Outline

Mon

Introduction, basic concepts, random networks

Tue

<u>Small-world networks, Scale-free networks</u>

Wed

Advanced network analysis, weighted & social networks

Thu

Weighted & social nets (cont'd), percolation on networks

Fri

Communities and modularity, Dynamic networks

E-R Random Networks: The Degree Distribution

- "If a random vertex is picked, what is the probability that its degree equals k?"

- Denote by $p_i(k)$ the probability that vertex i has degree k

- For the whole network

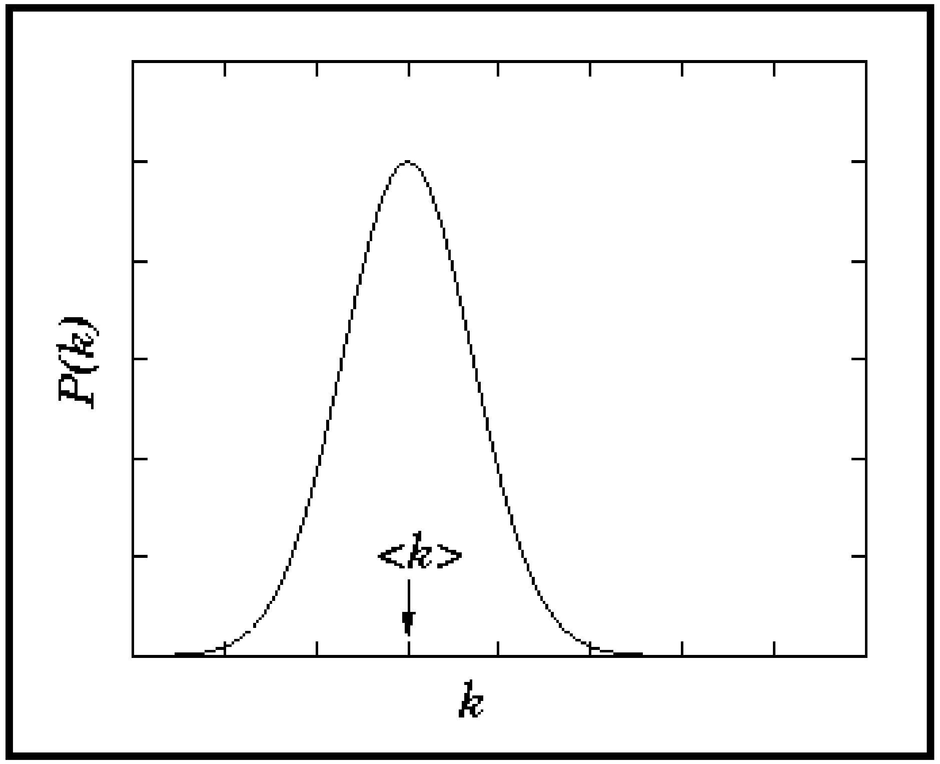
$$P(k) = \frac{1}{N} \sum_{i=1}^{N} p_i(k)$$

- For E-R networks, all vertices are alike, so $p_i(k) = P(k)$ for all i

- Erdös-Rényi networks:

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

(for large N)



$$\langle k \rangle = \frac{2\langle E \rangle}{N} = p\,(N-1) \approx pN$$

Average shortest path length in E-R networks

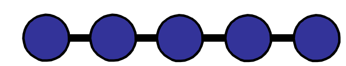
- Let us assume that there is a single connected component (a strong assumption, we'll get back to this)

- Then for E-R networks

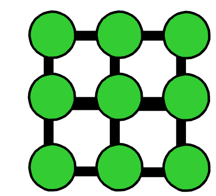
$$\langle l \rangle \propto \ln N$$

- The path length grows very slowly with network size; paths are short even for very large E-R networks

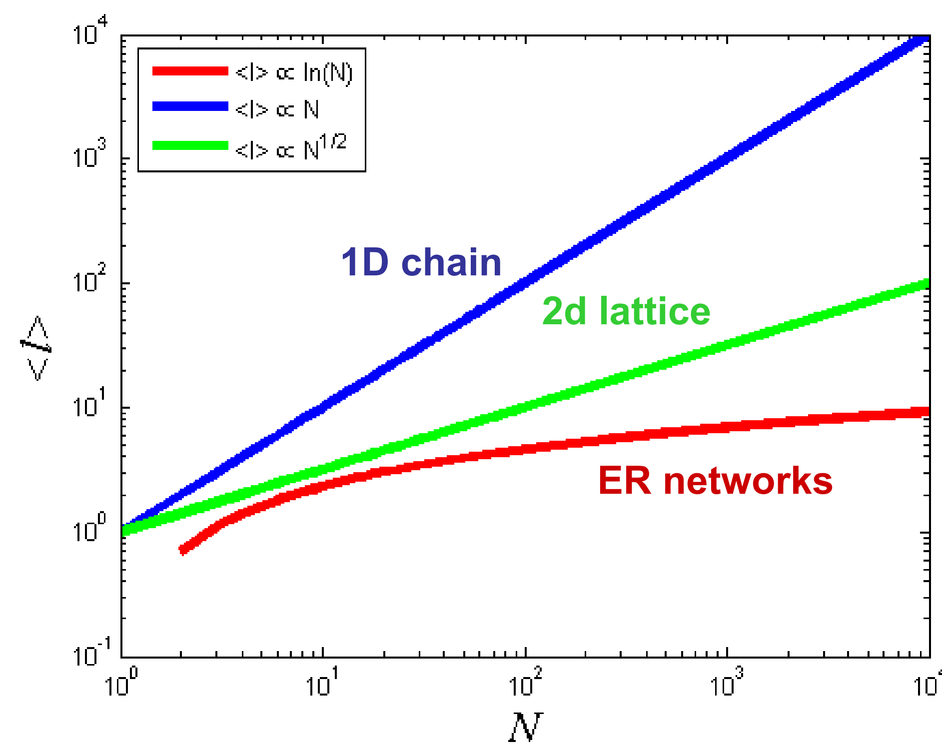
for comparison:



$$\langle l \rangle \propto N$$



$$\langle l \rangle \propto N^{1/2}$$



"E-R networks are infinite-dimensional"

It’s A Small World!

How Far Are You From Anyone Else?

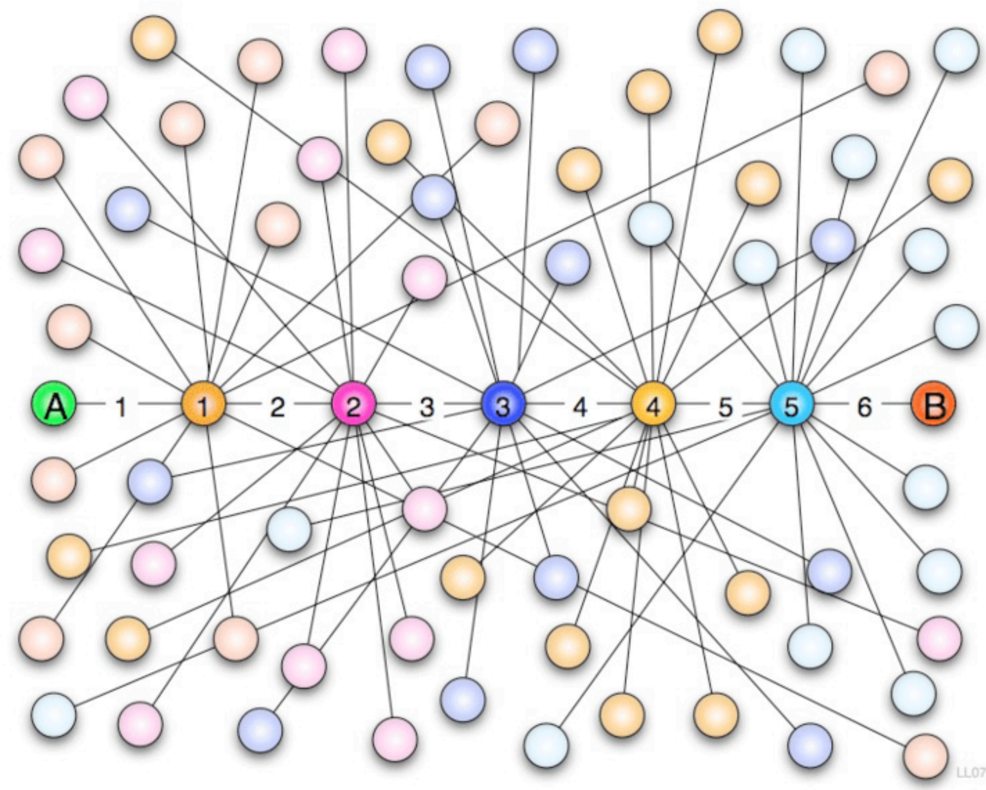
Frigyes Karinthy: Chains (1929)

- Classic short story
- Karinthy believed that the modern world was 'shrinking' due to ever-increasing connectedness of human beings
- Excerpt: "A fascinating game grew out of this discussion. One of us suggested performing the following experiment to prove that the population of the Earth is closer together now than they have ever been before. We should select any person from the 1.5 billion inhabitants of the Earth—anyone, anywhere at all.
  He bet us that, using no more than five individuals, one of whom is a personal acquaintance, he could contact the selected individual using nothing except the network of personal acquaintances."
- These ideas had a strong influence on social sciences

Six Degrees of Separation

Stanley Milgram,
"The Small World Problem",
Psychology Today, 1967, Vol. 2, 60-67

- Milgram picked 296 individuals in Nebraska and Boston
- Everyone was given a letter to be delivered to a target individual in Massachusetts
- Instructions: "If you know the target person, give the letter to him, otherwise give it to someone who you think is closer to the target."
- 64 letters reached the target, through 5.2 intermediaries



The Birth of Complex Networks Science:
Small-World Networks

D.J. Watts and S. Strogatz,
”Collective dynamics of 'small-world' networks",
Nature **393**, 440–442, 1998

- This paper practically launched the science of complex networks
- Elaborates on the topic of short path lengths
- Probably the simplest model ever published in Nature
- Has been cited >3500 times!!


letters to nature

typically slower than ~1 km s^{-1}) might differ significantly from what is assumed by current modelling efforts[27]. The expected equation-of-state differences among small bodies (ice versus rock, for instance) presents another dimension of study; having recently adapted our code for massively parallel architectures (K. M. Olson and E.A, manuscript in preparation), we are now ready to perform a more comprehensive analysis.

The exploratory simulations presented here suggest that when a young, non-porous asteroid (if such exist) suffers extensive impact damage, the resulting fracture pattern largely defines the asteroid's response to future impacts. The stochastic nature of collisions implies that small asteroid interiors may be as diverse as their shapes and spin states. Detailed numerical simulations of impacts, using accurate shape models and rheologies, could shed light on how asteroid collisional response depends on internal configuration and shape, and hence on how planetesimals evolve. Detailed simulations are also required before one can predict the quantitative effects of nuclear explosions on Earth-crossing comets and asteroids, either for hazard mitigation[28] through disruption and deflection, or for resource exploitation[29]. Such predictions would require detailed reconnaissance concerning the composition and internal structure of the targeted object. □

Received 4 February; accepted 18 March 1998.

Collective dynamics of 'small-world' networks

Duncan J. Watts* & Steven H. Strogatz

Department of Theoretical and Applied Mechanics, Kimball Hall, Cornell University, Ithaca, New York 14853, USA

Networks of coupled dynamical systems have been used to model biological oscillators[1–4], Josephson junction arrays[5,6], excitable media[7], neural networks[8–10], spatial games[11], genetic control networks[12] and many other self-organizing systems. Ordinarily, the connection topology is assumed to be either completely regular or completely random. But many biological, technological and social networks lie somewhere between these two extremes. Here we explore simple models of networks that can be tuned through this middle ground: regular networks 'rewired' to introduce increasing amounts of disorder. We find that these systems can be highly clustered, like regular lattices, yet have small characteristic path lengths, like random graphs. We call them 'small-world' networks, by analogy with the small-world phenomenon[13,14] (popularly known as six degrees of separation[15]). The neural network of the worm *Caenorhabditis elegans*, the power grid of the western United States, and the collaboration graph of film actors are shown to be small-world networks. Models of dynamical systems with small-world coupling display enhanced signal-propagation speed, computational power, and synchronizability. In particular, infectious diseases spread more easily in small-world networks than in regular lattices.

To interpolate between regular and random networks, we consider the following random rewiring procedure (Fig. 1). Starting from a ring lattice with n vertices and k edges per vertex, we rewire each edge at random with probability p. This construction allows us to 'tune' the graph between regularity ($p = 0$) and disorder ($p = 1$), and thereby to probe the intermediate region $0 < p < 1$, about which little is known.

We quantify the structural properties of these graphs by their characteristic path length $L(p)$ and clustering coefficient $C(p)$, as defined in Fig. 2 legend. Here $L(p)$ measures the typical separation between two vertices in the graph (a global property), whereas $C(p)$ measures the cliquishness of a typical neighbourhood (a local property). The networks of interest to us have many vertices with sparse connections, but not so sparse that the graph is in danger of becoming disconnected. Specifically, we require $n \gg k \gg \ln(n) \gg 1$, where $k \gg \ln(n)$ guarantees that a random graph will be connected[16]. In this regime, we find that $L \sim n/2k \gg 1$ and $C \sim 3/4$ as $p \to 0$, while $L \approx L_{random} \sim \ln(n)/\ln(k)$ and $C \approx C_{random} \sim k/n \ll 1$ as $p \to 1$. Thus the regular lattice at $p = 0$ is a highly clustered, large world where L grows linearly with n, whereas the random network at $p = 1$ is a poorly clustered, small world where L grows only logarithmically with n. These limiting cases might lead one to suspect that large C is always associated with large L, and small C with small L.

On the contrary, Fig. 2 reveals that there is a broad interval of p over which $L(p)$ is almost as small as L_{random} yet $C(p) \gg C_{random}$. These small-world networks result from the immediate drop in $L(p)$ caused by the introduction of a few long-range edges. Such 'short cuts' connect vertices that would otherwise be much farther apart than L_{random}. For small p, each short cut has a highly nonlinear effect on L, contracting the distance not just between the pair of vertices that it connects, but between their immediate neighbourhoods, neighbourhoods of neighbourhoods and so on. By contrast, an edge

*Present address: Paul F. Lazarsfeld Center for the Social Sciences, Columbia University, 812 SIPA Building, 420 W118 St, New York, New York 10027, USA.

