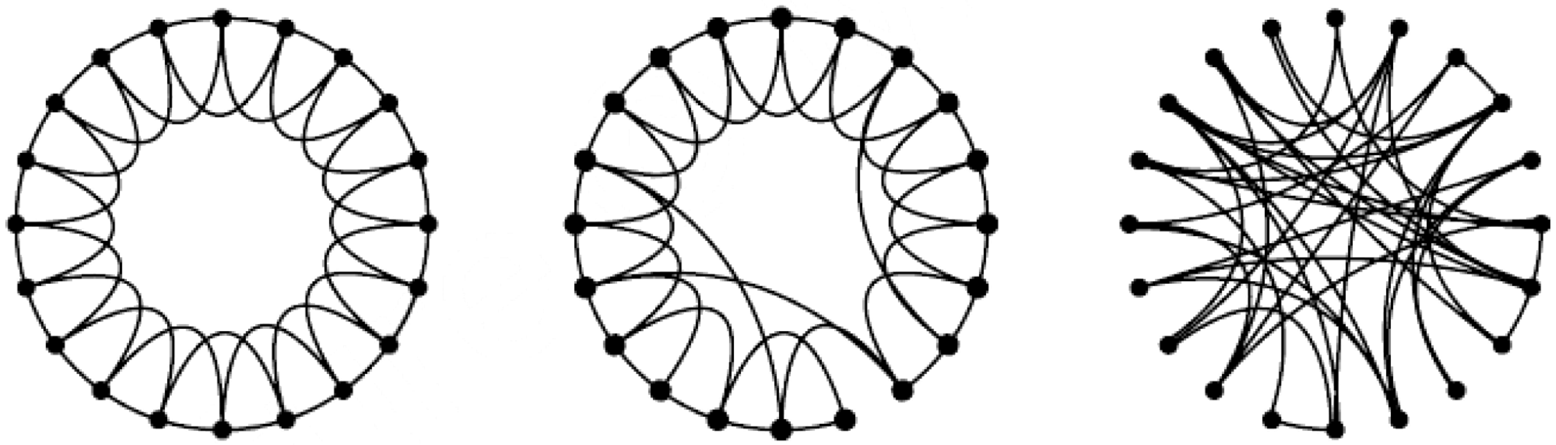
DATA SCIENCE SUMMER SCHOOL
A SERIES OF ONE-DAY WORKSHOPS ON DATA SCIENCE METHODS AND APPLICATIONS

CIVICA
THE EUROPEAN UNIVERSITY OF SOCIAL SCIENCES

Network Science

Part 2

Network Models



Márton Karsai
Central European University

Central quantities in network analysis

- Degree distribution: $P(k)$
- Clustering coefficient: C
- Average path length: $<d>$

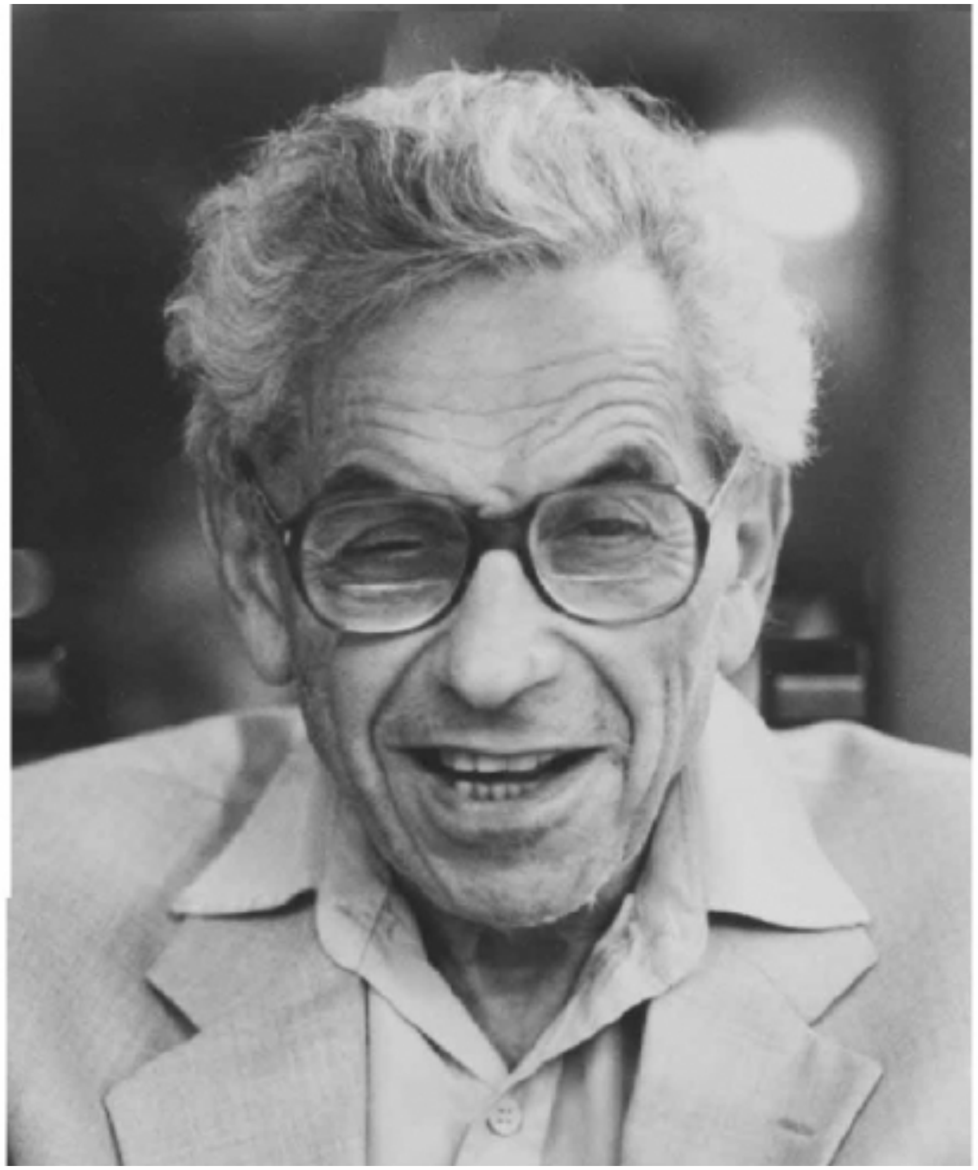
Central quantities in network analysis

- Degree distribution: $P(k)$
- Clustering coefficient: C
- Average path length: $<d>$

Network	Degree distribution	Path length	Clustering coefficient
Real world networks	broad	short	large

The Erdős-Rényi Random Graph model

Random Graphs



Pál Erdős
(1913-1996)

Alfréd Rényi
(1921-1970)

"If we do not know anything else than the number *N* of nodes and the number *L* of links, the simplest thing to do is to put the links at random (no correlations)"

P. Erdős and A. Rényi. On random graphs, I. Publicationes Mathematicae (Debrecen), 6:290-297, 1959.
P. Erdős and A. Rényi. On the evolution of random graphs. Publ. Math. Inst. Hung. Acad. Sci., 5:17-61, 1960.

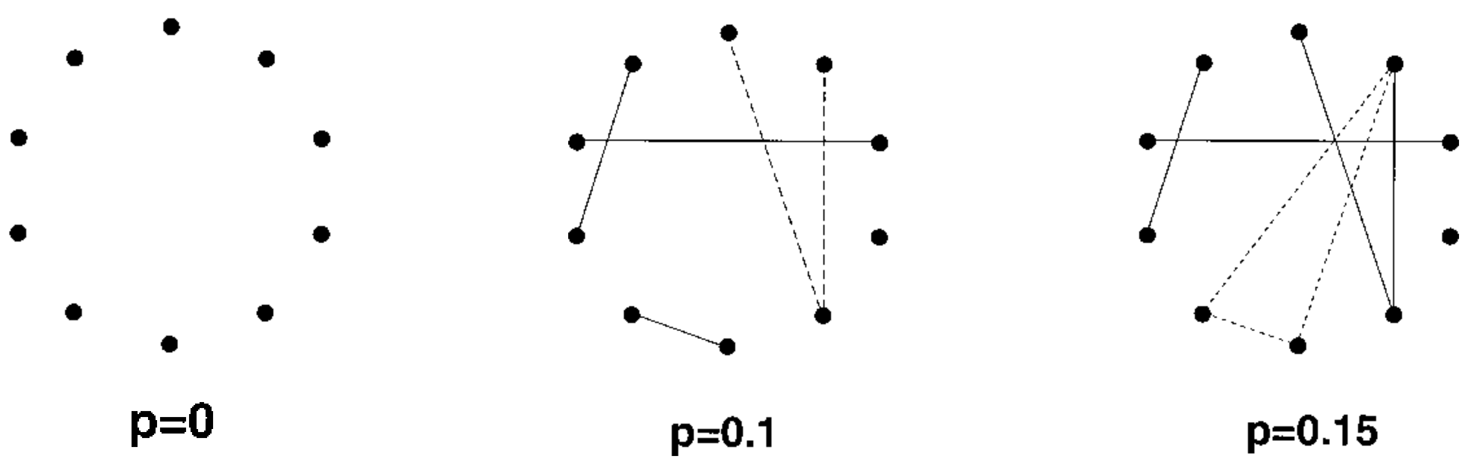
ER Random Graphs

Erdős-Rényi model: simple way to generate random graphs

- The *G(N,L)* definition
  1. Take N disconnected nodes
  2. Add L edges uniformly at random

Alternatively:

- pick uniformly randomly a graph from the set of all graphs with N nodes and L links



ER Random Graphs

Erdős-Rényi model: simple way to generate random graphs

- The $G(N,L)$ definition

  1. Take N disconnected nodes
  2. Add L edges uniformly at random

  Alternatively:

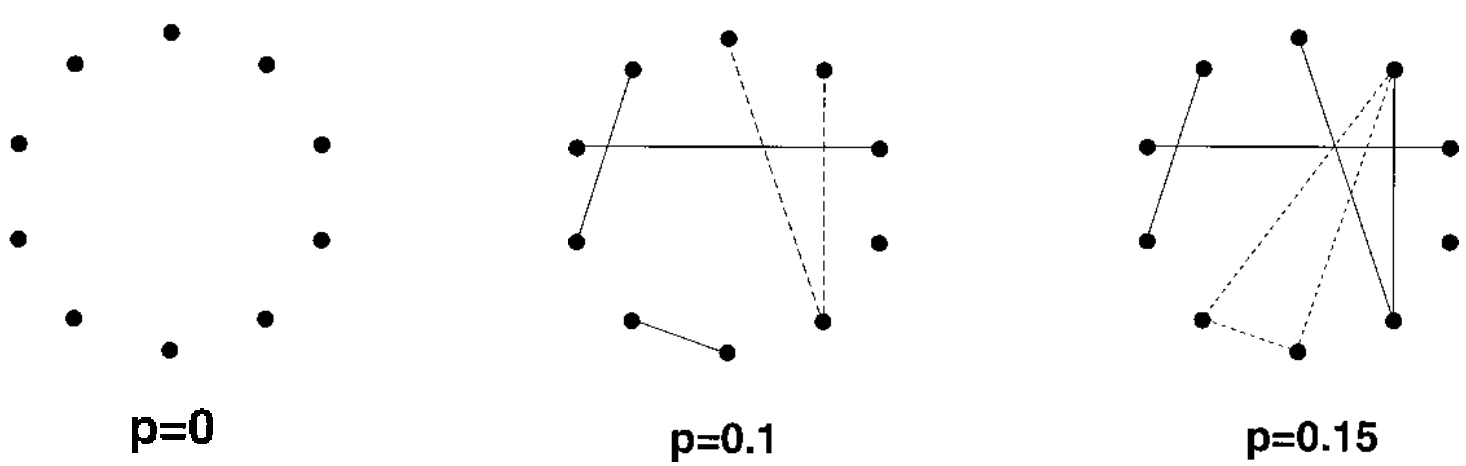
  - pick uniformly randomly a graph from the set of all graphs with N nodes and L links

- The $G(N,p)$ definition

  1. Take N disconnected nodes
  2. Add an edge between any of the nodes independently with probability p

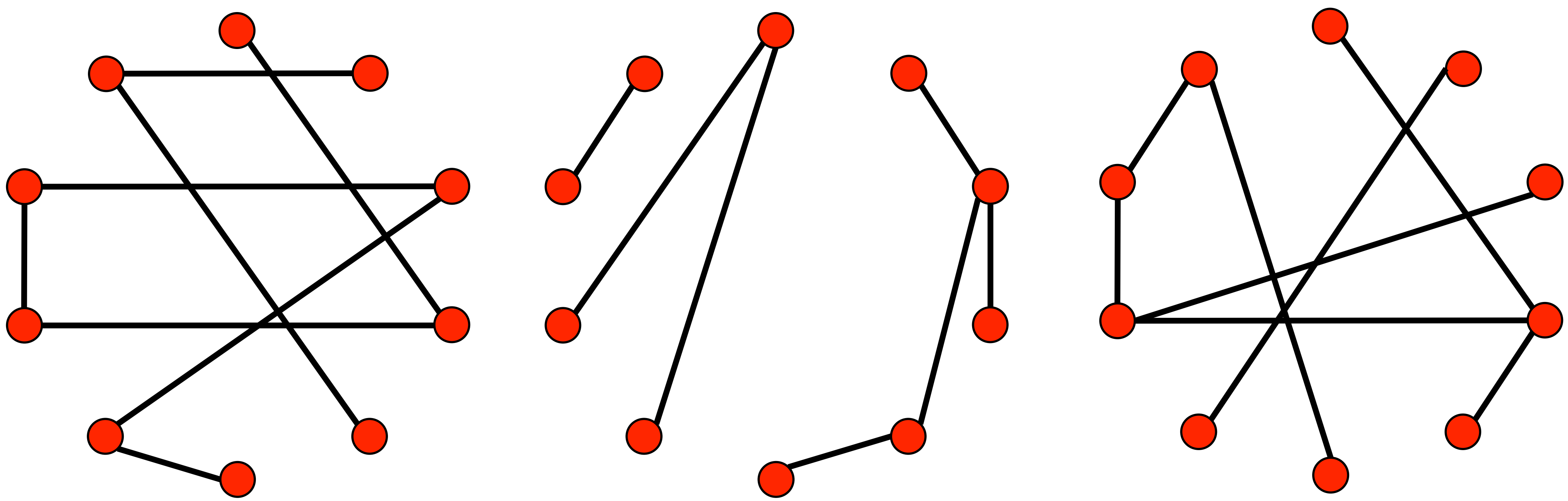
  Alternatively:

  - pick with probability $p^L(1-p)^{\binom{N}{2}-L}$ a network from the set of all networks with size N



Random Graphs

N and ***p*** do not uniquely define the network– we can have many different realizations of it.



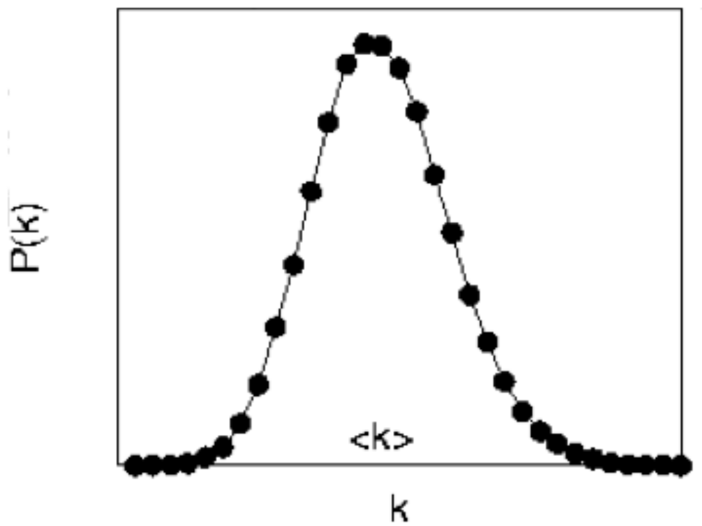
N=10
p=1/6

The probability to form a ***particular*** graph **G(N,p)** is

$$P(G(N,p)) = p^L (1-p)^{\frac{N(N-1)}{2}-L}$$

That is, each graph **G(N,p)** appears with probability **P(G(N,p))**.

Degree distribution - Random Graphs



$$P(k) = \binom{N-1}{k} p^k (1-p)^{(N-1)-k}$$

Select k
nodes from N-1

probability of
having ***k*** edges

probability of
missing N-1-k
edges

$$< k >= p(N-1) \qquad \sigma_k^2 = p(1-p)(N-1)$$

$$\frac{\sigma_k}{< k >} = \left[\frac{1-p}{p}\frac{1}{(N-1)}\right]^{1/2} \approx \frac{1}{(N-1)^{1/2}}$$

As the network size increases, the distribution becomes increasingly narrow—we are increasingly confident that the degree of a node is in the vicinity of <k>.

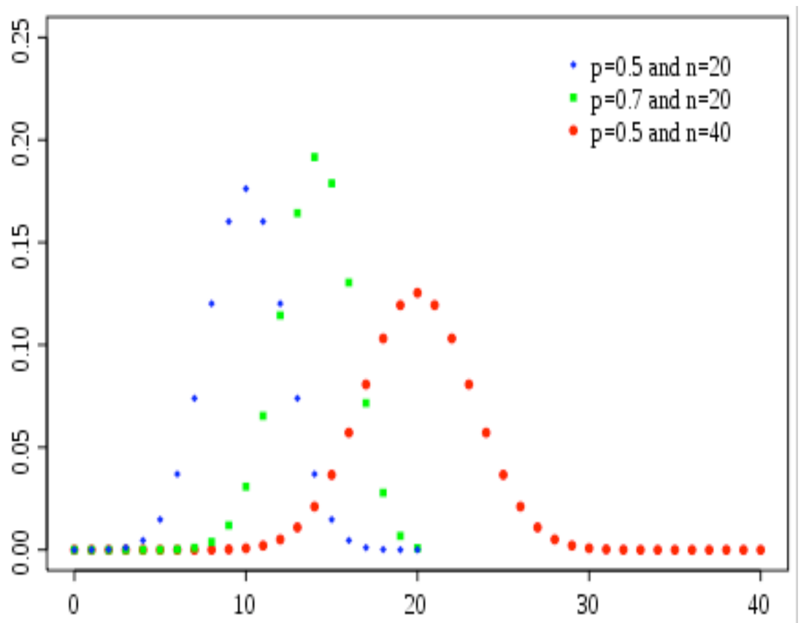
Degree distribution - Random Graphs

Probability Distribution Function (PDF)

Exact Result

-binomial distribution-

$$P(k) = \binom{N-1}{k} p^k (1-p)^{(N-1)-k}$$



$$<k> = (N-1)p$$

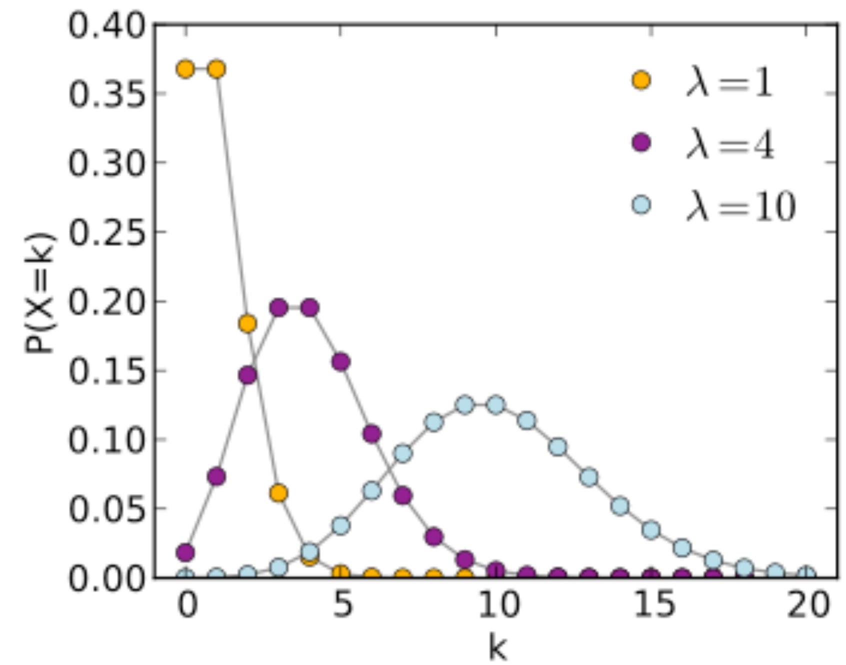
$$<k^2> = p(1-p)(N-1) + p^2(N-1)^2$$

$$\sigma_k = (<k^2> - <k>^2)^{1/2} = [p(1-p)(N-1)]^{1/2}$$

Large N limit

-Poisson distribution-

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



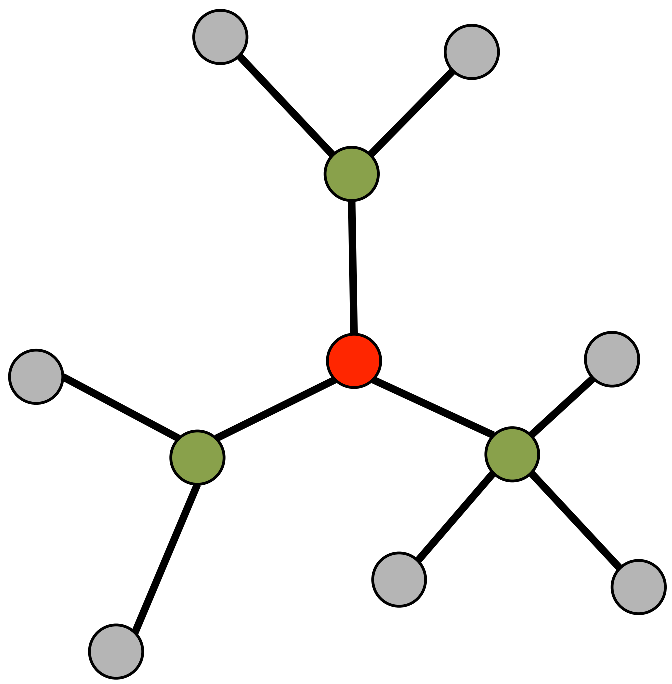
$$<k> = <k>$$

$$<k^2> = <k>(1 + <k>)$$

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

Distance - Random Graphs

Random graphs tend to have a tree-like topology with almost constant node degrees.



