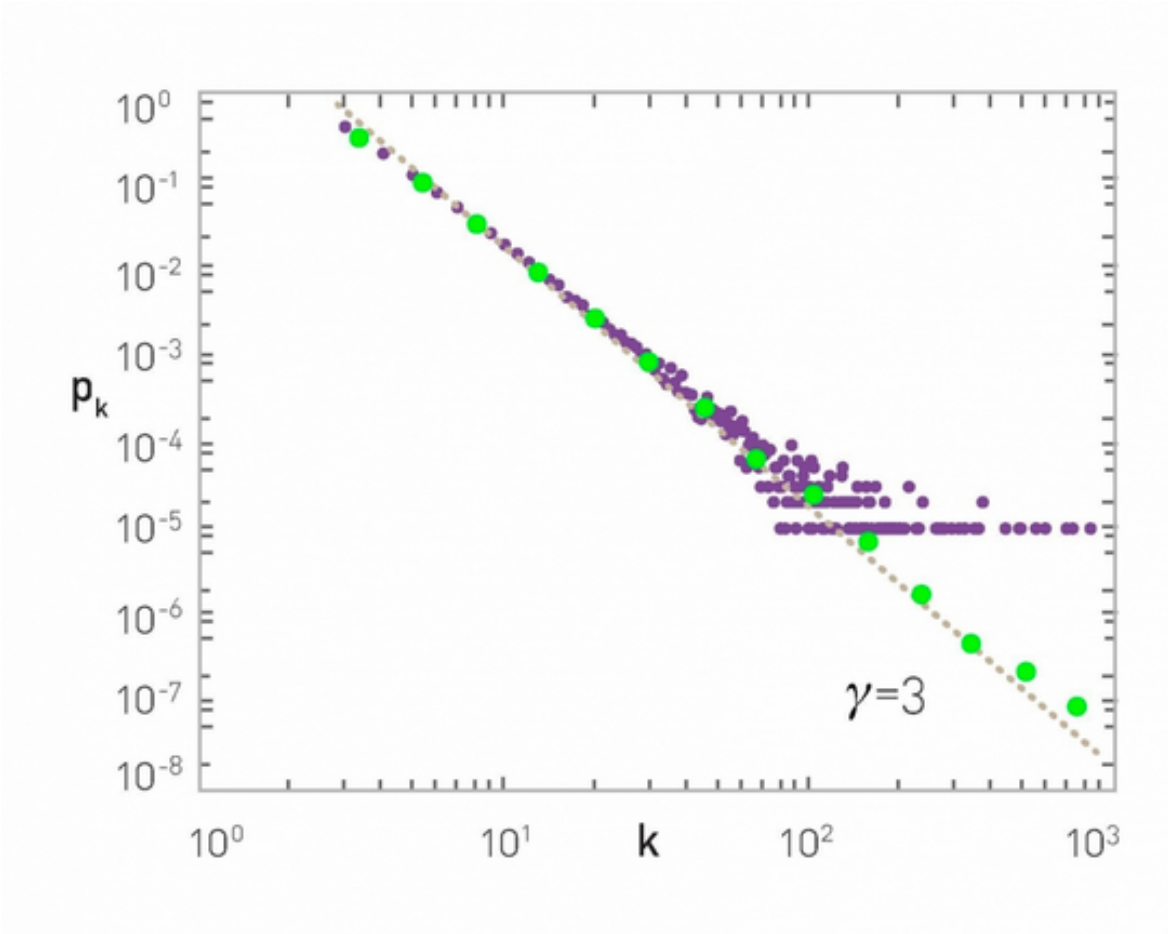
PREFERENTIAL ATTACHMENT:

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

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



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



The degree distribution of a network generated by the Barabási-Albert model.
The figure shows p_k for a single network of size N=100,000 and m=3. It shows both the linearly-binned (purple) and the log-binned version (green) of p_k. The straight line is added to guide the eye and has slope γ=3, corresponding to the network's predicted degree exponent.

