The recent network-hype in physics

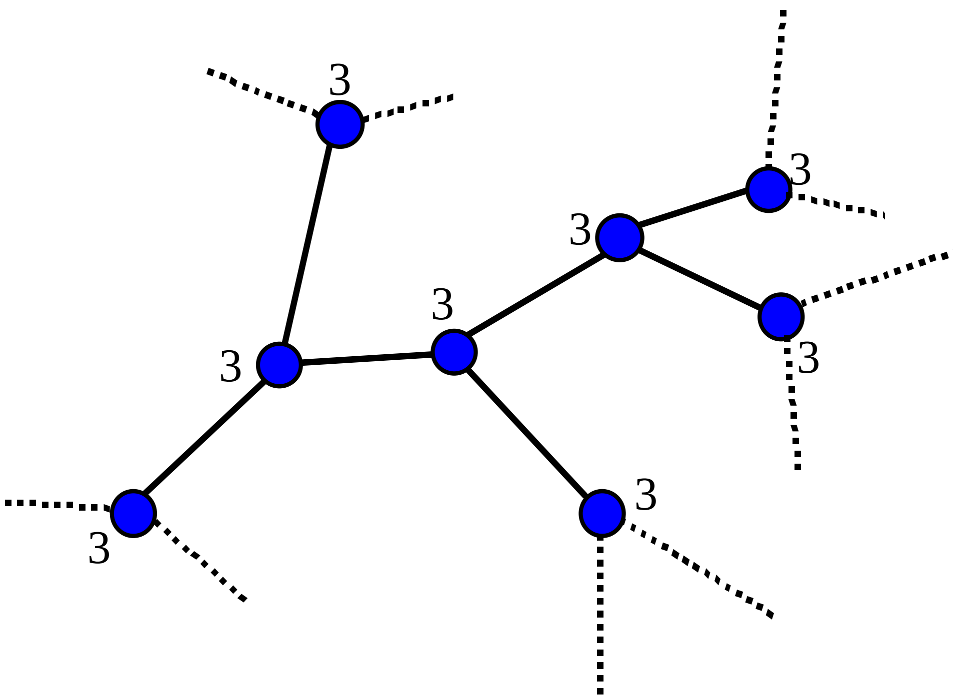
ComplexityScience.org
Thematic Institute Vienna 14.9.2004

www.AndreasKrueger.de/networks
networks@AndreasKrueger.de

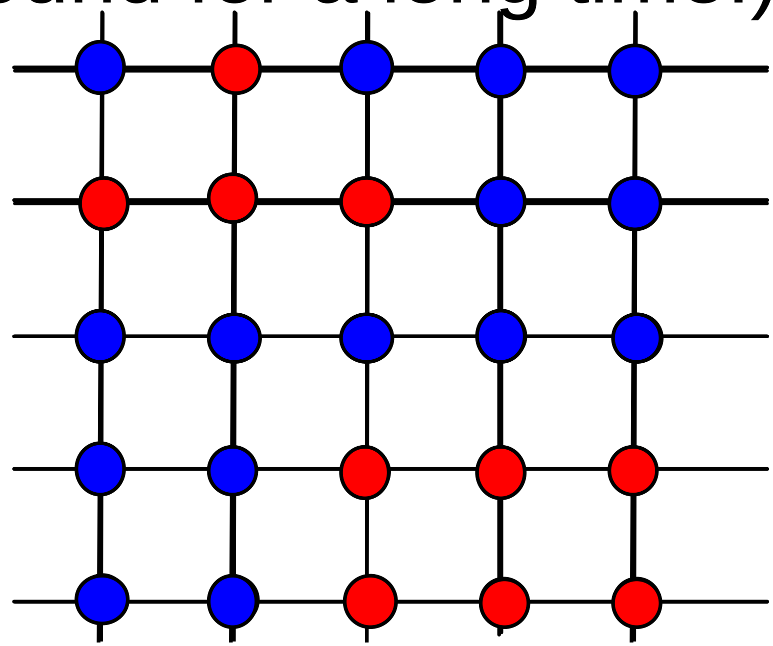
Structure of this quick'n dirty short talk

- Graphs in statistical mechanics models
- definitions of 3 graph measures that were our first important observables
- The „old“ graph model of Erdös-Renyi
- 1998 paradigm shift with random networks
  → new empirical findings
  → new random graph models
  → new community of physicists entering networks

Special cases of graphs: Trees, Lattices are *regular* graphs (and around for a long time!)



Caley-Tree with coordination number (degree) z=3
... branching process ...

