Network Analysis:

The Hidden Structures behind the Webs We Weave

17-338 / 17-668

Scale-Free Networks

Tuesday, October 29, 2024

Patrick Park & Bogdan Vasilescu

Carnegie Mellon University
School of Computer Science

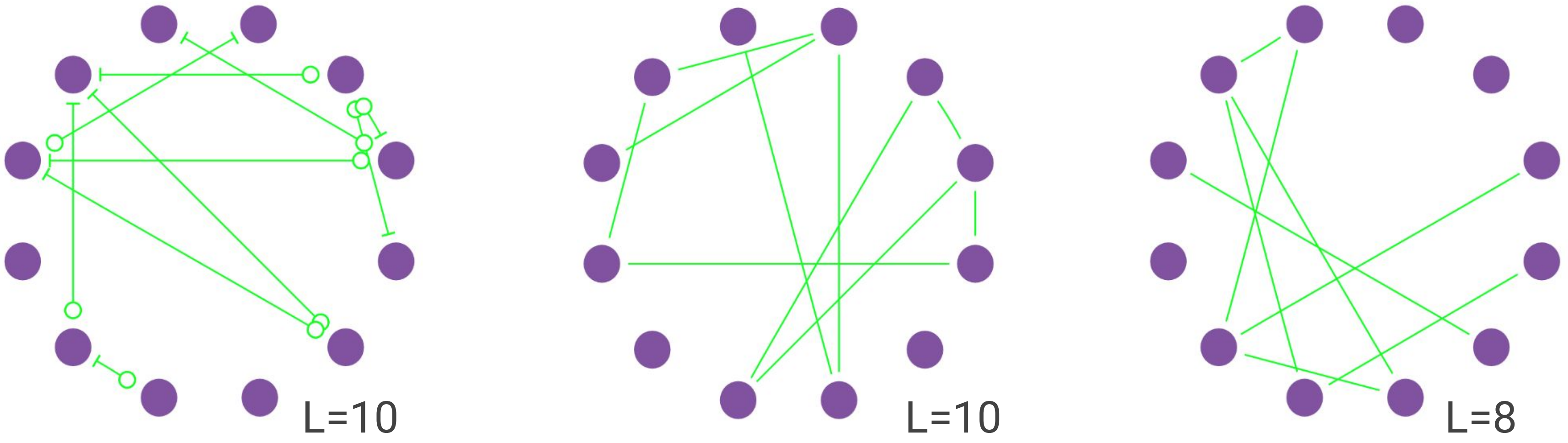
S3D
Software and Societal Systems Department

2-min Quiz, on Canvas

Recall: The Random Network Model

A random network consists of N nodes where each node pair is connected with probability p

Aka “Erdős-Rényi network” – from random graph theory (1959–1968)



Three realizations of a random network generated with the same parameters p=1/6 and N=12.

The Degree Distribution of Random Networks

In a given realization of a random network some nodes gain many links, while others acquire only a few or no links.

These differences are captured by the degree distribution, p_k, which is the probability that a randomly chosen node has degree k.

In a random network the probability that node i has exactly k links is the product of three terms

- The probability that k of its links are present, or p^k.
- The probability that the remaining (N-1-k) links are missing, or $(1-p)^{N-1-k}$.
- The number of ways we can select k links from N- 1 potential links a node can have, or $\binom{N-1}{k}$

Consequently the degree distribution of a random network follows the binomial distribution

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

Aside: The Binomial Distribution

If we toss a fair coin N times, tails and heads occur with probability p = 1/2.

The binomial distribution provides the probability p_x that we obtain exactly x heads in a sequence of N throws.

In general, the binomial distribution describes the number of successes in N independent experiments with two possible outcomes, in which the probability of one outcome is p, and of the other is 1-p.

$$p_x = \binom{N}{x} p^x (1-p)^{N-x}$$

Aside: The Binomial Distribution - Useful Properties

The mean of the distribution (first moment) is

$$\langle x \rangle = \sum_{x=0}^{N} x p_x = Np$$

Its second moment is

$$\langle x^2 \rangle = \sum_{x=0}^{N} x^2 p_x = p(1-p)N + p^2 N^2$$

providing its standard deviation as

$$\sigma_x = \left(\langle x^2 \rangle - \langle x \rangle^2 \right)^{\frac{1}{2}} = [p(1-p)N]^{\frac{1}{2}}$$

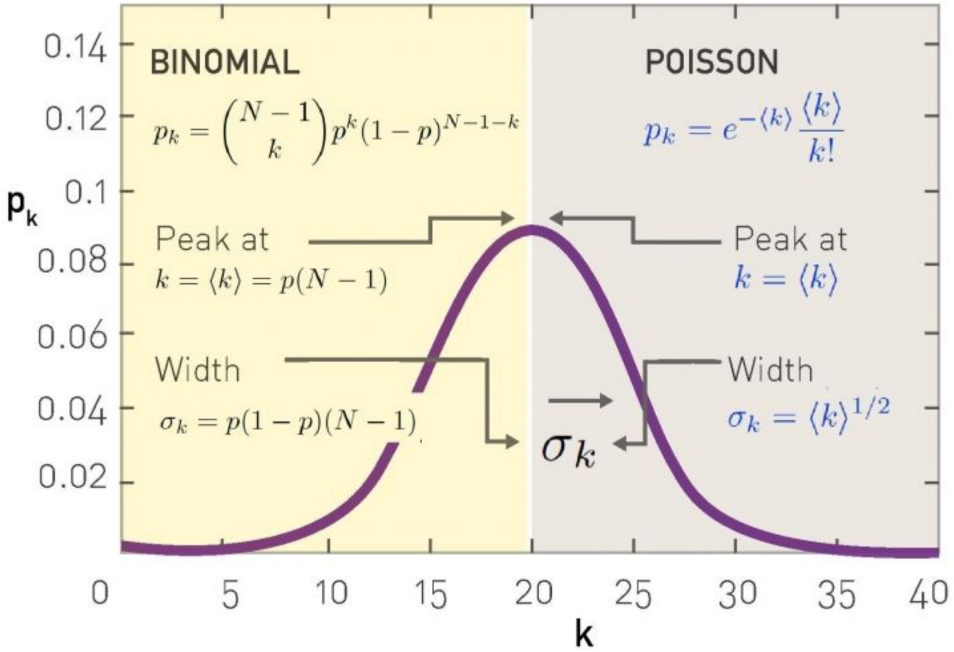
Binomial vs. Poisson Degree Distribution

The exact form of the degree distribution of a random network is the binomial distribution.

For N ≫ <k> the binomial is well approximated by a Poisson distribution, expressed in terms of different parameters:

- binomial depends on p and N
- Poisson depends on <k>

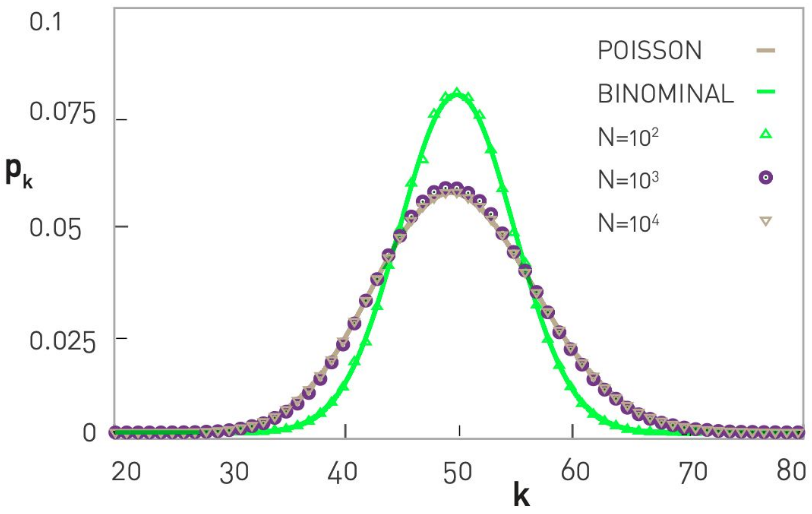
It is this simplicity that makes the Poisson form preferred in calculations.



Degree Distribution is Independent of the Network Size

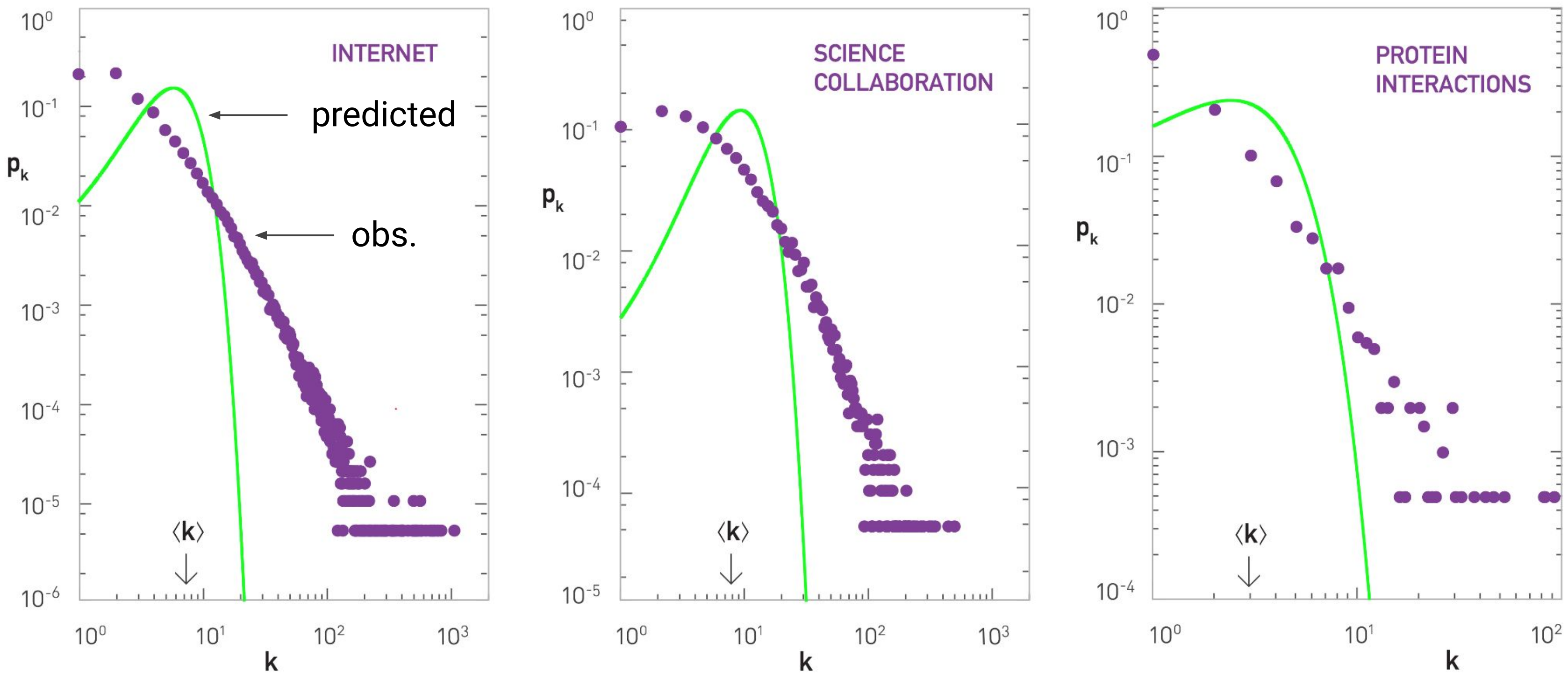
Small Networks: Binomial
For a small network ($N = 10^2$) the degree distribution deviates significantly from the Poisson form (recall assumption N » <k>).