- nr. of first neighbors: $N \simeq \langle k \rangle$
- nr. of second neighbors: $N \simeq \langle k \rangle^2$
- nr. of neighbours at distance d: $N \simeq \langle k \rangle^d$
- estimate maximum distance:

$$\underbrace{N = 1 + \langle k \rangle + \langle k \rangle^2 + \ldots + \langle k \rangle^d}_{\text{geometric series}} = \frac{\langle k \rangle^{d+1} - 1}{\langle k \rangle - 1} \approx \langle k \rangle^d \quad \Rightarrow \quad d = \frac{\log N}{\log \langle k \rangle}$$

Sum of all nodes in the network within maximum distance d

Distance - ER Random Networks

- **Logarithmically short distance among nodes**

$$d = \frac{logN}{log\langle k \rangle}$$

Real-world networks

Network	Size	$\langle k \rangle$	ℓ	ℓ_{rand}	C	C_{rand}	Reference
WWW, site level, undir.	153 127	35.21	3.1	3.35	0.1078	0.00023	Adamic, 1999
Internet, domain level	3015–6209	3.52–4.11	3.7–3.76	6.36–6.18	0.18–0.3	0.001	Yook *et al.*, 2001a, Pastor-Satorras *et al.*, 2001
Movie actors	225 226	61	3.65	2.99	0.79	0.00027	Watts and Strogatz, 1998
LANL co-authorship	52 909	9.7	5.9	4.79	0.43	1.8×10^{-4}	Newman, 2001a, 2001b, 2001c
MEDLINE co-authorship	1 520 251	18.1	4.6	4.91	0.066	1.1×10^{-5}	Newman, 2001a, 2001b, 2001c
SPIRES co-authorship	56 627	173	4.0	2.12	0.726	0.003	Newman, 2001a, 2001b, 2001c
NCSTRL co-authorship	11 994	3.59	9.7	7.34	0.496	3×10^{-4}	Newman, 2001a, 2001b, 2001c
Math. co-authorship	70 975	3.9	9.5	8.2	0.59	5.4×10^{-5}	Barabási *et al.*, 2001
Neurosci. co-authorship	209 293	11.5	6	5.01	0.76	5.5×10^{-5}	Barabási *et al.*, 2001
E. coli, substrate graph	282	7.35	2.9	3.04	0.32	0.026	Wagner and Fell, 2000
E. coli, reaction graph	315	28.3	2.62	1.98	0.59	0.09	Wagner and Fell, 2000
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06	Montoya and Solé, 2000
Silwood Park food web	154	4.75	3.40	3.23	0.15	0.03	Montoya and Solé, 2000
Words, co-occurrence	460.902	70.13	2.67	3.03	0.437	0.0001	Ferrer i Cancho and Solé, 2001
Words, synonyms	22 311	13.48	4.5	3.84	0.7	0.0006	Yook *et al.*, 2001b
Power grid	4941	2.67	18.7	12.4	0.08	0.005	Watts and Strogatz, 1998
C. Elegans	282	14	2.65	2.25	0.28	0.05	Watts and Strogatz, 1998

Albert, R. et.al. Rev. Mod. Phy. (2002)

Clustering - Random Graphs

$$C_i \equiv \frac{2n_i}{k_i(k_i-1)}$$

where n_i is the number of links between the neighbours of node i

- Edges are independent and have the same probability p $\quad n_i \cong p\frac{k_i(k_i-1)}{2}$
- Earlier we showed $\quad p = \frac{<k>}{N-1}$

if N>>1 and <k>≈1

$$C_i = \frac{2<k>}{N-1}\frac{k_i(k_i-1)}{2}\frac{1}{k_i(k_i-1)} = \frac{<k>}{N-1} \approx \frac{<k>}{N}$$

- For fixed average degree C is decreasing as N goes large

➡️Low clustering coefficient

➡️It is vanishing with the system size

Clustering - ER Random Networks

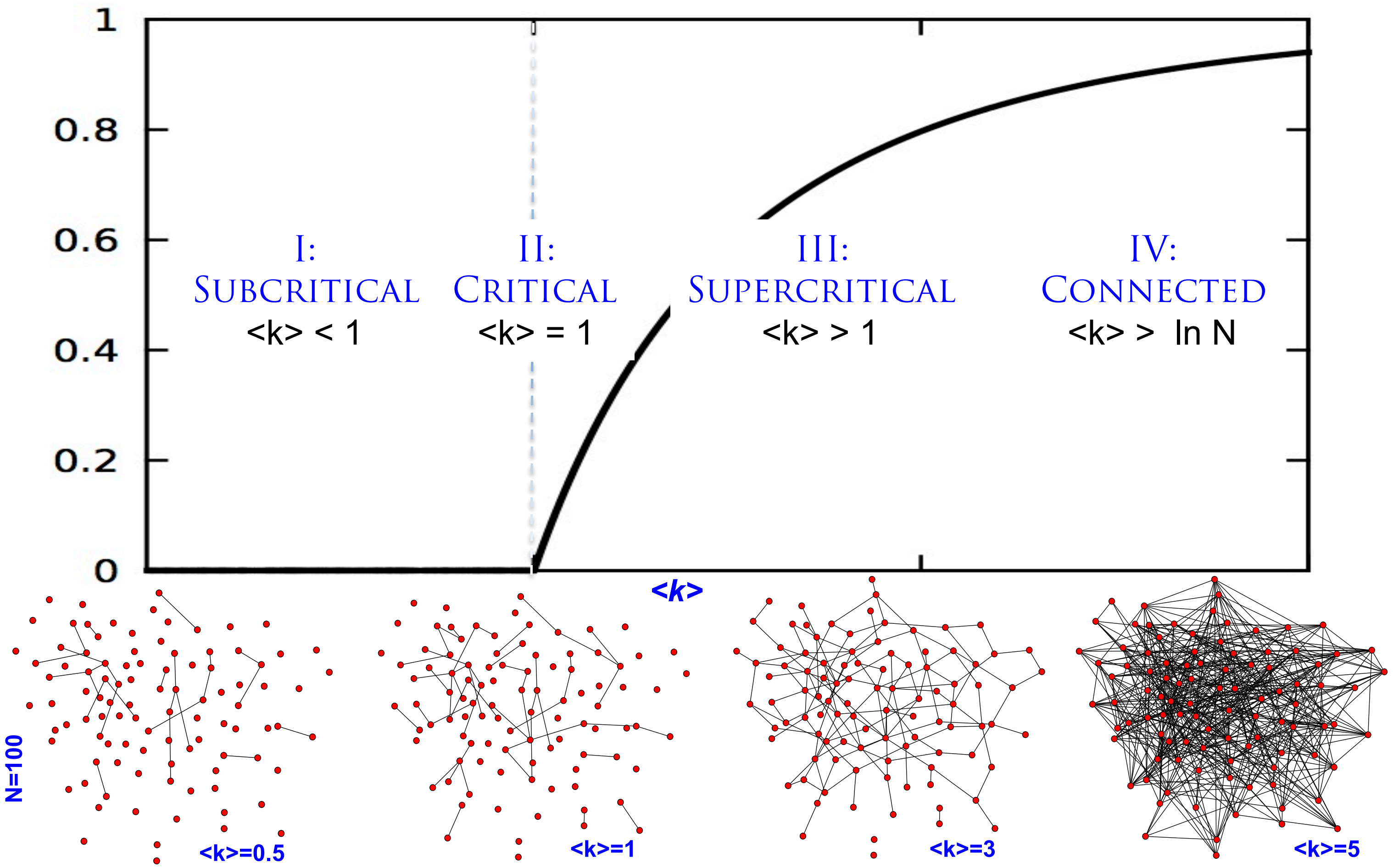
- **Small clustering coefficient** $$C_i \equiv \frac{1}{N} < k >= p$$

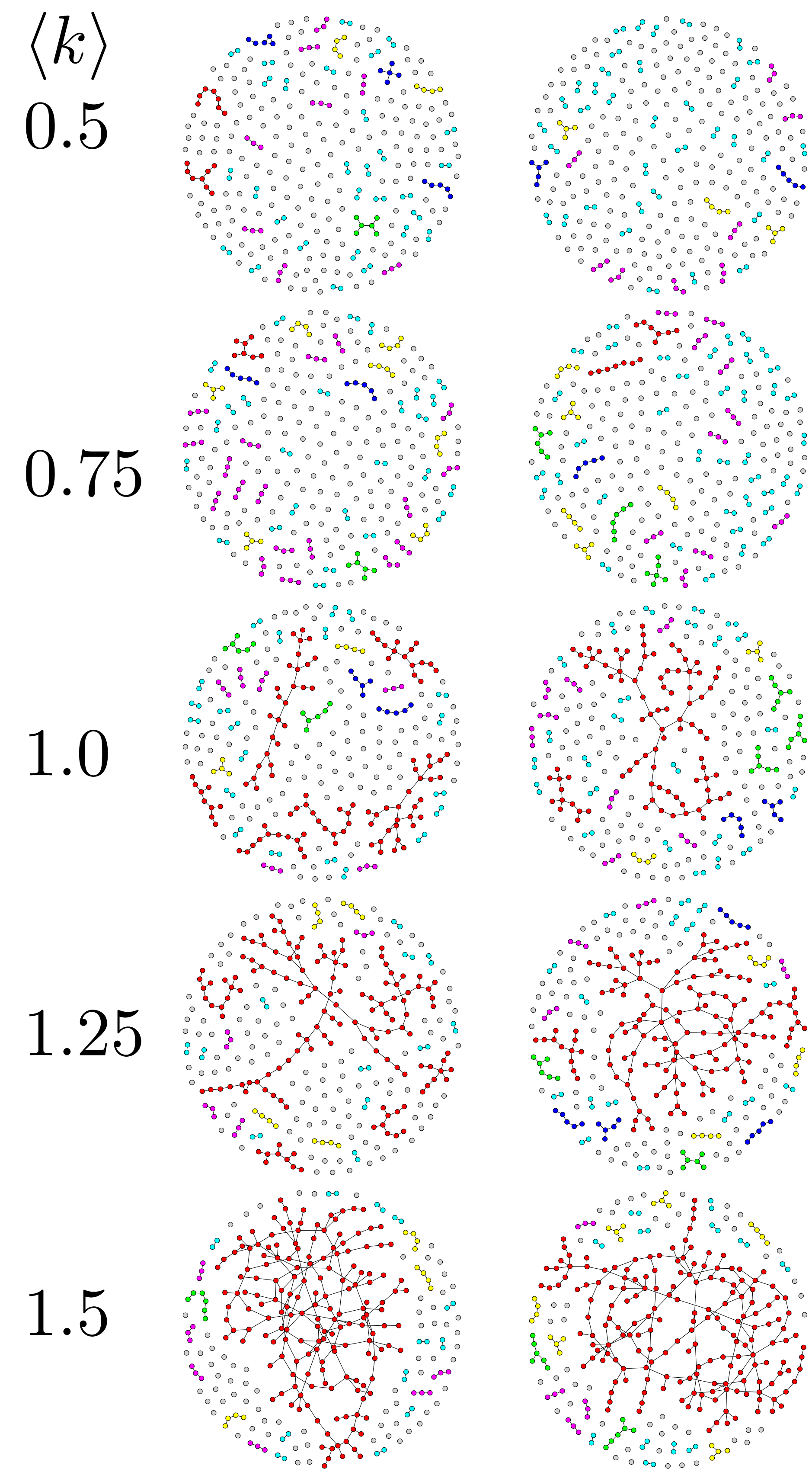
Real-world networks

Network	Size	$\langle k \rangle$	ℓ	ℓ_{rand}	C	C_{rand}	Reference
WWW, site level, undir.	153 127	35.21	3.1	3.35	0.1078	0.00023	Adamic, 1999
Internet, domain level	3015–6209	3.52–4.11	3.7–3.76	6.36–6.18	0.18–0.3	0.001	Yook *et al.*, 2001a, Pastor-Satorras *et al.*, 2001
Movie actors	225 226	61	3.65	2.99	0.79	0.00027	Watts and Strogatz, 1998
LANL co-authorship	52 909	9.7	5.9	4.79	0.43	1.8×10^{-4}	Newman, 2001a, 2001b, 2001c
MEDLINE co-authorship	1 520 251	18.1	4.6	4.91	0.066	1.1×10^{-5}	Newman, 2001a, 2001b, 2001c
SPIRES co-authorship	56 627	173	4.0	2.12	0.726	0.003	Newman, 2001a, 2001b, 2001c
NCSTRL co-authorship	11 994	3.59	9.7	7.34	0.496	3×10^{-4}	Newman, 2001a, 2001b, 2001c
Math. co-authorship	70 975	3.9	9.5	8.2	0.59	5.4×10^{-5}	Barabási *et al.*, 2001
Neurosci. co-authorship	209 293	11.5	6	5.01	0.76	5.5×10^{-5}	Barabási *et al.*, 2001
E. coli, substrate graph	282	7.35	2.9	3.04	0.32	0.026	Wagner and Fell, 2000
E. coli, reaction graph	315	28.3	2.62	1.98	0.59	0.09	Wagner and Fell, 2000
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06	Montoya and Solé, 2000
Silwood Park food web	154	4.75	3.40	3.23	0.15	0.03	Montoya and Solé, 2000
Words, co-occurrence	460.902	70.13	2.67	3.03	0.437	0.0001	Ferrer i Cancho and Solé, 2001
Words, synonyms	22 311	13.48	4.5	3.84	0.7	0.0006	Yook *et al.*, 2001b
Power grid	4941	2.67	18.7	12.4	0.08	0.005	Watts and Strogatz, 1998
C. Elegans	282	14	2.65	2.25	0.28	0.05	Watts and Strogatz, 1998

Albert, R. et.al. Rev. Mod. Phy. (2002)

Evolution of Random Graphs





- Network structure goes through a transition

Structural (percolation) phase transition at <k>=1
(or equivalently when p=1/N)

ER Random Network - catch up

Basic characteristics

- Degree distribution

$$p_k = \binom{n-1}{k} p^k (1-p)^{n-1-k} \xrightarrow[N \to \infty]{} p_k = \frac{\langle k \rangle^k}{k!} e^{-\langle k \rangle}$$

Binomial distribution — Poisson distribution

Degree distribution with exponential tail

- Clustering

$$C_i \equiv \frac{1}{N} \langle k \rangle = p$$

Vanishing clustering coefficient for large size

- Path length

$$l = \frac{logN}{log\langle k \rangle}$$

Logarithmically short distance among nodes

ER Random Network - catch up

Network	Degree distribution	Path length	Clustering coefficient
Real world networks	broad	short	large
ER random networks	Poissonian	short	small

It is not capturing the properties of any real system

BUT

it serves as a reference system for any other network model

Six degrees

of

separation

Six degrees - the idea

Frigyes Karinthy: Chains (1929)

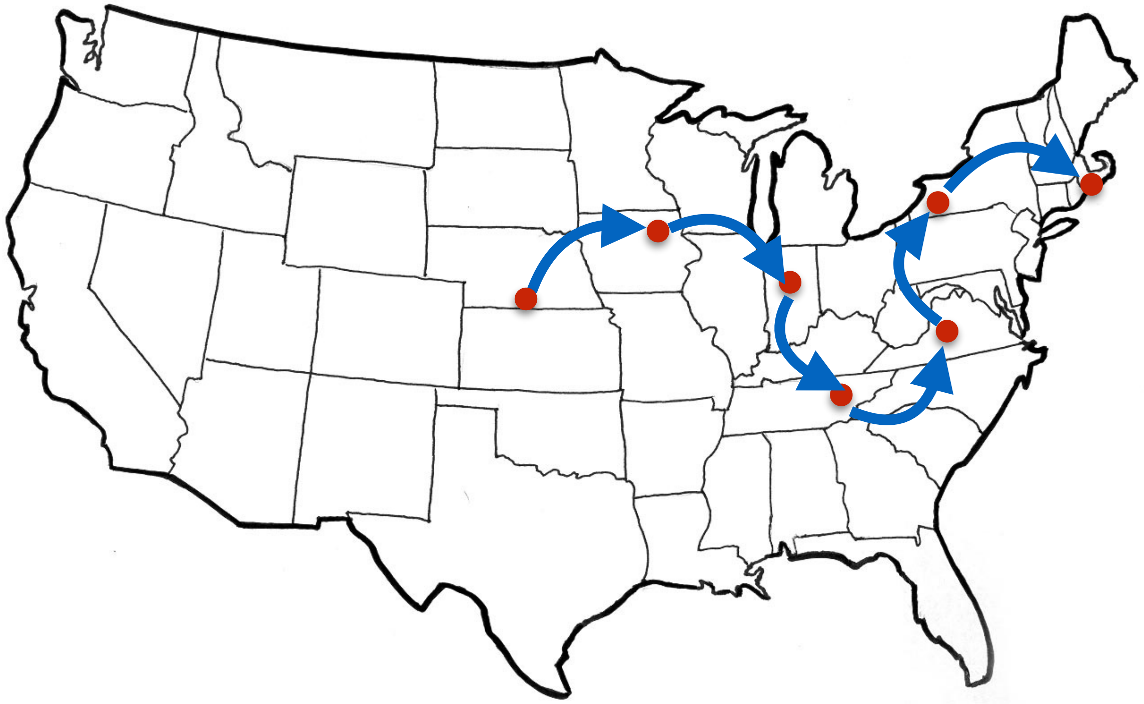
- Classic short story
- Karinthy believed that the modern world was 'shrinking' due to ever-increasing connectedness of human beings
- Excerpt: “A fascinating game grew out of this discussion. One of us suggested performing the following experiment to prove that the population of the Earth is closer 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 - the experiment

Milgram small-world experiment (1967)

Chose people randomly from Omaha (Nebraska) and asked them the following:

- 1. ADD YOUR NAME TO THE ROSTER AT THE BOTTOM OF THIS SHEET, so that the next person who receives this letter will know who it came from.
- 2. DETACH ONE POSTCARD. FILL IT AND RETURN IT TO HARVARD UNIVERSITY. No stamp is needed. The postcard is very important. It allows us to keep track of the progress of the folder as it moves toward the target person.
- 3. IF YOU KNOW THE TARGET PERSON ON A PERSONAL BASIS, MAIL THIS FOLDER DIRECTLY TO HIM (HER). Do this only if you have previously met the target person and know each other on a first name basis.
- 4. IF YOU DO NOT KNOW THE TARGET PERSON ON A PERSONAL BASIS, DO NOT TRY TO CONTACT HIM DIRECTLY. INSTEAD, MAIL THIS FOLDER (POST CARDS AND ALL) TO A PERSONAL ACQUAINTANCE WHO IS MORE LIKELY THAN YOU TO KNOW THE TARGET PERSON. You may send the folder to a friend, relative or acquaintance, but it must be someone you know on a first name basis.



