CPSC 572/672: Fundamentals of (Social) Network Analysis and Data Mining

Real Networks:
Scale-free property

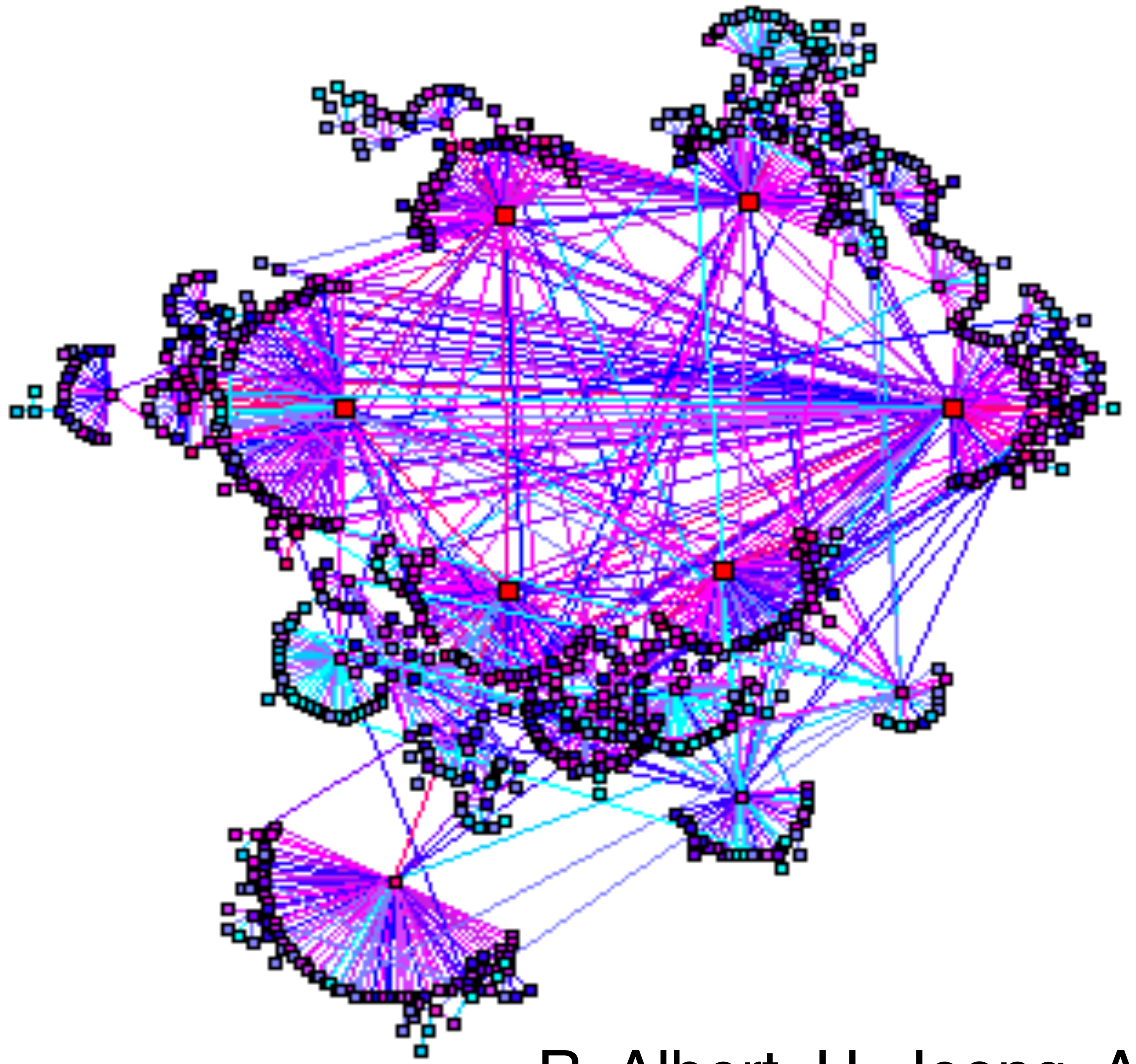
Power laws and scale-free networks

WORLD WIDE WEB

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

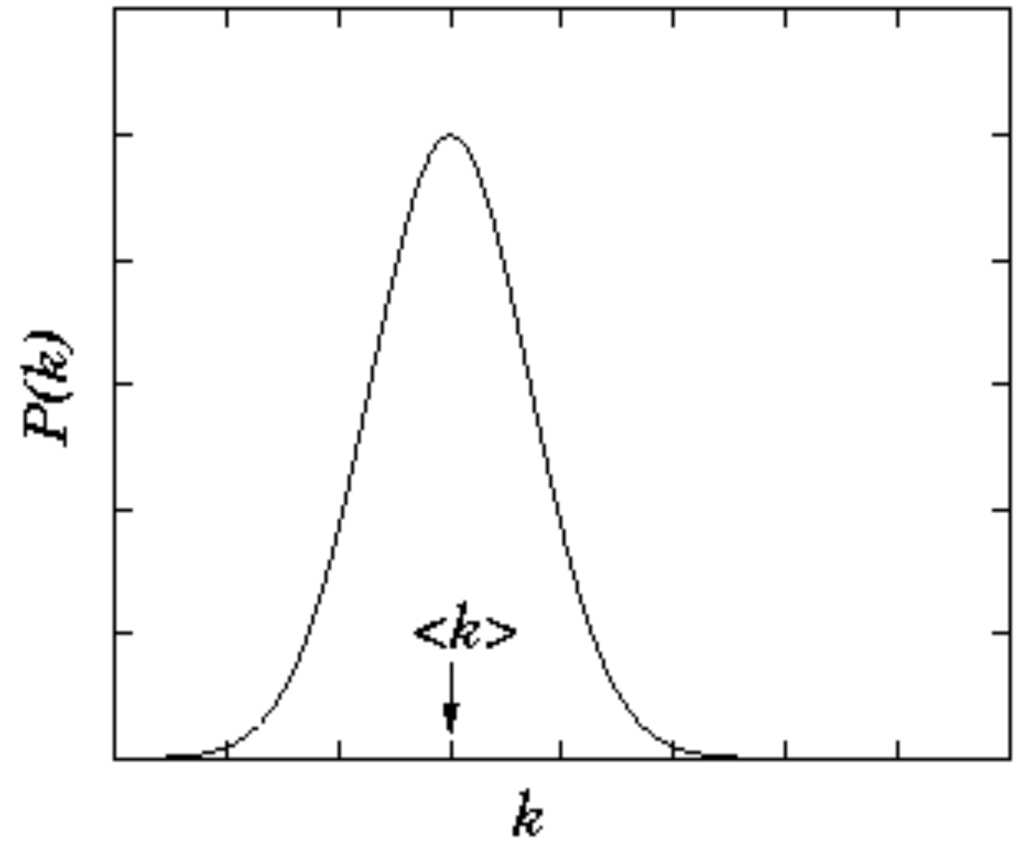
Over 3 billion documents

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

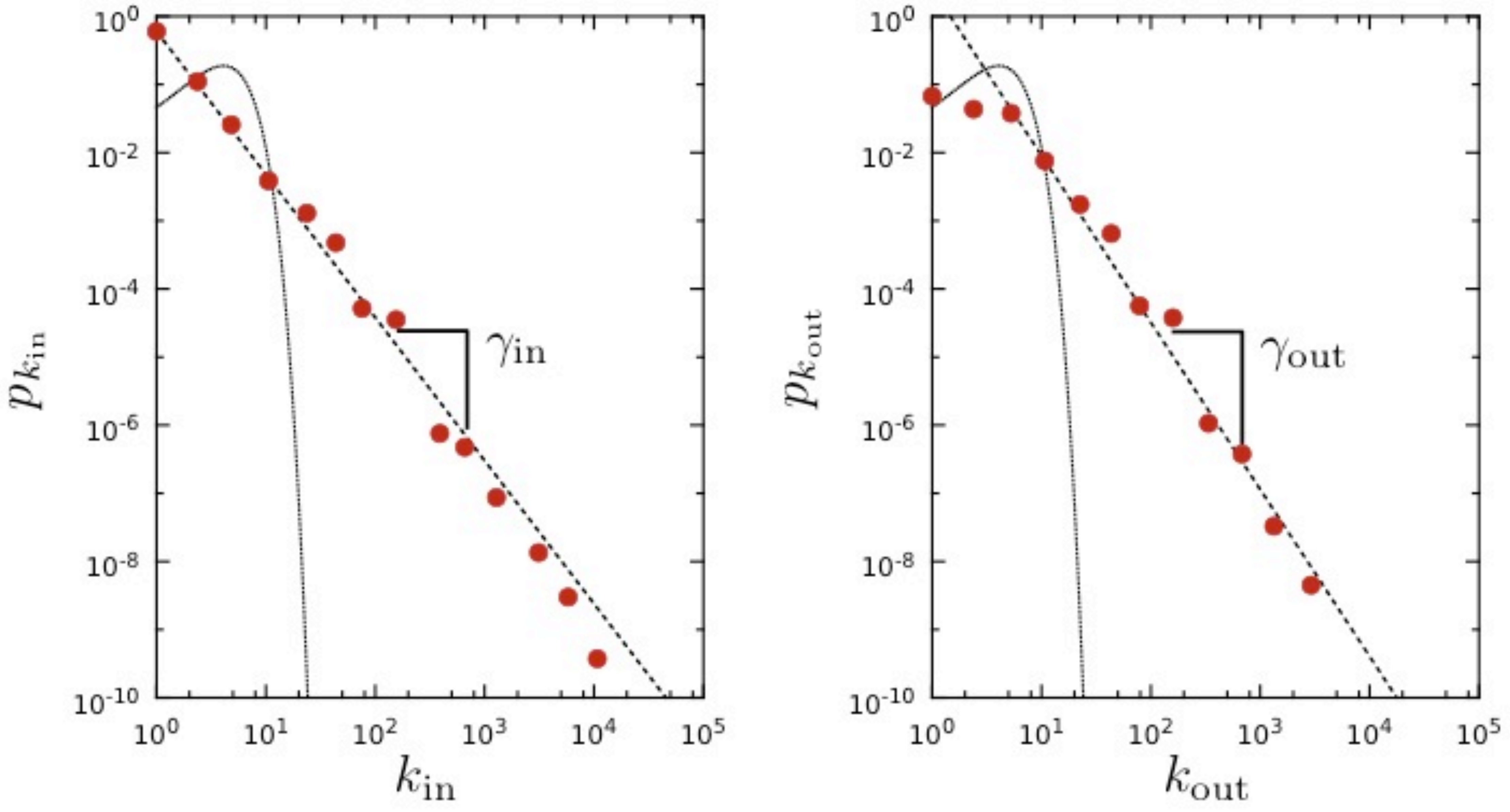


$$p_k \sim k^{-\gamma}$$

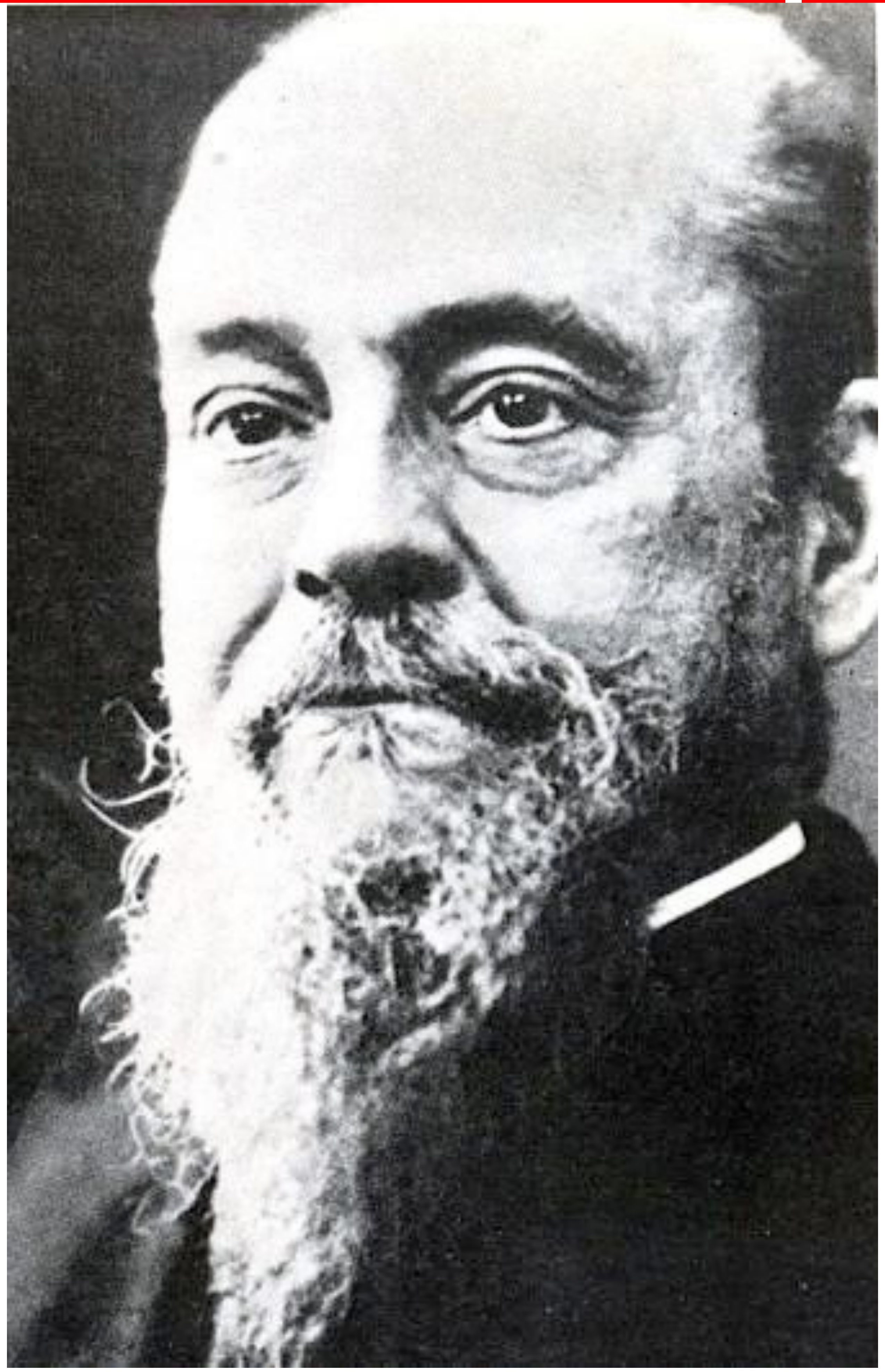
$$\log p_k \sim -\gamma \log k$$

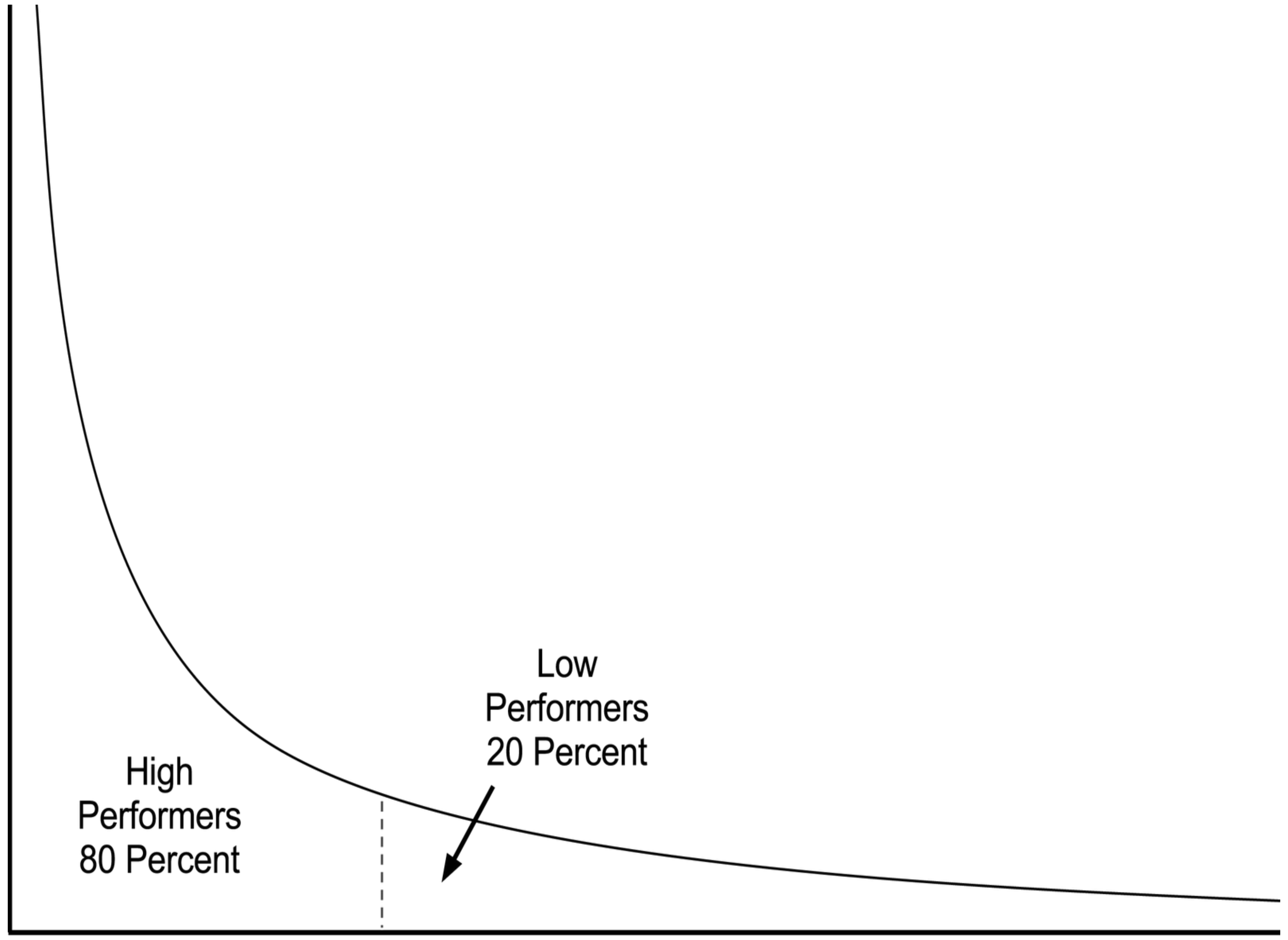


Expected



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

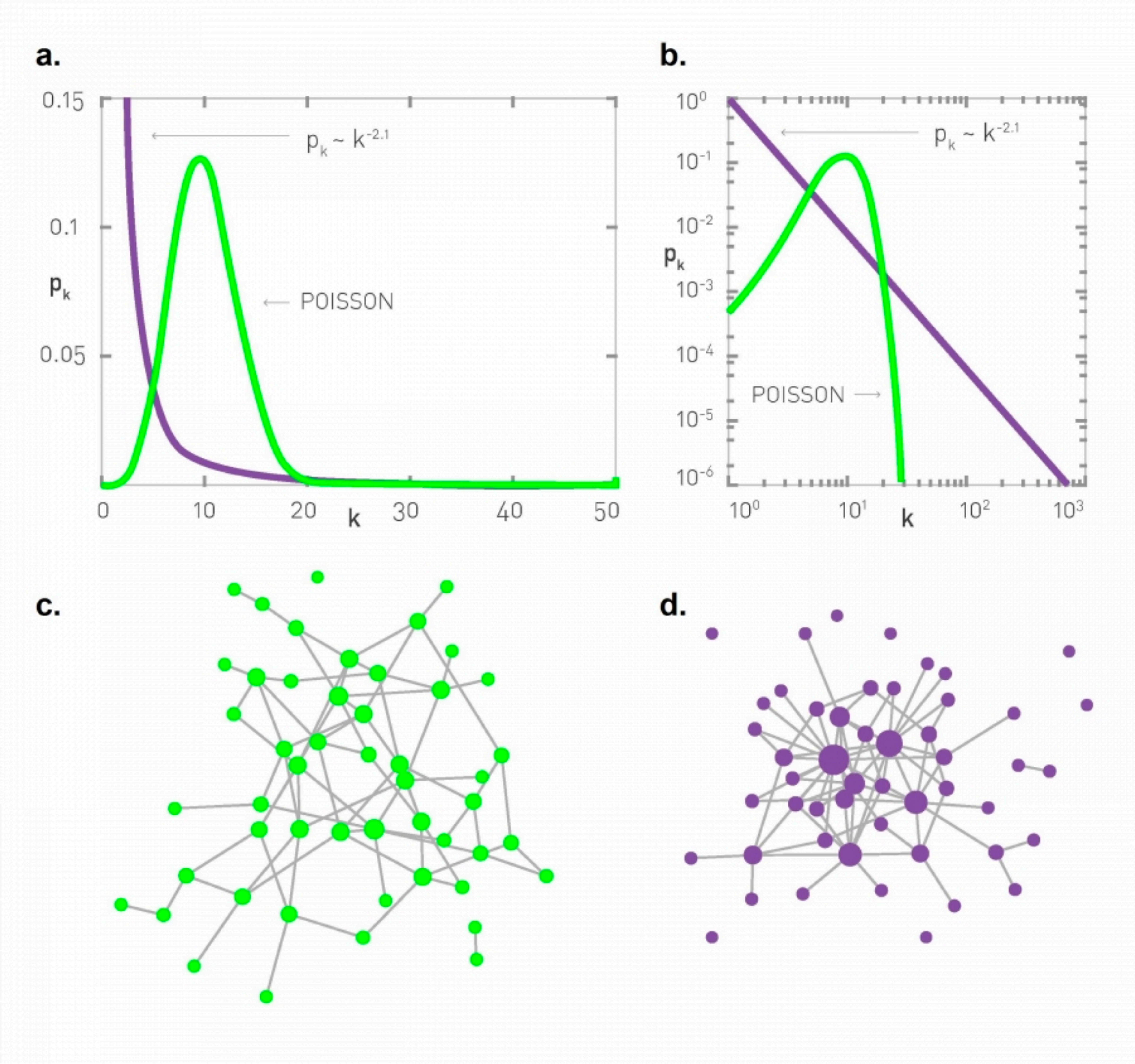




Vilfredo Federico Damaso Pareto (1848 – 1923), Italian economist, political scientist and philosopher, who had important contributions to our understanding of income distribution and to the analysis of individuals choices. A number of fundamental principles are named after him, like Pareto efficiency,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

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 where N* ≈ 10^{12} (current estimation).
The probability to have a node with k~100 is

- *About* $p_{100} \approx 10^{-94}$ *in a Poisson distribution*

