INTRODUCTION TO NETWORK SCIENCE

János Kertész

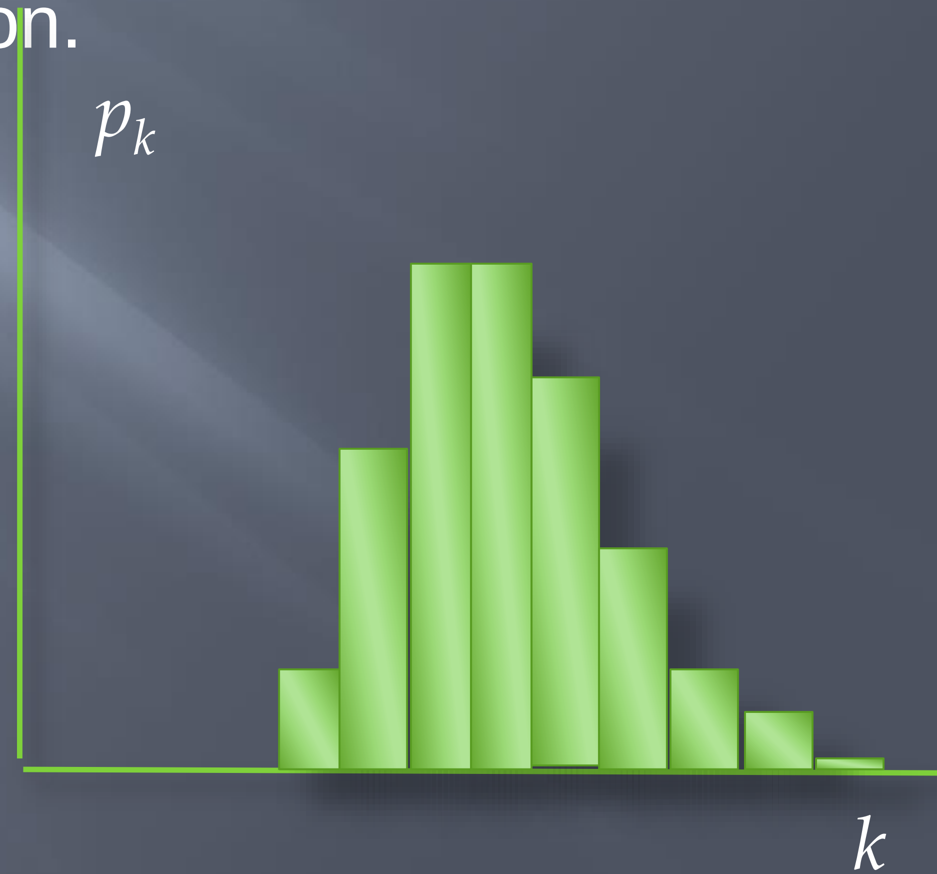
janos.kertesz@gmail.com

5. SCALE FREE NETWORKS AND THE CONFIGURATION MODEL

Small world model (WS)

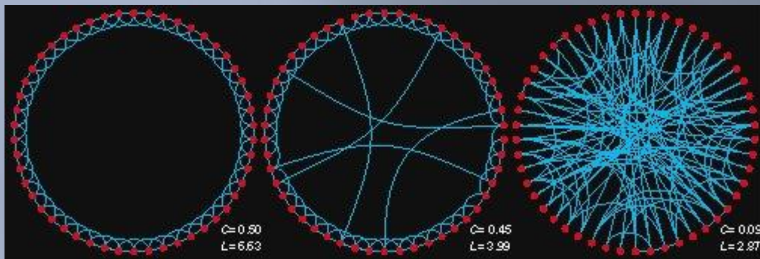
Degree distribution: Added links form an ER NW with prob *p.* If the original lattice has coordination number k_0 we finally get for the distribiution of the total degree k a shifted Poisson distribution.

$$p_k = e^{-\langle k - k_0 \rangle} \frac{\langle k - k_0 \rangle^{k - k_0}}{(k - k_0)!}$$



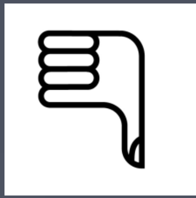
Sharply peaked, shifted Poisson

Small world model (WS)



Summary of the WS model:

- Combines large clustering of some lattices with short average distance due to cross links 
- Reflects some aspects of social networks (communities with high clustering connected by long distance links). 

- It has a sharp degree distribution – in contrast with real world networks 

Inhomogeneities on all scales

Wealth, fame, status etc. is unevenly distributed.
Why?

The Matthew effect:

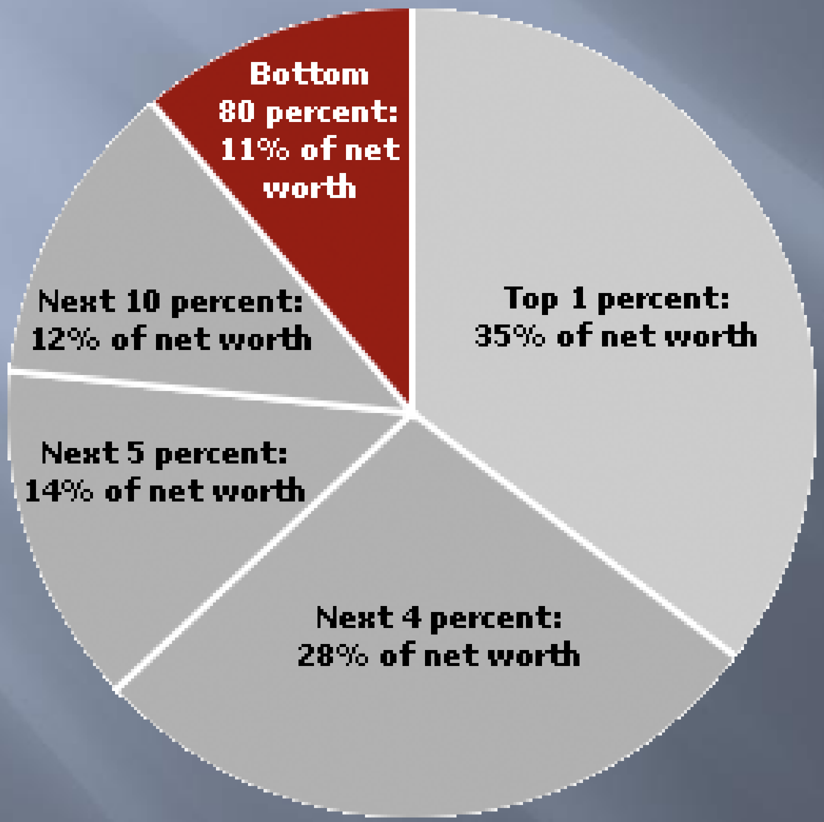
"For unto every one that hath, shall be given, and he shall have abundance: but from him that hath not shall be taken even that which he hath."

Matthew 25:29

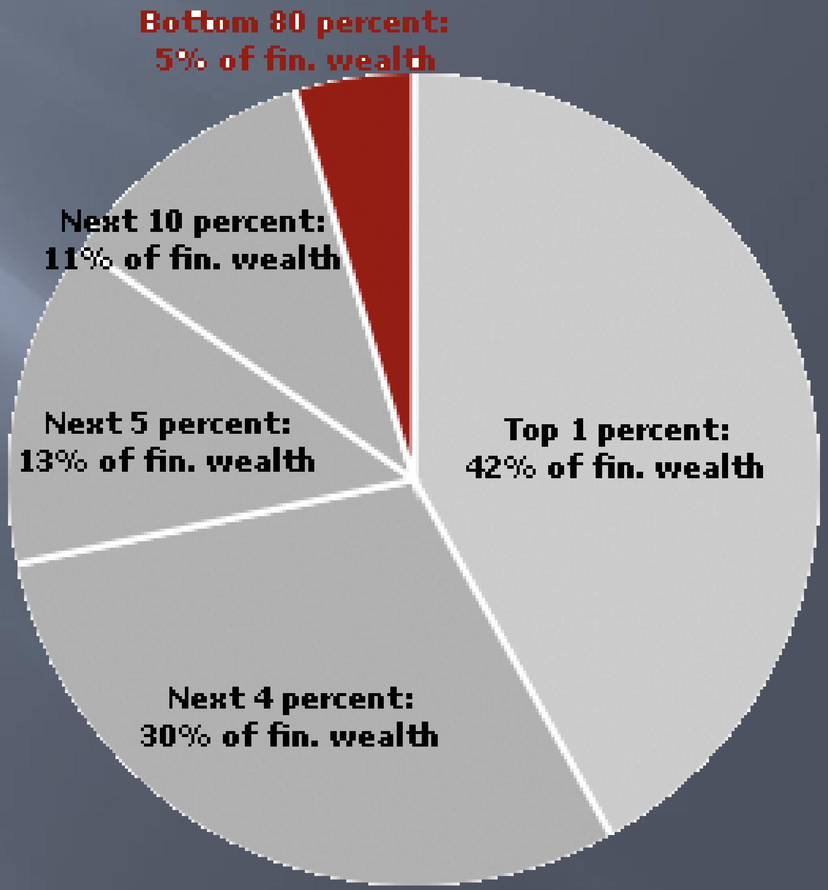
Apostle St. Matthew (El Greco)

Inhomogeneities on all scales

Wealth distribiution:



Net worth
distribution, 2010



Financial wealth
distribution, 2010

http://www2.ucsc.edu/whorulesamerica/power/wealth.html

Inhomogeneities on all scales

http://archikron.blogspot.hu/

The Guardian Nov 11, 2011

Inhomogeneities on all scales

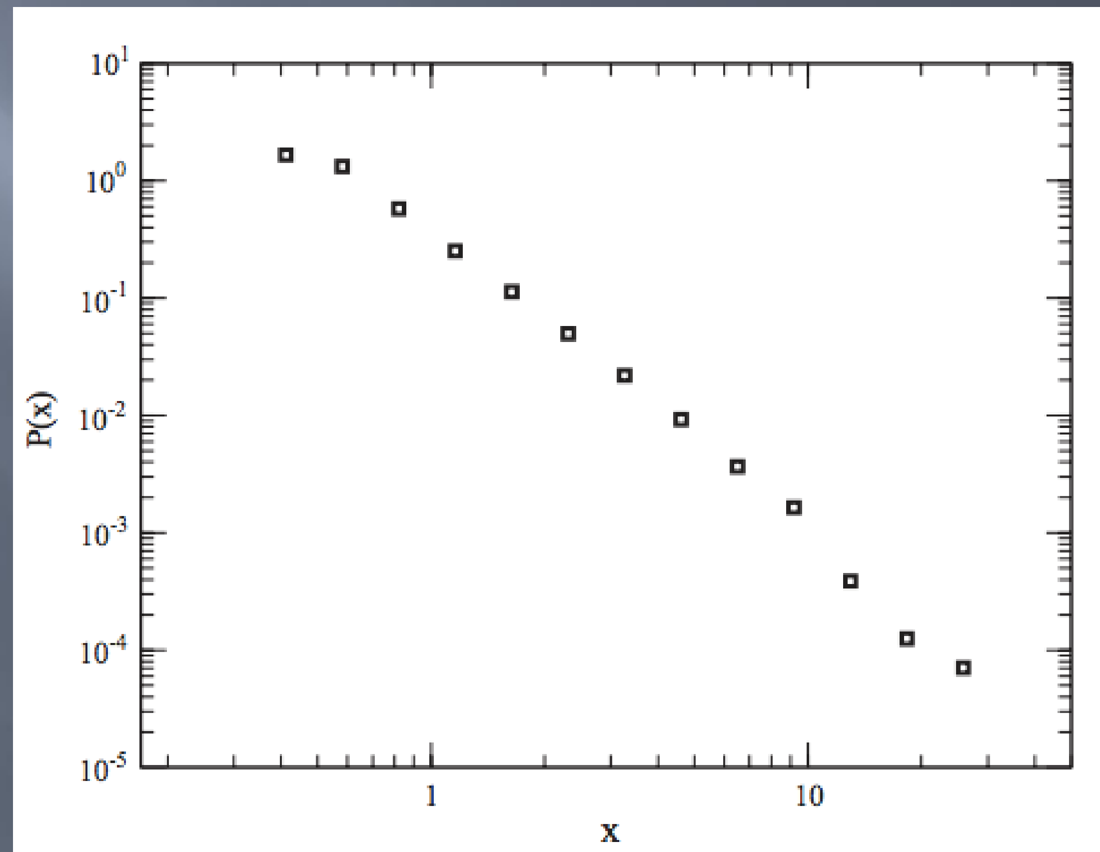
“We are the 99%”? Not quite.