- Results:
  - ~20% of the letters reached the target
  - For these, there were on average 5.5 inter-mediaries
- Conclusions:
  - Short chains exist...and people somehow manage to find them!

Small-world phenomena - six degrees

The Erdős number

- 1525 publications, 511 coauthors
- Erdős number: describe collaborative distance between mathematicians and P. Erdős
- Definition:

- 511 people with EN = 1 and 8162 with EN=2
- Pauli number, Bacon number, Erdős-Bacon number, ...

Small-world phenomena - six degrees

The Erdős number

- 1525 publications, 511 coauthors
- Erdős number: describe collaborative distance between mathematicians and P. Erdős
- Definition:

- 511 people with EN = 1 and 8162 with EN=2
- Pauli number, Bacon number, Erdős-Bacon number, ...

(…my Erdős Number is 2 (via László Lovász))

Small-world networks

- On of the first paper of Network Science…

  D.J. Watts and S. Strogatz,

  ”Collective dynamics of 'small-world' networks”, Nature 393, 440–442, 1998

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. □

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

- Observation in real world networks:

Table 1 Empirical examples of small-world networks

	L_{actual}	L_{random}	C_{actual}	C_{random}	N
Film actors	3.65	2.99	0.79	0.00027	22500
Power grid	18.7	12.4	0.080	0.005	4941
C. elegans	2.65	2.25	0.28	0.05	282

Contradiction: Real-world networks have

High clustering coefficient AND Short distances

Clustering vs. Interconnectedness

Random networks

- **Logarithmically short distance among nodes** $d = \frac{logN}{log\langle k \rangle}$ ✓
- Vanishing clustering coefficient for large size $C_i \equiv \frac{1}{N}\langle k \rangle = p$

Real-world networks

Network	Size	$\langle k \rangle$	ℓ	ℓ_{rand}	C	C_{rand}	Reference
WWW, site level, undir.	153 127	35.21	3.1	3.35	0.1078	0.00023	Adamic, 1999
Internet, domain level	3015–6209	3.52–4.11	3.7–3.76	6.36–6.18	0.18–0.3	0.001	Yook *et al.*, 2001a, Pastor-Satorras *et al.*, 2001
Movie actors	225 226	61	3.65	2.99	0.79	0.00027	Watts and Strogatz, 1998
LANL co-authorship	52 909	9.7	5.9	4.79	0.43	1.8×10^{-4}	Newman, 2001a, 2001b, 2001c
MEDLINE co-authorship	1 520 251	18.1	4.6	4.91	0.066	1.1×10^{-5}	Newman, 2001a, 2001b, 2001c
SPIRES co-authorship	56 627	173	4.0	2.12	0.726	0.003	Newman, 2001a, 2001b, 2001c
NCSTRL co-authorship	11 994	3.59	9.7	7.34	0.496	3×10^{-4}	Newman, 2001a, 2001b, 2001c
Math. co-authorship	70 975	3.9	9.5	8.2	0.59	5.4×10^{-5}	Barabási *et al.*, 2001
Neurosci. co-authorship	209 293	11.5	6	5.01	0.76	5.5×10^{-5}	Barabási *et al.*, 2001
E. coli, substrate graph	282	7.35	2.9	3.04	0.32	0.026	Wagner and Fell, 2000
E. coli, reaction graph	315	28.3	2.62	1.98	0.59	0.09	Wagner and Fell, 2000
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06	Montoya and Solé, 2000
Silwood Park food web	154	4.75	3.40	3.23	0.15	0.03	Montoya and Solé, 2000
Words, co-occurrence	460.902	70.13	2.67	3.03	0.437	0.0001	Ferrer i Cancho and Solé, 2001
Words, synonyms	22 311	13.48	4.5	3.84	0.7	0.0006	Yook *et al.*, 2001b
Power grid	4941	2.67	18.7	12.4	0.08	0.005	Watts and Strogatz, 1998
C. Elegans	282	14	2.65	2.25	0.28	0.05	Watts and Strogatz, 1998

Albert, R. et.al. Rev. Mod. Phy. (2002)

Clustering vs. Interconnectedness

Random networks

- Logarithmically short distance among nodes $d = \frac{logN}{log\langle k\rangle}$ ✓
- **Vanishing clustering coefficient for large size** $C_i \equiv \frac{1}{N}\langle k\rangle = p$ ✗

Real-world networks

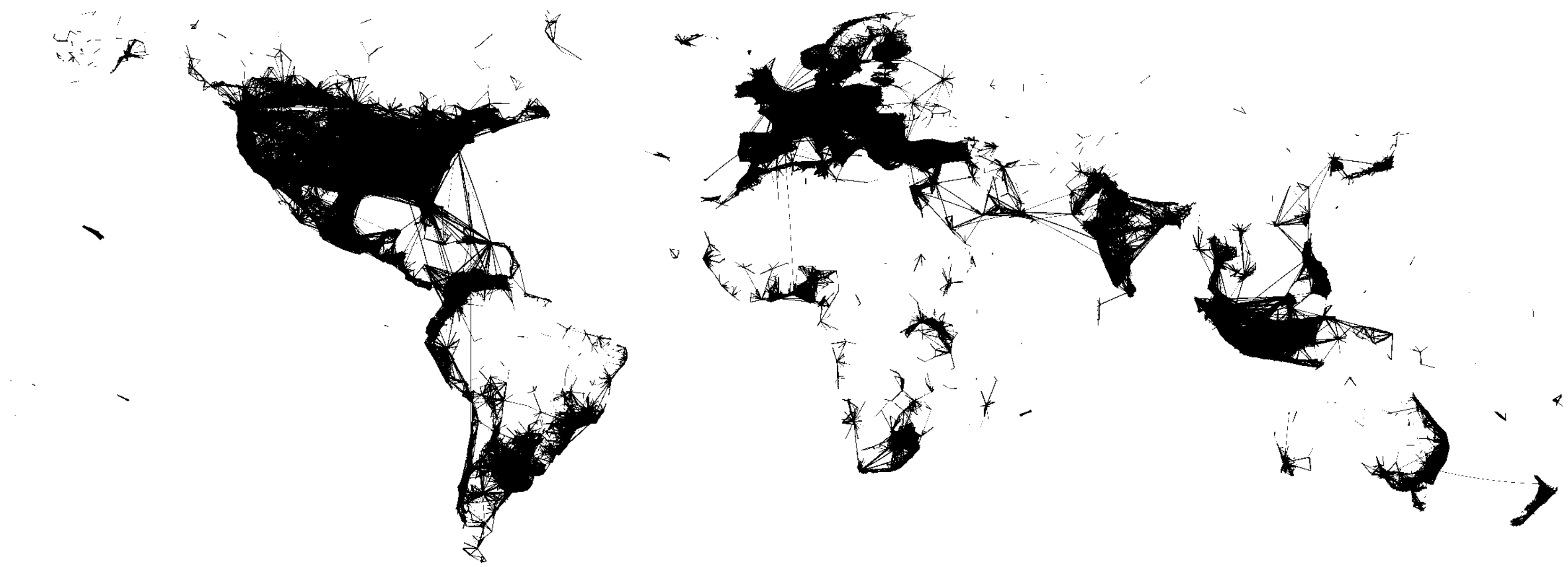
Network	Size	$\langle k\rangle$	ℓ	ℓ_{rand}	C	C_{rand}	Reference
WWW, site level, undir.	153 127	35.21	3.1	3.35	0.1078	0.00023	Adamic, 1999
Internet, domain level	3015–6209	3.52–4.11	3.7–3.76	6.36–6.18	0.18–0.3	0.001	Yook *et al.*, 2001a, Pastor-Satorras *et al.*, 2001
Movie actors	225 226	61	3.65	2.99	0.79	0.00027	Watts and Strogatz, 1998
LANL co-authorship	52 909	9.7	5.9	4.79	0.43	1.8×10^{-4}	Newman, 2001a, 2001b, 2001c
MEDLINE co-authorship	1 520 251	18.1	4.6	4.91	0.066	1.1×10^{-5}	Newman, 2001a, 2001b, 2001c
SPIRES co-authorship	56 627	173	4.0	2.12	0.726	0.003	Newman, 2001a, 2001b, 2001c
NCSTRL co-authorship	11 994	3.59	9.7	7.34	0.496	3×10^{-4}	Newman, 2001a, 2001b, 2001c
Math. co-authorship	70 975	3.9	9.5	8.2	0.59	5.4×10^{-5}	Barabási *et al.*, 2001
Neurosci. co-authorship	209 293	11.5	6	5.01	0.76	5.5×10^{-5}	Barabási *et al.*, 2001
E. coli, substrate graph	282	7.35	2.9	3.04	0.32	0.026	Wagner and Fell, 2000
E. coli, reaction graph	315	28.3	2.62	1.98	0.59	0.09	Wagner and Fell, 2000
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06	Montoya and Solé, 2000
Silwood Park food web	154	4.75	3.40	3.23	0.15	0.03	Montoya and Solé, 2000
Words, co-occurrence	460.902	70.13	2.67	3.03	0.437	0.0001	Ferrer i Cancho and Solé, 2001
Words, synonyms	22 311	13.48	4.5	3.84	0.7	0.0006	Yook *et al.*, 2001b
Power grid	4941	2.67	18.7	12.4	0.08	0.005	Watts and Strogatz, 1998
C. Elegans	282	14	2.65	2.25	0.28	0.05	Watts and Strogatz, 1998

Albert R. et.al. Rev. Mod. Phy. (2002)

Clustering vs. Interconnectedness

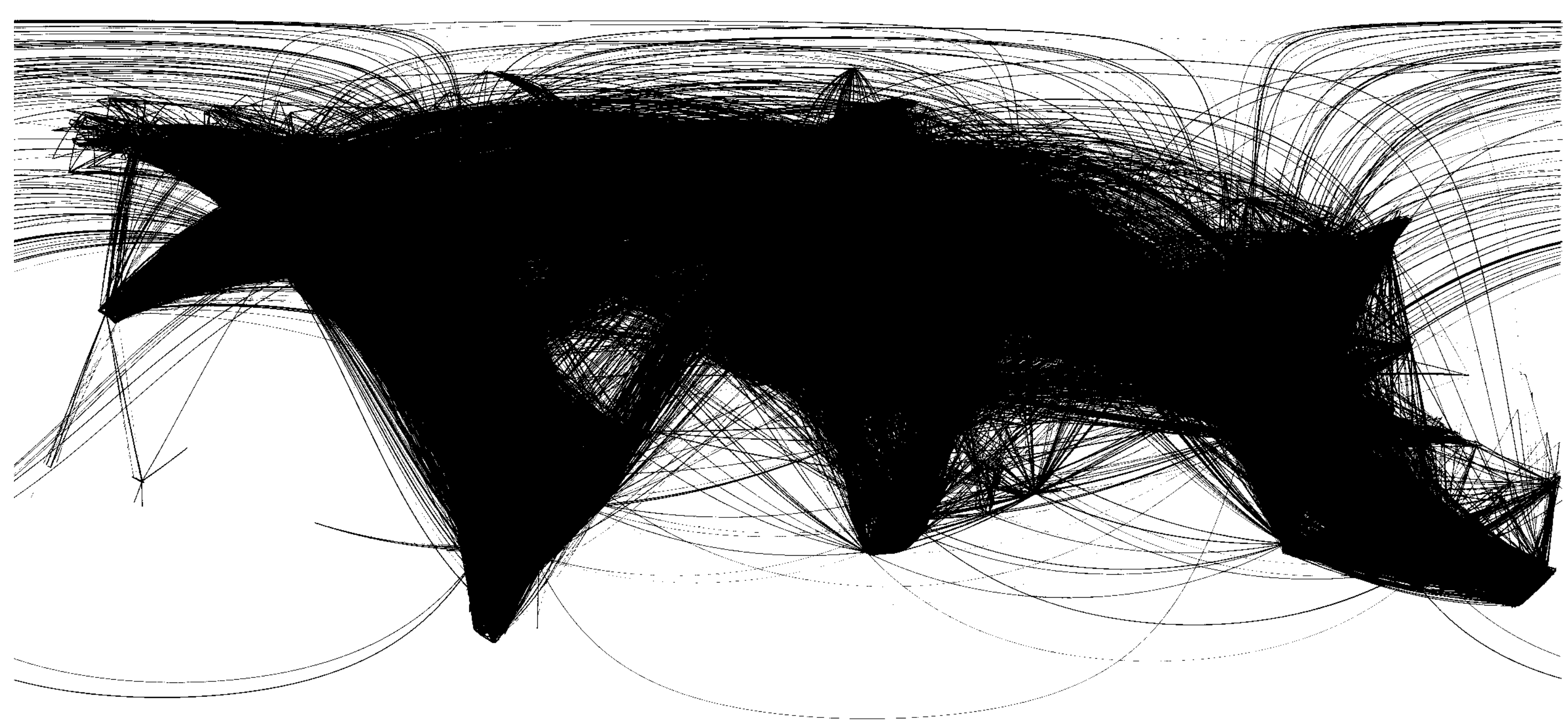
High clustering

- Locally structured
- No connections between nodes apart



Random

- Globally interconnected
- Low clustering



Clustering vs. Interconnectedness

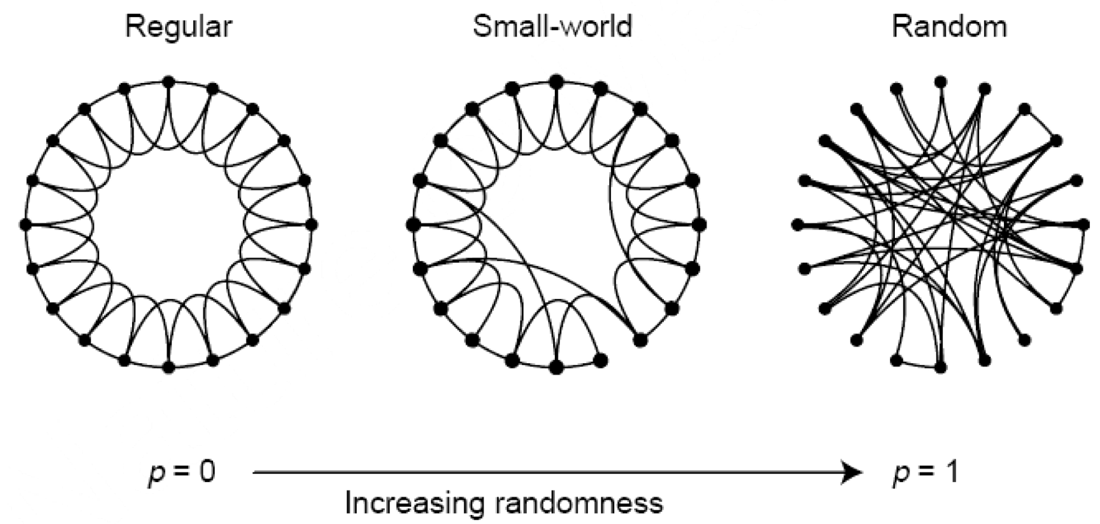
Real networks have high clustering and short distances

Watts-Strogatz
model of
small-world networks

The Watts-Strogatz model

A model to capture large clustering coefficient and short distances observed in real networks

- It interpolates between an ordered finite lattice and a random graph
- Fixed parameters:
  - N - system size
  - K - initial coordination number
- Variable parameters:
  - p - rewiring probability



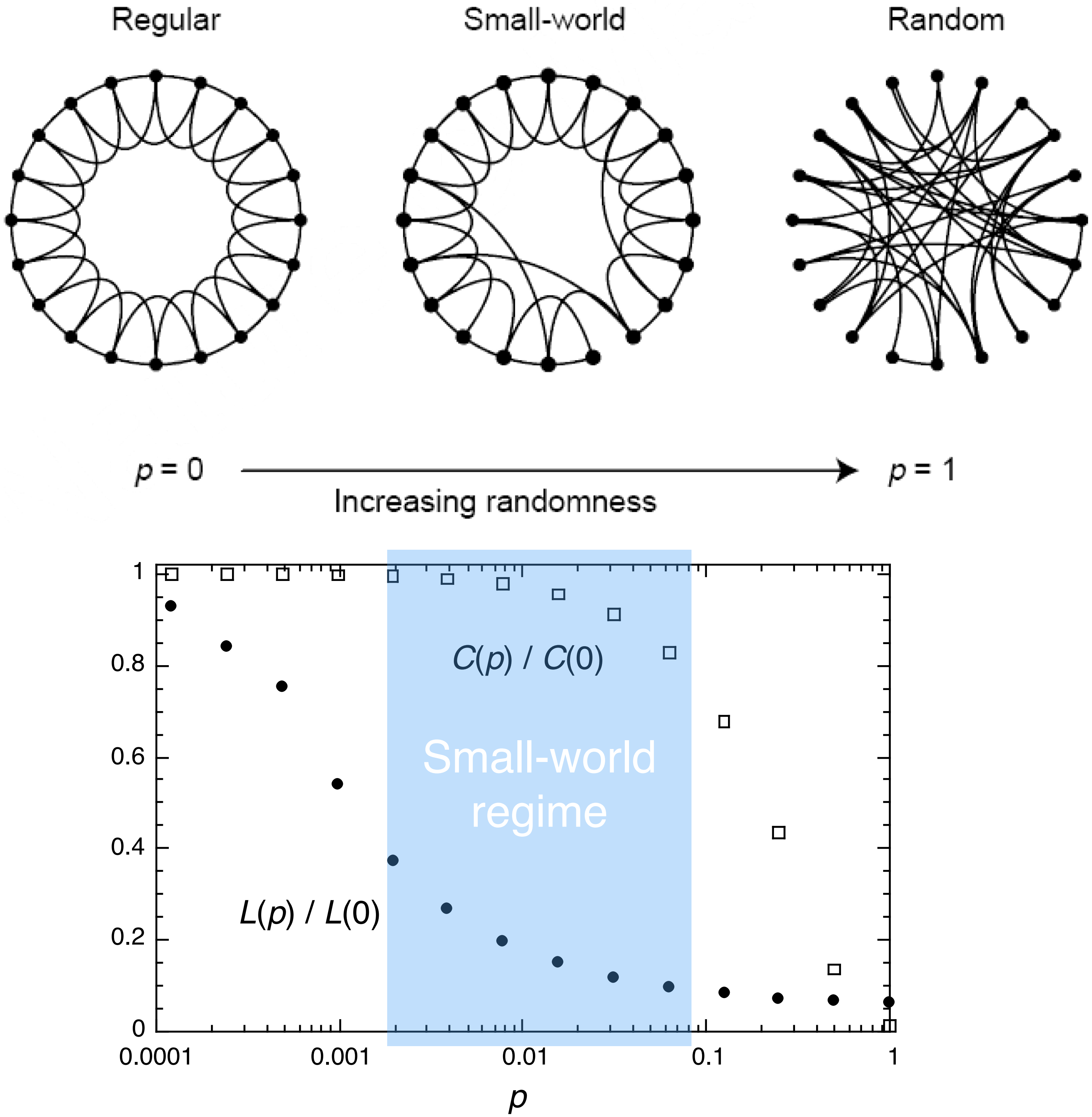
D.J. Watts and S. Strogatz, Nature (1998)

- Algorithm:
  1. Start with a ring lattice with N nodes in which every node is connected to its first K neighbours ($K/2$ on either side).
  2. Randomly rewire each edge of the lattice with probability p such that self-connections and duplicate edges are excluded.