Large Networks: Poisson
For larger networks ($N = 10^3, 10^4$) the degree distribution becomes indistinguishable from the Poisson prediction. Therefore for large N the degree distribution is independent of the network size.



The degree distribution of a random network with <k> = 50 and $N = 10^2, 10^3, 10^4$. (Averaged over 1,000 independently generated random networks)

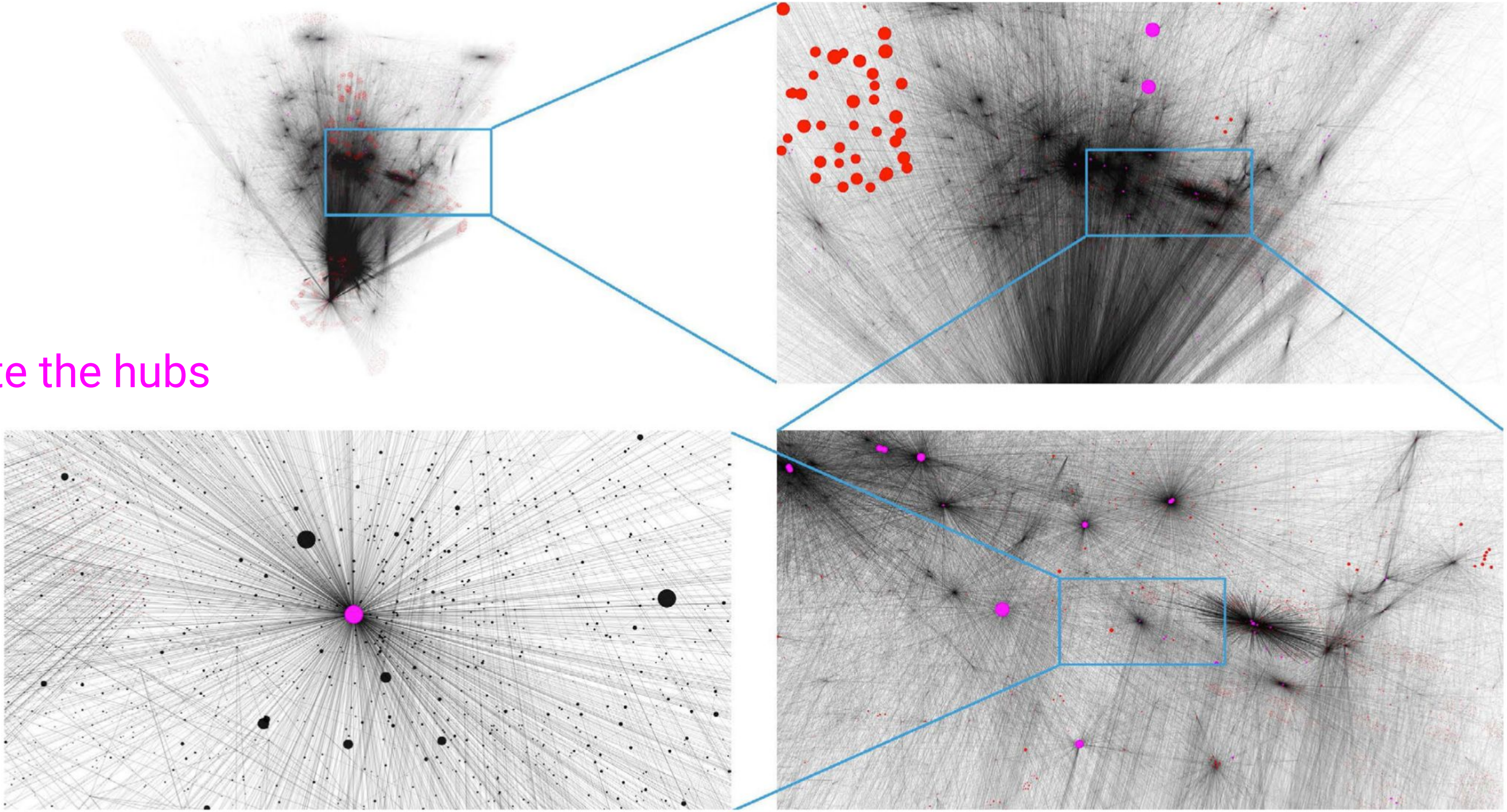
The random network model underestimates the size and frequency of the high degree nodes, and the number of low degree nodes.



(Barabasi Ch. 3.5)

The Scale-Free Property

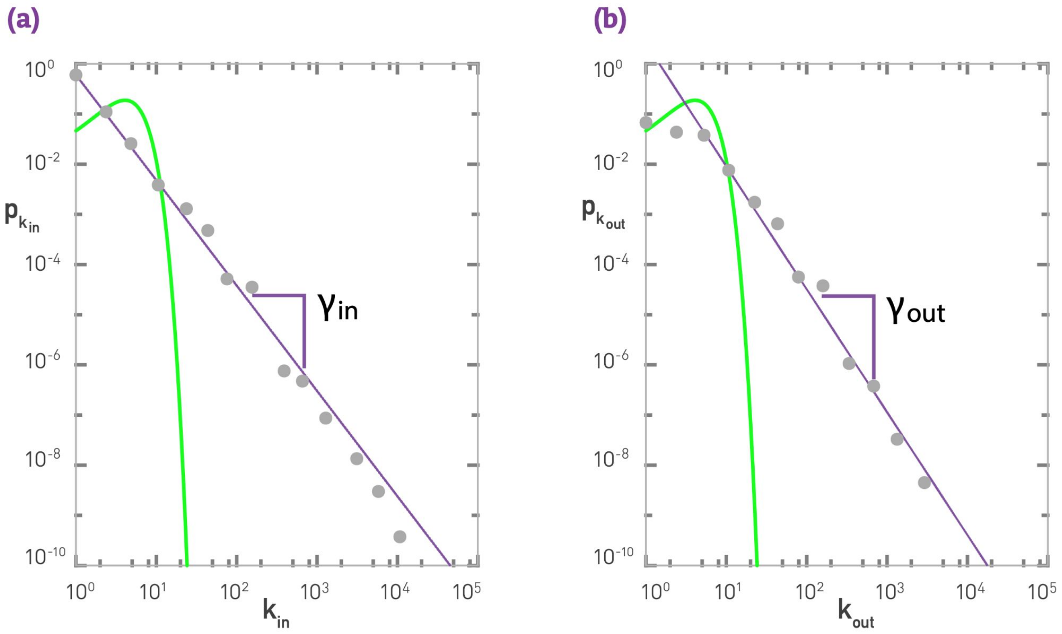
The Topology of the World Wide Web in 1998 (nodes are pages, links are URLs; map of the nd.edu domain, ~300k pages and 1.5M links)



Note the hubs

H. Jeong, R.Albert, and A.-L. Barabási. Internet: Diameter of the world-wide web. Nature, 401:130-131, 1999

The Poisson form offers a poor fit for the WWW's degree distribution



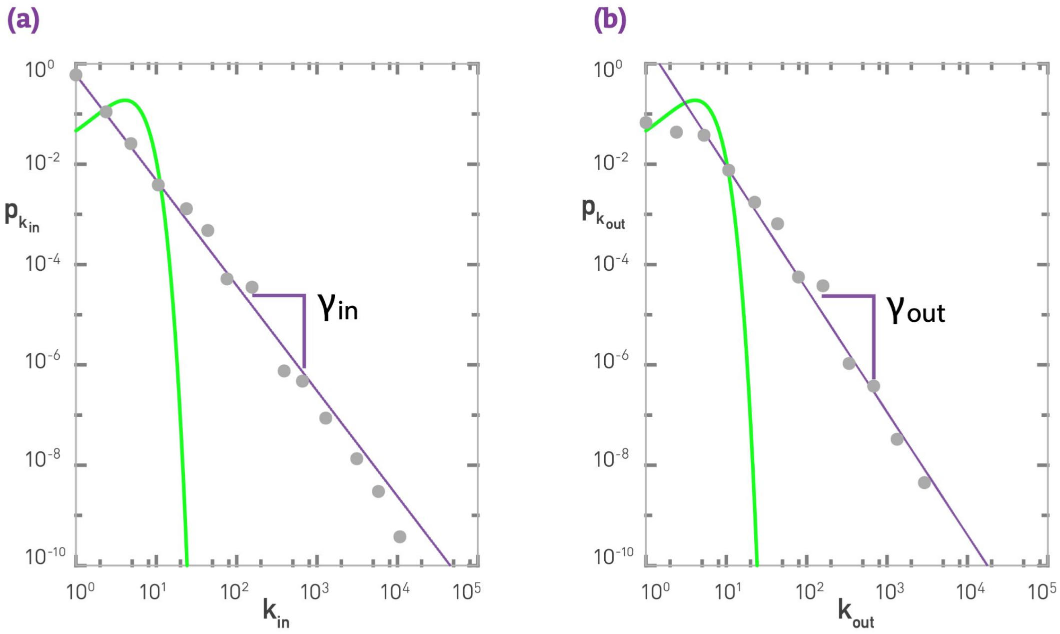
The incoming (a) and outgoing (b) degree distribution of the previous WWW sample. Note the log-log plot, in which a power law follows a straight line.

The dots correspond to the empirical data and the line corresponds to the power-law fit, with degree exponents $\gamma_{in} = 2.1$ and $\gamma_{out} = 2.45$.

The green line is the degree distribution predicted by a Poisson function with the average degree $\langle k_{in} \rangle = \langle k_{out} \rangle = 4.60$.

Instead, the degree distribution is well approximated with: $p_k \sim k^{-\gamma}$

The Poisson form offers a poor fit for the WWW's degree distribution



The incoming (a) and outgoing (b) degree distribution of the previous WWW sample. Note the log-log plot, in which a power law follows a straight line.

The dots correspond to the empirical data and the line corresponds to the power-law fit, with degree exponents γ_{in} = 2.1 and γ_{out} = 2.45.

