The Scale-Free property

Alberto Paccanaro
EMAp – FGV

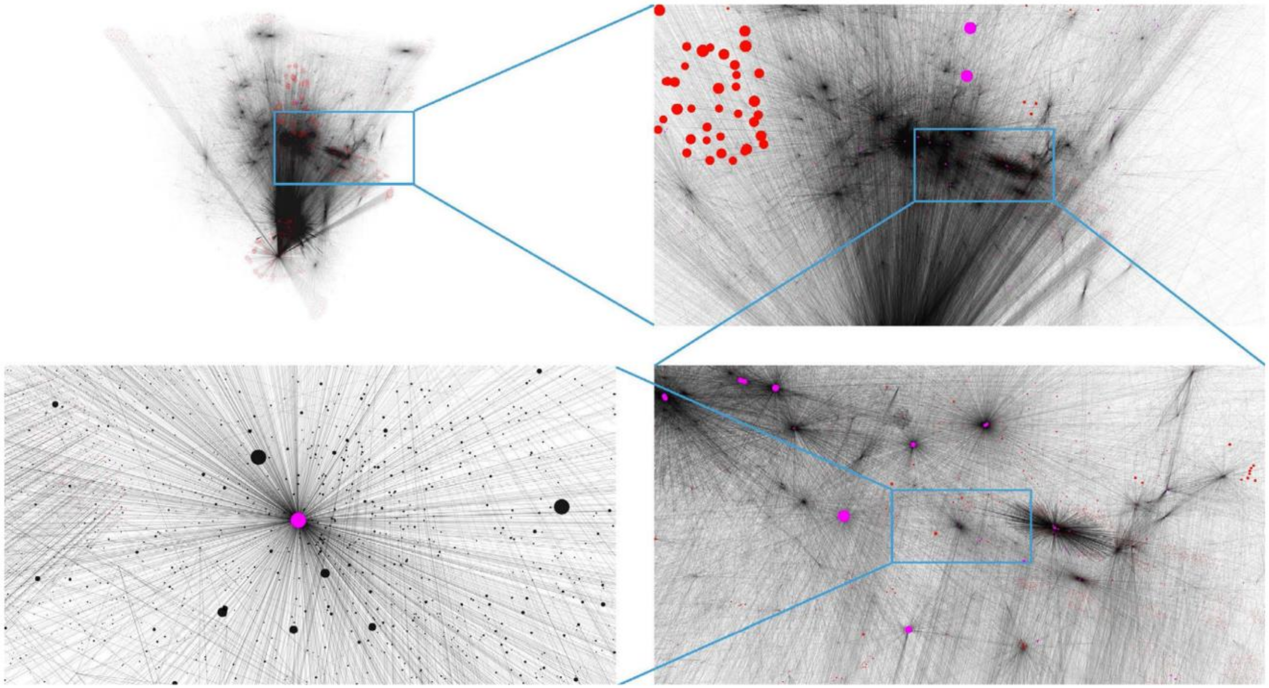
www.paccanarolab.org

Some material and images are from (or adapted from):
A. Barabási, and M. Pósfai. Network science, Cambridge University Press, 2016

Hubs in the WWW

In 1998 we believed that the WWW could be well approximated by a random network.

more than 50 links
more than 500 links



In a random network highly connected nodes, or hubs, are effectively forbidden.

Hubs represent a signature of *scale-free property*

A.PACCANARO – FGV

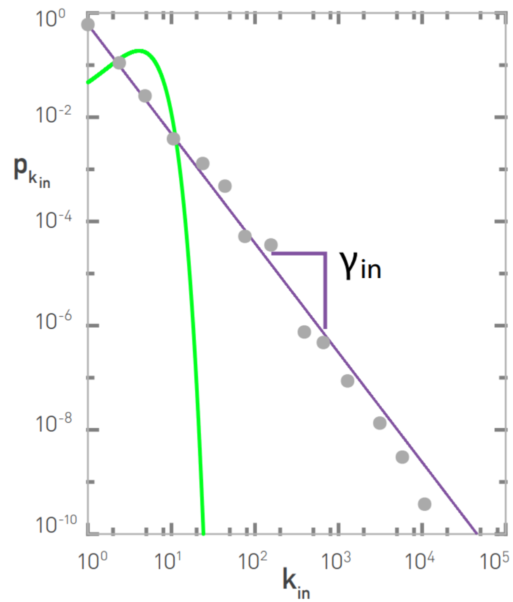
Power laws

For the degrees WWW documents, on a log-log scale the data points form an approximate straight line.

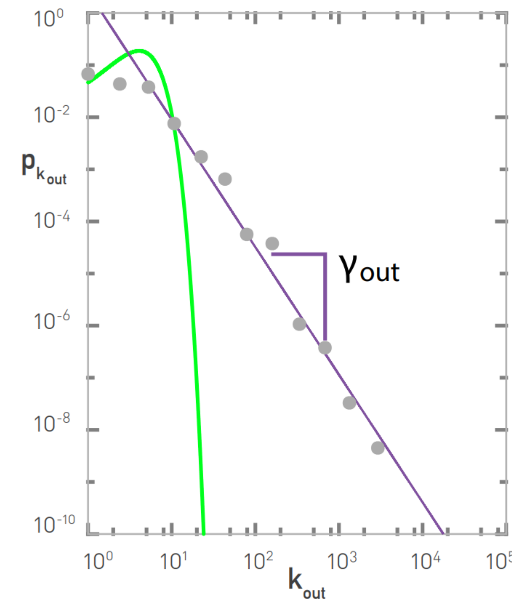
$$p_k \sim k^{-\gamma}$$

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

power law distribution



$$\gamma_{in} \approx 2.1$$



$$\gamma_{out} \approx 2.45$$

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

Discrete Formalism

Probability that a node has k links:

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

We know that: $\sum_{k=1}^{\infty} p_k = 1$

Hence: $C\sum_{k=1}^{\infty} k^{-\gamma} = 1$

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

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

where $\zeta(\gamma)$ is the Riemann-zeta function

Continuous Formalism (degrees can have any positive real value)

Probability that a node has k links:

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

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

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

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

where $k_{\min}$ is the smallest degree for which the power law holds

A change in perspective…

Before we had a model

From the model we came up with distribution of node degrees, etc… Some conclusion were unexpected.

Now we don't start form a model.

