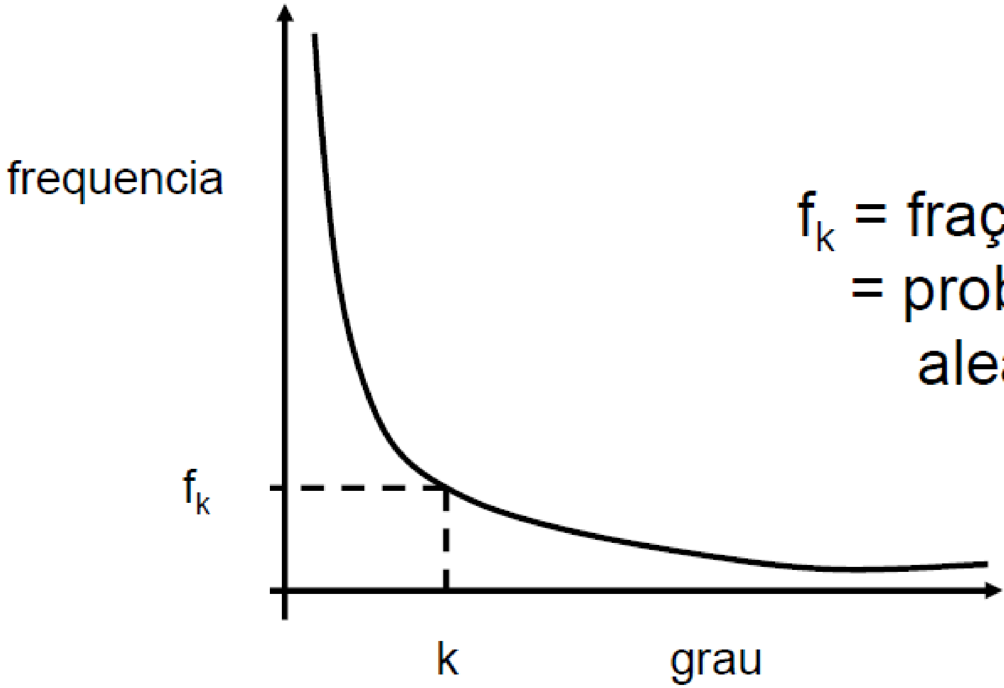
Power Laws and Rich-Get-Richer Phenomena

Ana Paula
Fabrício Benevenuto

Distribuição dos graus



f_k = fração de nós com grau k
= probabilidade de nó selecionado aleatoriamente ter grau k

- Problema: determine a distribuição de probabilidade que melhor ajuste ("best-fit") a um conjunto de dados experimentais de uma rede. Proponha uma coleta de dados e faça o fitting!

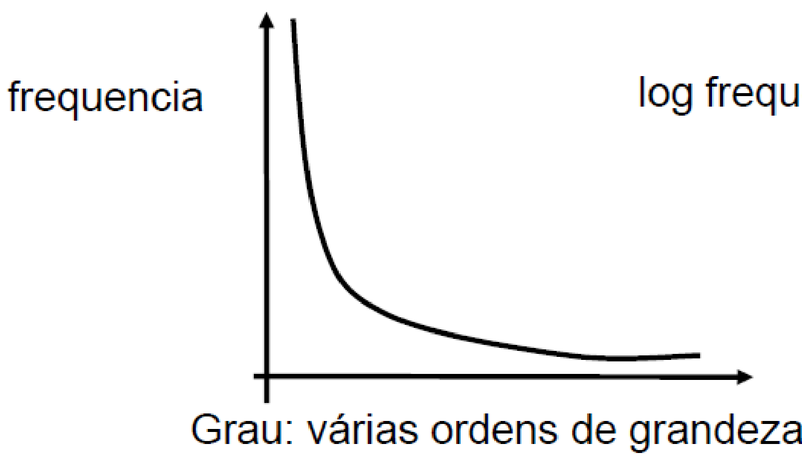
Distribuições Power Law

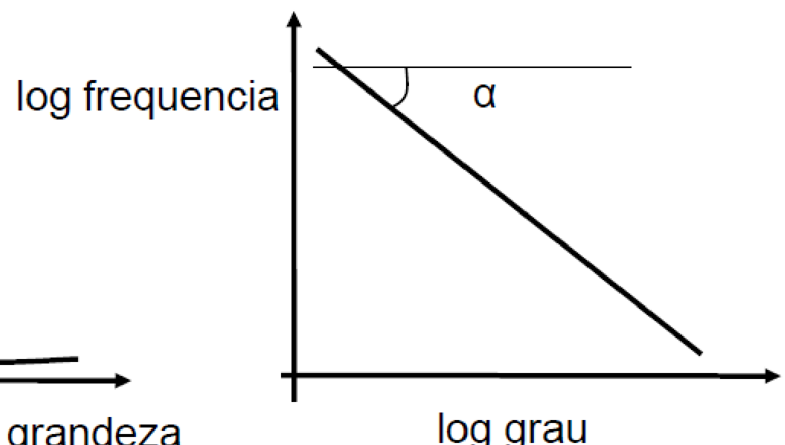
- A distribuição do grau k de uma rede real muitas vezes segue uma power law.

$$p(k) = Ck^{-\alpha}$$

  - Distribuição Power-law leva a uma linha reta num gráfico log-log.