The green line is the degree distribution predicted by a Poisson function with the average degree $\langle k_{in}\rangle=\langle k_{out}\rangle$= 4.60.

log p_k is expected to depend linearly on log k, the slope of this line being the degree exponent γ:

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

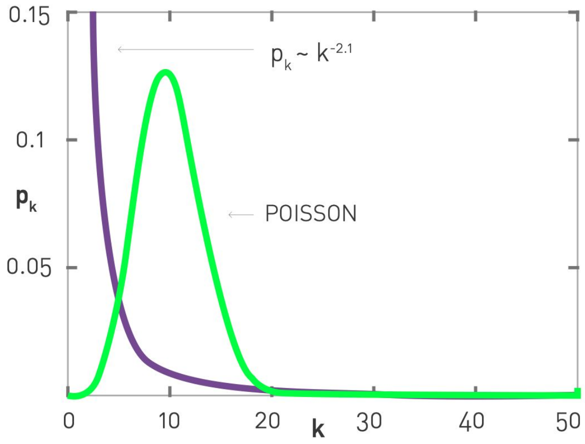
A scale-free network is a network whose degree distribution follows a power law

Poisson vs. Power-law Distributions

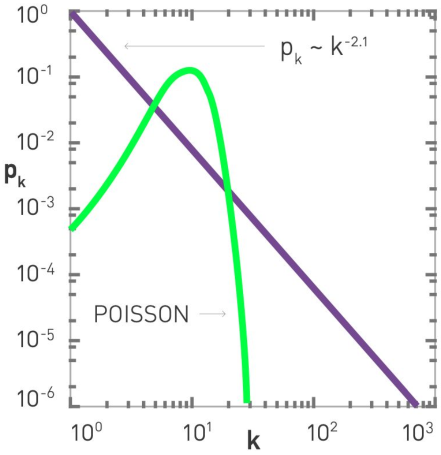
(a) Comparing a Poisson function with a power-law function (γ= 2.1) on a linear plot. Both distributions have ⟨k⟩= 11.

(b) The same curves as in (a), but shown on a log-log plot, allowing us to inspect the difference between the two functions in the high-k regime.

(a)



(b)

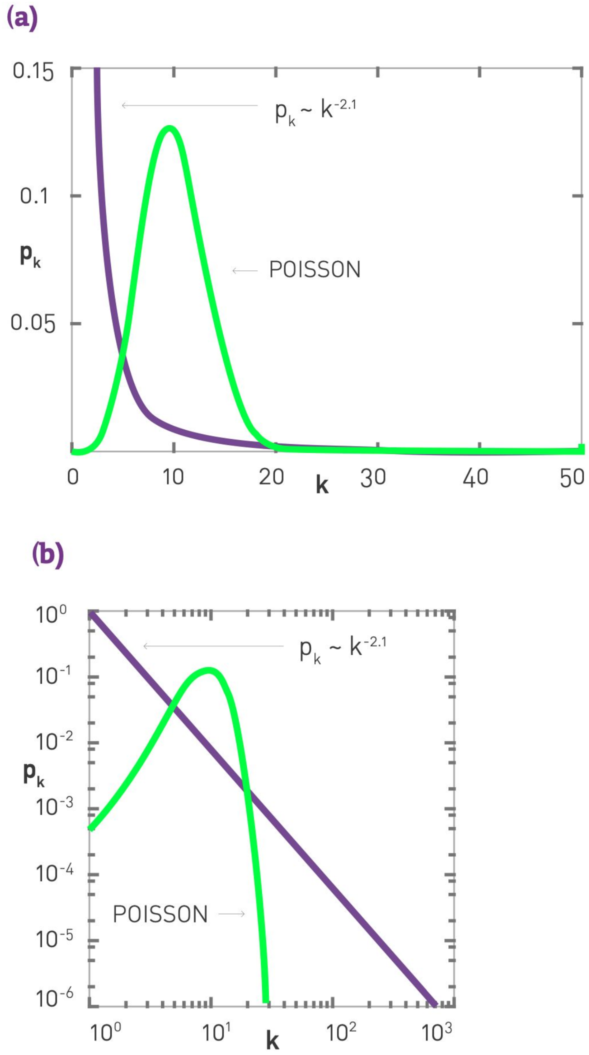


Poisson vs. Power-law Distributions

Small k: power law is above the Poisson → a scale-free network has a **large number of small degree nodes**, most of which are absent in a random network.

k around⟨k⟩: the Poisson is above the power law → in a random network there is an **excess of nodes with degree k ≈⟨k⟩**

Large k: power law is again above the Poisson → **observing a high-degree node, or hub, is orders of magnitude more likely in a scale-free network**.



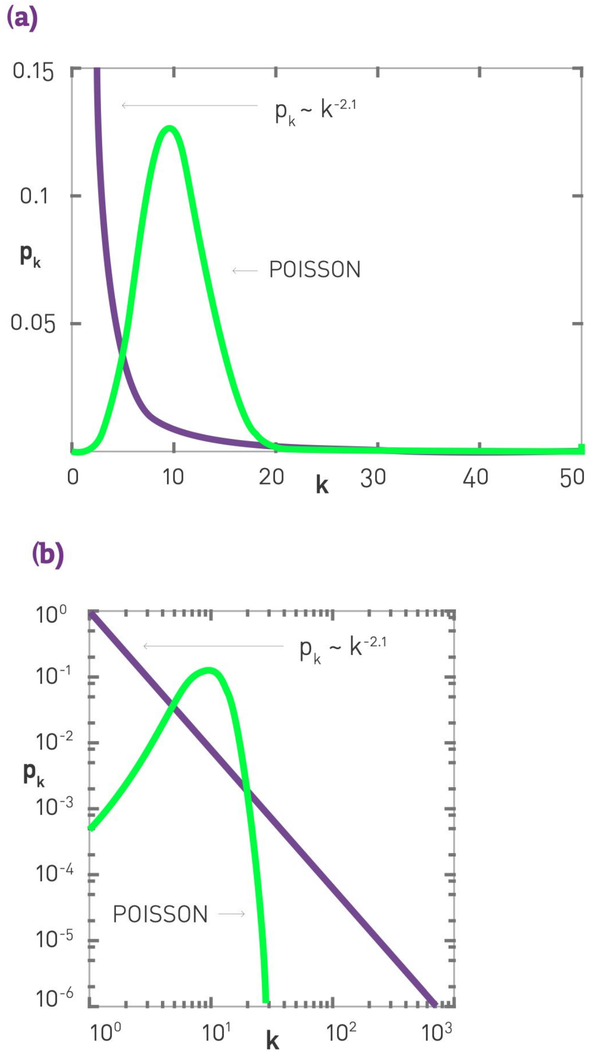
Poisson vs. Power-law Distributions

If the WWW were to be a random network with <k>=4.6 and size N≈10^{12}, we would expect

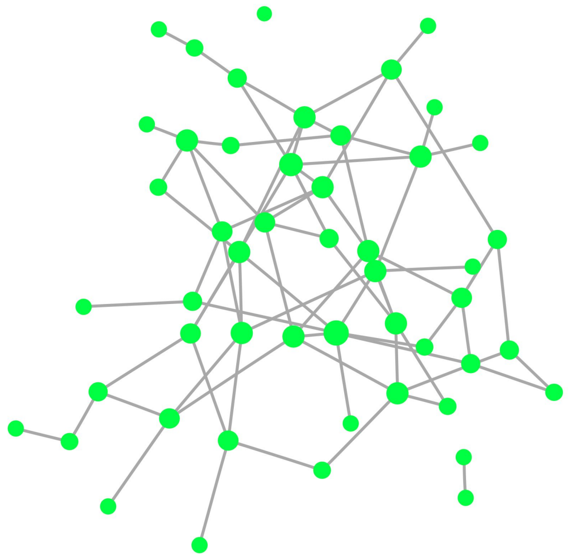
$$N_{k\geq 100} = 10^{12} \sum_{k=100}^{\infty} \frac{(4.6)^k}{k!} e^{-4.6} \simeq 10^{-82}$$

nodes with at least 100 links, or effectively none.

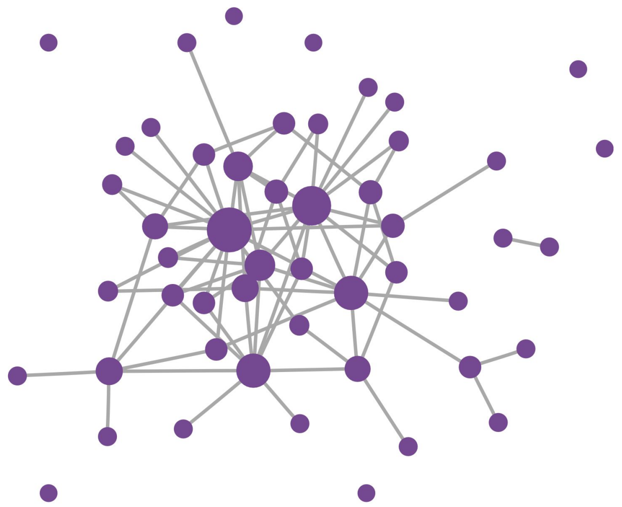
In contrast, given the WWW's power law degree distribution, with $\gamma_{in} = 2.1$ we have $N_{k\geq 100} = 4\times 10^9$, i.e. more than four billion nodes with degree k ≥100.



Poisson vs. Power-law Distributions



A random network with ⟨k⟩= 3 and N = 50, illustrating that most nodes have comparable degree k≈⟨k⟩.



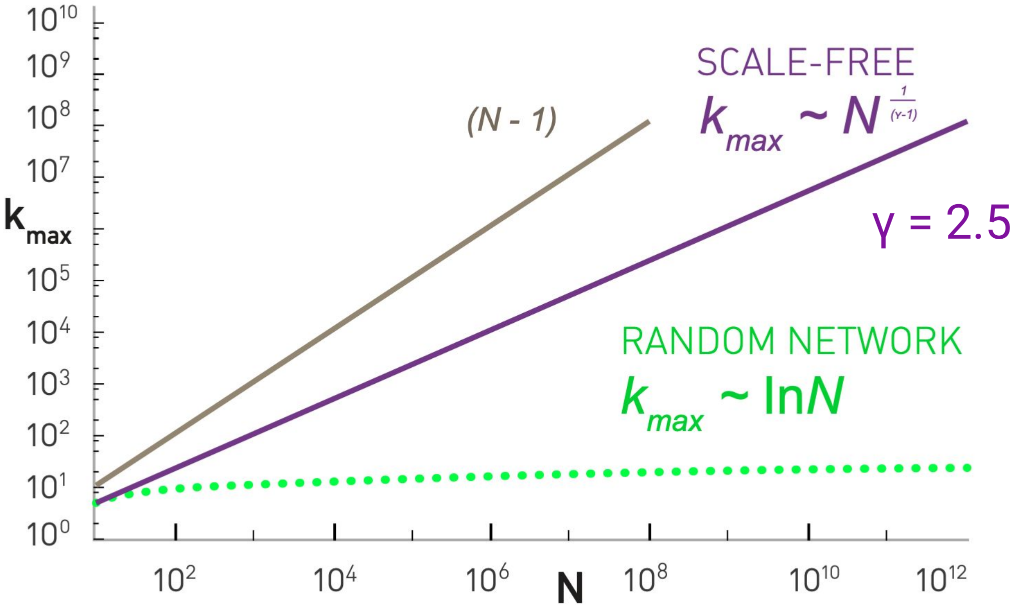
A scale-free network with γ=2.1 and ⟨k⟩= 3, illustrating that numerous small-degree nodes coexist with a few highly connected hubs.