- *About* $p_{100} \approx 4\text{x}10^{-4}$ *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^{-82} *k>100 degree nodes, or none.*

- *For a power law degree distribution, we expect about* $4\text{x}10^{9}$ *k>100 degree nodes*

The size of the biggest 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: the probability to have a node larger than k_{max} should not exceed the prob. to have one node, i.e. *1/N* fraction of all nodes

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

The size of the biggest 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 of Image 4.1, 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 λ=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 k_{max} 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*.

Finite scale-free networks

Expected maximum degree, k_{max}

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

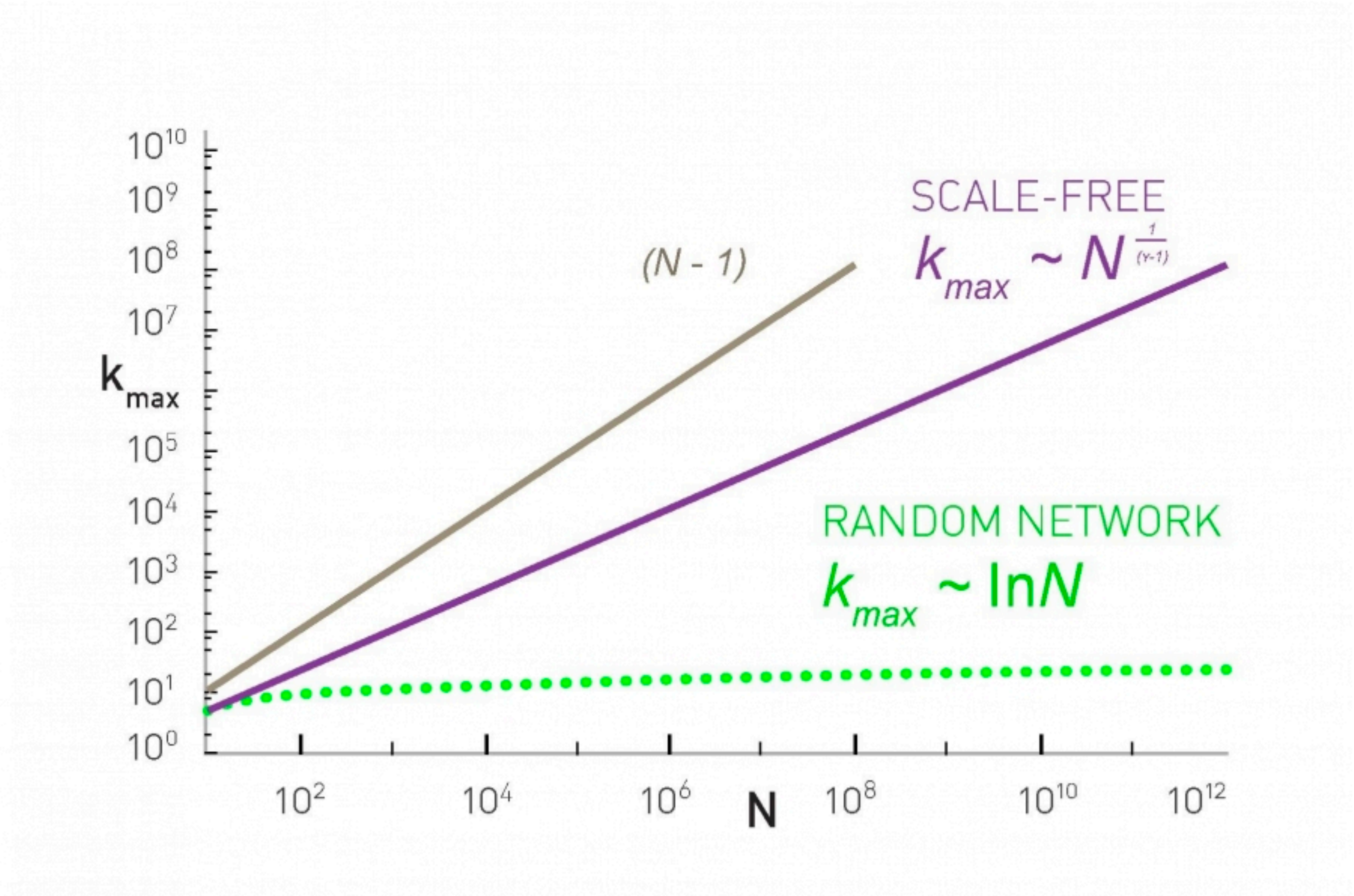
- k_{max}, increases with the size of the network
  → the larger a system is, the larger its biggest hub

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

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

- For $\gamma<2$ k_{max} increases faster than N
  → the largest hub will grab an increasing fraction of links. Anomaly!