There is a distribution of wealth and there are people with wealth on all scales

$p(x) \sim x^{-a}$

Pareto distribution

$a \approx 2.5$

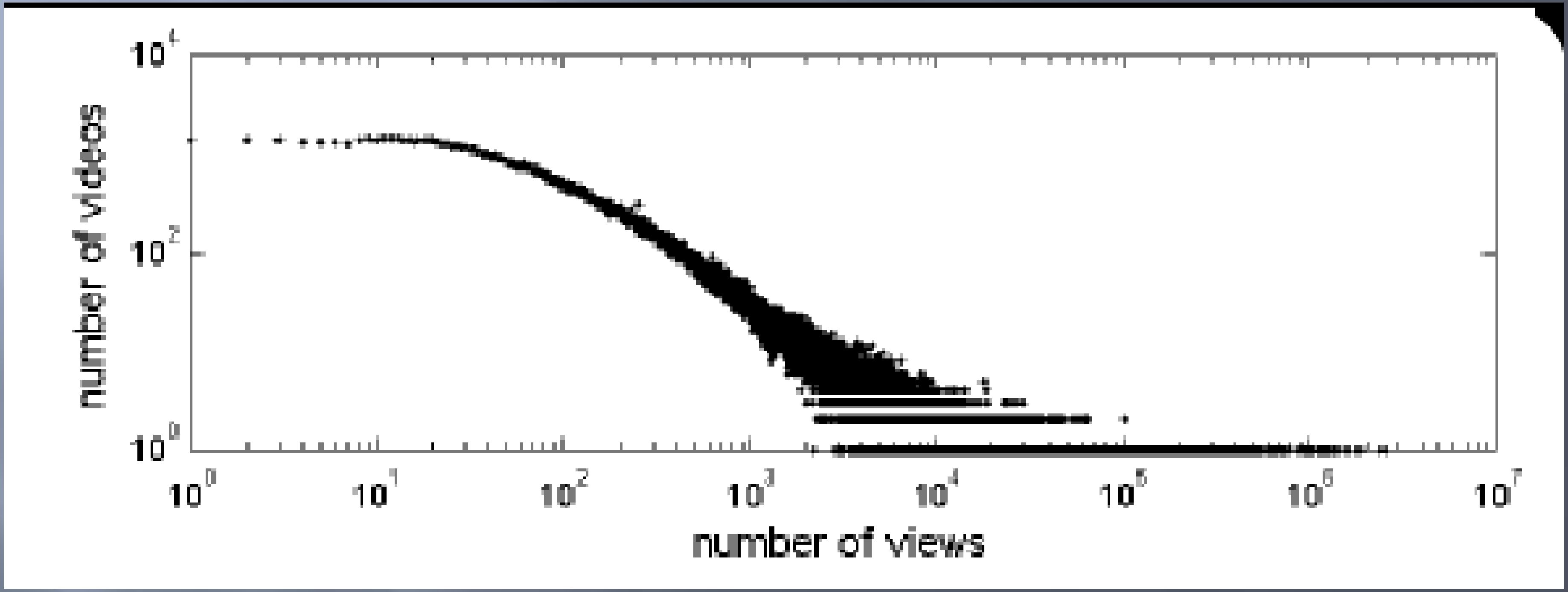


Forbes 400 (1988-2003); $x=w/<w>$

O.S. Klass et al. Economics Letters 90 (2006) 290 – 295

Inhomogeneities on all scales

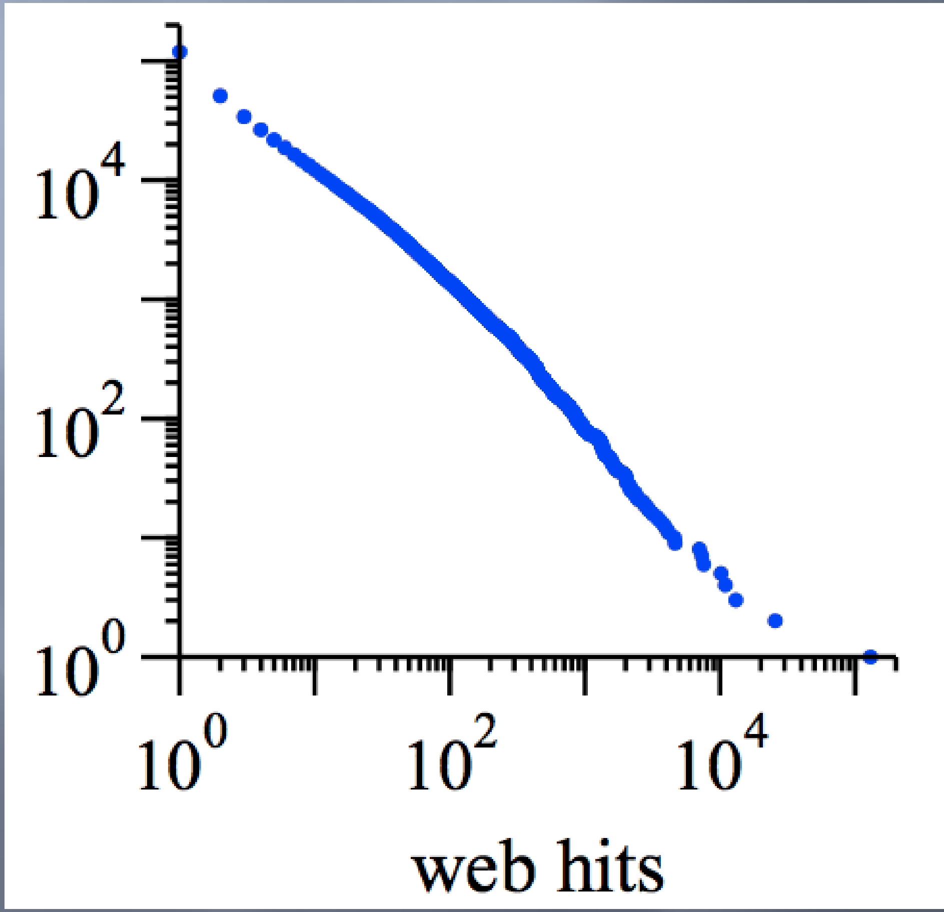
Popularity of youtube videos



M. Cha et al. KAIST (2007)

Inhomogeneities on all scales

Popularity of web pages

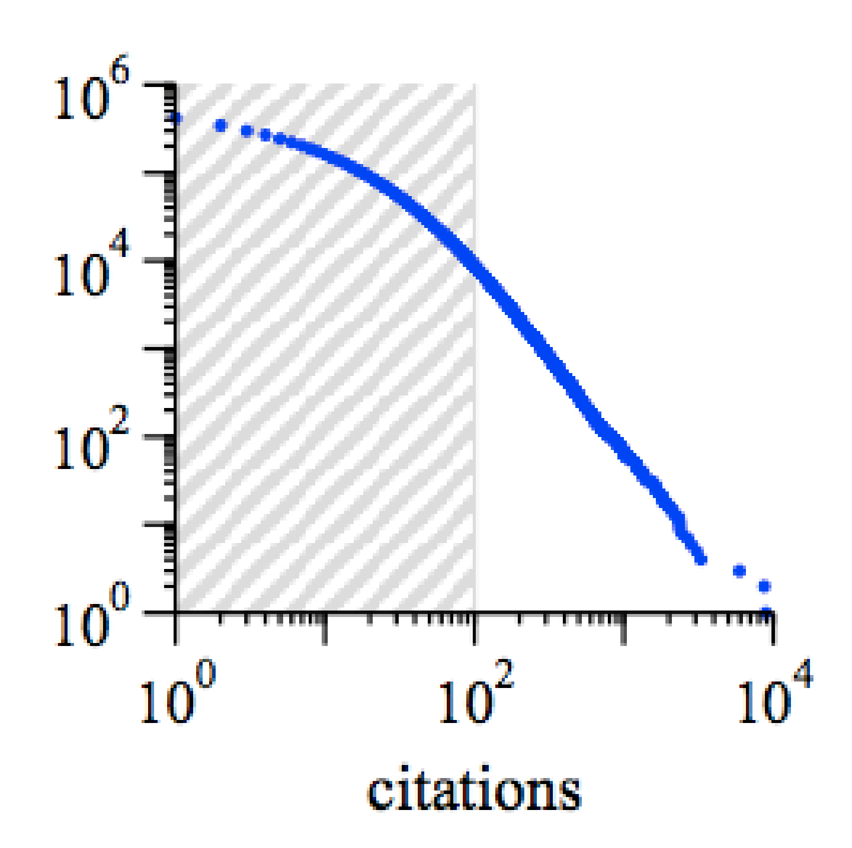


Distribution of pages with given # of clicks within given period of time.

L. A. Adamic and B. A. Huberman,. Quarterly J., of El. Commerce 1, 512 (2000).

Inhomogeneities on all scales

Popularity of scientific papers



(Independent) citations are scientometric measures often used in evaluation of papers and researchers.

S. Redner, Eur. Phys. J. B 4, 131134 (1998).

Power law distributions

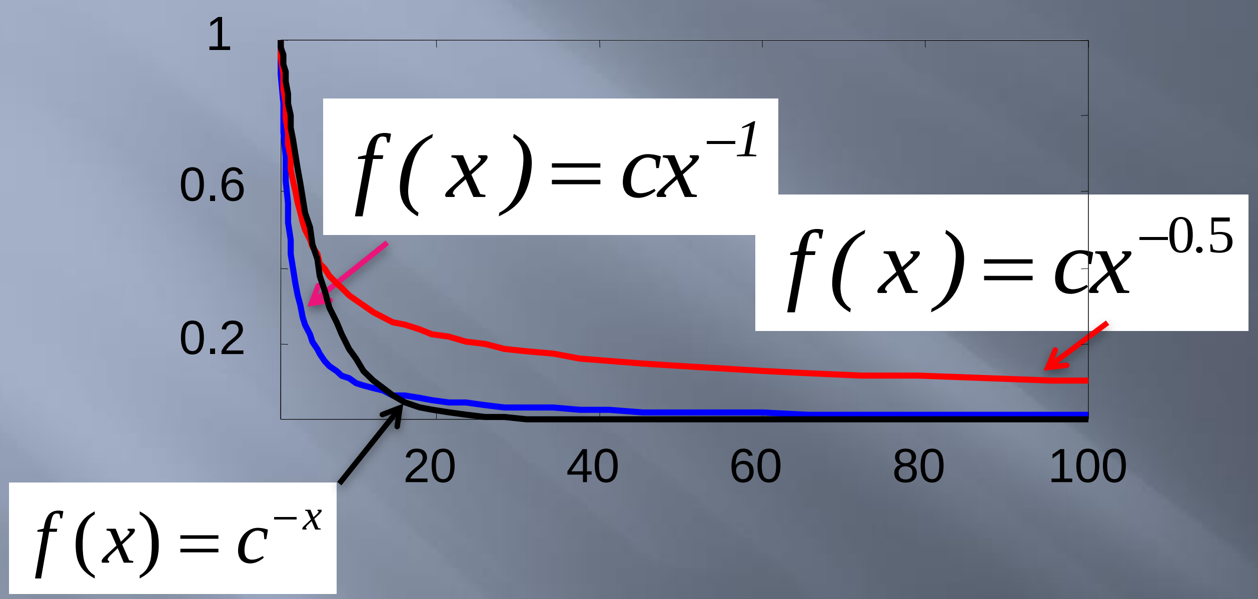
Remember: In percolation at criticality there is no characteristic length in the system → no scale → power laws.

If we have a distribution without a characteristic scale ("scale free" distribution) → power law.

Power laws are very inhomogeneous.

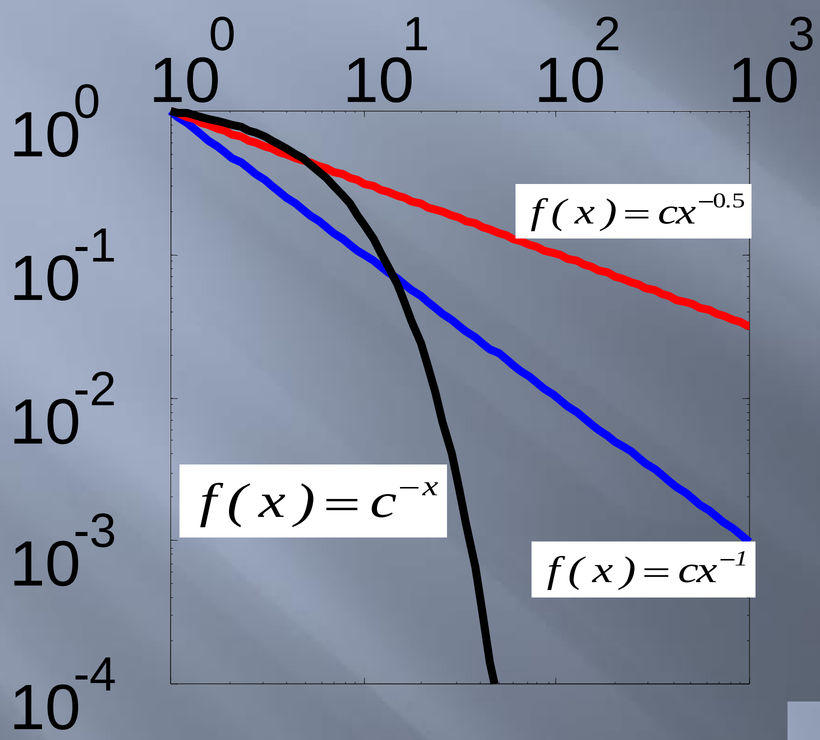
No scale? All scales!

Power law distributions

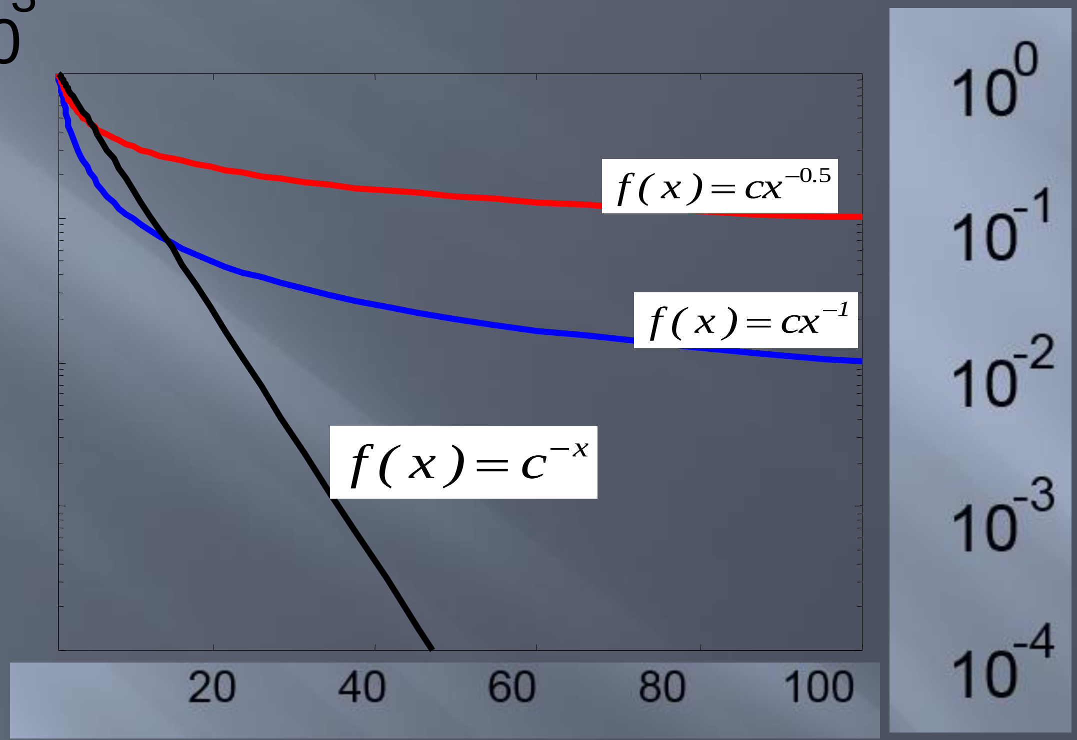


Above a certain x value, the power law is always higher than the exponential.