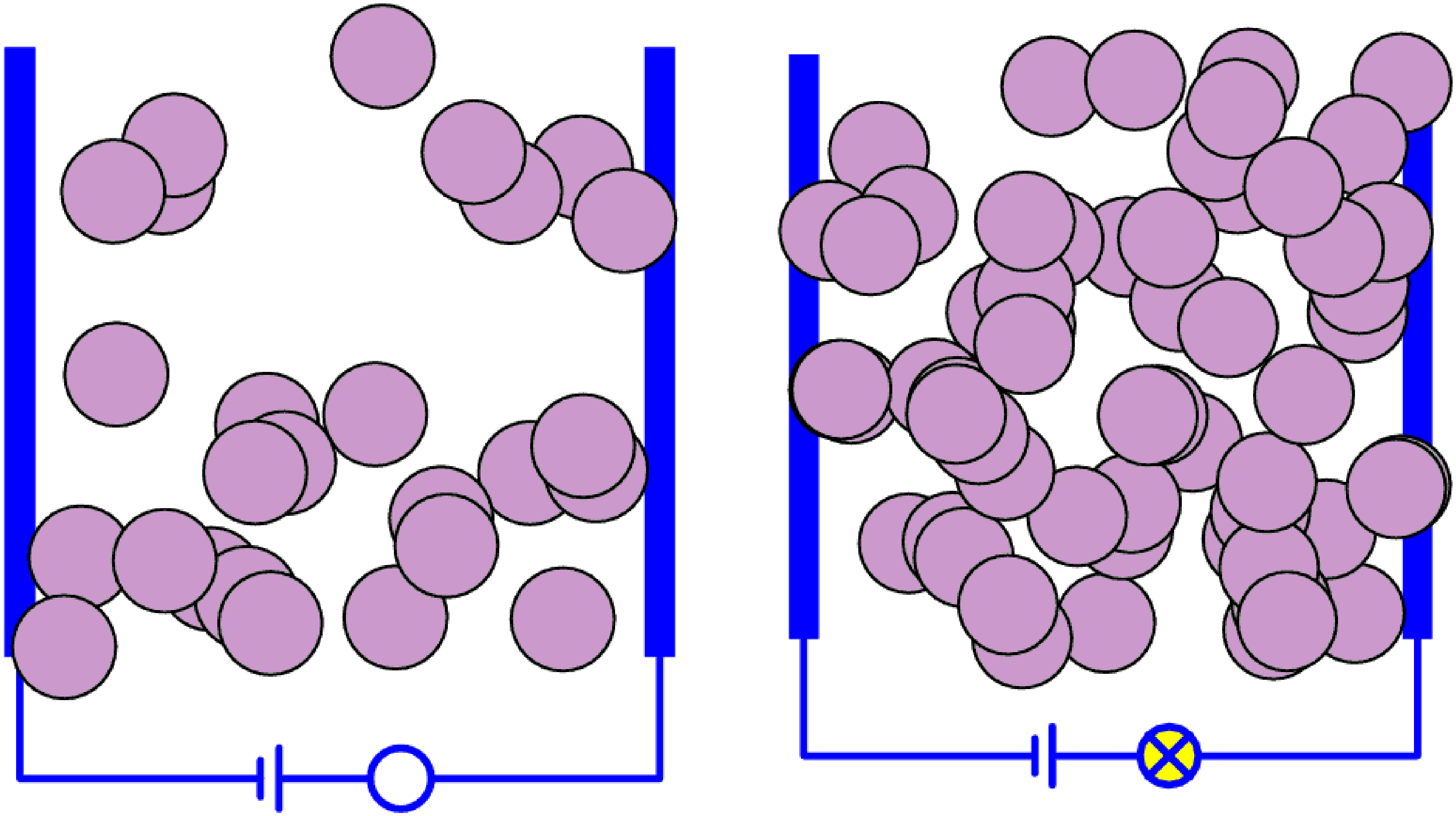


Lattice Z^2 - some processes:

- Ising (model magnet with *spins*)
- gauge field theory
  (elementary particle theory simulations)
- Self organized criticality (SOC)
- ...

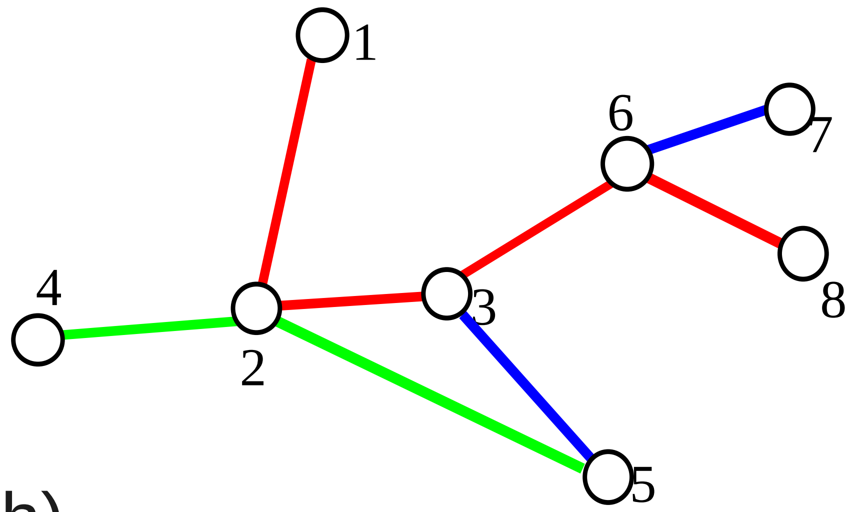
Percolation – geometric ***critical transition***

here: „continuum percolation“ (no lattice, but free coordinates)



www.AndreasKrueger.de/thesis

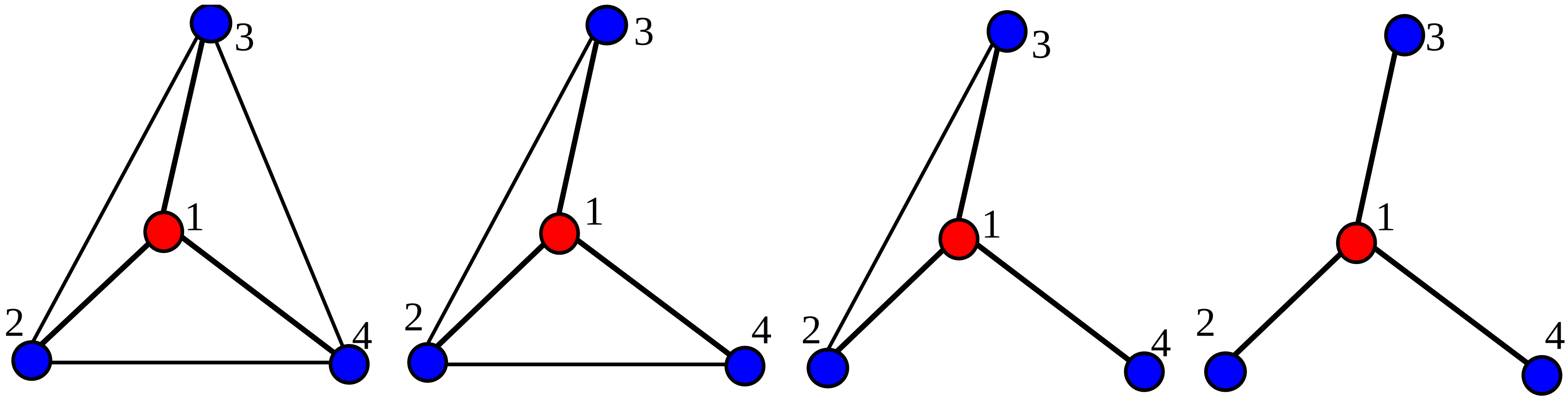
graph measures 1: diameter, paths



- pathlength (geodesic path)
  – **Shortest** connection between 2 nodes
  – Example pathlength(1,8) = 4