We go to real word networks.

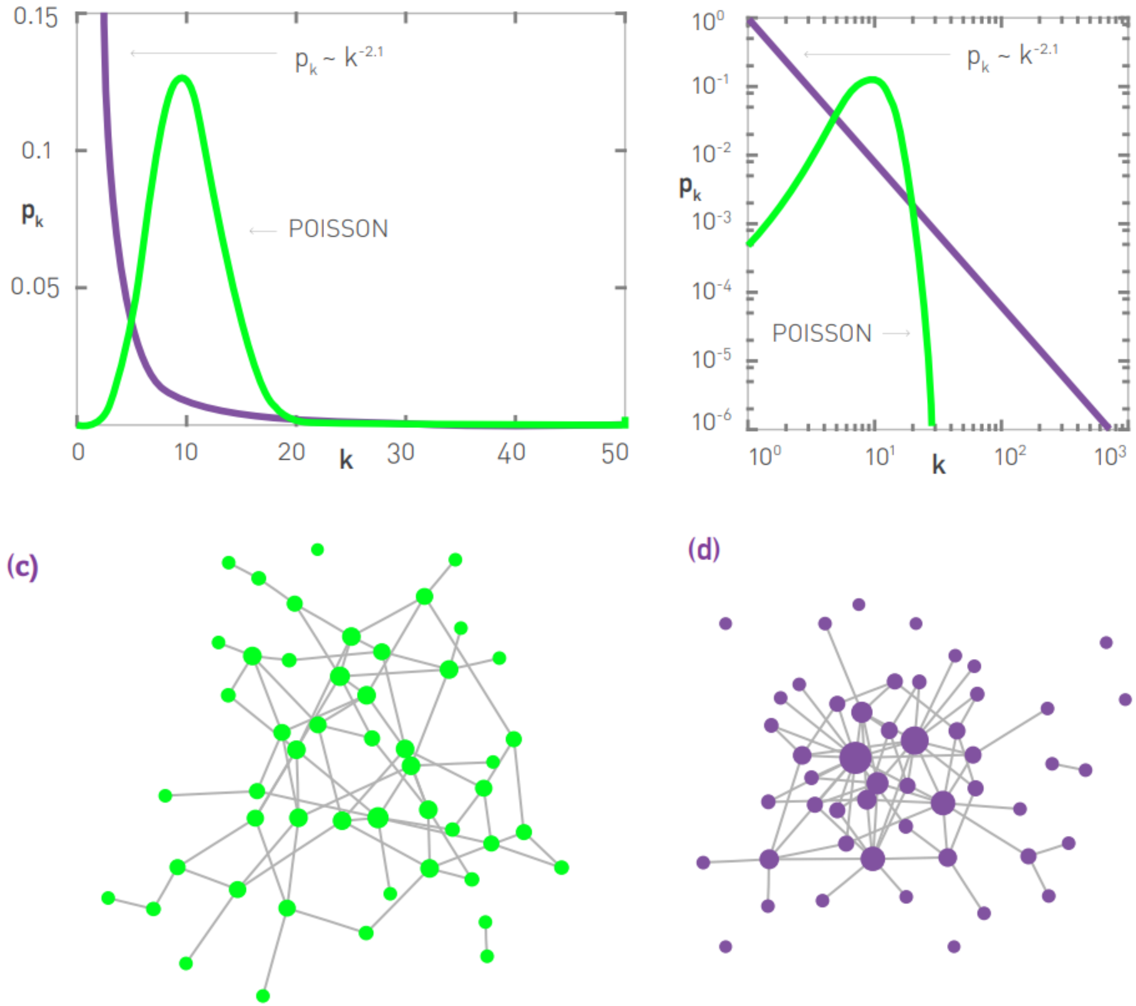
We look at their degree distribution (because we know it is important)

And then we say: *If the degree distribution of a network follows a power law, what can I predict about the properties of the network?* (that I don't know yet how to build)

Data comes first. Model comes last.

Hubs

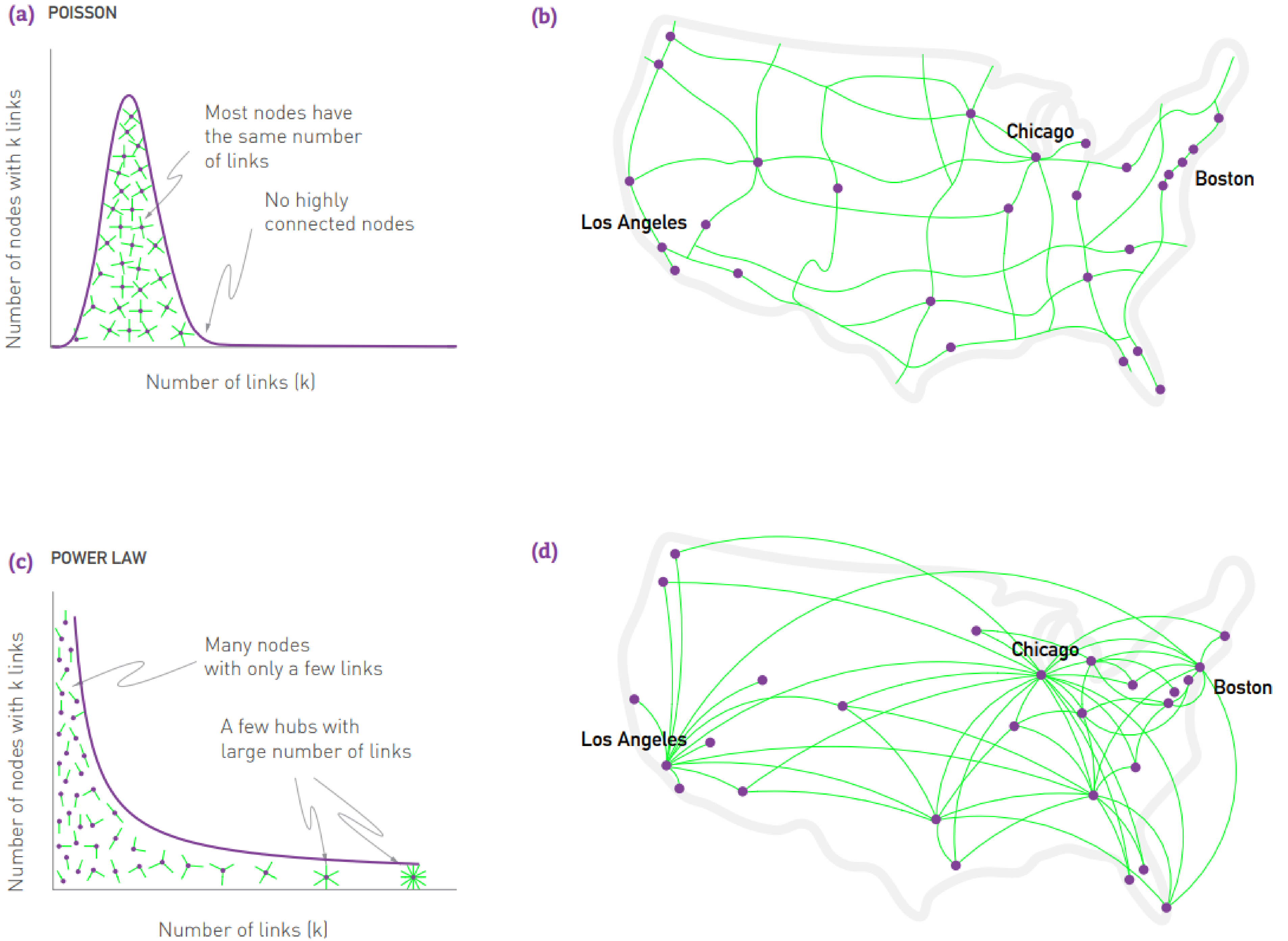
The main difference between a random and a scale-free network comes in the tail of the degree distribution.