Power law distributions



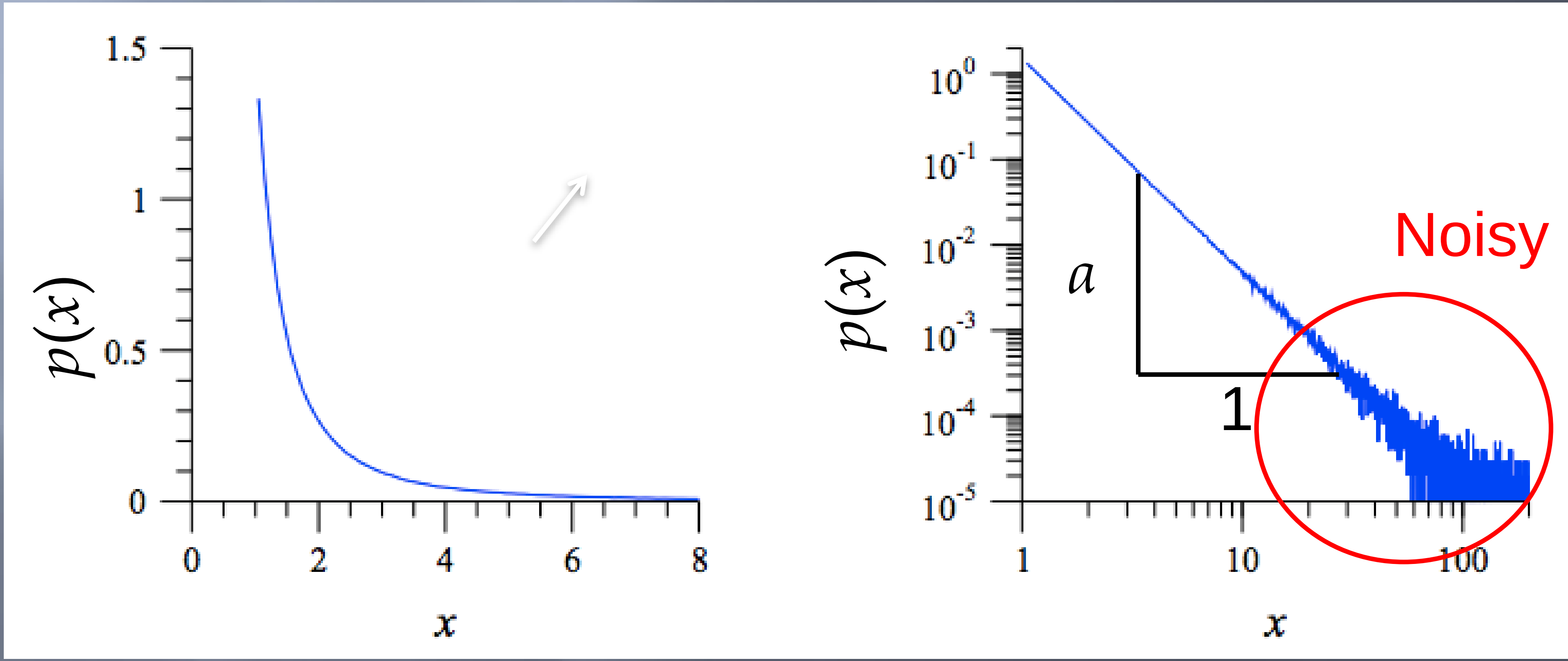
Log-log plot



Semilog plot

Power law distributions

We measure the empirical distribution by counting the frequency. $p(x) \sim x^{-a}$ $\log p(x) \sim -a \log x$

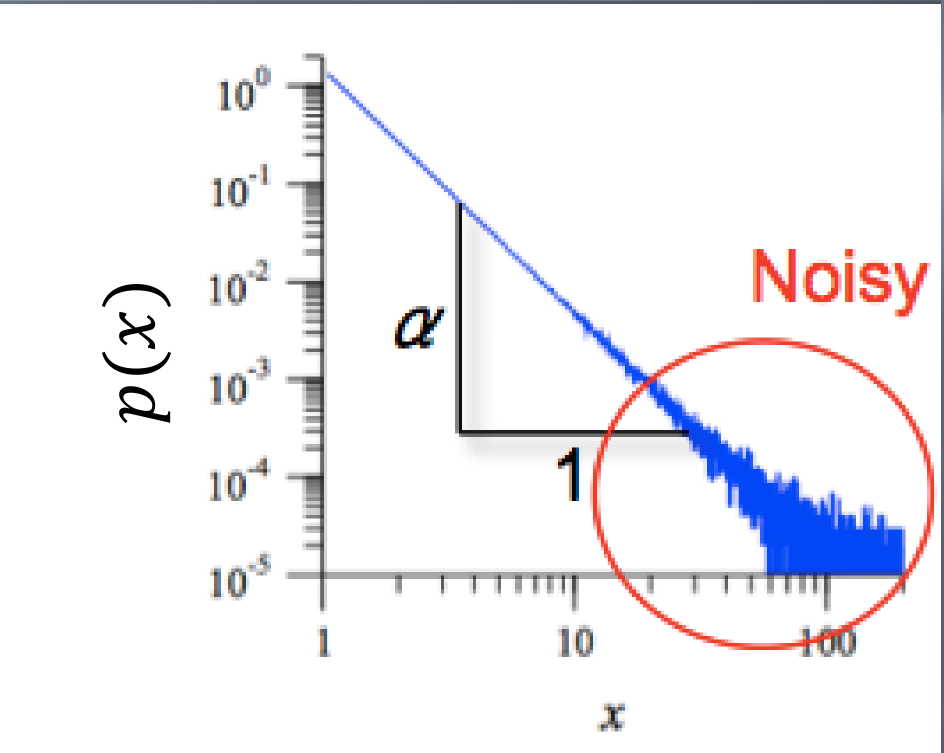


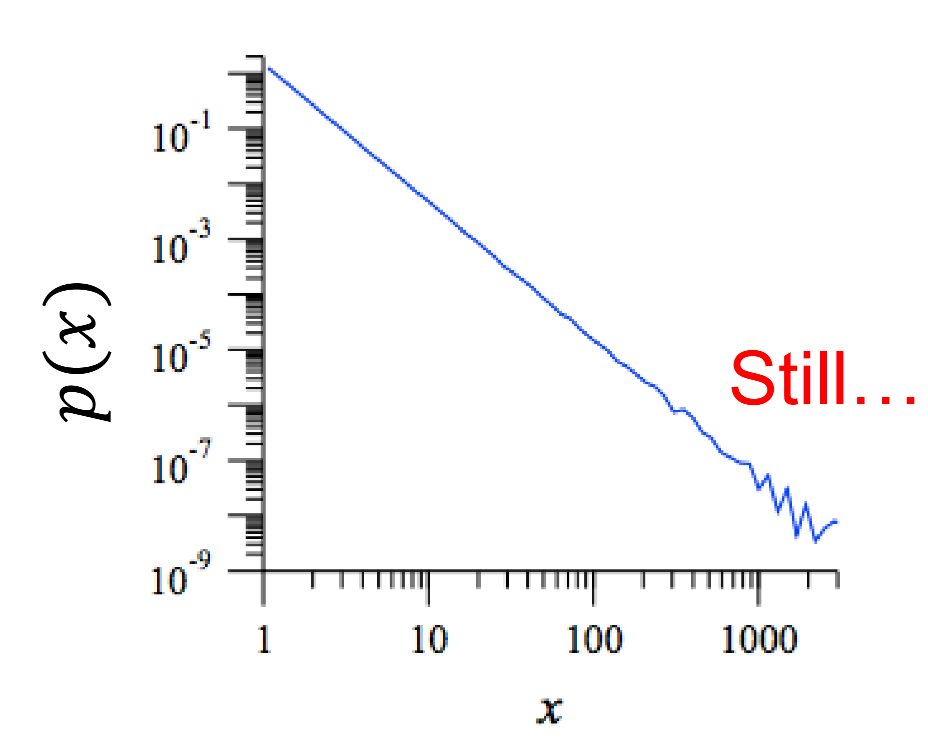
Lin-lin

Log-log

Power law distributions

The empirical distribution is obtained by making a histogram. If equidistant binning is used, there will be much fluctuations in the tail: Use log binning!





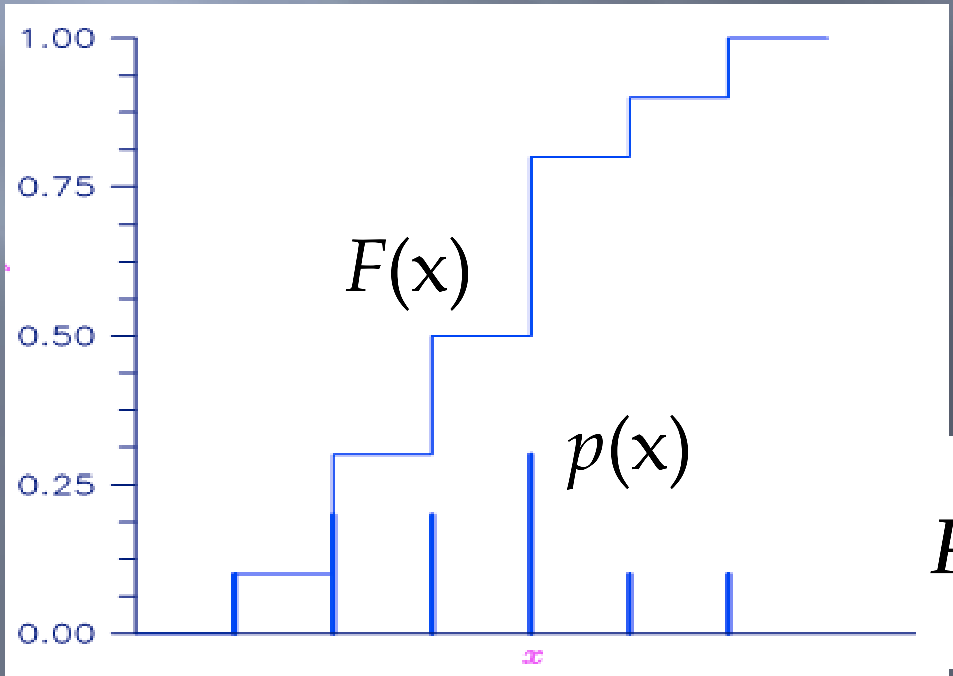
Lin. binning

Log. binning

Power law distributions

There is still too much fluctuation in the distribution function.

$$\sum_{x=x_{\min}}^{\infty} p(x) = 1; \qquad F(x) = P(x' \leq x) = \sum_{x'=x_{\min}}^{x} p(x')$$



Cumulative distribution

Or, alternatively

$$P(x' > x) = \sum_{x'=x}^{\infty} p(x') = 1 - F(x)$$

Power law distributions

If the random variable x is (quasi-) continuous, we have *probability density function,* denoted by $p(x)$

The probability that $a < x < b$ is then

$$P(a < x < b) = \int_a^b p(x)dx; \qquad \int_{-\infty}^{\infty} p(x)dx = 1$$

Cumulative distribution:

$$F(x) = P(x' \leq x) = \int_{-\infty}^{x} p(x')dx'$$

$$P(x' > x) = 1 - F(x) = \int_{x}^{\infty} p(x')dx'$$

Power law distributions

What if $p(x)$ is power law?

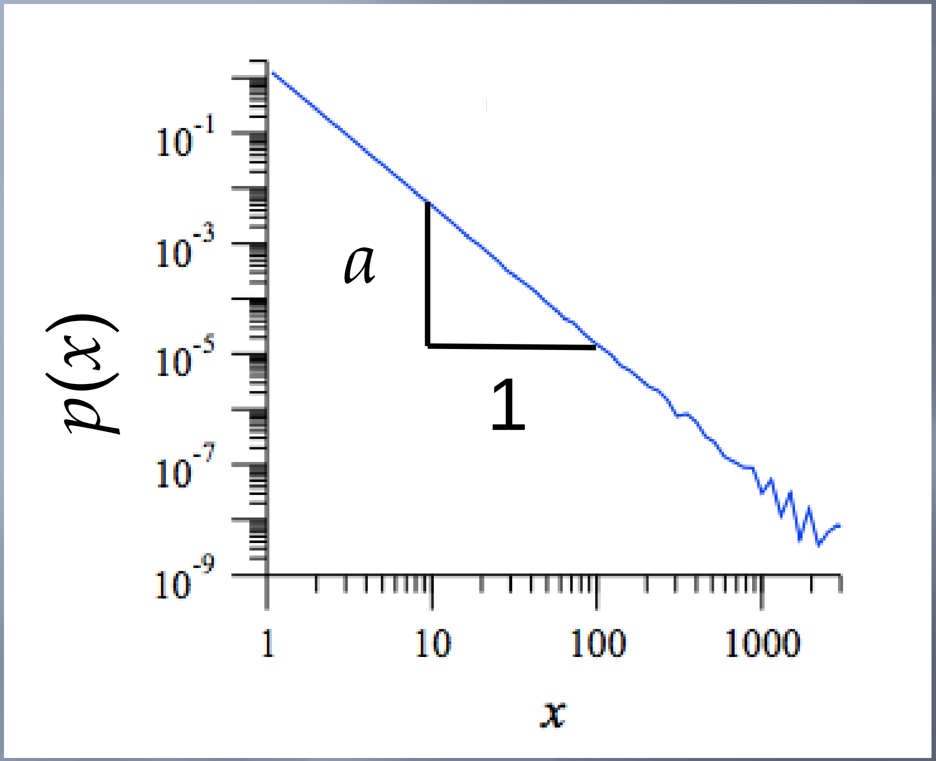
$$p(x) = Cx^{-\alpha}$$

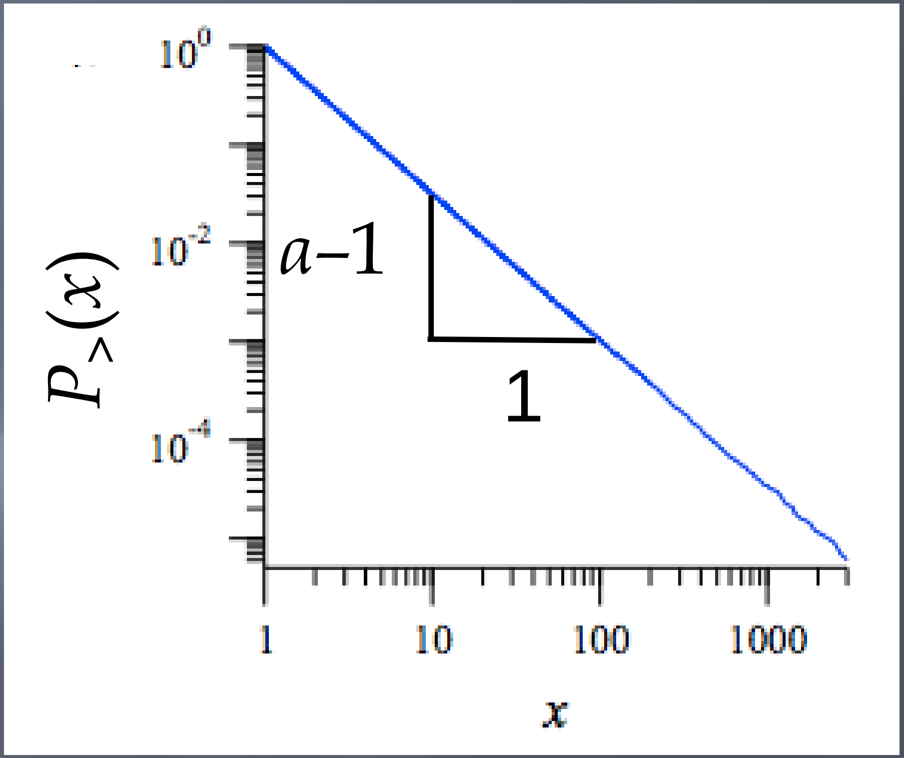
$$F(x) = \int_{x_{\min}}^{x} p(x')dx' = \frac{C}{\alpha - 1}\left[x_m^{-(\alpha-1)} - x^{-(\alpha-1)}\right]$$

$$P_>(x) = 1 - F(x) = \int_{x}^{\infty} p(x')dx' = \frac{C}{\alpha - 1} x^{-(\alpha-1)}$$

If the probability density decays as a power law with an exponenent α then the cumulative distribution function $P_>(x)$ will also decay as a power law with an exponent α–1.

