Scale-Free Networks

Introduction to Network Science
Carlos Castillo
Topic 09

Universitat Pompeu Fabra Barcelona

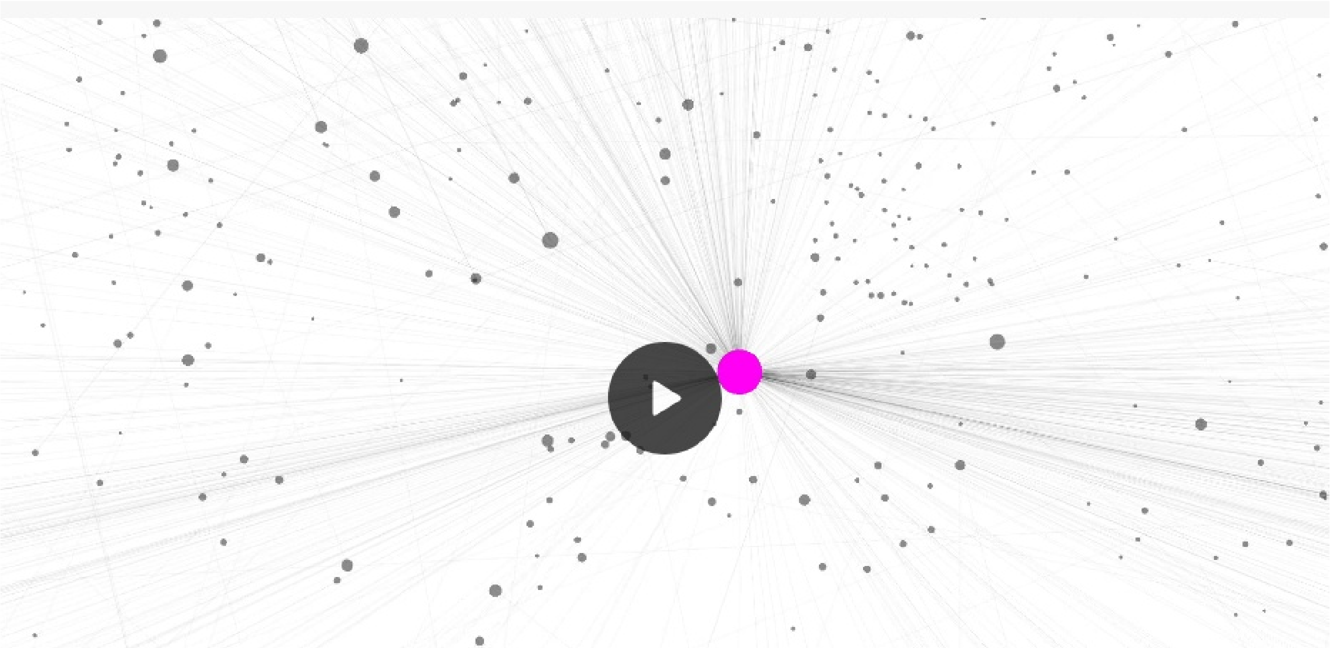
Contents

- 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