Hubs are Large in Scale-free Networks

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

For comparison, for the linear behavior, k_{max} ~ N − 1.

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⟩.



(Barabasi Ch. 4.3)

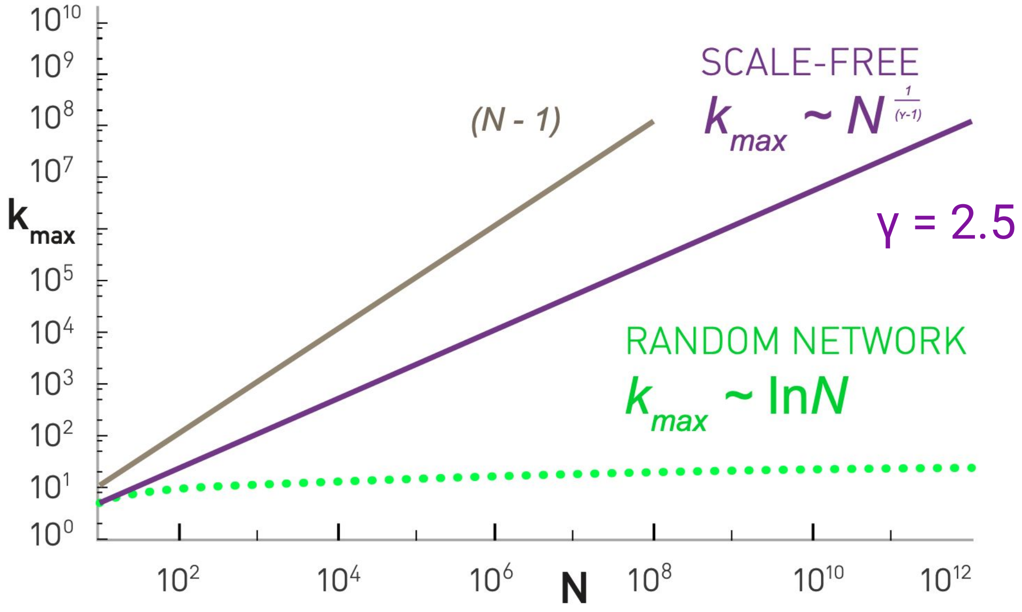
Hubs are Large in Scale-free Networks

In the WWW sample ($N \approx$ 300k nodes):

If the degree distribution were to follow an exponential, $k_{max} \approx 14$ for $\lambda=1$.

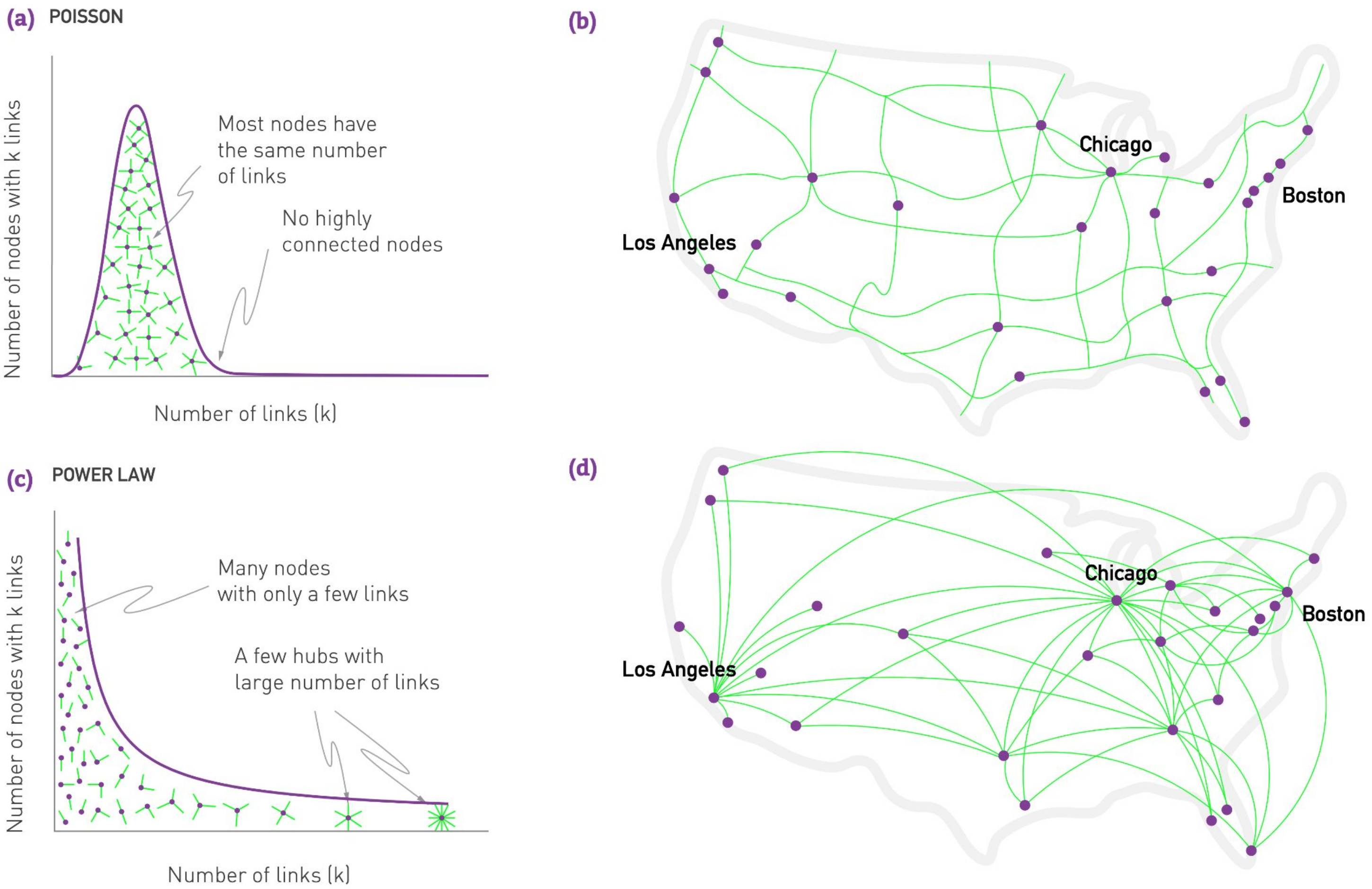
If scale-free with $\gamma = 2.1$, $k_{max} \approx 95{,}000$.

Real (observed) $k_{max} = 10{,}721$, which is comparable to k_{max} predicted by a scale-free network.



(Barabasi Ch. 4.3)

Summary: Random vs. Scale-free Networks



Summary

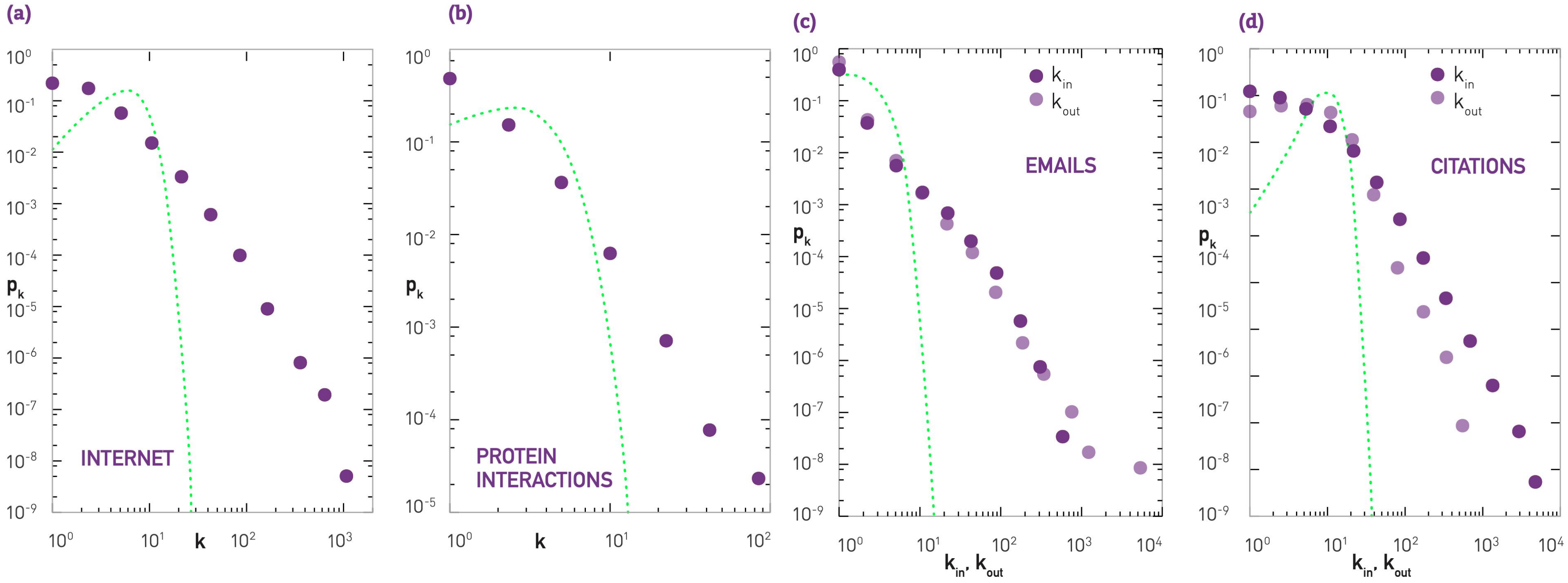
In a random network hubs are effectively forbidden, while in scale-free networks they are naturally present.

The more nodes a scale-free network has, the larger are its hubs. Indeed, the size of the hubs grows polynomially with network size.

In contrast, in a random network the size of the largest node grows logarithmically or slower with N, implying that hubs will be tiny even in a very large random network.