Power law distributions





Log. binning

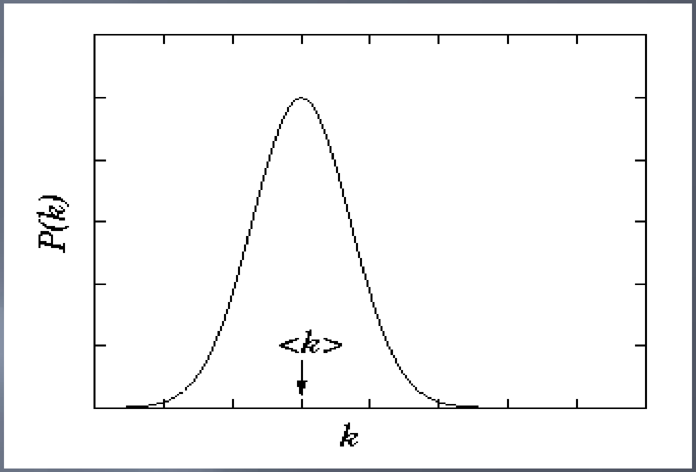
Slope: $-a$

Cumulative distribution

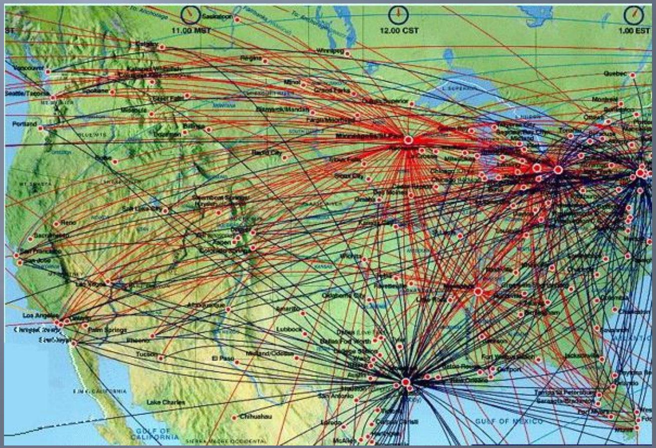
Slope: $-(a-1)$

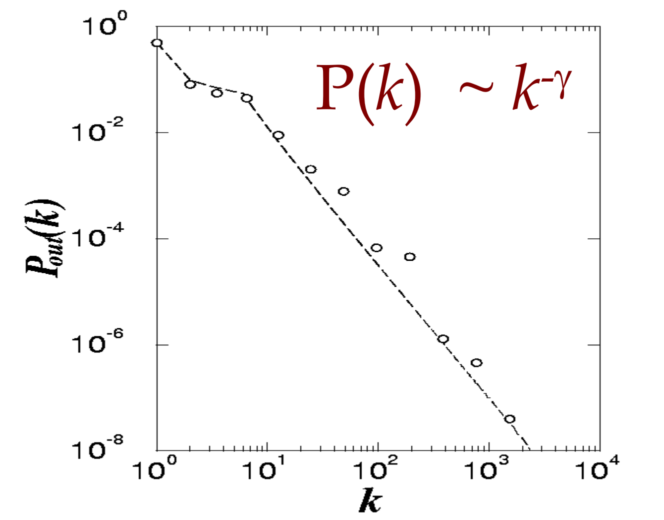
More sophisticated methods: Clauset et al., *SIAM Review* **51**(4), 661-703 (2009)

Inhomogeneities in complex networks



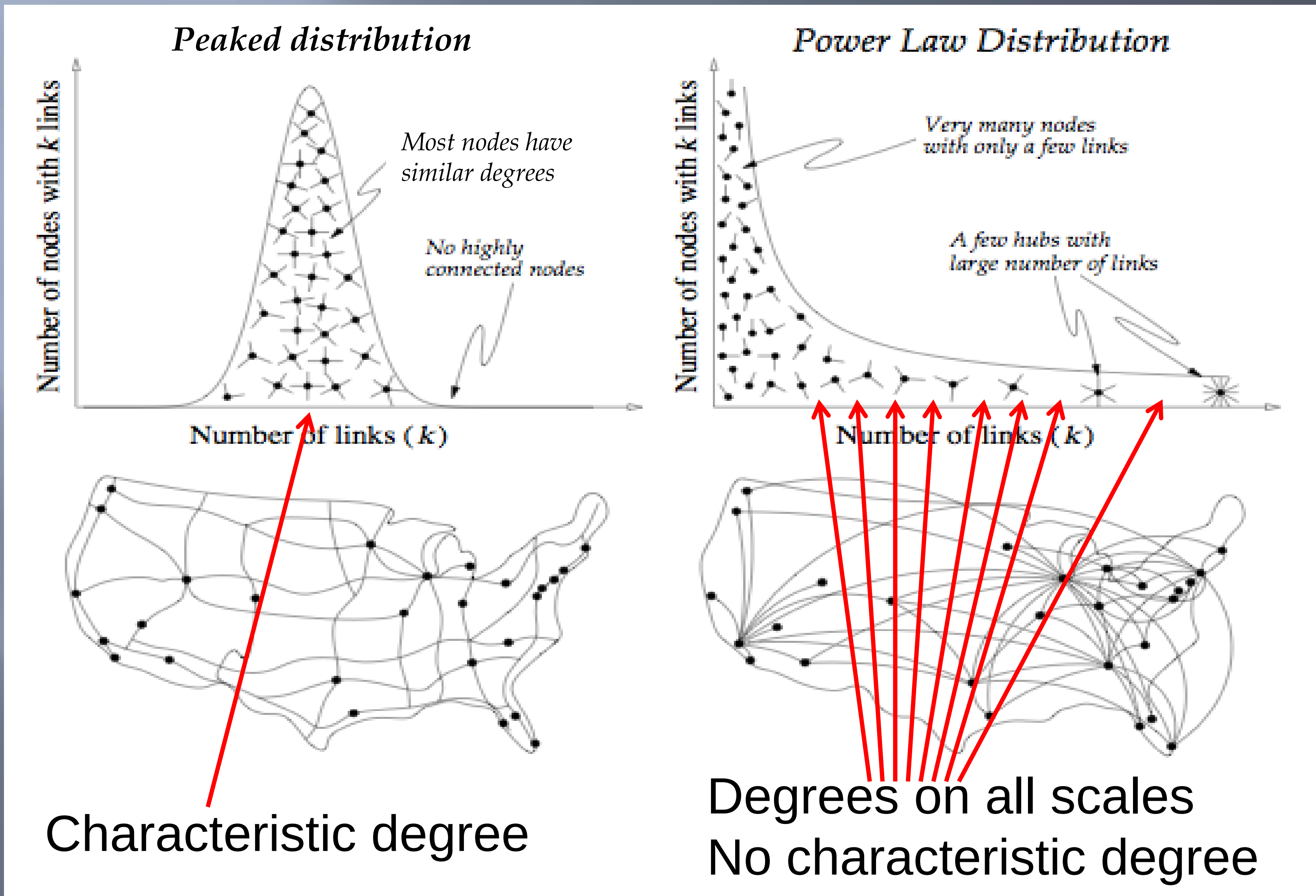
Homogeneous





Inhomogeneous

Inhomogeneities in complex networks



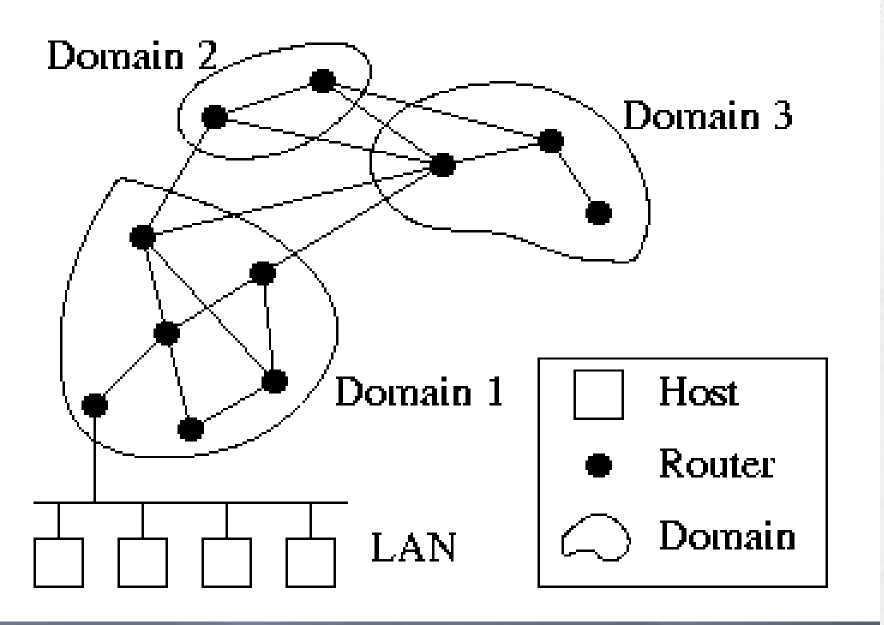
Scale free networks

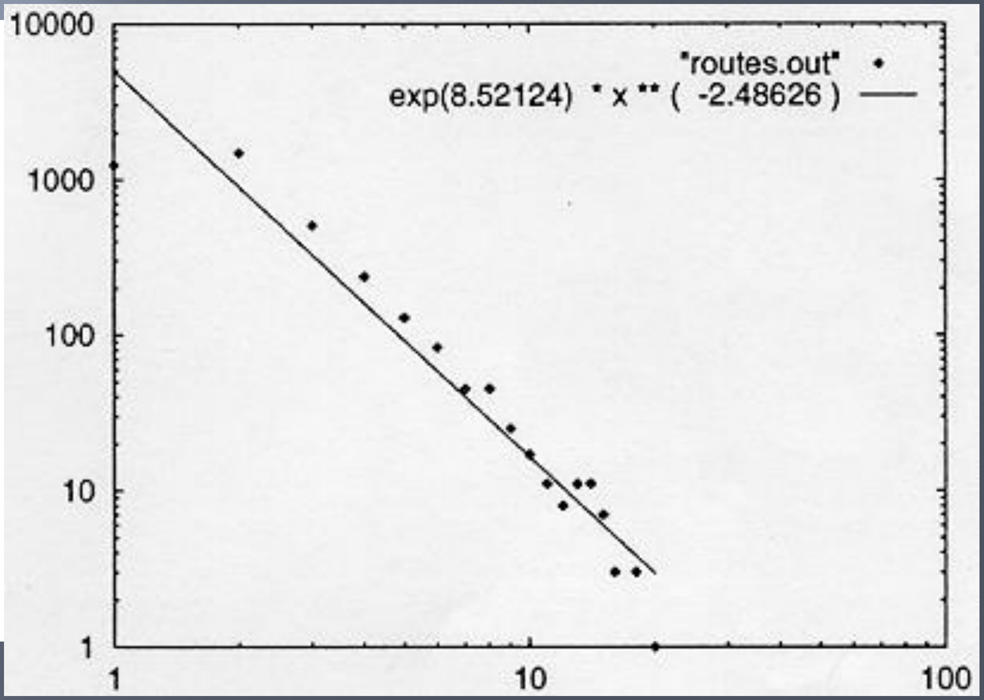
Networks with a degree distribution having a power law tail are called scale free networks

Internet

There are many!

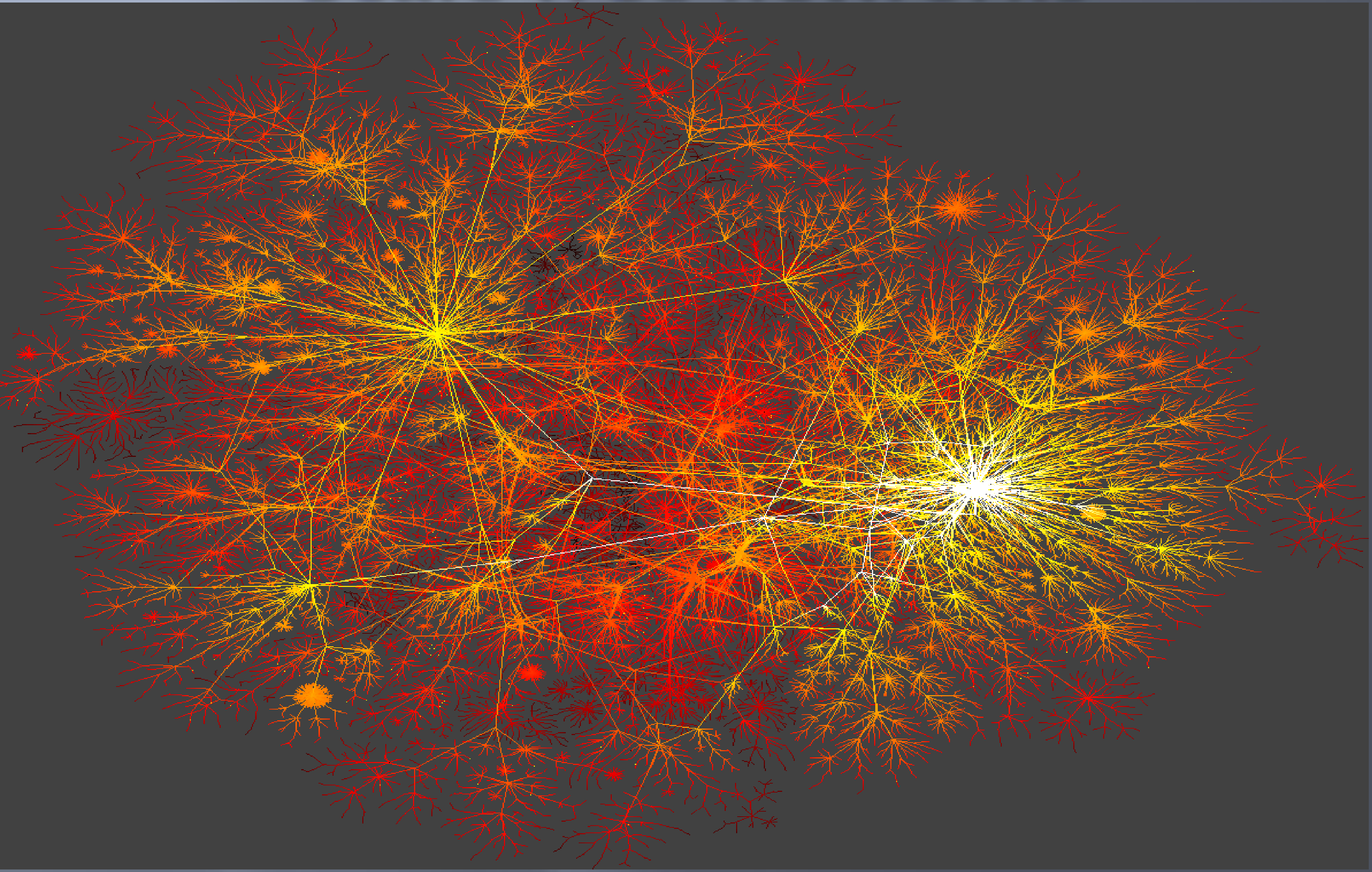
Nodes: computers, routers
Links: physical lines





(Faloutsos, Faloutsos and Faloutsos, 1999)

Scale free networks

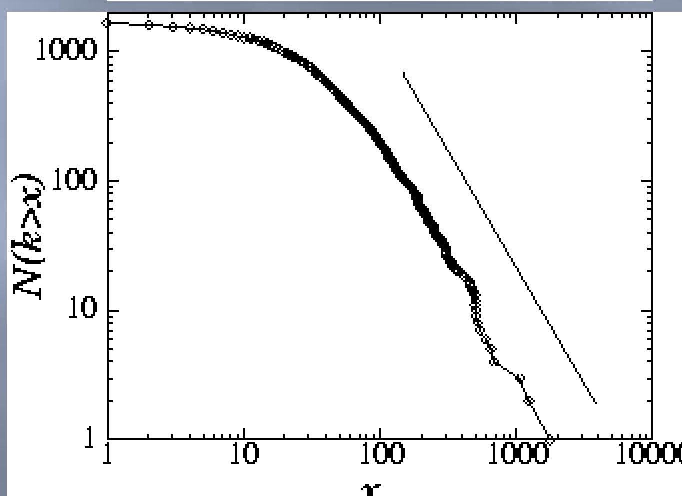


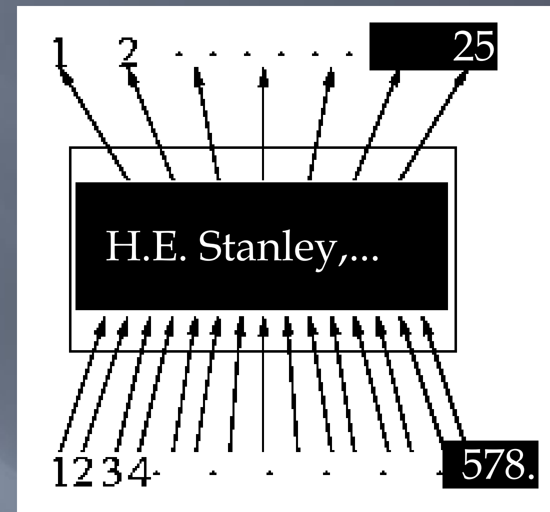
The presence of hubs is apparent!