Let us use the WWW to illustrate the magnitude of these differences. The probability to have a node with k=100 is about $p_{100} \approx 10^{-94}$ in a Poisson distribution while it is about $p_{100} \approx 4x10^{-4}$ if p_k follows a power law. Consequently, if the WWW were to be a random network with $<k>=4.6$ and size $N \approx 10^{12}$, we would expect

$$N_{k \geq 100} = 10^{12} \sum_{k=100}^{\infty} \frac{(4.6)^k}{k!} e^{-4.6} \simeq 10^{-82} \qquad (4.14)$$

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 $N_{k \geq 100} = 4x10^9$, i.e. more than four billion nodes with degree $k \geq 100$.

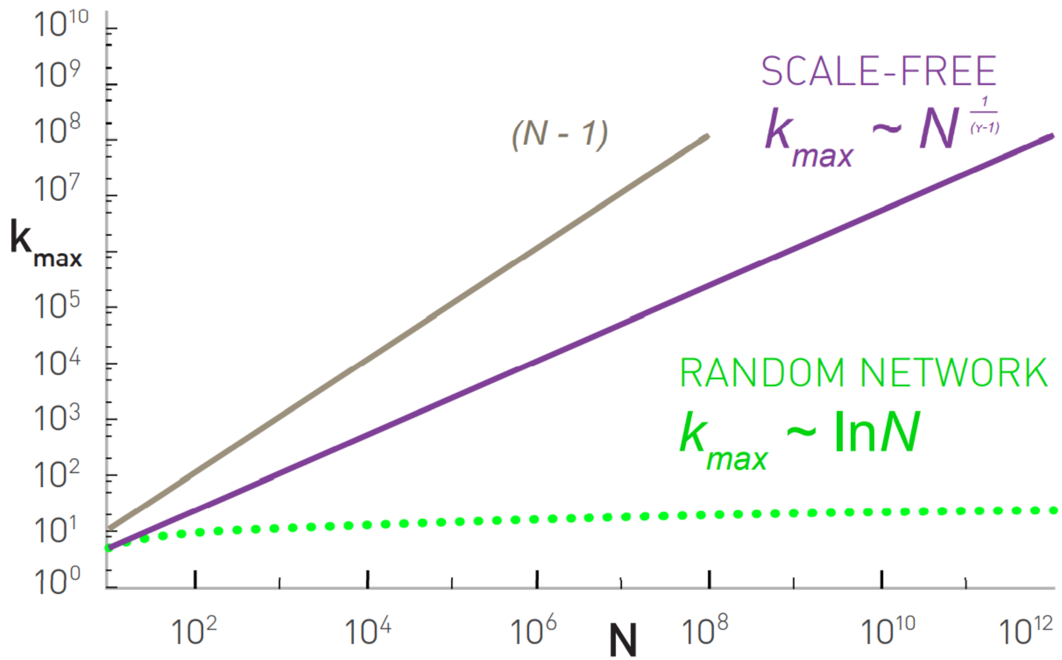

(a)
POISSON
Most nodes have the same number of links
No highly connected nodes
Number of nodes with k links
Number of links (k)
(b)
Chicago
Boston
Los Angeles
(c)
POWER LAW
Many nodes with only a few links
A few hubs with large number of links
Number of nodes with k links
Number of links (k)
(d)
Chicago
Boston
Los Angeles

How does the network size affect the hub size?

Natural cutoff: the expected size of the largest hub in a network

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

hubs in a scale-free network are several orders of magnitude larger than the biggest node in a random network



Conclusion: key difference between a random and a scale-free network

Shape of the Poisson vs the power-law function

In a random network:

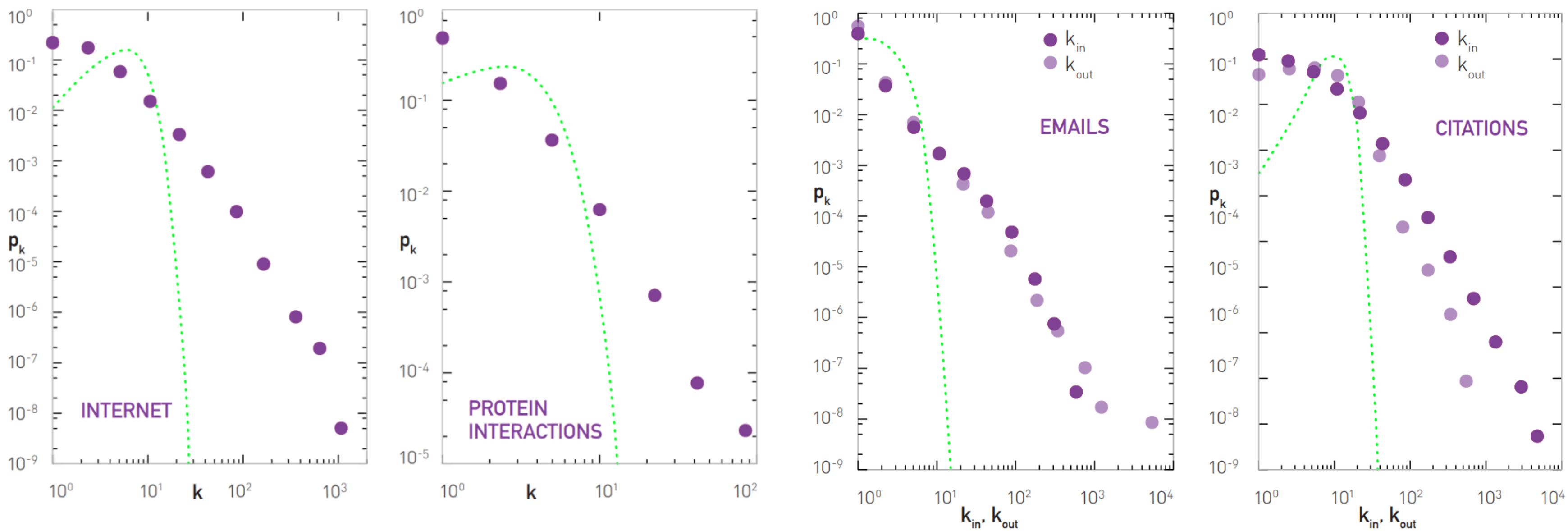
- most nodes have comparable degrees
- hubs are essentially forbidden (size of the largest node grows logarithmically with N)

Scale-free networks:

- hubs are expected
- larger N implies larger hubs (size of the hubs grows polynomially with N)

A closer look to real world networks

Many real networks were found to display the scale-free property.



Repository of network data:
https://snap.stanford.edu/data/index.html

The meaning of Scale Free

Let us look at the degree of a randomly chosen node:

$\langle k \rangle$ average degree

$\sigma_k^2 = \langle k^2 \rangle - \langle k \rangle^2$ variance (spread of the degree)

For many scale-free networks, the degree exponent γ is between 2 and 3 and for $N \to \infty$

$\langle k \rangle$ is finite

$\langle k^2 \rangle \to \infty$

What range do we expect for the degree of a randomly chosen node? $k = \langle k \rangle \pm \sigma_k$

Random Networks Have a Scale

Poisson degree distribution $\sigma_k = \langle k \rangle^{1/2} < \langle k \rangle$.

Degrees in the range $k = \langle k \rangle \pm \langle k \rangle^{1/2}$

Nodes have comparable degrees.

The average degree $\langle k \rangle$ serves as the "scale" of the network

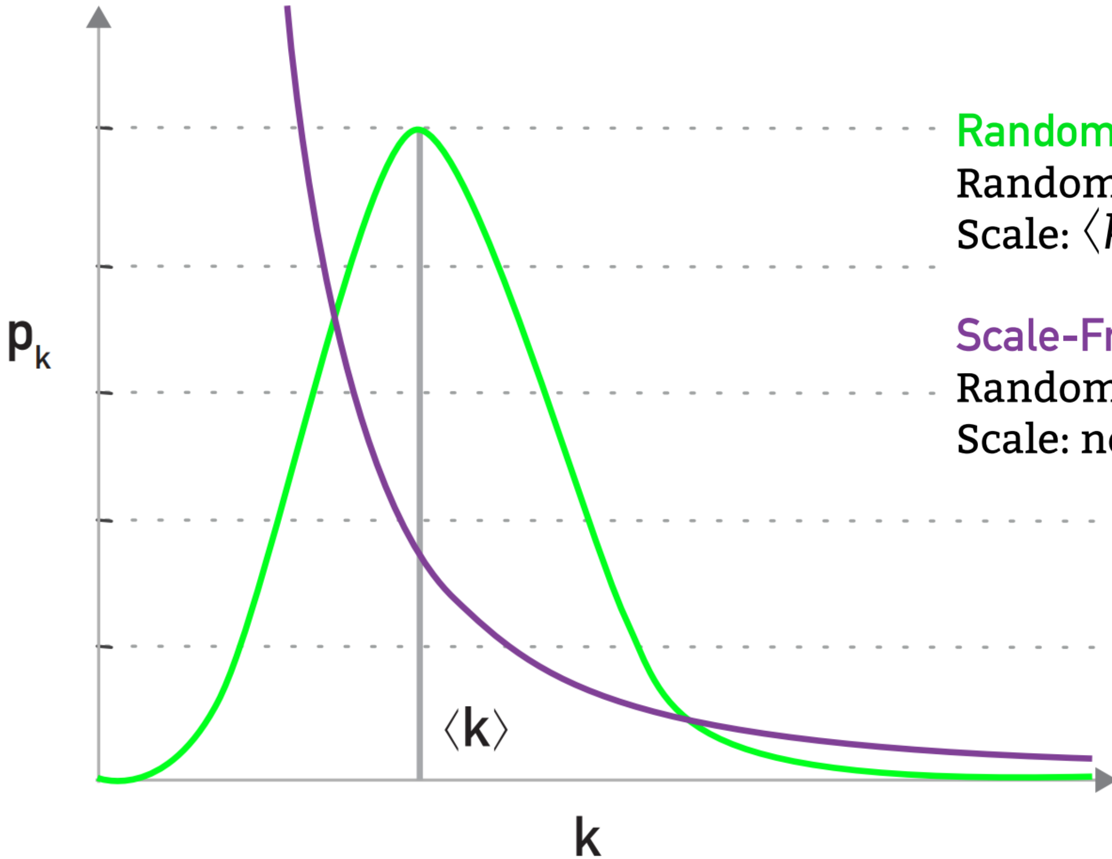
Scale-free Networks Lack a Scale

Power-law degree distribution (with $\gamma < 3$), $\sigma_k \rightarrow \infty$

Degrees in the range $k = \langle k \rangle \pm \infty$

Fluctuations around the average can be arbitrary large

Degree of a random node could be tiny or arbitrarily large, hence network does not have a meaningful internal scale, but are "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 **scale-free** name captures the *lack of an internal scale*, a consequence of the fact that nodes with widely different degrees coexist in the same network.