Universality

Many Real Networks are Scale-free

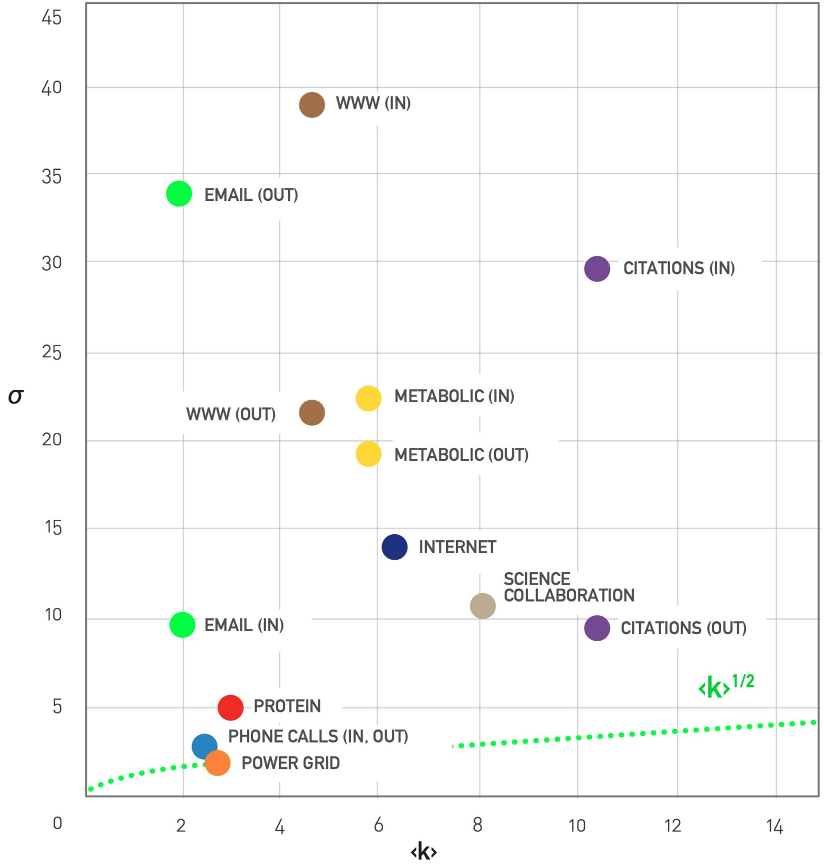


The green dotted line shows the Poisson distribution with the same $\langle k \rangle$ as the real network, illustrating that the random network model cannot account for the observed p_k.

Many Real Networks are Scale-free

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	93,439	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*

Standard Deviation is Large in Real Networks



For a random network the standard deviation follows $\sigma = \langle k \rangle^{1/2}$ shown as a green dashed line.

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.

(The actor network has a very large ⟨k⟩ and σ, and it is omitted)

The Ultra Small-World Property

Do hubs affect the small world property?

Intuitively “yes”: Airlines build hubs precisely to decrease the number of hops between two airports.

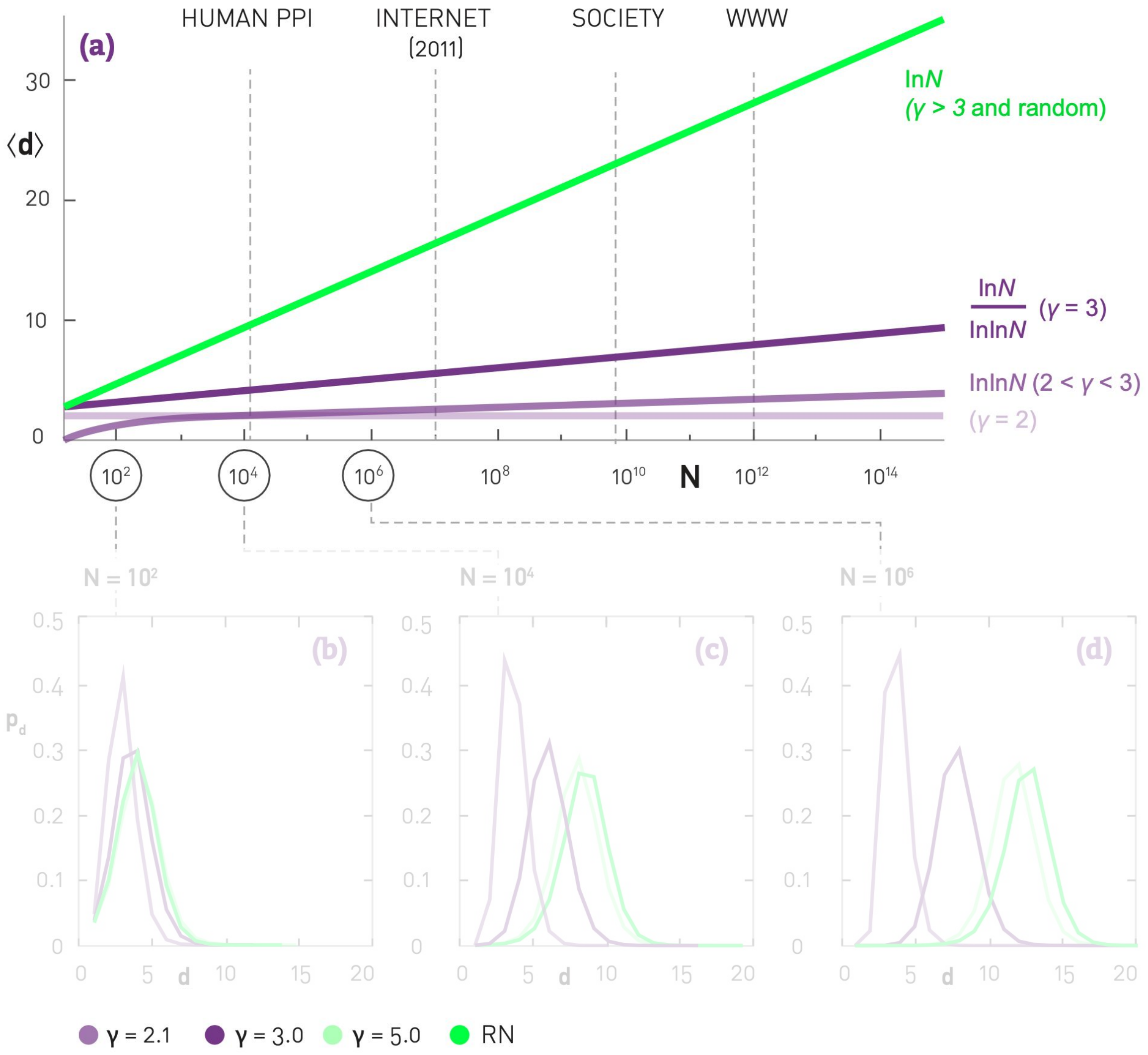
Do hubs affect the small world property?

Intuitively “yes”: Airlines build hubs precisely to decrease the number of hops between two airports.

Indeed, distances in a scale-free network are **smaller** than the distances observed in an equivalent random network.

The dependence of the average distance ⟨d⟩ on the system size N and the degree exponent γ are captured by:

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

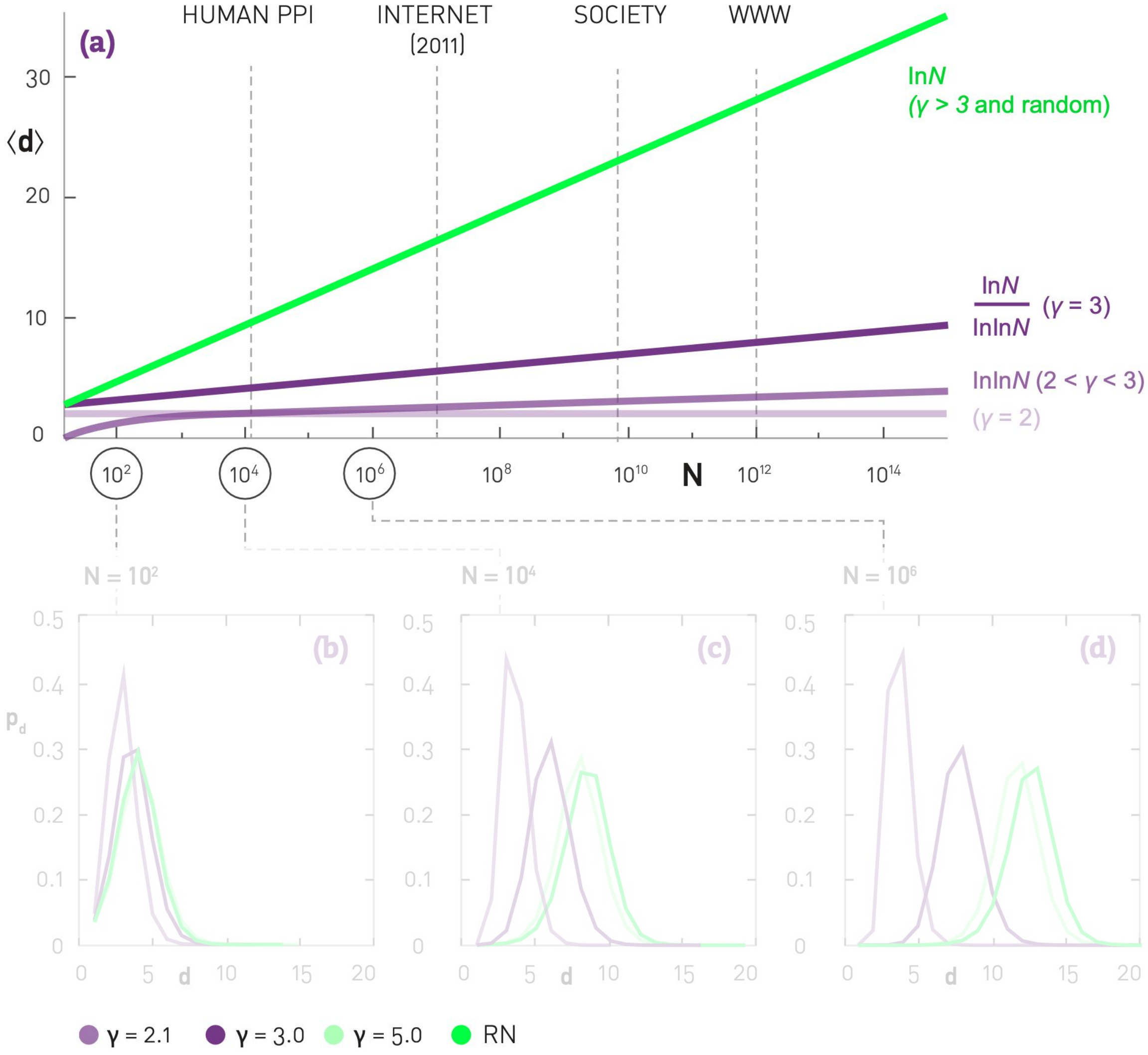


Anomalous Regime (γ = 2)

The degree of the biggest hub grows linearly with the system size, i.e. $k_{max} \sim N$.

This forces the network into a hub & spoke configuration: all nodes are close to each other because they all connect to the same central hub.