By varying p the network can be transformed from a completely ordered ($p=0$) to a completely random ($p=1$) structure

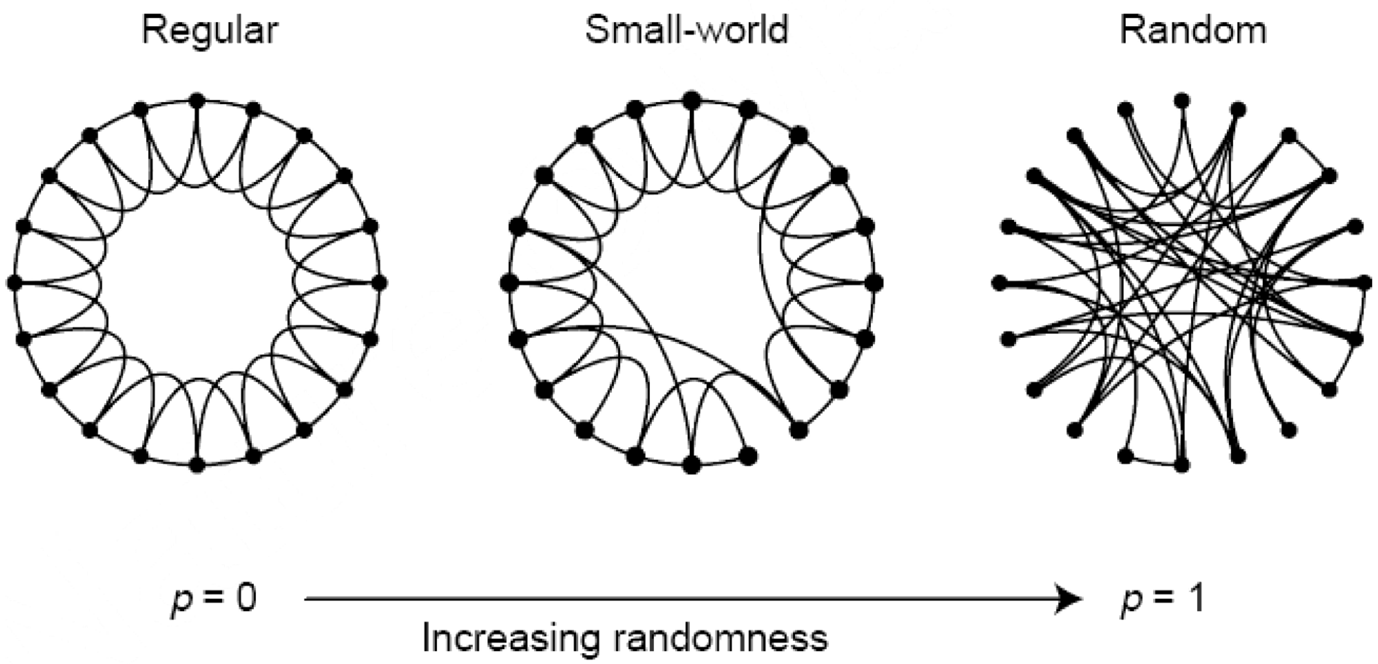
The Watts-Strogatz model

- N and K are chosen $N \gg K \gg \ln(N) \gg 1$ thus the random graph remains connected ($K \gg \ln(N)$)



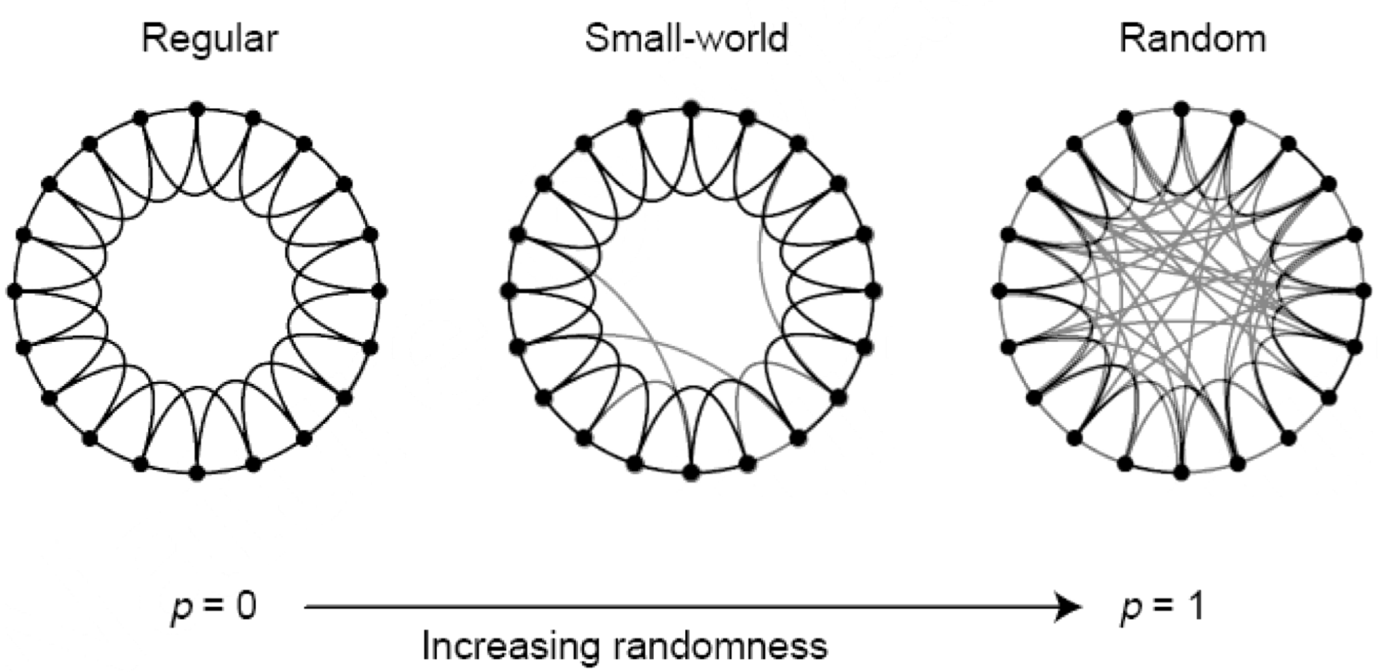
The Watts-Strogatz model

- N and K are chosen $N \gg K \gg \ln(N) \gg 1$ thus the random graph remains connected ($K \gg \ln(N)$)



- Definition 1:

  1. Start with a ring lattice with N nodes in which every node is connected to its first K neighbours ($K/2$ on either side).
  2. Randomly rewire each edge of the lattice with probability p such that self-connections and duplicate edges are excluded.

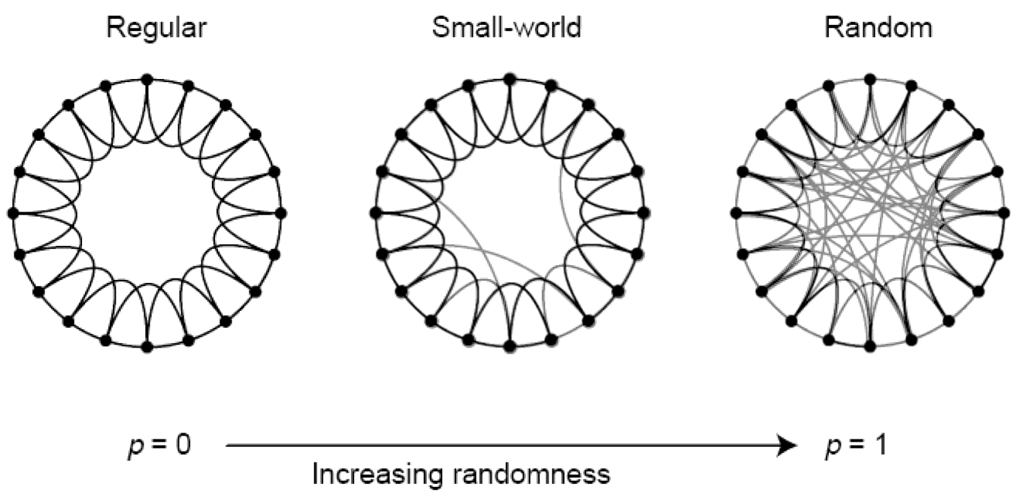


Definition 2:

1. Start with a ring lattice with N nodes in which every node is connected to its first K neighbours ($K/2$ on either side).
2. For every edge in the network we add an additional edge with independent probability p, connected two nodes selected uniformly at random

The Watts-Strogatz model

Degree distribution (Definition 2)



- p=0 - each node has the same degree K
- p>0 - each node has degree K + shortcut links
    - Number of shortcut edges: $s = \frac{1}{2} NK \times p$
    - Each node will have on average *Kp* number of shortcuts
    - Probability of having *s* number of shortcut edges: $P(s) = e^{-Kp} \frac{(Kp)^s}{s!}$
    - Total degree *k= K + s*, so *s=k-K* and the degree distribution is

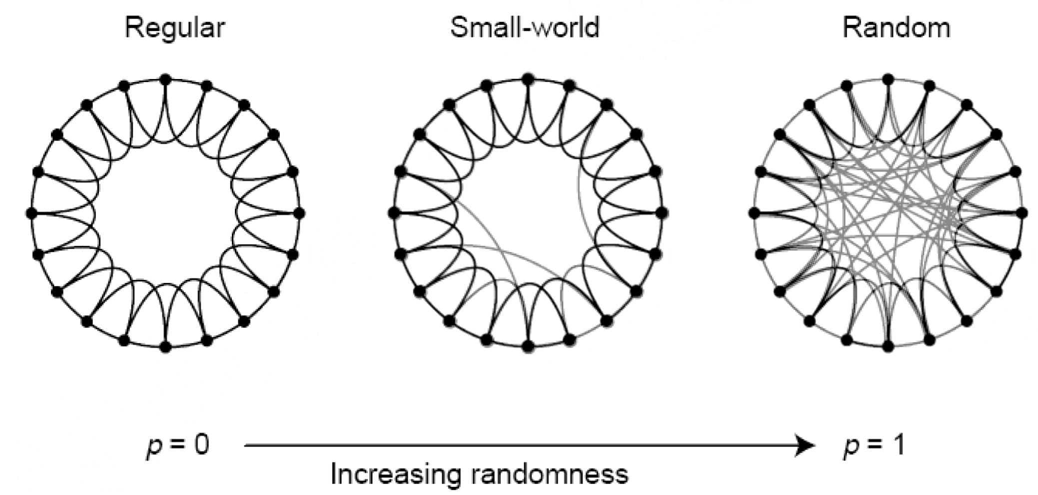
$$\boxed{P(k) = e^{-Kp} \frac{(Kp)^{(k-K)}}{(k-K)!}}$$

if *k≥K* and *P(k)=0* if *k<K*

- p=0 - it is a delta function at K
- p>0 - approximates a **Poisson distribution** just like a random network

The Watts-Strogatz model

(Global) Clustering coefficient (Definition 2)



- *p=0* - regular ring with constant clustering: $C = \dfrac{3(K-2)}{4(K-1)}$
  - *$0 \leq C \leq 3/4$*
  - Independent of *N*
- *p>0* - we can count triangles and tuples

Number of triangles

- We use Definition 2 so the number of triangles in the ring is constant: $\dfrac{1}{4}NK(\dfrac{1}{2}K-1)$
- Number of triangles created by 1, 2 or 3 shortcut links can be neglected as their number goes to zero as N goes large (their number is constant $\sim Kp$)

Number of triplets

- Triplets formed by links in the ring: $\dfrac{1}{2}NK(K-1)$
- Triplets formed by shortcut and ring links: NK^2p
- Triplets formed by two shortcut links: $\dfrac{1}{2}NK^2p^2$

Global clustering coefficient

$$\boxed{C = \frac{\frac{1}{4}NK(\frac{1}{2}K-1)\times 3}{\frac{1}{2}NK(K-1)+NK^2p+\frac{1}{2}NK^2p^2} = \frac{3(K-2)}{4(K-1)+8Kp+4Kp^2}}$$

- Independent of N
- if p→0 it recovers the ring value
- if p→1 it well approximates 1

The Watts-Strogatz model

Average path length (Definition 2)

- No closed form solution
- General functional dependence (Barthalémy&Amaral)

$$\ell = \frac{N}{K} f(NKp)$$

- From numerical simulations the approximated function:

$$f(x) = \frac{2}{\sqrt{x^2+4x}} \tanh^{-1} \sqrt{\frac{x}{x+4}}$$

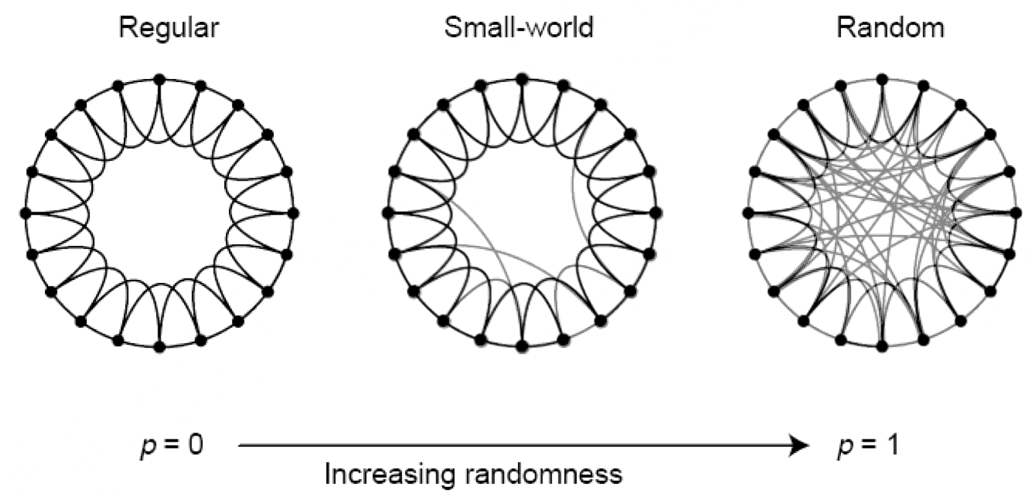
- Using identity of *tanh-1*: $\tanh^{-1} u = \frac{1}{2} \ln \frac{1+u}{1-u}$

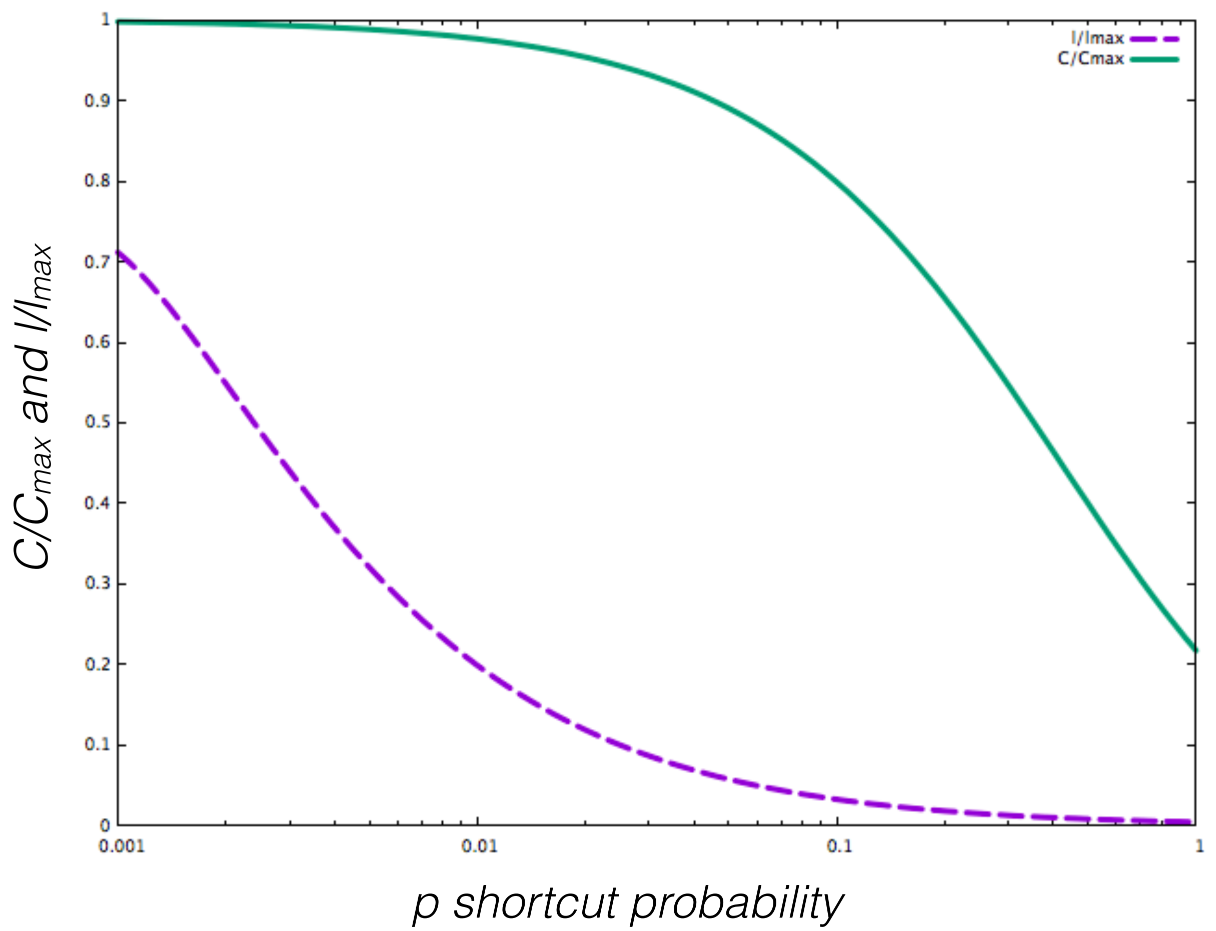
leading to $$f(x) = \frac{2}{\sqrt{x^2+4x}} \ln \frac{\sqrt{1+4/x}+1}{\sqrt{1+4/x}-1}$$

- In the large x limit it becomes: $f(x) = \frac{\ln x}{x}$

- average path length $$\boxed{\ell = \frac{\ln(NKp)}{K^2 p}}$$

WS networks are small-worlds





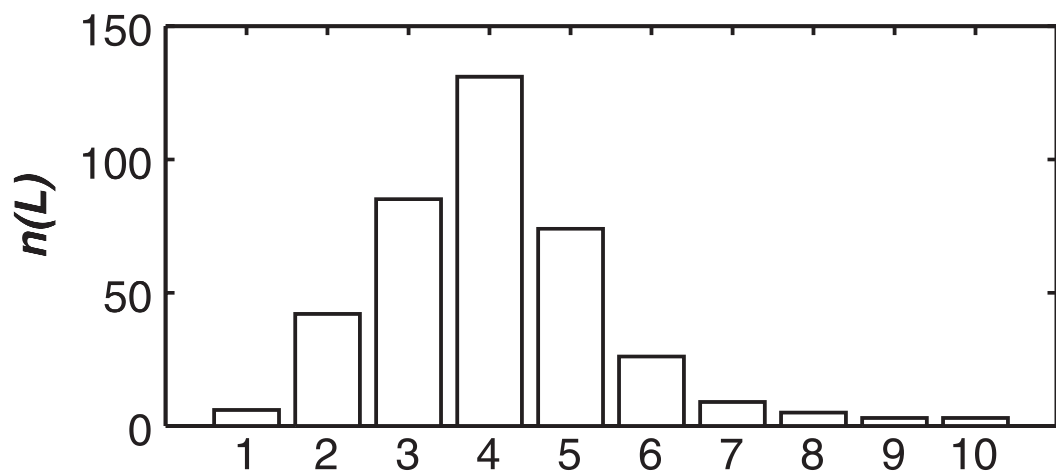
The Watts experiment

- Repeated the experiment of Milgram on email networks
- 18 targets in 13 countries
- 61,168 starters signed on, and 24,163 chains were begun. Of those, only 384 were completed
- $\langle l \rangle = 4$ (6 when accounting for broken chains)

An Experimental Study of Search in Global Social Networks

Peter Sheridan Dodds,[1] Roby Muhamad,[2] Duncan J. Watts[1,2]*

We report on a global social-search experiment in which more than 60,000 e-mail users attempted to reach one of 18 target persons in 13 countries by forwarding messages to acquaintances. We find that successful social search is conducted primarily through intermediate to weak strength ties, does not require highly connected "hubs" to succeed, and, in contrast to unsuccessful social search, disproportionately relies on professional relationships. By accounting for the attrition of message chains, we estimate that social searches can reach their targets in a median of five to seven steps, depending on the separation of source and target, although small variations in chain lengths and participation rates generate large differences in target reachability. We conclude that although global social networks are, in principle, searchable, actual success depends sensitively on individual incentives.