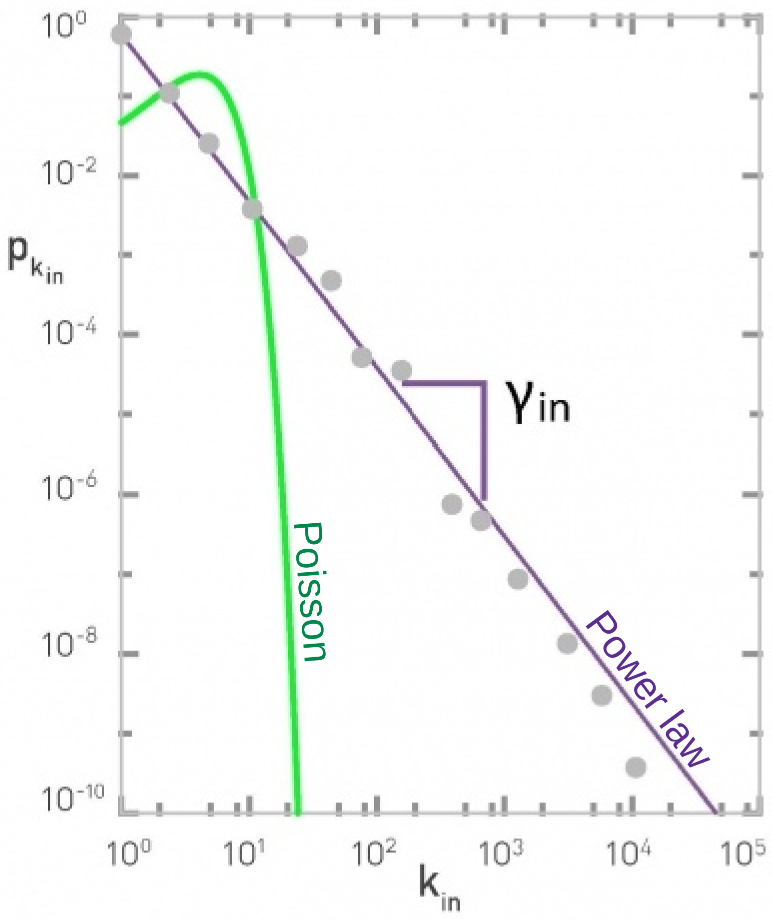


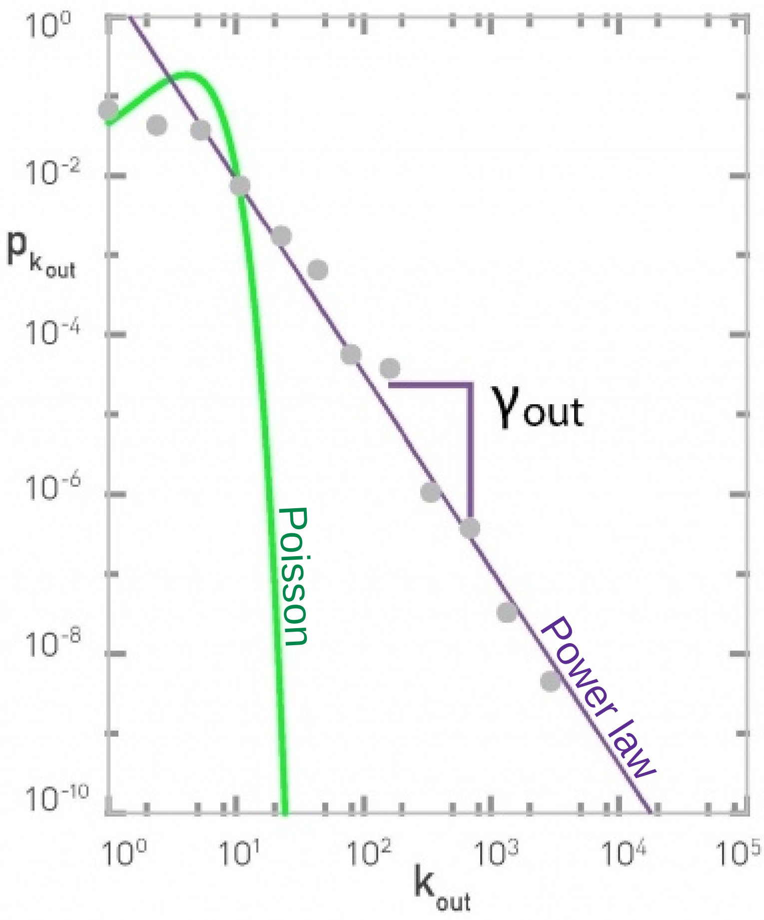
http://networksciencebook.com/images/ch-04/video-4-1.mov

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 observed degree distribution