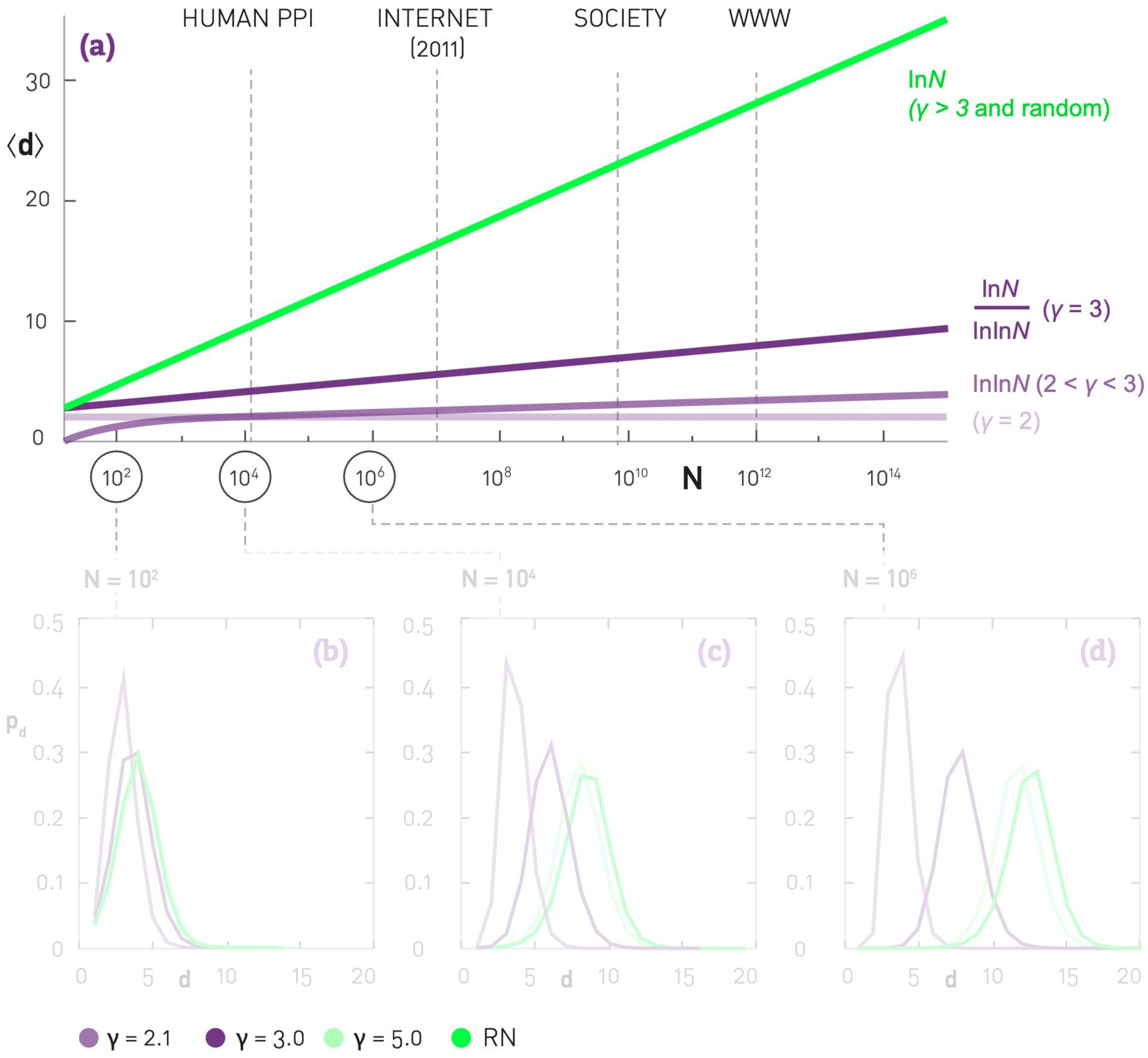
The average path length does not depend on N.



Ultra-Small World ($2 < \gamma < 3$)

The average distance increases as ln lnN, a significantly slower growth than the lnN derived for random networks.

“Ultra small”: The hubs radically reduce the path length by linking to many small-degree nodes, creating short distances between them.

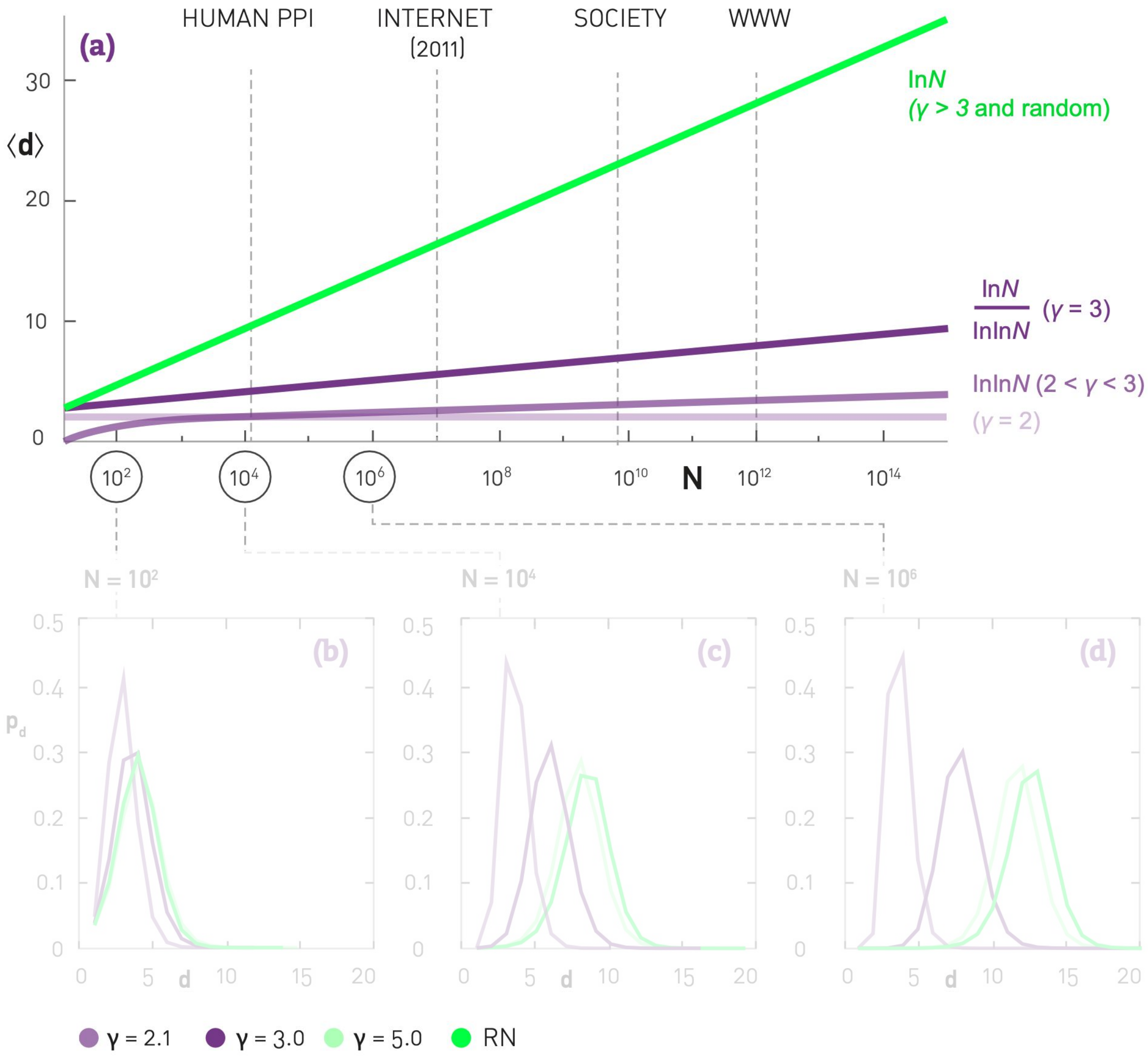


Critical Point (γ = 3)

(When the second moment of the degree distribution does not diverge any longer)

The lnN dependence encountered for random networks returns.

Yet, the calculations indicate the presence of a double log correction *ln lnN*, which shrinks the distances compared to a random network of similar size.

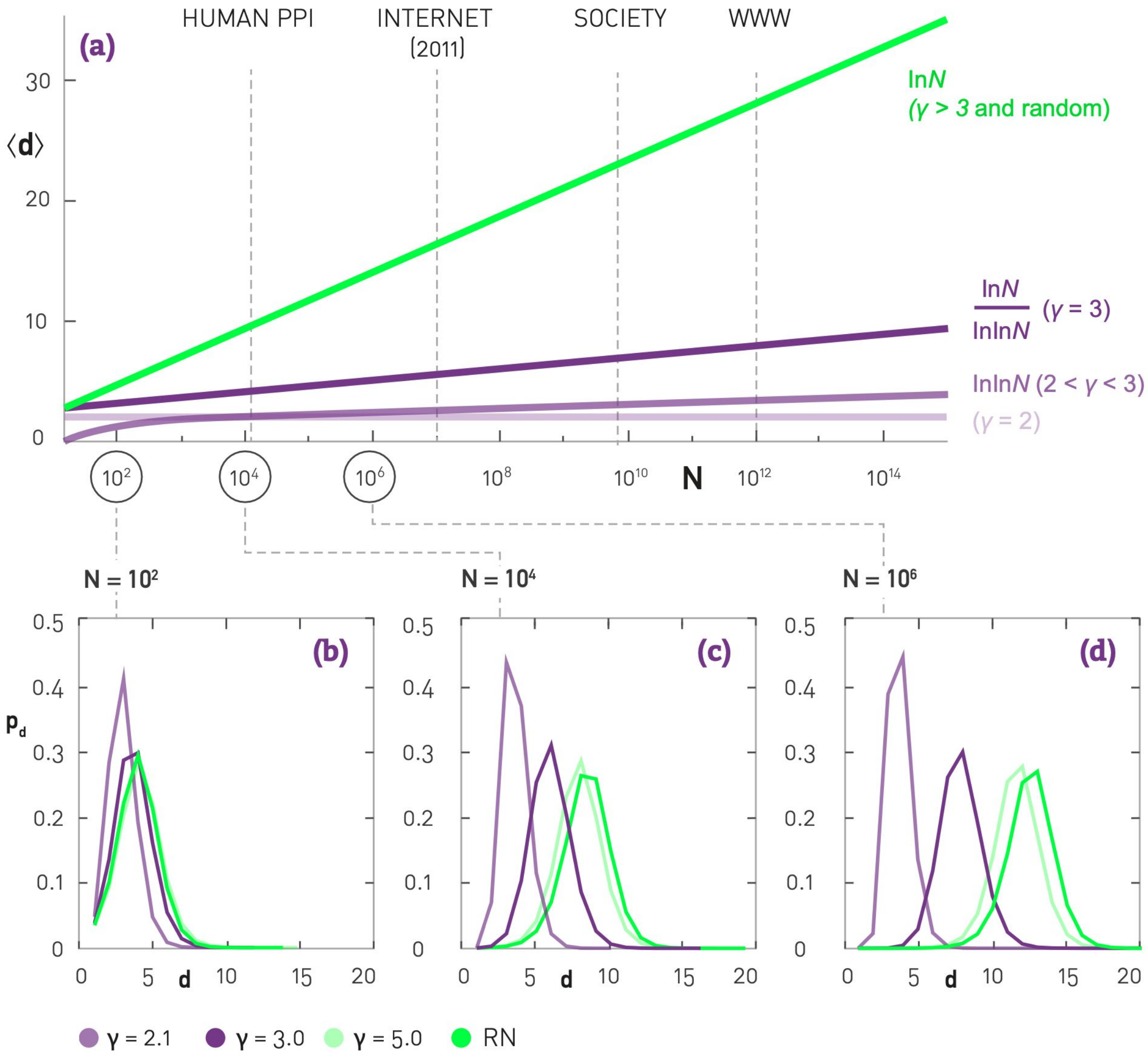


Small World (γ > 3)

(When the second moment of the degree distribution is finite **does not diverge** any longer)

The average distance follows the small world result derived for random networks.