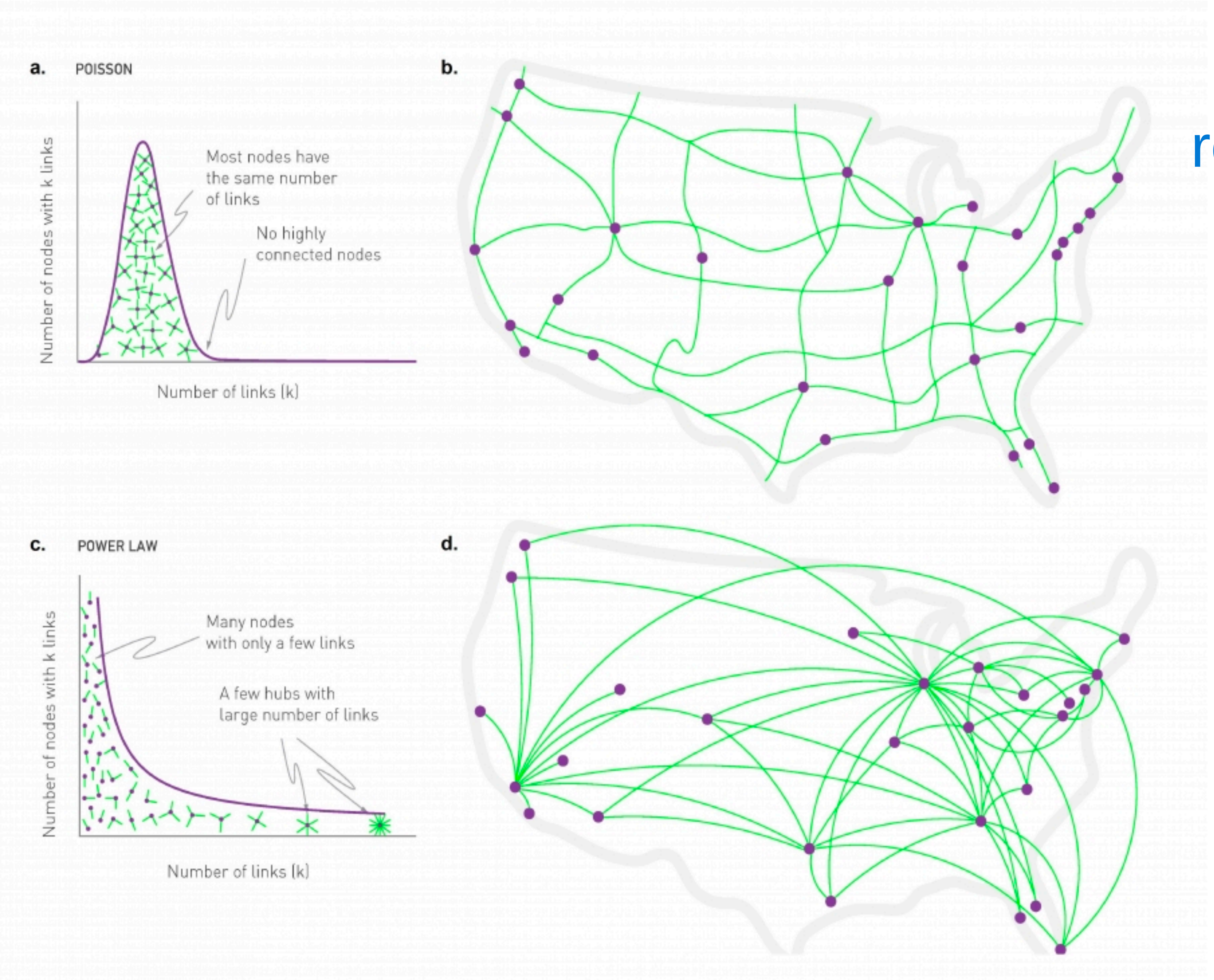
The size of the largest hub



growth of a random network is really slow

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

Random vs. Scale-free Network



road network

airplane network

The meaning of scale-free

Divergences in scale-free distributions

$$P(k) = Ck^{-\gamma} \qquad k = [k_{\min}, \infty)$$

power distribution with limits

Divergences in scale-free distributions

$$P(k) = Ck^{-\gamma} \qquad k = [k_{\min}, \infty)$$

$$\langle k^n \rangle = \sum_{k_{\min}}^{\infty} k^n p_k \approx \int_{k_{\min}}^{\infty} k^n p(k) dk$$

Divergences in scale-free distributions

$$P(k) = Ck^{-\gamma} \qquad k = [k_{\min}, \infty)$$

$$\langle k^n \rangle = \sum_{k_{\min}}^{\infty} k^n p_k \approx \int_{k_{\min}}^{\infty} k^n p(k) dk$$

$$\vdots$$

$$\langle k^n \rangle = \int_{k_{\min}}^{k_{\max}} k^n p(k) dk = C \frac{k_{\max}^{n-\gamma+1} - k_{\min}^{n-\gamma+1}}{n-\gamma+1}$$

- If $n - \gamma + 1 \leq 0$ then the first term on the r.h.s. of (4.20), $k_{max}^{n-\gamma+1}$, goes to zero as k_{max} increases. Therefore all moments that satisfy $n \leq \gamma - 1$ are finite.
- If $n - \gamma + 1 > 0$ then $\langle k^n \rangle$ goes to infinity as $k_{max} \to \infty$. Therefore all moments larger than $\gamma - 1$ diverge.

DIVERGENCE OF THE HIGHER MOMENTS

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*-

Scale-free networks: Definition

Definition:

Networks with a power law tail in their degree distribution are called 'scale-free networks'

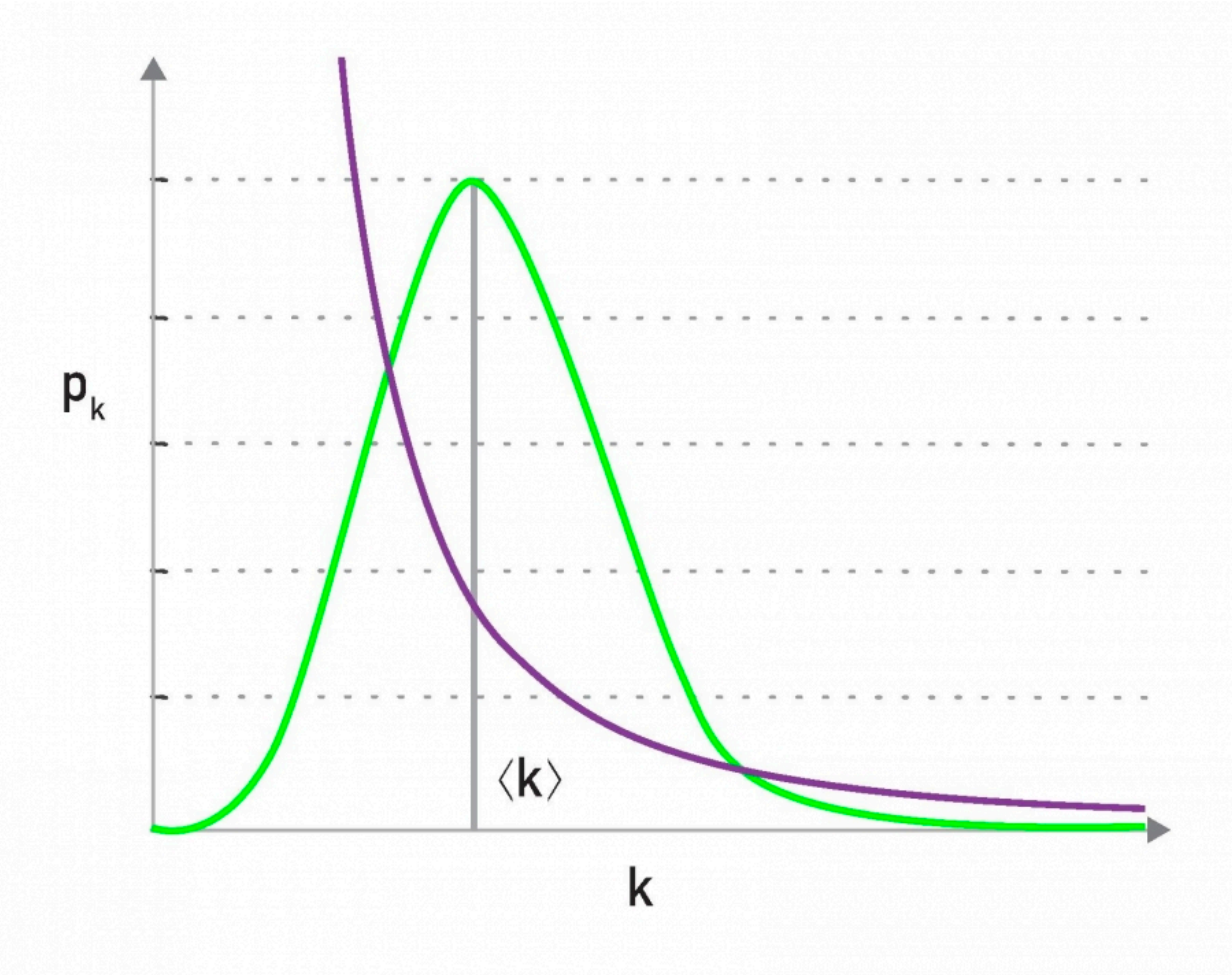
Where does the name come from?