440 Nature © Macmillan Publishers Ltd 1998 NATURE | VOL 393 | 4 JUNE 1998

Small Worlds: Path Lengths Revisited

- E-R random networks: shortest path lengths

$$\langle l \rangle \propto \ln N$$

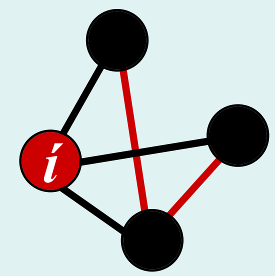
- Compatible with the six-degrees idea
- Turns out that path lengths in real-world networks are indeed short!
- Example networks in Watts & Strogatz (1998):
  - Film actors: movie collaborations from the Internet Movie Database
  - US power grid
  - Neurons of C. Elegans, a tiny worm which is one of the model organisms in biology

Network	N	$<l>$	$<l_{rand}>$
film actors	22500	**3.65**	2.99
US power grid	4941	**18.7**	12.4
C. Elegans neurons	282	**2.65**	2.25

- N = network size
- $<l>$ = avg shortest path length
- $<l_{rand}>$ = -”- in E-R networks with same N, $<k>$

The Clustering Coefficient

- Measures "cliquishness" in local network neighbourhoods, i.e. deviations from randomness
- "What is the probability that two of my friends are also friends?"
- Defined for node i (of degree k_i) as the number of links between neighbours divided by the possible number of such links $k_i(k_i\text{-}1)/2$
- $C_i \in [0,1]$



$$C_i = \frac{2 \times 2}{3\,(3-1)} = \frac{2}{3}$$

$$\boxed{C_i = \frac{2E_i}{k_i\,(k_i-1)}}$$

E_i = # of edges among i's neighbours

$$\boxed{C(k) = \frac{1}{N_k} \sum_{i,k_i=k} C_i}$$

average of C_i for vertices of degree k

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

average of C_i over the network

Clustering Coefficient for E-R Networks

- The clustering coefficient

$$\boxed{C_i = \frac{2E_i}{k_i\left(k_i - 1\right)}}$$

can be viewed as the probability that two neighbours of *i* are connected

- For E-R networks

$$C_{ER} = C_{i,ER} = p = \frac{\langle k \rangle}{N}$$

(remember $<k>=pN$)

- Hence

$$\lim_{N \to \infty} C_{ER} = \frac{\langle k \rangle}{N} = 0$$

(limit taken such that $<k>$=const)

- **Large random networks are treelike - there is practically no clustering!**

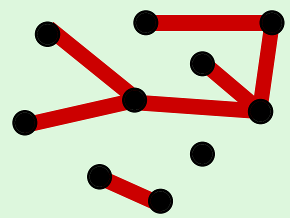
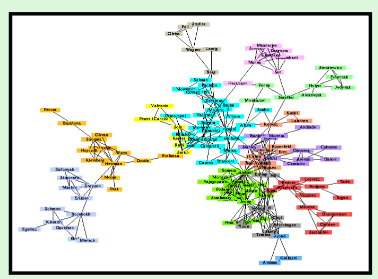
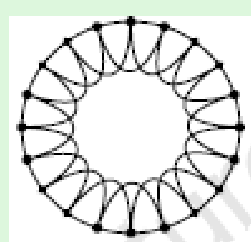
Clustering in Real-World Networks

Network	N	$<l>$	$<l_{rand}>$	$<C>$	$<C_{rand}>$
film actors	22500	3.65	2.99	**0.79**	0.00027
US power grid	4941	18.7	12.4	**0.08**	0.006
C. Elegans neurons	282	2.65	2.25	**0.28**	0.06

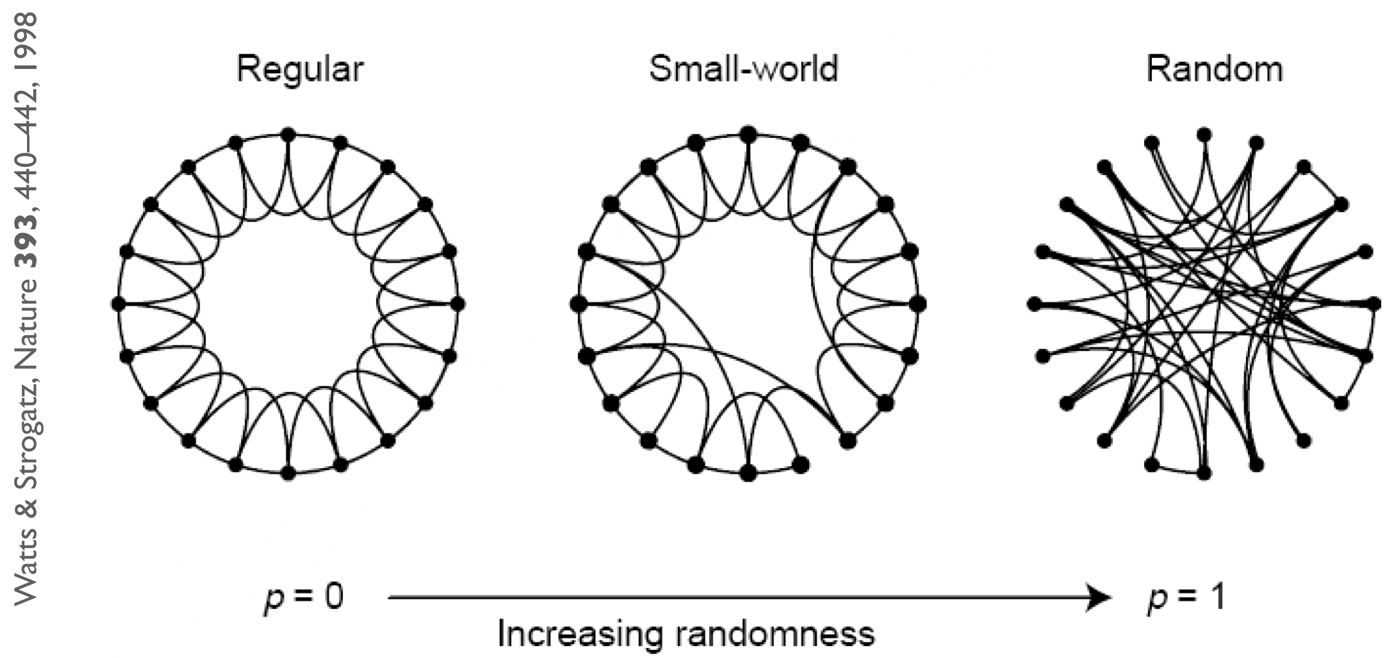
Average clustering coefficient values are orders of magnitude higher than in E-R networks!

Note: Nowadays one would not use E-R networks as the random reference, but construct random networks with the same degree sequence than in data. This would not change the results here, though.

Summary: Path Lengths and Clustering in Real-World Networks

	shortest paths	clustering
Random networks 	short	low
Real-world networks 	**short**	**high**
Regular networks 	long	high

The Watts-Strogatz Small World Model



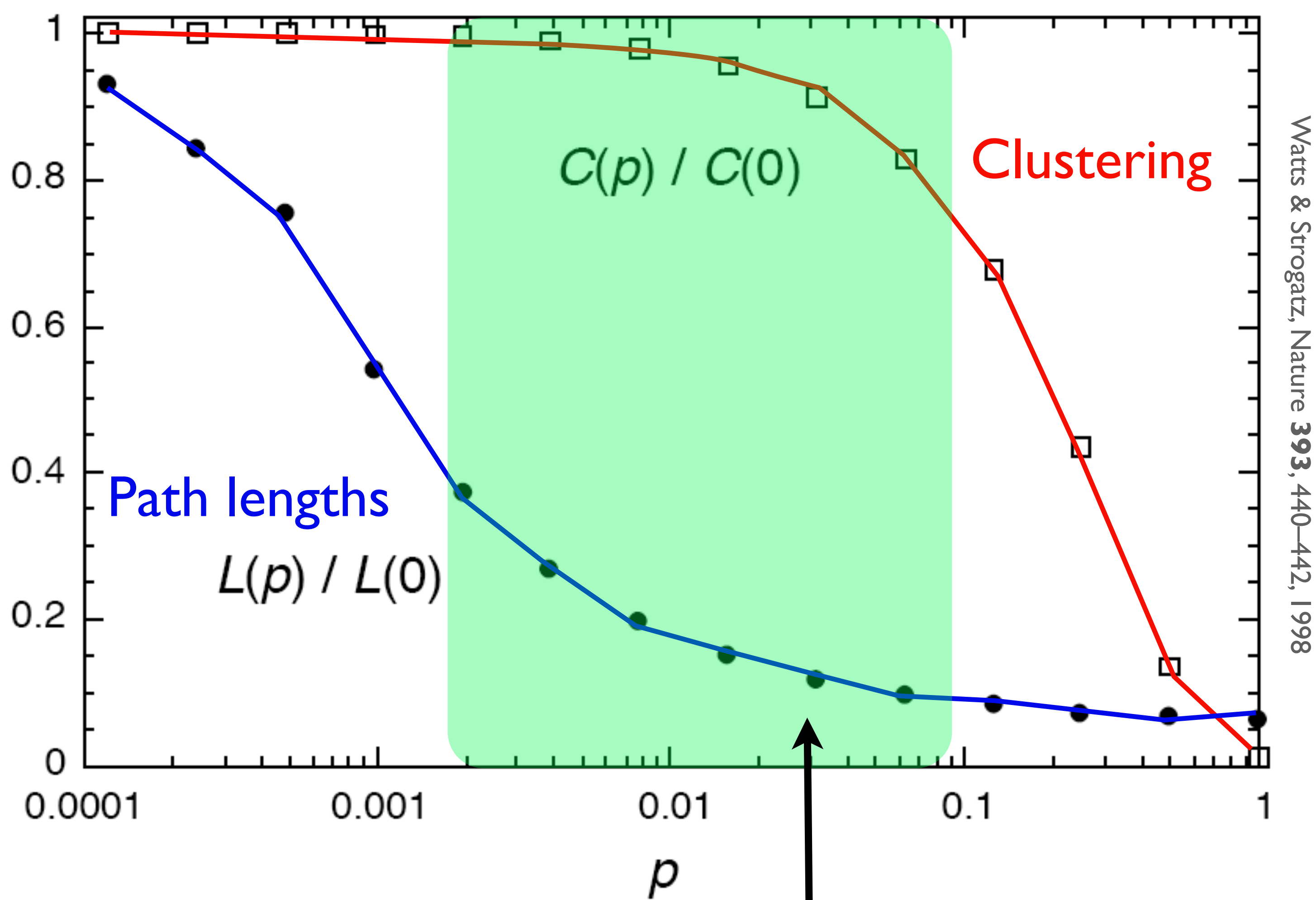
- A simple model for interpolating between regular and random networks

- Randomness controlled by a single tuning parameter

The model:

- Take a regular clustered network
- Rewire the endpoint of each link to a random node with probability p

Path lengths and clustering in the WS model



The "Small-World" regime: paths short, clustering high

Dynamics on Small-World Networks

- Crucial observation: **network structure heavily affects processes taking place on networks!**

  Examples:

- Spreading or contact processes: small number of shortcuts enormously speeds up the process
- Synchronization: shortcuts give rise to rapid synchronization of oscillators

1) Many real-world processes take place on networks:
  - disease spreading
  - computer virus spreading
  - cascading power grid failures
  - information transmission
  - metabolism, genetic regulation
  - everything inside your brain

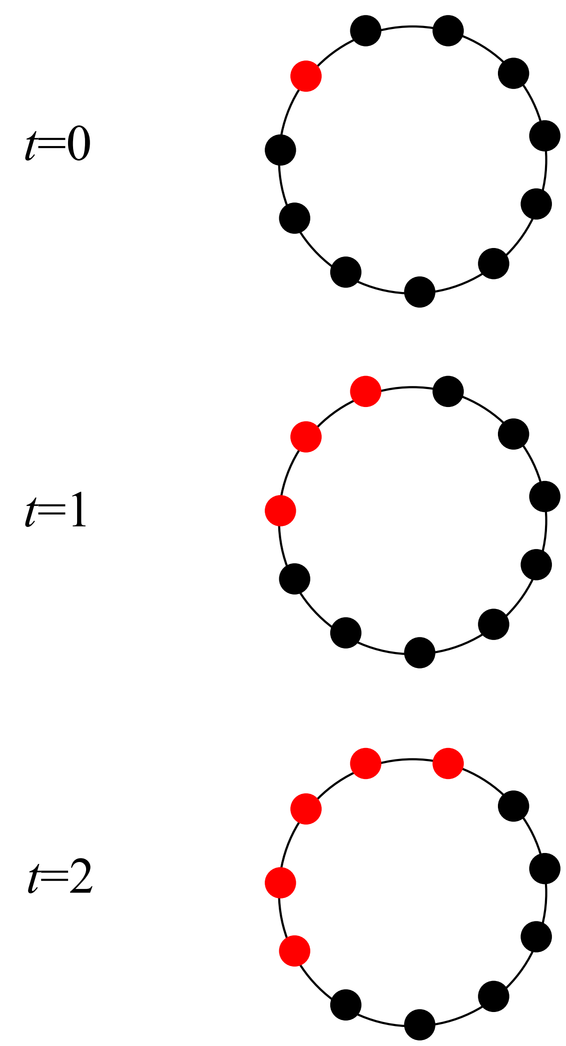
2) Network structure has important consequences on such processes

1) & 2) → **To understand such processes it is imperative to understand the underlying networks!**

Example: Dynamics of Spreading (SI)

