

# Scale-free networks

Introduction to Network Science

Carlos Castillo

Topic 04

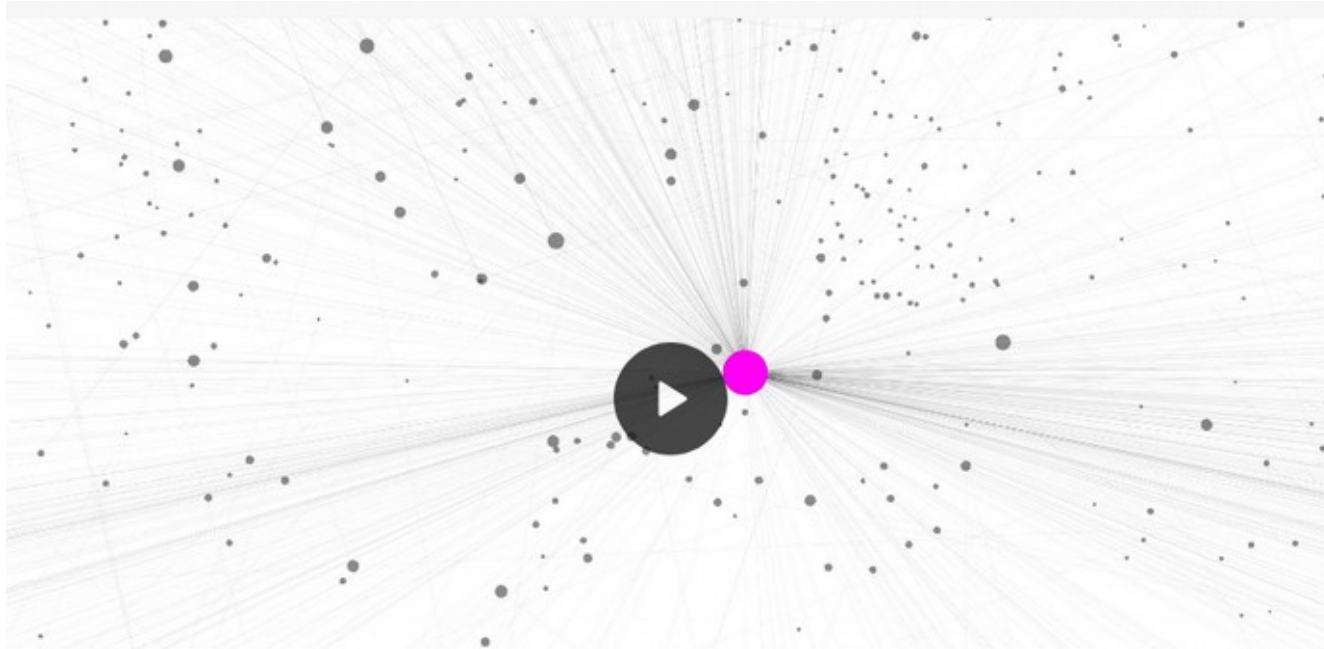
# Contents

- Characteristics of scale-free networks
- Degree distribution of scale-free networks
- Distance distribution of scale-free networks

# Sources

- Albert-László Barabási: Network Science. Cambridge University Press, 2016.
  - Follows almost section-by-section chapter 04
- URLs cited in the footer of specific slides

# nd.edu in 1998 (N=300K, L=1.5M) nd1998



<http://networksciencebook.com/images/ch-04/video-4-1.mov>

# What the Web Graph has but random networks don't have

- Large “hubs”
  - Nodes with a very high degree
  - Very unlikely in a random (ER) graph
- We have already seen the Poisson distribution is a bad approximation of the degree distribution

# A good approximation of degree in real networks

- Straight descending line in log-log plot

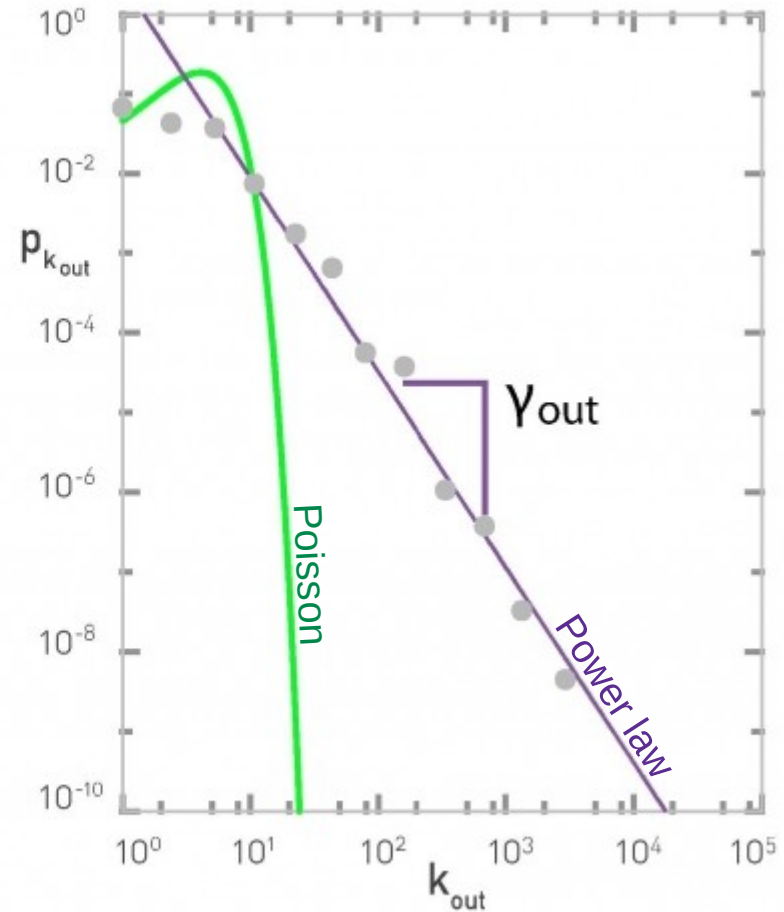
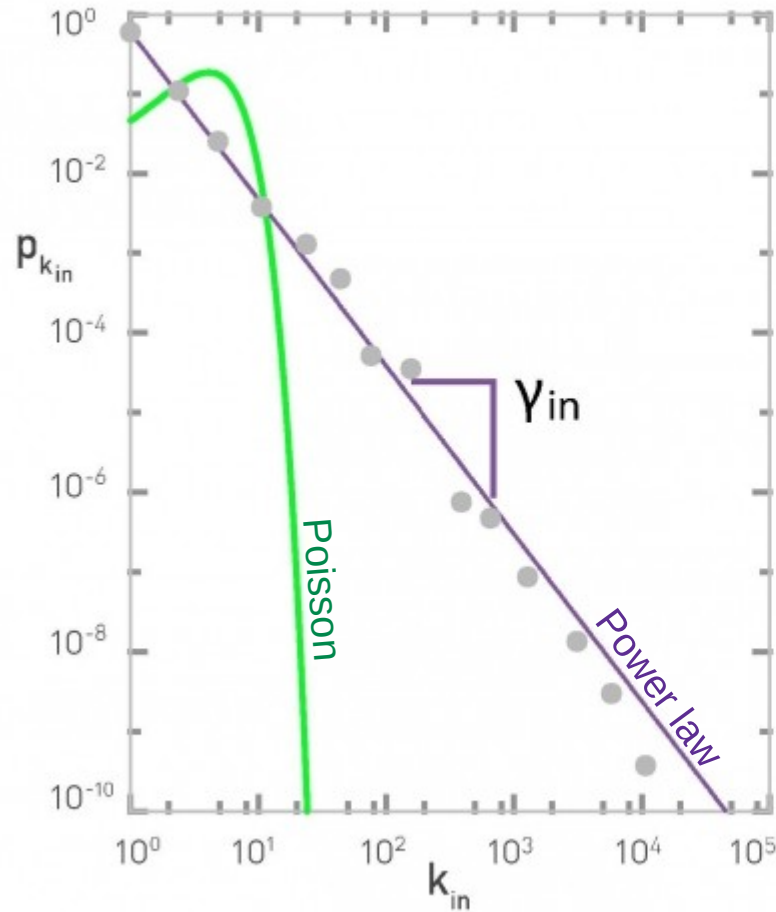
$$\log p_k \sim -\gamma \log k$$