www.sciencemag.org SCIENCE VOL 301 8 AUGUST 2003

Real small-world networks

Network	Size	$\langle k \rangle$	ℓ	ℓ_{rand}	C	C_{rand}	Reference
WWW, site level, undir.	153 127	35.21	3.1	3.35	0.1078	0.00023	Adamic, 1999
Internet, domain level	3015–6209	3.52–4.11	3.7–3.76	6.36–6.18	0.18–0.3	0.001	Yook *et al.*, 2001a, Pastor-Satorras *et al.*, 2001
Movie actors	225 226	61	3.65	2.99	0.79	0.00027	Watts and Strogatz, 1998
LANL co-authorship	52 909	9.7	5.9	4.79	0.43	1.8×10^{-4}	Newman, 2001a, 2001b, 2001c
MEDLINE co-authorship	1 520 251	18.1	4.6	4.91	0.066	1.1×10^{-5}	Newman, 2001a, 2001b, 2001c
SPIRES co-authorship	56 627	173	4.0	2.12	0.726	0.003	Newman, 2001a, 2001b, 2001c
NCSTRL co-authorship	11 994	3.59	9.7	7.34	0.496	3×10^{-4}	Newman, 2001a, 2001b, 2001c
Math. co-authorship	70 975	3.9	9.5	8.2	0.59	5.4×10^{-5}	Barabási *et al.*, 2001
Neurosci. co-authorship	209 293	11.5	6	5.01	0.76	5.5×10^{-5}	Barabási *et al.*, 2001
E. coli, substrate graph	282	7.35	2.9	3.04	0.32	0.026	Wagner and Fell, 2000
E. coli, reaction graph	315	28.3	2.62	1.98	0.59	0.09	Wagner and Fell, 2000
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06	Montoya and Solé, 2000
Silwood Park food web	154	4.75	3.40	3.23	0.15	0.03	Montoya and Solé, 2000
Words, co-occurrence	460.902	70.13	2.67	3.03	0.437	0.0001	Ferrer i Cancho and Solé, 2001
Words, synonyms	22 311	13.48	4.5	3.84	0.7	0.0006	Yook *et al.*, 2001b
Power grid	4941	2.67	18.7	12.4	0.08	0.005	Watts and Strogatz, 1998
C. Elegans	282	14	2.65	2.25	0.28	0.05	Watts and Strogatz, 1998

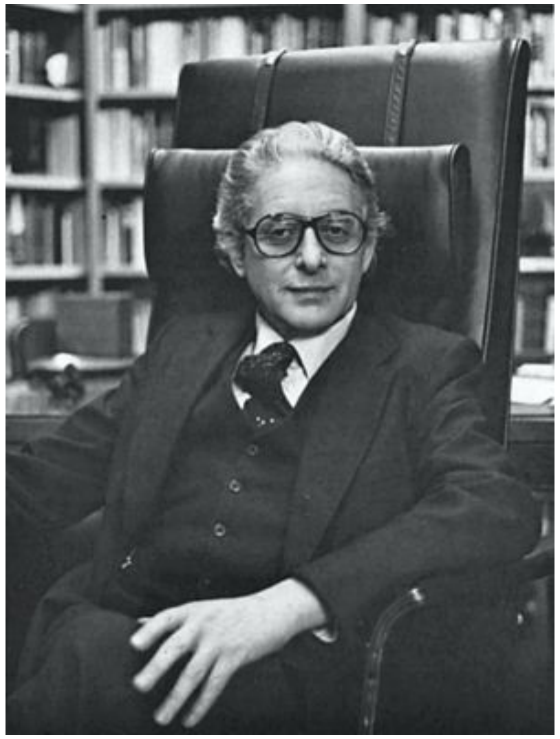
Albert, R. et.al. Rev. Mod. Phy. (2002)

WS networks

Network	Degree distribution	Path length	Clustering coefficient
Real world networks	broad	short	large
ER random networks	Poissonian	short	small
WS small-world networks	Exponential	short	large

Scale-free networks

Scale-free networks - first observations



Networks of scientific papers **Derek J. de Solla Price**, Science (1965)

- Nodes: scientific papers, Links: citations between them
- Number of citations to scientific papers shows a heavy-tailed distribution
- It can be characterised as a Pareto distribution or power-law distribution

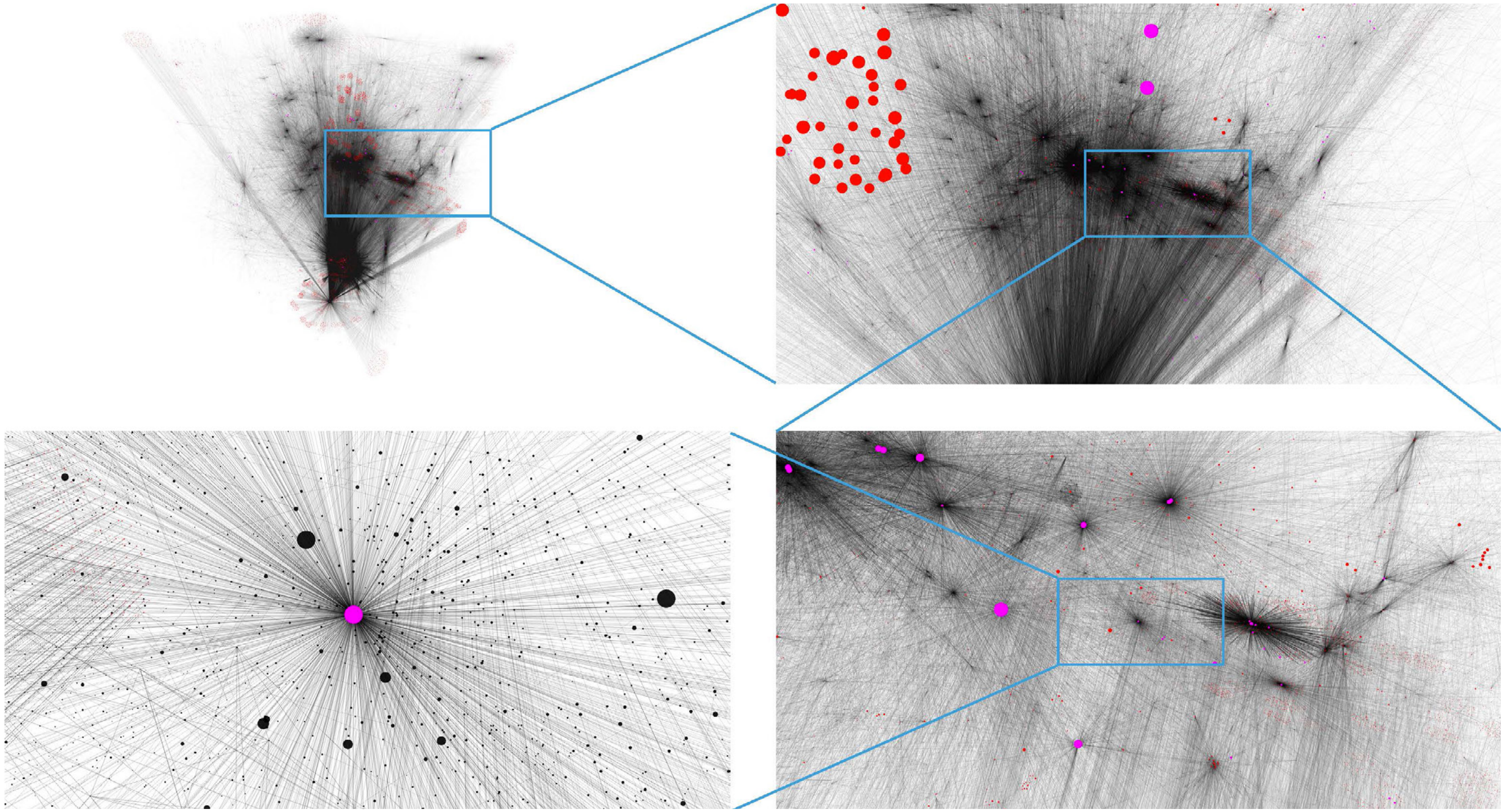
Structure of the WWW **R. Albert, H. Jeong, A-L Barabási**, Nature (1999)

- Nodes: WWW documents, Links: URL links
- More than 3 billions of documents
- Collection by a robot which explores all URL links in a document (web site) and follow them recursively
- They found a heavy-tailed degree distribution which could be well approximated with a power-law function

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

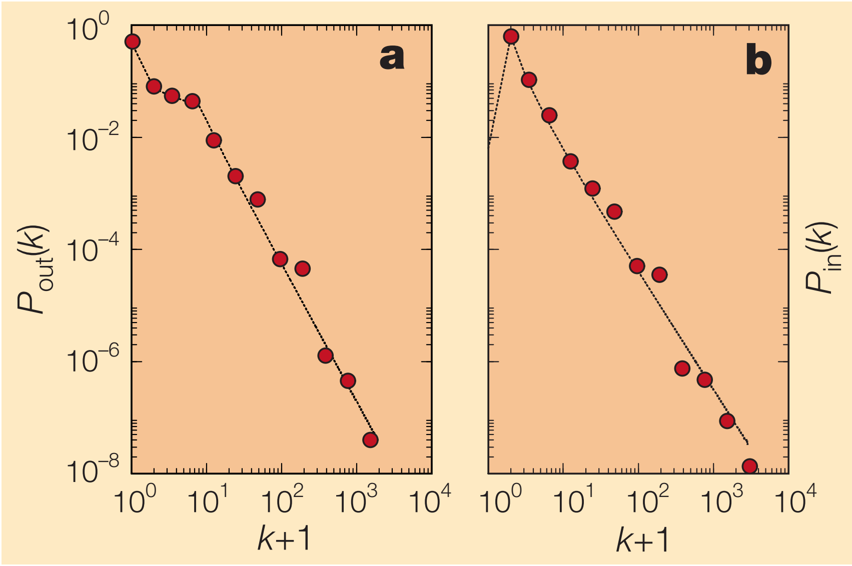
- It is a scale-free network

Scale-free networks - first observations



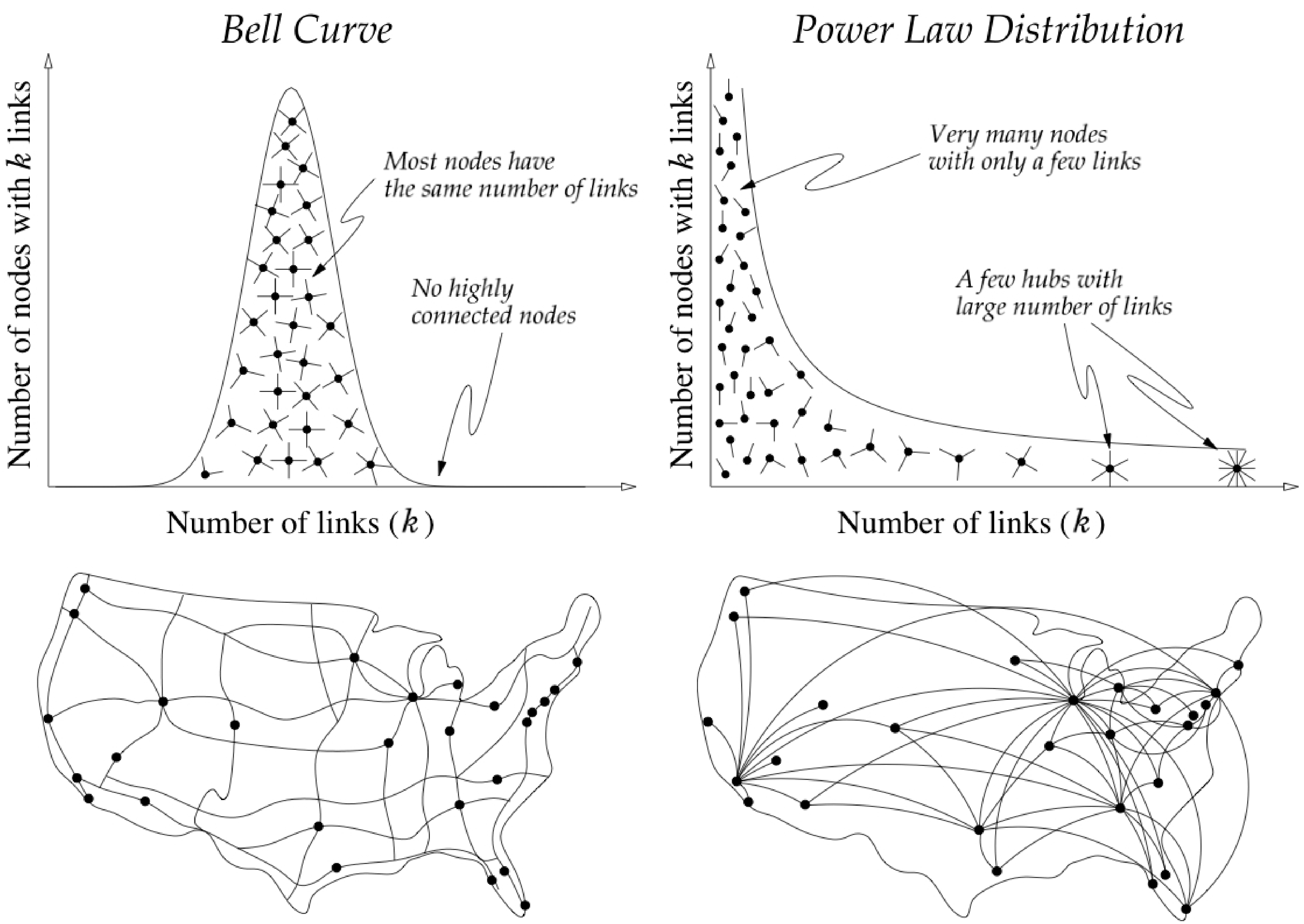
AL Barabási, Network Science Book (2013)

R. Albert, H. Jeong, A-L Barabási, Nature (1999)



Scale-free distribution

What does it mean?



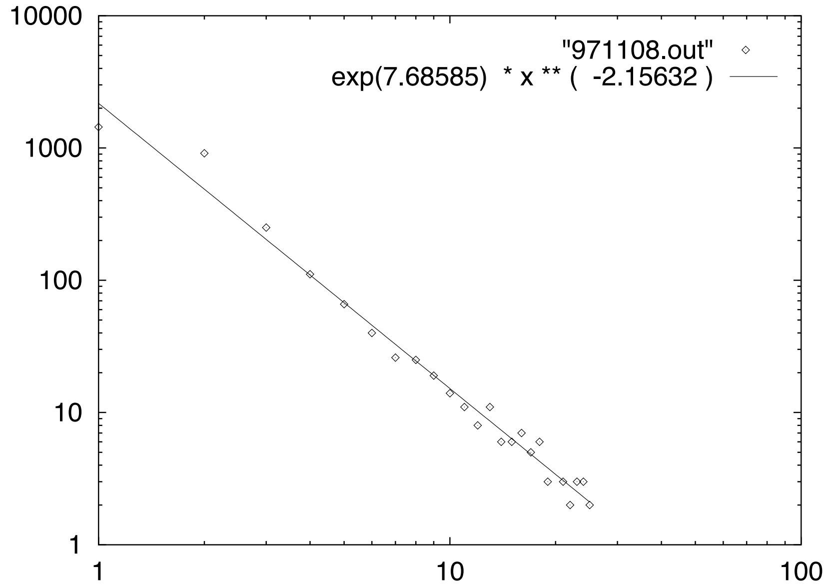
AL. Barabási, Linked (2002)

Degree fluctuations have no characteristic scale (scale invariant)

Scale-free networks - other examples

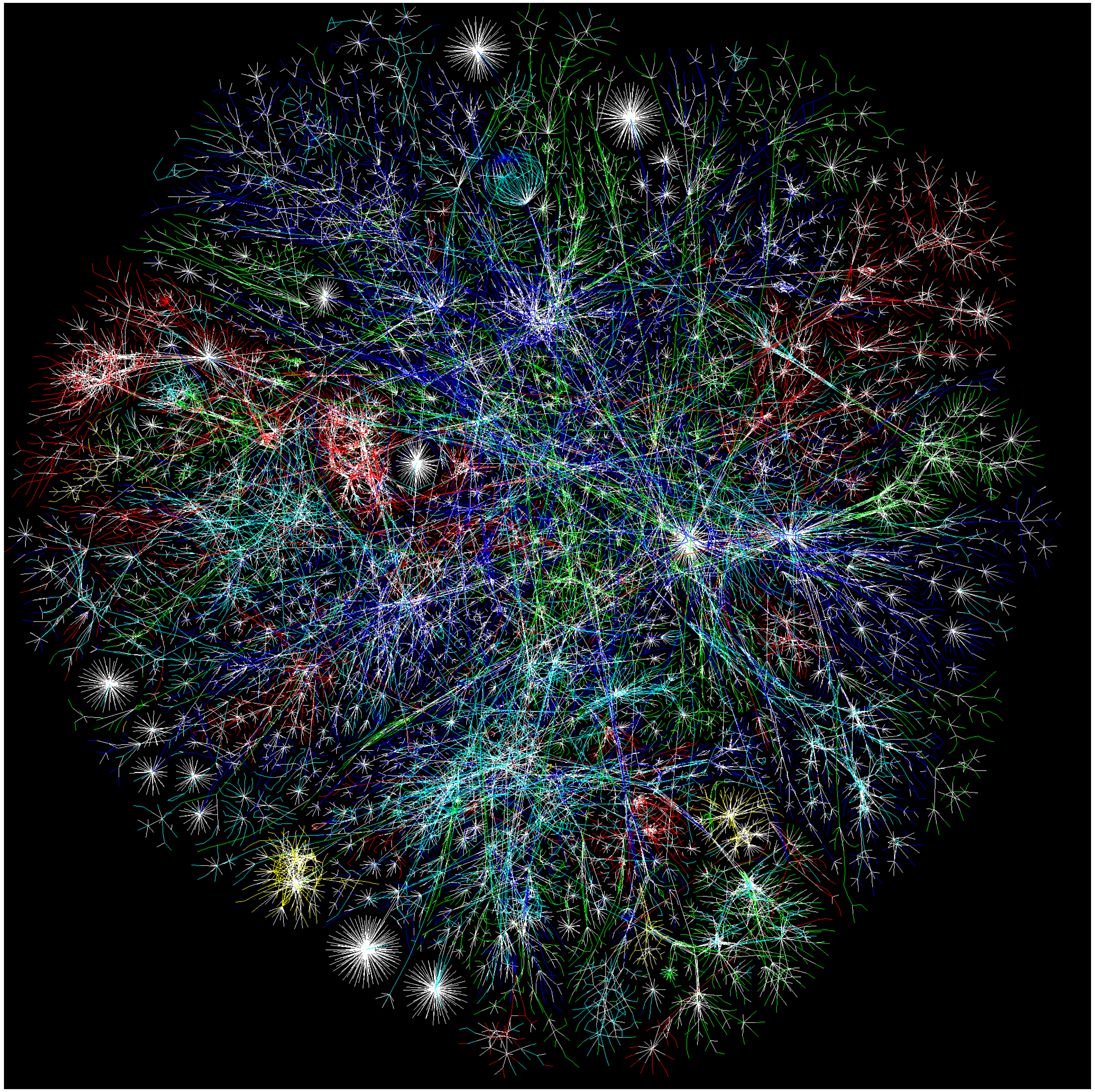
The internet

- Nodes: routers
- Links: Physical wires



(a) Int-11-97

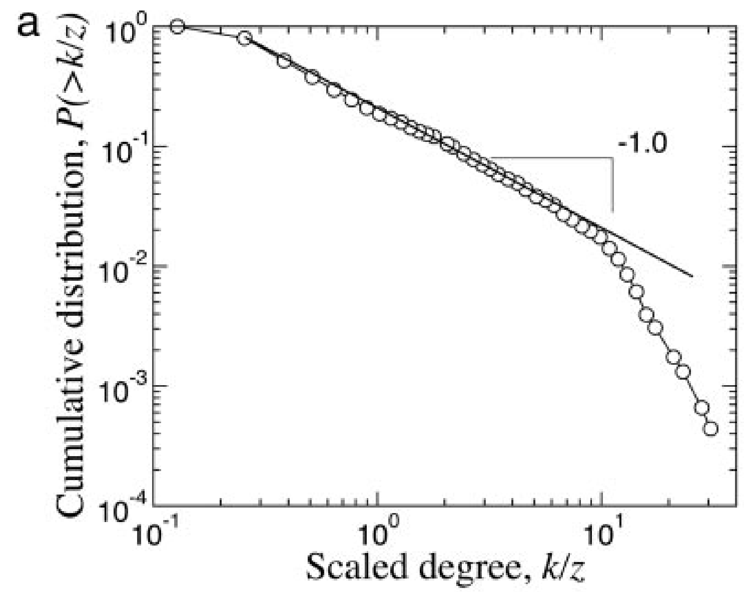
Faloutsos, Faloutsos and Faloutsos (1999)