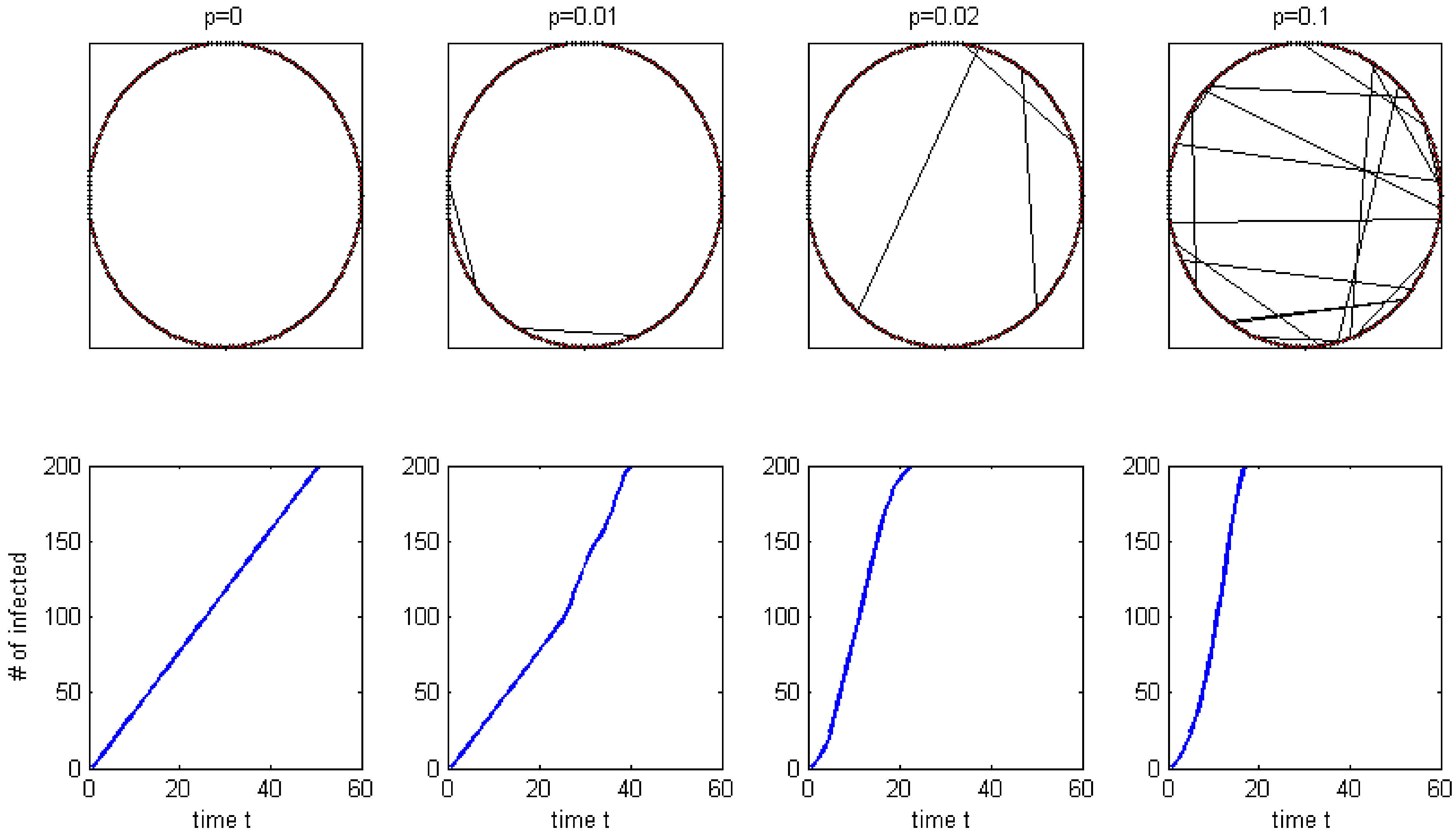
- Simplest possible model: Susceptible-Infected (SI)
- Initially all nodes susceptible
- Introduce an infected node
- Each infected node infects its susceptible neighbours, each with probability *p* per time step

SI in a regular ring lattice



Example: Dynamics of Spreading (SI)

SI in a regular ring with added shortcuts



number of infected

time

Example: Dynamics of Spreading (SIR)

- More realistic model: Susceptible-Infected-Recovered (SIR)

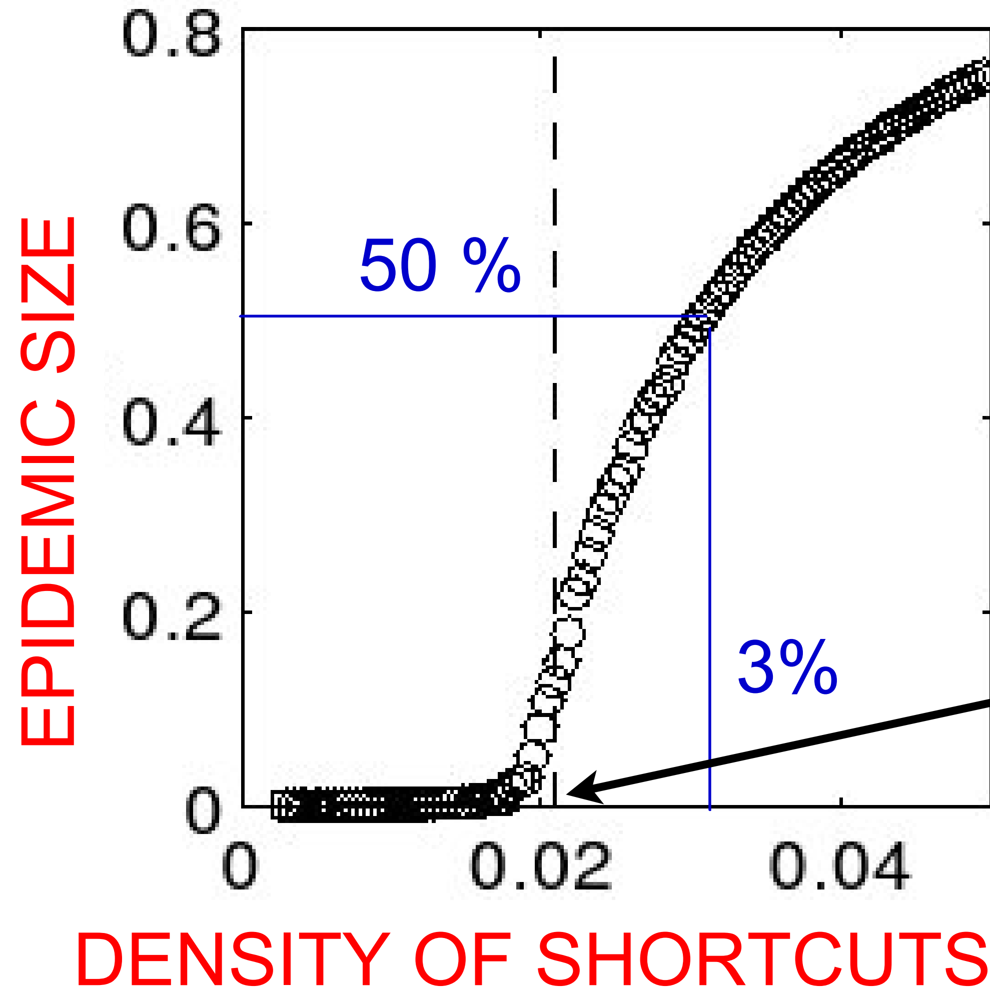
- Each infected node infects its susceptible neighbours, each with probability p per time step

- Each infected node recovers and becomes immune with probability q per time step

- Three different outcomes:
  1) the disease dies out,
  2) the disease becomes endemic (doesn't die, doesn't spread),
  3) a large fraction of the system gets infected

- These depend on p, q, and network structure

Example: Dynamics of Spreading (SIR)



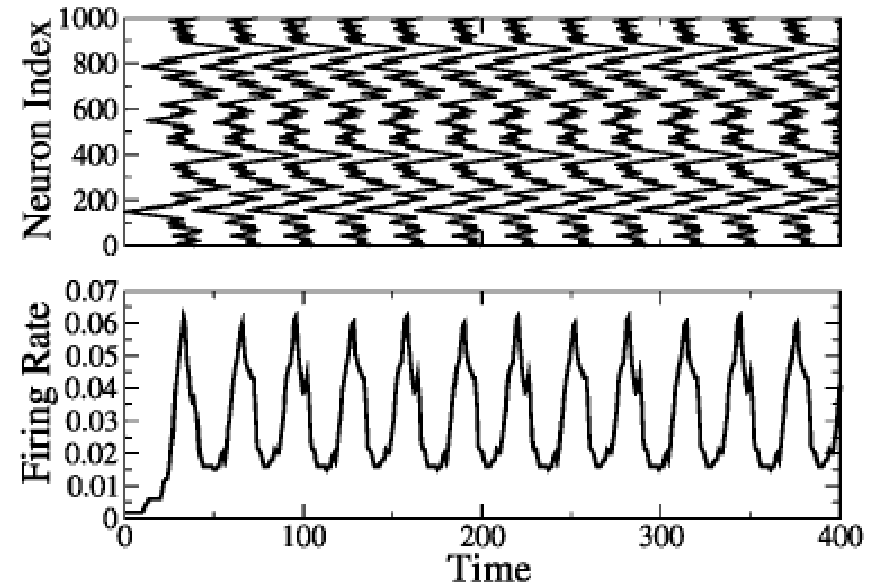
At a critical fraction of shortcuts, the disease suddenly becomes an epidemic

Note: this fraction depends on the spreading model's parameters

Example: Dynamics of Neurons

- Roxin et al, *Phys. Rev. Lett.* **92**, 198101 (2004)
- Neurons obey integrate-and-fire dynamics:
  - Input from neighbours excites neuron
  - Excitement fades with time
  - If excited enough, fire a signal to neighbours

$$\tau_m \frac{dV_i}{dt} = -V_i + I_{\mathrm{ext}} + g_{\mathrm{syn}} \sum_{j,m} w_{ij} \delta(t - t_j^{(m)} - \tau_D)$$

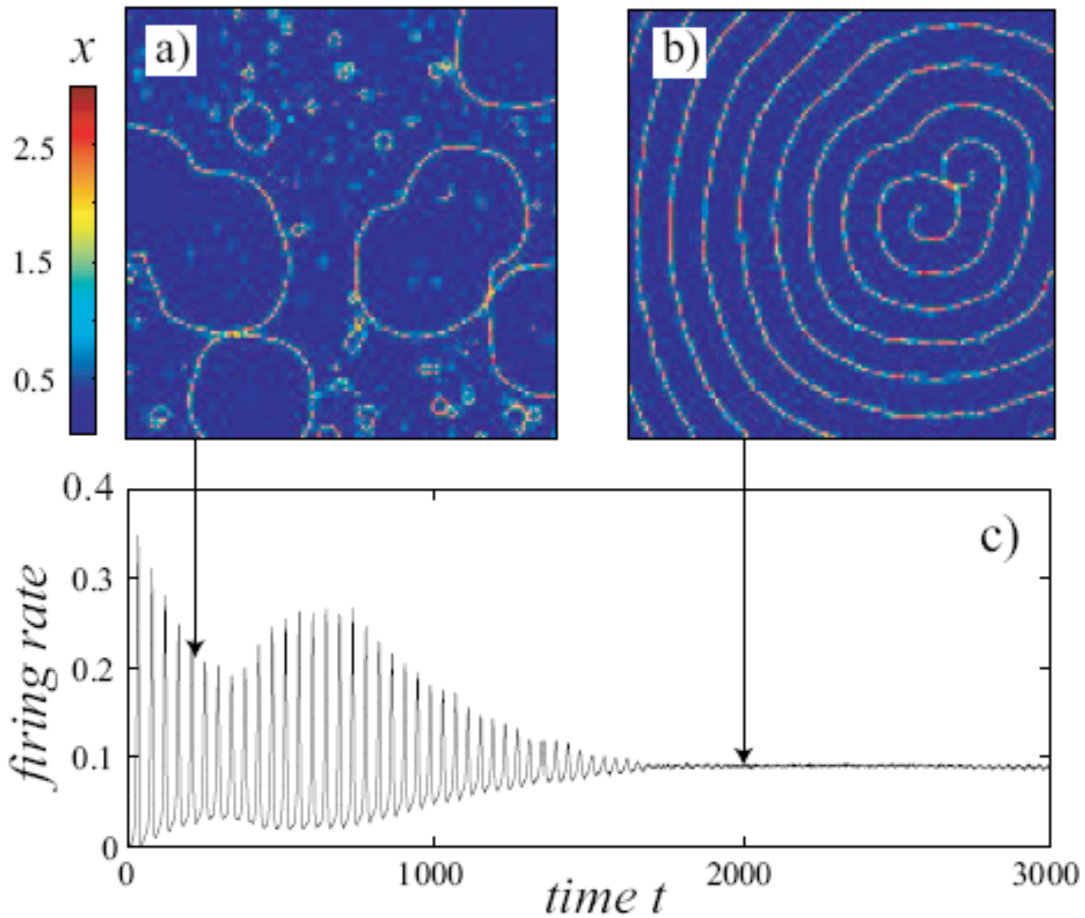


p=0.1

In a regime of shortcut densities, the system starts oscillating

Example:
2D small world of excitable media

S. Sinha, J. Saramäki, K. Kaski, cond-mat/0701121



- Each element behaves like a neuron (integrate, fire, recover)
- Random shortcuts added on top of the 2D plane
- Different dynamics arise as function of shortcut density