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

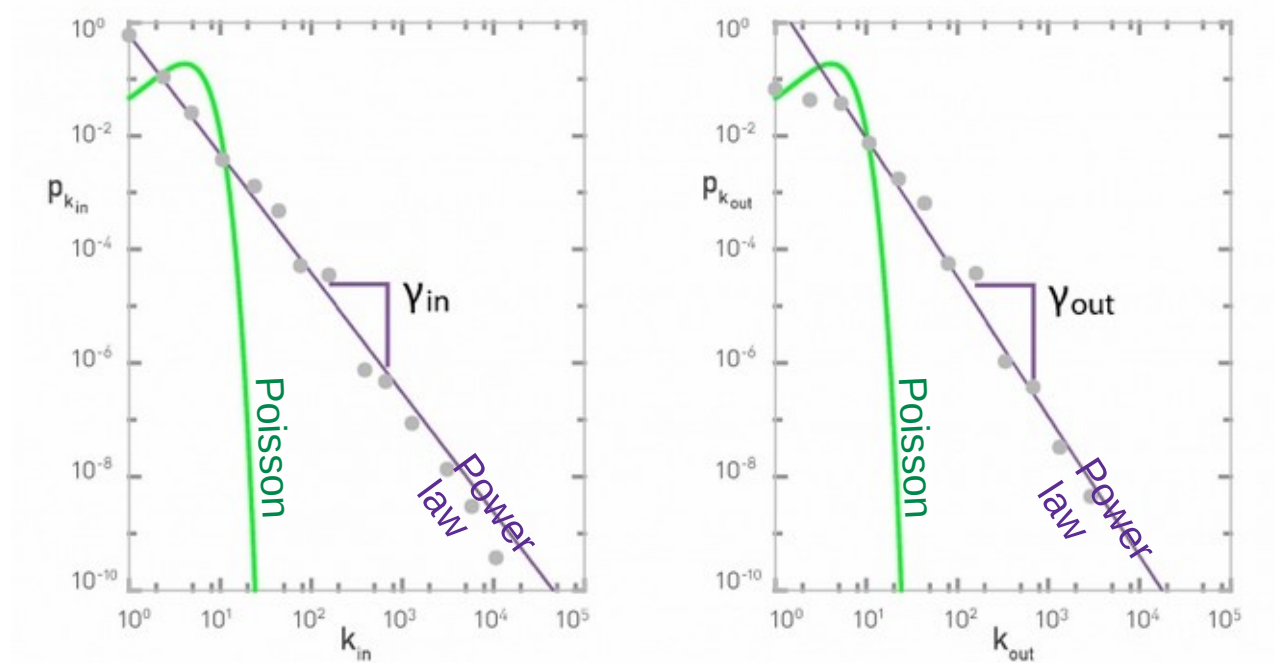
- Parameter  $\gamma$  is the exponent of the power law

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

# Degree distributions in nd1998



# Degree distributions in nd1998



What kind of values of gamma reduce the “long tail” of the power law?



# Parenthesis: 80/20 and Pareto

- Vilfredo Pareto in the 19<sup>th</sup> century noted 80% of money was earned by 20% of people
- More recently ...
  - 80 percent of links on the Web point to only 15 percent of pages;
  - 80 percent of citations go to only 38 percent of scientists;
  - 80 percent of links in Hollywood are to 30 percent of actors
- A debate that is still open: the wealth of the 1% and the 0.1%

# In directed networks ...

- Each node has two degrees:  $k_{\text{in}}$  and  $k_{\text{out}}$
- In general they may differ, hence

$$p_{k_{\text{in}}} \sim k^{-\gamma_{\text{in}}}$$

$$p_{k_{\text{out}}} \sim k^{-\gamma_{\text{out}}}$$

- In nd1998,  $\gamma_{\text{in}} \approx 2.1$ ,  $\gamma_{\text{out}} \approx 2.4$

# Formally (discrete)

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

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

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

Riemann's  
zeta

This formalism assumes there are no nodes with degree zero

# Formally (continuous)

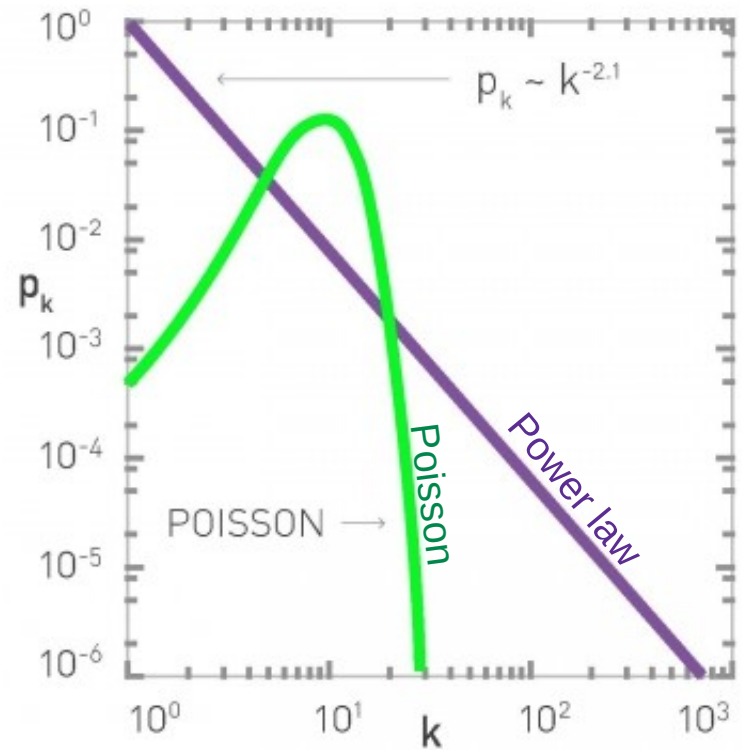
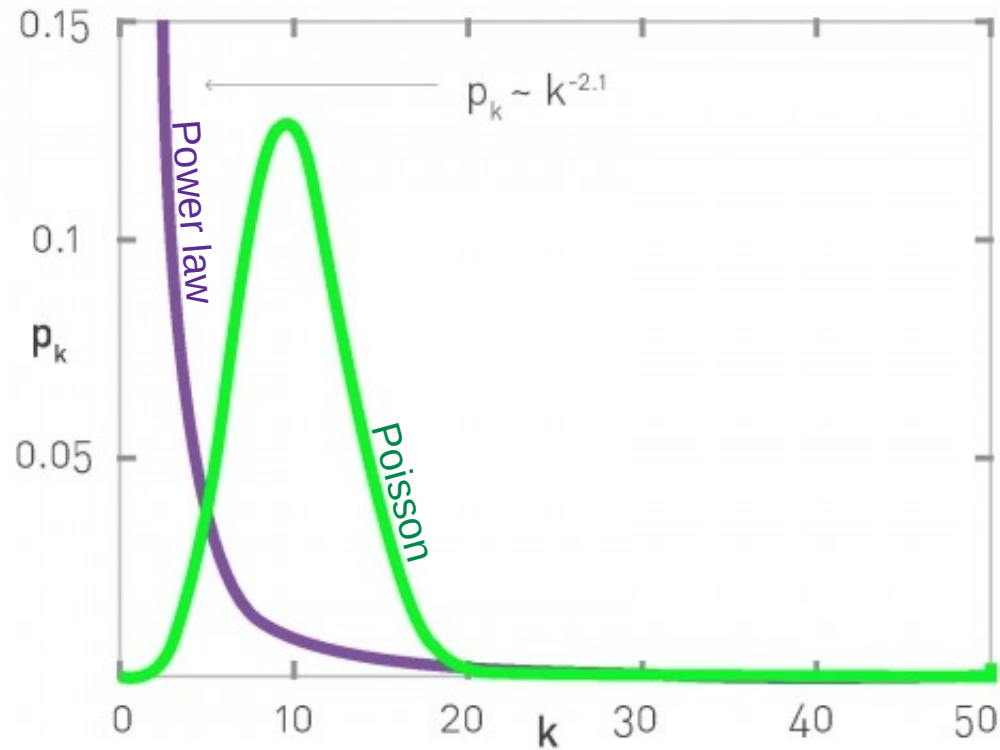
$$p_k = Ck^{-\gamma}$$