Scale free networks

Nodes: papers
Links: citations

1736 PRL papers (1988)





$$P(k) \sim k^{-\gamma}$$

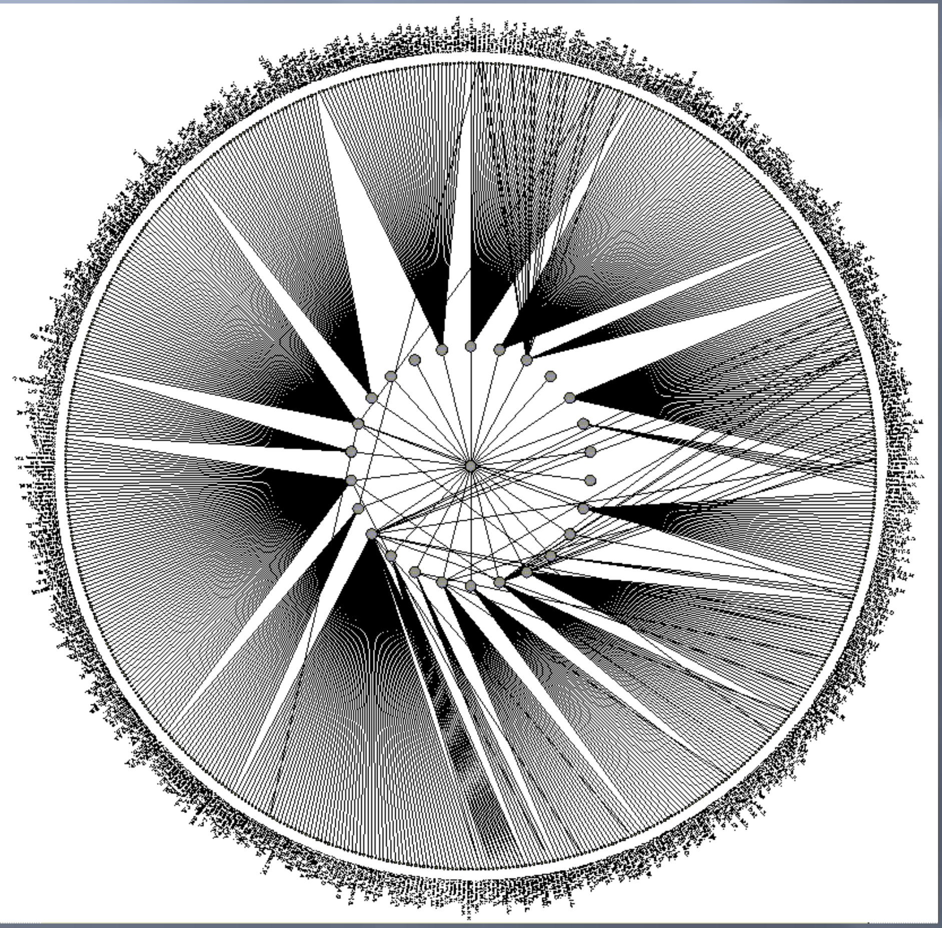
$(\gamma \sim 3)$

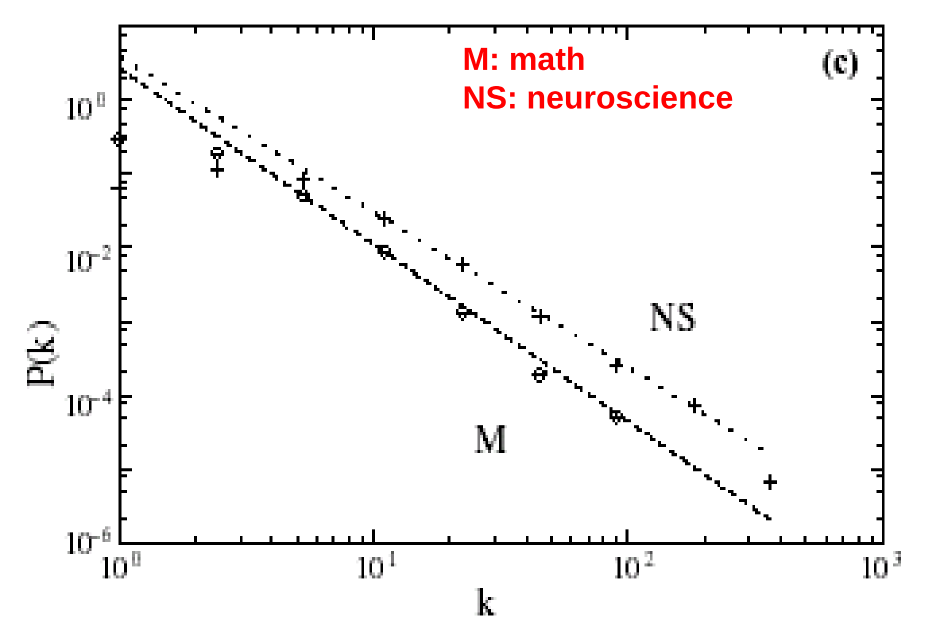
(S. Redner, 1998)

Scale free networks

Collaboration network

Nodes: scientist (authors)
Links: joint publication

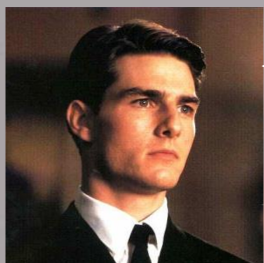


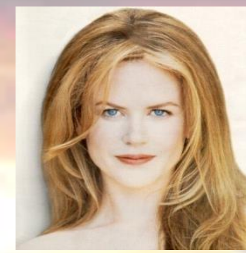


(Newman, 2000, Barabasi et al 2001)

Scale free networks

Actor network





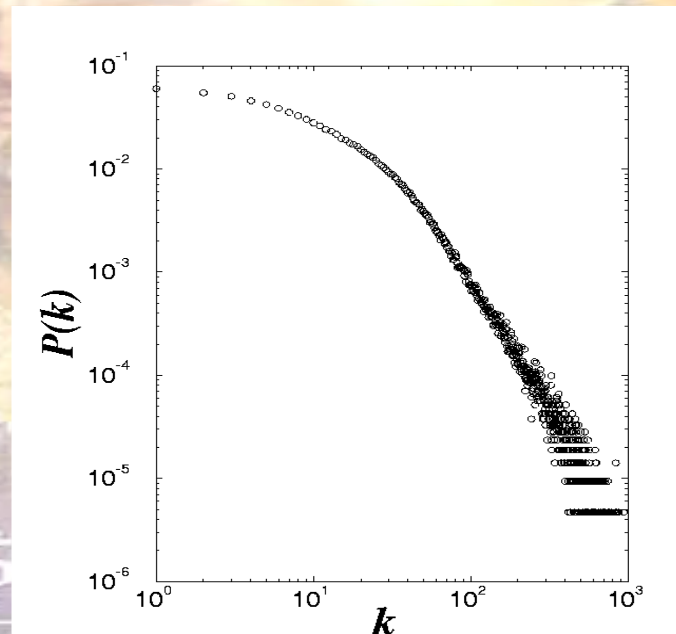
Nodes: actors
Links: cast jointly

Days of Thunder *(1990)*
Far and Away *(1992)*
Eyes Wide Shut *(1999)*

N = 212,250 actors
$\langle k \rangle = 28.78$

$P(k) \sim k^{-\gamma}$

$\gamma = 2.3$



IMDb Internet Movie Database

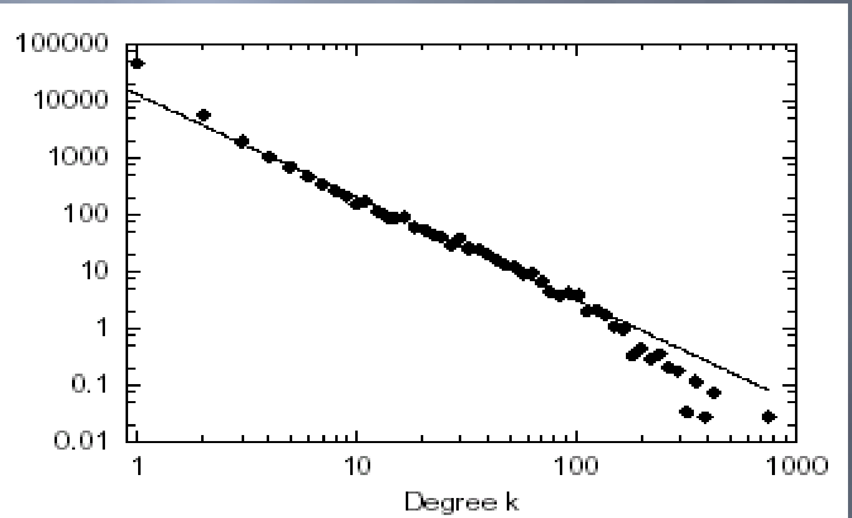
Scale free networks

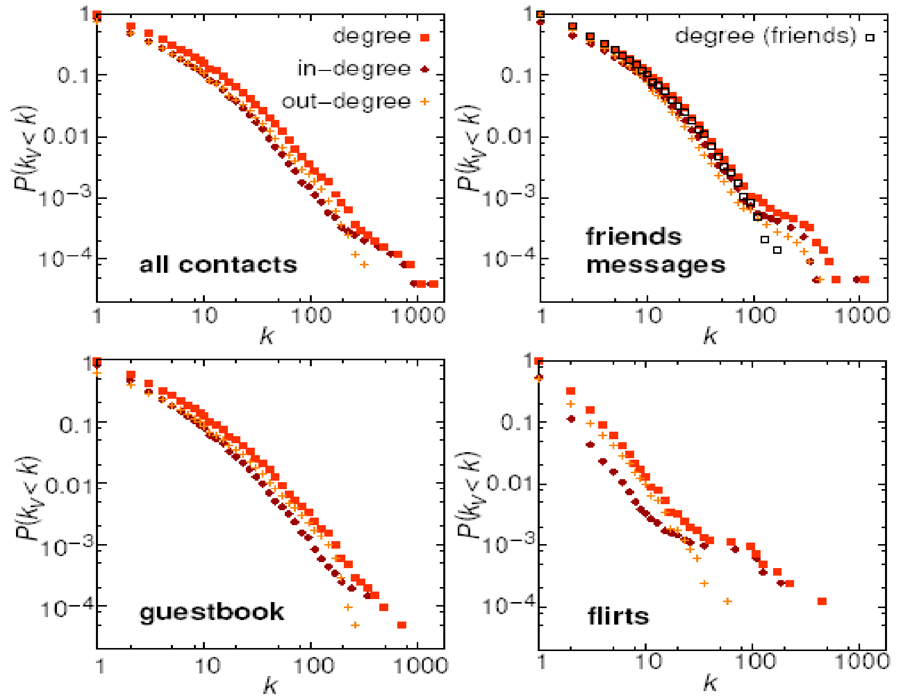
Social networks

Pussokram.com online dating community;
512 days, 25,000 users.

Nodes: online user
Links: email contact

Kiel University log files
112 days, N=59,912 nodes





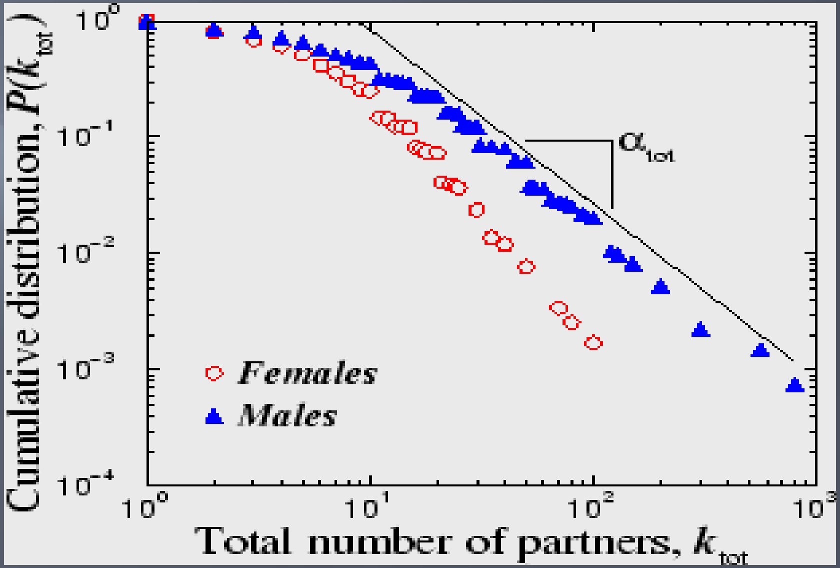
Ebel, Mielsch, Bornholdtz, PRE 2002.

Holme, Edling, Liljeros, 2002.