The Erdös Number

- Erdös authored 512 publications
- “Collaborative distance” to Pál Erdös is known as the Erdös number
- Average number is less than 5, almost everyone with a finite number has a number <8

http://www.ams.org/mathscinet/collaborationDistance.html

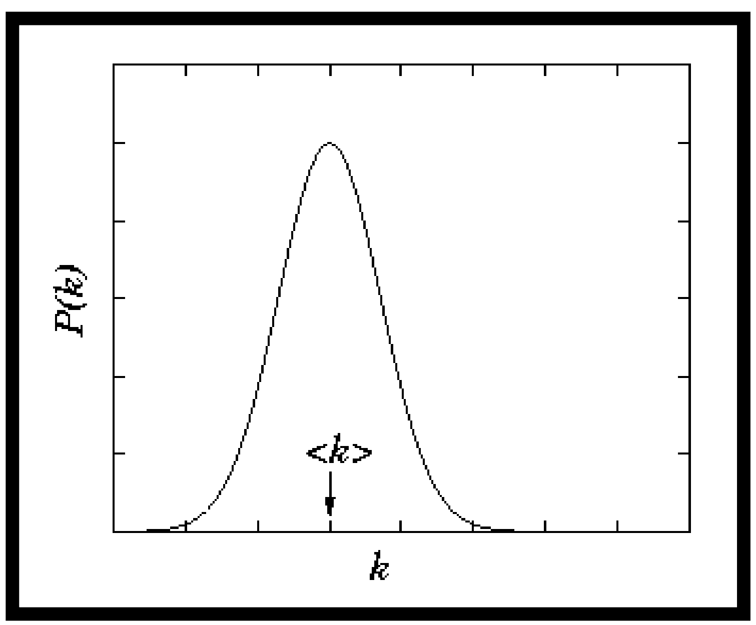
Scale-Free Networks

All Nodes Are Not Equal

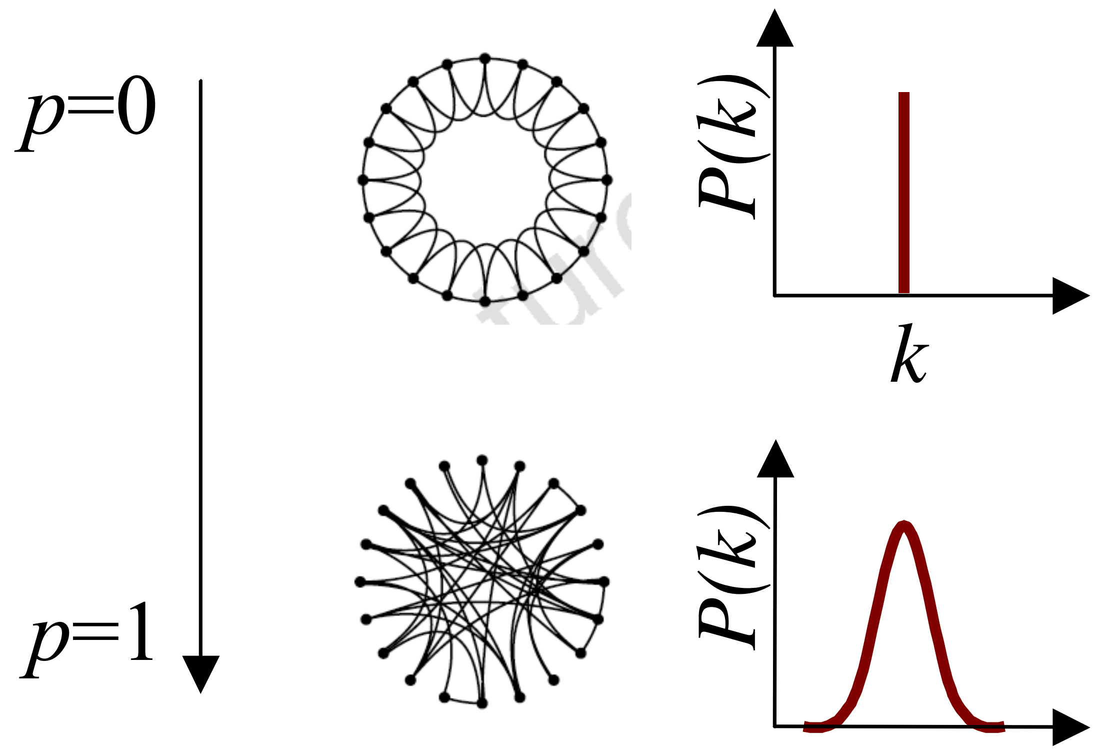
Degree Distributions, Again

- Erdös-Renyi:

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



- Small-World Networks:



Degree Distributions in Real-World Networks

A.-L. Barabási & R. Albert,

Emergence of Scaling in Random Networks, *Science* **286**, 509 (1999)

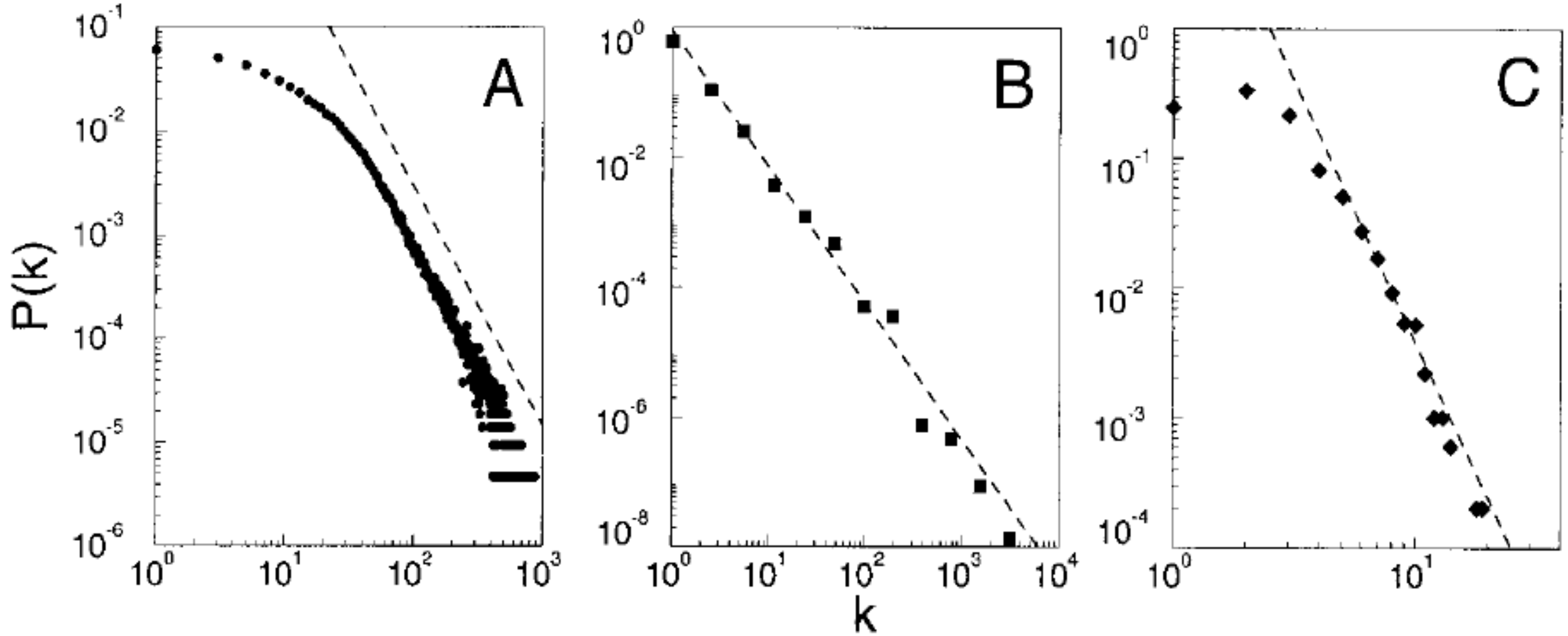
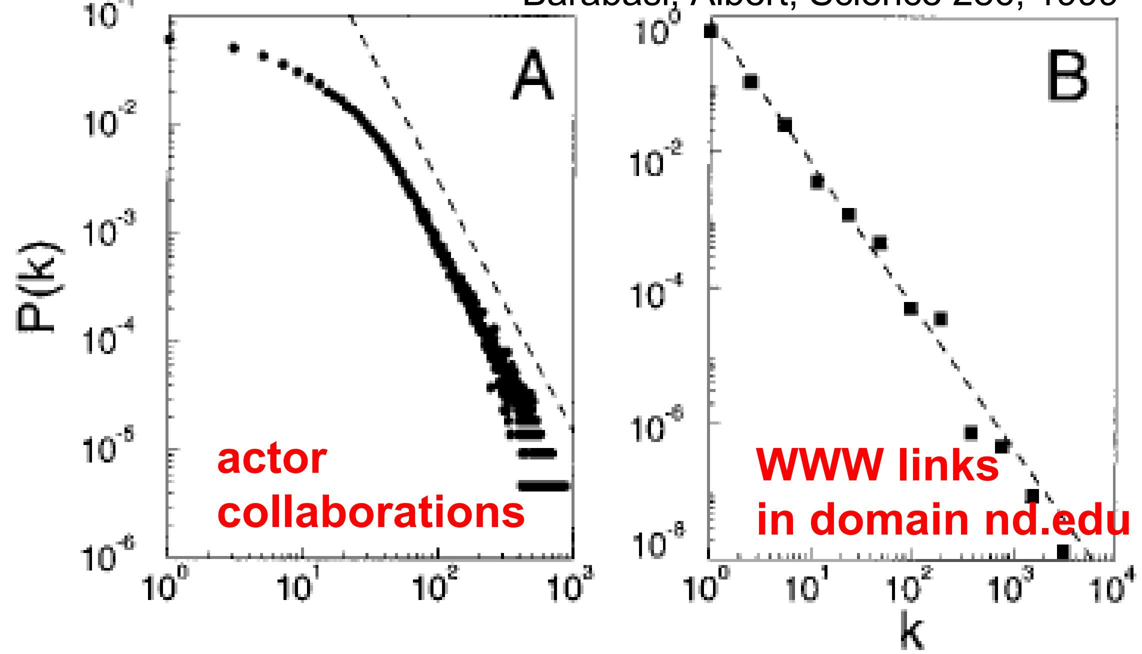


Fig. 1. The distribution function of connectivities for various large networks. (**A**) Actor collaboration graph with N = 212,250 vertices and average connectivity $\langle k \rangle$ = 28.78. (**B**) WWW, N = 325,729, $\langle k \rangle$ = 5.46 (6). (**C**) Power grid data, N = 4941, $\langle k \rangle$ = 2.67. The dashed lines have slopes (A) γ_{actor} = 2.3, (B) γ_{www} = 2.1 and (C) γ_{power} = 4.

Degree Distributions in Real-World Networks

Barabási, Albert, Science 286, 1999

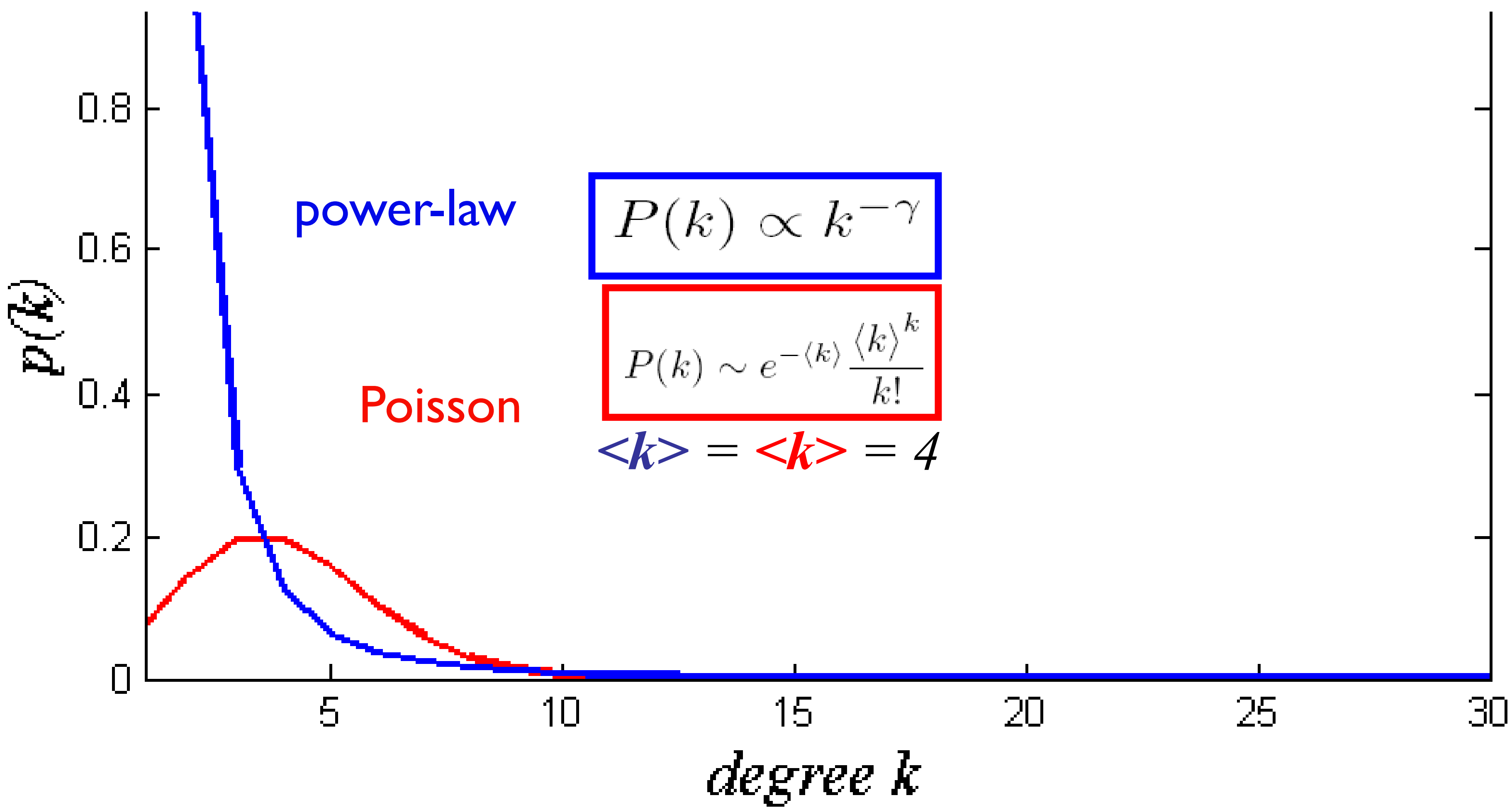


$$\boxed{P(k) \propto k^{-\gamma}}$$

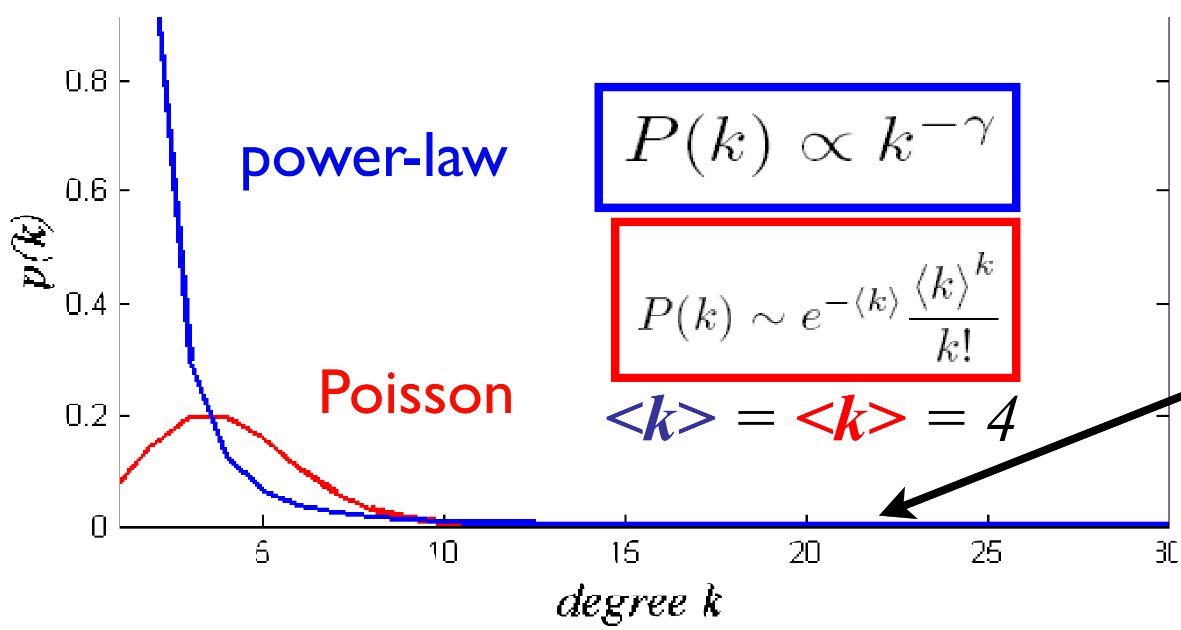
- Note the logarithmic axes:

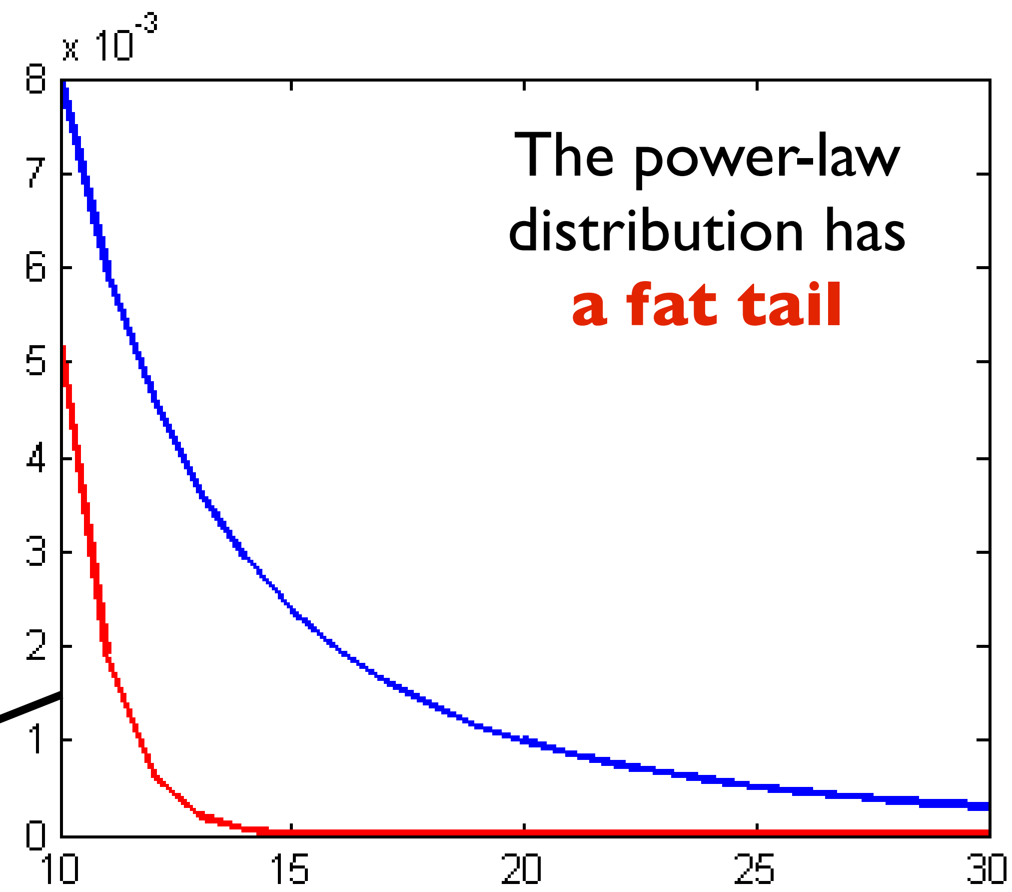
$$\log P(k) = -A \log k + B \Rightarrow P(k) = e^B k^{-A}$$

Power-Law Distributions

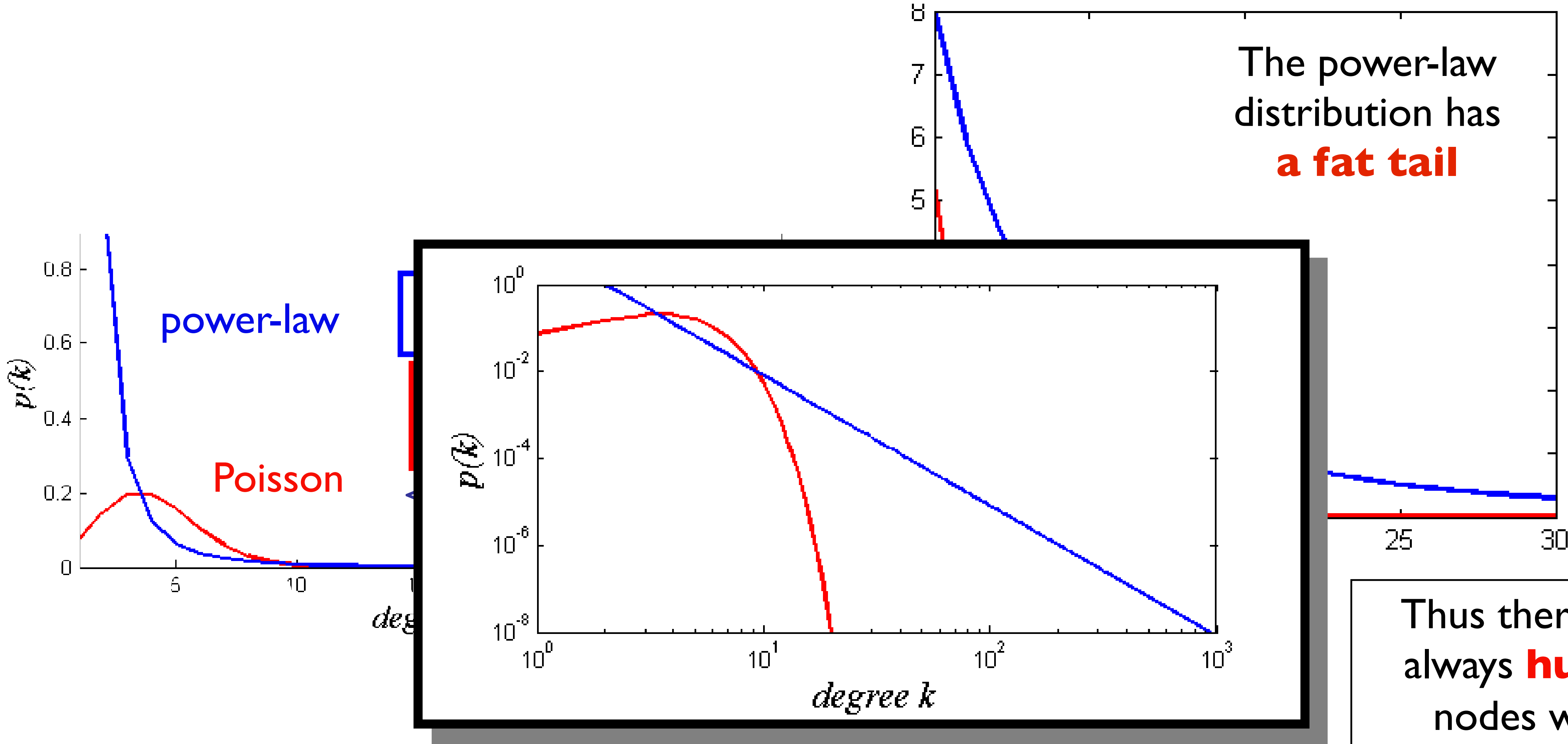


Power-Law Distributions



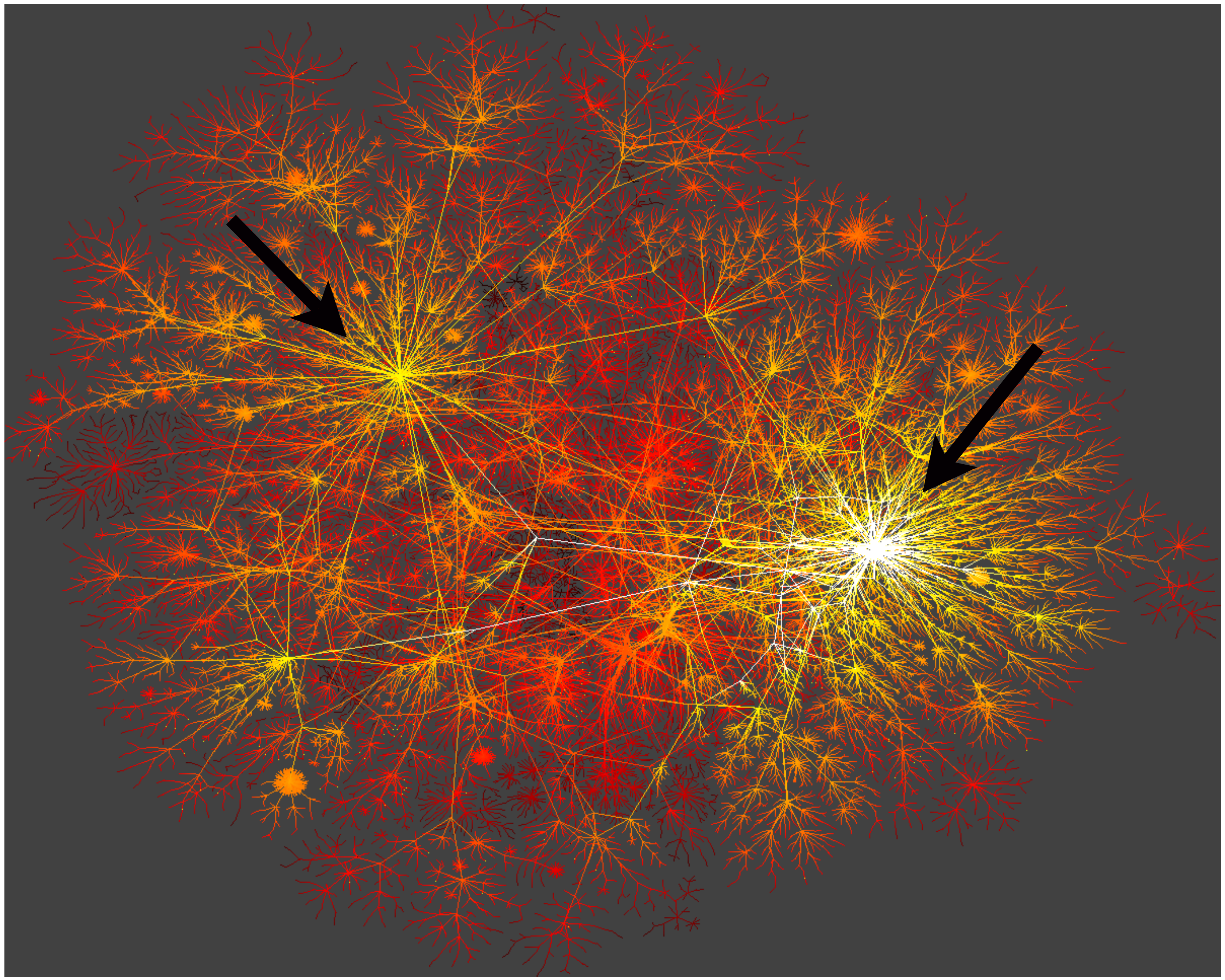


Power-Law Distributions



Thus there are always **hubs** - nodes with excessively high degrees

Example: The Internet



K. Claffy

More Real-World Examples

- Here, cumulative distributions have been plotted:

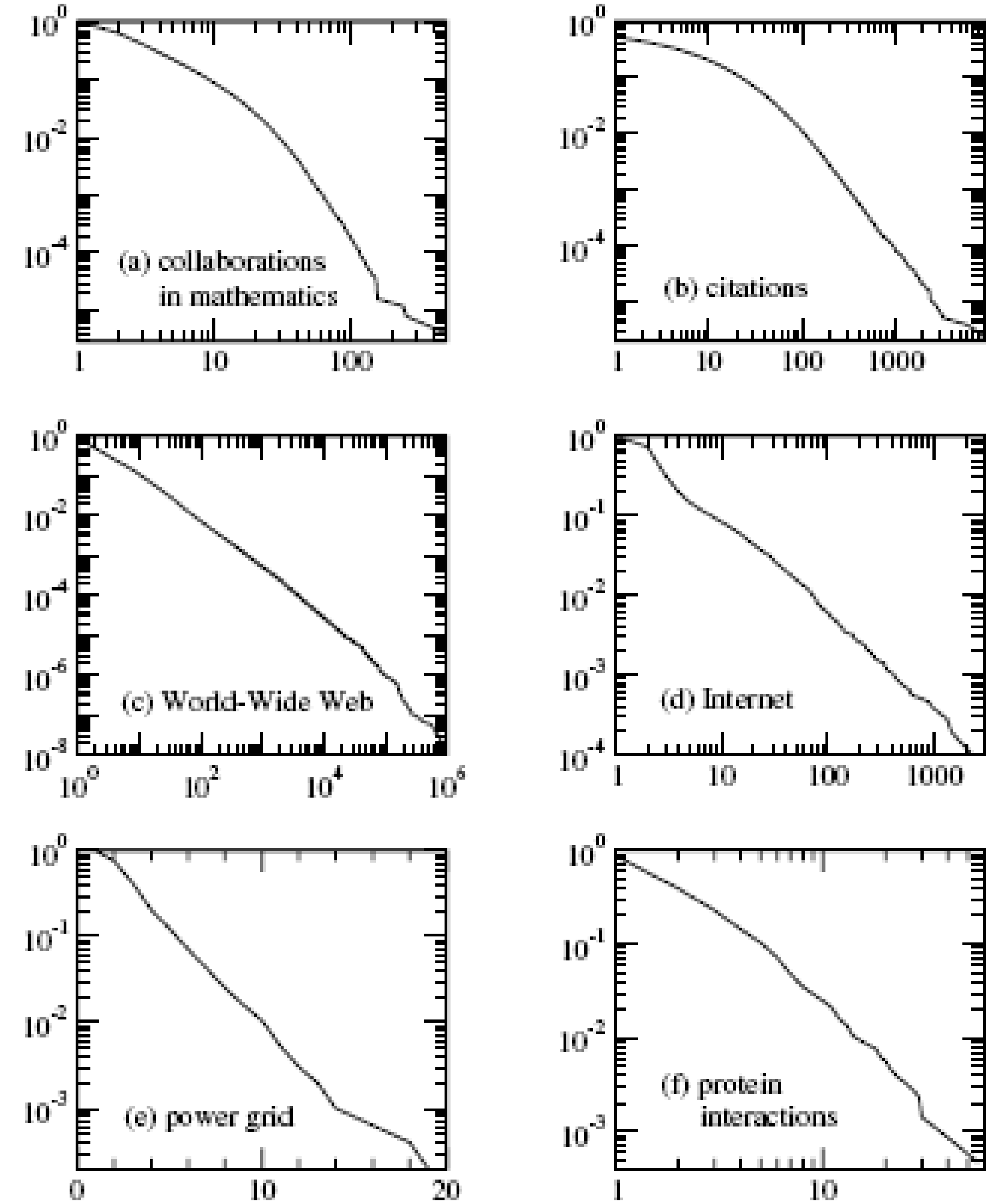
$$P\left(k > k'\right) \sim \sum_{k'=k}^{\infty} k'^{-\gamma} \sim k'^{-(\gamma-1)}$$

- For most real-world networks,

$$\gamma \in [2, 3]$$

- However note that there are serious difficulties in estimating power-law exponents!