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

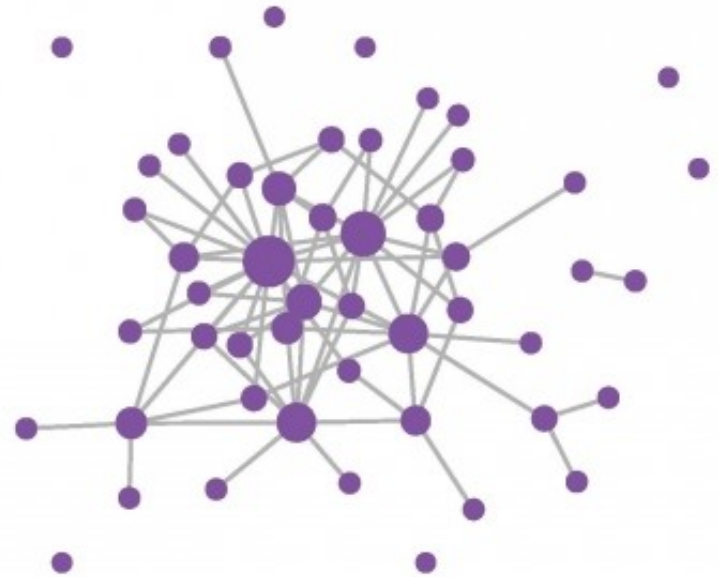
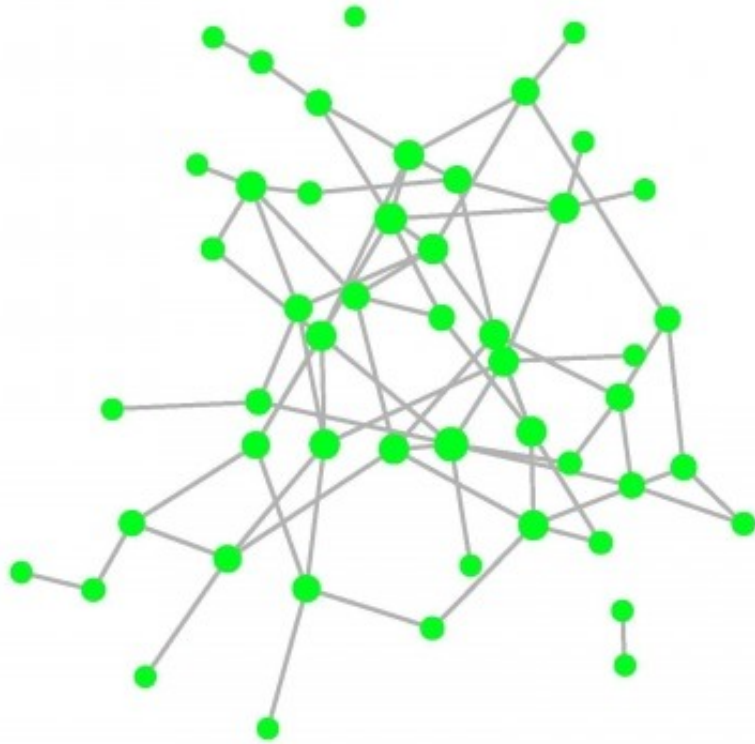
$$p_k = (\gamma - 1)k_{\min}^{\gamma-1} k^{-\gamma}$$