Degree distributions in nd1998





A good approximation of degree in real networks

- Straight descending line in log-log plot

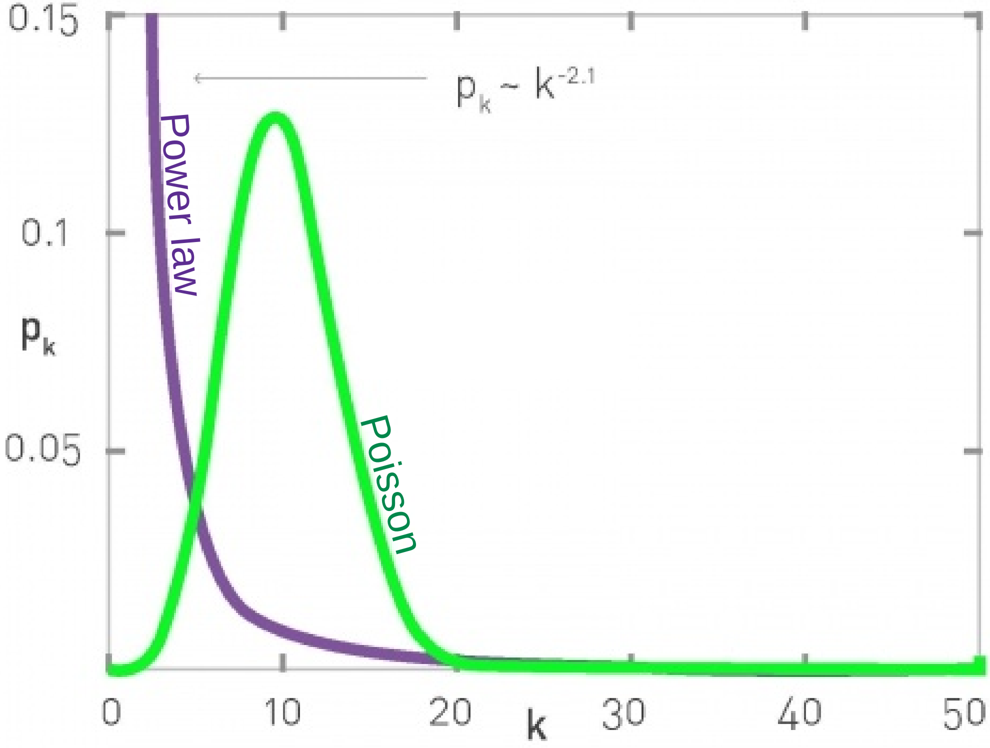
$$\log p_k \sim -\gamma \log k$$

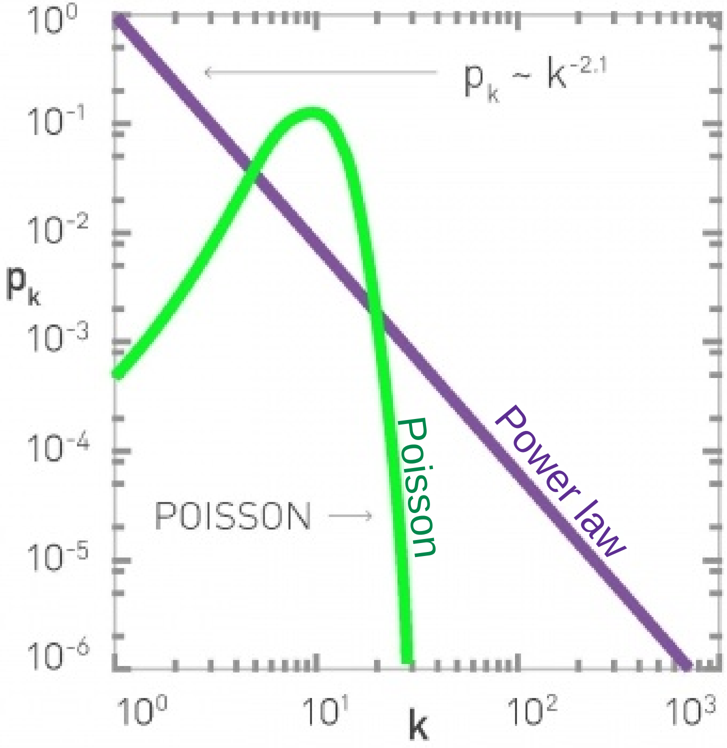
$$p_k \sim k^{-\gamma}$$

- Parameter γ is the exponent of the power law

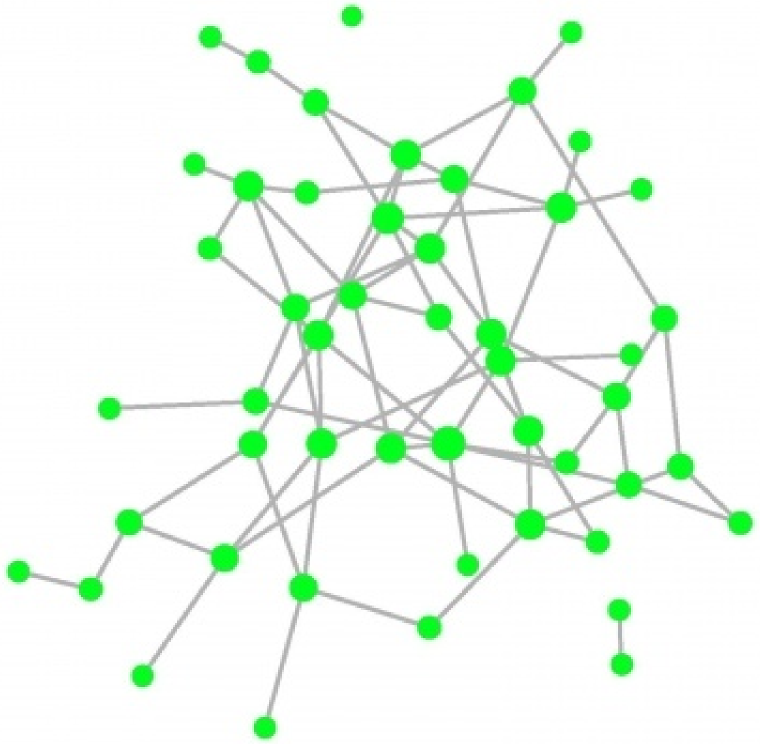
A scale-free network is a network whose degree distribution follows a power law