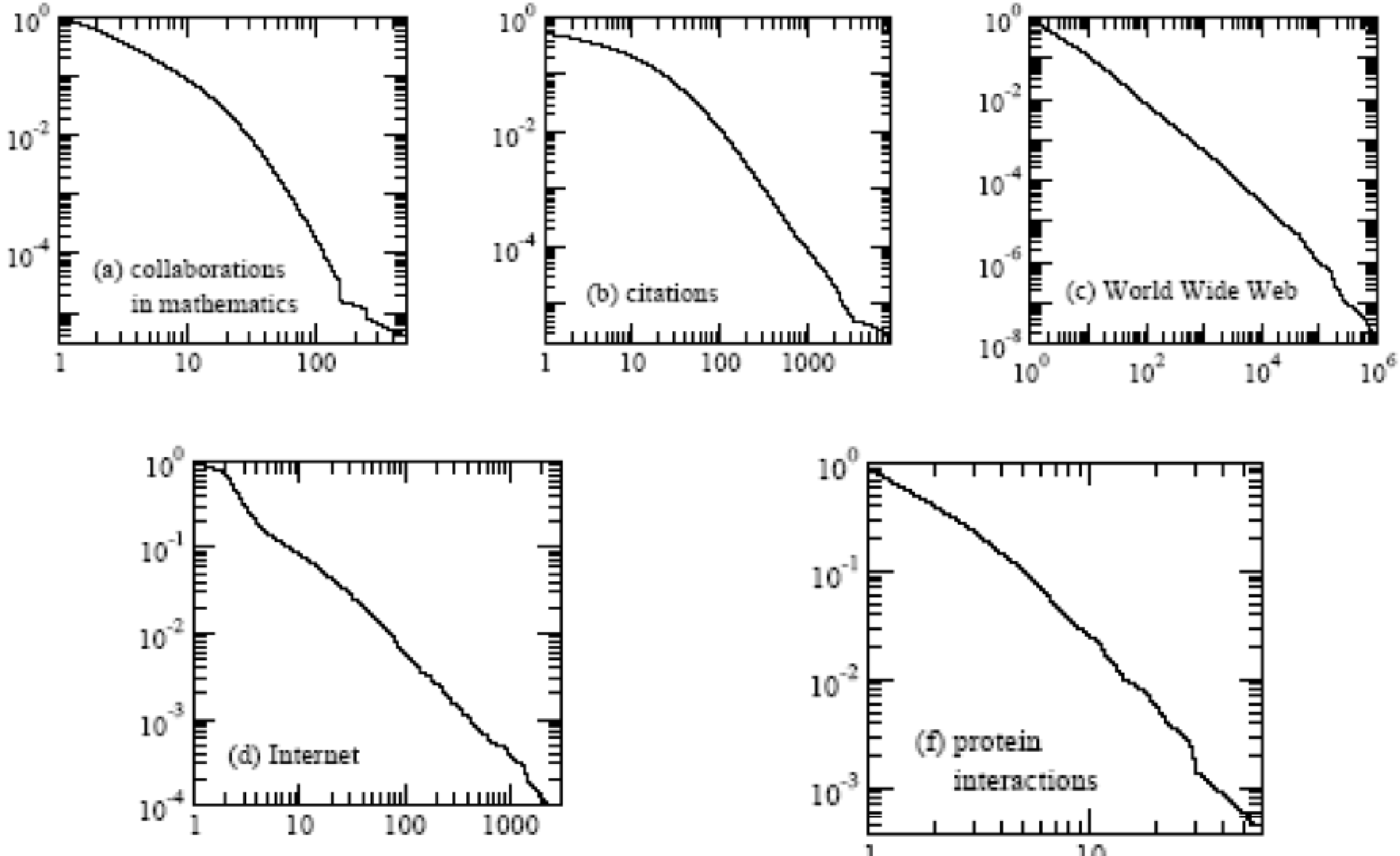
$$\log p(k) = -\alpha \log k + \log C$$





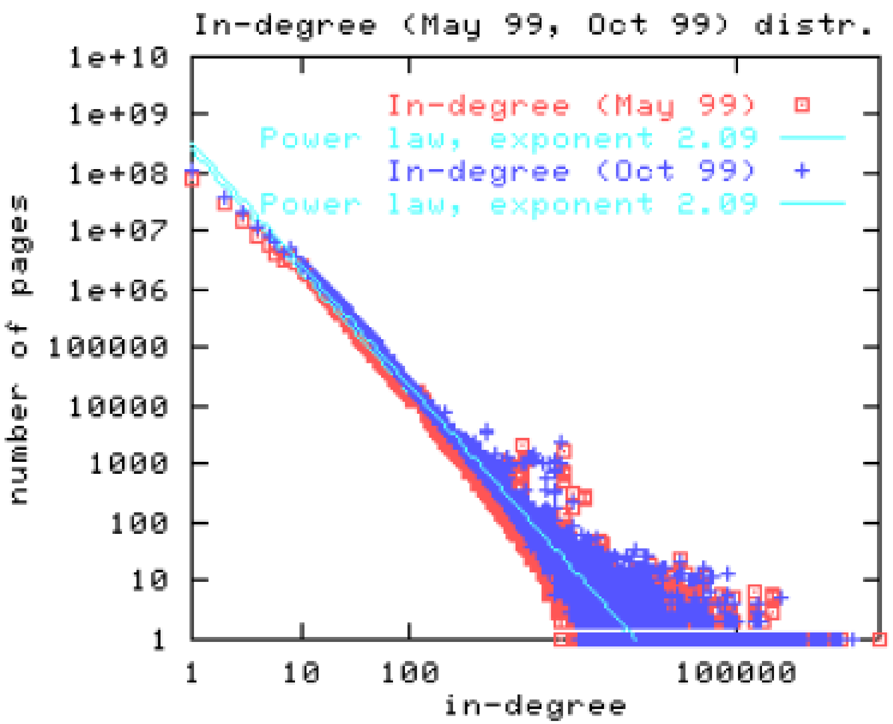
  - α : expoente power-law (tipicamente $2 \leq \alpha \leq 3$)

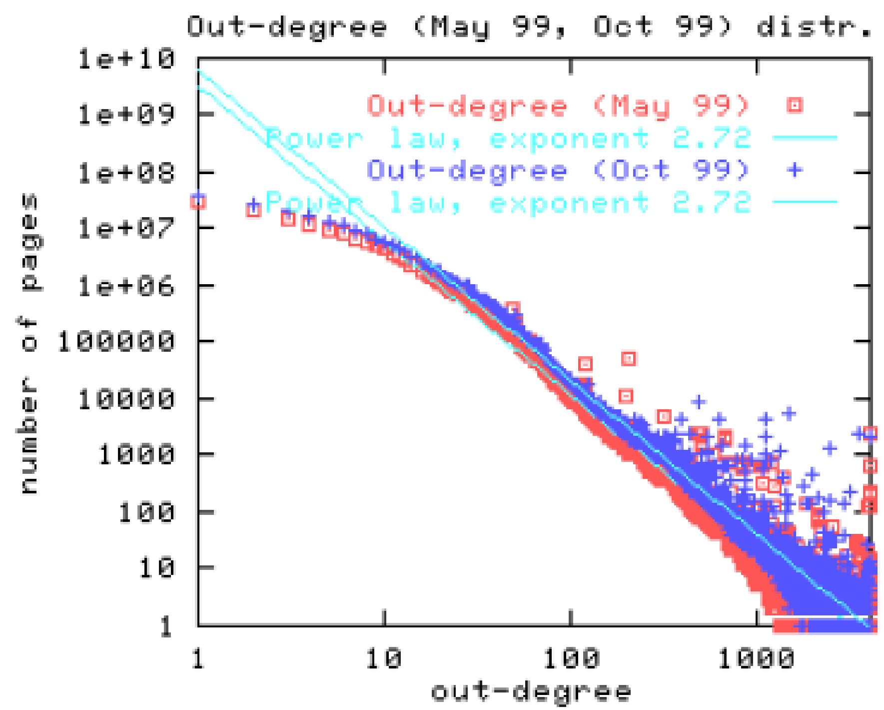
Exemplos



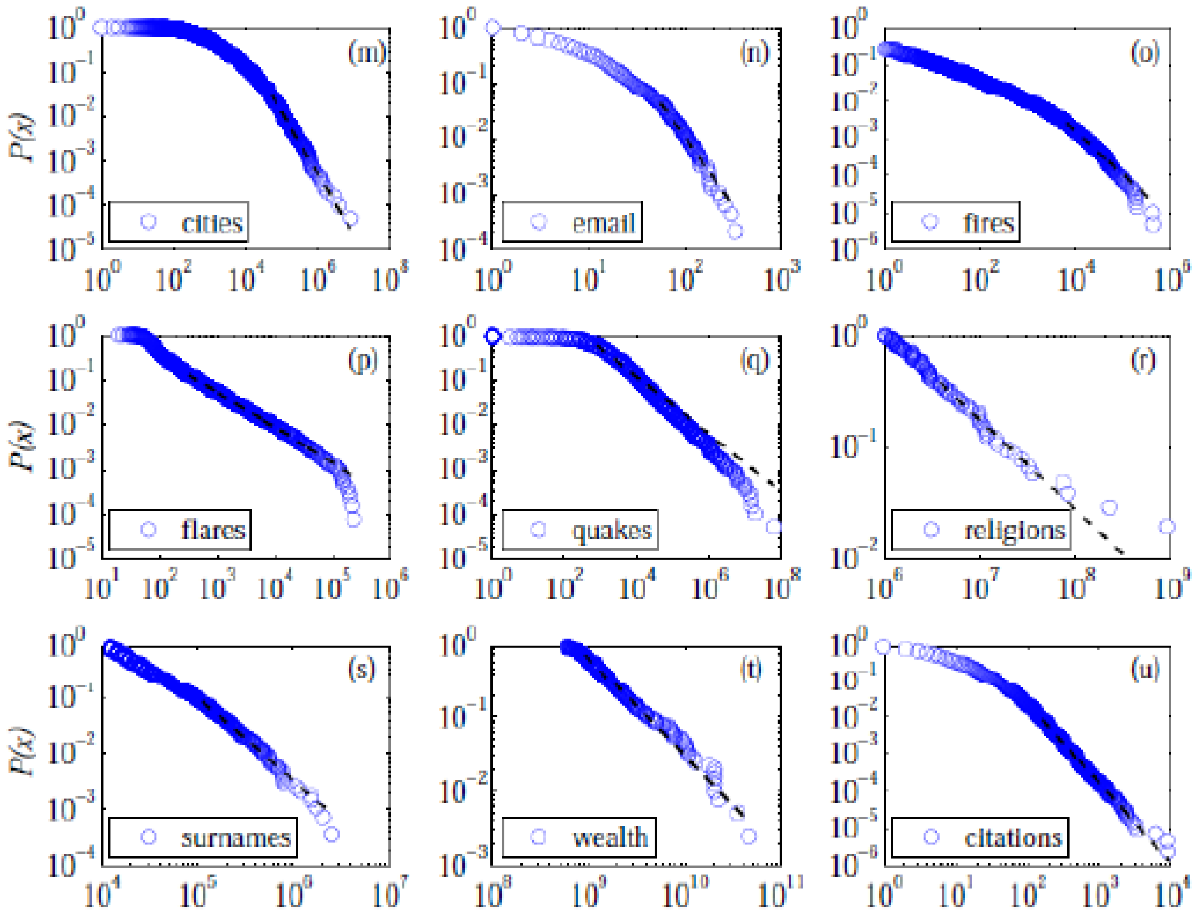
[Newman 2003]