$k_{\min}$  is the smaller degree found in the network

# Comparing Poisson to power law

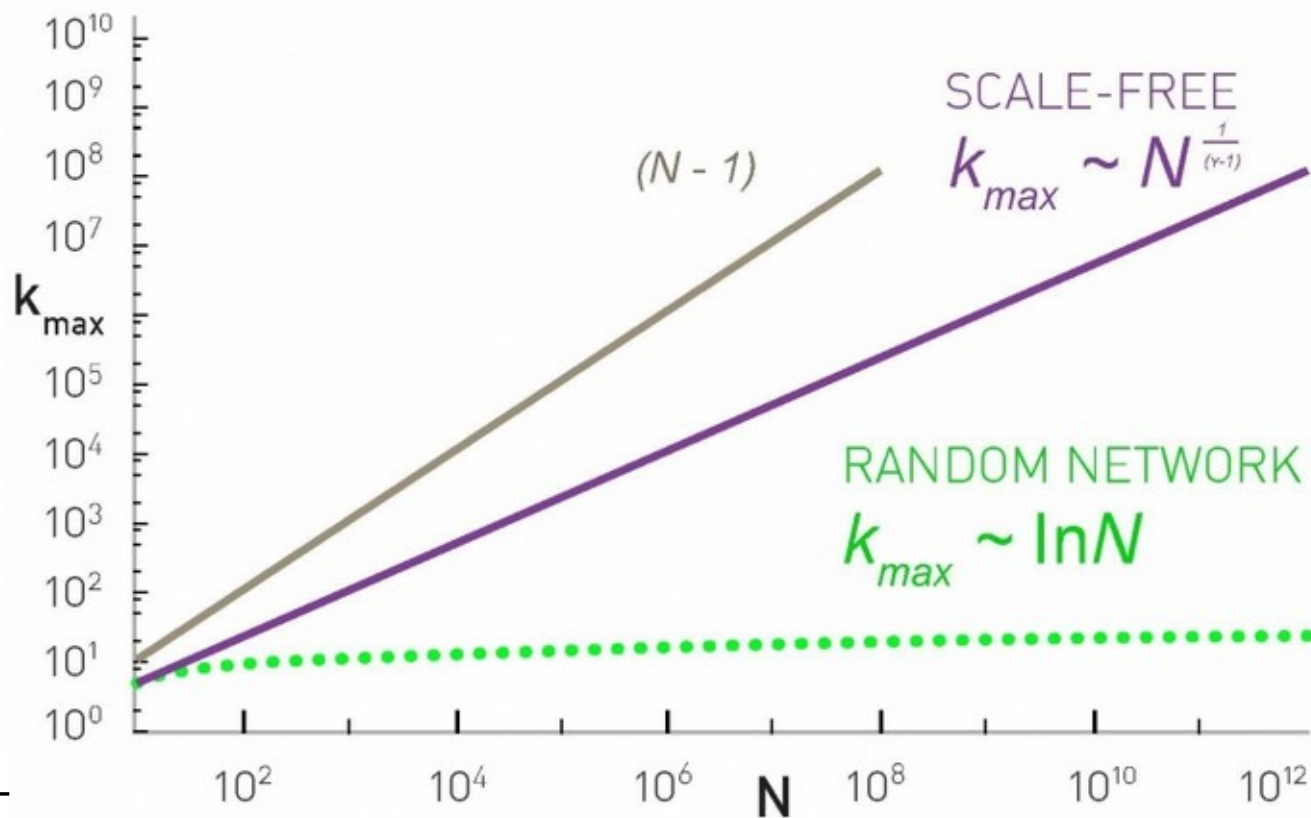


# Comparing Poisson to power law



# The natural cut-off of the degree

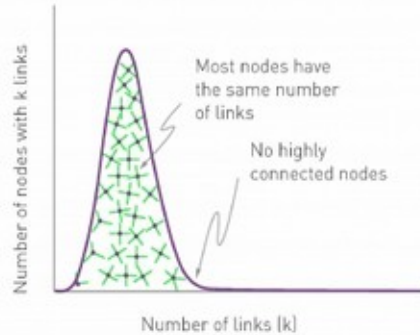
The largest hub cannot have more than  $N-1$  connections



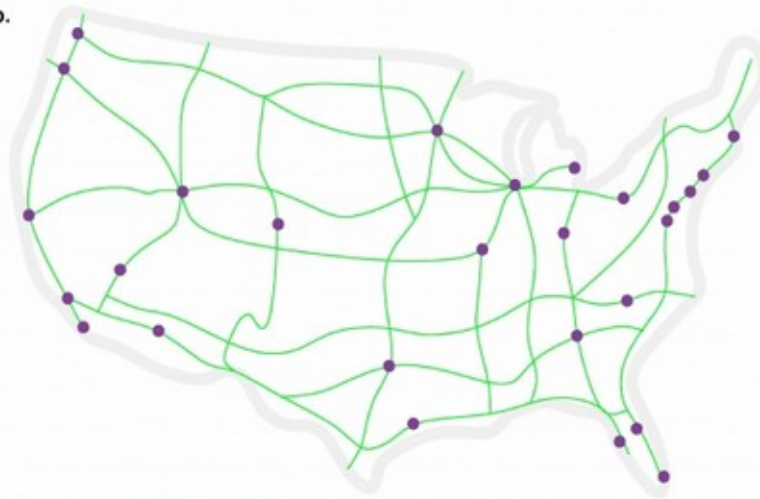
$$k_{\max} = k_{\min} N^{\frac{1}{\gamma-1}}$$

# Random vs scale-free networks

a. POISSON

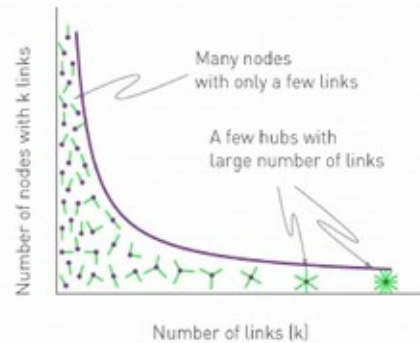


b.



Ground transportation

c. POWER LAW



d.



Air transportation



# What does it mean “scale-free”?

- A distribution has a “scale” if values are close to each other, for instance in a random network  $\sigma_k = \langle k \rangle^{1/2}$
- Hence, most nodes are in the range  $\langle k \rangle \pm \langle k \rangle^{1/2}$
- However in scale-free networks ...

# What does it mean “scale-free”?

- Moments of degree distribution

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

$$C = (\gamma - 1) k_{\min}^{\gamma-1}$$

# What does it mean “scale-free”?

$$\sigma_k = \langle k^2 \rangle - \langle k \rangle^2$$

- In a scale-free network

$$\langle k^2 \rangle = \int_{k_{\min}}^{k_{\max}} k^2 p_k dk = C \frac{k_{\max}^{3-\gamma} - k_{\min}^{3-\gamma}}{3 - \gamma}$$

- This diverges as  $k_{\max} \rightarrow \infty$  if  $\gamma < 3$
- Hence there is no “typical” scale

# What does it mean “scale-free”?

$$\sigma_k = \langle k^2 \rangle - \langle k \rangle^2$$

- In a scale-free network

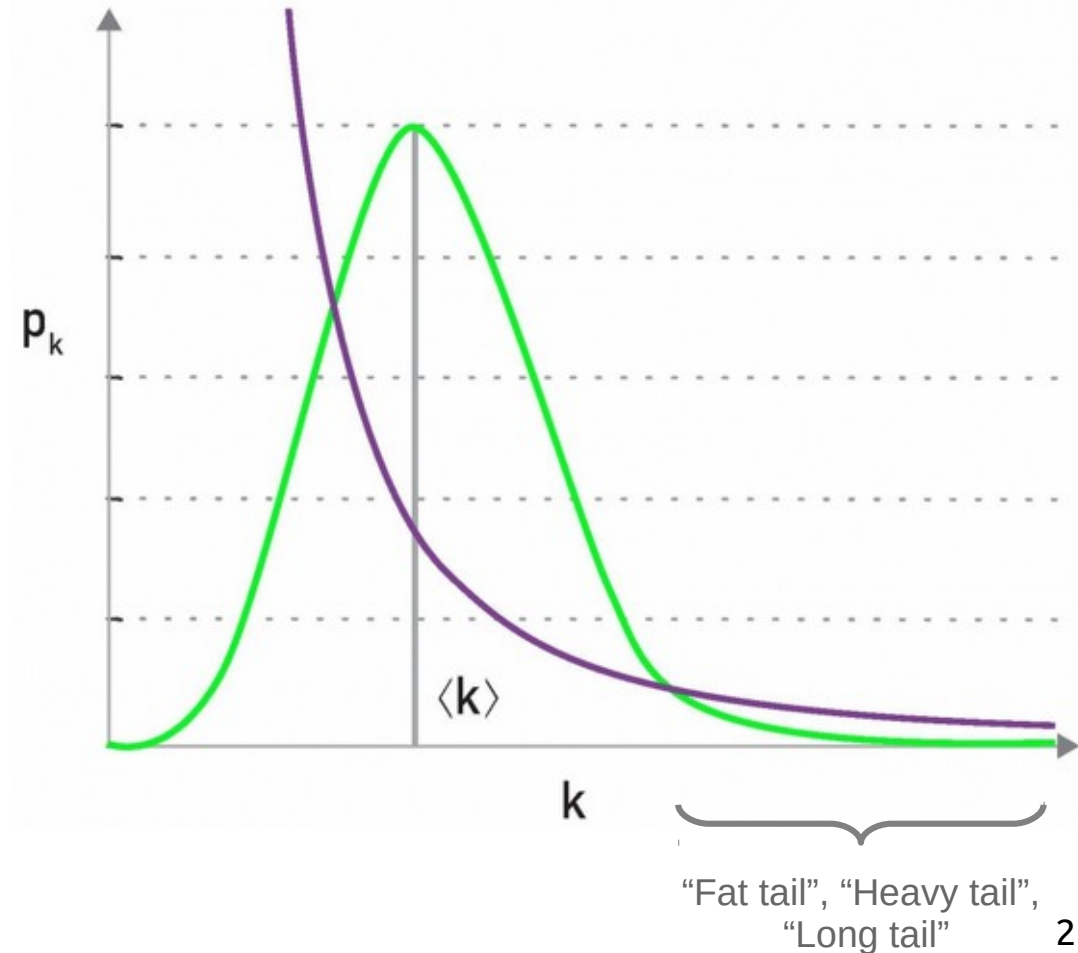
$$\langle k^2 \rangle = \int_{k_{\min}}^{k_{\max}} k^2 p_k dk = C \frac{k_{\max}^{3-\gamma} - k_{\min}^{3-\gamma}}{3 - \gamma}$$