Critical Phenomena and scale-invariance

(a detour)

Slides after **Dante R. Chialvo**

Lack of an Internal Scale



Random Network

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

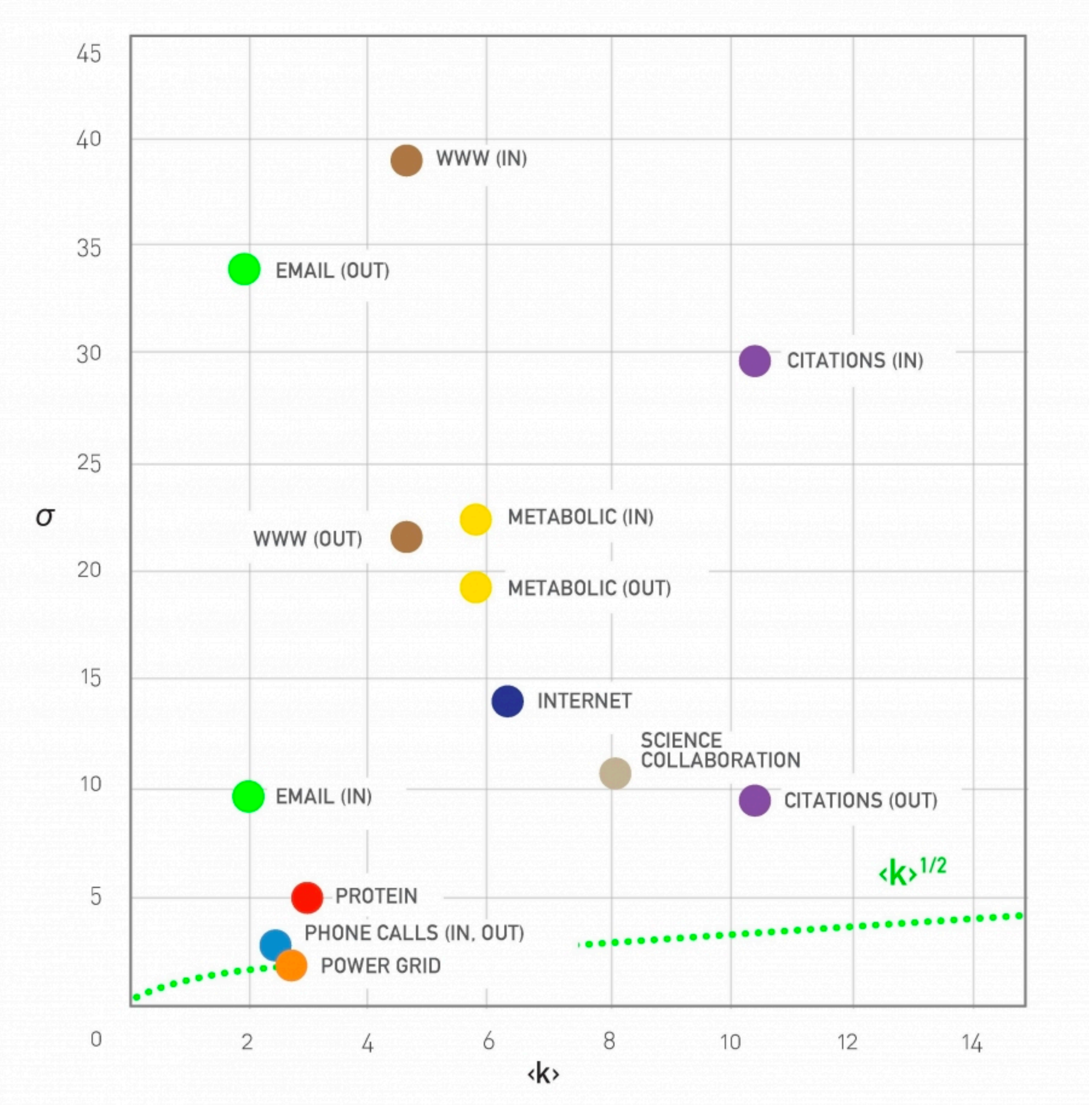
Scale: $\langle k \rangle$

Scale-Free Network

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

Scale: none

Standard Deviation in Real Networks

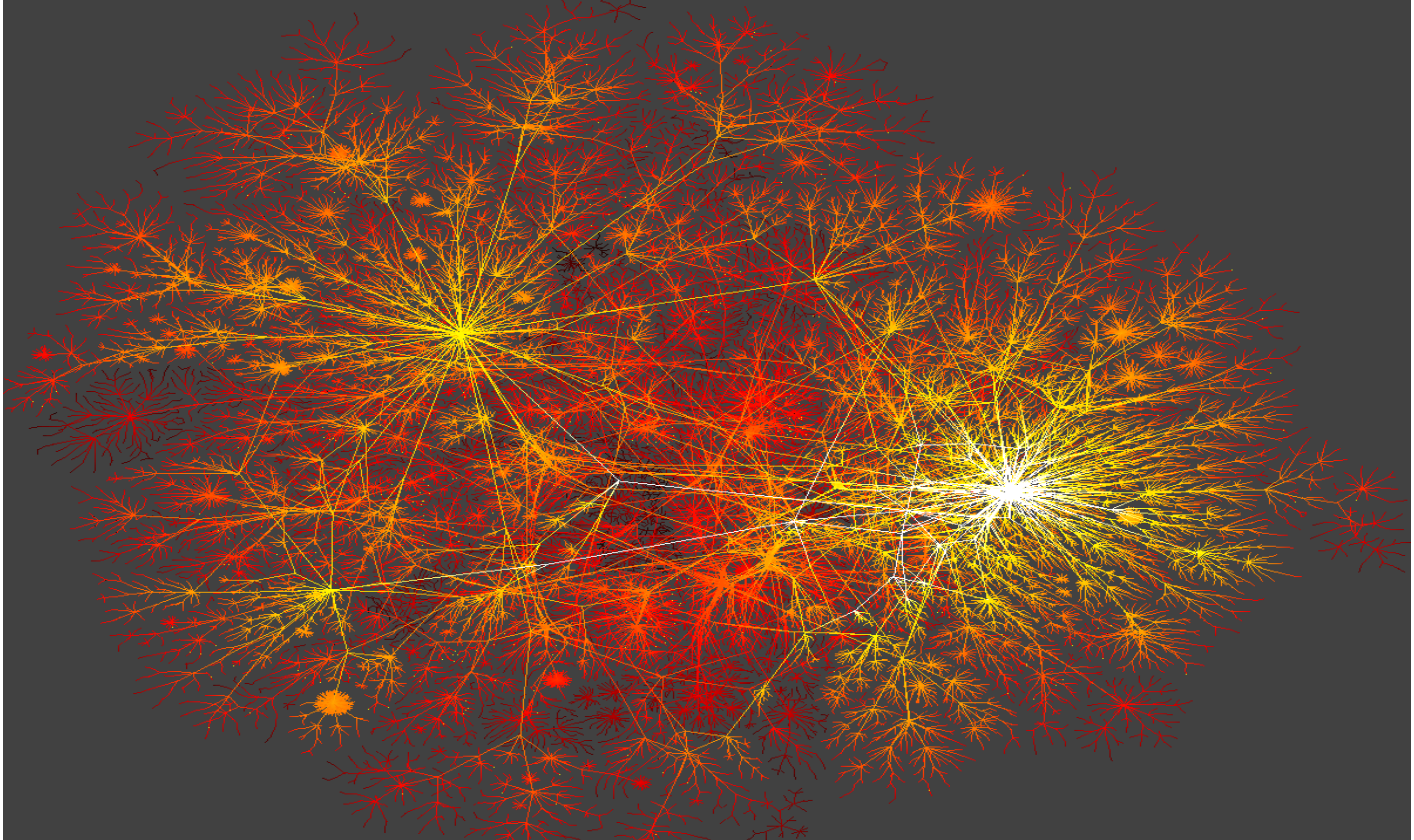


CRITICAL PHENOMENA

- Correlation length diverges at the critical point: the whole system is correlated!

- **Scale invariance:** there is no characteristic scale for the fluctuation **(scale-free behaviour).**

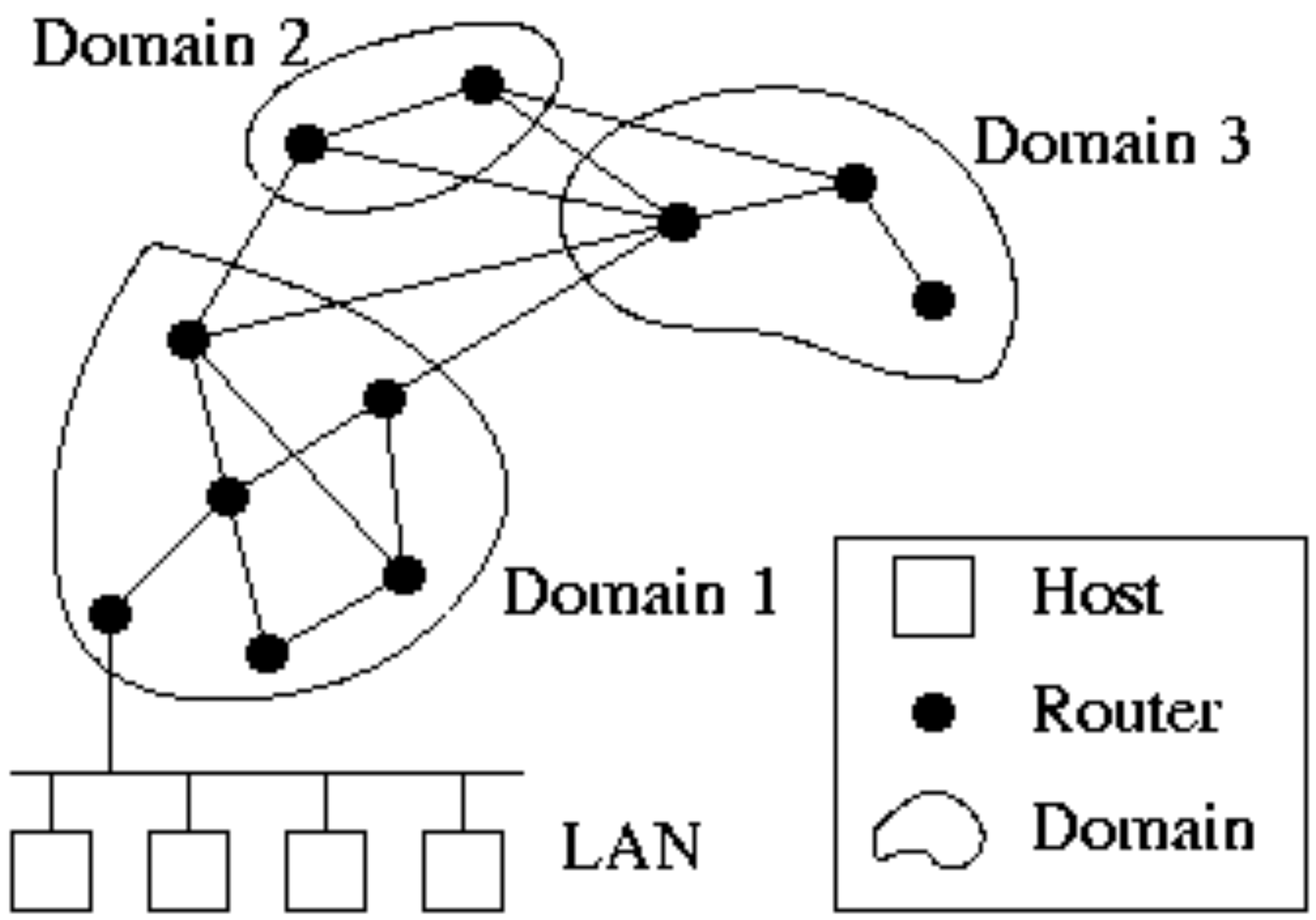
- **Universality:** exponents are independent of the system's details.