Degree distribution on the Web





Power-laws are everywhere

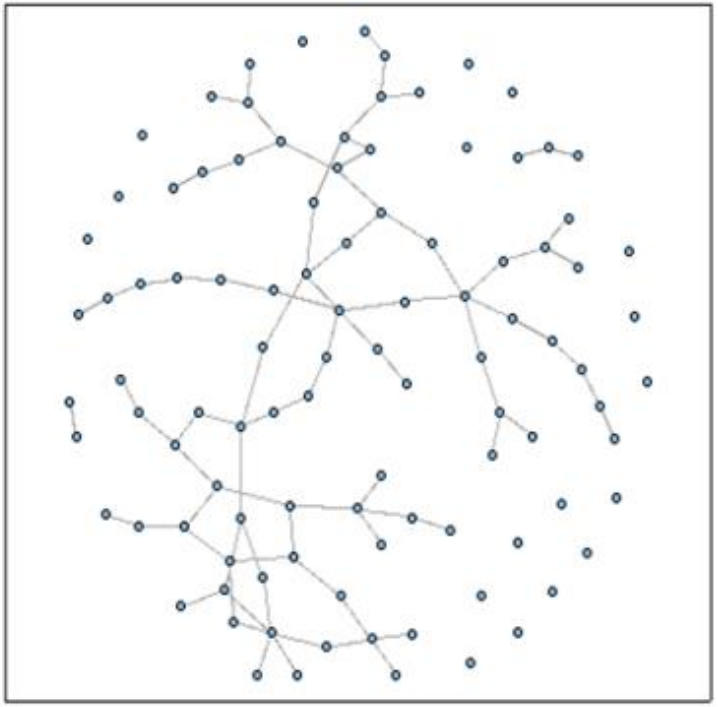


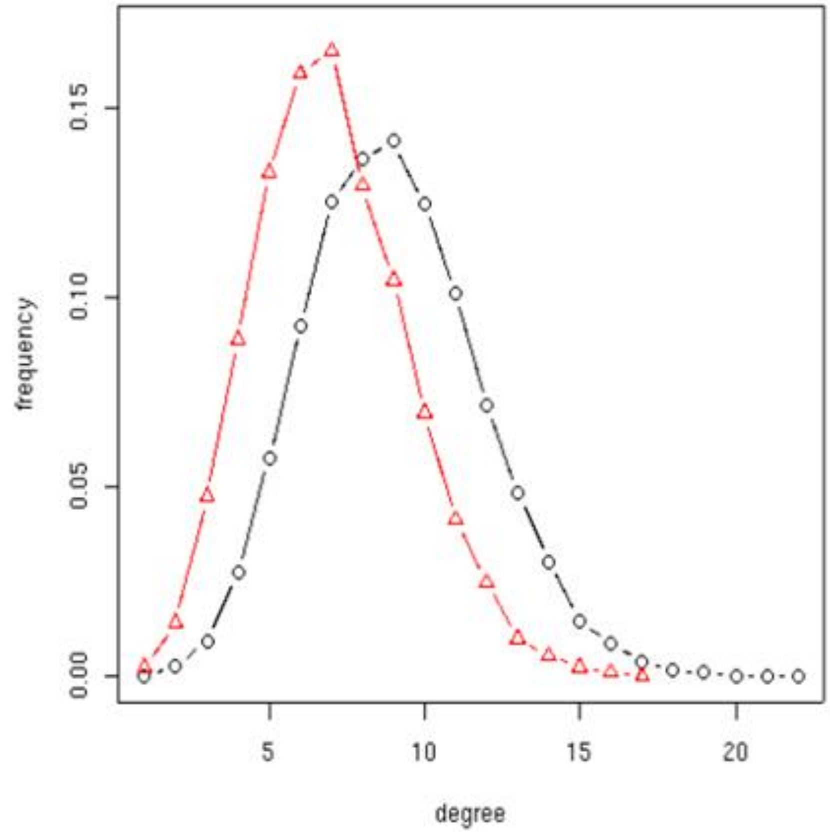
Many other quantities follow heavy-tailed distributions

Not everyone likes power-laws ☺

CMU students protesting at the G20 meeting in Pittsburgh in Sept 2009

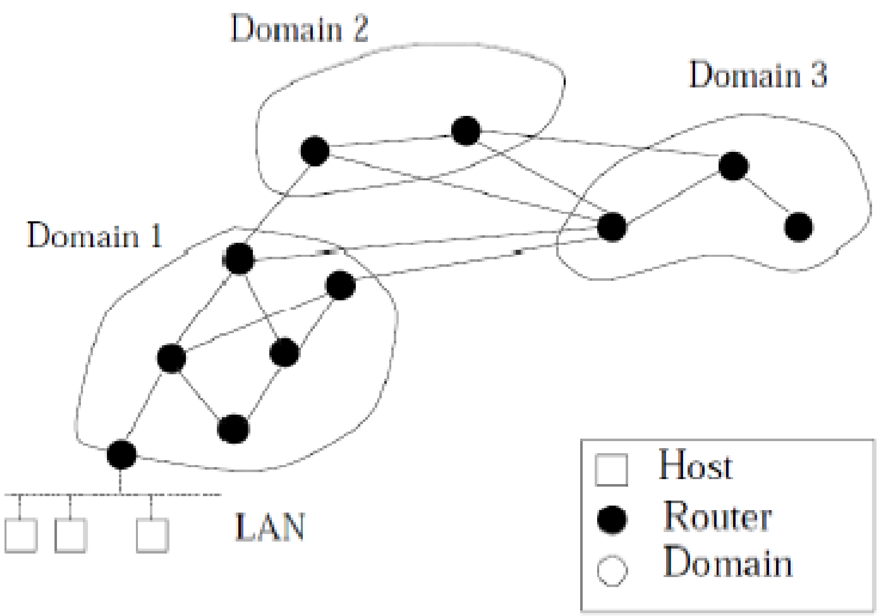
Rede Aleatória ---------------------------- Distribuição de Gráus

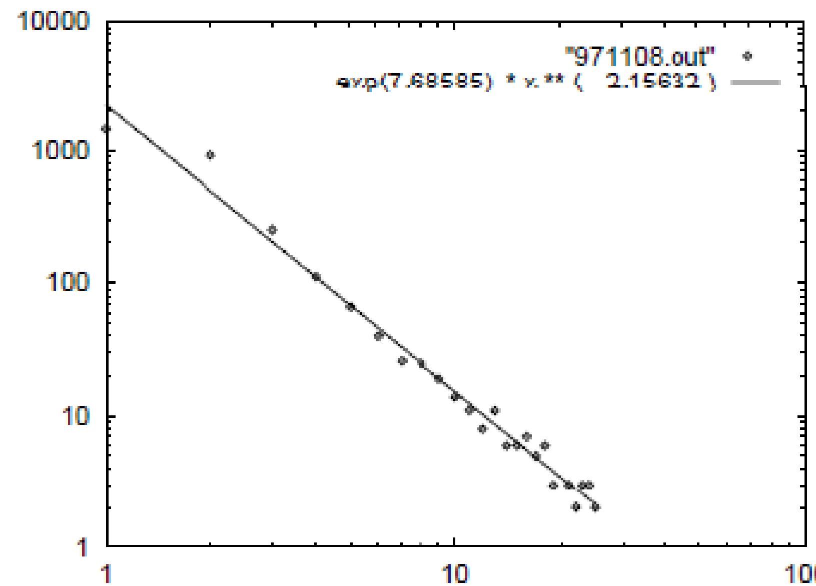




Faloutsos³

- [Faloutsos, Faloutsos and Faloutsos, 1999]