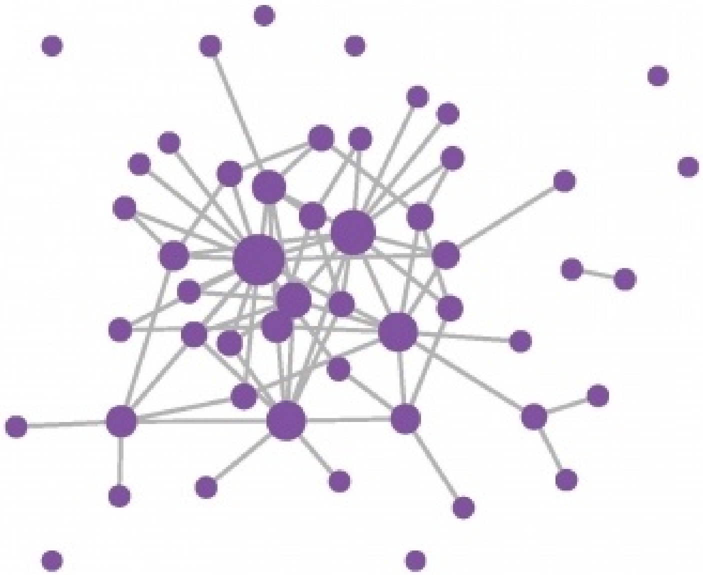
Comparing Poisson to power law



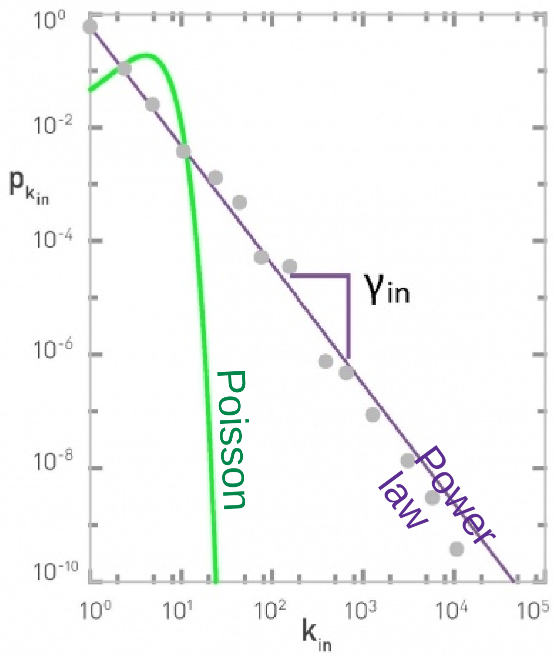


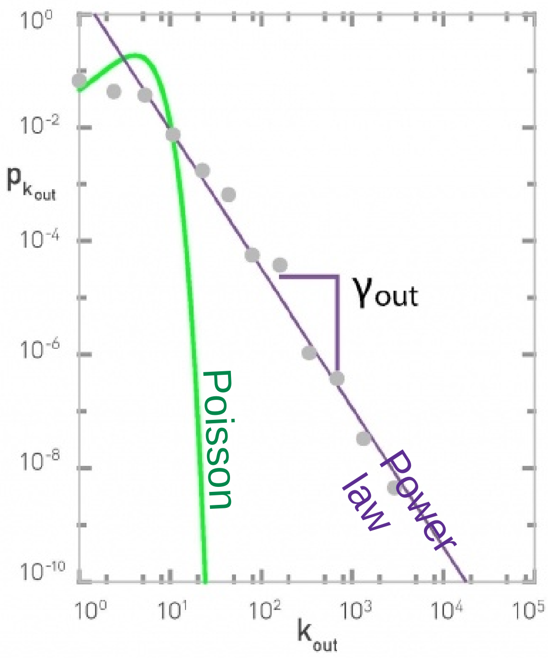
Comparing Poisson to power law





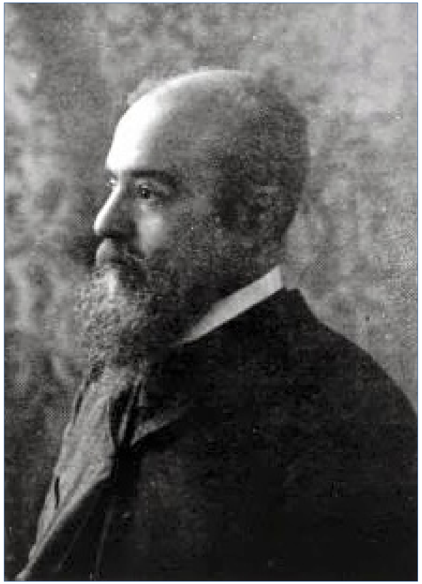
Degree distributions in nd1998





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

Parenthesis: Pareto



- Italian economist Vilfredo Pareto in the 19th 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_{in} and k_{out}
- In general they may differ, hence

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

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

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

Formally (discrete)

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

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

Riemann’s zeta

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

This formalism assumes there are no nodes with degree zero

Formally (continuous approx.)