Degree distribution

$$k_i(t) = m\left(\frac{t}{t_i}\right)^{\beta} \qquad \beta = \frac{1}{2}$$

$$P(k) = \frac{2m(m+1)}{k(k+1)(k+2)} \qquad \boxed{\gamma = 3}$$

$$P(k) \sim k^{-3} \qquad \text{for large } k$$

(i) The degree exponent is independent of m.

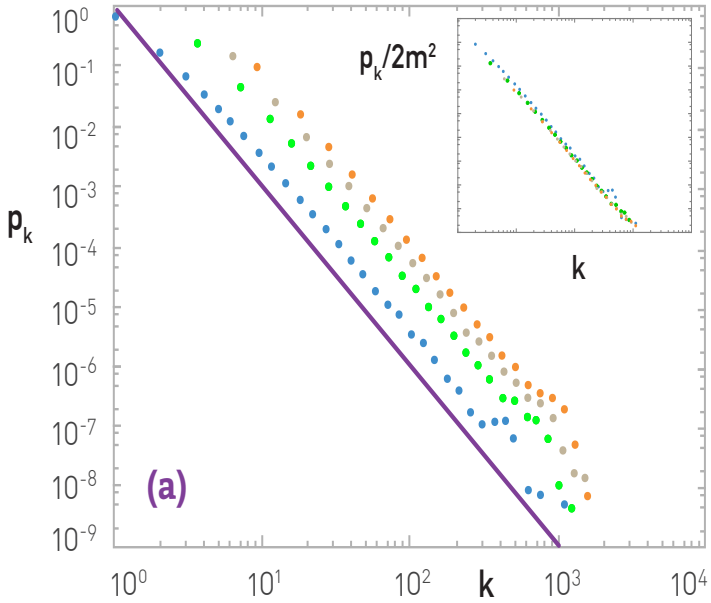
(ii) As the power-law describes systems of rather different ages and sizes, it is expected that a correct model should provide a time-independent degree distribution. Indeed, asymptotically the degree distribution of the BA model is independent of time (and of the system size N)

→ the network reaches a stationary scale-free state.

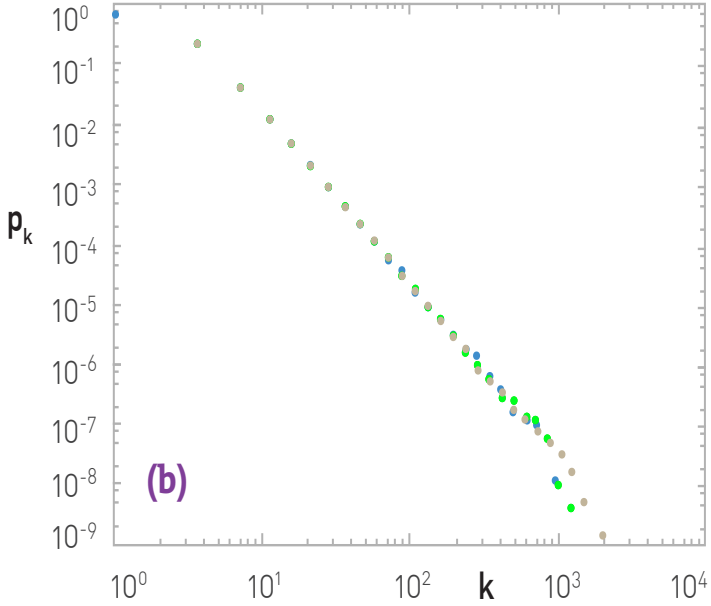
(iii) The coefficient of the power-law distribution is proportional to m^2.

NUMERICAL SIMULATION OF THE BA MODEL

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



(a) We generated networks with N=100,000 and m_0=m=1 (blue), 3 (green), 5 (grey), and 7 (orange). The fact that the curves are parallel to each other indicates that γ is independent of m and m_0. The slope of the purple line is -3, corresponding to the predicted degree exponent γ=3. Inset: (5.11) predicts $p_k \sim 2m^2$, hence $p_k/2m^2$ should be independent of m. Indeed, by plotting $p_k/2m^2$ vs. k, the data points shown in the main plot collapse into a single curve.