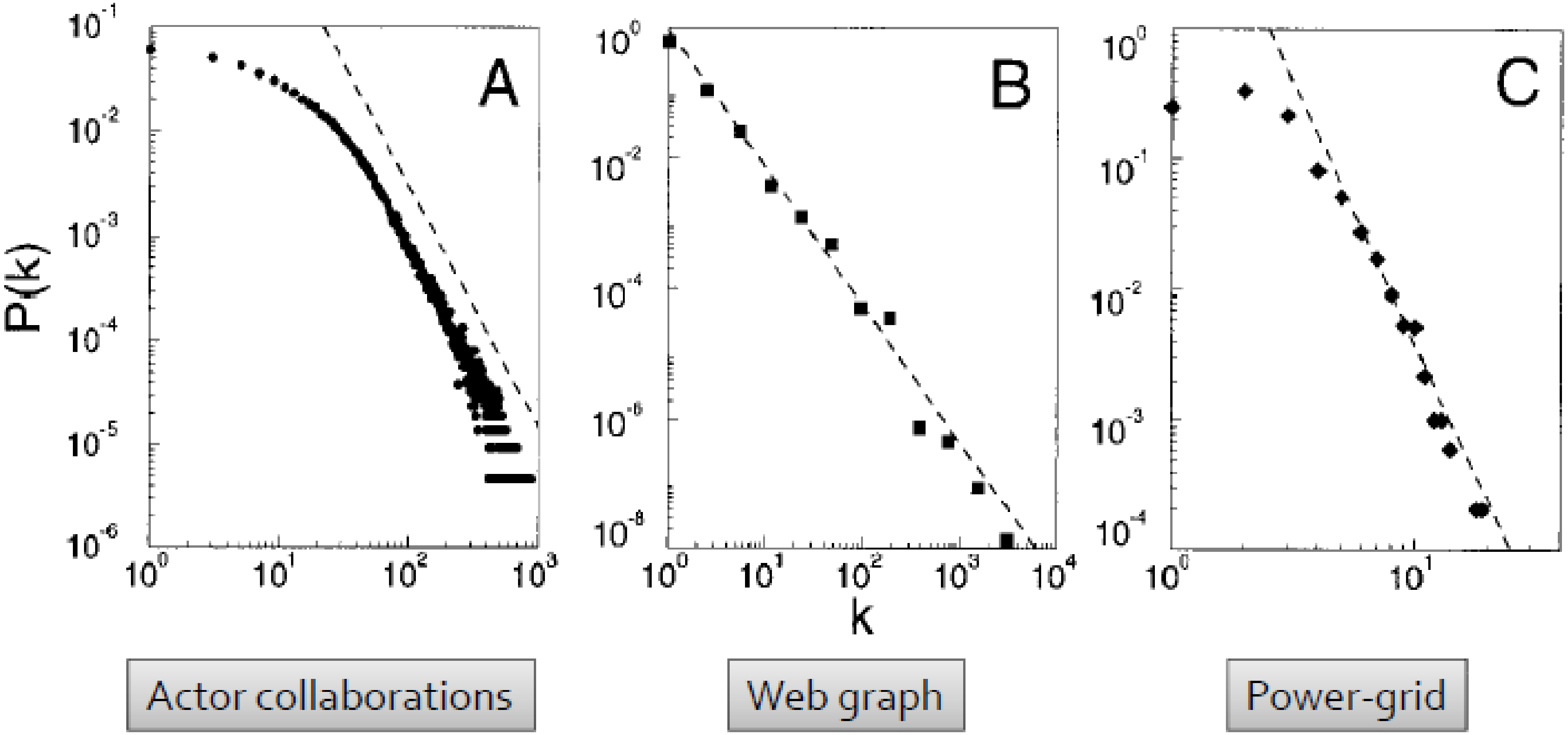




(a) Int-11-97

Barabasi&Albert

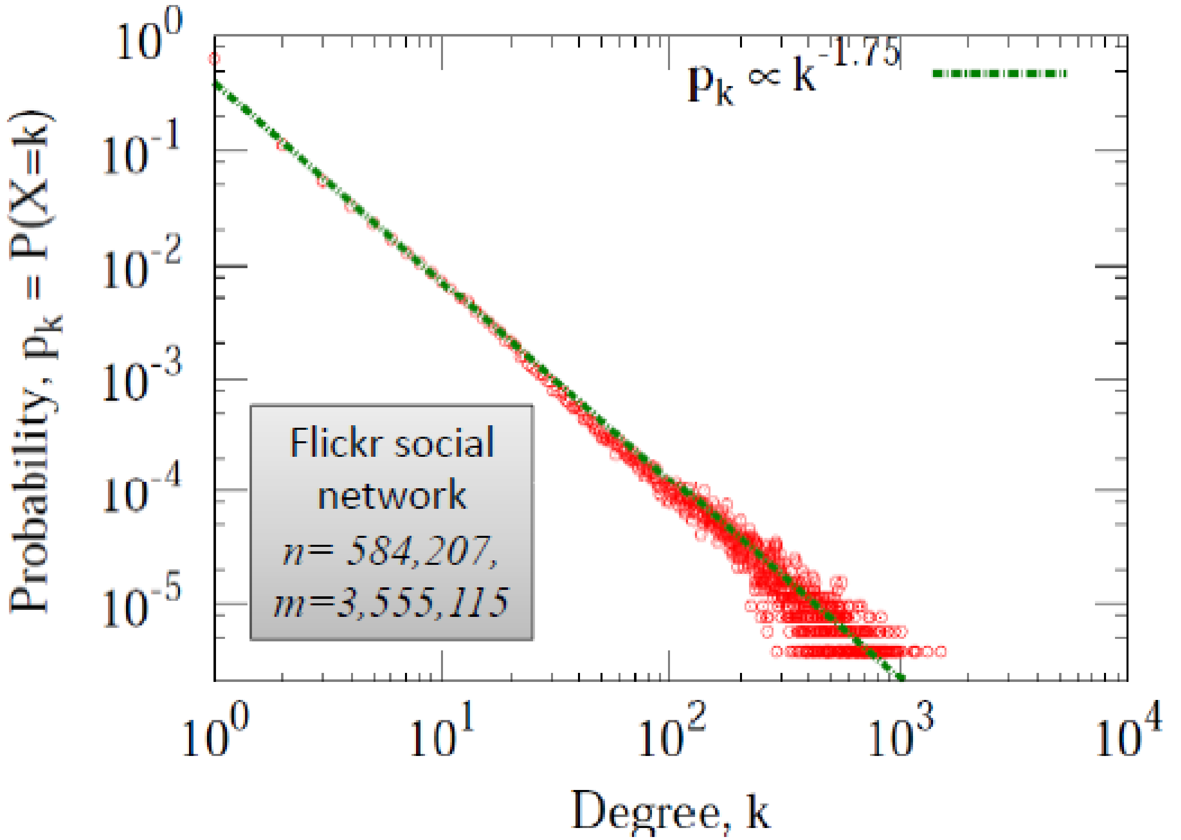
- [Barabasi-Albert, 1999]