- global graph-properties
  – *Diameter* = **longest** geodesic path (here 4)
  – *characteristic pathlength* = Average of all paths (i,j)

graph measures 2: Cluster-Coefficient, Triangle Number



$C_1=1$ $C_1=2/3$ $C_1=1/3$ $C_1=0$

$$C_i = \frac{\#T_i}{k_i(k_i-1)/2}$$

$$C = \frac{1}{N}\sum_i C_i$$

(global)

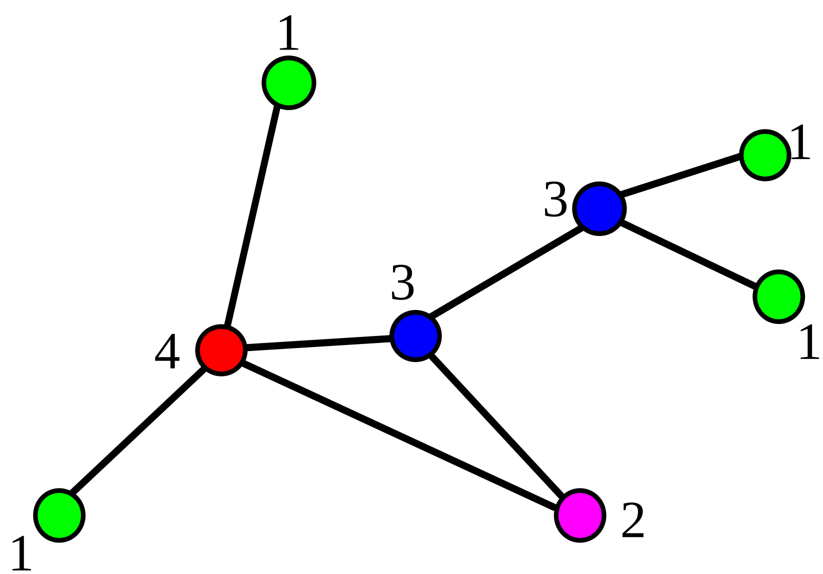
$\#T_i$ = Number of Triangles around vertex i

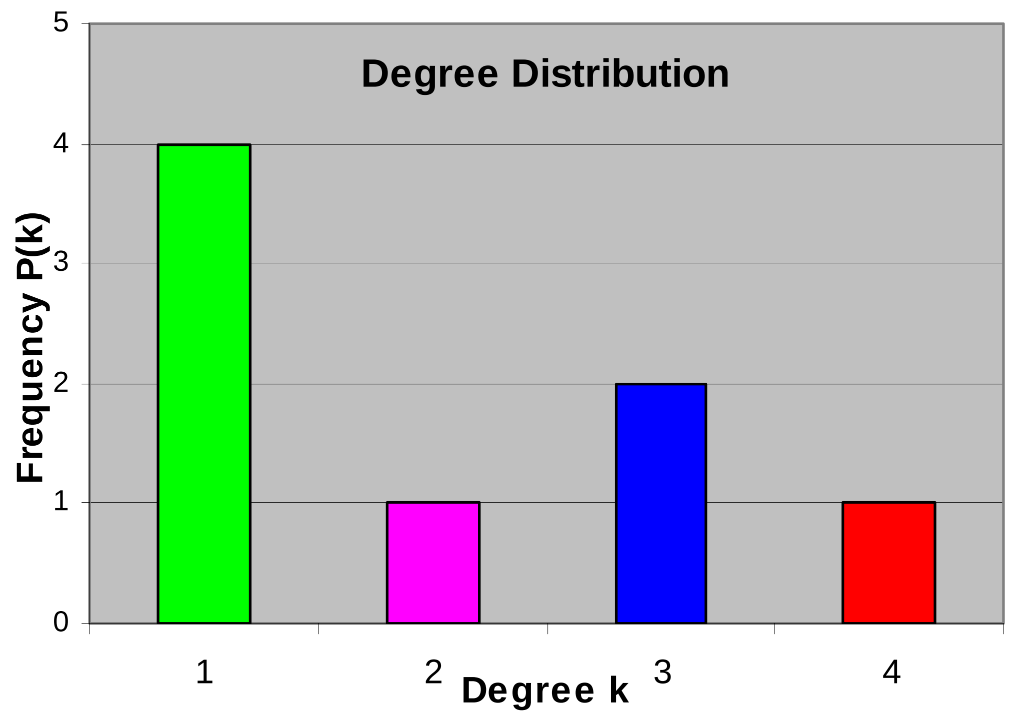
C_i: Estimator for **local density of connections**, "*how many of my friends are friends to each other?*"

graph measure 3: **degree** of a node

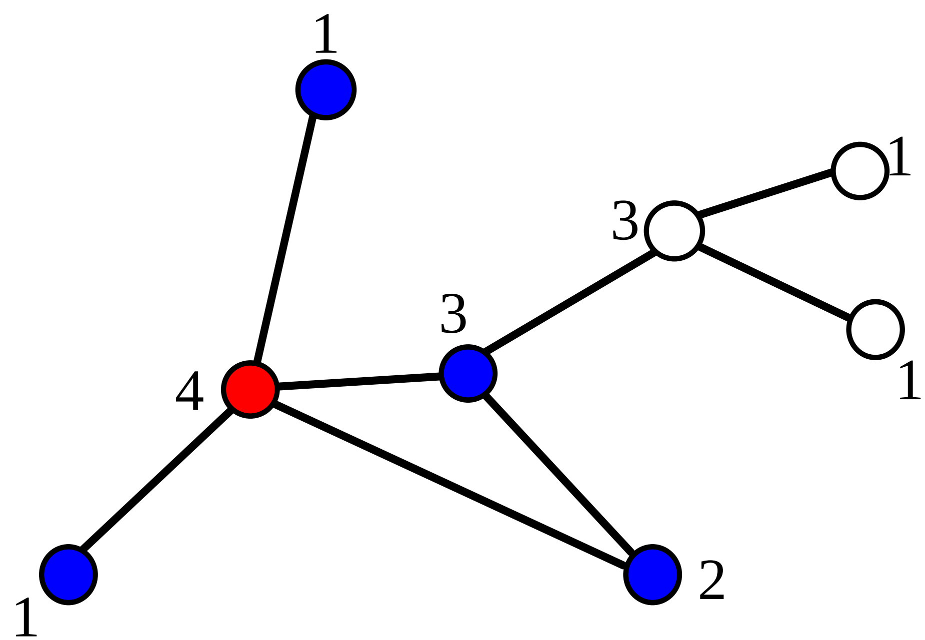
k_x = deg(x) = $|N_1(x)|$ = how many N_1-Neighbours has x