(b) The Barabási-Albert model predicts that p_k is independent of N. To test this we plot p_k for N = 50,000 (blue), 100,000 (green), and 200,000 (grey), with m_0=m=3. The obtained p_k are practically indistinguishable, indicating that the degree distribution is stationary, i.e. independent of time and system size.

absence of growth and preferential attachment

Limiting cases

Model A: retains growth but does not include preferential attachment.

The probability of a new node connecting to any pre-existing node is equal.
The resulting degree distribution in this limit is geometric.

Model B: retains preferential attachment but eliminates growth.

The model begins with a fixed number of disconnected nodes and adds links, preferentially choosing high degree nodes as link destinations.
Though the degree distribution early in the simulation looks scale-free, the distribution is not stable, and it eventually becomes nearly Gaussian as the network nears saturation.

Growth and preferential attachment are needed simultaneously to reproduce the stationary power-law distribution observed in real networks.

Diameter and clustering coefficient

Diameter

$$D \sim \frac{\log N}{\log\log N}$$

The average distance ⟨d⟩ scales in a similar fashion.

Indeed, for small N the ln N term captures the scaling of ⟨d⟩ with N, but for large N(≥10^4) the impact of the logarithmic correction ln ln N becomes noticeable.

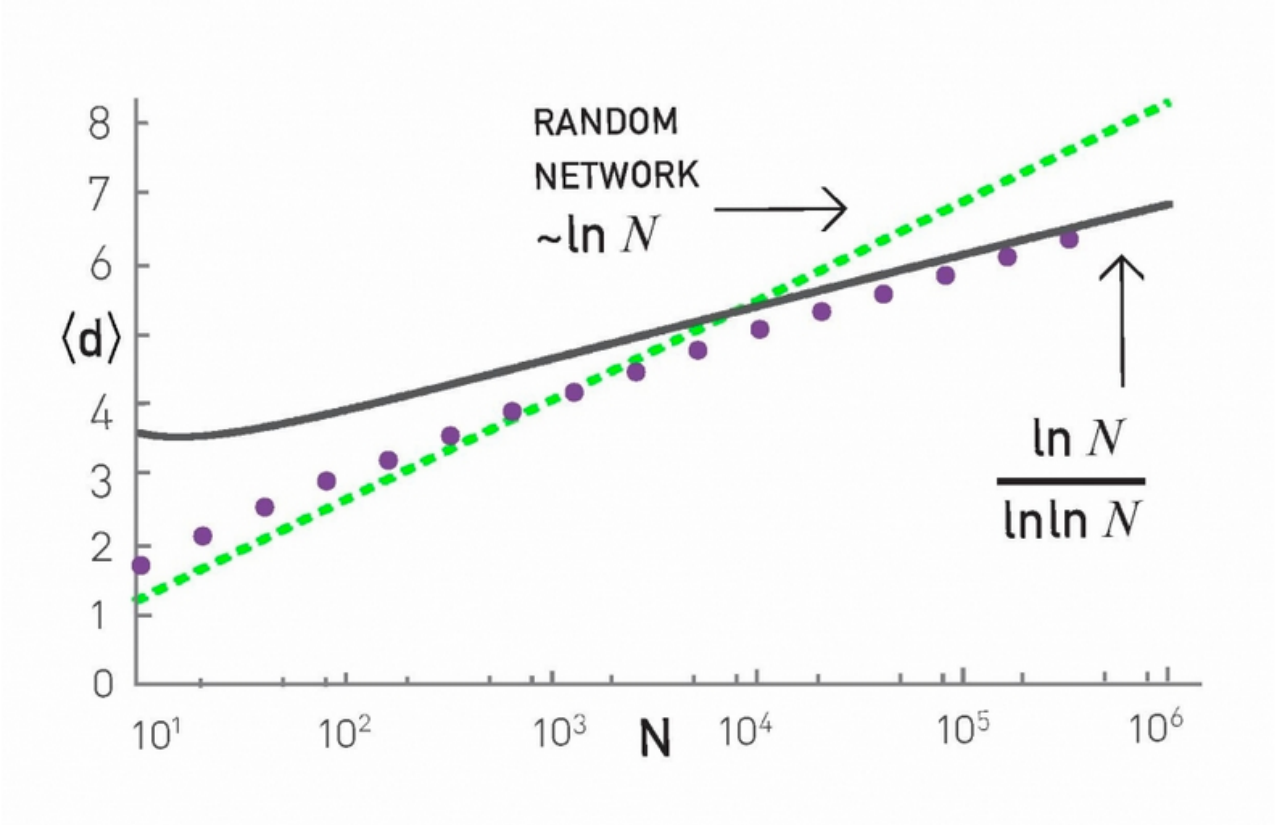


Image 5.18

Average Distance

The dependence of the average distance on the system size in the Barabási-Albert model. The continuous line corresponds to the exact result (5.29), while the dotted line corresponds to the prediction (3.19) for a random network. The analytical predictions do not provide the exact perfactors, hence the lines are not fits, but indicate only the predicted *N*-dependent trends. The results were averaged for ten independent runs for $m = 2$.

Clustering coefficient

Reminder: for a random graph we have:

$$C_{rand} = \frac{< k >}{N} \sim N^{-1}$$

What is the functional form of C(N)?

$$C = \frac{m}{8} \frac{(\ln N)^2}{N}$$

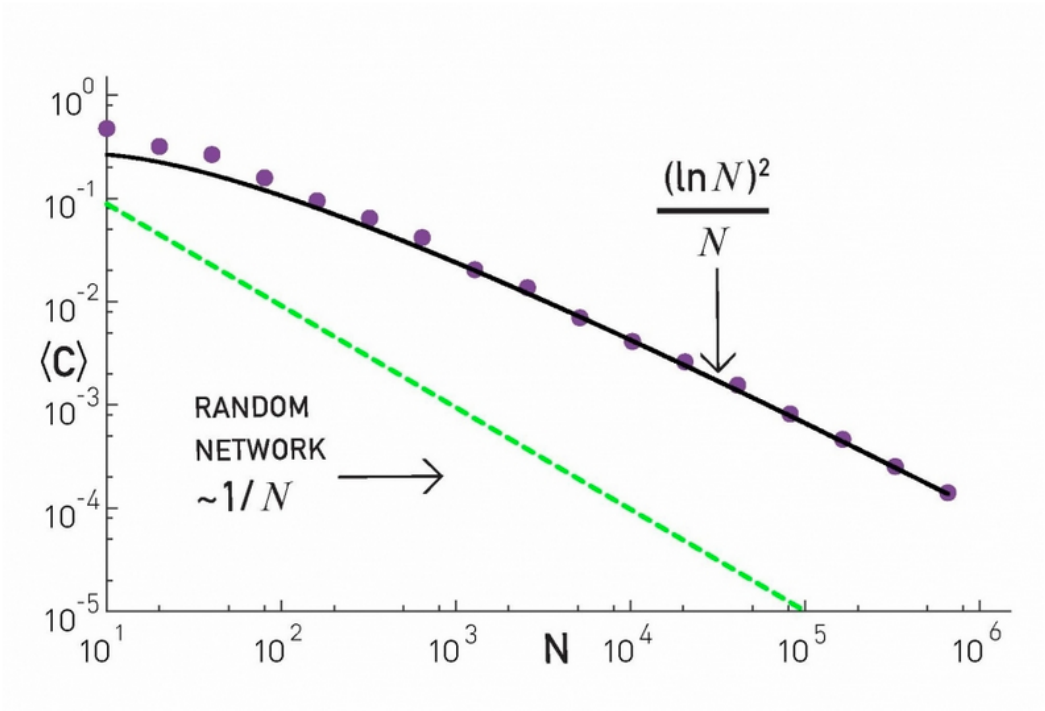


Image 5.19

Clustering Coefficient

The dependence of the average clustering coefficient on the system size N for the Barabási-Albert model. The continuous line corresponds to the analytical prediction (5.30), while the dotted line corresponds to the prediction for a random network, for which $\langle C \rangle \sim 1/N$. The results are averaged for ten independent runs for $m = 2$. The dashed and continuous curves are not fits, but are drawn to indicate the predicted N dependent trends.