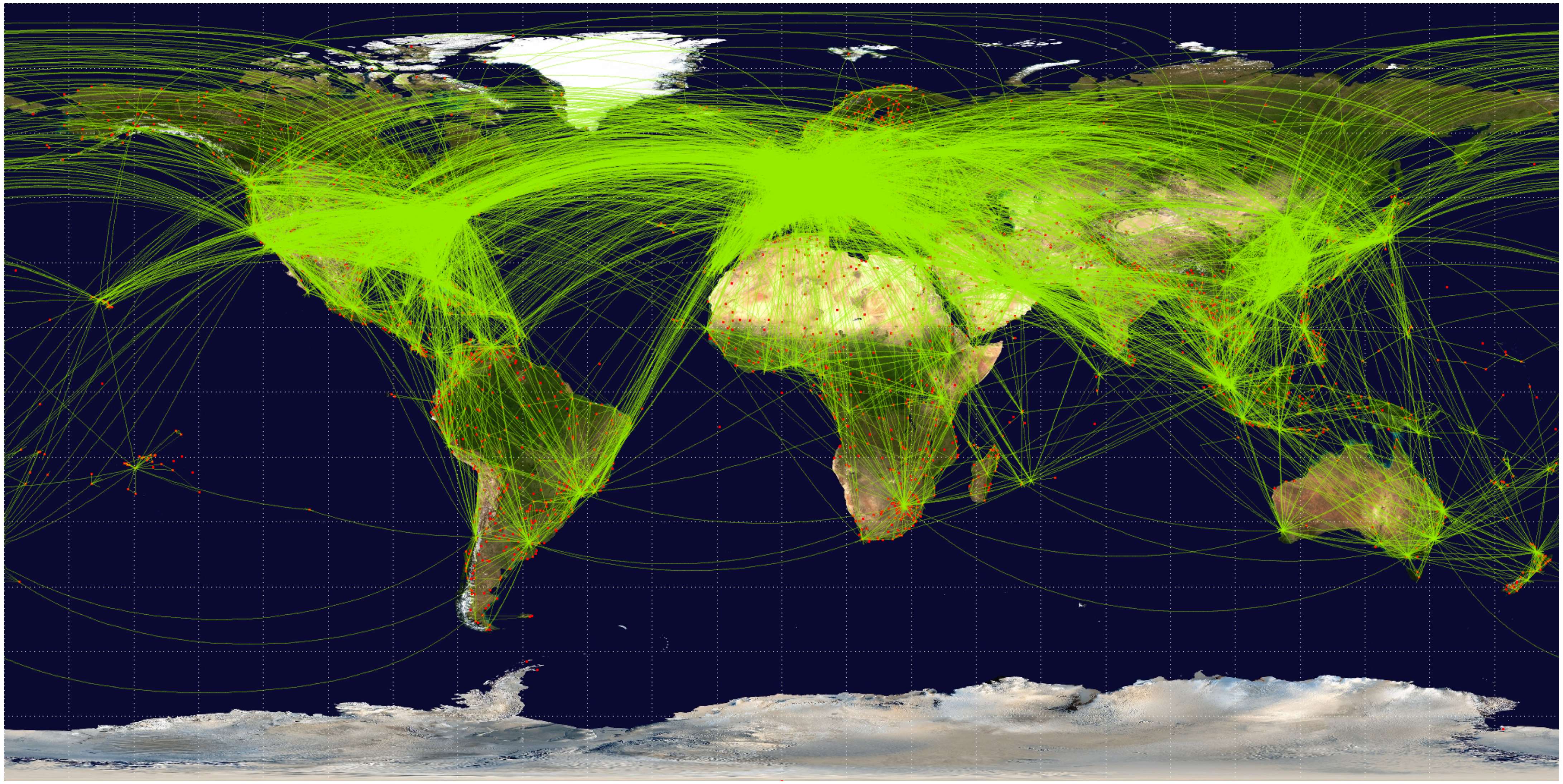
Scale-free networks - other examples

Airline route map network

- Nodes: airports
- Links: airplane connections



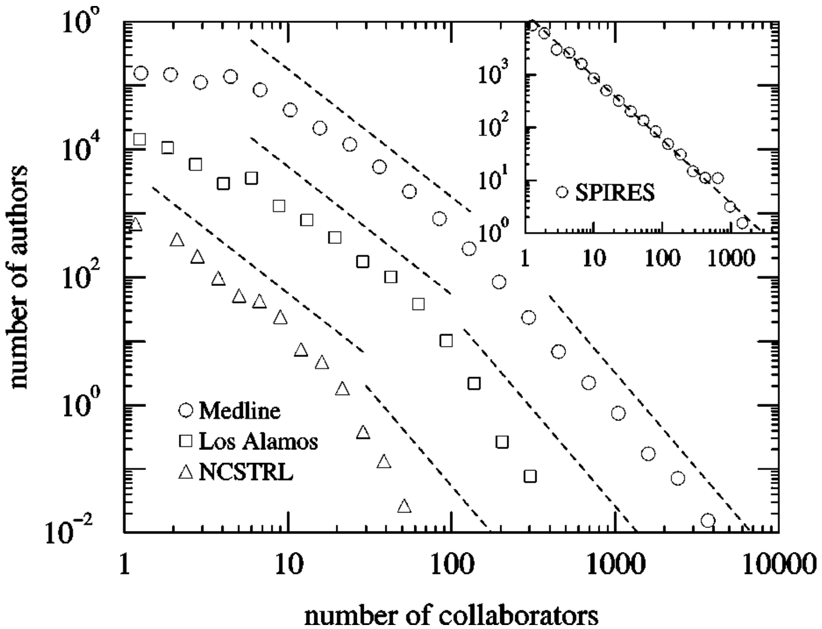
Guimera et.al. (2004)



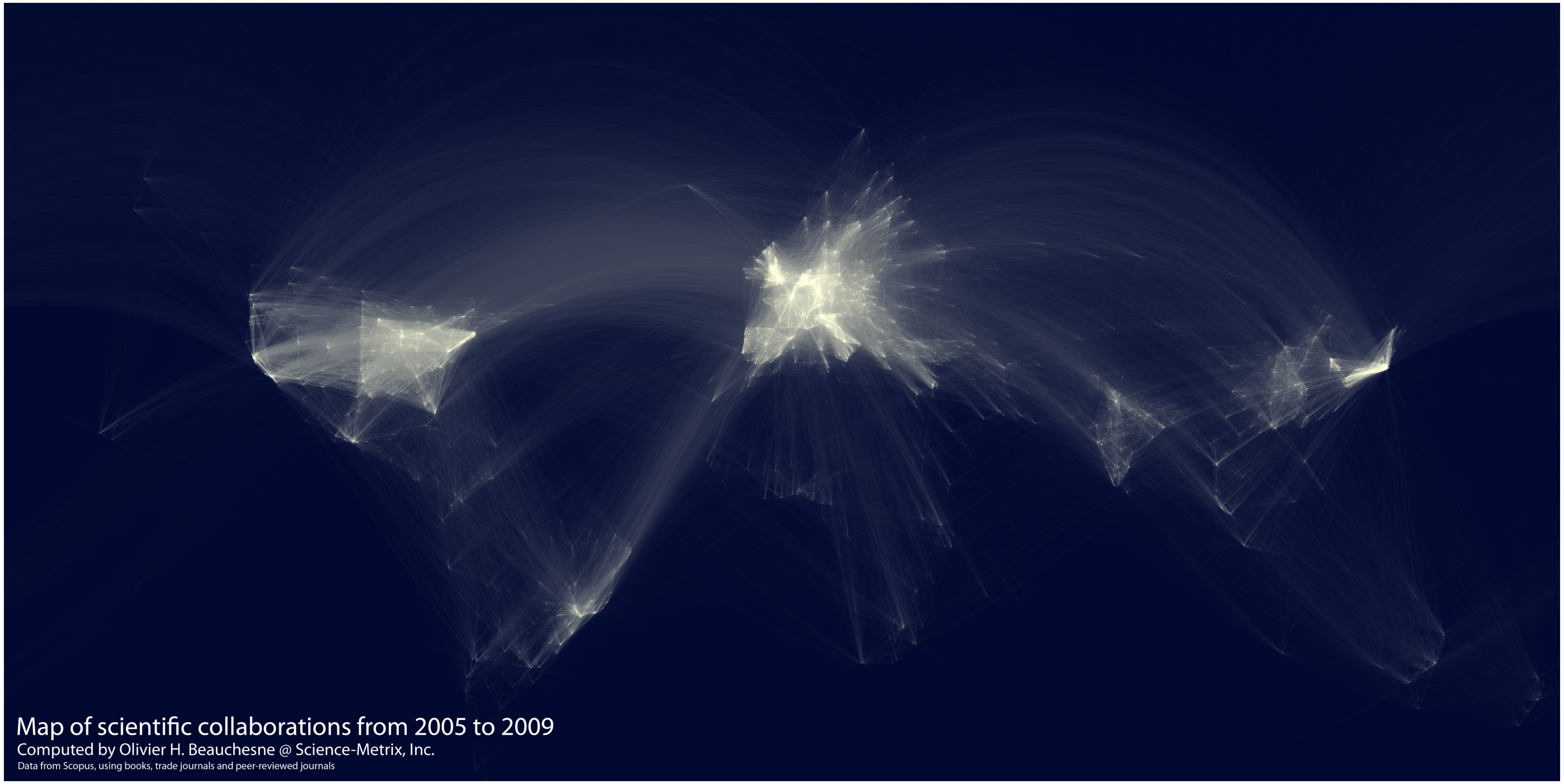
Scale-free networks - other examples

Scientific collaborations

- Nodes: scientists (here geo-localised)
- Links: common papers



Newman (2001)

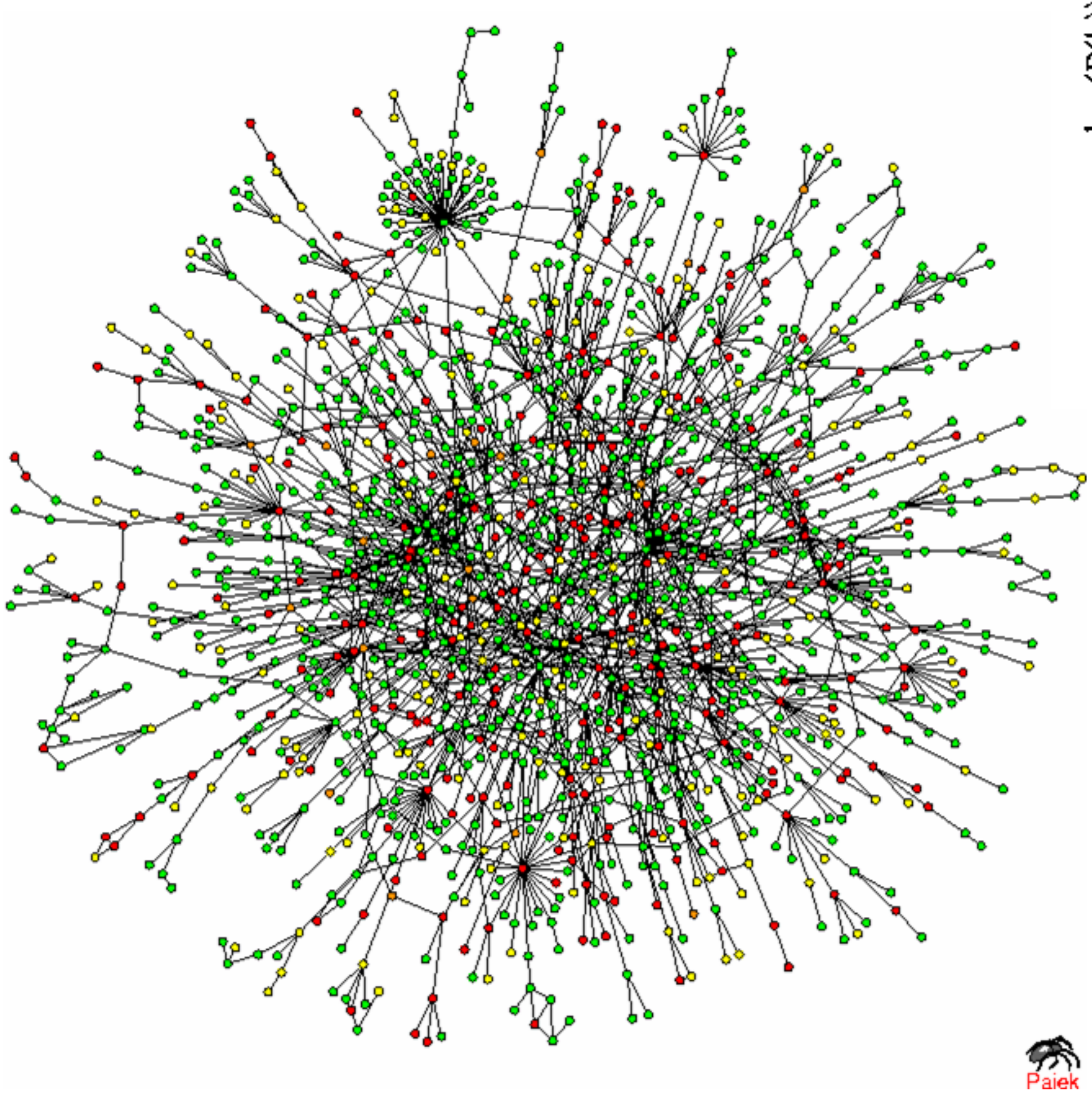


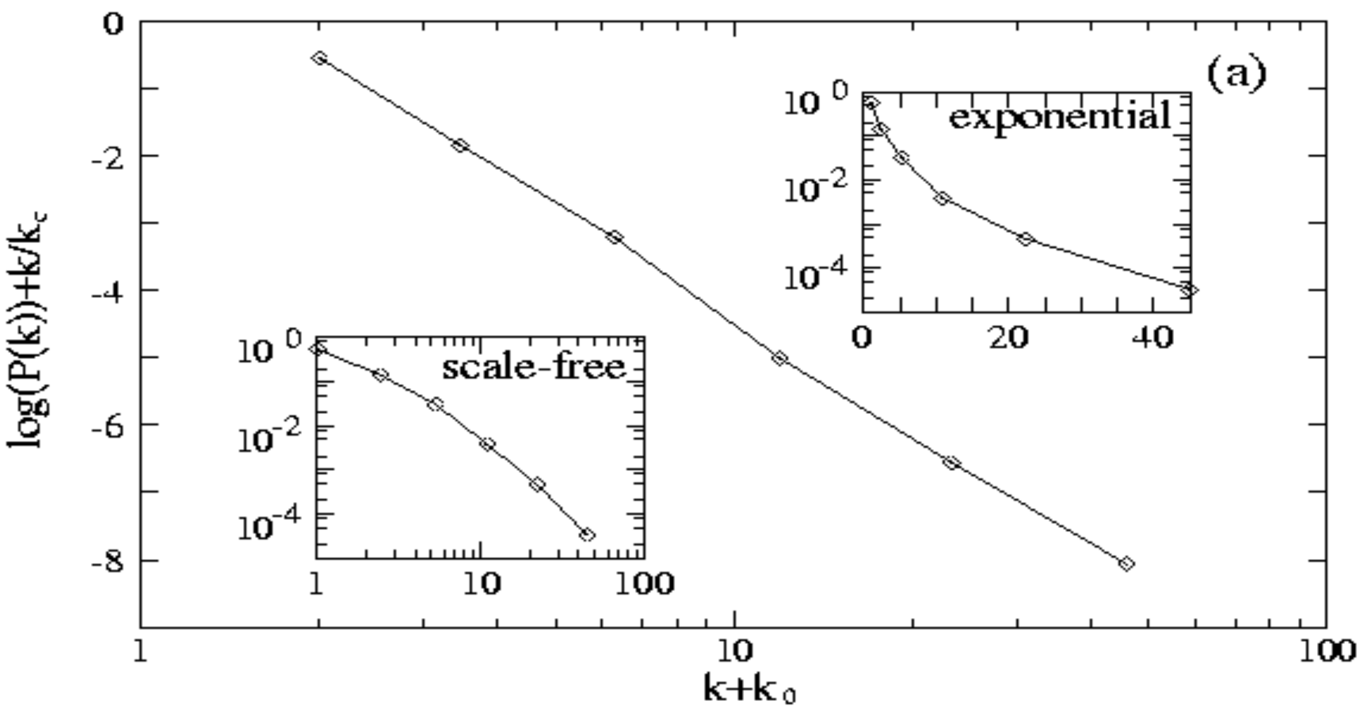
Scale-free networks - other examples

Protein networks

Jeong et.al. (2001)

- Nodes: proteins
- Links: physical interactions-binding





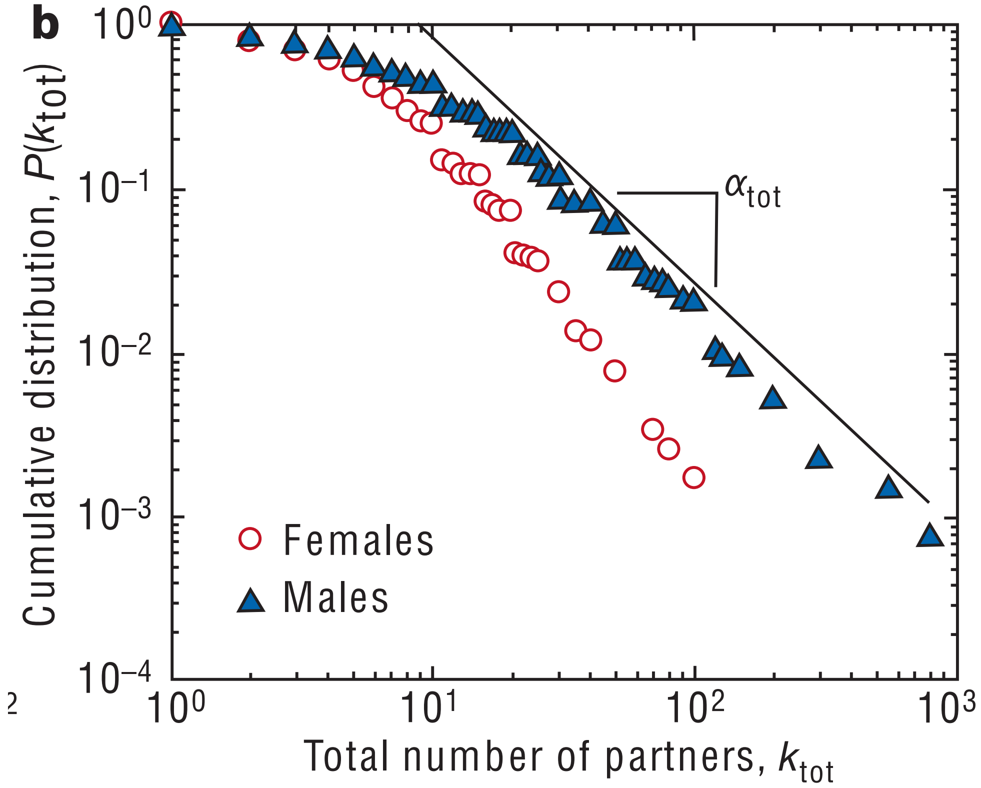
$$P(k) \sim (k+k_0)^{-\gamma} \exp\left(-\frac{k+k_0}{k_\tau}\right)$$