P(k) = *Degree-Distribution*= **number of nodes** with deg=k





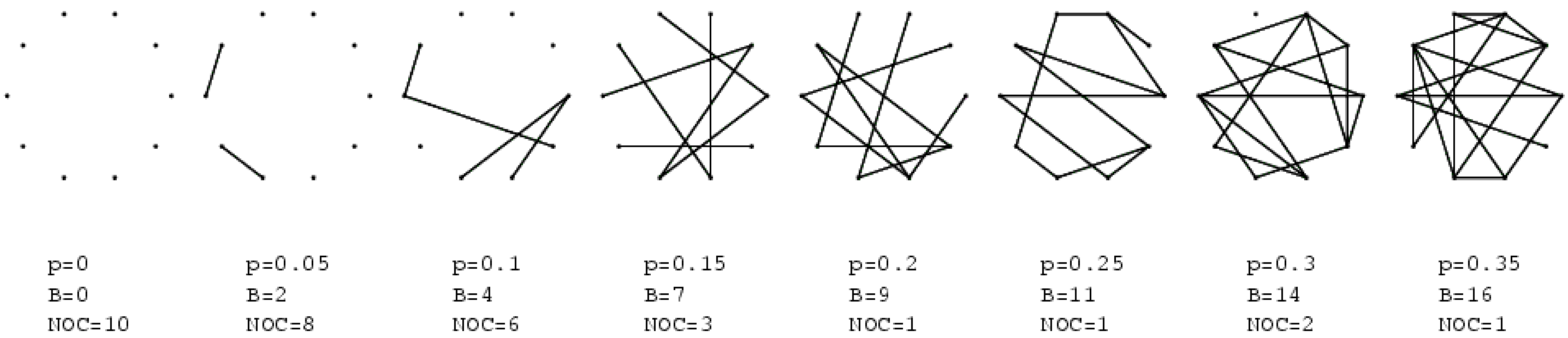
Hubs and *Authorities*



- *Hub* ~ high degree
  - e.g. plane traffic: Chicago, Frankfurt Main
  - E.g. mathematics: Erdös
- *Authority* ~ linked by a Hub

THE *random* graph model: Erdös Renyi RandomGraph (~19**60**)

- G(N,p) random graph
- N vertices → # possible edges: $M_{max} = \frac{N(N-1)}{2}$
- Independent Probability p for each edge (Bernoulli-process) $p \in [0,1]$



Degree Distribution with ER G(N,p) is ~ Poisson

$$\langle k \rangle = (N-1)\,p$$

$$= (N-1)\frac{M}{N(N-1)/2} = \frac{2M}{N}$$

$$= \mu$$

Average is good estimator for the whole distribution (bell shaped)

The degree has a binomial distribution. For N>>1 it becomes Poissonian:

$$P(k) = e^{-\mu}\frac{\mu^k}{k!}$$

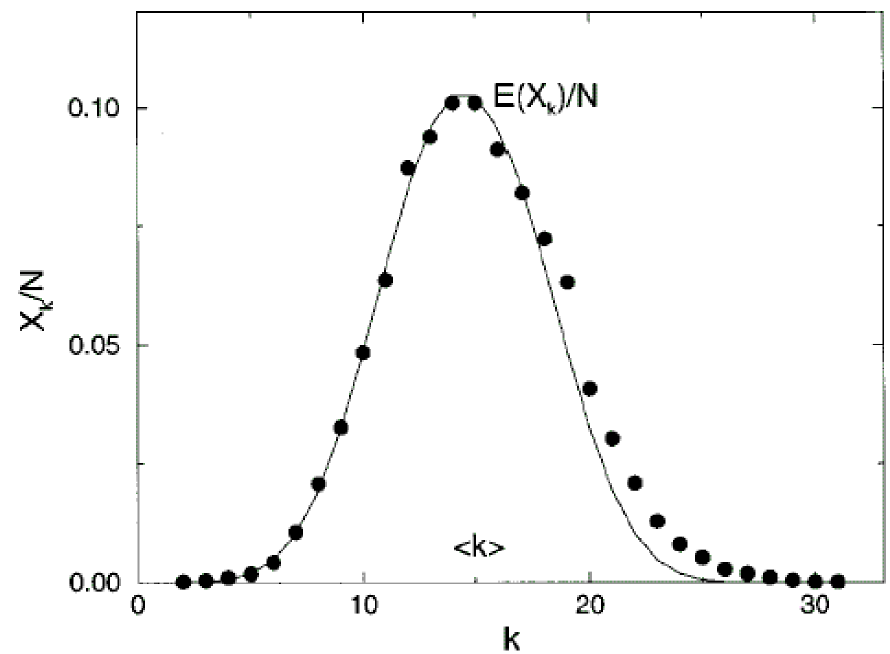


FIG. 7. The degree distribution that results from the numerical simulation of a random graph. We generated a single random graph with $N=10\,000$ nodes and connection probability $p=0.0015$, and calculated the number of nodes with degree k, X_k. The plot compares X_k/N with the expectation value of the Poisson distribution (13), $E(X_k)/N=P(k_i=k)$, and we can see that the deviation is small.

paradigm shift 1998/99
static random → grown networks

Starting Papers

- **Watts**, D. J. and **Strogatz** S. H.,
  Collective dynamics of small-world networks,
  1998.06.04 Nature, 393, 440.
- **Barabasi**, A.-L. and **Albert**, R.,
  Emergence of scaling in random networks,
  1999, Science 286, 509–512 .
- Albert, R., **Jeong**, H. and Barabasi, A.-L.,
  The diameter of the world-wide web,
  1999, Nature (London) 401, 130-131; cond-mat/9907038.
- Barabasi, A.-L., Albert, R., and Jeong, H.,
  Mean-field theory for scale-free random networks,
  1999, Physica A 272, 173–187.