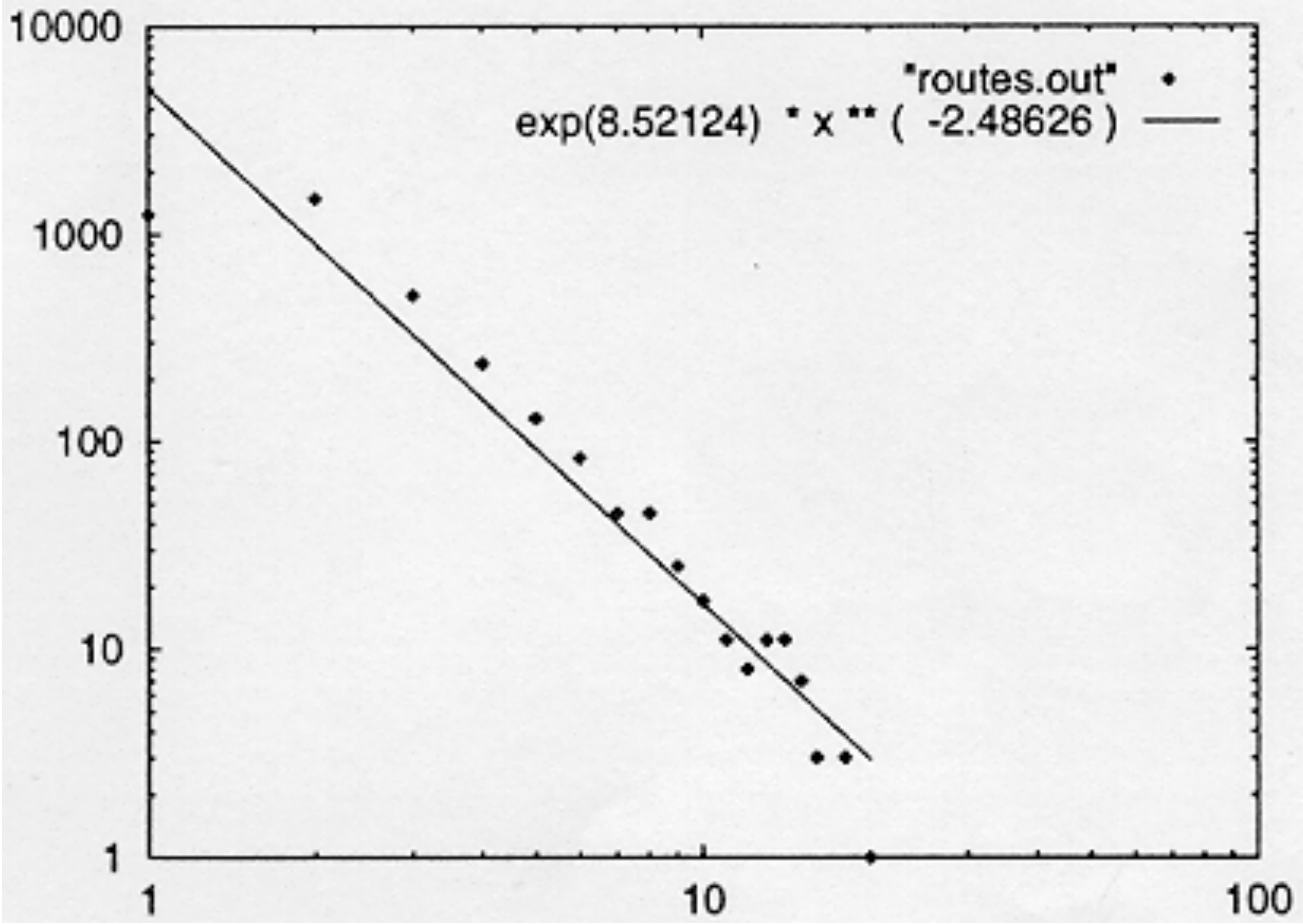


Network Science: Scale-Free Property

INTERNET BACKBONE

Nodes: computers, routers
Links: physical lines

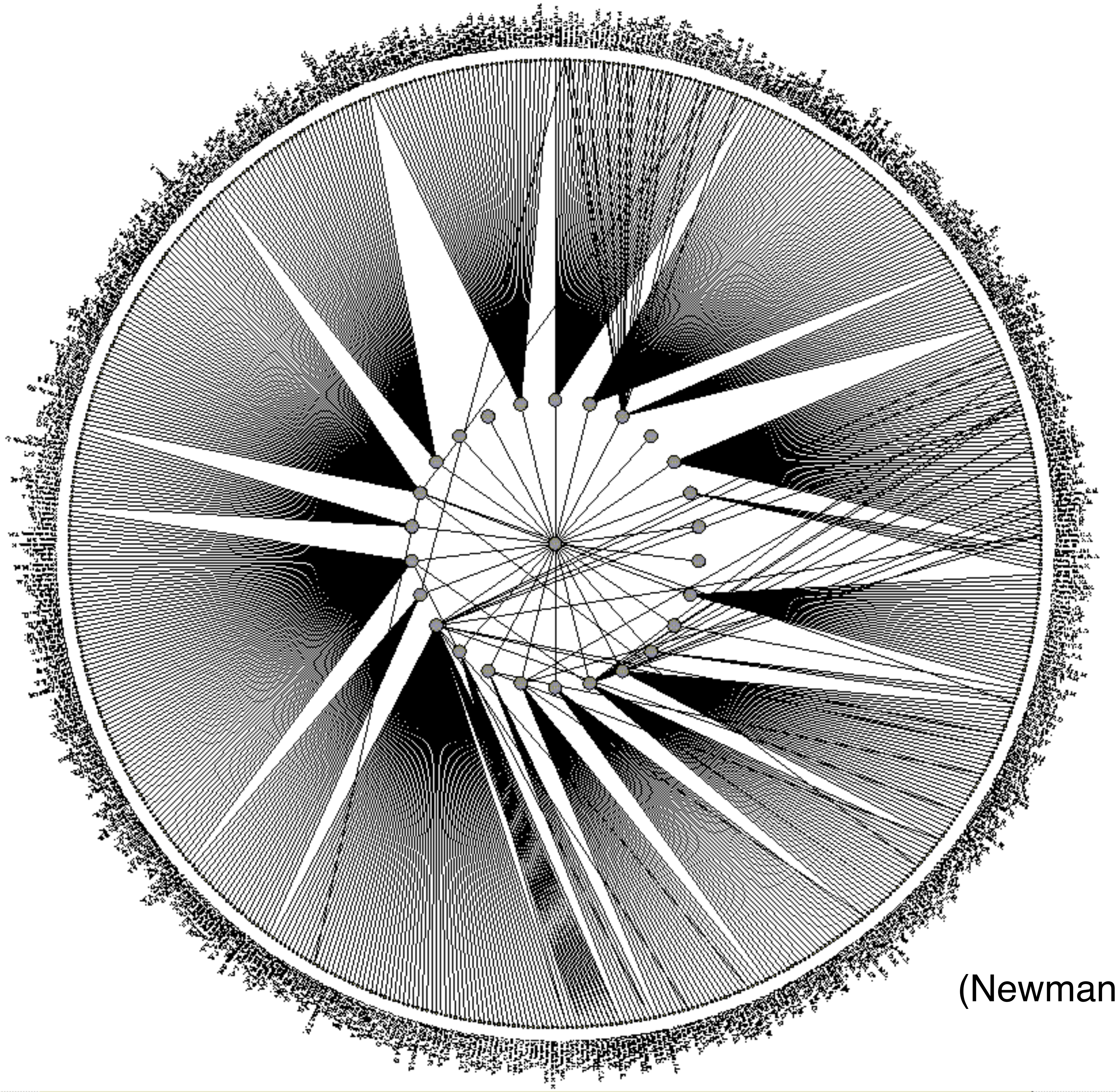


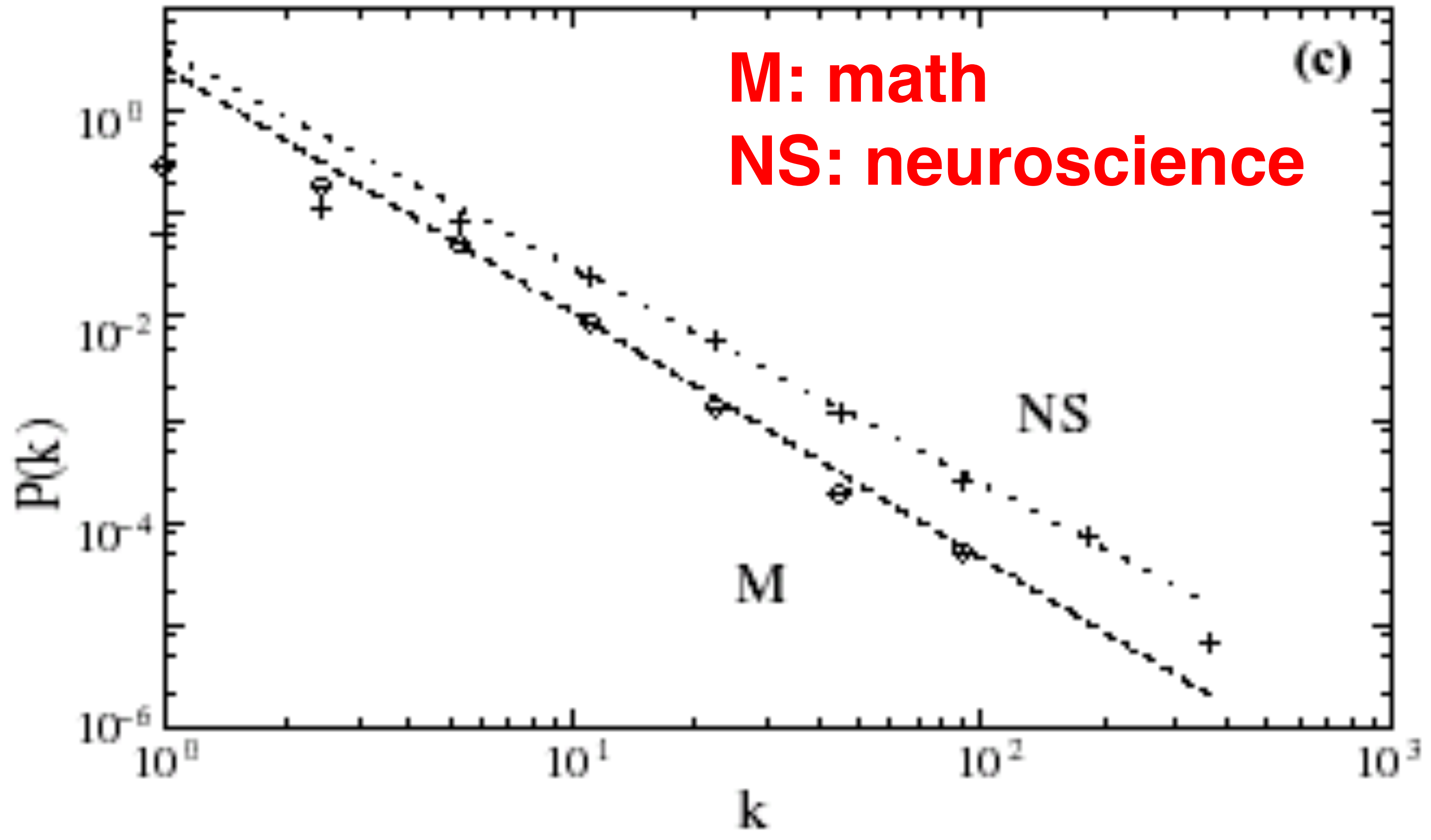


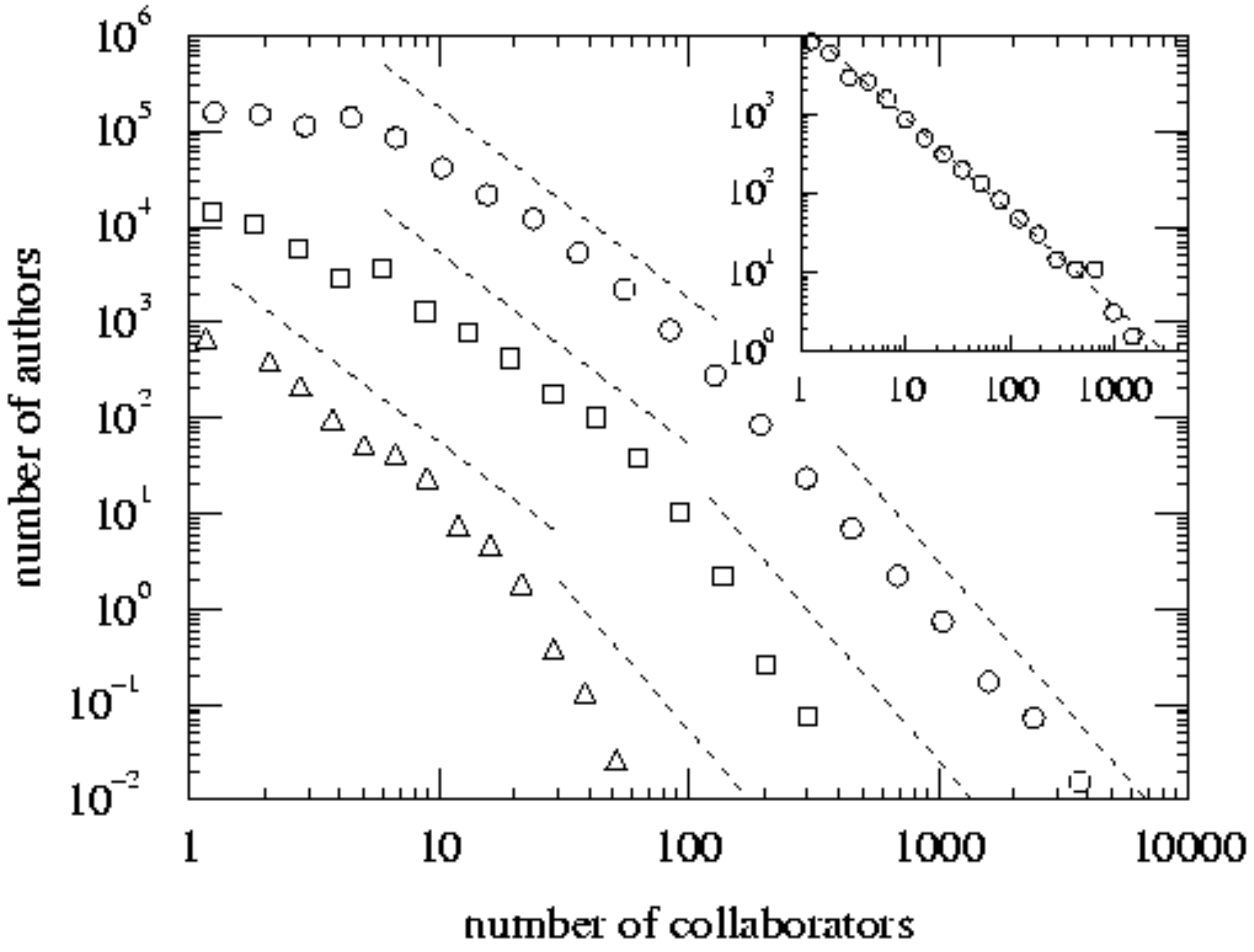
(Faloutsos, Faloutsos and Faloutsos, 1999)