Scale free networks

Network of sexual contacts

Nodes: people (Females; Males)
Links: sexual relationships

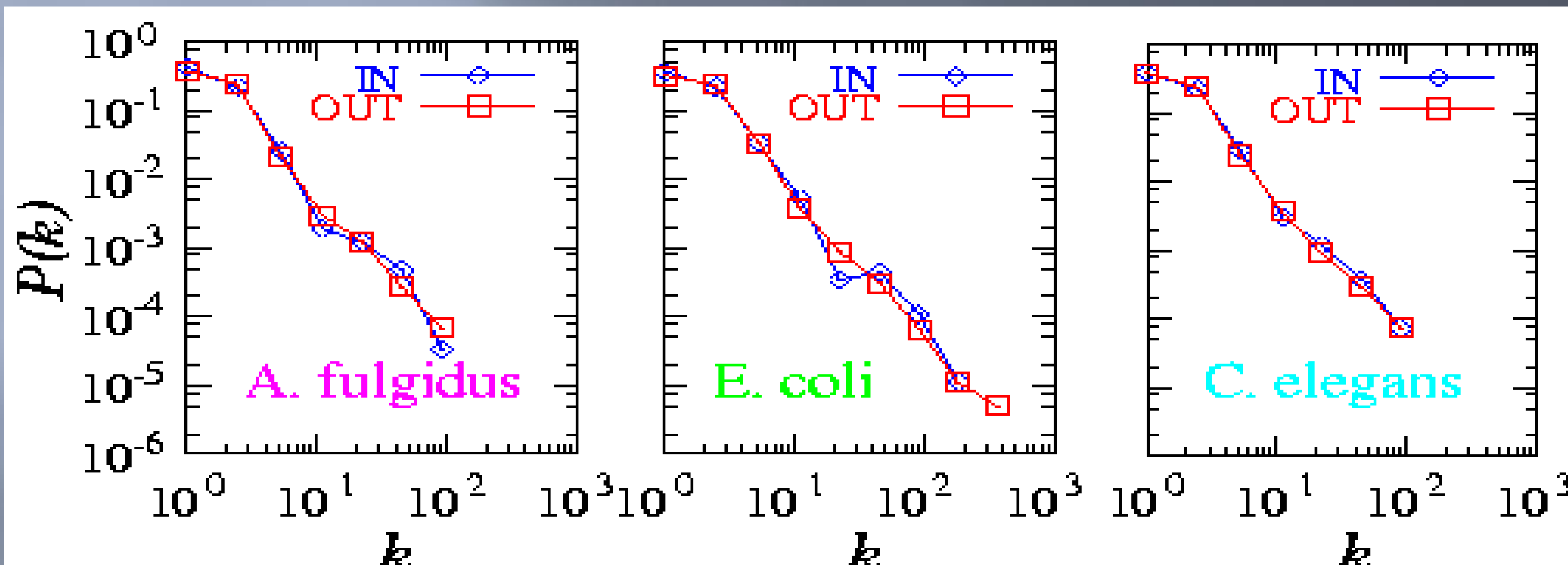


4781 Swedes; 18-74;
59% response rate.

Liljeros et al. Nature 2001

Scale free networks

Metabolic network



Archaea Bacteria Eukaryotes

Organisms from all three domains of life are scale-free!

$$P_{in}(k) \approx k^{-2.2}$$

$$P_{out}(k) \approx k^{-2.2}$$

H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, and A.L. Barabasi, *Nature*, 407 651 (2000)

Scale free networks

Network	Size	$\langle k \rangle$	κ	γ_{out}	γ_{in}
WWW	325 729	4.51	900	2.45	2.1
WWW	4×10^7	7		2.38	2.1
WWW	2×10^8	7.5	4000	2.72	2.1
WWW, site	260 000				1.94
Internet, domain*	3015–4389	3.42–3.76	30–40	2.1–2.2	2.1–2.2
Internet, router*	3888	2.57	30	2.48	2.48
Internet, router*	150 000	2.66	60	2.4	2.4
Movie actors*	212 250	28.78	900	2.3	2.3
Co-authors, SPIRES*	56 627	173	1100	1.2	1.2
Co-authors, neuro.*	209 293	11.54	400	2.1	2.1
Co-authors, math.*	70 975	3.9	120	2.5	2.5
Sexual contacts*	2810			3.4	3.4
Metabolic, *E. coli*	778	7.4	110	2.2	2.2
Protein, *S. cerev.**	1870	2.39		2.4	2.4
Ythan estuary*	134	8.7	35	1.05	1.05
Silwood Park*	154	4.75	27	1.13	1.13
Citation	783 339	8.57			3
Phone call	53×10^6	3.16		2.1	2.1
Words, co-occurrence*	460 902	70.13		2.7	2.7
Words, synonyms*	22 311	13.48		2.8	2.8

<u>Networks</u>:
The exponents vary from system to system.
Most are between 2 and 3

Universality?

Properties of power law distributions

$$p_k = Ck^{-g}$$

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

$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s} = \frac{1}{1^s} + \frac{1}{2^s} + \frac{1}{3^s} + \cdots \qquad s \in \mathbb{R}, \quad s > 1$$

Riemann Zeta function

$$p_k = \frac{k^{-g}}{z(g)}$$

for $k>0$ (i.e. we assume that there are no disconnected nodes in the network)

Properties of power law distributions

In continuous formalism:

$$p(k) = Ck^{-\gamma} \qquad k = [K_{\min}, \infty)$$

$$\int_{K_{\min}}^{\infty} p(k)dk = 1$$

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

$$p(k) = (\gamma - 1)K_{\min}^{\gamma-1}k^{-\gamma}$$

Since a distribution has to be normalized, $\gamma > 1$

Properties of power law distributions

m-th moment of the degree distribution: $$< k^m >= \int_{K_{\min}}^{\infty} k^m p(k) dk$$

$$p(k) = (\gamma - 1) K_{\min}^{\gamma - 1} k^{-\gamma}$$

$$k = [K_{\min}, \infty)$$

$$< k^m >= (\gamma - 1) K_{\min}^{\gamma - 1} \int_{K_{\min}}^{\infty} k^{m-\gamma} dk = \frac{(\gamma - 1)}{(m - \gamma + 1)} K_{\min}^{\gamma - 1} \left[k^{m-\gamma+1} \right]_{K_{\min}}^{\infty}$$

If $m - \gamma + 1 < 0$: $$< k^m >= - \frac{(g - 1)}{(m - g + 1)} K_{\min}^m$$

If $m - \gamma + 1 >= 0$, the integral diverges.
For a fixed γ this means that all moments with $m >= \gamma - 1$ diverge.

Properties of power law distributions

Most degree exponents are smaller than 3 → $\langle k^2 \rangle$ diverges!!!

$$\sigma_k = (< k^2 > - < k >^2)^{1/2} \rightarrow \infty$$

$$k =< k > \pm s_k$$

Network	Size	$\langle k \rangle$	κ	γ_{out}	γ_{in}
WWW	325 729	4.51	900	2.45	2.1
WWW	4×10^7	7		2.38	2.1
WWW	2×10^8	7.5	4000	2.72	2.1
WWW, site	260 000				1.94
Internet, domain*	3015–4389	3.42–3.76	30–40	2.1–2.2	2.1–2.2
Internet, router*	3888	2.57	30	2.48	2.48
Internet, router*	150 000	2.66	60	2.4	2.4
Movie actors*	212 250	28.78	900	2.3	2.3
Co-authors, SPIRES*	56 627	173	1100	1.2	1.2
Co-authors, neuro.*	209 293	11.54	400	2.1	2.1
Co-authors, math.*	70 975	3.9	120	2.5	2.5
Sexual contacts*	2810			3.4	3.4
Metabolic, *E. coli*	778	7.4	110	2.2	2.2
Protein, *S. cerev.**	1870	2.39		2.4	2.4
Ythan estuary*	134	8.7	35	1.05	1.05
Silwood Park*	154	4.75	27	1.13	1.13
Citation	783 339	8.57			3
Phone call	53×10^6	3.16		2.1	2.1
Words, co-occurrence*	460 902	70.13		2.7	2.7
Words, synonyms*	22 311	13.48		2.8	2.8

WWW $k = \langle 7 \rangle \pm \infty$

Internet $k = \langle 3.5 \rangle \pm \infty$

Metabolic $k = \langle 7.4 \rangle \pm \infty$

Phone call $k = \langle 3.16 \rangle \pm \infty$

Due to the huge fluctuations empirical $\langle k \rangle$ looses meaning as an estimator.

Properties of power law distributions

Finite scale free networks
(real networks are always finite)

There will be a maximum degree: $K_{\max}$

$$\int_{K_{\max}}^{\infty} p(k)dk \approx \frac{1}{N}$$

The probability to have a node with degree larger than $K_{\max}$ should not exceed the prob. to have one node, i.e. $1/N$ fraction of all

$$\int_{K_{\max}}^{\infty} p(k)dk = (\gamma-1)K_{\min}^{\gamma-1}\int_{K_{\max}}^{\infty} k^{-\gamma}dk = \frac{(\gamma-1)}{(-\gamma+1)}K_{\min}^{\gamma-1}\left[k^{-\gamma+1}\right]_{K_{\max}}^{\infty} = \frac{K_{\min}^{\gamma-1}}{K_{\max}^{\gamma-1}} \approx \frac{1}{N}$$

$$K_{\max} = K_{\min} N^{\frac{1}{g-1}}$$

Distances in scale free networks

How do Kevin Bacon, Erdős, etc. games work?

These are scale free networks
Find a path to a hub (usually short) →
Find the path to the target (also short)

Due to the presence of hubs, scale free networks are automatically small worlds!

The mechanism is different from that of the ER or the Watts-Strogatz model.

Distances in scale free networks

$$< l > \sim \begin{cases} const. & \gamma = 2 \\ \dfrac{\ln\ln N}{\ln(\gamma - 1)} & 2 < \gamma < 3 \\ \dfrac{\ln N}{\ln\ln N} & \gamma = 3 \\ \ln N & \gamma > 3 \end{cases}$$

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

Small World ($\gamma > 3$)

$$K_{max} = K_{min} N^{\frac{1}{g-1}}$$

Degree of the biggest hub is of order O(N). Most nodes can be connected within two layers of it, thus the average path length will be independent of the system size.

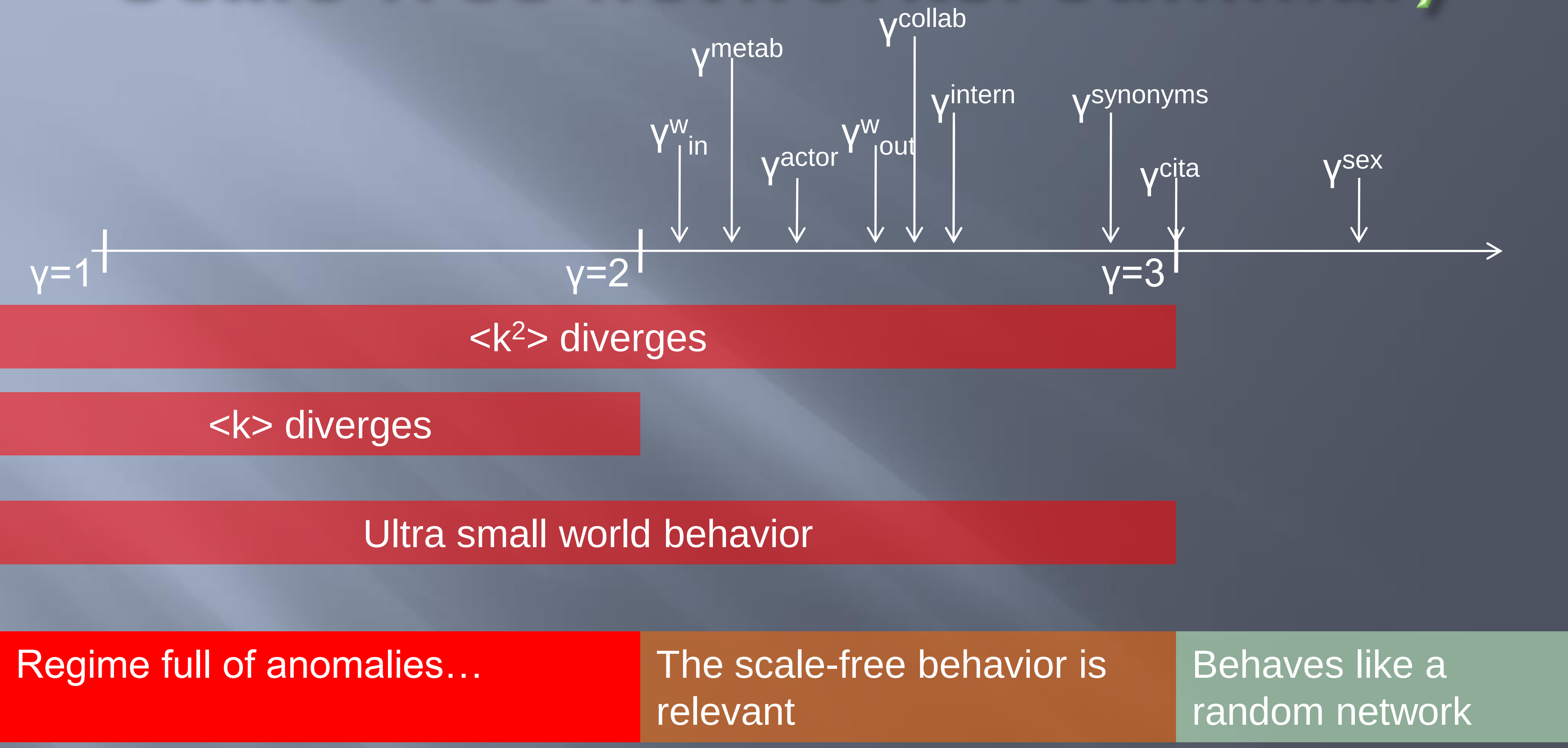
The average path length increases in a double log manner so it is much slower than logarithmic. In a random network all nodes have comparable degree, thus most paths will have comparable length. In a scale-free network the vast majority of the path go through the few high degree hubs, reducing the distances between nodes.

Some key models produce γ=3, so the result is of particular importance for them. This was first derived by Bollobas and Riordan for the network diameter in the context of a dynamical model, but it holds for the average path length as well.

The second moment of the distribution is finite, thus in many ways the network behaves as a random network. Hence the average path length follows the result that we derived for the random network model earlier.

Cohen, Havlin 2003, Bollobas, Riordan, 2002;

Scale free networks: summary



Scale free networks: summary

Practical remarks:

- The tail of the distribution follows often a power law causing divergence of the moments

- Since the low *k* regime “does not matter” and the network is always finite, we usually have a lower and an upper cutoff for the power law (in slang: scaling)

A form which reflects both cutoffs: $P(k) \sim (k + k_0)^{-g} \exp(-\frac{k + k_0}{k_t})$

Properties of large real world networks

- Small worldness
- High clustering
- Scale free degree distribution

Erdős-Rényi: Small world, low clustering,
narrow degree-distribution
Watts Strogatz: Small world, high clustering
narrow degree distribution

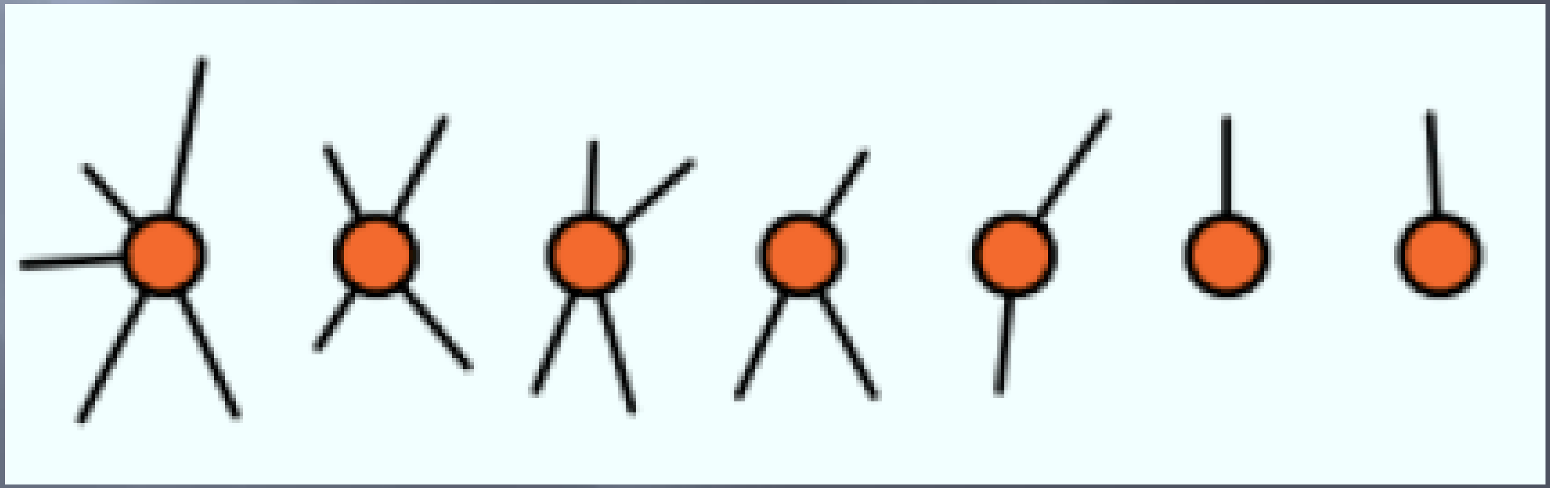
How to construct models with prescribed properties?