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*

How do we know if a network is scale-free?

We need tools to fit the p_k distribution and to estimate γ.

Practical issues:

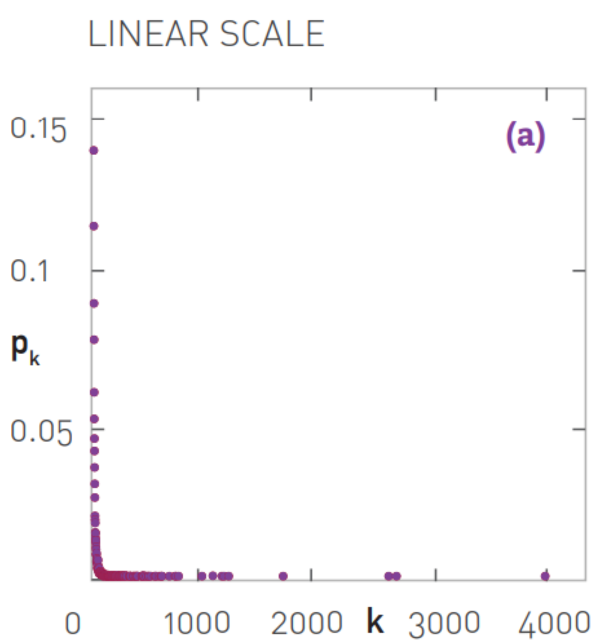
1. Plotting the Degree Distribution
2. Measuring the Degree Exponent

1. Plotting the Degree Distribution

To start:

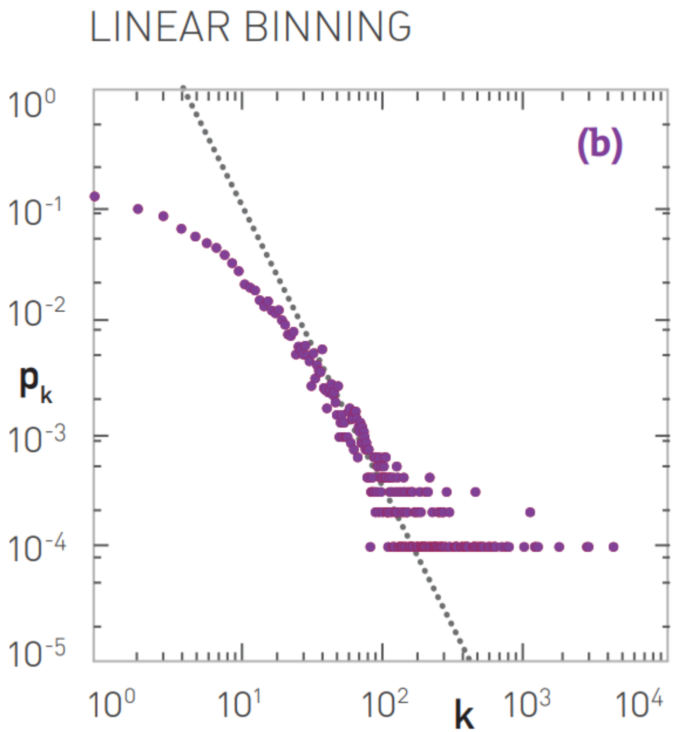
- Get N_k , the number of nodes with degree k.
- Calculate $p_k = N_k / N$

a) Log-log plot



A linear k-axis compresses the numerous small degree nodes in the small-k region, rendering them invisible. Similarly, for p_k.

Plot $p_k = N_k / N$ on a log-log plot
Linear binning (bin has size Δk = 1)



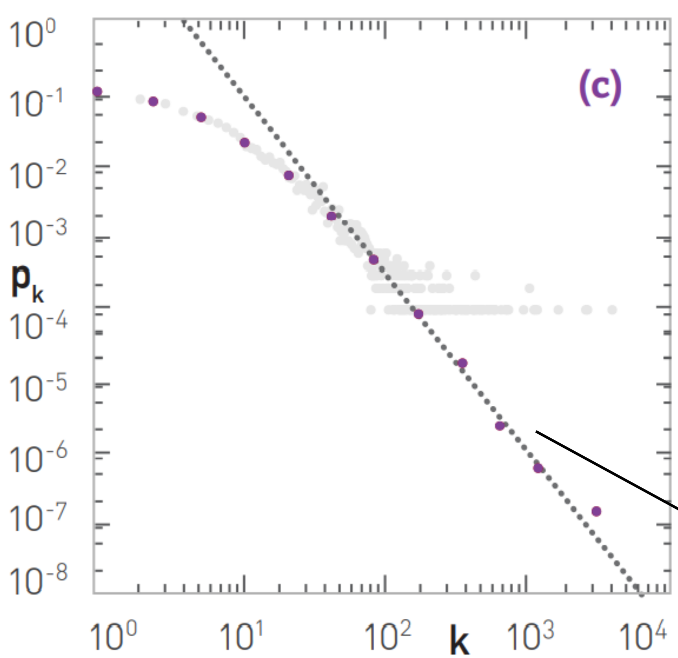
The tail of the distribution is visible but there is a plateau for high k

Often, only one copy of each high degree node so $N_k = 0$ or $N_k = 1$
➔ linear binning $p_k = 0$ (not shown) or $p_k = 1/N$, for all hubs, generating a plateau at $p_k = 1/N$. ☺

b) Logarithmic binning – so each datapoint has sufficient number of observations

The bin sizes increase with the degree ➔ *each bin has a comparable number of nodes*.