- What happens with the variance of the degree for networks with high max degree?

# Example: nd1998

$$k_{\text{in}} = 4.60 \pm 1546$$

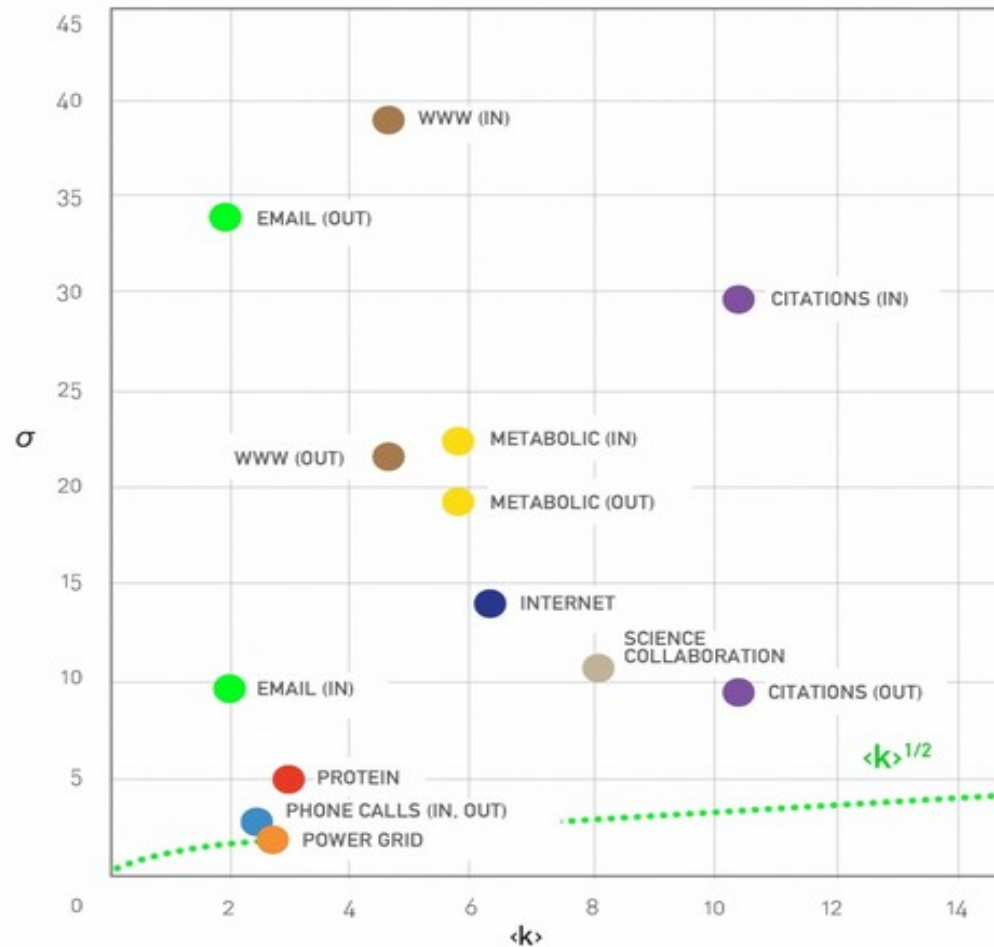
In general, the  
average degree is  
not very informative  
in scale-free  
networks



# Real network examples

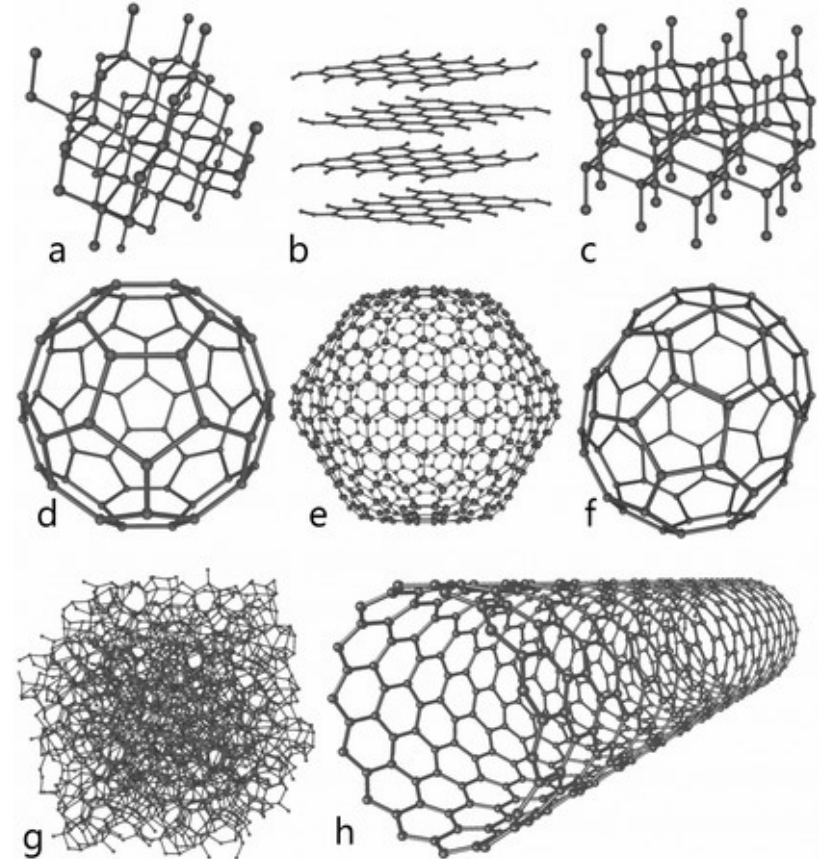
Network	$N$	$L$	$\langle k \rangle$	$\langle k_{in}^2 \rangle$	$\langle k_{out}^2 \rangle$	$\langle k^2 \rangle$	$\gamma_{in}$	$\gamma_{out}$	$\gamma$
Internet	192,244	609,066	6.34	-	-	240.1	-	-	3.42*
WWW	325,729	1,497,134	4.60	1546.0	482.4	-	2.00	2.31	-
Power Grid	4,941	6,594	2.67	-	-	10.3	-	-	Exp.
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*-