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

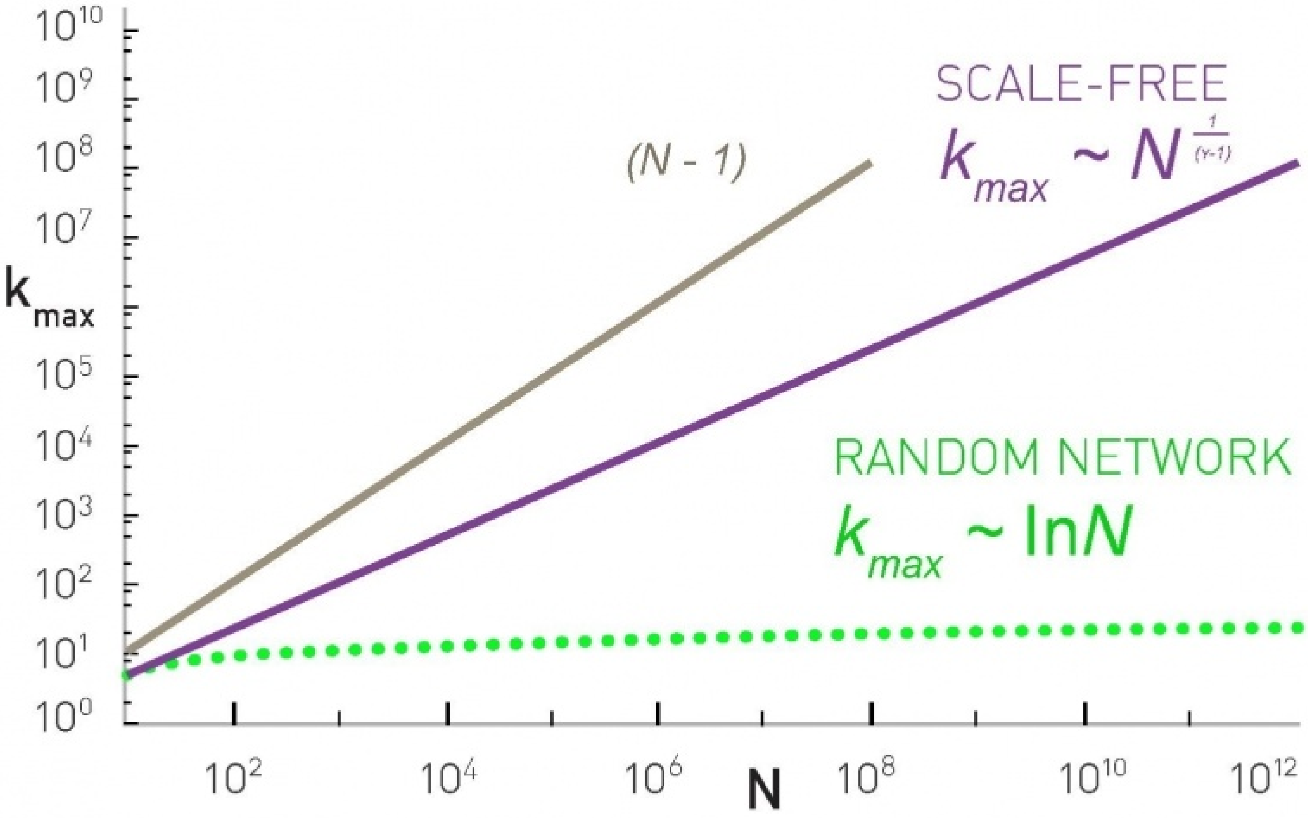
$$\int_{k=k_{\min}}^{\infty} p_k = 1 \longrightarrow C = \frac{1}{\int_{k=k_{\min}}^{\infty} k^{-\gamma}} = (\gamma - 1)k_{\min}^{\gamma-1}$$