LOG-BINNING



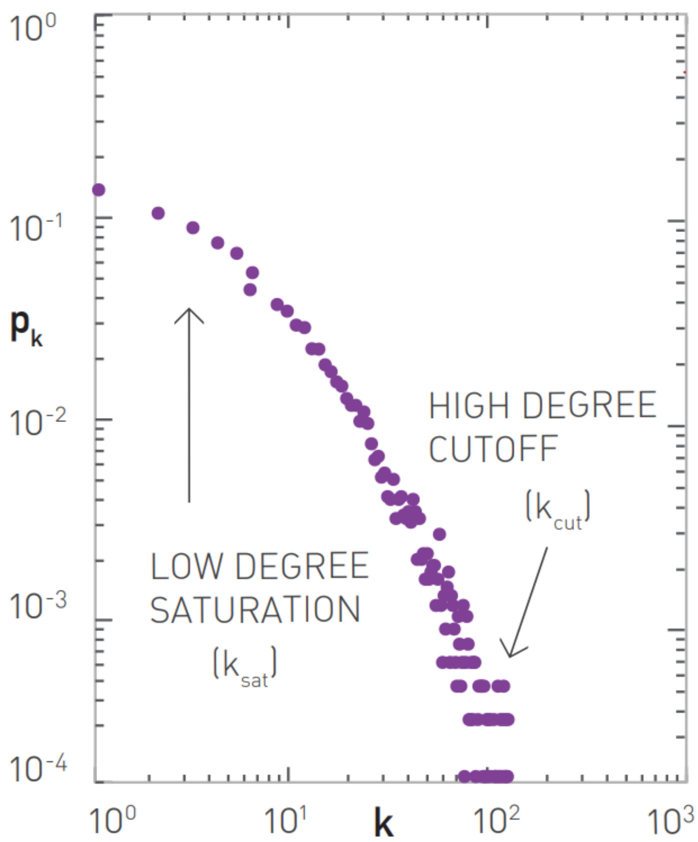
$p\langle k_n \rangle = N_n / b_n$

N_n is the number of nodes found in the bin n of size b_n

$\langle k_n \rangle$ is the average degree of the nodes in bin b_n.

The scaling extends into high k (invisible under linear binning).

c) Low degree saturation and high degree cutoff



Low-degree saturation:
a flattened p_k for $k < k_{sat}$

High-degree cutoff:
a rapid drop in p_k for $k > k_{cut}$
(fewer high-degree nodes than expected)

We can use a curtailed distribution.

Note: presence of cutoffs indicates the presence of additional phenomena.

http://tuvalu.santafe.edu/~aaronc/powerlaws/

Not all networks are scale free

- Networks appearing in material science
- Neural network of the C. elegans worm
- Power grid, consisting of generators and switches connected by transmission lines.

For the scale-free property: nodes need to have the capacity to link to an arbitrary number of other nodes.

The Ultra-small world property

How are distances in a network that has a scale free distribution of node degrees ?

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

$\gamma = 2$ Anomalous regime

Degree of the biggest hub grows linearly with the system size, i.e. $k_{\max} \sim N$. It is a *hub and spoke* configuration

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

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

Average distance increases as *ln ln N*
i.e. slower than the *ln N* for random networks.
The hubs radically reduce the path length