# Real network examples



# When you don't observe the scale-free property

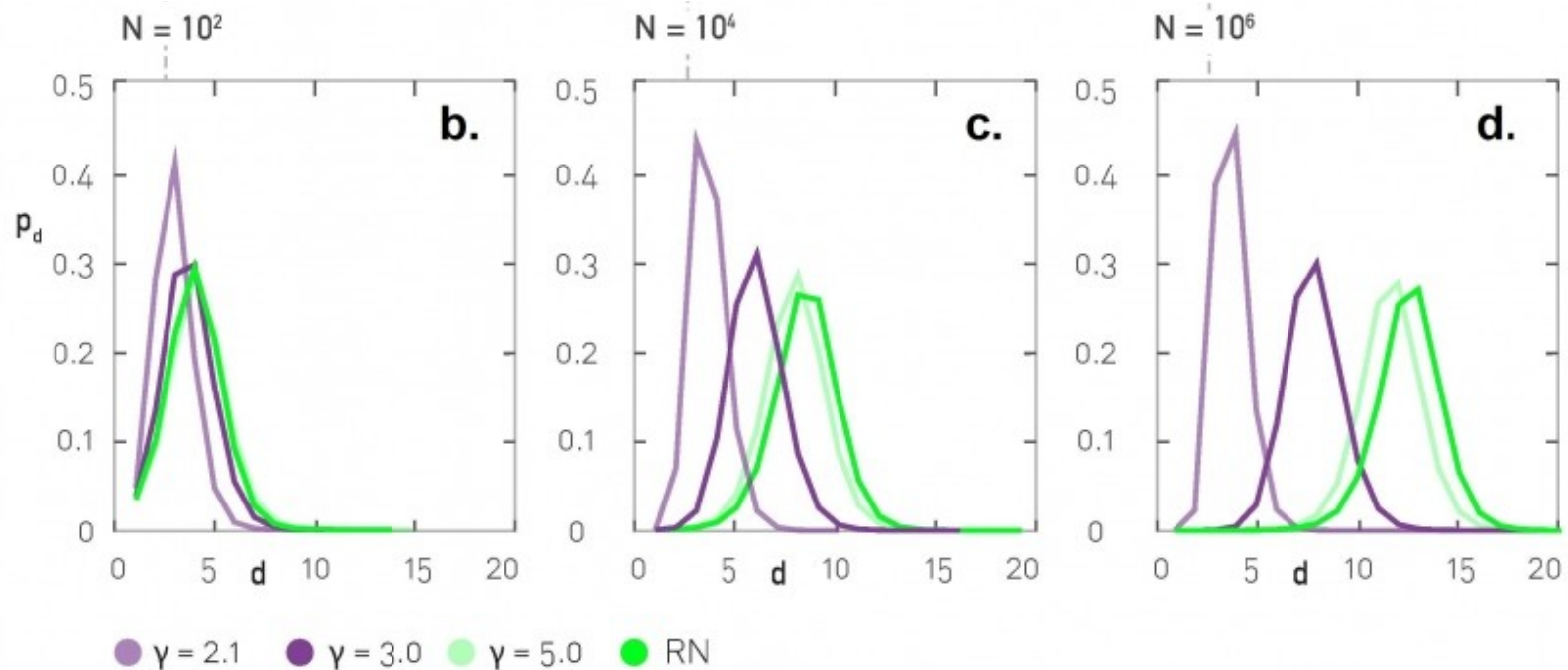
- In general, when there is a limit to  $k_{\max}$
- Out-degree in some social networks
- Materials networks