- Barabasi, A.-L.,
  Linked: The New Science of Networks,
  Perseus, Cambridge, MA (2002).

Property 1: small world

1998: Watts-Strogatz random rewiring

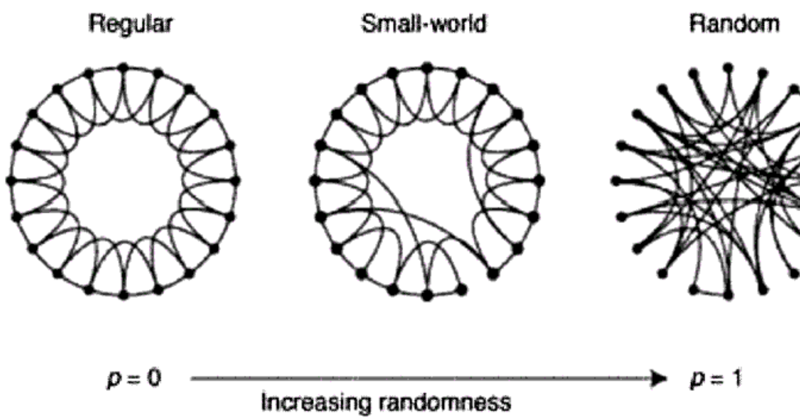


FIG. 15. The random rewiring procedure of the Watts-Strogatz model, which interpolates between a regular ring lattice and a random network without altering the number of nodes or edges. We start with $N=20$ nodes, each connected to its four nearest neighbors. For $p=0$ the original ring is unchanged; as p increases the network becomes increasingly disordered until for $p=1$ all edges are rewired randomly. After Watts and Strogatz, 1998.

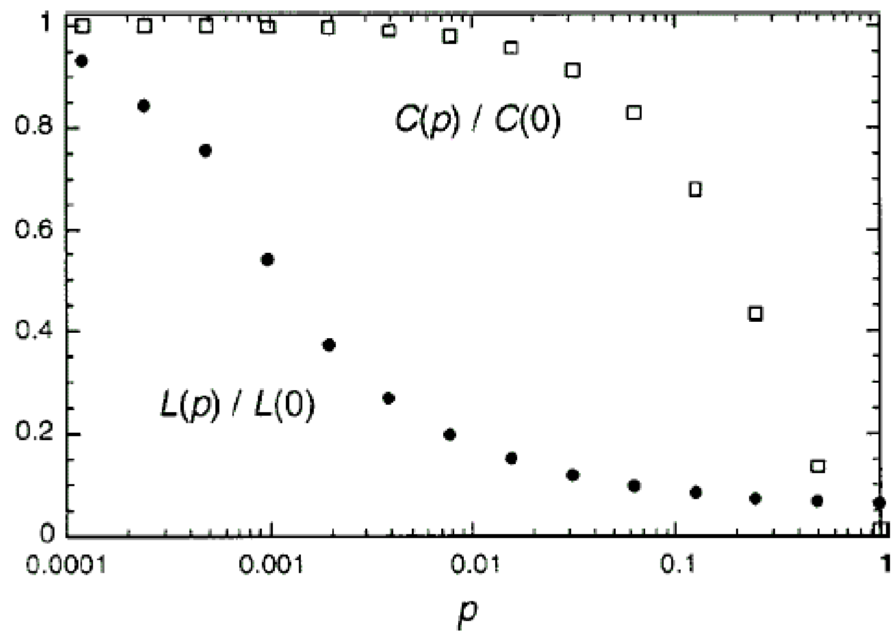


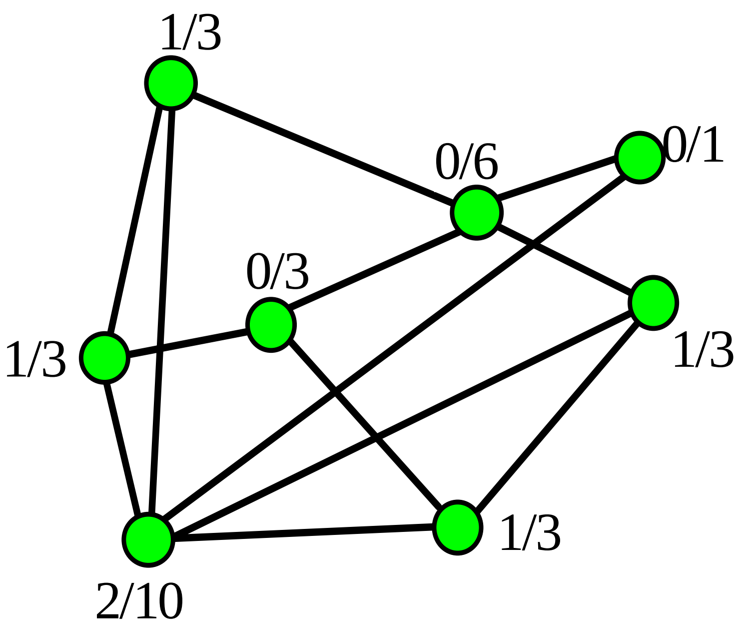
FIG. 16. Characteristic path length $\ell(p)$ and clustering coefficient $C(p)$ for the Watts-Strogatz model. The data are normalized by the values $\ell(0)$ and $C(0)$ for a regular lattice. A logarithmic horizontal scale resolves the rapid drop in $\ell(p)$, corresponding to the onset of the small-world phenomenon. During this drop $C(p)$ remains almost constant, indicating that the transition to a small world is almost undetectable at the local level. After Watts and Strogatz, 1998.