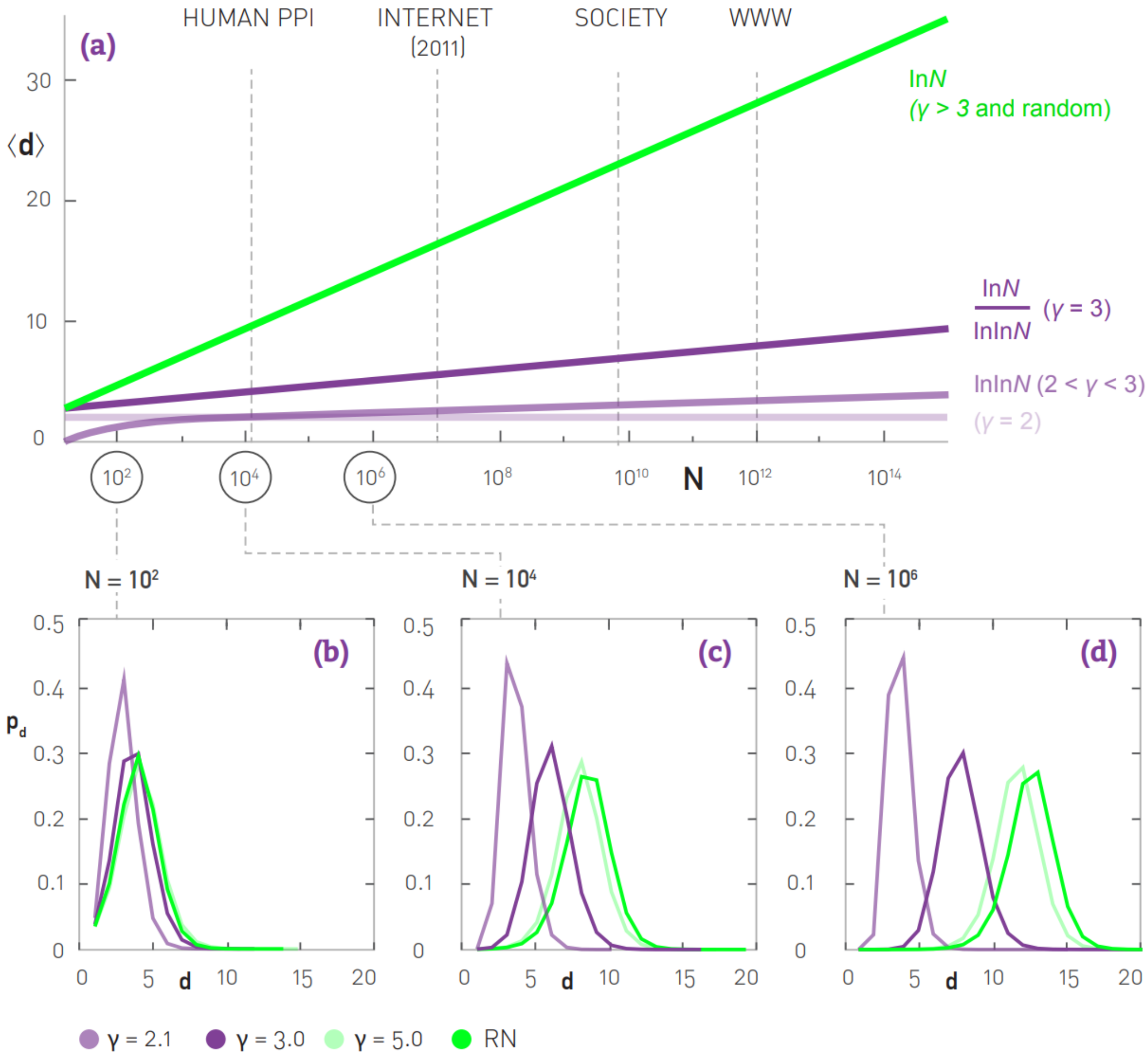
$\gamma = 3$ Critical Point

$\langle k^2 \rangle$ is finite
The *ln N* dependence comes back

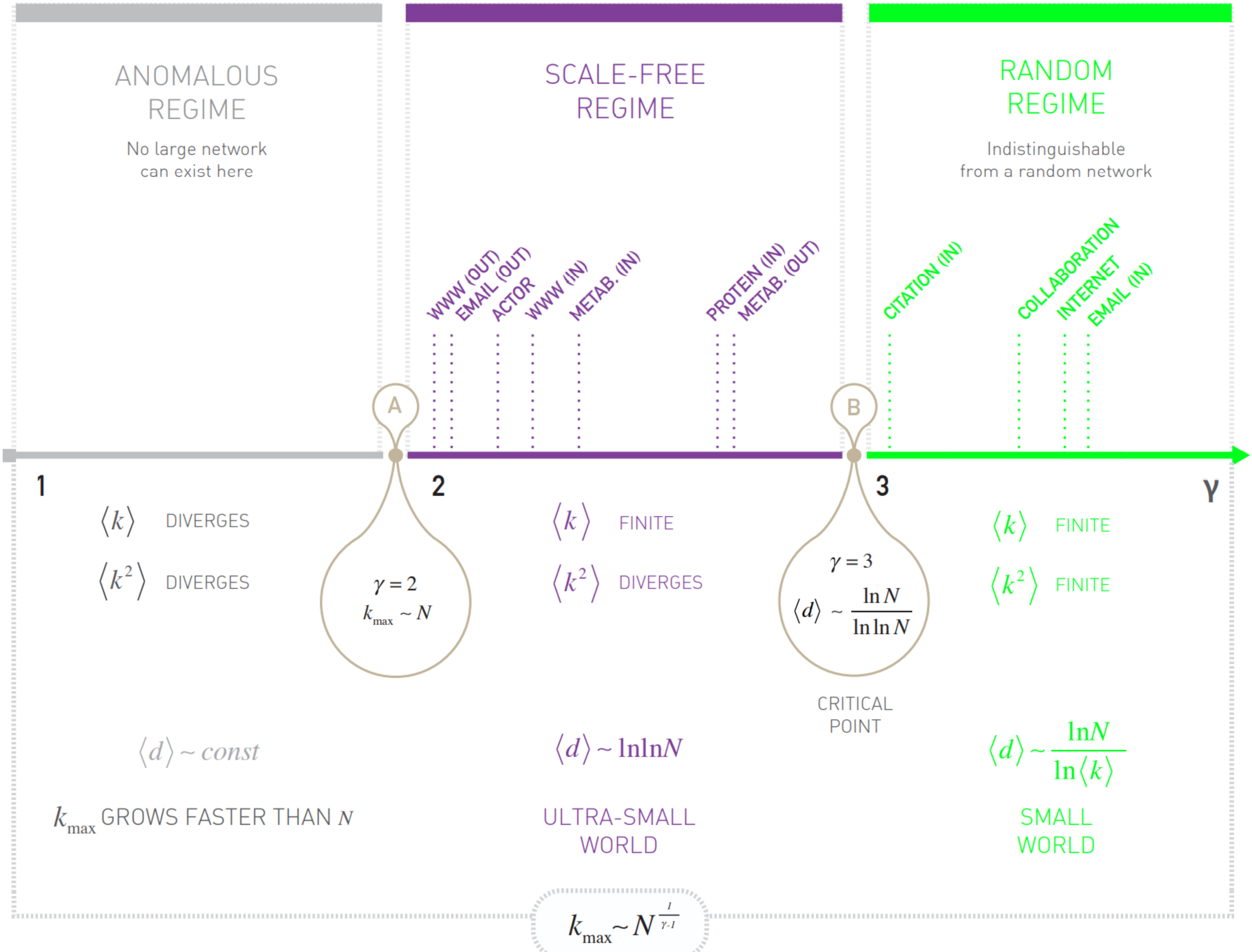
$\gamma > 3$ Small world

$\langle k^2 \rangle$ does not diverge
ln N dependence, hubs are present but not sufficiently large and numerous to have impact

Hubs effectively shrink the distances between nodes

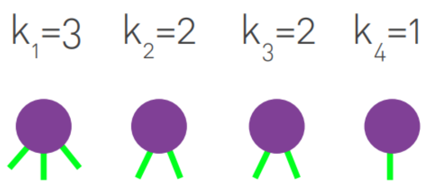


A.PACCANARO – FGV INTERNAL USE ONLY – DO NOT DISTRIBUTE


ANOMALOUS REGIME
No large network can exist here
SCALE-FREE REGIME
RANDOM REGIME
Indistinguishable from a random network
WWW (OUT)
EMAIL (OUT)
ACTOR
WWW (IN)
METAB. (IN)
PROTEIN (IN)
METAB. (OUT)
CITATION (IN)
COLLABORATION
INTERNET
EMAIL (IN)
A
B
1
2
3
γ
$\langle k \rangle$ DIVERGES
$\langle k^2 \rangle$ DIVERGES
$\gamma = 2$
$k_{max} \sim N$
$\langle k \rangle$ FINITE
$\langle k^2 \rangle$ DIVERGES
$\gamma = 3$
$\langle d \rangle \sim \frac{\ln N}{\ln \ln N}$
$\langle k \rangle$ FINITE
$\langle k^2 \rangle$ FINITE
CRITICAL POINT
$\langle d \rangle \sim const$
$\langle d \rangle \sim \ln \ln N$
$\langle d \rangle \sim \frac{\ln N}{\ln \langle k \rangle}$
k_{max} GROWS FASTER THAN N
ULTRA-SMALL WORLD
SMALL WORLD
$k_{max} \sim N^{\frac{1}{\gamma-1}}$

Generating networks with an arbitrary degree distribution

Configuration model



1. Assign a degree to each node, represented as stubs or half-links (even number of stubs ☺ !)
2. Randomly select a stub pair and connect them.

Depending on the selection order, we obtain different networks

Each node has a pre-defined degree k_i , but otherwise the network is wired randomly.