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

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

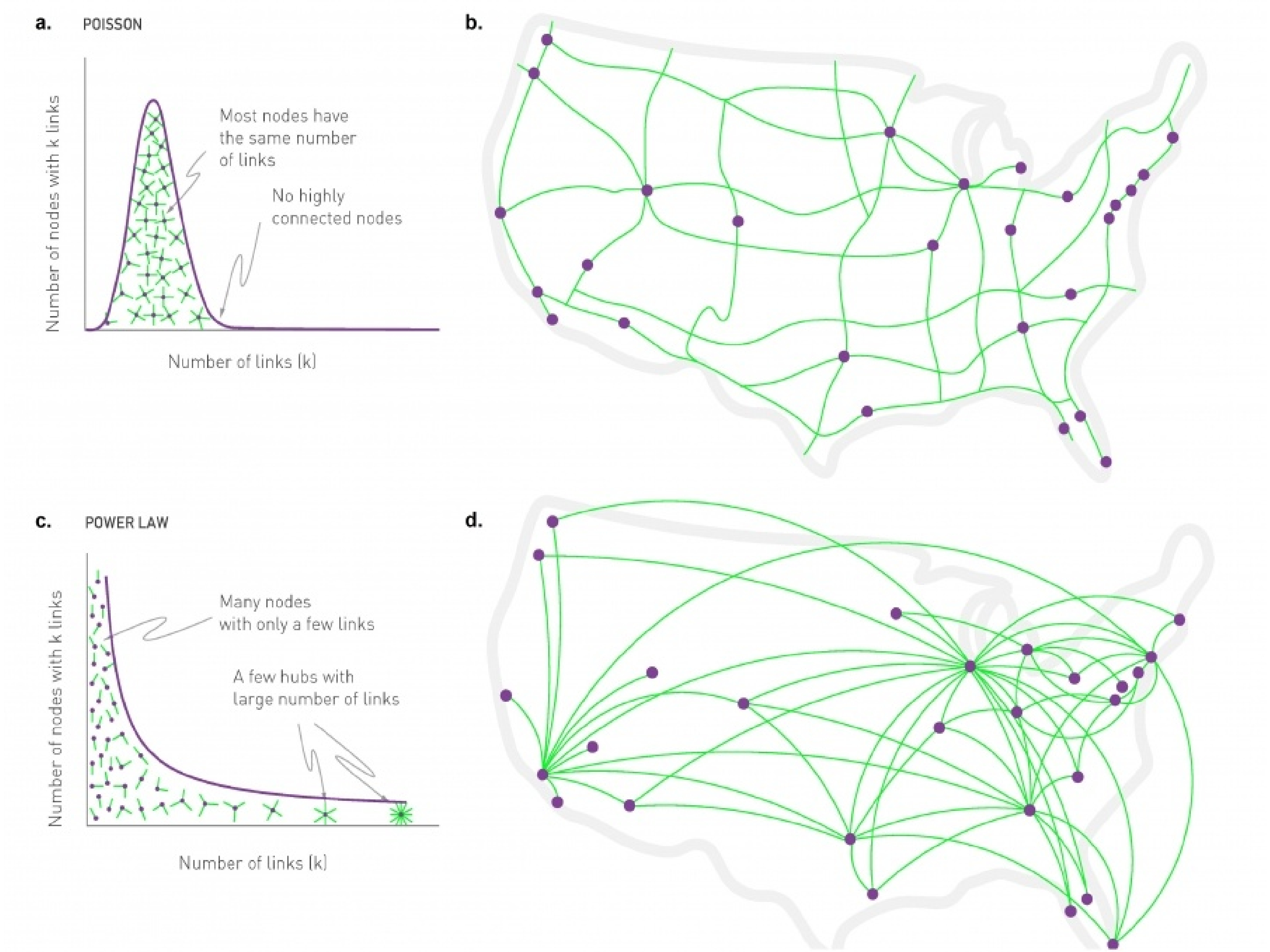
The natural cut-off of the degree

The largest hub cannot have more than N-1 connections



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

Random vs scale-free networks



Ground transportation

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”?

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

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

What does it mean “scale-free”?