1967: Milgram “6 degrees of separation”

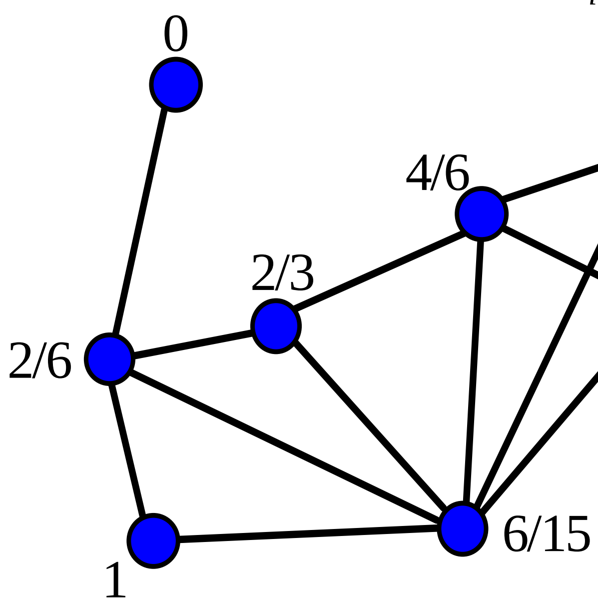
$L \sim \log N$

Property 2: high clustering

$$C_i = \frac{\#T_i}{k_i(k_i - 1)/2}$$



C=0.1917



C=0.6333

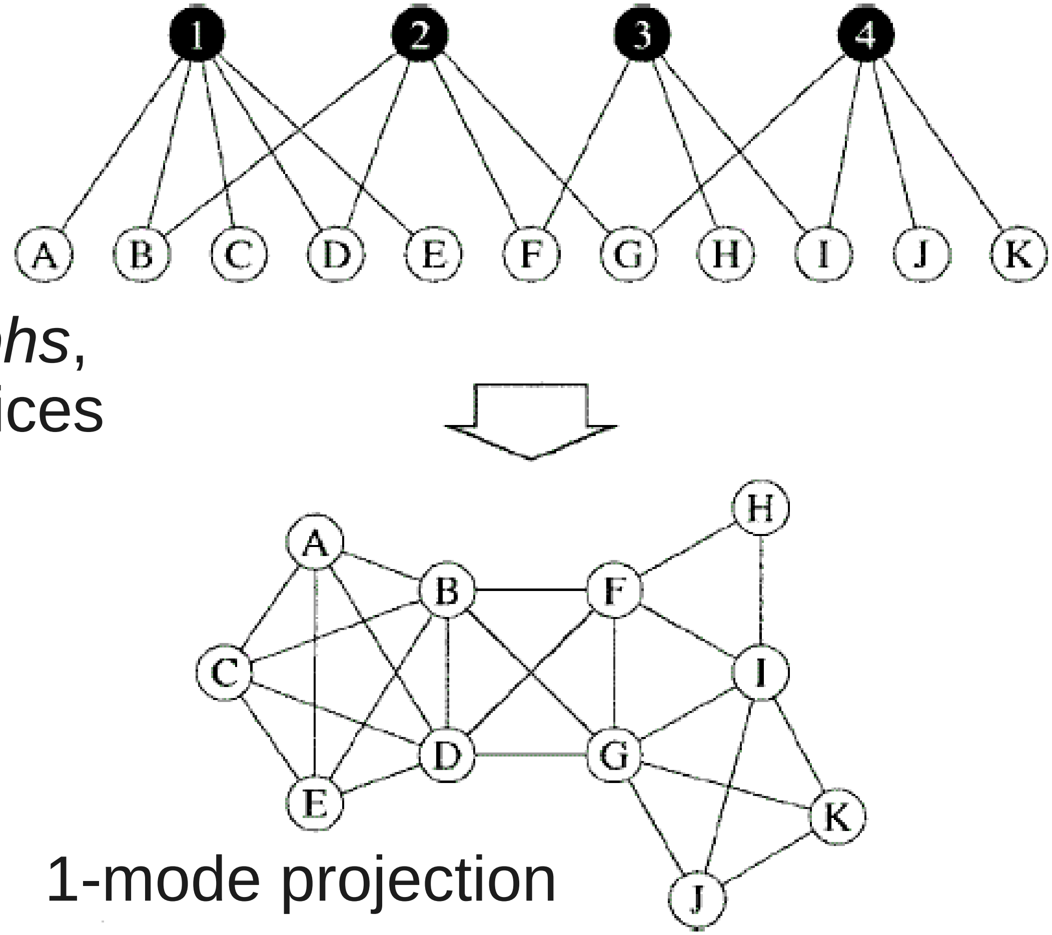
In both cases M=13 and N=8, but in the *right* picture many more friends are themselves direct friends to each other

! “Empirical Networks” have a significantly **higher clustering-coefficient** than ErdosRenyi-RandomGraphs !

Bipartite Graphs

Up to now we have only seen so called *1-mode graphs*, i.e. there is **one** type of vertices

Now imagine for example 4 **films** (black) and 11 playing **Actors** (white).



1-mode projection

FIG. 14. A schematic representation of a bipartite graph, such as the graph of movies and the actors who have appeared in them. In this small graph we have four movies, labeled 1 to 4, and eleven actors, labeled A to K, with edges joining each movie to the actors in its cast. The bottom figure shows the one-mode projection of the graph for the eleven actors. After Newman, Strogatz, and Watts (2001).