- There are also other distributions which look like power laws!



M.E.J. Newman

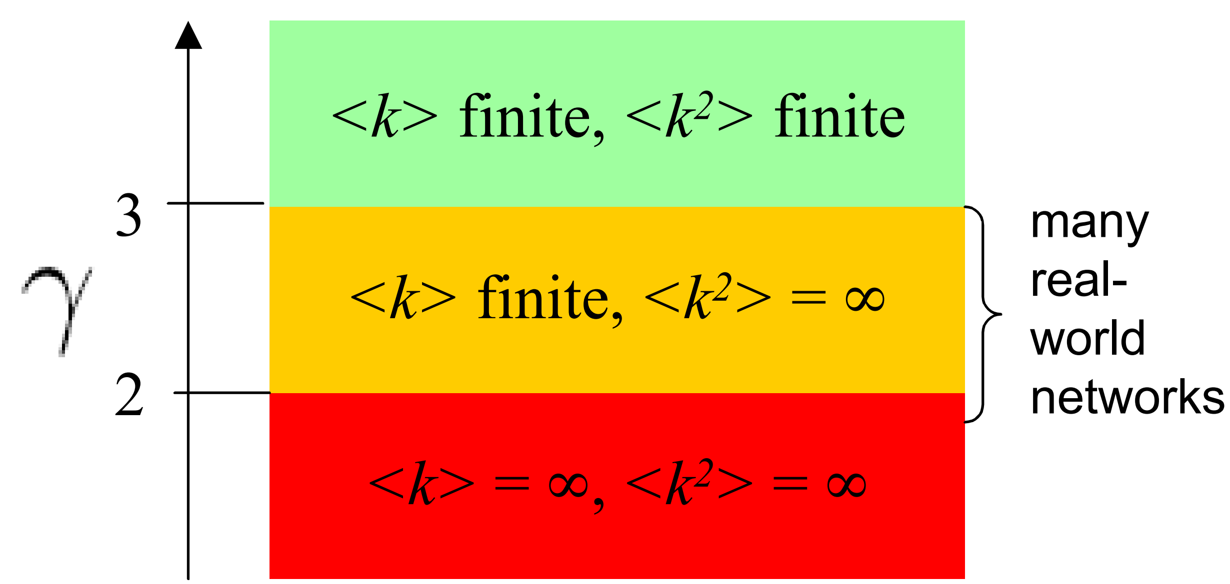
Moments of Power-Law Distributions

- What are the consequences of different ranges of γ?
- First moment = mean = average degree $<k>$:

$$
\begin{aligned}
\langle k \rangle &\sim \int_{k_0}^{\infty} k \times k^{-\gamma}\, dk \\
&= \int_{k_0}^{\infty} k^{-\gamma+1} dk \\
&= \lim_{k \to \infty} \frac{1}{2-\gamma} \left(\frac{1}{k^{\gamma-2}} - \frac{1}{k_0^{\gamma-2}} \right) \\
&\begin{cases} = & \infty, \text{ if } \gamma \le 2 \\ = & \text{const., if } \gamma > 2 \end{cases}
\end{aligned}
$$

- Second moment = variance = $<k^2>$:

$$
\langle k^2 \rangle = \begin{cases} = & \infty, \text{ if } \gamma \le 3 \\ = & \text{const., if } \gamma > 3 \end{cases}
$$



Scale-Free Networks

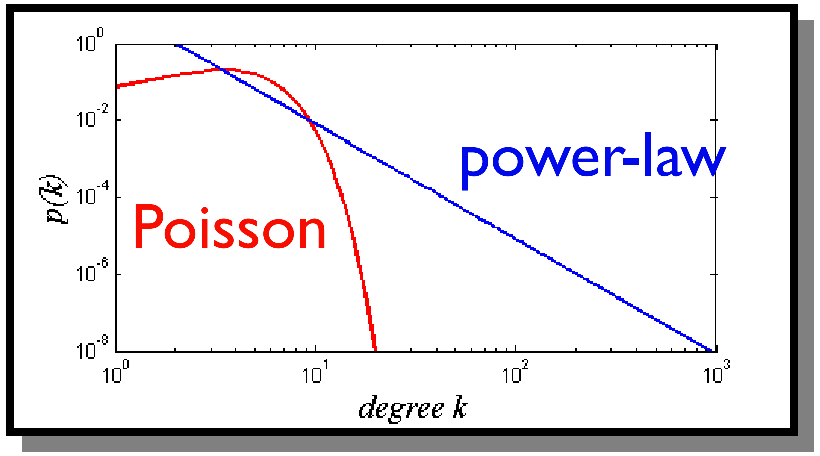
- Networks with power-law degree distributions are called **scale-free networks**
- This is because there is no characteristic scale in the distribution
- If degrees are rescaled, the form of the distribution does not change:

$$P(\alpha k) \propto (\alpha k)^{-\gamma} = \alpha^{-\gamma} P(k)$$

$$\Rightarrow \quad \frac{P(k=20)}{P(k=2)} = \frac{P(k=200)}{P(k=20)} = \frac{P\left(k = 2\times 10^6\right)}{P\left(k=2\times 10^5\right)} = \ldots$$

- For comparison, the Poisson distribution behaves like this:

$$\begin{aligned} \frac{P(\alpha k)}{P(k)} &= \frac{\langle k\rangle^{\alpha k}\, k!}{(\alpha k)!\,\langle k\rangle^{k}} \\ &= \langle k\rangle^{(\alpha-1)} \frac{k!}{(\alpha k)!} \\ &\sim 0, \text{when } \alpha > 1, k \gg \langle k\rangle \end{aligned}$$

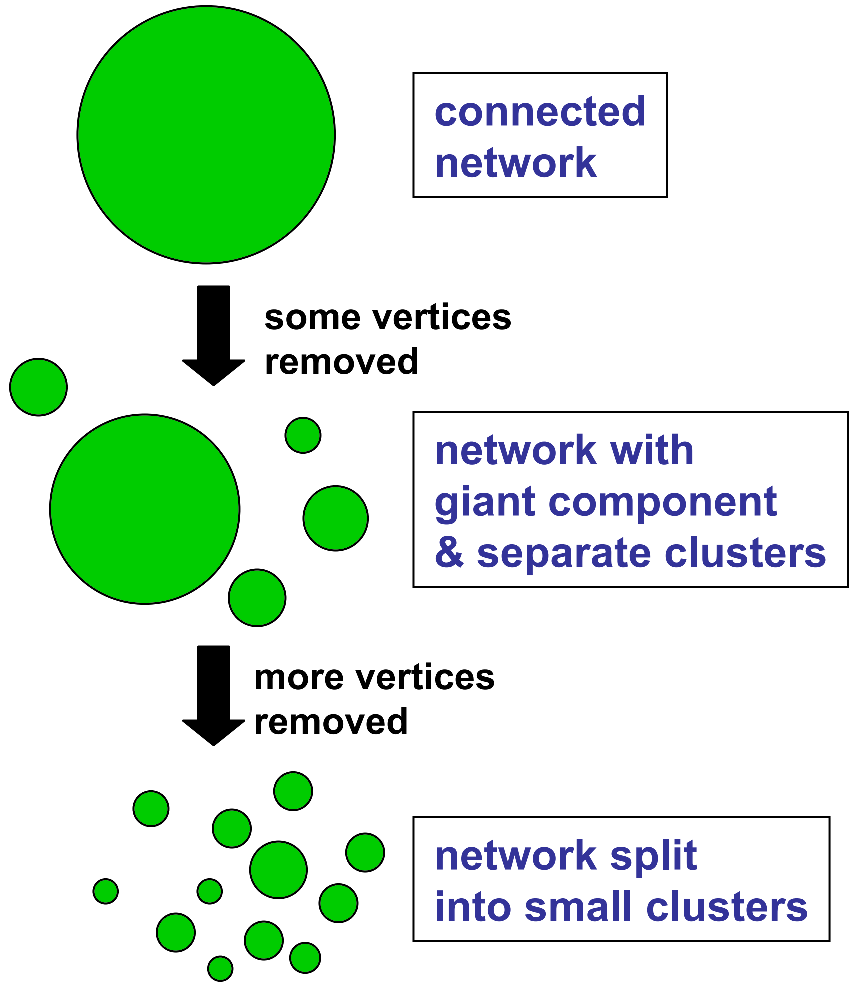


Importance of hubs

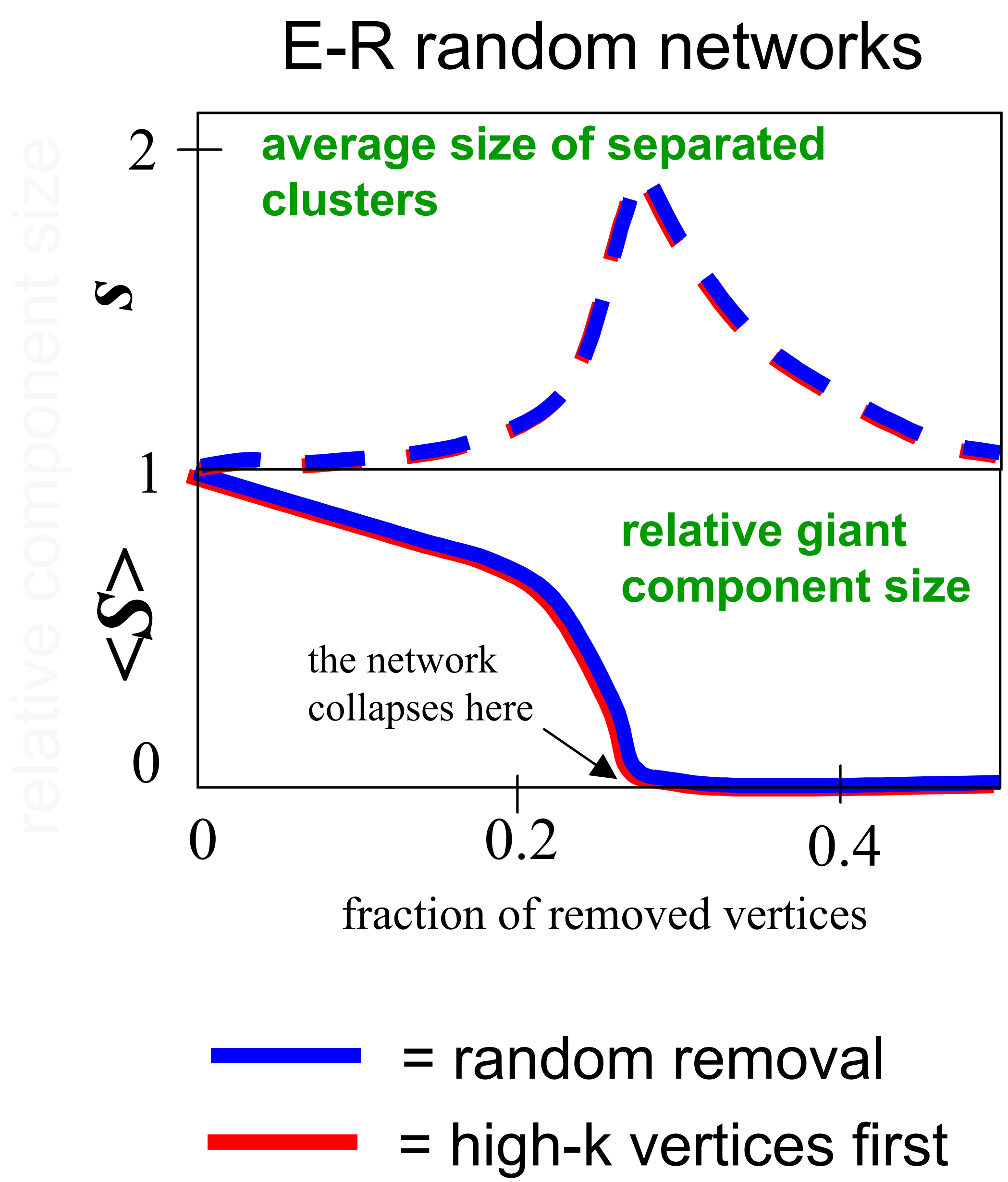
Numerical experiment:

1. Take a connected network
2. Start removing vertices one by one
3. Keep track of the size of the largest connected component

- Do this such that
  a) Vertices are removed *randomly*,
  b) Vertices are removed *in order of degree*, starting with the hubs