In concreto: With a given degree distribution?

Configuration model

How to generate scale free (power law) degree distribution?

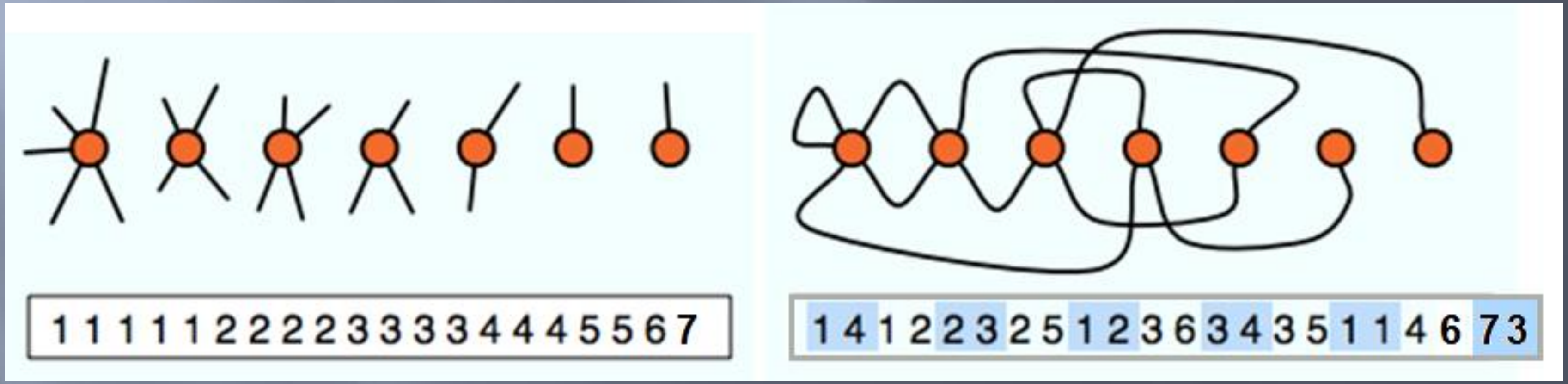
Instead of taking a degree distribution we make a model for a prescribed set of degrees. Let us have N nodes, where the i-th has degree k_i.



Configuration model

The degrees of the nodes are indicated by “half links” or “stubs”.

The network is constructed then by pairwise connecting the stubs. One possible set of pairings:



This figure indicates the algorithm too.
Clearly, one needs even number of stubs to be able to pair them.

A. Clauset

Configuration model

This is a model for a degree sequence and not for a (given, theoretical) degree distribution. However, if N is large, the degree sequence taken from the distribution can be considered as representative. I.e., we generate a sequence from the distribution and from that the network.

The degree sequence itself defines an ensemble. For a given degree sequence, all possible pairings have the same probability. As the pairings are entirely random, there will be no correlations. (E.g., no (dis)assortativity). “Most random network with a given degree sequence.”

Configuration model

What is the weight of a given network?

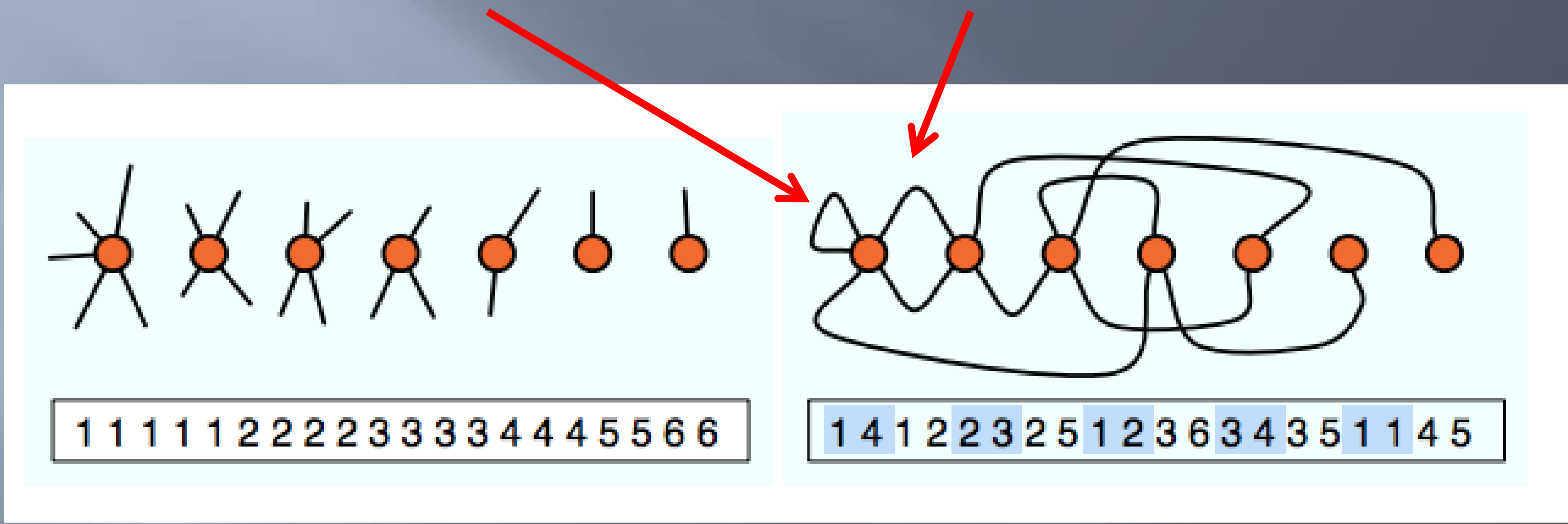
The number of permutations at a node is: $k_i!$
The total number of possible permutations in the network is then $N_{\text{perm}}\left(\{k_i\}\right) = \prod_i k_i!$

Since the degree sequence is constant in the ensemble, this means that all networks we construct have the same weight.

There is a little problem here!

Configuration model

Problems: self-links and multiple links



These are usually unwelcome. (We want to have a simple graph.) Moreover, they influence the number of permutations, e.g., exchanging the ends of a self link is not a new permutation; similarly, we over-count if there are multiple links.

Prohibit such pairings?

Configuration model

No!
This would mess up the statistics and even block the construction (what if there are no other possibilities than those we want to avoid?!).

If we are interested in large networks then this is usually a minor problem. Why?

For nice degree distributions the probability of self-links and multiple links decreases rapidly with the size N of the system! We can simply disregard them.

Caution is needed for power law distributions with exponents smaller than 3.

Configuration model

What is the expected number L_{self} self edges?

The probability of having a self edge at node i with degree k_i is (we choose two stubs at i for all $2L$-1 trials)

$$p_{ii} = \frac{k_i(k_i - 1)/2}{2L - 1} \approx k_i(k_i - 1)/4L$$

from which follows:

$$L_{\text{self}} = \sum_i p_{ii} = \sum_i k_i(k_i - 1)/4L = \sum_i k_i(k_i - 1)/2N\langle k \rangle = \frac{\langle k^2 \rangle - \langle k \rangle}{2\langle k \rangle}$$

For finite first and second moments L_{self} remains finite even in the N, $L \to \infty$ limit → it becomes negligible.

(Similar reasoning for multiple links.)

What if $2 < \gamma < 3$? We had $K_{\max} = K_{\min} N^{\frac{1}{\gamma - 1}}$

$$\langle k^2 \rangle \sim K_{\max}{}^{3-\gamma} \sim N^{\frac{3-\gamma}{\gamma-1}} \Rightarrow \lim_{N\to\infty} L_{\text{self}}/N \to 0$$

Still OK

Configuration model

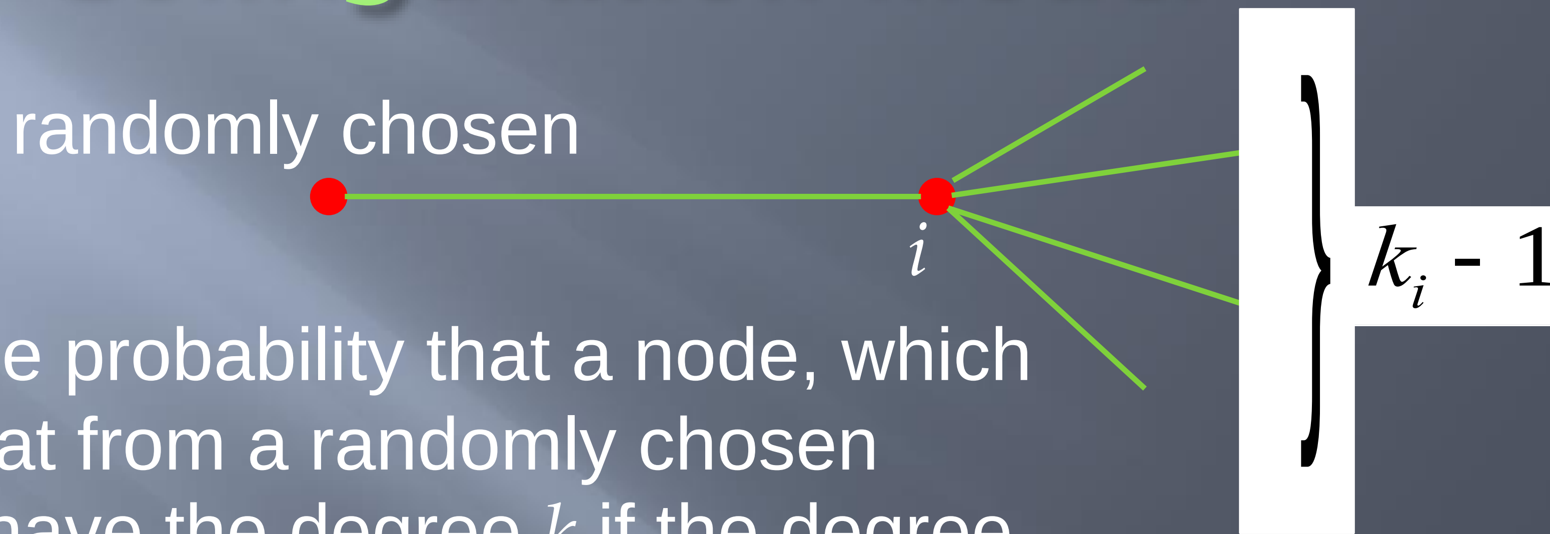
What is the probability p_{ij} in the config. model to have a link between node i and j?

Let us take a stub from node i. It has $2L$-1 possible pairing points. Out of these k_j are from node *j*. Thus the probability of “landing” at *j* is $k_j\,/\,(2L-1)$. But there are k_i different possibilities to chose the starting stub at i. The final result is then:

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

for large networks

Configuration model



What is the probability that a node, which we arrive at from a randomly chosen node will have the degree k if the degree distribution is p_k?

If i has degree k the probability of landing there is $k\,/\,(2L-1) \approx k\,/\,2L$ (for large networks). There are Np_k nodes which have k degrees. Thus the probability that we land at any of them starting from an arbitrary node is

$$p_{\mathrm{nn}}(k) = \frac{k}{2L}, \quad Np_k = \frac{kp_k}{\langle k \rangle}$$

proportional to kp_k not to p_k only!

Configuration model

Let us assume that a friendship network can be described by the configuration model. What is the average number of friends of your friends.

$$\langle k \rangle_{\text{nn}} = \sum_k k p_{\text{nn}}(k) = \sum_k \frac{k^2 p_k}{\langle k \rangle} = \frac{\langle k^2 \rangle}{\langle k \rangle}$$

The average degree is just $\langle k \rangle$. The above formula tells that $\langle k \rangle_{\text{nn}} > \langle k \rangle$ because:

$$\frac{\langle k^2 \rangle}{\langle k \rangle} - \langle k \rangle = \frac{\langle k^2 \rangle - \langle k \rangle^2}{\langle k \rangle} = \frac{s^2}{\langle k \rangle} > 0$$

"Your friend has more friends than you do."

S. Feld, Am. J. Sociol. 96, 1464 (1991)

Configuration model

Collaboration networks and Internet:

Network	n	Average degree	Average neighbor degree	$\frac{\langle k^2 \rangle}{\langle k \rangle}$
Biologists	1 520 252	15.5	68.4	130.2
Mathematicians	253 339	3.9	9.5	13.2
Internet	22 963	4.2	224.3	261.5

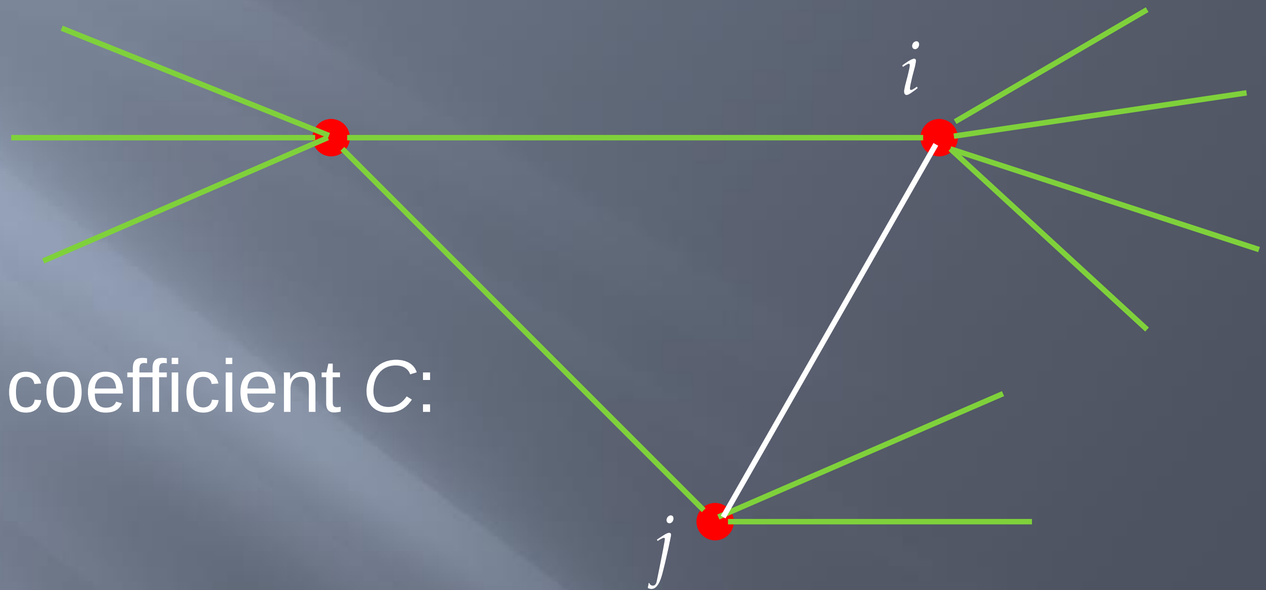
Config. model is not exact (see last column) but captures an important aspect.