Section 11: Summary

Number of Nodes

$N = t$

Number of Links

$N = mt$

Average Degree

$\langle k \rangle = 2m$

Degree Dynamics

$k_i(t) = m\,(t/t_i)^{\beta}$

Dynamical Exponent

$\beta = 1/2$

Degree Distribution

$p_k \sim k^{\gamma}$

Degree Exponent

$\gamma = 3$

Average Distance

$\langle d \rangle \sim \log N / \log \log N$

Clustering Coefficient

$\langle C \rangle \sim (\ln N)^2 / N$

The network grows, but the degree distribution is stationary.

Section 11: Summary

Number of Nodes

$N = t$

Number of Links

$N = mt$

Average Degree

$\langle k \rangle = 2m$

Degree Dynamics

$k_i(t) = m\,(t/t_i)^{\beta}$

Dynamical Exponent

$\beta = 1/2$

Degree Distribution

$p_k \sim k^{\gamma}$

Degree Exponent

$\gamma = 3$

Average Distance

$\langle d \rangle \sim \log N / \log \log N$

Clustering Coefficient

$\langle C \rangle \sim (\ln N)^2 / N$

Consequently, the modeling philosophy behind the model is simple: *to understand the topology of a complex system, we need to describe how it came into being.*

The network grows, but the degree distribution is stationary.

Section 11: Summary

Number of Nodes

$N = t$

Number of Links

$N = mt$

Average Degree

$\langle k \rangle = 2m$

Degree Dynamics

$k_i(t) = m\,(t/t_i)^{\beta}$

Dynamical Exponent

$\beta = 1/2$

Degree Distribution

$p_k \sim k^{\gamma}$

Degree Exponent

$\gamma = 3$

Average Distance

$\langle d \rangle \sim \log N / \log\log N$

Clustering Coefficient

$\langle C \rangle \sim (\ln N)^2 / N$

- The model predicts γ=3 while the degree exponent of real networks varies between 2 and 5 (Table 4.2).

- Many networks, like the WWW or citation networks, are directed, while the model generates undirected networks.

- Many processes observed in networks, from linking to already existing nodes to the disappearance of links and nodes, are absent from the model.

- The model does not allow us to distinguish between nodes based on some intrinsic characteristics, like the novelty of a research paper or the utility of a webpage.

- While the Barabási-Albert model is occasionally used as a model of the Internet or the cell, in reality it is not designed to capture the details of any particular real network. It is a minimal, proof of principle model whose main purpose is to capture the basic mechanisms responsible for the emergence of the scale-free property. Therefore, if we want to understand the evolution of systems like the Internet, the cell or the WWW, we need to incorporate the important details that contribute to the time evolution of these systems, like the directed nature of the WWW, the possibility of internal links and node and link removal.

Can latecomers make it?

The Bianconi-Barabasi Model

- **Growth**

  In each timestep a new node j with m links and fitness η_j is added to the network, where η_j is a random number chosen from a *fitness distribution* $\rho(\eta)$. Once assigned, a node's fitness does not change.

- **Preferential Attachment**

  The probability that a link of a new node connects to node i is proportional to the product of node i's degree k_i and its fitness η_i,

$$\Pi_i = \frac{\eta_i k_i}{\sum_j \eta_j k_j}. \qquad (6.1)$$

Degree distribution of the Bianconi–Barabási model depends on the fitness distribution.

Two scenarios:

- If the fitness distribution has a finite domain, then the degree distribution will have a power-law just like the BA model.
- If the fitness distribution has an infinite domain, then the node with the highest fitness value will attract a large number of nodes and show a winners-take-all scenario (monopoly dominance).

LESSONS LEARNED: evolving network models

1. There is no universal exponent characterizing all networks.

2. Growth and preferential attachment are responsible for the emergence of the scale-free property.

3. The origins of the preferential attachment is system-dependent.

4. Modeling real networks:
   - identify the microscopic processes that take place in the system
   - measure their frequency from real data
   - develop dynamical models that capture these processes.