While hubs continue to be present, for γ > 3 they are not sufficiently large and numerous to have a significant impact on the distance between the nodes.



Now let's look at the path length distribution ((b)-(d)) for scale-free networks with different γ and N.

While for small networks ($N = 10^2$) the distance distributions are not too sensitive to γ, for large networks ($N = 10^6$) p_d and ⟨d⟩ change visibly with γ.

The larger the degree exponent γ, the larger are the distances between the nodes.

Summary

The scale-free property has several effects on network distances:

- Shrinks the average path lengths. Therefore most scale-free networks of practical interest are not only “small”, but are “ultra-small”. This is a consequence of the hubs, that act as bridges between many small degree nodes.
- Changes the dependence of $\langle d \rangle$ on the system size. The smaller is γ, the shorter are the distances between the nodes.
- Only for $\gamma > 3$ we recover the $\ln N$ dependence, the signature of the small-world property characterizing random networks.

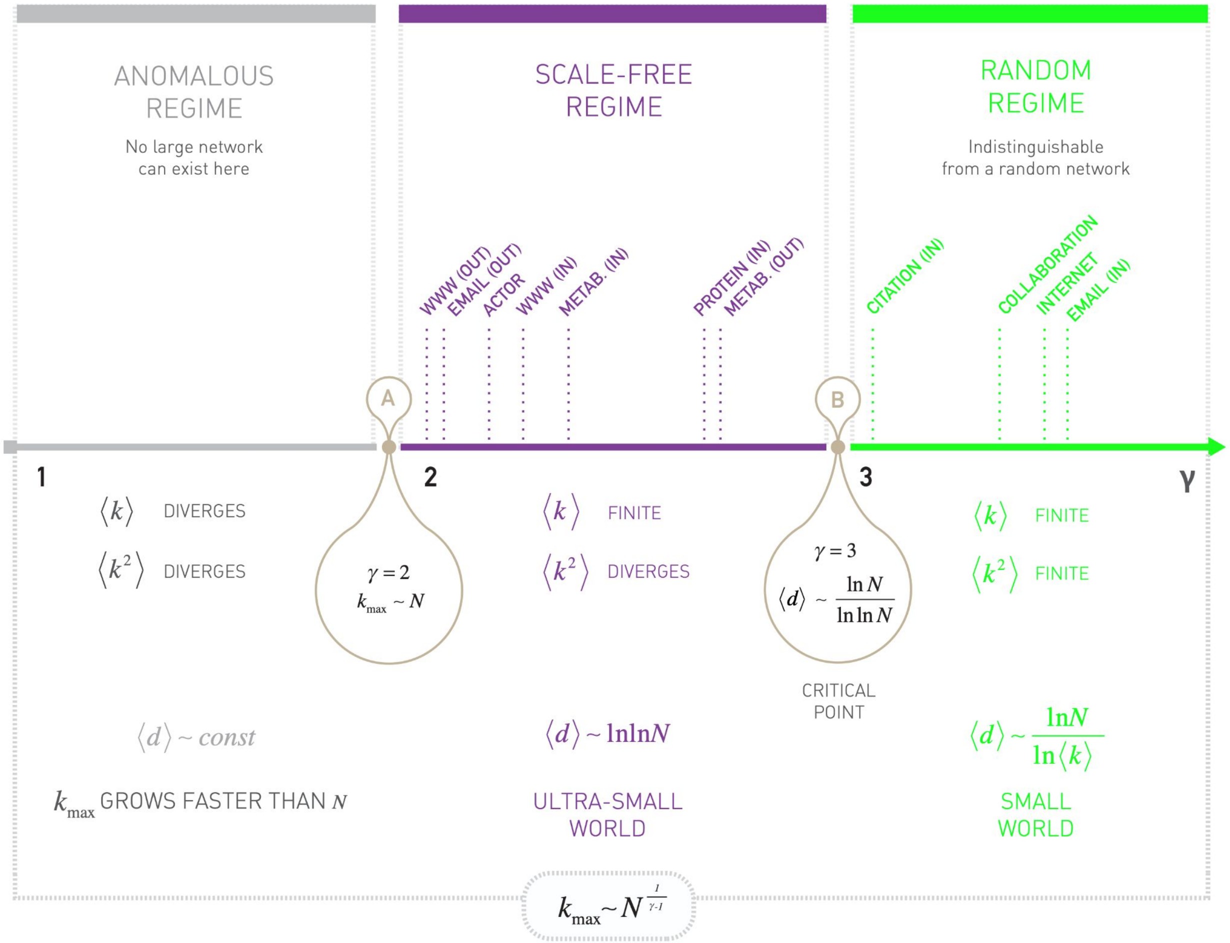
The Role of the Degree Exponent

Many properties of a scale-free network depend on the value of γ

Anomalous Regime (γ≤ 2)

For γ< 2 the exponent 1/(γ− 1) is larger than one, hence the number of links connected to the largest hub grows faster than the size of the network.

For sufficiently large N the degree of the largest hub must exceed the total number of nodes in the network, hence it will run out of nodes to connect to.

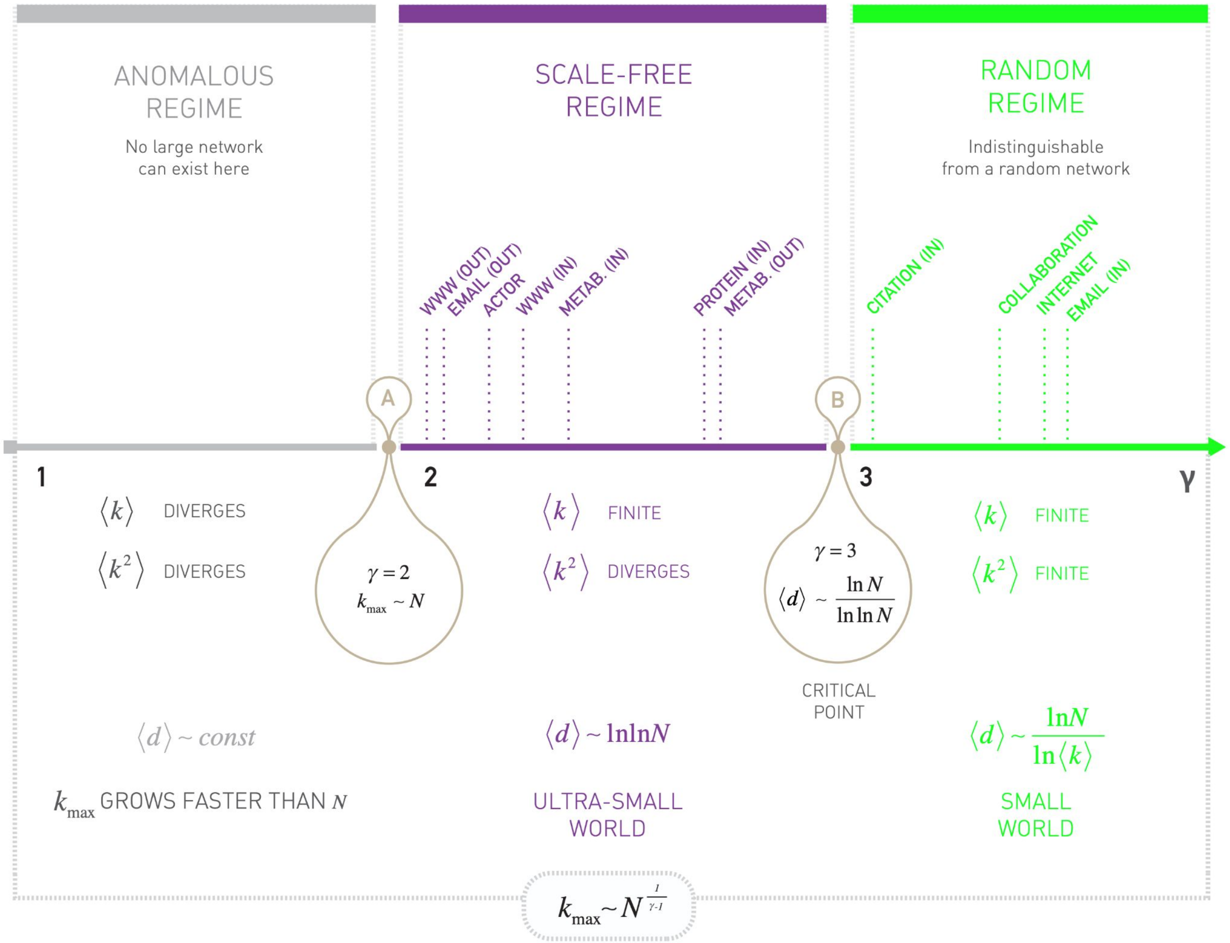


Many properties of a scale-free network depend on the value of γ

Anomalous Regime (γ≤ 2)

Many other anomalous features of scale-free networks in this regime.

→ Large scale-free network with γ < 2, that lack multi-links, cannot exist.