SCIENCE COAUTHORSHIP

Nodes: scientist (authors)
Links: joint publication



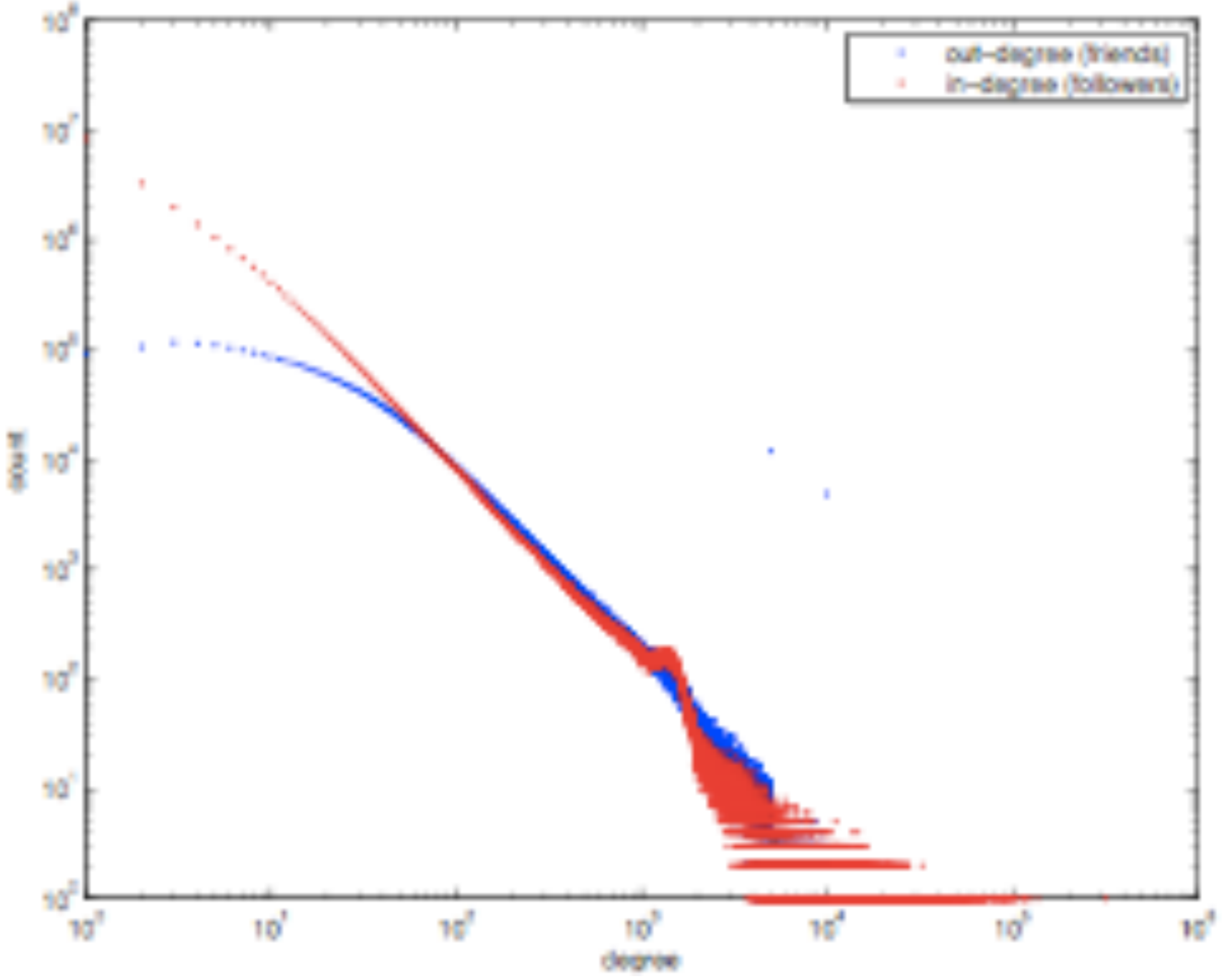




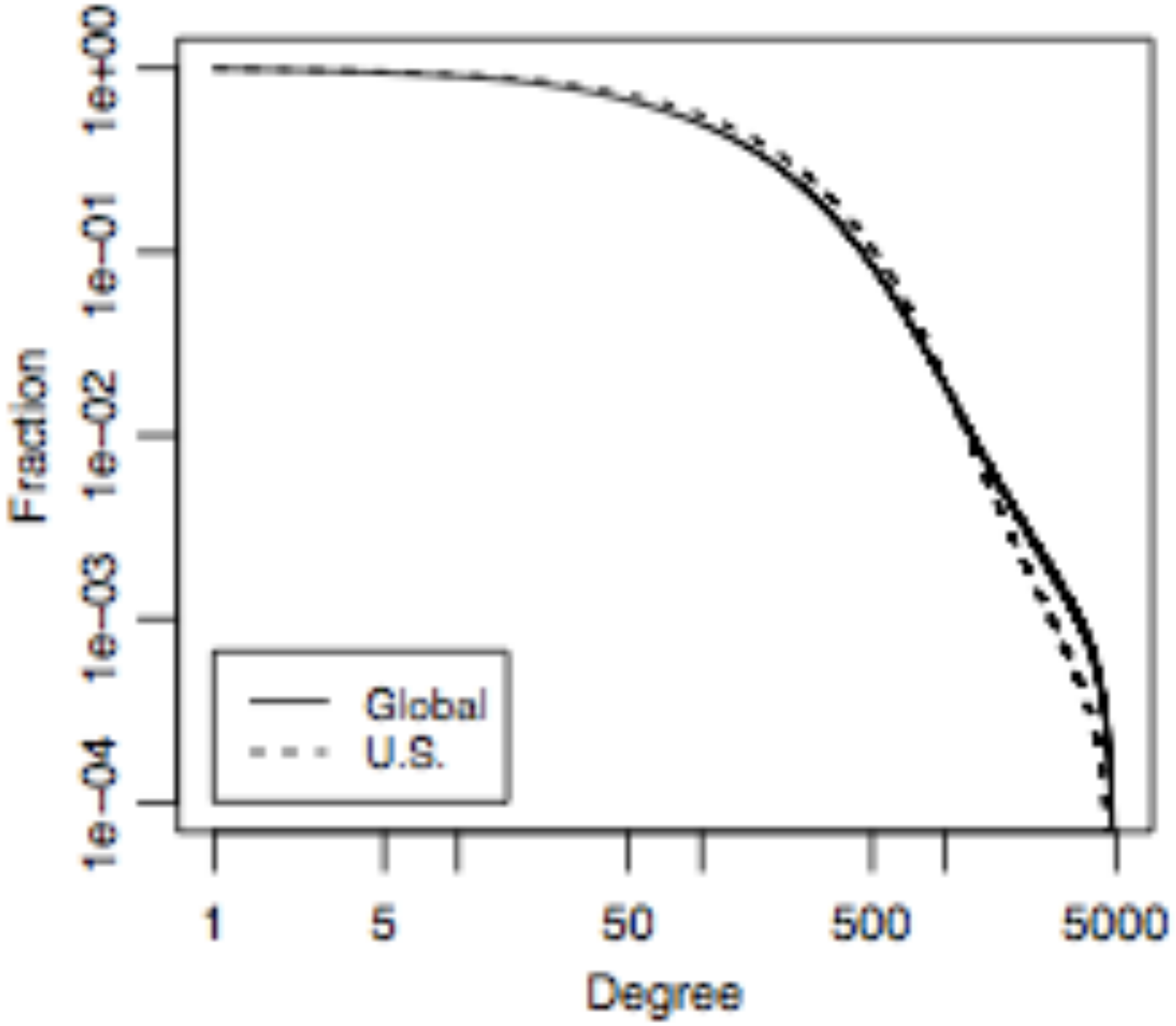
(Newman, 2000, Barabasi et al 2001)

ONLINE COMMUNITIES

Twitter:

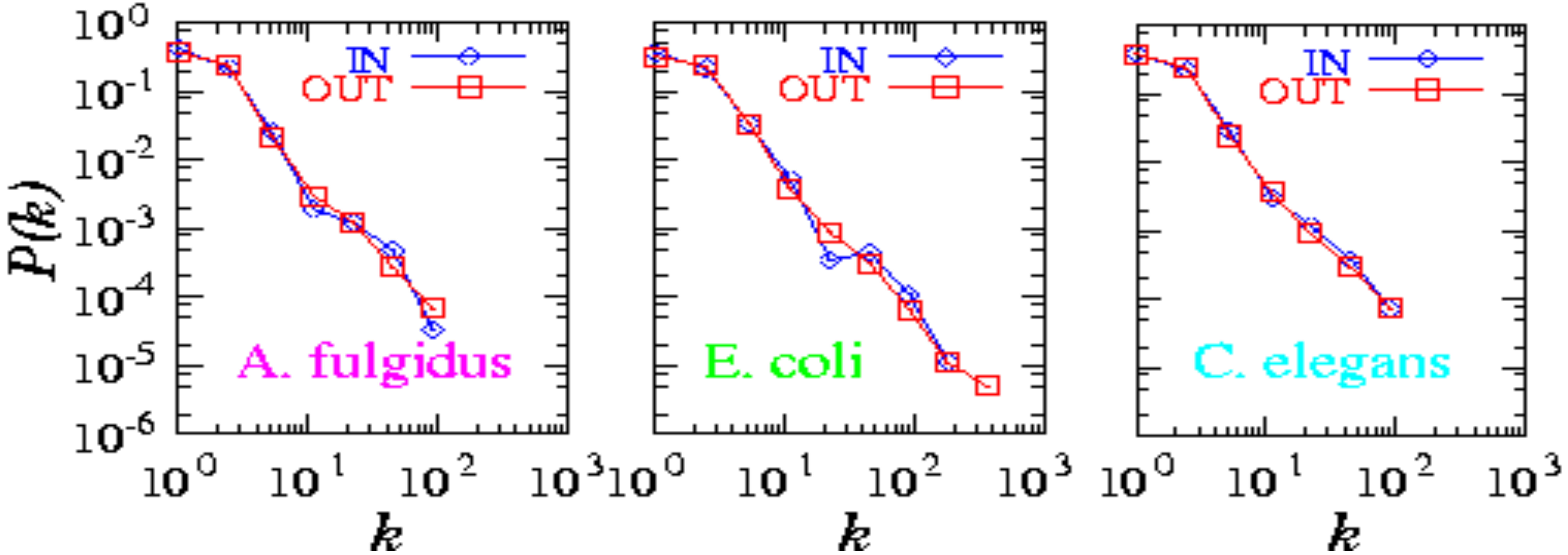


Facebook



Brian Karrer, Lars Backstrom, Cameron Marlowm 2011

METABOLIC NETWORK



Archaea Bacteria Eukaryotes

Organisms from all three domains of life are scale-free!