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

Degree preserving randomization

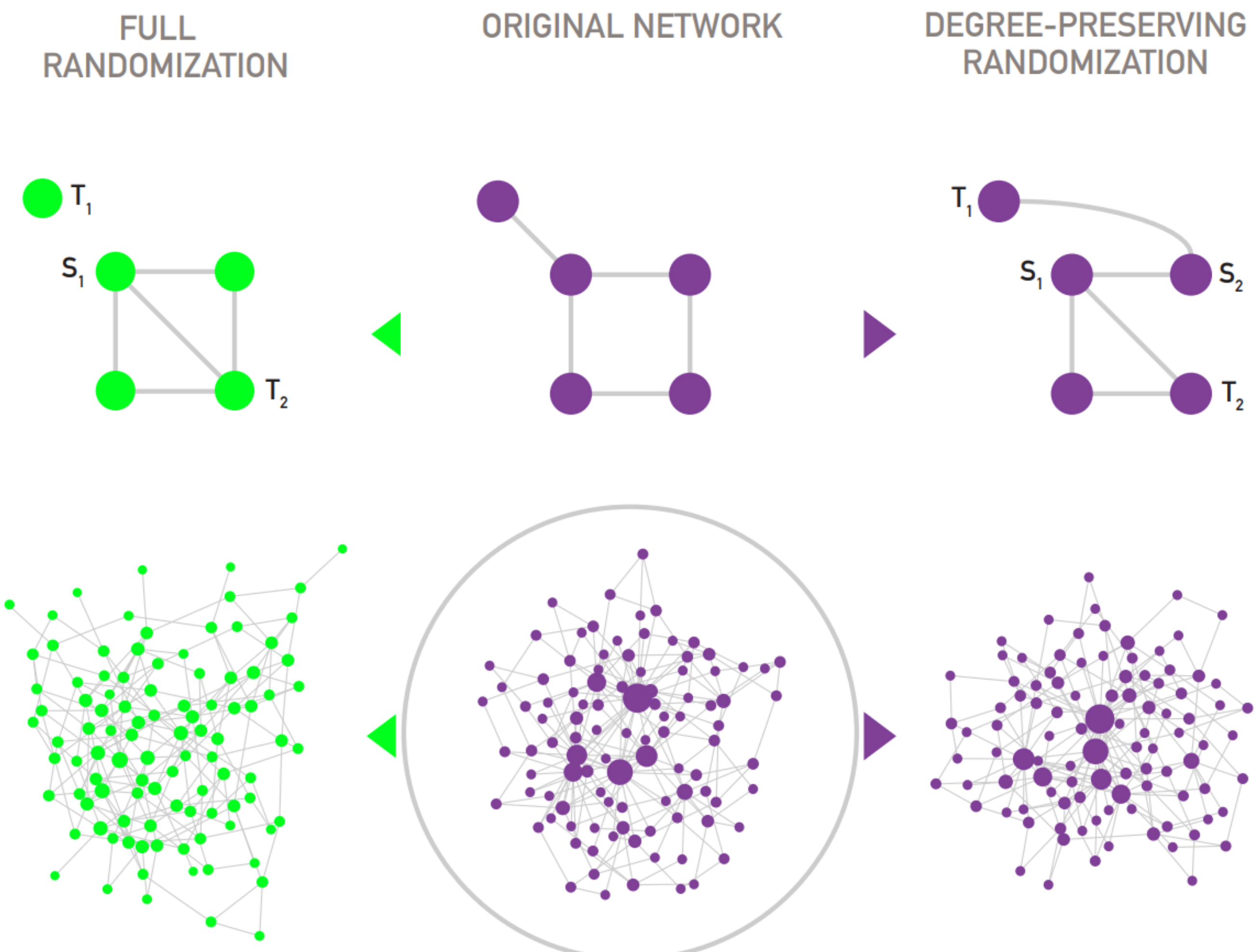
Important practical question: *given a network property, is it predicted by its degree distribution alone? or does it represents some additional property not explained by p_k?*

We need to generate networks that are wired randomly, with p_k is identical to the original network.

1. select two source (S1 , S2) and two target nodes (T1 , T2), such that S1-T1 and S2-T2 are linked.
2. swap the two links, creating an S1 -T2 and an S2-T1 link. repeat this procedure until we rewire each link at least once.

Hence the degree of each of the four involved nodes in the swap remains unchanged

Degree preserving randomization is different from full randomization (which generates Erdős-Rényi network with a Poisson degree distribution).



How to test the small scale property

Are the distances observed in a real network comparable with the distances observed in a randomized network with the same degree distribution?

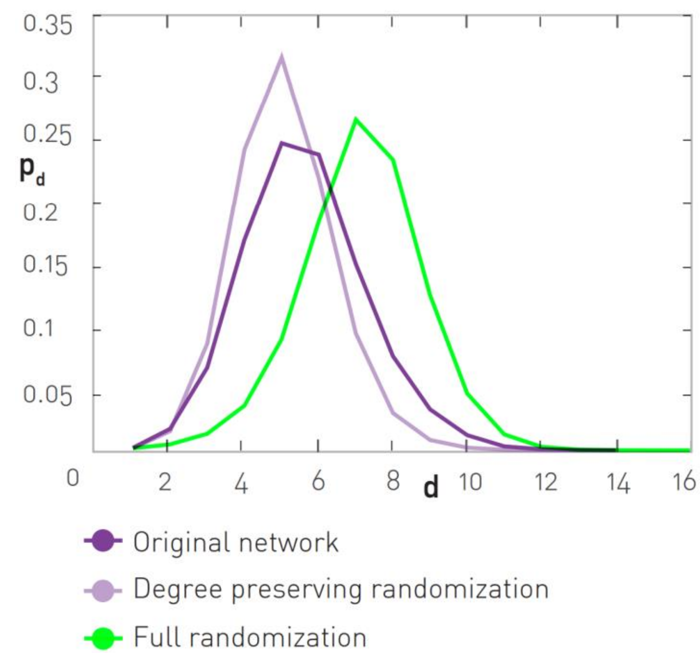
We measure the distance distribution p_d

1) on the original network

2) on a random network with the same N and L (here $\langle d \rangle = \frac{\ln N}{\ln \langle k \rangle}$)

3) after degree-preserving randomization

PPI network



$\langle d \rangle$=5.61±1.64 (original)
$\langle d \rangle$=7.13 ± 1.62 (full randomization)
$\langle d \rangle$=5.08 ± 1.34 (degree-preserving randomization)

Scale free networks 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

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

MOMENTS OF p_k for $N \to \infty$

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

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

DISTANCES

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