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

- In a scale-free network

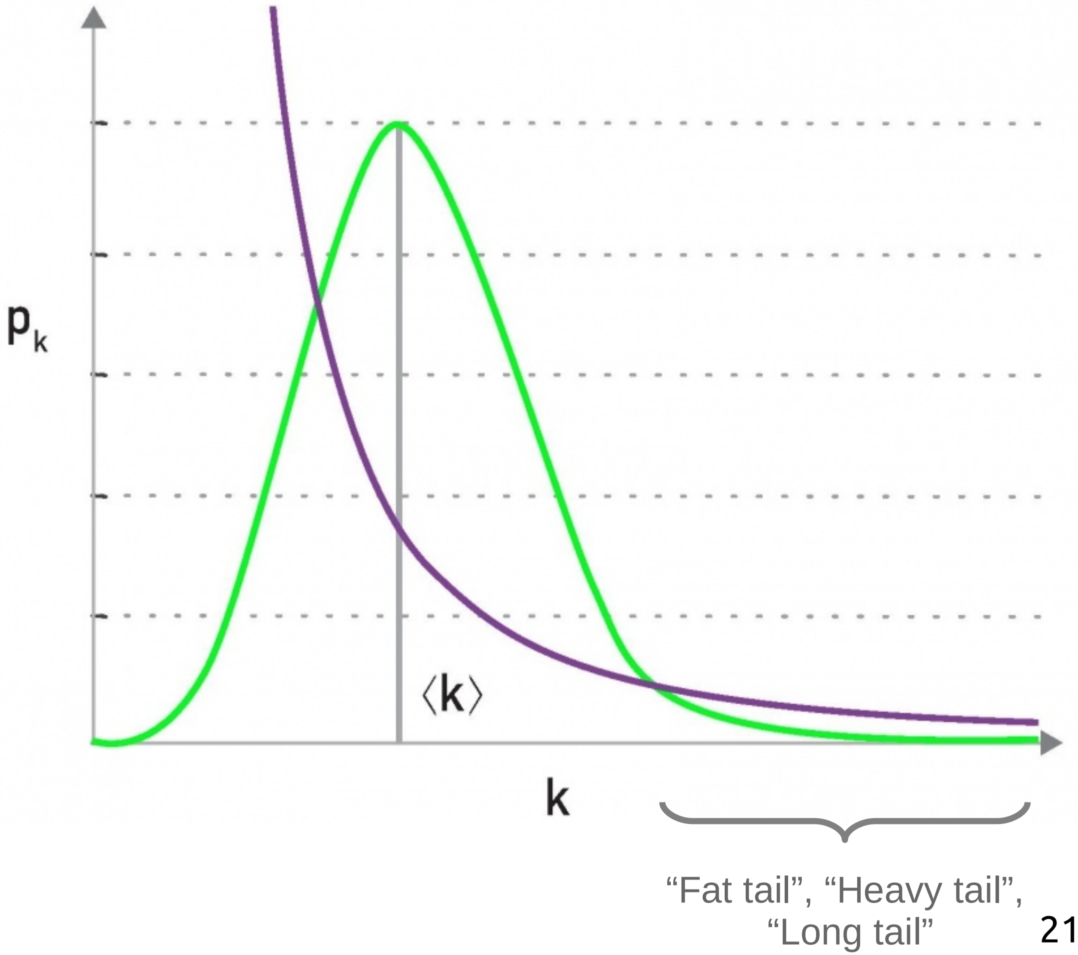
$$\langle k^2 \rangle = \int_{k_{\min}}^{k_{\max}} k^2 p_k dk = C \frac{k_{\max}^{3-\gamma} - k_{\min}^{3-\gamma}}{3-\gamma}$$

- 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	$\langle k \rangle$	$\langle k_{in}^2 \rangle$	$\langle k_{out}^2 \rangle$	$\langle k^2 \rangle$	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*-

Exercise

Answer in Nearpod Collaborate
https://nearpod.com/student/
Code to be given during class