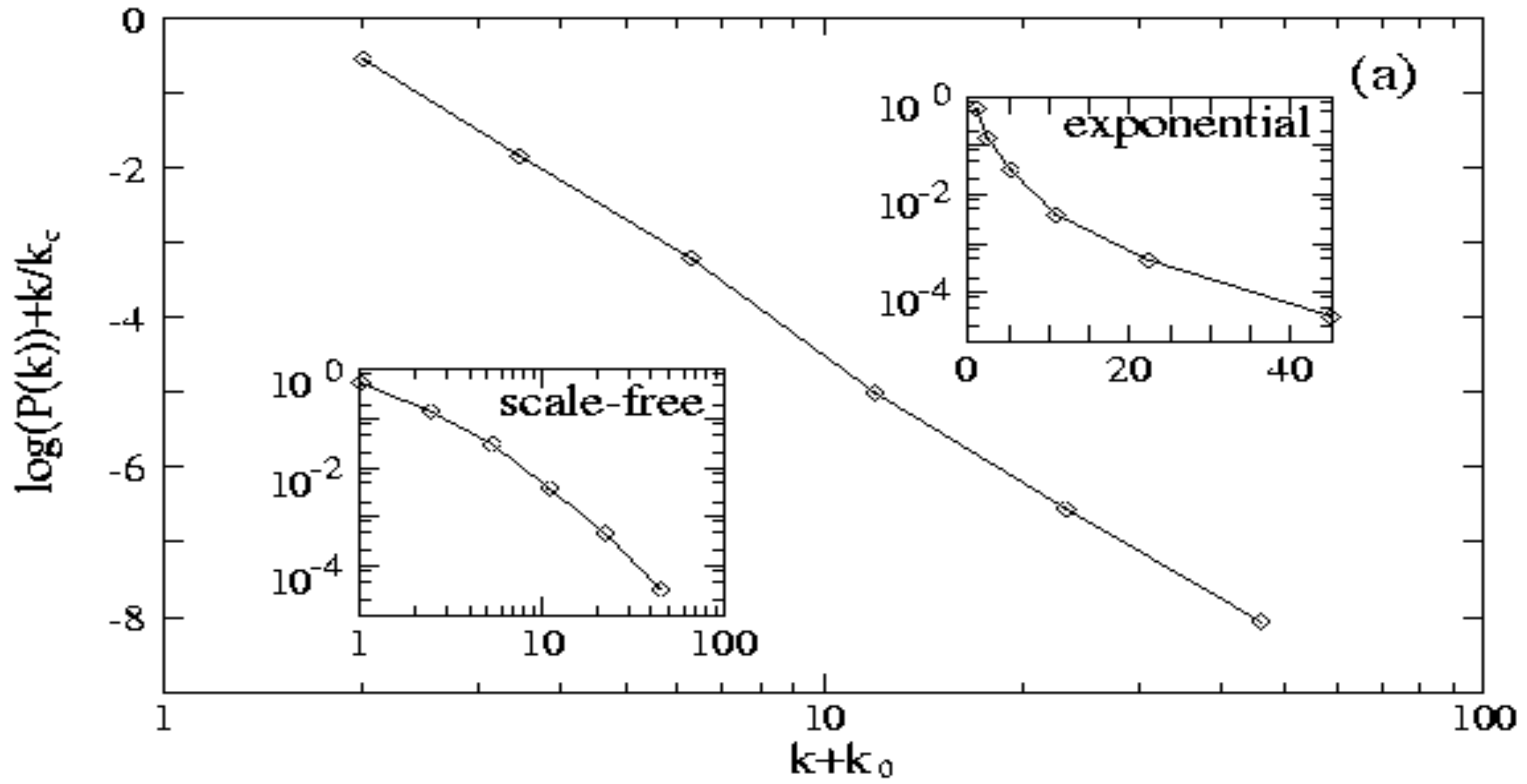
$$P_{in}(k) \approx k^{-2.2}$$

$$P_{out}(k) \approx k^{-2.2}$$

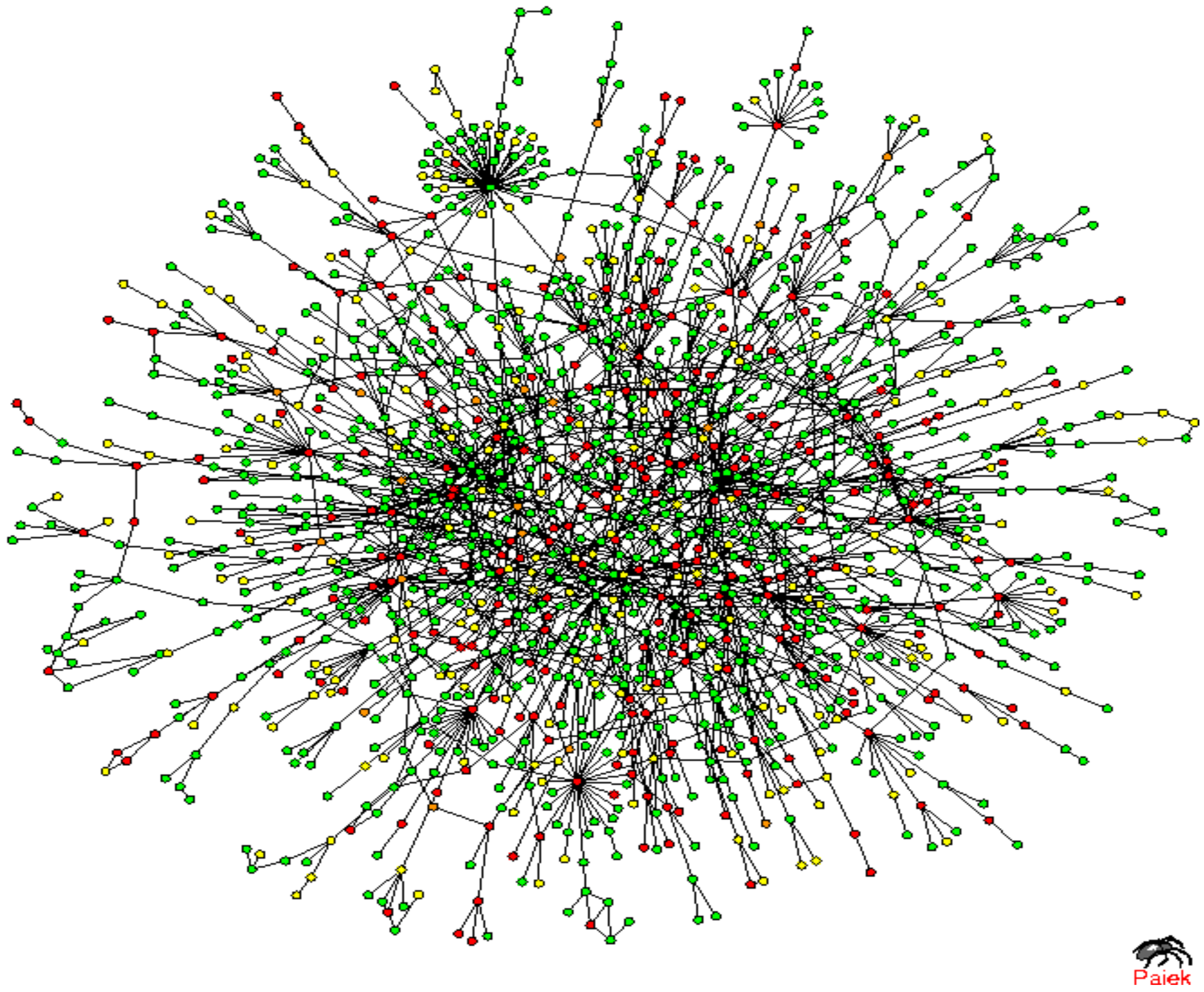
H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, and A.L. Barabasi, *Nature*, 407 651 (2000)

TOPOLOGY OF THE PROTEIN NETWORK

Nodes: proteins
Links: physical interactions-binding

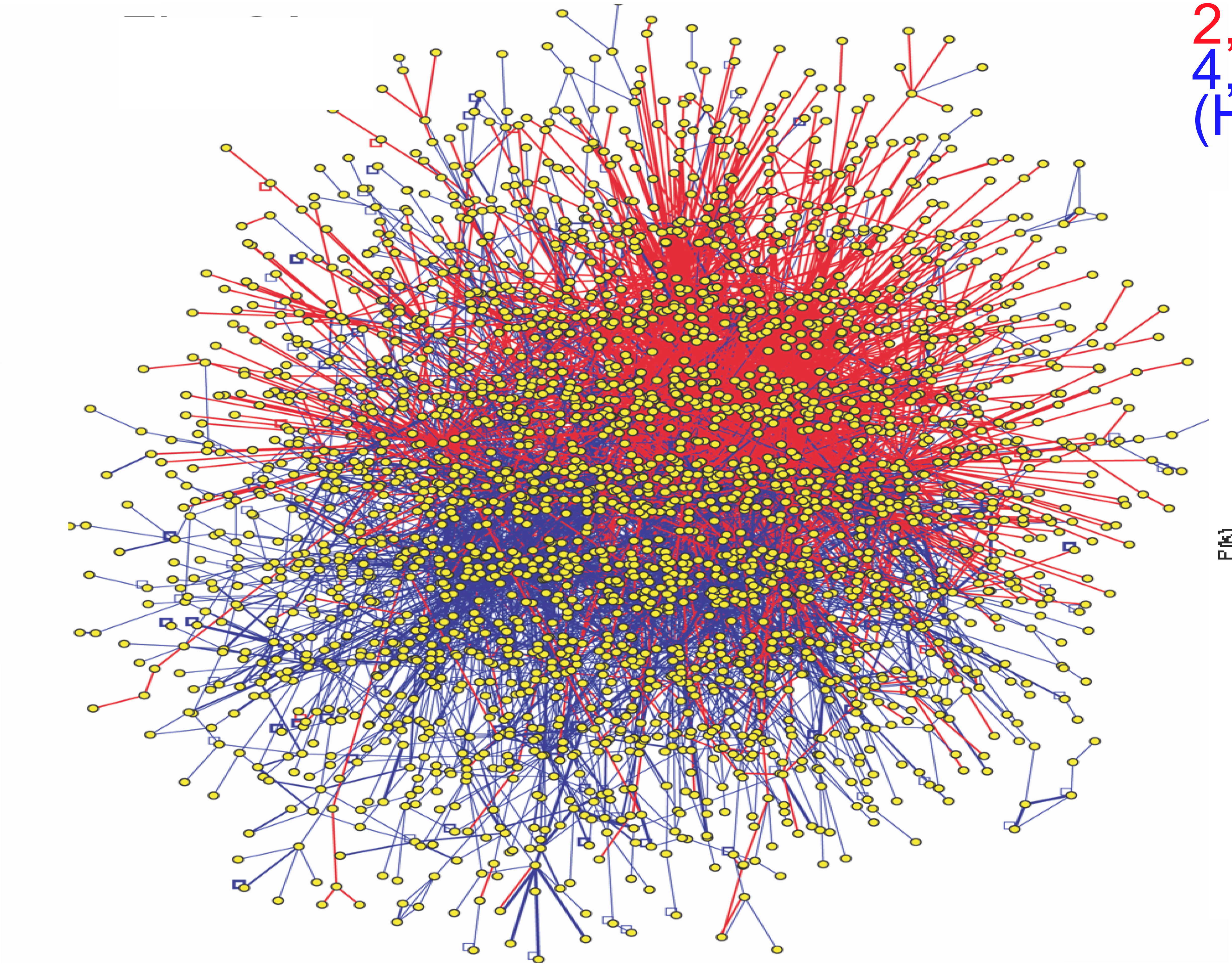


$$P(k) \sim (k + k_0)^{-\gamma} \exp(-\frac{k + k_0}{k_\tau})$$

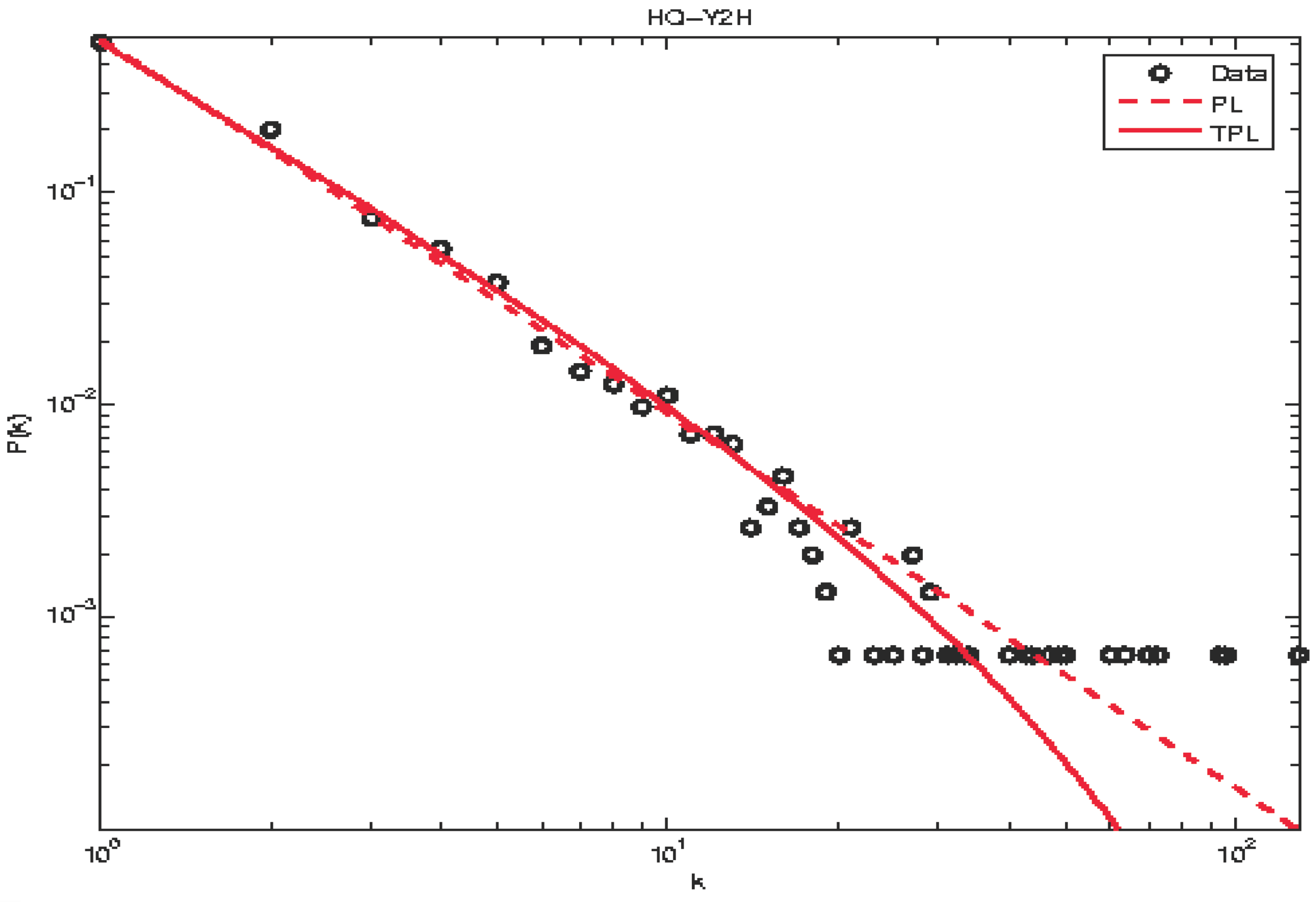


H. Jeong, S.P. Mason, A.-L. Barabasi, Z.N. Oltvai, Nature 411, 41-42 (2001)

HUMAN PROTEIN-PROTEIN INTERACTION NETWORK



2,800 Y2H interactions
4,100 binary LC interactions
(HPRD, MINT, BIND, DIP, MIPS)