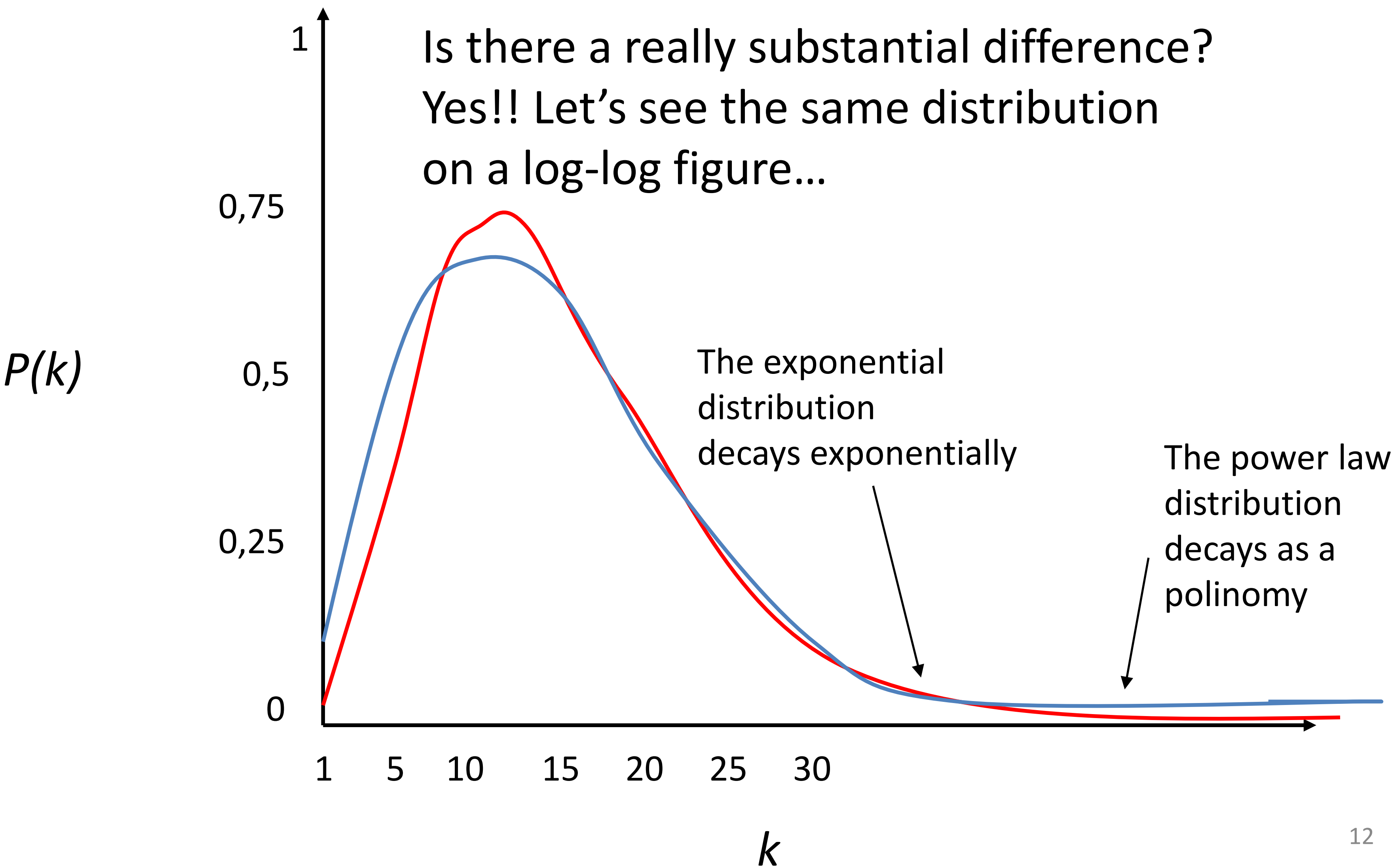
[Leskovec et al. KDD '08]

Degrees in real networks (2)

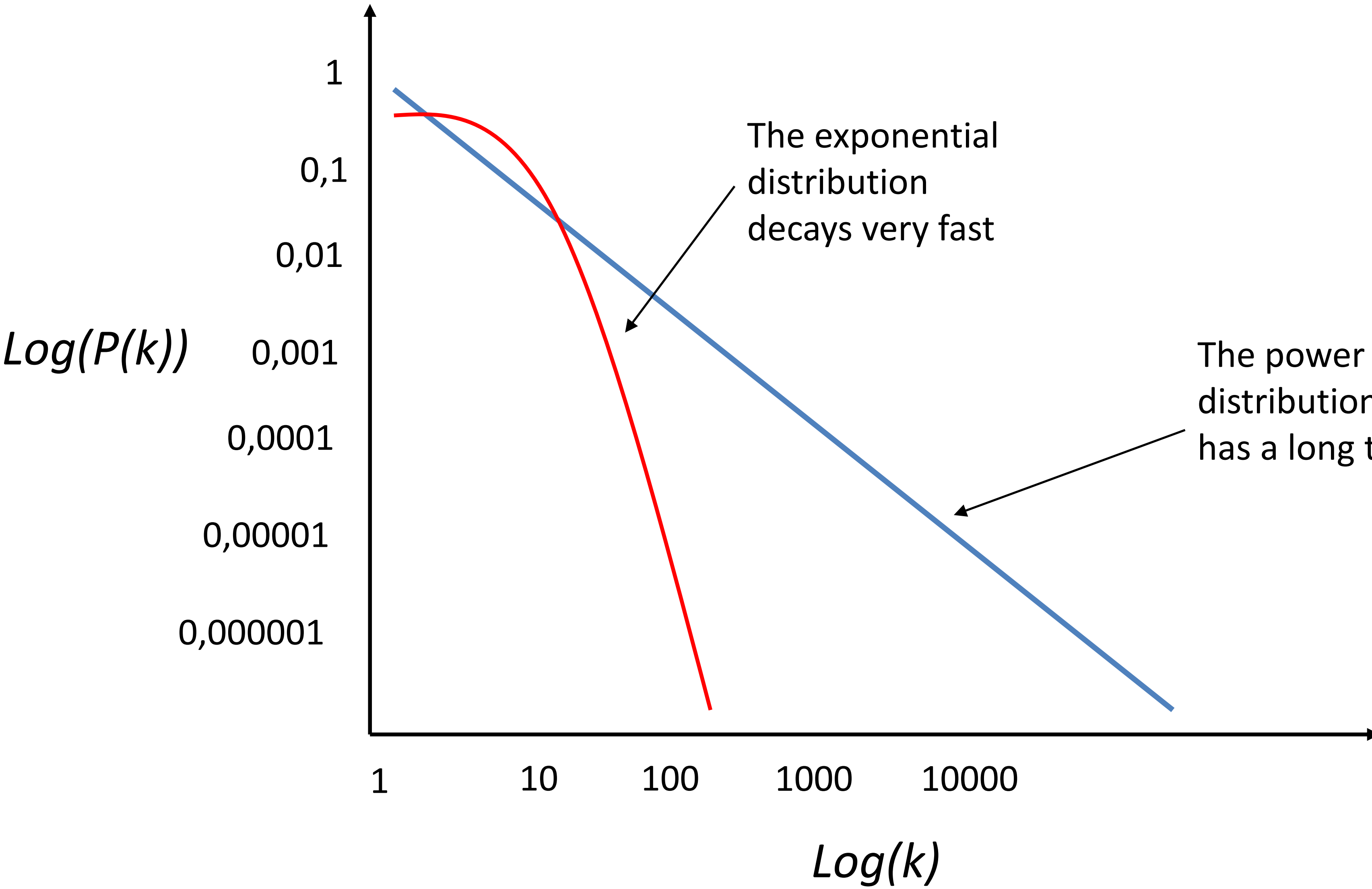
- Plot the same data on *log-log* axis:



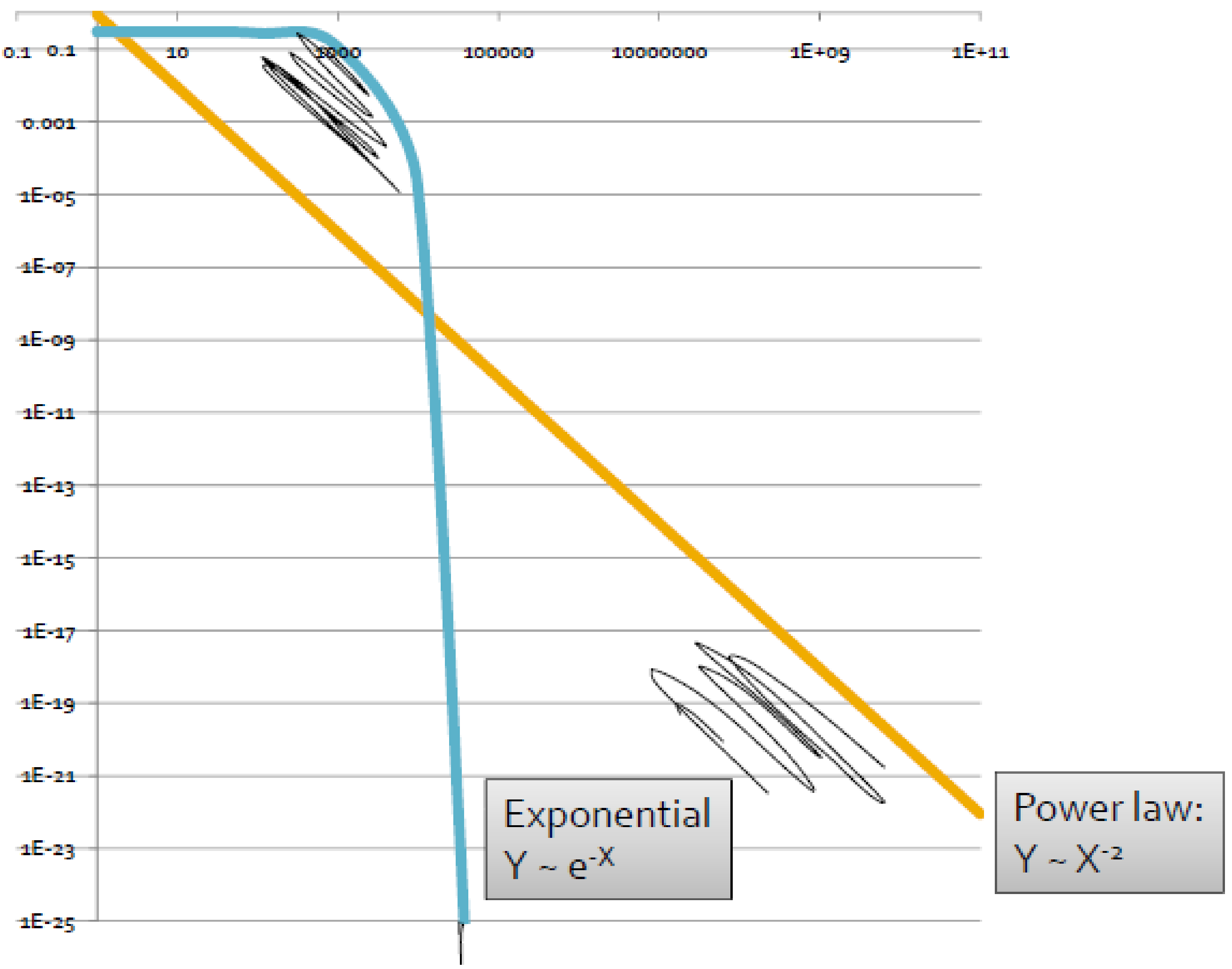
Power vs. Exponential Distribution



Power vs. Exponential Distribution



Exponential tail vs. Power-law tail

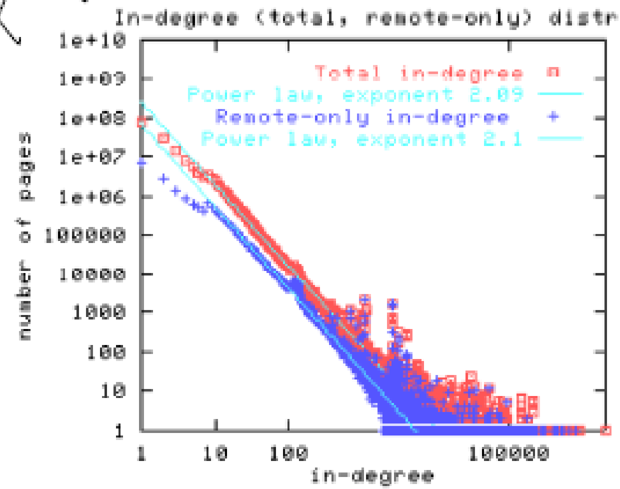


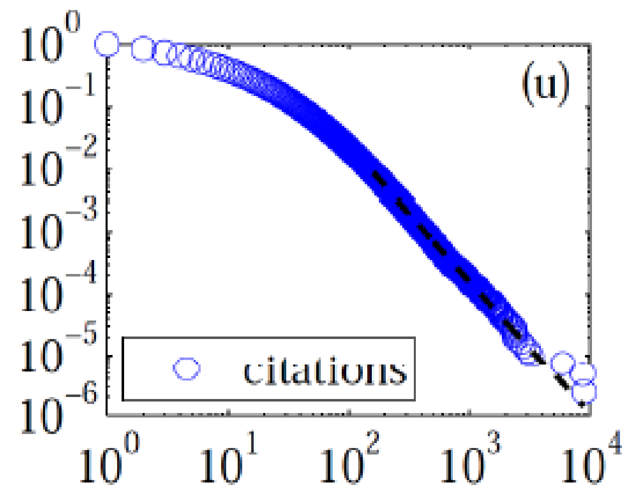
The Heavy Tail

- The power law implies that
  - The probability to have elements very far from the average is not neglectable
  - The big number counts, the tail of the distribution counts
  - On the other hand, the “area” of “big *ks*” in an exponential distribution tend to zero with $k \rightarrow \infty$

Power law degree exponents

- Power law degree exponent is $y \propto x^{-\alpha}$ typically $2 < \alpha < 3$
  - Web graph [Broder et al. 00]:
    - $\alpha_{in} = 2.1$, $\alpha_{out} = 2.4$
  - Autonomous systems [Faloutsos et al. 99]:
    - $\alpha = 2.4$
  - Actor collaborations [Barabasi-Albert 00]:
    - $\alpha = 2.3$
  - Citations to papers [Redner 98]:
    - $\alpha \approx 3$
  - Online social networks [Leskovec et al. 07]:
    - $\alpha \approx 2$





The Power Law in Real Networks

Average k

Power law exponents

Network	Size	$\langle k \rangle$	κ	γ_{out}	γ_{in}	ℓ_{real}	ℓ_{rand}	ℓ_{pow}	Reference
WWW	325 729	4.51	900	2.45	2.1	11.2	8.32	4.77	Albert, Jeong, and Barabási 1999
WWW	4×10^7	7		2.38	2.1				Kumar *et al.*, 1999
WWW	2×10^8	7.5	4000	2.72	2.1	16	8.85	7.61	Broder *et al.*, 2000
WWW, site	260 000				1.94				Huberman and Adamic, 2000
Internet, domain*	3015–4389	3.42–3.76	30–40	2.1–2.2	2.1–2.2	4	6.3	5.2	Faloutsos, 1999
Internet, router*	3888	2.57	30	2.48	2.48	12.15	8.75	7.67	Faloutsos, 1999
Internet, router*	150 000	2.66	60	2.4	2.4	11	12.8	7.47	Govindan, 2000
Movie actors*	212 250	28.78	900	2.3	2.3	4.54	3.65	4.01	Barabási and Albert, 1999
Co-authors, SPIRES*	56 627	173	1100	1.2	1.2	4	2.12	1.95	Newman, 2001b
Co-authors, neuro.*	209 293	11.54	400	2.1	2.1	6	5.01	3.86	Barabási *et al.*, 2001
Co-authors, math.*	70 975	3.9	120	2.5	2.5	9.5	8.2	6.53	Barabási *et al.*, 2001
Sexual contacts*	2810			3.4	3.4				Liljeros *et al.*, 2001
Metabolic, *E. coli*	778	7.4	110	2.2	2.2	3.2	3.32	2.89	Jeong *et al.*, 2000
Protein, *S. cerev.**	1870	2.39		2.4	2.4				Jeong, Mason, *et al.*, 2001
Ythan estuary*	134	8.7	35	1.05	1.05	2.43	2.26	1.71	Montoya and Solé, 2000
Silwood Park*	154	4.75	27	1.13	1.13	3.4	3.23	2	Montoya and Solé, 2000
Citation	783 339	8.57			3				Redner, 1998
Phone call	53×10^6	3.16		2.1	2.1				Aiello *et al.*, 2000
Words, co-occurrence*	460 902	70.13		2.7	2.7				Ferrer i Cancho and Solé, 2001
Words, synonyms*	22 311	13.48		2.8	2.8				Yook *et al.*, 2001b