In the actor network, N=702,388, γ=2.12

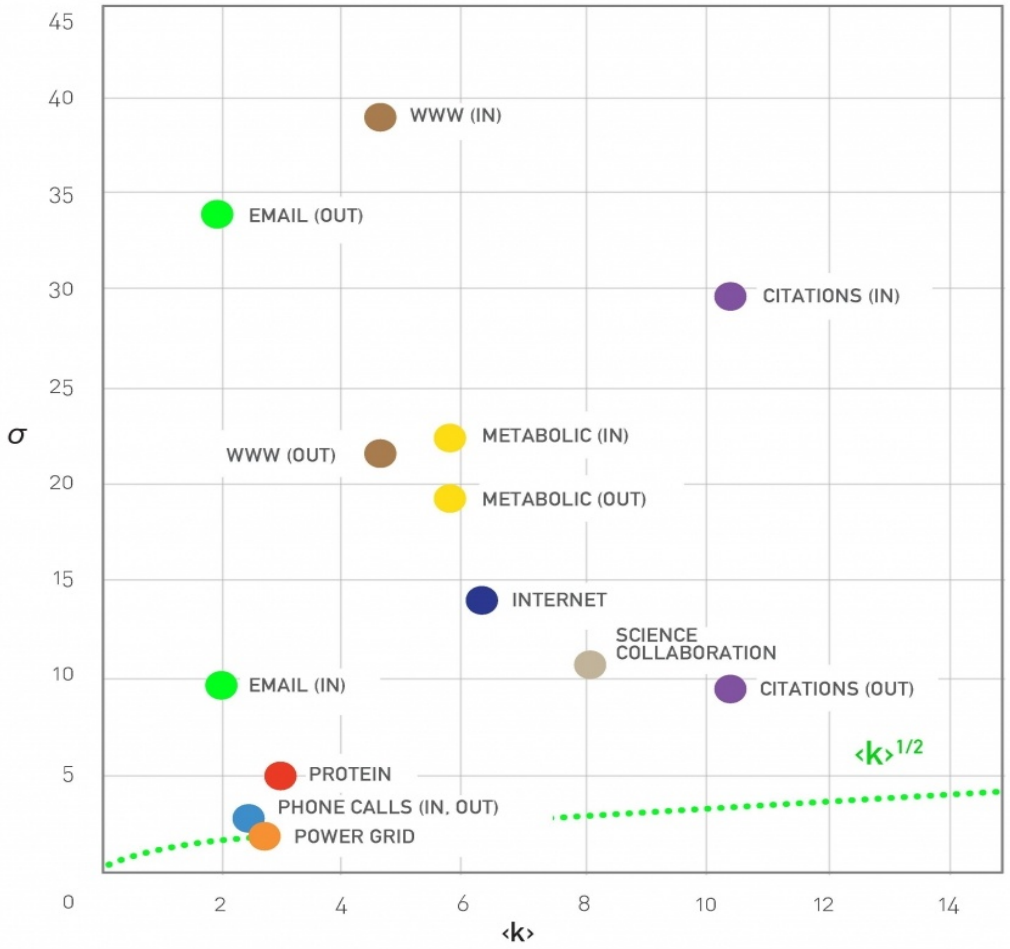
1. How many actors do we expect to have …
   - 1 other co-star?
   - 10 other co-stars?
   - 100 other co-stars?

https://www.wolframalpha.com/ recognizes `x*y, x/y, Zeta(x), x^(-y),` etc.

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

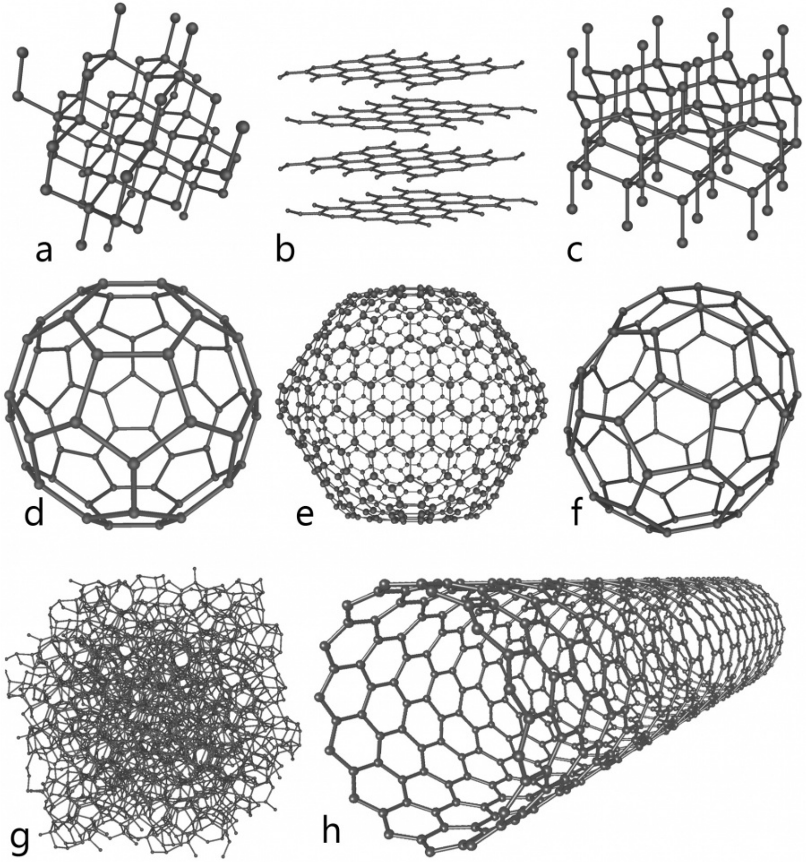
2. For how many co-stars do we still expect to have one actor that has that many co-stars?

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



Summary

Things to remember

- Definition of scale-free
- Power law
- Formulas for degree distribution
  - Discrete formula
  - Continuous formula
- Formula for k_{max}

Practice on your own

- **(Somewhat) difficult, try to solve it ON YOUR OWN**
- Imagine a connected scale-free graph with 1 million nodes and average degree 5

  If we draw 100 nodes from this graph, how many will have degree 1?

  Remember, if the graph is connected, $k_{min}=1$

  If you cannot clear the unknown in a formula, plot it
- Solution in next slide (shown only in .odp, not .pdf)