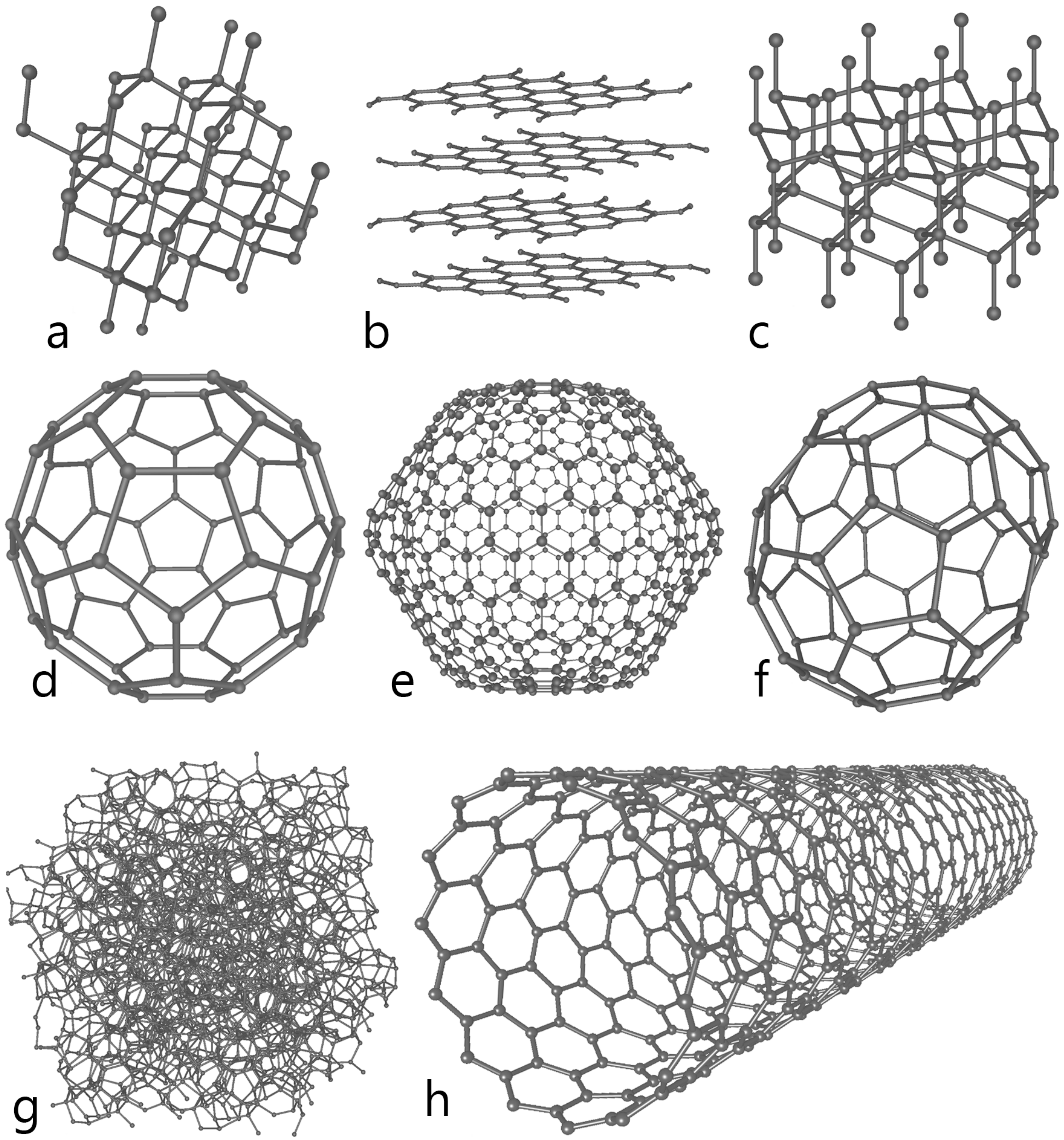
Rual *et al.* Nature 2005; Stelze *et al.* Cell 2005

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.*

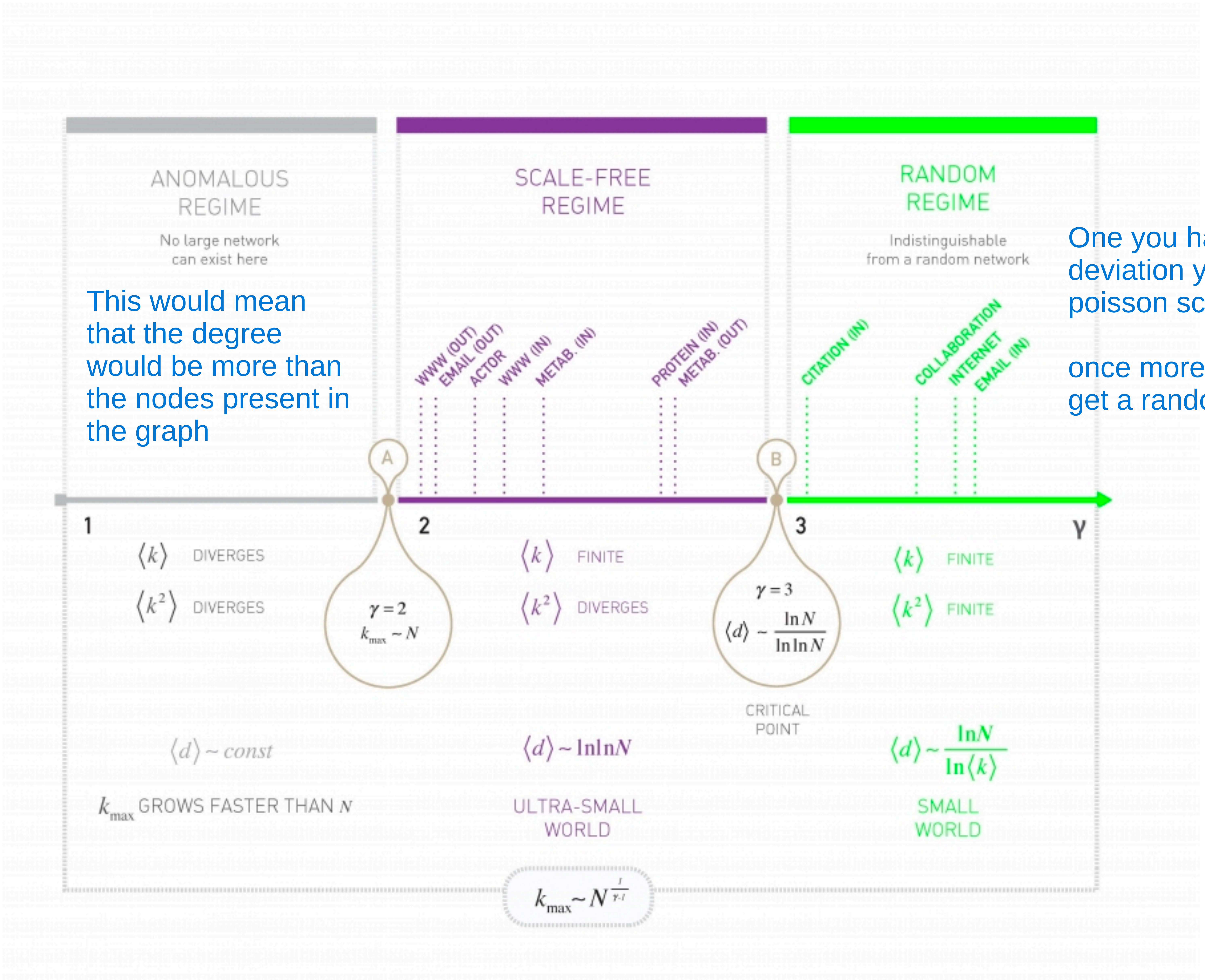
•*The neural network of the C.elegans worm.*

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



The role of the degree exponent

SUMMARY OF THE BEHAVIOR OF SCALE-FREE NETWORKS

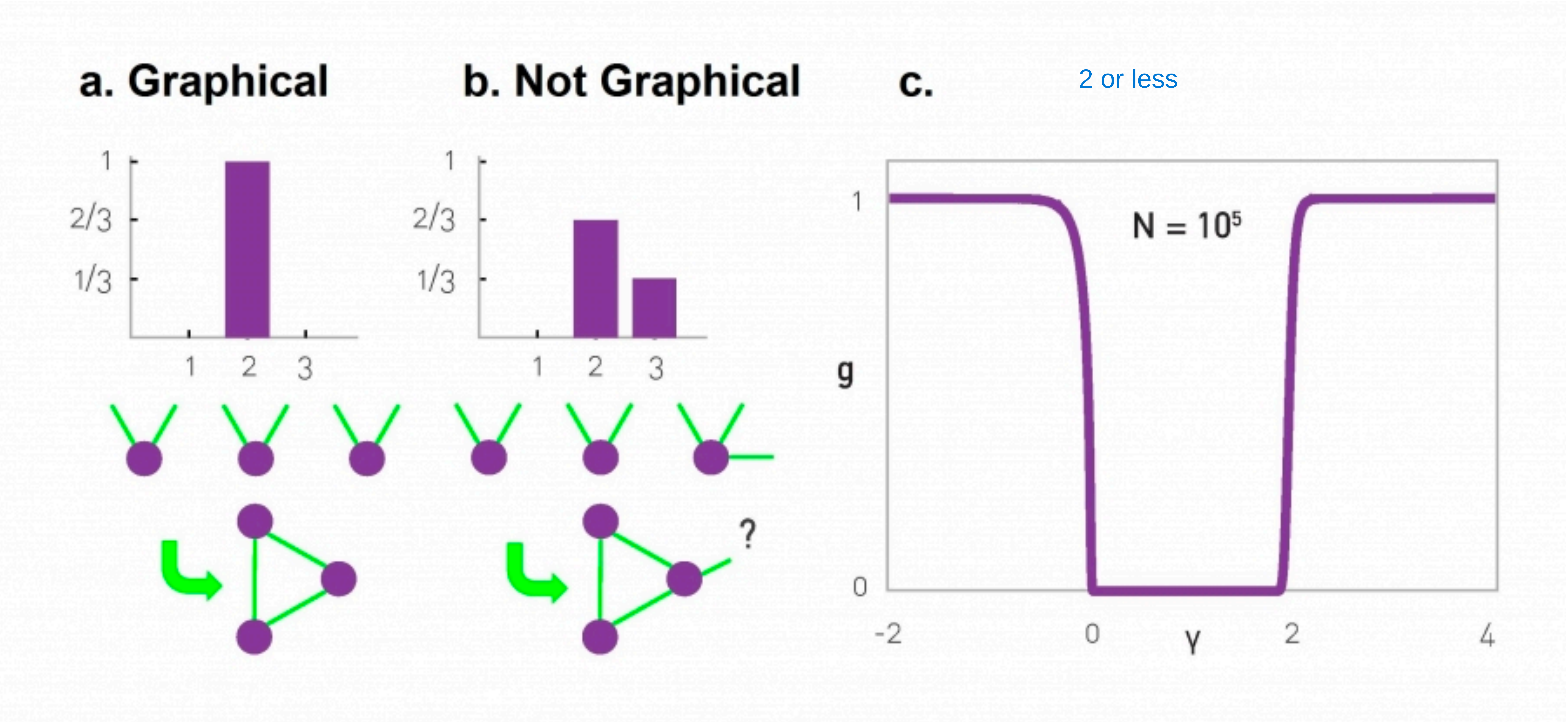


This would mean that the degree would be more than the nodes present in the graph

One you have a standard deviation you say you have a poisson scale.

once more than 3 you usually get a random regime

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

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

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

it is possible to have networks in a higher degree but at the moment we dont have the data for it.

for example you can put the www in the 3rd degree just because it has trillions of nodes

the network just needs to be large enough

Onella et al. PNAS 2007

Questions?