Scale-free networks - other examples

Sexual-interaction networks

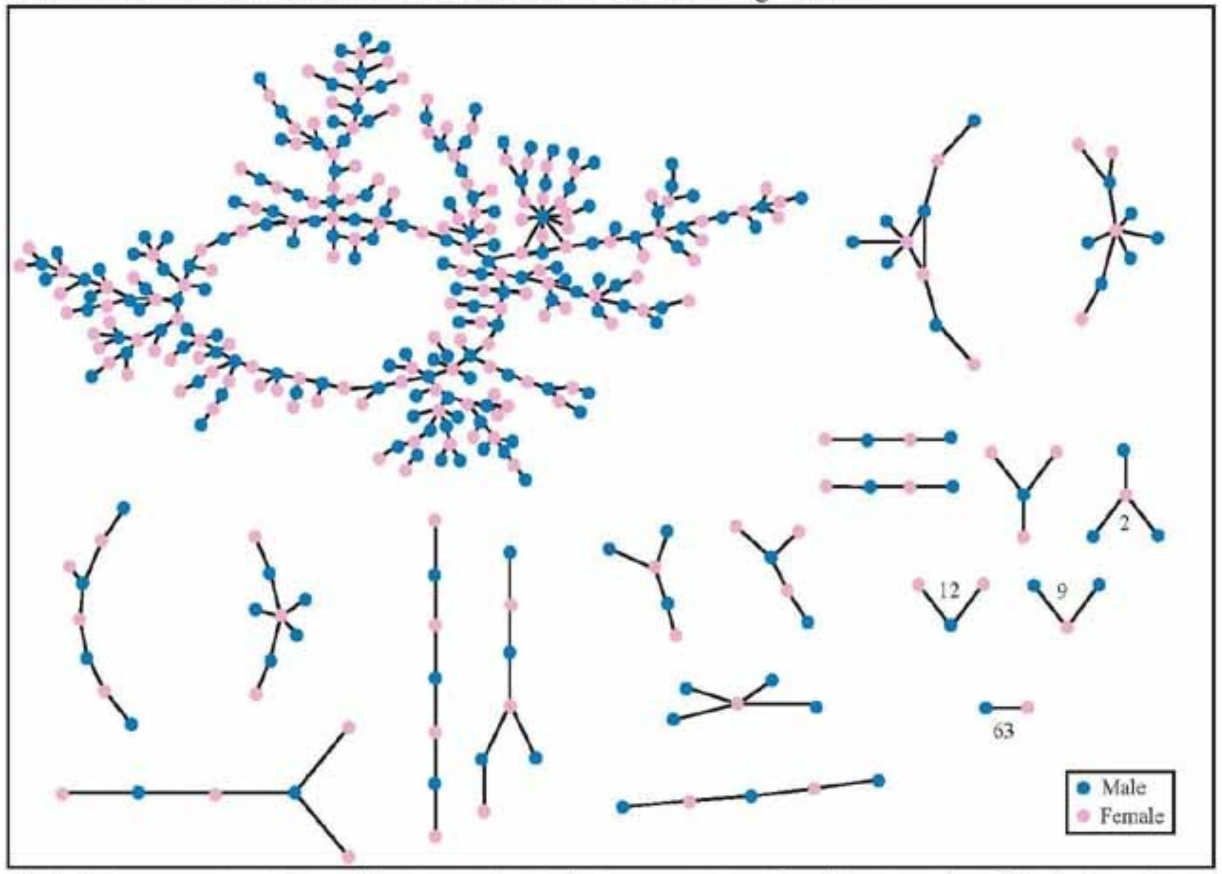
- Nodes: individuals
- Links: sexual incursion

Liljeros et.al. (2001)



Bearman et.al. (2004)

The Structure of Romantic and Sexual Relations at "Jefferson High School"

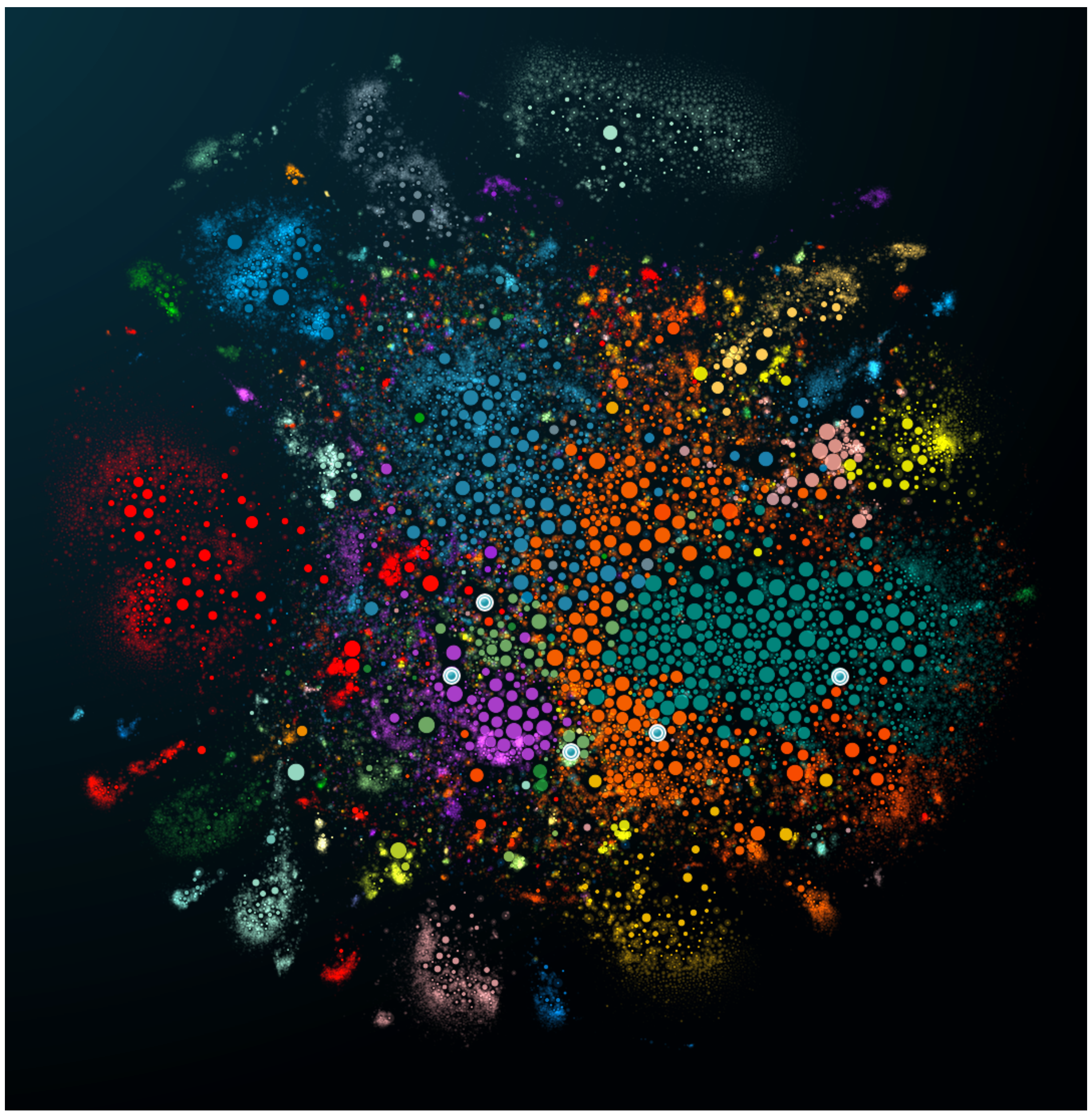


Each circle represents a student and lines connecting students represent romantic relations occuring within the 6 months preceding the interview. Numbers under the figure count the number of times that pattern was observed (i.e. we found 63 pairs unconnected to anyone else).

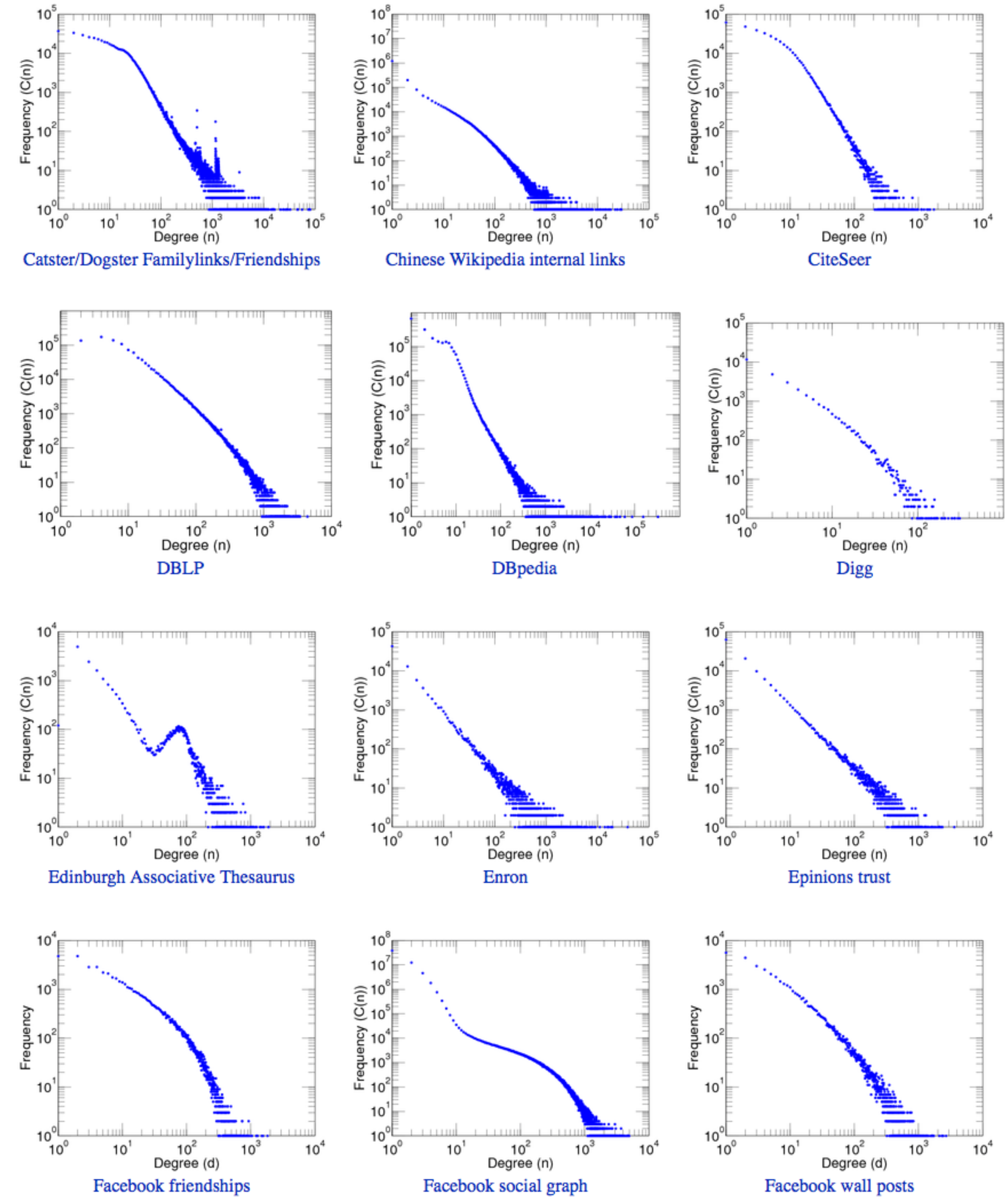
Scale-free networks - other examples

Online social networks

- Nodes: individuals
- Links: online interactions



Social network of Steam
http://85.25.226.110/mapper



Scale-free distribution - continuous formalism

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

Albert, R. et.al. Rev. Mod. Phy. (2002)

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

- Exponents of real-world networks are usually between 2 and 3
- $\Rightarrow \langle k^2 \rangle$ diverges if $N\to\infty$
- Consequently:

$$\sigma_k = (\langle k^2 \rangle - \langle k \rangle^2)^{1/2} \to \infty$$

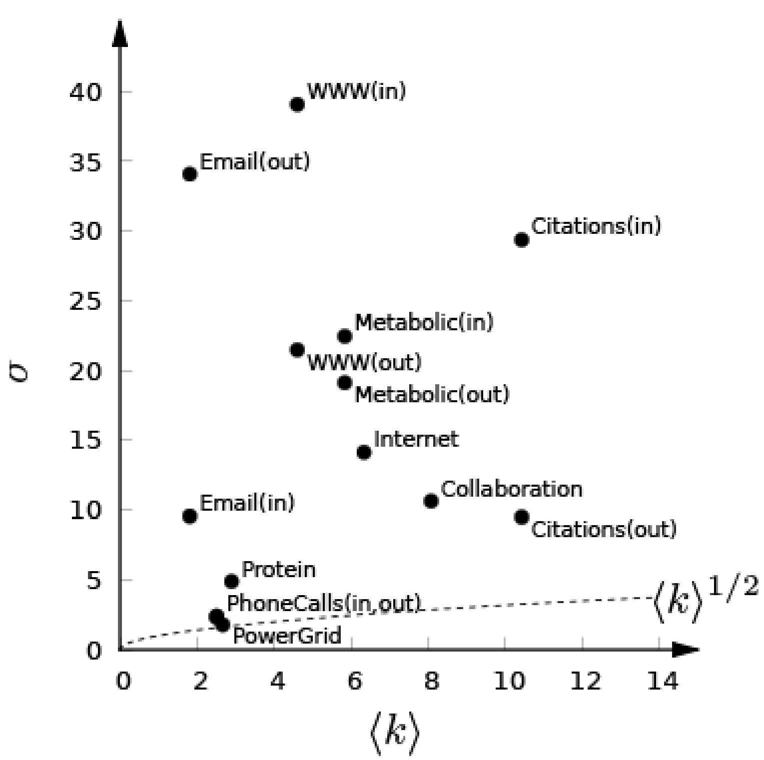
$$k = \langle k \rangle \pm \sigma_k = \langle k \rangle \pm \infty$$

- Average values are meaningless since the fluctuations are are infinitely large

Scale-free networks

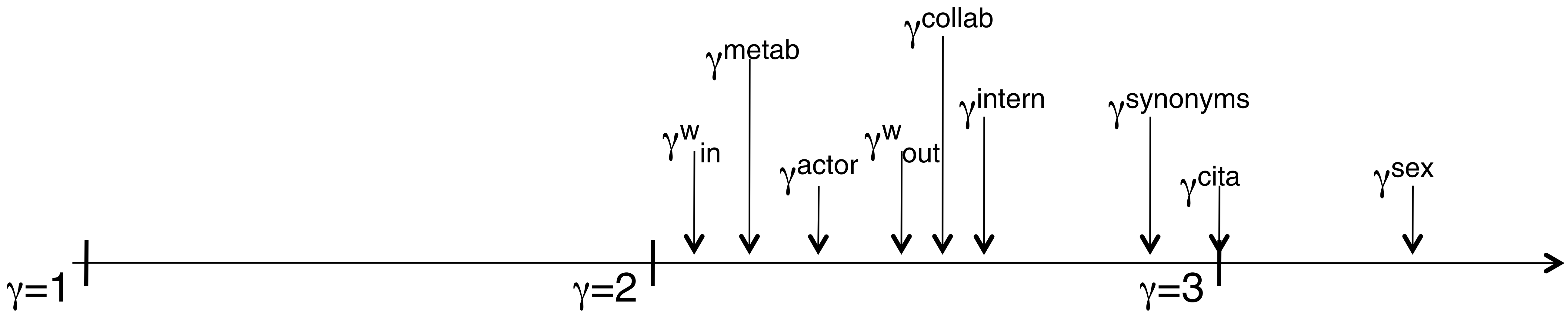
NETWORK	NL		$\langle k \rangle$ $\langle k_{in} \rangle = \langle k_{out} \rangle$	σ_{in}	σ_{out}	σ	γ_{in}	γ_{out}	γ
Internet	192,244	609,066	6.34	-	-	14.14	-	-	3.42*
WWW	325,729	1,497,134	4.60	39.05	21.48	-	2.31	2.00	-
Power Grid	4,941	6,594	2.67	-	-	1.79	-	-	Exp.
Mobile Phone Calls	36,595	91,826	2.51	2.39	2.32	-	4.69*	5.01*	-
Email	57,194	103,731	1.81	9.56	34.07	-	3.43*	2.03	-
Science Collaboration	23,133	93,439	8.08	-	-	10.63	-	-	3.35
Actor Network	702,388	29,397,908	83.71	-	-	200.86	-	-	2.12
Citation Network	449,673	4,689,479	10.43	29.37	9.49	-	3.03**	4.00	-
E. Coli Metabolism	1,039	5,802	5.58	22.46	19.12	-	2.43	2.90	-
Yeast Protein Interactions	2,018	2,930	2.90	-	-	4.88	-	-	2.89*

AL Barabási, Network Science Book (2013)



- Each network has larger degree fluctuation than equivalent random networks
- Exceptions:
    - Power-grid network - not scale-free
    - Mobile-call network - very high degree exponent

Scale-free networks - summary



$\langle k^2 \rangle$ diverges	$\langle k^2 \rangle$ finite
$\langle k \rangle$ diverges	$\langle k \rangle$ finite
Ultra small world behavior	Small world

Regime full of anomalies…	The scale-free behavior is relevant	Behaves like a random network

Why most of the real networks are in this regime?

Slide from CCNR Complex Networks Course
A. L. Barabási 2012

The Barabási-Albert model

of

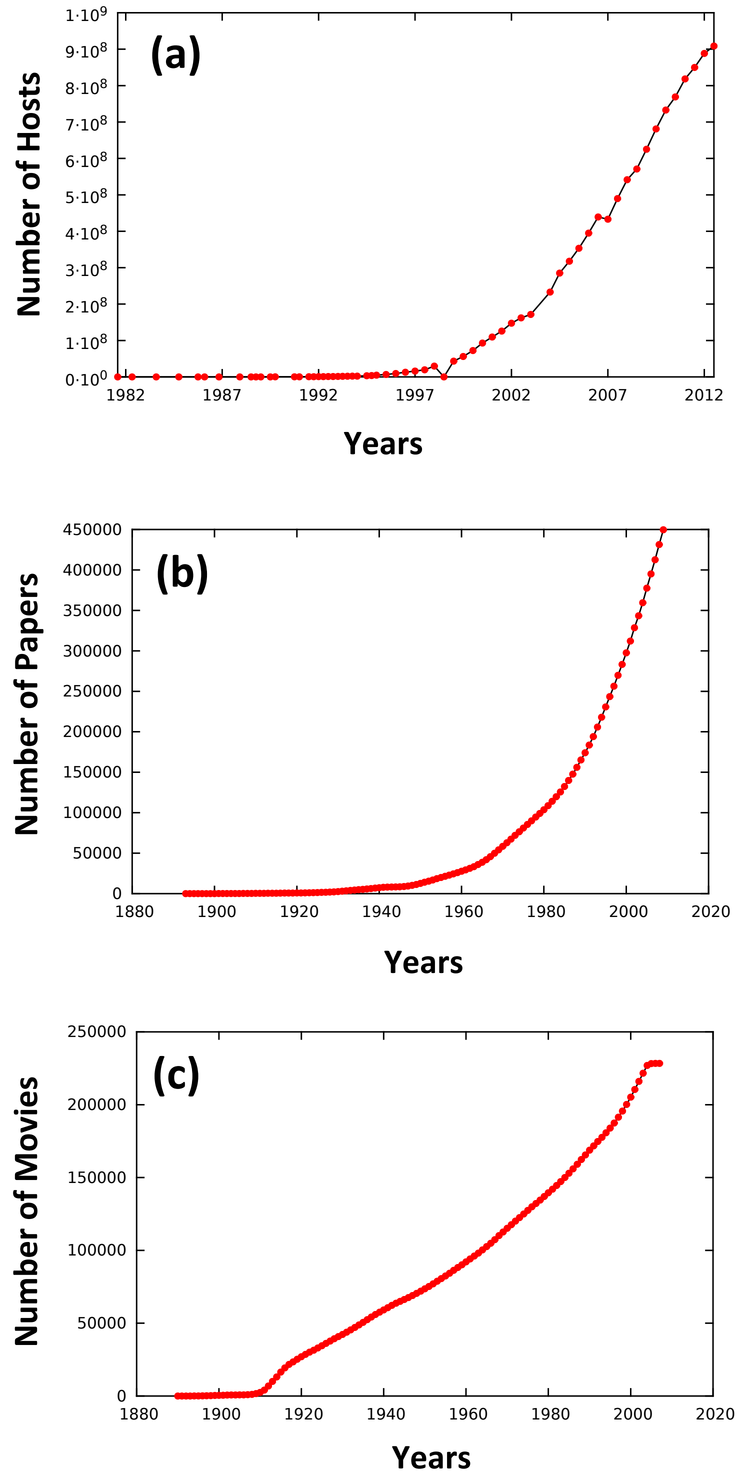
scale-free networks

Emergence of hubs

What did we miss with the earlier network models?

1. Networks are evolving
   - Networks are not static but growing in time as new nodes are entering the system

2. Preferential attachement
   - Nodes are not connected randomly but tends to link to more attractive nodes
     - Pólya urn model (1923)
     - Yule process (1925)
     - Zipf's law (1941)
     - Cumulative advantage (1968)
     - Preferential attachement (1999)
     - Pareto's law - 80/20 rule
     - The rich get richer phenomena
     - etc.