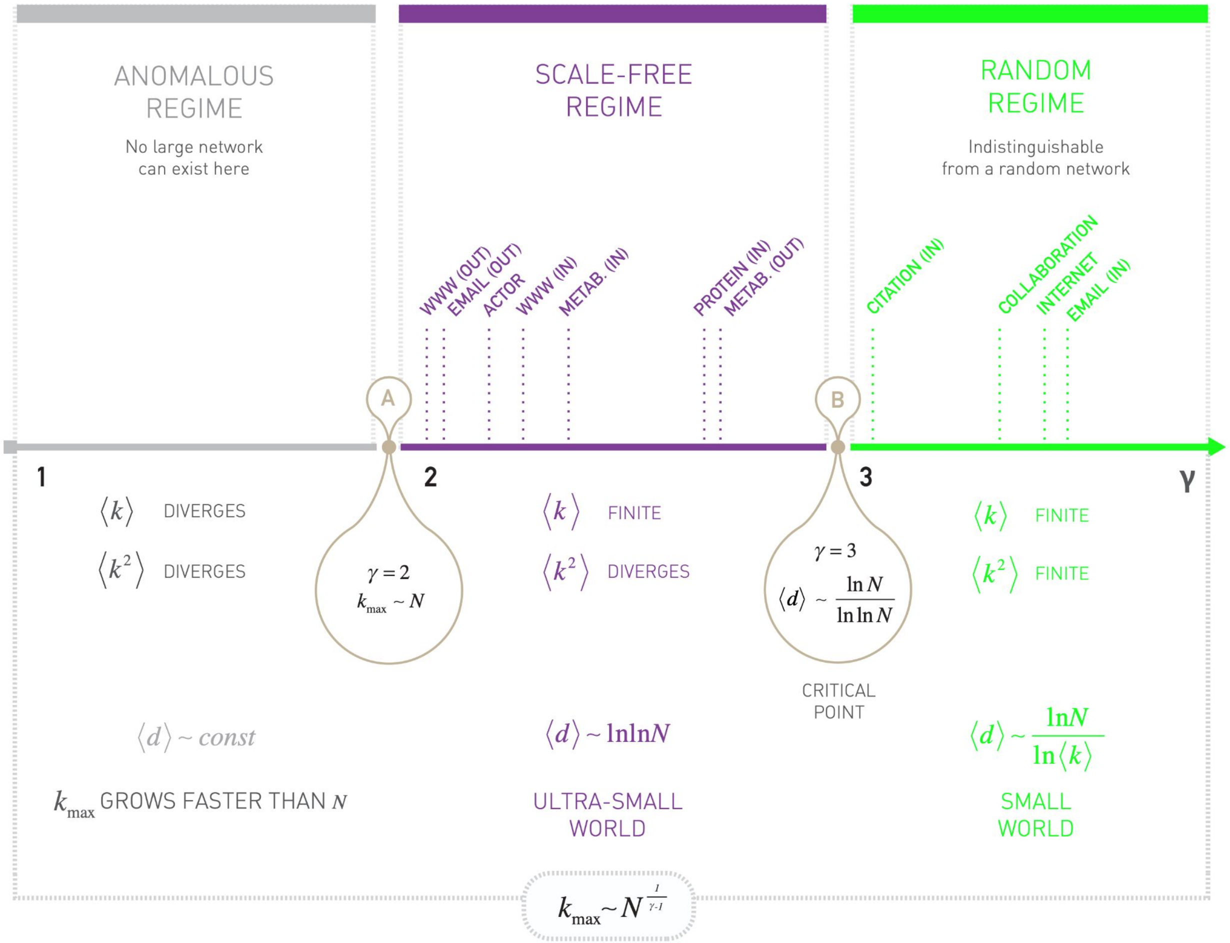


Many properties of a scale-free network depend on the value of γ

Scale-Free Regime (2 < γ < 3)

Scale- free networks in this regime are ultra-small: k_{max} grows with the size of the network with exponent 1/ (γ - 1) which is smaller than one.

Hence the market share of the largest hub, k_{max} /N, decreases as k_{max} /N ~ $N^{-(\gamma-2)/(\gamma-1)}$.

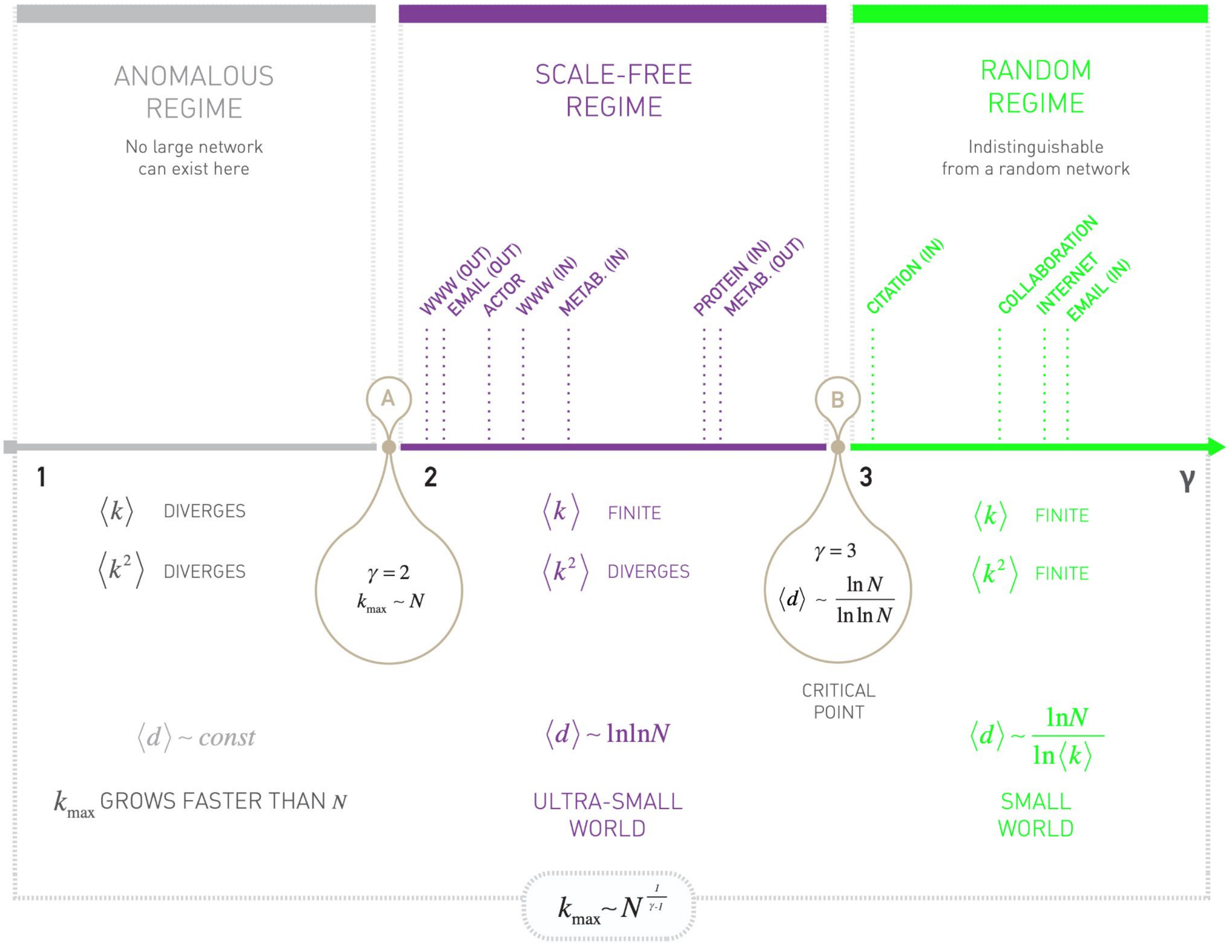


Many properties of a scale-free network depend on the value of γ

Random Regime (γ > 3)

For all practical purposes the properties of a scale-free network in this regime are difficult to distinguish from the properties a random network of similar size.

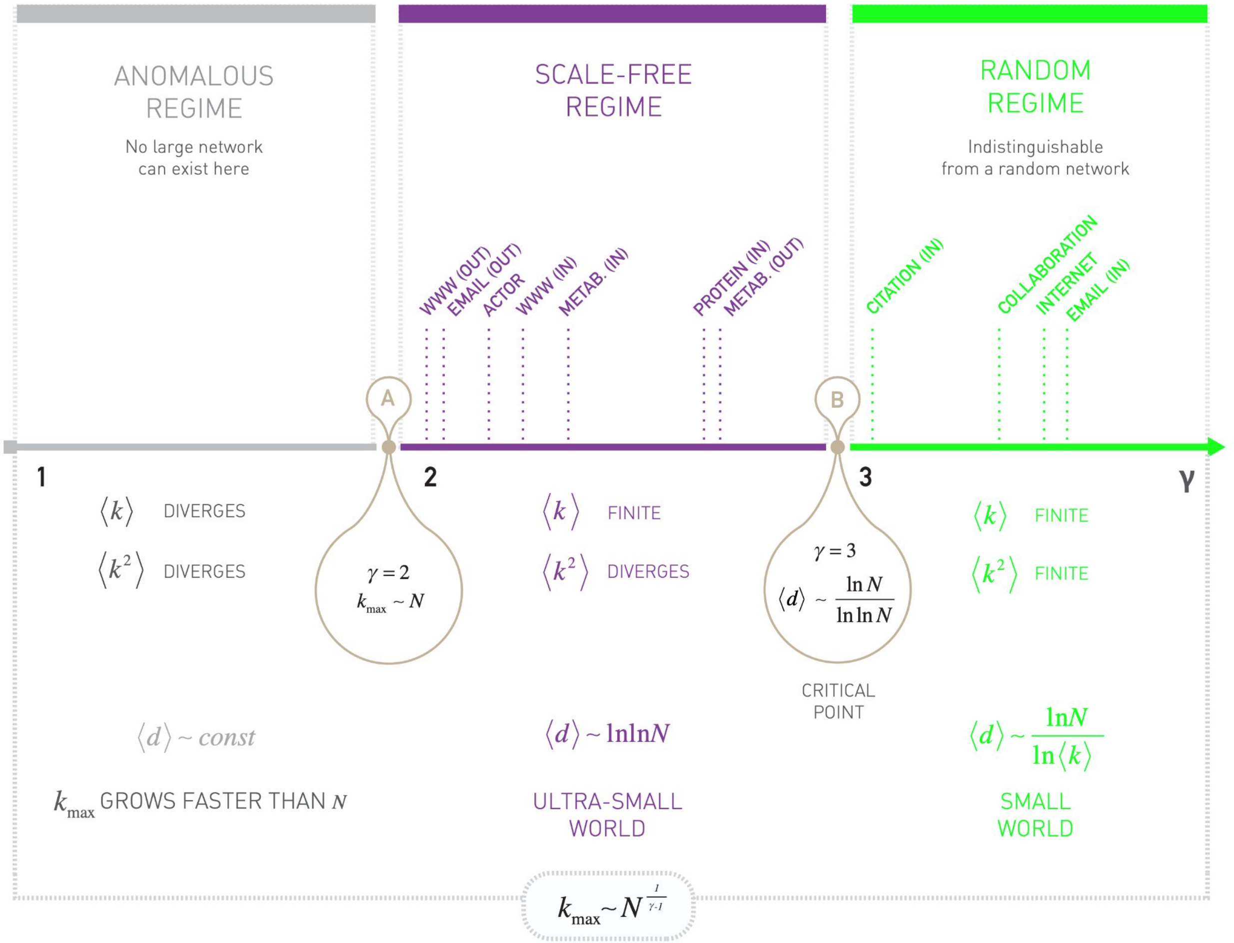
E.g., the average distance between the nodes converges to the small-world formula derived for random networks.



Many properties of a scale-free network depend on the value of γ

Random Regime (γ > 3)

The reason is that for large γ the degree distribution p_k decays sufficiently fast to make the hubs small and less numerous.



Scale-free networks with large γ are hard to distinguish from a random network.

For a scale-free network, the natural cutoff is:

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

By inverting the formula, we can estimate the network size necessary to observe the desired scaling regime:

$$N = \left(\frac{k_{max}}{k_{min}} \right)^{\gamma - 1}$$

Scale-free networks with large γ are hard to distinguish from a random network.

To document the presence of a power-law degree distribution we ideally need 2-3 orders of magnitude of scaling, which means that k_{max} should be at least 10^2 - 10^3 times larger than k_{min}.

For example, to document the scale-free nature of a network with $\gamma = 5$, requiring scaling that spans at least two orders of magnitudes (e.g. $k_{min} \sim 1$ and $k_{max} \simeq 10^2$), the size of the network must exceed $N > 10^8$!

There are very few network maps of this size. Therefore, there may be many networks with large degree exponent. Given, however, their limited size, it is difficult to obtain convincing evidence of their scale-free nature.

Summary

The scale-free property has played an important role in the development of network science for two main reasons:

Many networks of scientific and practical interest, from the WWW to the subcellular networks, are scale-free.

Once the hubs are present, they fundamentally change the system's behavior. More on this later.