From the 2-mode graph we can generate a 1-mode graph by projection (under information loss)

Property 3: scale free

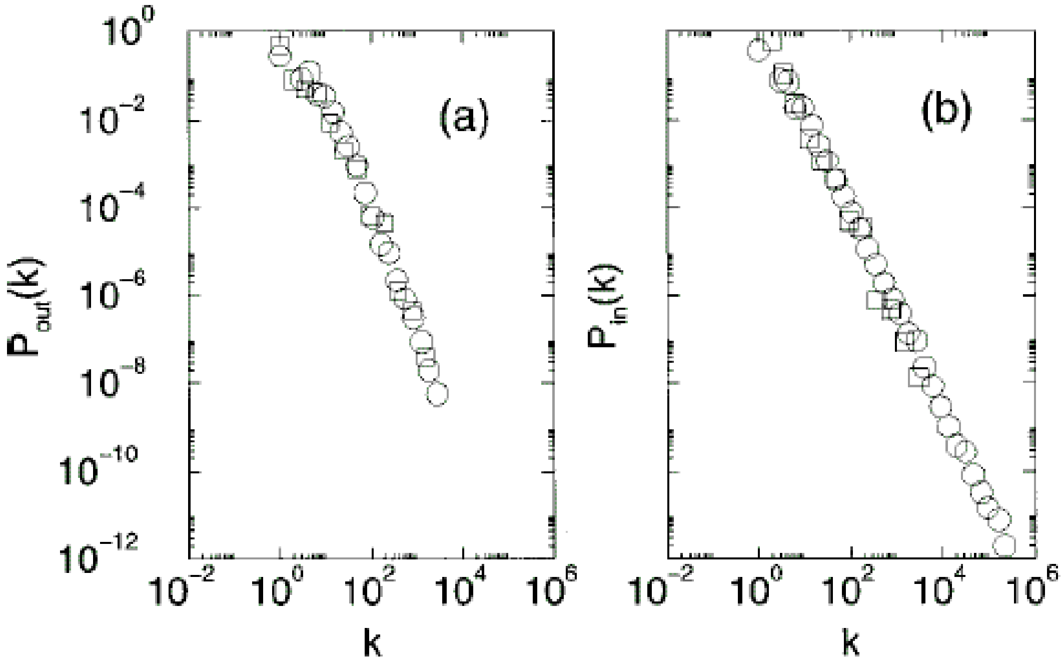


FIG. 2. Degree distribution of the World Wide Web from two different measurements: □, the 325 729-node sample of Albert *et al.* (1999); ○, the measurements of over 200 million pages by Broder *et al.* (2000); (a) degree distribution of the outgoing edges; (b) degree distribution of the incoming edges. The data have been binned logarithmically to reduce noise. Courtesy of Altavista and Andrew Tomkins. The authors wish to thank Luis Amaral for correcting a mistake in a previous version of this figure (see Mossa *et al.*, 2001).

MEASURED networks:

for k large:
degree distribution is **not** Poissonian (with exponential tail)

but "fat tail"
→ falling **power-law**

$$P(k) = \frac{1}{k^{\gamma}}$$

$$\gamma \sim 2.5$$

An average <k> doesn‘t really make sense here
= no *built-in scale*

→ „scale-free“

property 3: scale free

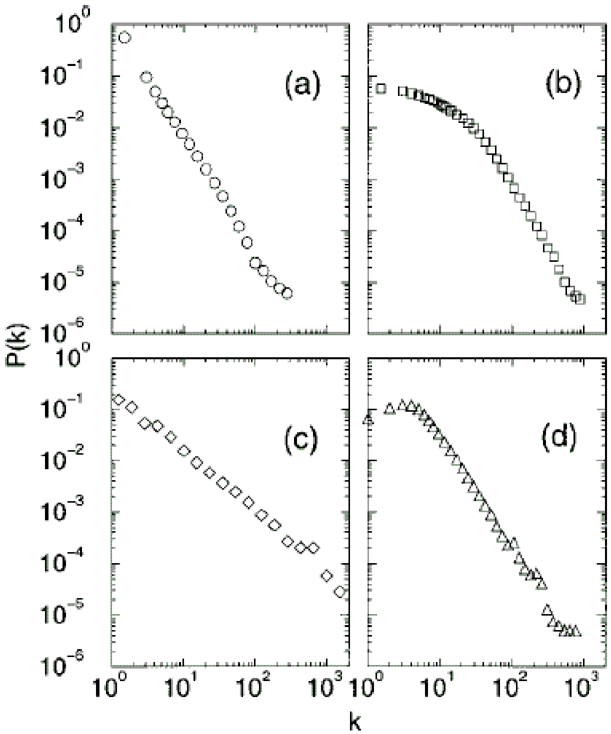


FIG. 3. The degree distribution of several real networks: (a) Internet at the router level. Data courtesy of Ramesh Govindan; (b) movie actor collaboration network. After Barabási and Albert 1999. Note that if TV series are included as well, which aggregate a large number of actors, an exponential cut-off emerges for large k (Amaral *et al.*, 2000); (c) co-authorship network of high-energy physicists. After Newman (2001a, 2001b); (d) co-authorship network of neuroscientists. After Barabási *et al.* (2001).

Now an incredible run on real life data started...

Almost identical scale-free distributions were measured in totally different objects, here e.g.

a) Internet Router

b) Actor-Movie-network

c) coauthors high energy physics

d) coauthors neuro sciences

! The <u>measurements(!)</u> almost perfectly lie on a straight line!

! And the power-law exponents differ only a little!

Property 3: scale free –
Some objects that seem to have a scale-free degree

- WWW
- Internet-Routing
- Protein-Protein-docking
- citations
- collaborations
  - publications
  - Movie-Actor-Network
- Human Sexuality Networks
- Telefone calls
- brains
  - Caenorhabditis elegans
  - Humans
- computer code
- The Word Web of language