The 20-80 Rule

- Examples
  - The 20% of the Web sites gests the 80% of the visits
  - The 20% of the Internet routers handles the 80% of the total Internet traffic
  - The 20% of world industries hold the 80% of the world’s income
  - The 20% of the world population consumes the 80% of the world’s resources
  - The 20% of the earthquakes caused the 80% of the victims
  - The 20% of the rivers in the world carry the 80% of the total sweet water
  - The 20% of the proteins handles the 80% of the most critical metabolic processes

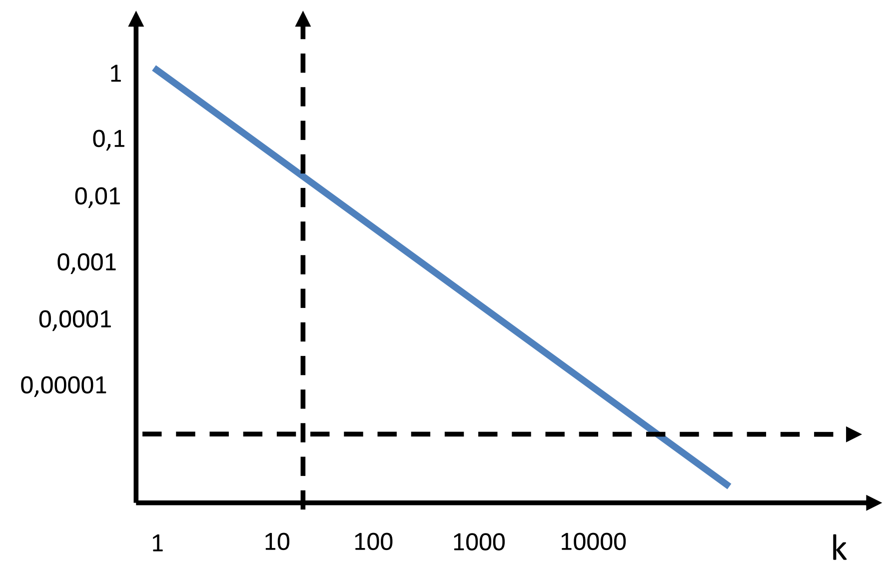
- Does this derive from the power law distribution? YES!

Hubs and Connectors

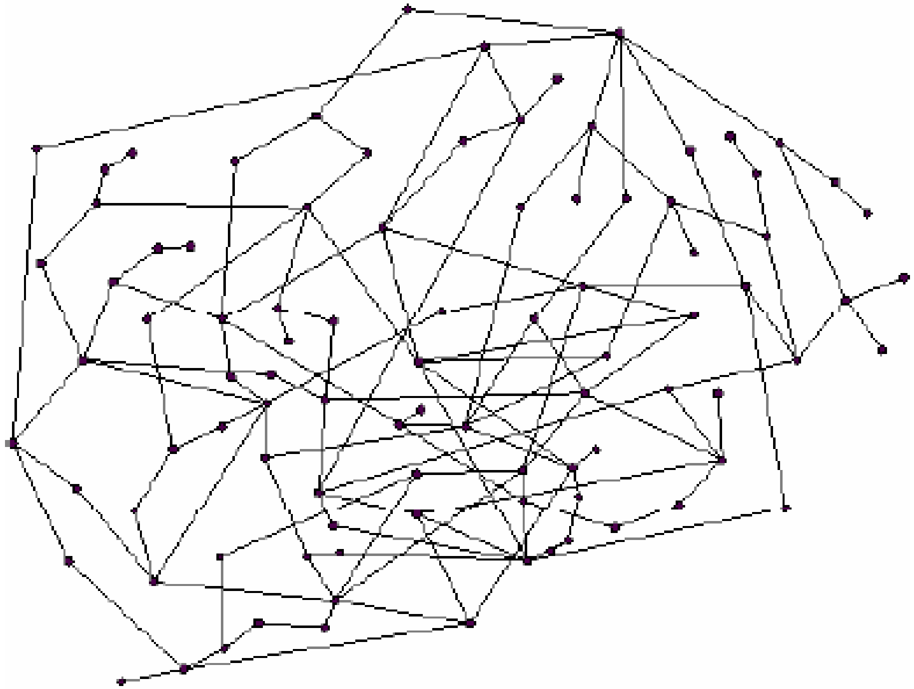
- Scale free networks exhibit the presence of nodes that
  - Act as hubs, i.e., as point to which most of the other nodes connects to
  - Act as connectors, i.e., nodes that make a great contributions in getting great portion of the network together
  - “smaller nodes” exists that act as hubs or connectors for local portion of the network
- This may have notable implications, as detailed below

Why “Scale-Free” Networks

- Why networks following a power law distribution for links are called “**scale free**”?
  - Whatever the scale at which we observe the network
  - The network looks the same, i.e., it looks similar to itself
- The overall properties of the network are preserved independently of the scale
- In particular:
  - If we cut off the details of a network – skipping all nodes with a limited number of links – the network will preserve its power-law structure
  - If we consider a sub-portion of any network, it will have the same overall structure of the whole network

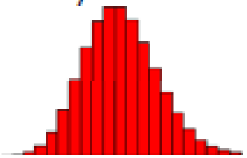


Random vs. Scale-free network

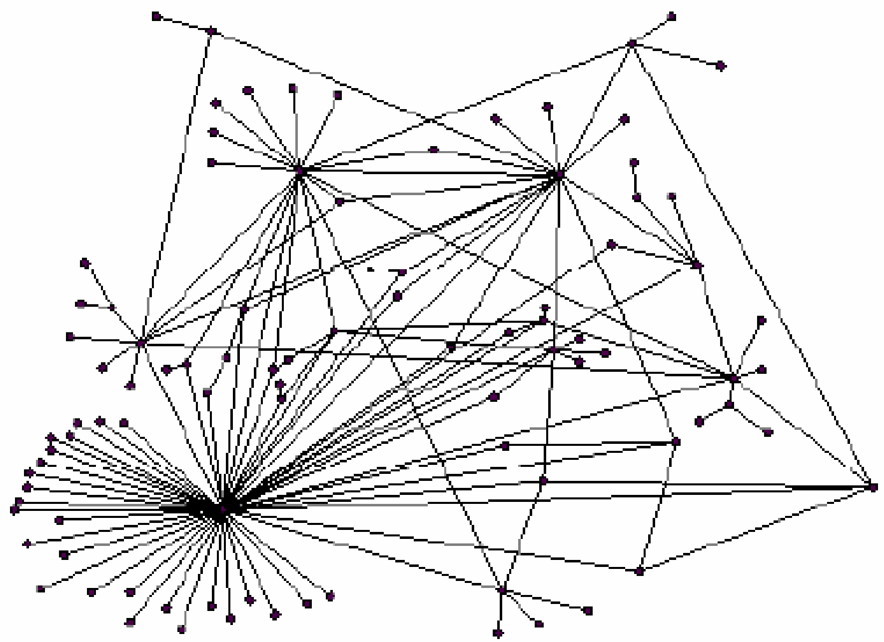


Random network

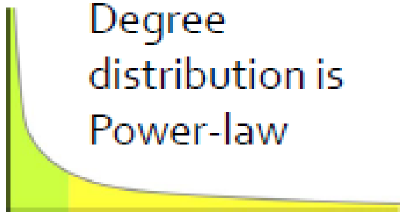
(Erdos-Renyi random graph)



Degree distribution is Binomial



Scale-free (power-law) network



Degree distribution is Power-law

Function is scale free if:
$f(ax) = c\, f(x)$

Evolving Networks

- More in general…
  - Networks grows AND
  - Network evolves
- The evolution may be driven by various forces
  - Connection age
  - Connection satisfaction
- What matters is that connections can change during the life of the network
  - Not necessarily in a random way
  - But following characteristics of the network…

- Qual seria um modelo que leva a uma rede power law?