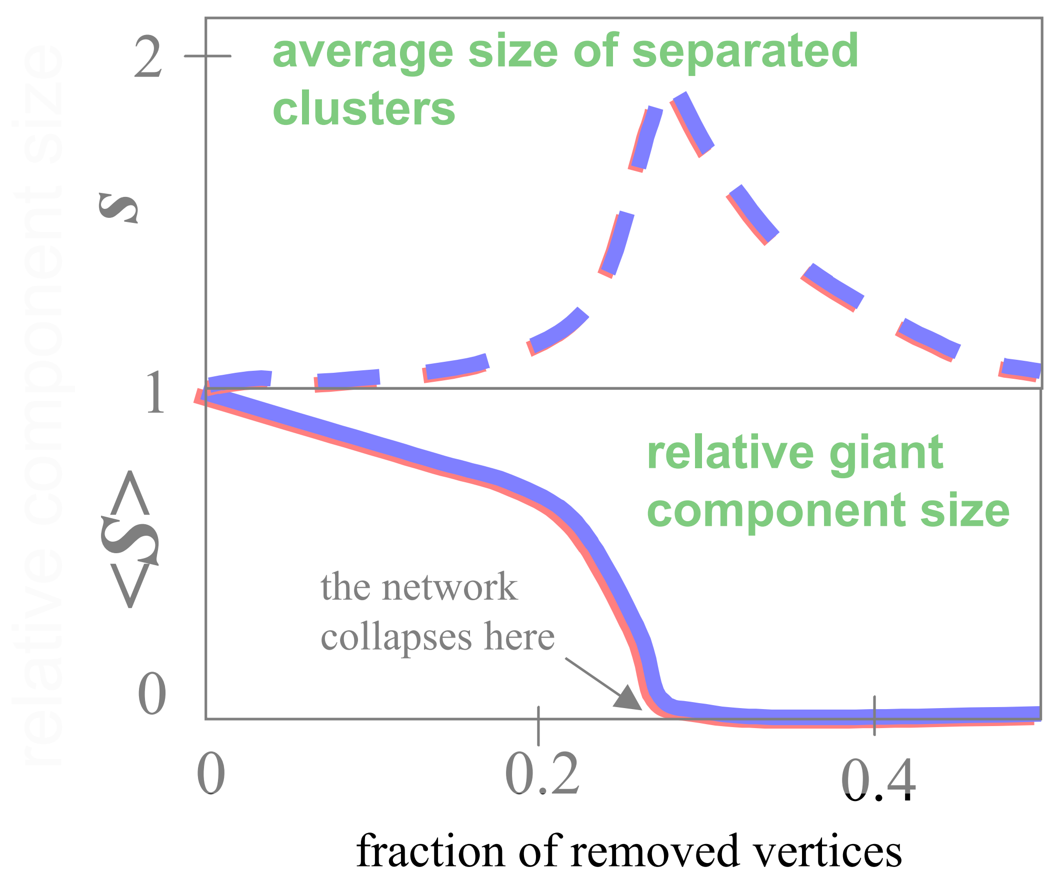
Importance of hubs: error tolerance



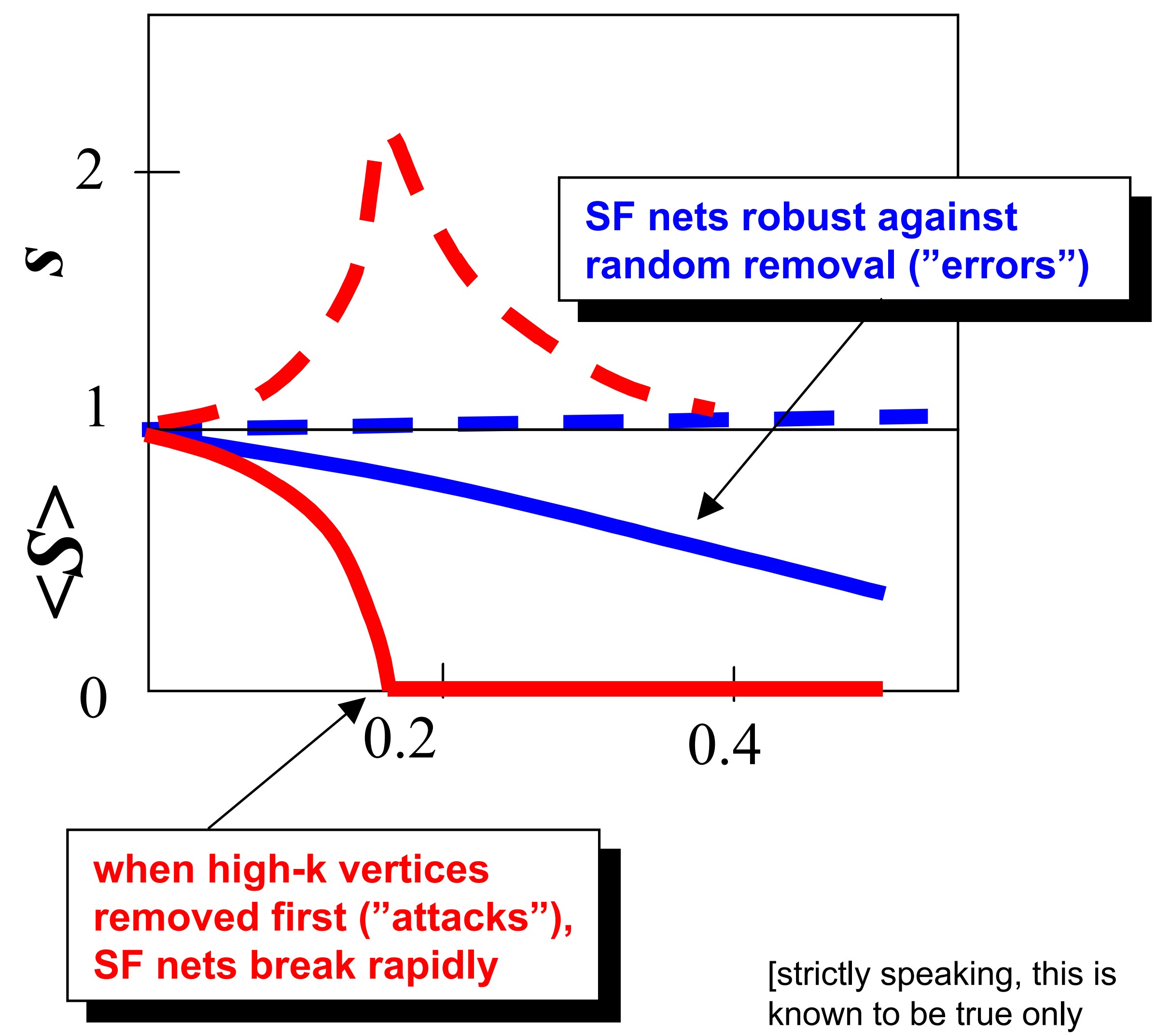
Albert, Jeong, Barabási, *Nature* **406**, 378 (2000)

Importance of hubs: error tolerance

E-R random networks



Scale-free networks



= random removal

= high-k vertices first

[strictly speaking, this is known to be true only for uncorrelated SF networks]

Albert, Jeong, Barabási, *Nature* **406**, 378 (2000)

The Role of Hubs

- This result indicates that scale-free networks are **very resilient to random damage**
  - E.g. the Internet is very robust against server breakdowns
  - This has been suggested as one reason for the ubiquity of scale-freeness
- However, there is a cost: **targeted attacks** will easily destroy such networks!

- There are deep connections to many observed phenomena
- E.g. spreading of biological and electronic viruses - as long as hubs exist, these will always find a way to spread

Where Do Networks Come From?

- For understanding network structure, the following point is essential:
- NATURAL NETWORKS ARE NOT STATIC OR IN EQUILIBRIUM!
- Instead, they are dynamic entities which are constantly growing and altering their wiring

- Of especial importance is the **growth** of networks

Network	Changes due to
Social networks	People being born & dying, people moving, ...
WWW	New hyperlinks
Protein interactions	Biological evolution
Scientific collaborations	New scientists, new papers

The Barabási-Albert Scale Free Model

- A model of **network growth**
- Based on the principle of **preferential attachment** - "rich get richer!"
- Yields networks with a power-law degree distribution

$$\boxed{P(k) = \frac{2m^2}{k^3}}$$

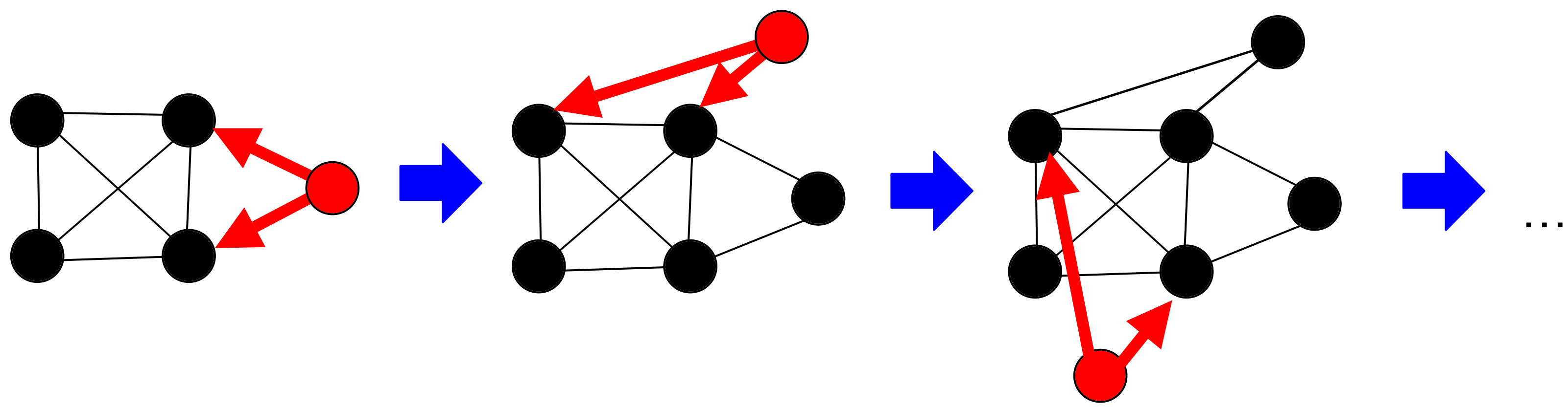
(average degree <k>=$2m$)

1. Take a small seed network, e.g. a few connected nodes
2. Let a new node of degree m enter the network
3. Connect the new node to existing nodes such that the probability of connecting to node i of degree k_i is

$$\boxed{\pi_i = \frac{k_i}{\sum_i k_i}}$$

4. Repeat 2.-3. until N nodes.

(Early) Motivation for Preferential Attachment



- Websites with many hyperlinks are easily found, and so people tend to link to them
- Popular people get to know more people
- Important proteins interact with more and more proteins produced by evolution

Some prior similar models

- The Simon Model (1955)
- The Price Model (1976) for scientific citations

The BA model: properties

- Average shortest path lengths:

$$\langle l \rangle \propto \frac{\ln N}{\ln\left(\ln N\right)}$$

- "Ultrasmall worlds!"

- Clustering coefficient:

$$C\left(k,N\right) \propto \frac{\left(\ln N\right)^2}{N}$$

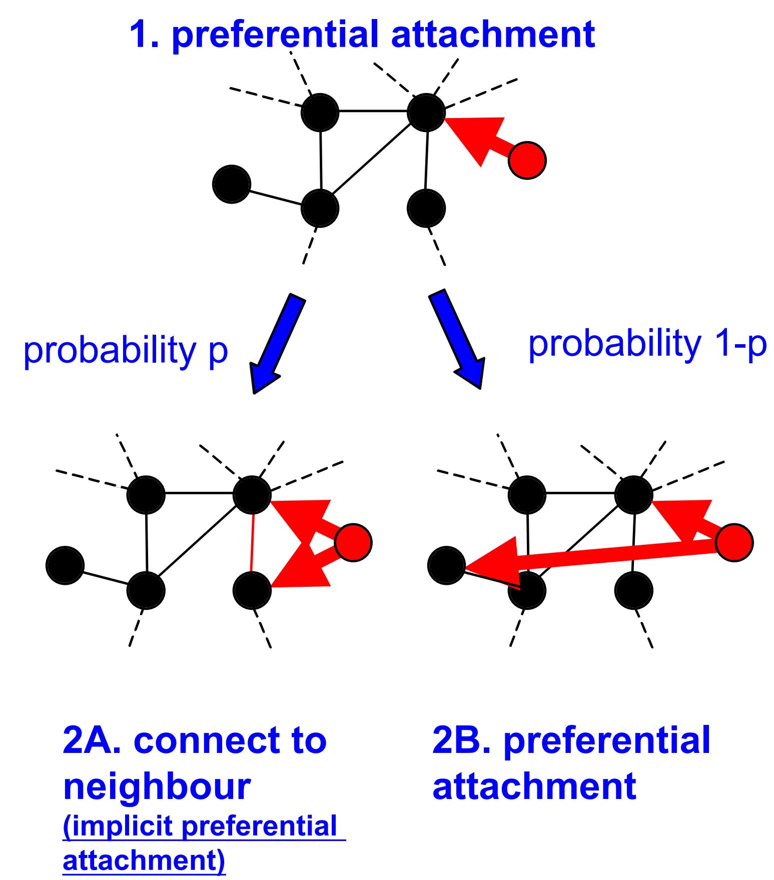
- The clustering coefficient is **independent of degree**, **decreases with network size**, and is **unrealistically small** for large networks!