What is the probability q_k that an arbitrary node is connected to another one with k degrees in excess to the link between them? (Excess degree distribution)

$$q_k = p_{\mathrm{nn}}(k+1) = \frac{(k+1)p_{k+1}}{\langle k \rangle}$$

Newman book

Configuration model



Global clustering coefficient *C*:

$q_{k_i} q_{k_j}$ will be the distribution that nodes i and j have k_i and k_j excess degrees, respectively. Since the probabiliy of having a bond between two nodes having k_i and k_j free degrees is $k_i k_j / 2L$, we have

$$C = \sum_{k_i, k_j = 0}^{\infty} q_{k_i} q_{k_j} \frac{k_i k_j}{2L} = \frac{1}{2L} \left(\sum_{k=0}^{\infty} k q_k\right)^2 = \frac{1}{2L} Const = \frac{1}{N} const$$

Configuration model

$$C = \frac{1}{N} const(p_k)$$

where the "const" depends on the moments of the distribution.

We see that in the large *N* limit the average clustering coefficient becomes small.

Most (especially social) networks have high clustering!

Three important features:
1. Short average distance
2. High clustering
3. Broad (in the tail often power law) distribution

3. Automatically fulfilled (by construction)
2. Fails

What about 1?

Configuration model

One might think that power law implies hubs and hubs were needed for small worldness → configuration model with power law degree distribution will automatically be a small world.

This reasoning assumes a single component or at least a giant component (the “world”, which is expected to be small).

Nothing assures a priori that there is a giant component in the configuration network with power law distribution of degrees.

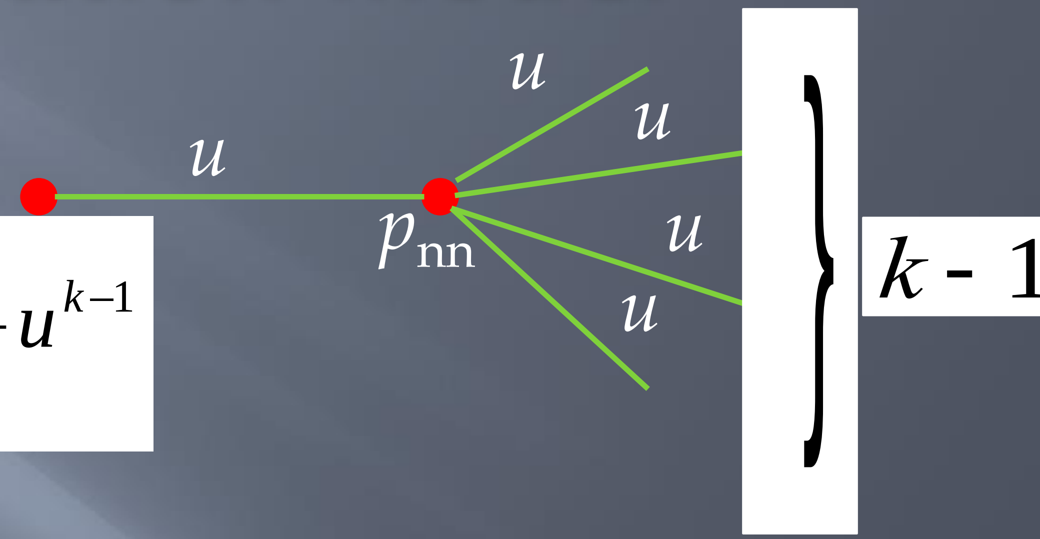
Configuration model

In fact, this is not always the case. If the exponent of the power law is too large, that means the decay of the probability of finding high degree nodes is too fast, there will be only isolates.

We calculate generally for the configuration model the probability of having a giant component following the ideas we used for the ER graph.

Let *u* be the probability that a *link* does not lead to a giant (infinite) component

Configuration model



$$u = \sum_{k=1}^{\infty} p_{\text{nn}}(k) u^{k-1} = \sum_{k=1}^{\infty} \frac{k p_k}{\langle k \rangle} u^{k-1}$$

$$u = g(u) = \frac{1}{\langle k \rangle} \sum_{k=1}^{\infty} k p_k u^{k-1}$$

Trivial solution: $u = 1$ since $\langle k \rangle = \sum_{k=1}^{\infty} k p_k$

Is there any other solution? (Needed for having a giant component.)

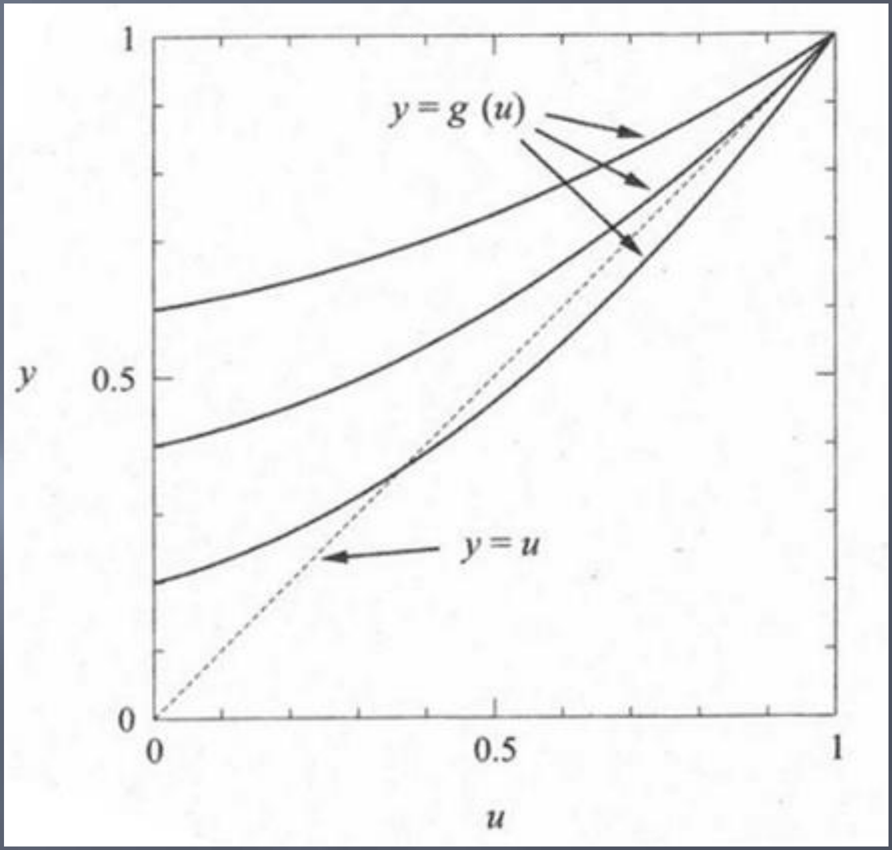
For $p_k = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$ the ER result is retrieved (check!)

Configuration model

$$u = g(u) = \frac{1}{\langle k \rangle} \sum_{k=1}^{\infty} k p_k u^{k-1}$$

The tipping point:
$g'(u)=1$

For $g'(u=1)>1$ there is a giant component, because there is a solution $u < 1$



Consequently, the probability of leading to a giant component is $1 - u > 0$.

Configuration model

For $g'(u)>1$ there is a giant component, because there is a solution $u<1$. What does it mean?

$$g'(u)=\frac{d}{du}\left\{\frac{1}{\langle k\rangle}\sum_{k=1}^{\infty}kp_k u^{k-1}\right\}=\frac{1}{\langle k\rangle}\sum_{k=1}^{\infty}k(k-1)p_k u^{k-2}>1$$

At $u=1$:

$$\frac{1}{\langle k\rangle}\sum_{k=1}^{\infty}k(k-1)p_k=\frac{1}{\langle k\rangle}\sum_{k=1}^{\infty}k^2 p_k-\frac{1}{\langle k\rangle}\sum_{k=1}^{\infty}kp_k>1$$

$$\frac{1}{\langle k\rangle}\sum_{k=1}^{\infty}k^2 p_k-\frac{1}{\langle k\rangle}\sum_{k=1}^{\infty}kp_k=\frac{1}{\langle k\rangle}\langle k^2\rangle-\frac{1}{\langle k\rangle}\langle k\rangle>1$$

From which the result for random network follows: There is a giant component if

$$\langle k^2\rangle-2\langle k\rangle>0$$

Molloy-Reed 1995

Configuration model

$$\langle k^2 \rangle - 2\langle k \rangle > 0$$

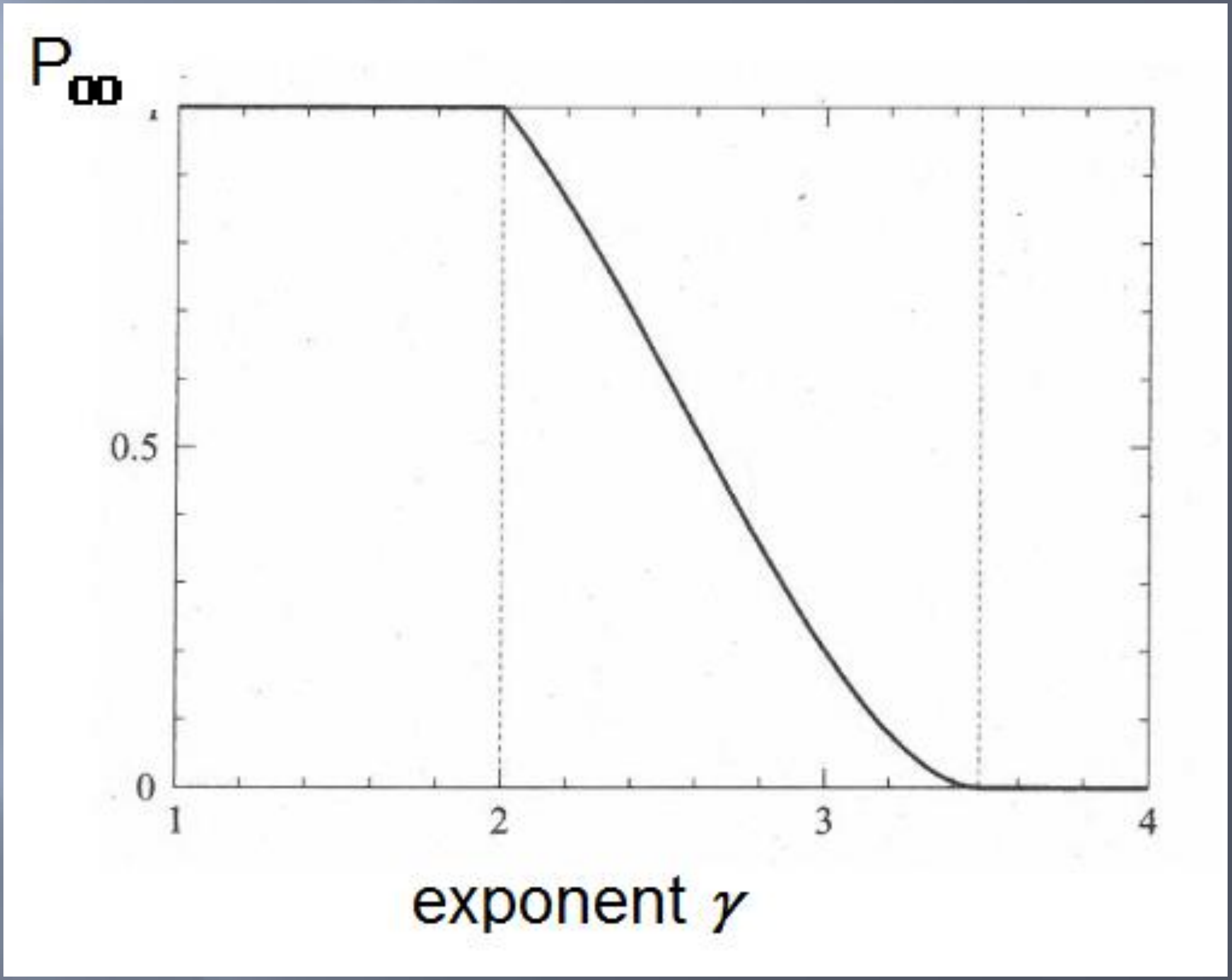
This is the general Molloy-Reed criterion for the existence of a giant component.

What does it mean for power law degree distributions?

Usually we have only a power law in the tail. The small *k* values do not matter from the point of view of the asymptotic behavior but they influence the values of the moments.

If $p_k \sim Ak^{-\gamma}$ at least asymptotically, the second moment diverges for $\gamma \leq 3$. Therefore for these values the MR inequality is automatically satisfied. In fact, one can show that for small enough γ there is only one component in an infinite system. (The prob. to find an isolate → 0.)

Configuration model



Assuming power law from $k = 1$.

Configuration model

Low clustering is a problem!

Can we take the brute force approach as for the degree distribution?

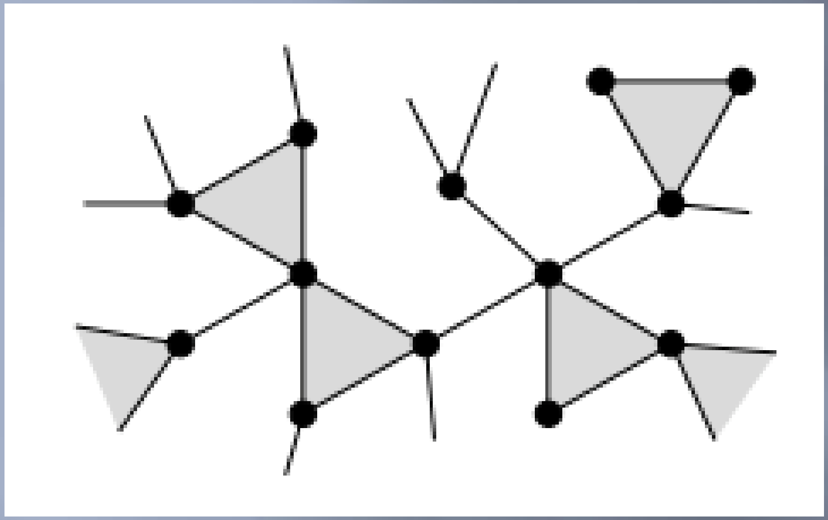
Yes!

Instead of nodes with stubs only, we take nodes with stubs and corners of triangles!

P_{st} will be the probability of having a node with s stubs and t corners. (The total number of stubs must be a multiple of 2, that of the corners a multiple of 3.)

Newman 2009

Configuration model



Of course, the total degree is given from contributions by the stubs and the corners (with multiplicity 2).

$$p_k = \sum_{s,t=0}^{\infty} p_{st} \delta(k, s+2t)$$

where

$$\delta(a,b) = \begin{cases} 1 \text{ if } & a=b \\ 0 & \text{otherwise} \end{cases}$$

(Kronecker delta)

Several properties can be calculated, e.g., percolation threshold:

$$\left[\frac{\langle s^2\rangle}{\langle s\rangle} - 2\right]\left[2\frac{\langle t^2\rangle}{\langle t\rangle} - 3\right] = 2\frac{\langle st\rangle^2}{\langle s\rangle\langle t\rangle}$$

This replaces Maloy-Reed

Configuration model

Further refinements are possible. E.g., correlations between degree and clustering (which indeed do exist).

In principle, whenever we discover a new feature of a network, we may incorporate that into the random network model!

What can be learned from such a model?

Homework

In a regular graph all degrees k are the same.
Generate with the configuration model random regular graphs with $k = 1,2,$ and 3
Visualize and characterize the graphs.