Model: Preferential attachment

- Preferential attachment [Price 1965, Albert-Barabasi 1999]:
  - Nodes arrive in order
  - A new node j creates m out-links
  - Prob. of linking to a previous node i is proportional to its degree d_i

$$P(j \to i) = \frac{d_i}{\sum_k d_k}$$

The Preferential Attachment Algorithm

- Start with a limited number of initial nodes
- At each time step, add a new node that has *m* edges that link to m existing nodes in the system
- When choosing the nodes to which to attach, assume a probability P for a node i proportional to the number k_i of links already attached to it
- **It can be shown that this leads to a power law network!**

$$P(k_i) = \frac{k_i}{\sum_j k_j}$$

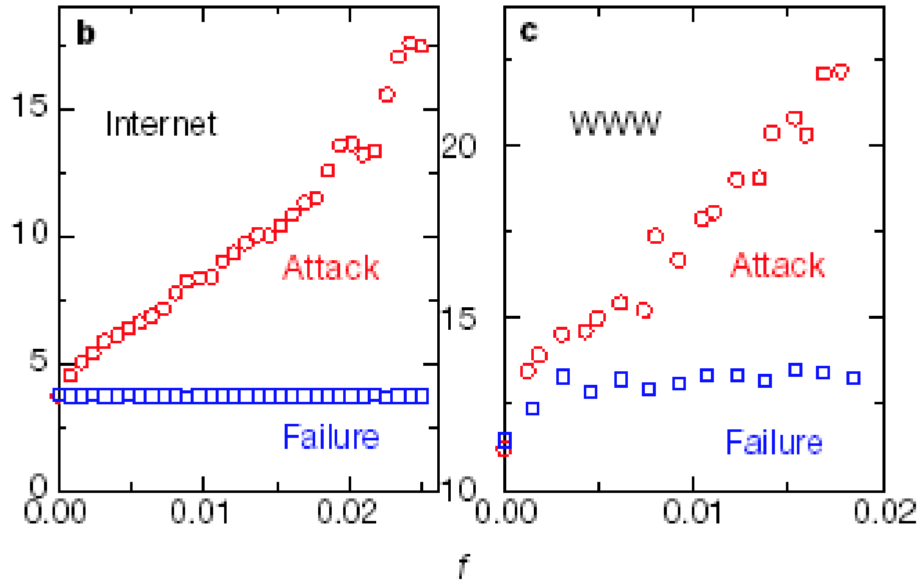
Rich-get-richer

- New nodes are more likely to link to nodes that already have high degree
- Herbert Simon’s result
  - Power-laws arise from “Rich get richer” (cumulative advantage)
- Examples [Price 65]:
  - Citations: new citations of a paper are proportional to the number it already has

Error Tolerance

- Scale free networks are very robust to errors
  - If nodes randomly “break” of disconnect to the network
  - The structure of the network, with high probability, will not be significantly affected by such errors
  - At least only a few small clusters of nodes will disconnect to the network
  - The average path length remains the same

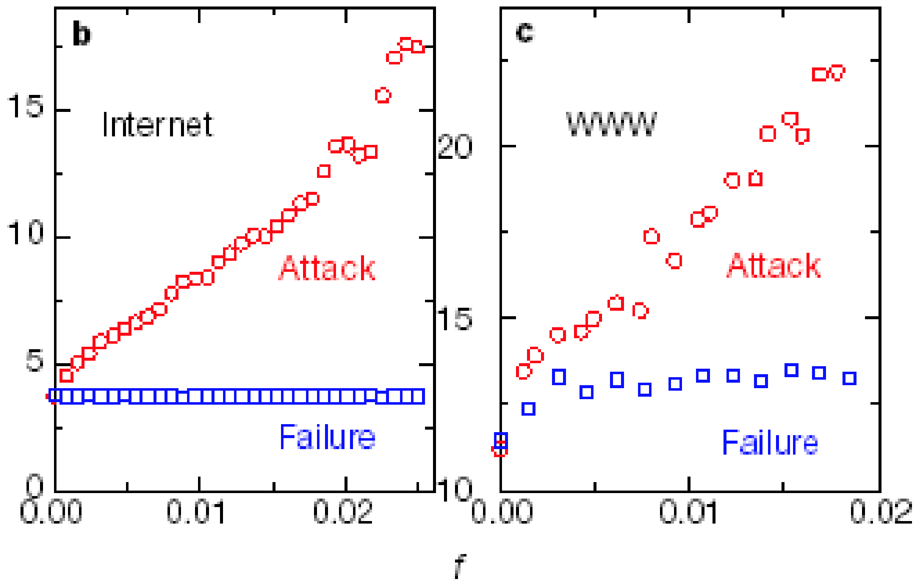
Characteristic Path Lenght



Attack Tolerance

- Scale free networks are very sensitive to targeted attacks
  - If the most connected nodes get deliberately chosen as targets of attacks
  - The average path length of the network grows very soon
  - It is very likely that the network will break soon into disconnected clusters
  - Although these independent clusters still preserves some internal connection

Characteristic
Path Lenght



Implications for Everyday Systems: Scale Free Networks and Terrorism

- The network of terrorism is growing
  - And it is a social network with a scale free structure
- How can we destroy such network?
  - Getting unimportant nodes will not significantly affect the network
  - Getting the right nodes, i.e., the hubs (as Bin Laden) is extremely important
  - But it may be very difficult to identify and get the hubs
  - In any case, even if we get the right nodes, other connected clusters will remains that will likely act in any case
- As far as breaking the information flow among terrorists
  - This is very difficult nowadays

Implications for Everyday Systems: Scale Free Networks and Trends

- Who decide what is in and what is “out” in music, fashion, etc.?
  - How can an industry have its products become “in”?
- Industries spend a lot of money in trying to influence the market
  - A lot of commercial advertising, a lot of “free trials”, etc.
  - Still, many new products fail and never have market success!

- Who is the king? Content or dissemination?

Does Content Determine Information Popularity in Social Media?

A Case Study of YouTube Videos’ Content and their Popularity

Flavio Figueiredo[1], **Jussara M. Almeida**[1], **Fabrício Benevenuto**[1]
[1]Department of Computer Science, UFMG, Brazil
{flaviov,jussara,fabricio}@dcc.ufmg.br

Krishna P. Gummadi[2]
[2]MPI-SWS, Germany
gummadi@mpi-sws.org

Máquinas de Busca e Sistemas de Recomendação

- Google foi tão bom em seu algoritmo, que ele mudou a forma como as pessoas usam a Web
  - Procurou no Google, não achou. Busca de novo
  - Não está na primeira página, busca de novo
  - Buscar na “long tail” requer uma query elaborada

Máquinas de Busca e Sistemas de Recomendação

- Recomendações e links para sites eram muito mais importantes no passado
- Máquinas de busca acabaram com o bookmarking e mudaram a navegação na Web
- Hoje as pessoas recomendam conteúdo pelas redes sociais. Talvez as máquinas de busca despertaram novamente essa necessidade
- O que vem depois?