# Distance distributions: simulation results

Scale-free networks of increasing size,  $\langle k \rangle = 3$



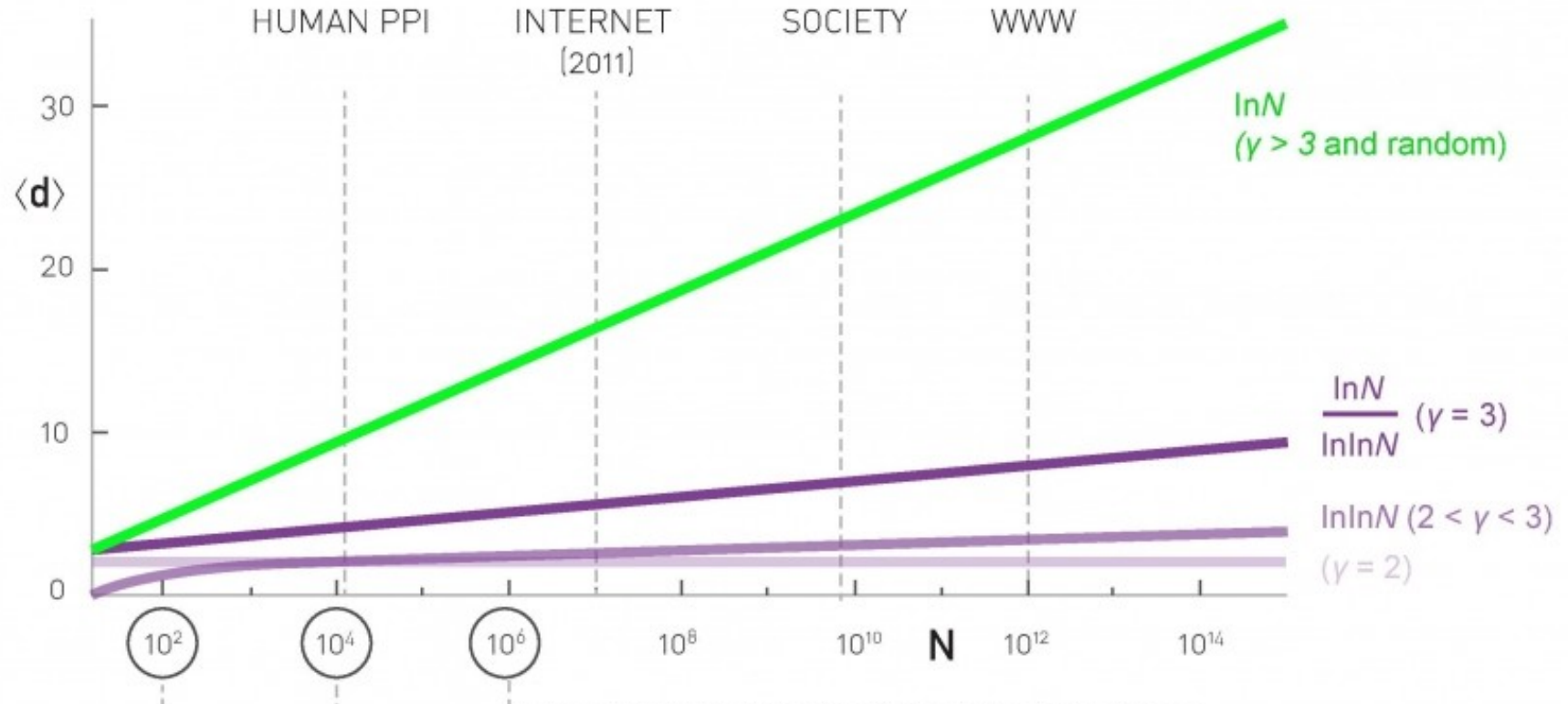
# Average distance

- Depends on  $\gamma$  and  $N$

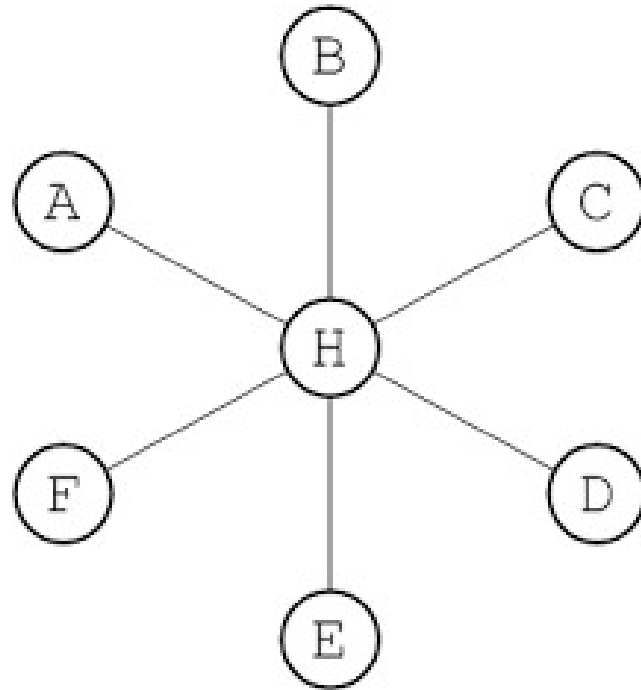
$$\langle d \rangle = \begin{cases} \text{const.} & \text{if } \gamma = 2 \\ \log \log N & \text{if } 2 < \gamma < 3 \\ \log N / \log \log N & \text{if } \gamma = 3 \\ \log N & \text{if } \gamma > 3 \end{cases}$$

Same as in  
ER graphs

# Average distance and N



# Anomalous regime $\gamma = 2$



# Ultra-small world $2 < \gamma < 3$

- Average distance follows  $\log(\log(N))$
- Example (humans):

$$N \approx 7 \times 10^9$$

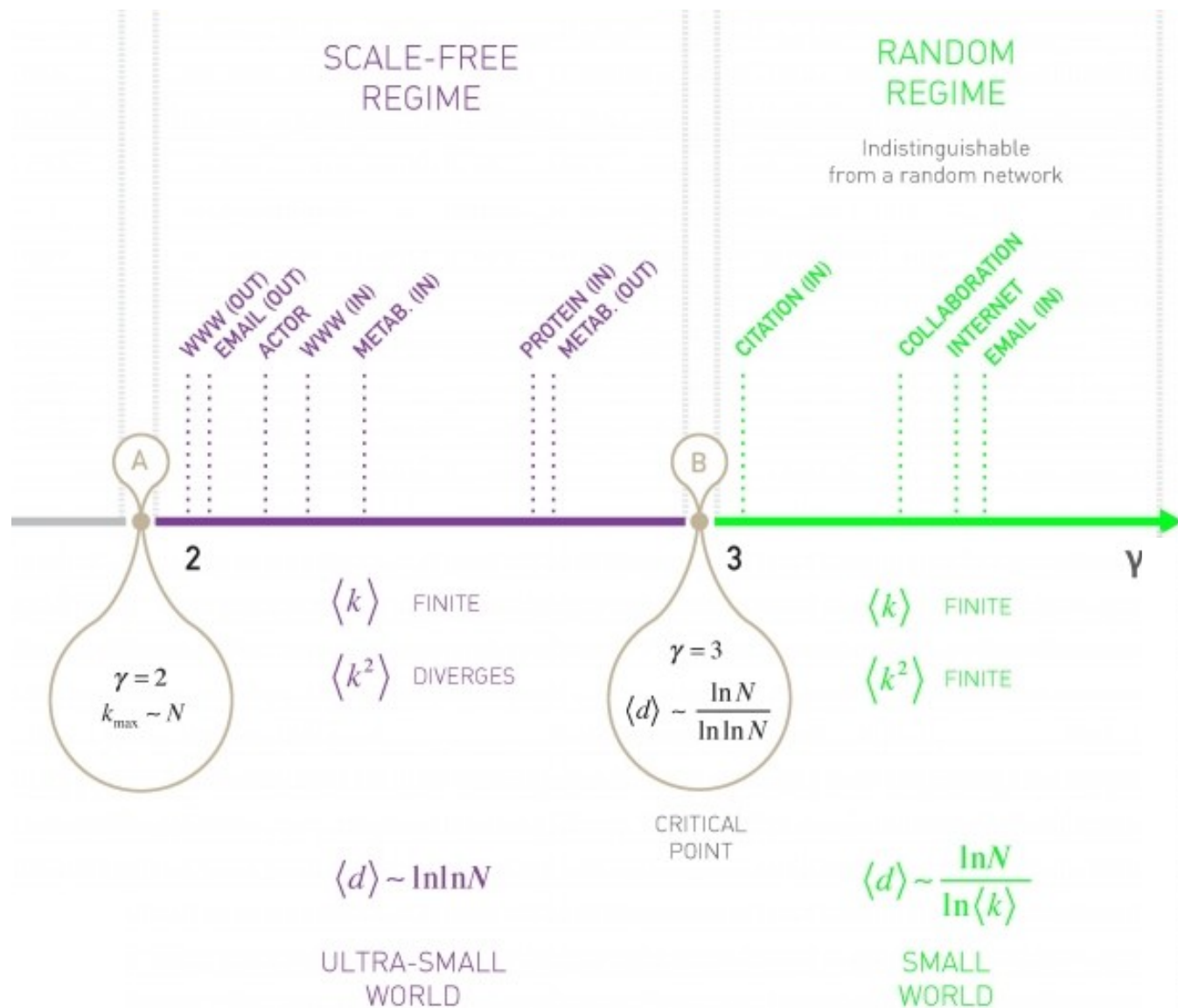
$$\log N \approx 22.66$$

$$\log \log N \approx 3.12$$

# Small world $\gamma > 3$

- Average distance follows  $\log(N)$
- Similar to ER graphs where it followed  $\log(N)/\log(\langle k \rangle)$

The degree distribution exponent plays an important role



# When $\gamma > 3$

- In this case it is hard to distinguish the case from an ER graph
- In most real complex networks (but not all)