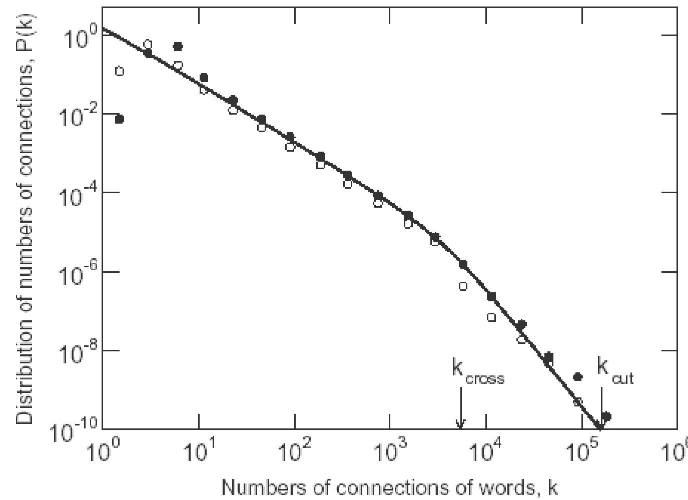


FIG. 9. The distribution of the numbers of connections (degrees) of words in the word web in a log-log scale [126]. Empty and filled circles show the distributions of the number of connections obtained in Ref. [126] for two different methods of the construction of the Word Web. The solid line is the result of theory of Ref. [127] (see Sec. IX J where the parameters of the Word Web, namely, the size $t \approx 470\,000$ and the average number of connections of a node, $\overline{k}(t) \approx 72$, were used. The arrows indicate the theoretically obtained point of crossover, k_{cross} between the regions with different power laws, and the cutoff k_{cut} due to the size effect. For a better comparison, the theoretical curve is displaced upward to exclude two experimental points with the smallest k (note that the comparison is impossible in the region of the smallest k where the empirical distribution essentially depends on the definition of the Word Web).

citations of the Phys. Rev. D **11-50** (1975-1994)	$24,296$	$351,872$	$\gamma_i = 3.0$	[27]
"————" (another fitting of the same data)			$\gamma_i = 2.6$	[99]
"————" (another estimate from the same data)			$\gamma_i = 2.3$	[94,95]
citations of the Phys. Rev. D (1982-June 1997)	—	—	$\gamma_i = 1.9$	[100]
collaboration network of movie actors	$212,250$	$61,085,555$	2.3	[55]
"————" (another fitting of the same data)			3.1	[102]
collaboration network of MEDLINE	$1,388,989$	$1.028\ 10^7$	2.5	[13]
collaboration net collected from mathematical journals	$70,975$	0.132×10^6	2.1	[15]
collaboration net collected from neuro-science journals	$209,293$	1.214×10^6	2.4	[15]
networks of metabolic reactions	$\sim 500 - 800$	$\sim 1500 - 3000$	$\gamma_i = 2.2$	[41]
			$\gamma_o = 2.2$	
net of protein-protein interactions (yeast proteome) [3]	1870	2240	~ 2.5	[44,45]
word web [4]	$470,000$	$17,000,000$	1.5	[126]
digital electronic circuits	2×10^4	4×10^4	3.0	[128]
telephone call graph [5]	47×10^6	8×10^7	$\gamma_i = 2.1$	[32]
web of human sexual contacts [6]	2810	—	3.4	[132]
food webs [7]	$93 - 154$	$405 - 366$	~ 1	[48,49]

TABLE I. Sizes and values of the γ exponent of the networks or subgraphs reported as having power-law (in-, out-) degree distributions. For each network (or class of networks) data are presented in more or less historical order, so that the recent exciting progress is visible. Errors are not shown (see the caption of Fig. 24). They depend on the size of a network and on the value of γ. We recommend our readers to look at the remark at the end of Sec. V C 2 before using these values. [1]The data for the network of operating AS was obtained for one of days in December 1999. [2]The value of the γ exponent was estimated from the degree distribution plot in Ref. [104]. [3]The network of protein-protein interaction is treated as undirected. [4]The value of the γ exponent for the word web is given for the range of degrees below the crossover point (see Fig. 9). [5]The out-degree distribution of the telephone call graph cannot be fitted by a power-law dependence (notice the remark in Sec. V F). [6]In fact, the data was collected from a small set of vertices of the web of human sexual contacts. These vertices almost surely have no connections between them. [7]These food webs are truly small. In Refs. [50,51] degree distributions of such food webs were interpreted as exponential-like.

First Model
Albert Barabasi: Preferential-Attachment

1. Growth ! Not in static systems...
   - Per time step **1 new node** and **m new edges**
2. Preferential Attachment
   - **y~x**: new node y, an old node x, but which one?
   - Probability to choose x linearly proportional to current degree of x:

$$\mathbf{P(\ deg(x)=k_i\ )\ =\ k_i\ /\ sum(k_i);}$$

„the rich get richer "

YES → scale-free, exponent gamma=3

YES → small world property

NO →high clustering

Some more topics: Topology analysis

- assortative (homophily) vs. dissassortative degree-degree correlations (human vs. technology/nature)
- community clustering algorithms (dozens of); modularity measure for ideal partitioning
- failure vs. attack:
  giant component / percolative situation
  (~any node can reach ~any node)
  destroyed difficult vs. easily

Some more topics:
Processes on networks

- Ising model, percolation
- Synchronisation, Voter models
- Epidemiology:
  How do infections spread on networks?
  - Already in the first Watts/Strogatz 1998 paper:
    infection time until all infected
    ~ mean path length L ~ log(N) in small worlds !
  - Infection threshold $\mathbf{A_c}$:
    - on ER random graph (or lattice) there is a positive critical infection rate $\mathbf{A_c}>0$ below which the epidemy dies out and above which the population dies out
    - BUT on scale-free graphs with gamma > 3
      epidemies are possible which don‘t die out with $\mathbf{A_c}\sim 0$!

This can only be a *very short* glance

- In a way, the whole subject was been born out of the Internet, which is not a small but at least a medium sized system (e.g. ~ 10^{10} webpages) and can be measured much easier than nature or society
- Our explanatory approach usually goes for the „thermodynamical limit“ of N→infinity; very different from the sociological viewpoint
- Within 6 years, some 3000(?) papers have been published about networks in the physics community
  `pre-prints on www.arxiv.org → cond-mat`
- Many scientists leave their (neighbouring) field and do research on networks now
- It resembles a little bit the hype of the 80ies „Fractals/Nonlinearity/Chaos„ – everything was a fractal back then, now everything is networks

Many thanks
for your attention!

networks@AndreasKrueger.de
www.AndreasKrueger.de/networks

<u>Suggestions for reading:</u>

Linked: The New Science of Networks,
Barabasi, A.-L.; Perseus Cambridge MA (**2002**).
Written very fluffily & easy to read, “Prosa”

Statistical Mechanics of Complex Networks
Reka Albert and Albert-Laszlo Barabasi, **2001**
arXiv:**cond-mat/0106096** **(www.arxiv.org)**
54 pages review-article; not new, but a very good introduction