- However, it is better than for E-R networks, where it is practically zero for networks of any size

Other Scale-Free Network Models

- **the Holme-Kim Model**
  - **motivation: to get realistic clustering**

1. Take a small seed network
2. Create a new vertex with m edges
3. Connect the first of the m edges to existing vertices with a probability proportional to their degree k (just like BA)
4. With probability p, connect the next edge to a random neighbour of the vertex of step 3., otherwise do 3. again
5. Repeat 2.-4. until the network has grown to desired size of N vertices



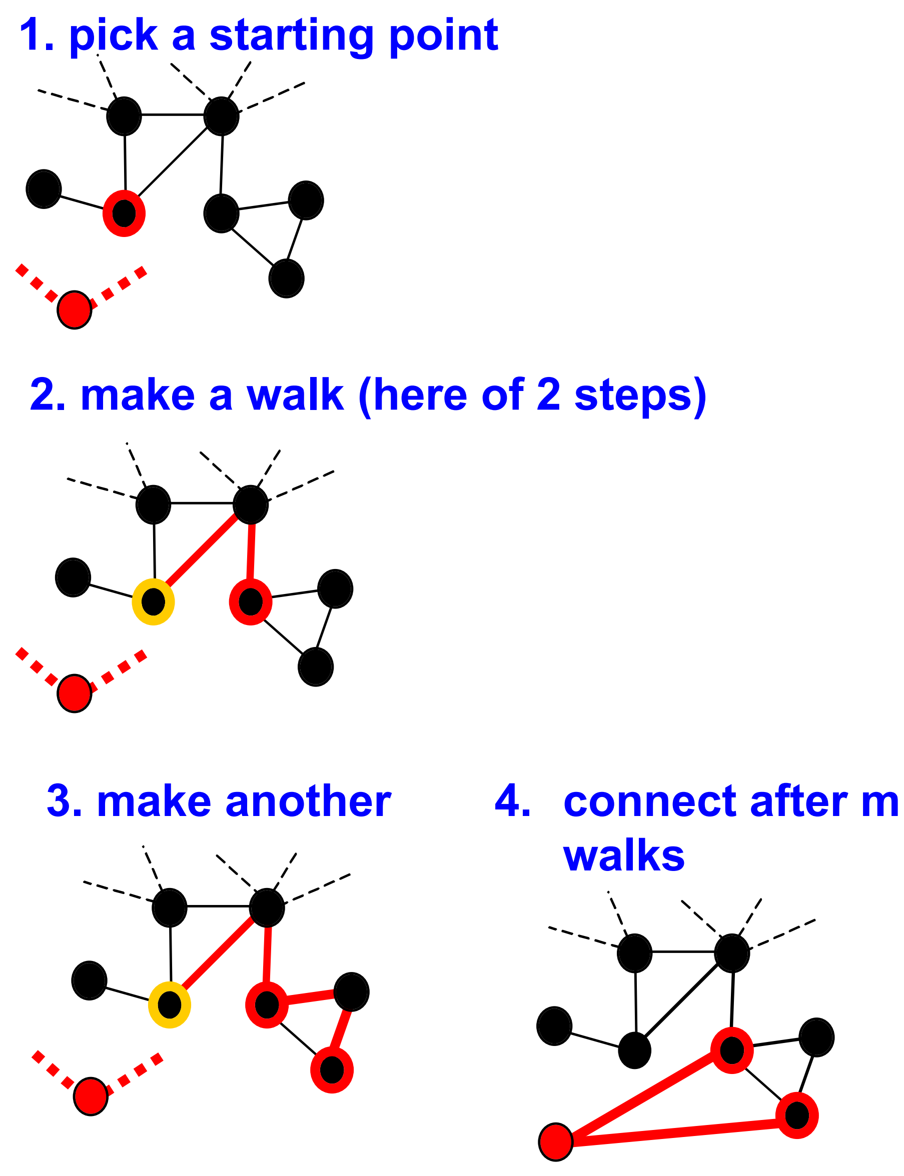
$$C(k) \propto \frac{1}{k}$$

for large N, ie clustering more realistic! This type of clustering is found in many real-world networks.

Other Scale-Free Network Models

- **Random walks (Vazquez, Evans, yours truly, et al.)**
  - **motivation: a "local" explanation to preferential attachment**
  - **e.g. people learn to know people through other people, which leads to popular people without looking for them**

1. Take a small seed network
2. Create a new vertex with m edges
3. Pick a random vertex
4. Make a l-step random walk starting from this vertex
5. Connect one of the edges of the new vertex to wherever you are
6. Repeat 3.-5. or 4.-5. m times
7. Repeat 2.-6. until N vertices

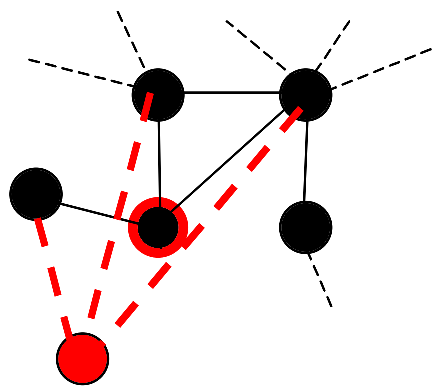


Other Scale-Free Network Models

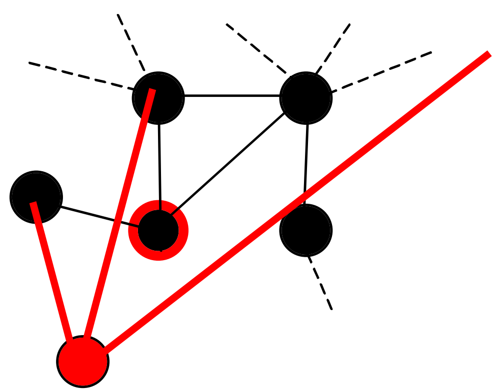
- **the vertex copying model (Kleinberg, Kumar):**
  - **motivation: citations or WWW link lists are often copied**
  - **a "local" explanation to preferential attachment**
  - **asymptotically SF with $\gamma \geq 3$**

1. Take a small seed network
2. Pick a random vertex
3. Make a copy of it
4. With probability p, move each edge of the copy to point to a random vertex
5. Repeat 2.-4. until the network has grown to desired size of N vertices

1. copy a vertex

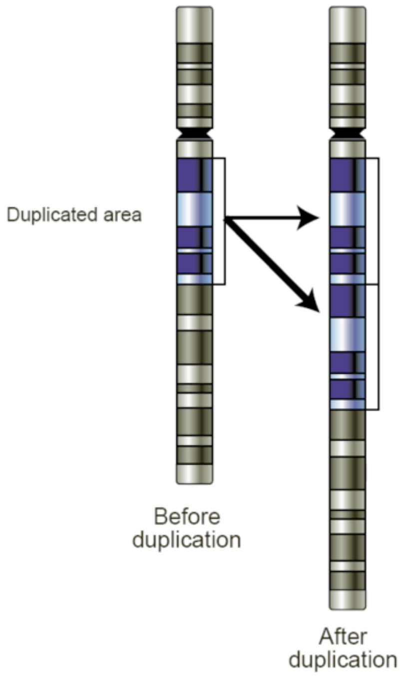


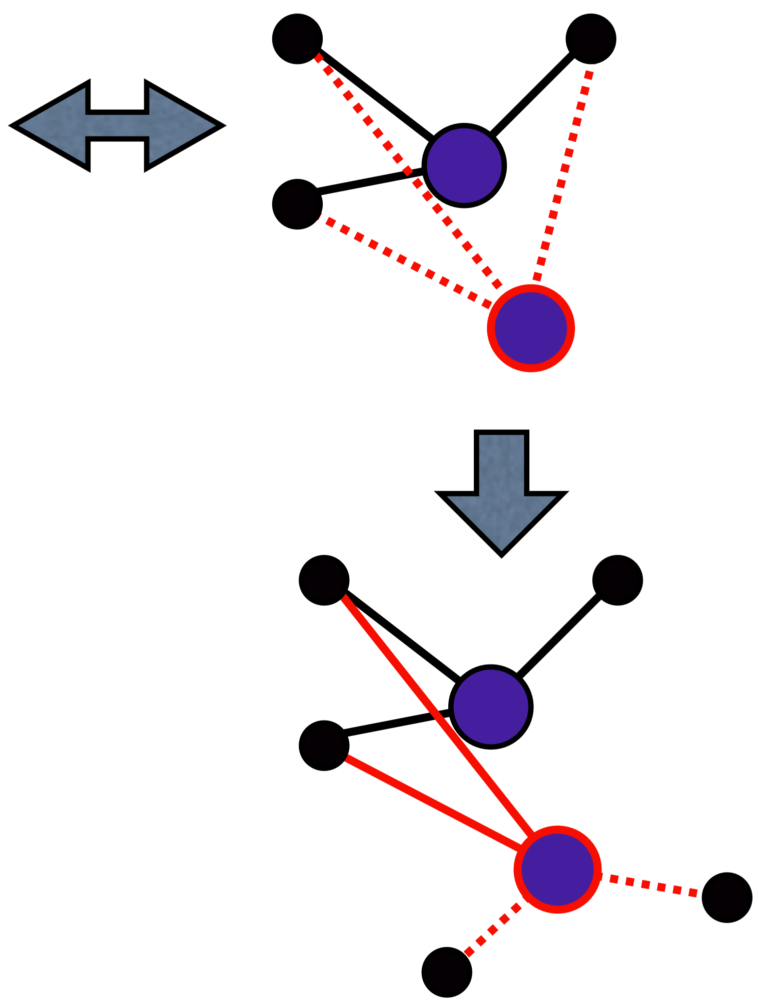
2. rewire edges with *p*



Evolution of Genome: Duplication & Divergence

- During evolution genes get accidentally duplicated
- A gene is duplicated = a node in the protein-protein interaction network is duplicated
- The duplicated gene can mutate more freely (no pressure to retain function)
- In terms of proteins, in course of time the duplicated protein loses some links and gains new ones
- This mechanism creates fat-tailed or scale-free degree distributions in protein-protein interaction networks!





Scale-Free Network Models: Summary

- For growing networks, **preferential attachment** yields power-law degree distributions

- To be exact, it has to be linear:

$$\boxed{\pi_i = \frac{k_i}{\sum_i k_i}}$$

(If *superlinear*, "winner takes it all" and in the end one node has ALL the links! If *sublinear*, we get a stretched exponential degree distribution)

(if *mixed*, e.g. combination of linear preferential and random attachment, we get exponents larger than 3!)

- The fundamental model: Barabási-Albert, where

$$\boxed{P(k) = \frac{2m^2}{k^3}}$$

- Several mechanisms lead to the preferential attachment principle!

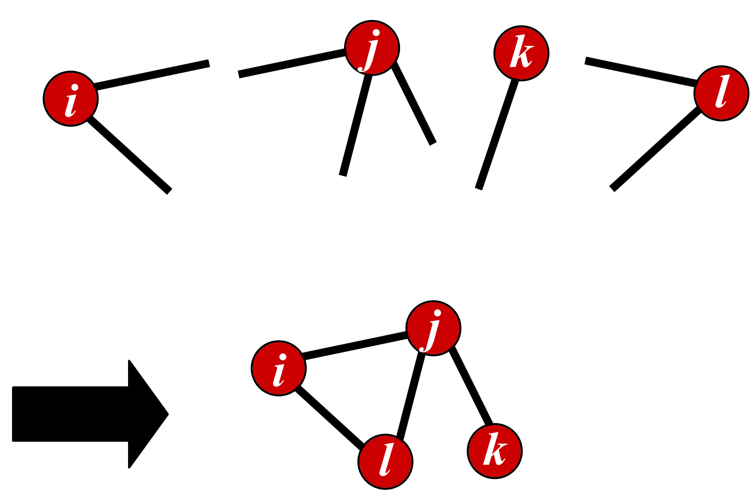
Generalized random graphs

- E-R networks: random with Poissonian degree distribution

- It may be useful to construct entirely random networks with a CHOSEN degree distribution $p(k)$

- Useful e.g. for comparisons:
  - Take a real-world network
  - Analyze it
  - Generate an entirely random network with the same degrees
  - Compare network characteristics (you'll learn these later...)
  - Why? To see if characteristics are not simply induced by degree distribution

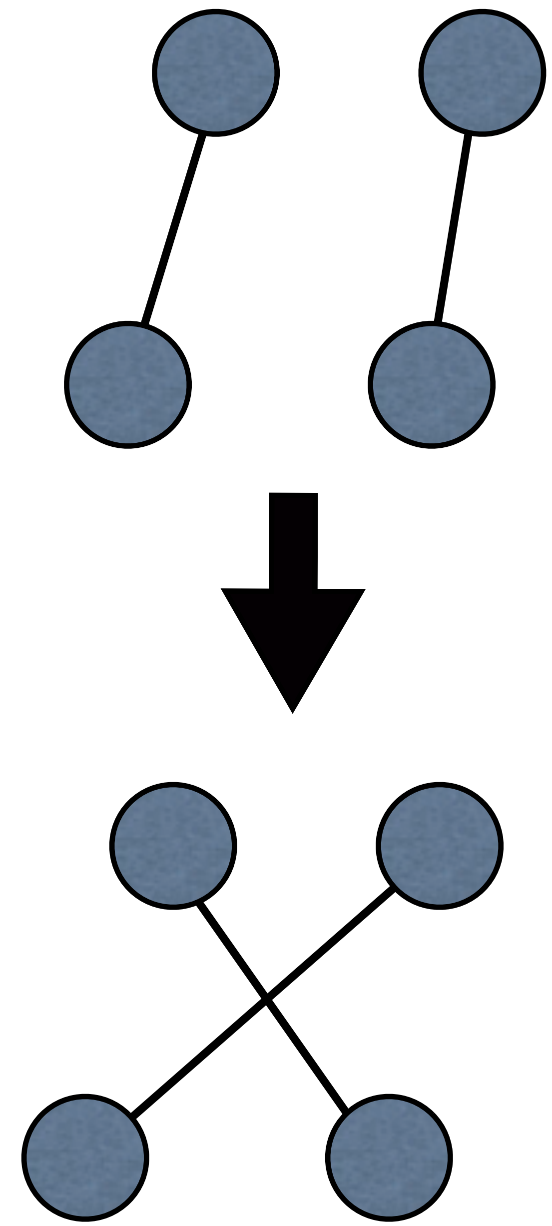
- The **configuration model:**
  - Take N vertices
  - Assign a degree to each
  - Join them randomly



Generalized random networks

Recipe II

- Take a network
- Pick two random edges
- Exchange their endpoints
- Repeat



Reading & Pointers

Review Papers

- M.E.J. Newman, The Structure and Function of Complex Networks, *SIAM Review* **45**, 167-256 (2003)
  - http://arxiv.org/abs/cond-mat/0303516
- Boccaletti et al, Complex networks: Structure and dynamics, *Physics Reports* **424**, 175 (2006)
  - http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.122.7590&rep=rep1&type=pdf

Books

- Popular science: A.-L. Barabási: Linked, D.J. Watts: Six Degrees
- Dorogovtsev & Mendes: Evolution of Networks
- Mark Newman, Albert-László Barabási, & Duncan J. Watts: The Structure and Dynamics of Networks (collection of key papers with editorial introductions)

Earlier network courses & exercises by me:

- http://www.lce.hut.fi/teaching/S-114.4150/
- https://noppa.tkk.fi/noppa/kurssi/s-114.4150