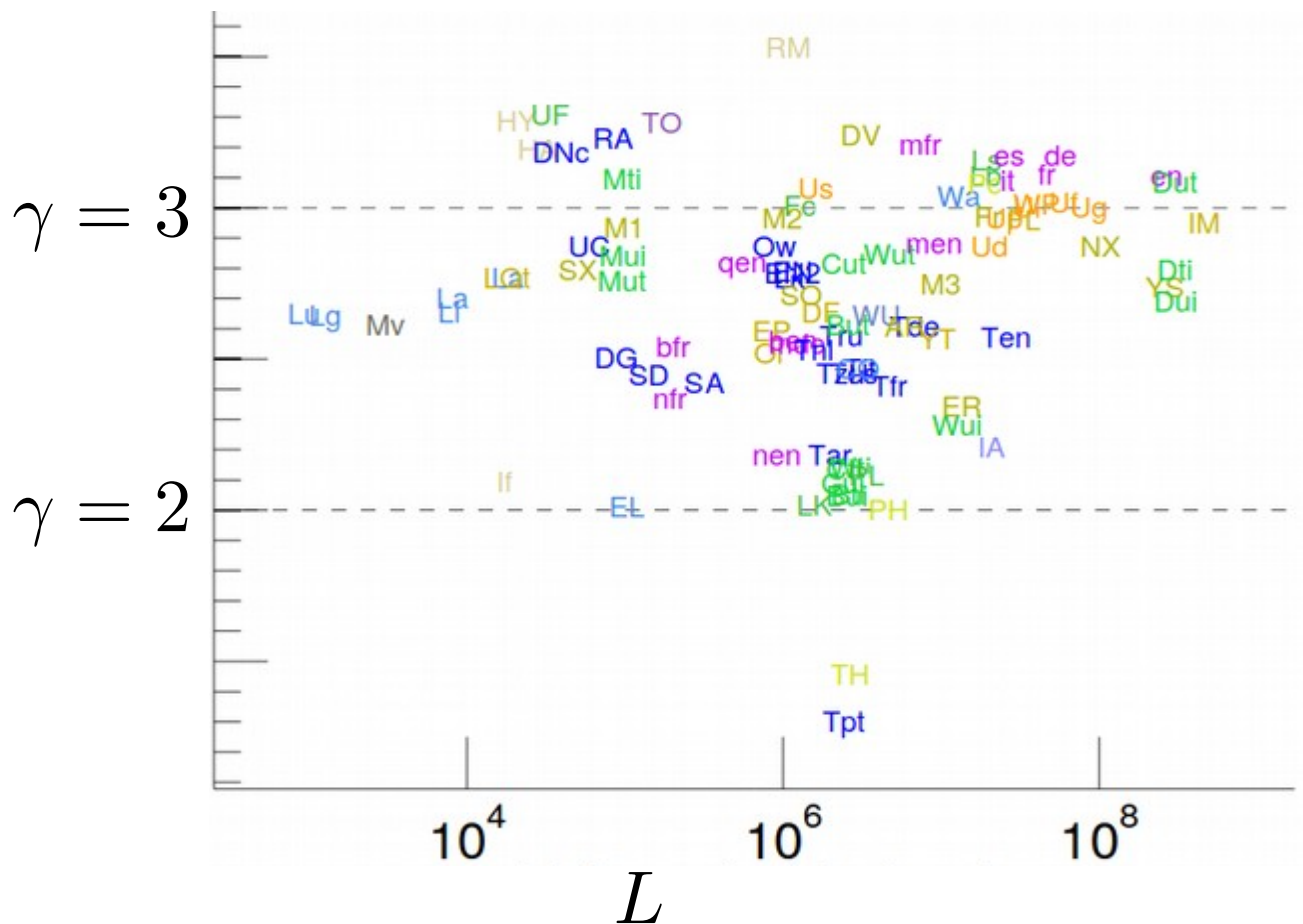
$$2 < \gamma < 3$$



# When $\gamma > 3$

- Remember  $k_{\max} = k_{\min} N^{\frac{1}{\gamma-1}}$        $N = \left( \frac{k_{\max}}{k_{\min}} \right)^{\gamma-1}$
- Observing the scale-free properties requires that  $k_{\max} \gg k_{\min}$ , e.g.  $k_{\max} = 10 k_{\min}$
- Then if  $\gamma = 5$ ,  $N > 10^8$
- Hence we won't find many such networks

# Examples



<b>EL</b>	Wikipedia elections
<b>LK</b>	Linux kernel mailing list threads
<b>Bul</b>	BibSonomy u-i
<b>Bti</b>	BibSonomy t-i
<b>Cul</b>	CiteULike u-i
<b>If</b>	Infectious
<b>PL</b>	Prosper loans
<b>Cti</b>	CiteULike t-i
<b>Wti</b>	Twitter t-i
<b>nen</b>	Wikinews (en)
<b>Tar</b>	Wikipedia talk, Arabic
<b>Wul</b>	Twitter u-i
<b>ER</b>	Epinions
<b>nfr</b>	Wikinews (fr)
<b>Tfr</b>	Wikipedia talk, French
<b>SD</b>	Slashdot
<b>Tzh</b>	Wikipedia talk, Chinese
<b>Tes</b>	Wikipedia talk, Spanish

Etc.

# Exercise [B. 2016, Ex. 4.10.2]

## "Friendship Paradox"

- Remember  $p_k$  is the probability that a node has  $k$  "friends"
- If we randomly select a link, the probability that a node at any end of the link has  $k$  friends is  $q_k = C k p_k$  where  $C$  is a normalization factor
  - (a) Find  $C$  (the sum of  $q_k$  must be 1)

# Exercise [B. 2016, Ex. 4.10.2]

## "Friendship Paradox"

- If we randomly select a link, the probability that a node at any end of the link has  $k$  friends is  $q_k = C k p_k$  where  $C$  is a normalization factor
  - (b)  $q_k$  is also the prob. that a randomly chosen node has a neighbor of degree  $k$ ; find its average

# Exercise [B. 2016, Ex. 4.10.2]

## "Friendship Paradox"

(c-d) Compute the expected number of friends of a neighbor of a randomly chosen node; compare with the expected number of friends of a randomly chosen node when

$$N = 10000$$

$$\gamma = 2.3$$

$$k_{\min} = 1$$

$$k_{\max} = 1000$$

$$\langle k^n \rangle = C \frac{k_{\max}^{n-\gamma+1} - k_{\min}^{n-\gamma+1}}{n - \gamma + 1}$$

$$C = (\gamma - 1) k_{\min}^{\gamma-1}$$

# Python code

```
def degree_moment(kmin, kmax, moment, gamma):  
    C = (gamma-1.0)*(kmin**(gamma-1.0))  
    numerator = (kmax**(moment-gamma+1.0) - kmin**(moment-gamma+1.0))  
    denominator = (moment-gamma+1.0)  
    return C * numerator / denominator
```

```
kavg = degree_moment(kmin=1, kmax=1000, moment=1, gamma=2.3)  
print(kavg)
```

3.787798988222529

```
ksqavg = degree_moment(kmin=1, kmax=1000, moment=2, gamma=2.3)  
print(ksqavg)
```

231.94329076177414

```
print(ksqavg / kavg)
```

61.23431879119234

# Practice on your own

- Remember the regimes of a graph given  $\langle k \rangle$   
(It's useful to know this by heart)
- Estimate degree distributions and distance distributions for some graphs
- Apply the friendship paradox to some graphs