AL Barabási, Network Science Book (2013)

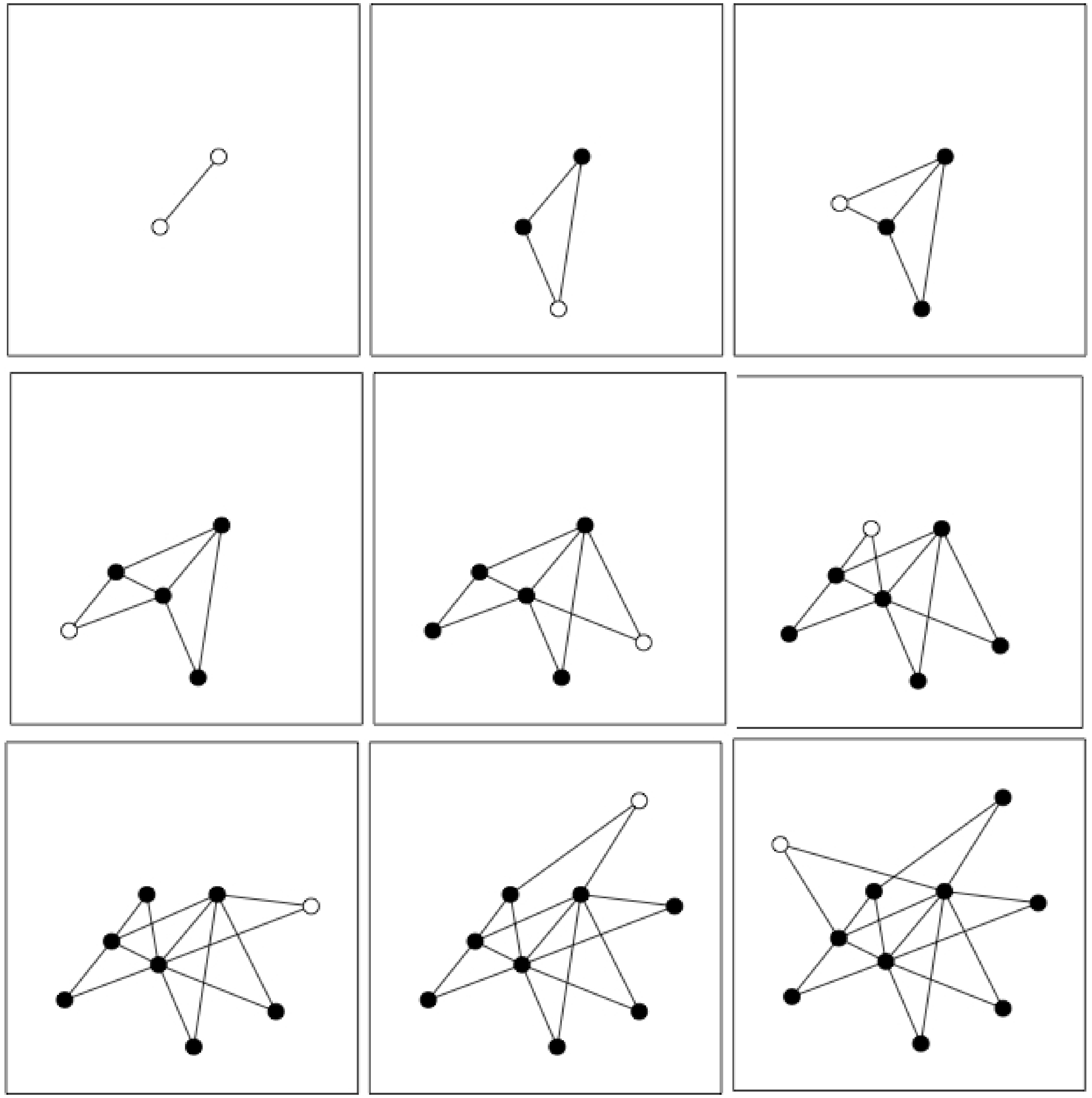
The Barabási-Albert model

1. Start with m_0 connected nodes

2. At each timestep we add a new node with m ($\leq m_0$) links that connect the new node to m nodes already in the network.

3. The probability $\pi(k)$ that one of the links of the new node connects to node i depends on the degree k_i of node i as

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

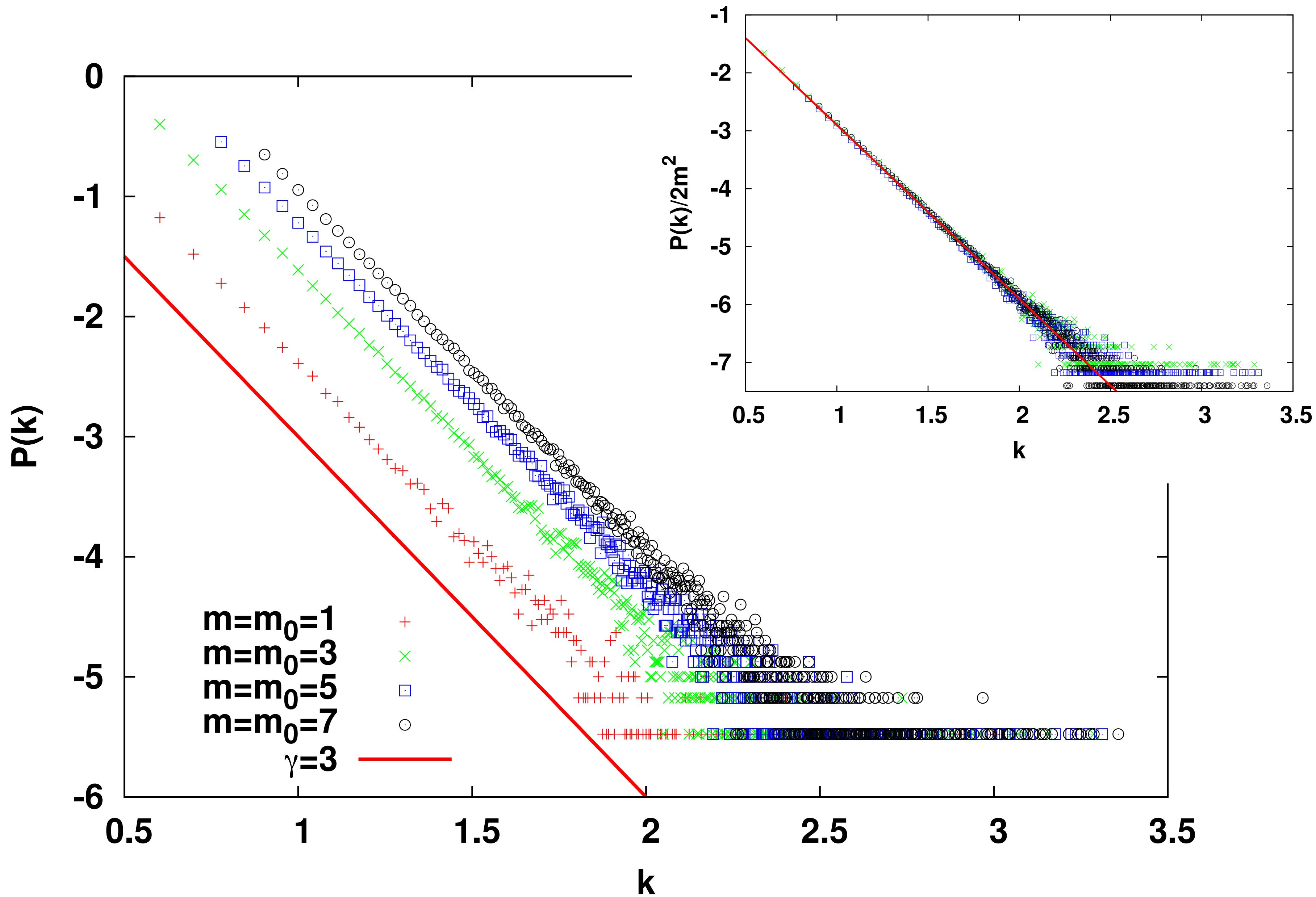


AL Barabási, Network Science Book (2013)

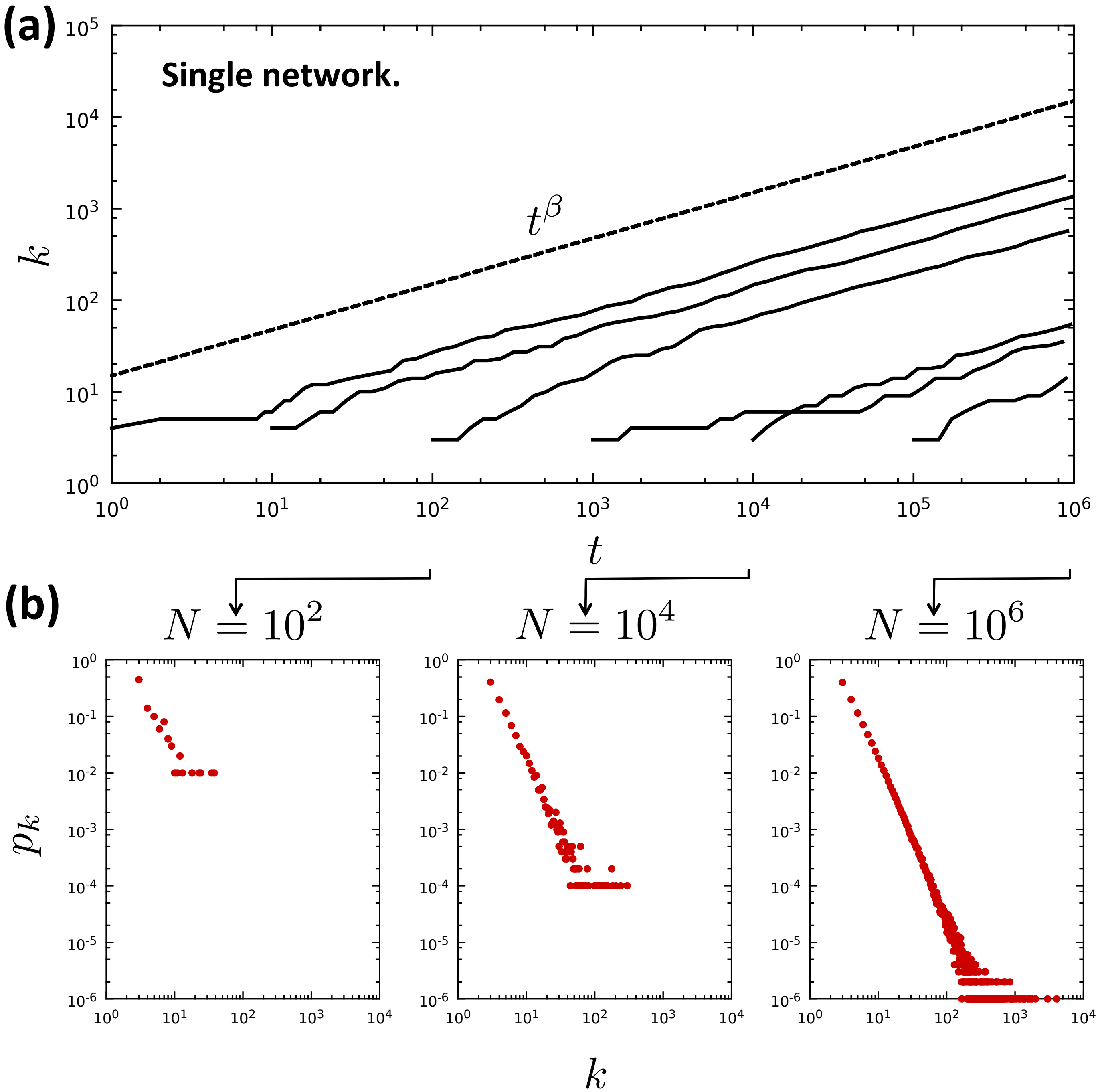
The BA model - degree distribution

- The degree exponent is independent of m
- The degree exponent is stationer in time and the degree distribution is time independent
- The exponent is very close to the exponents of real networks

$$P(k) \sim k^{-3}$$



The BA model - emergence of hubs



- The degree of each node increase as a power-law with exponent $\beta=1/2$
- The growth of the network is sub-linear
- Earlier a node was added larger its degree due to its earlier arrival and not because it grows feaster

Rich-get-richer mechanism

AL Barabási, Network Science Book (2013)

The BA model - path length

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

$\gamma = 2$: Size 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.

Ultra Small World ($2 < \gamma < 3$): The average path length increases slower than logarithmically. 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.

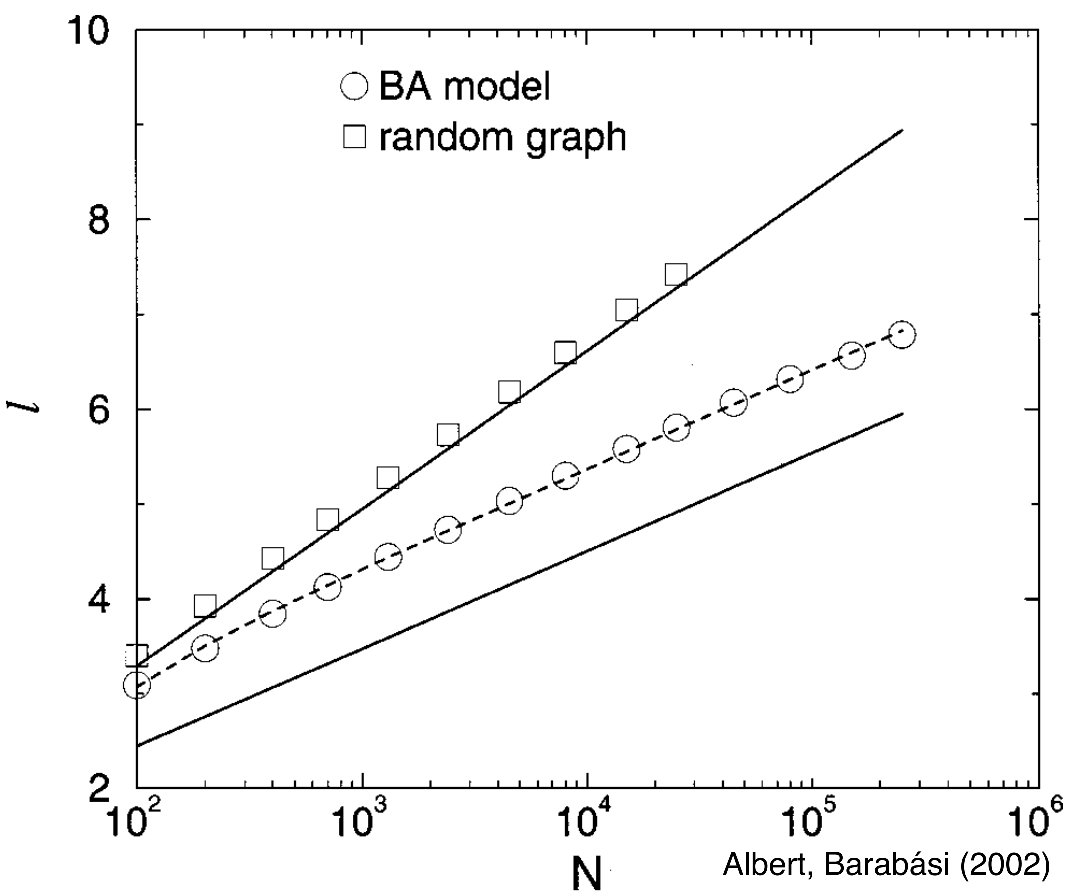
$\gamma = 3$: Some key models produce γ=3, so the result is of particular importance for them. This was first derived by Bollobas and collaborators for the network diameter in the context of a dynamical model, but it holds for the average path length as well.

Small World ($\gamma > 3$): 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.

$$\langle l \rangle = \frac{\ln N}{\ln\ln N}$$

Ultra Small World network

Bollobás, Riordan (2001)



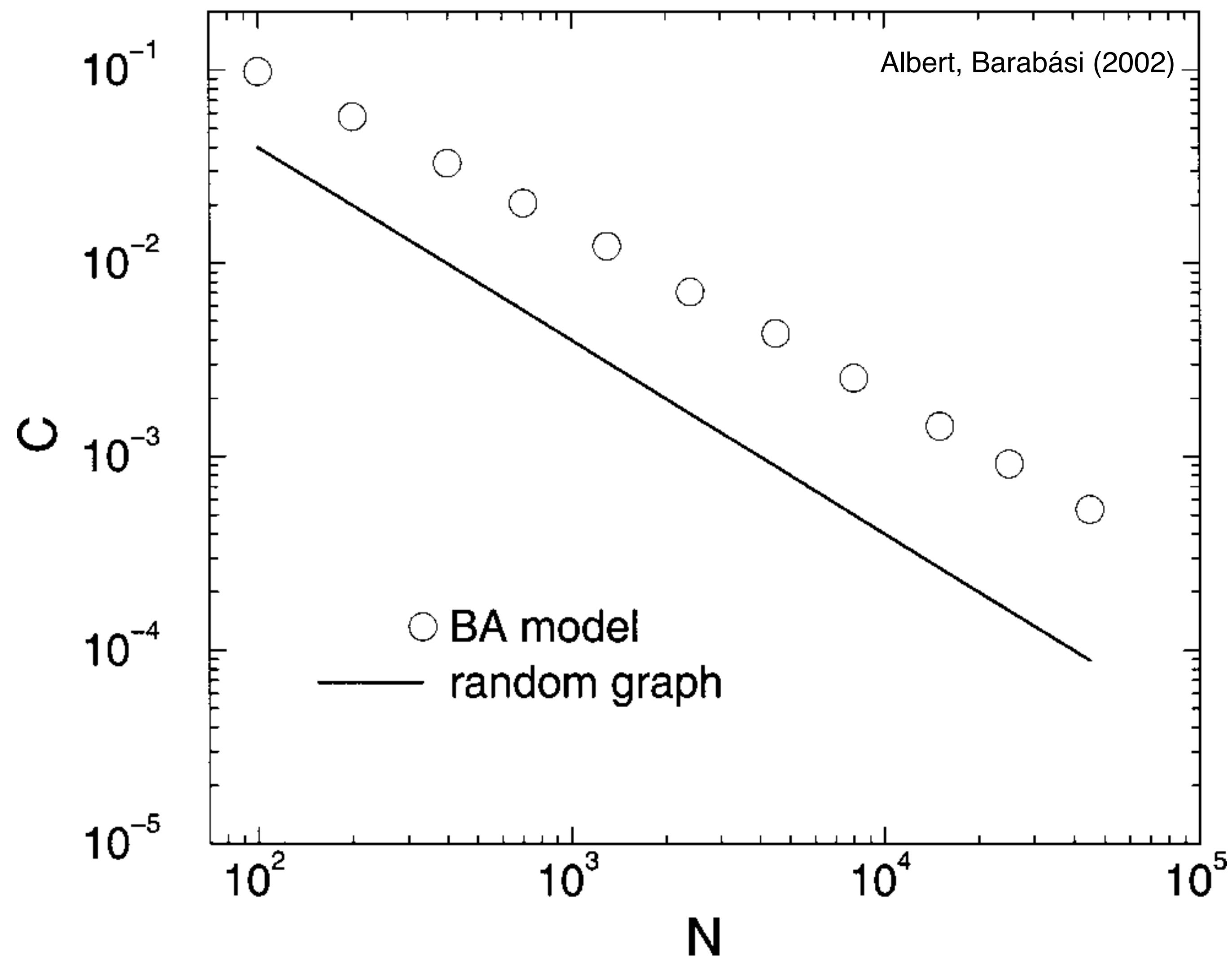
Albert, Barabási (2002)

The BA model - clustering coefficient

- The clustering coefficient decreases with the system size as

$$C = \frac{m}{4}\frac{(\ln N)^2}{N}$$

- It is still 5 times more than for random graphs



Degree correlations:

- The BA model is inducing non-trivial degree correlations due to its definition

$$n_{kl} \simeq k^{-2} l^{-2}$$

ER Random Network - catch up

Network	Degree distribution	Path length	Clustering coefficient
Real world networks	broad	short	large
ER random networks	Poissonian	short	small
WS small-world networks	exponential	short	large
BA scale-free networks	power-law	short	N dependent

Configuration model

Random graphs with specified degrees

Problem

- The ER Random Graph model has a Poisson degree distribution
- Most real-world networks have heavy-tailed degree distributions
- **We need to generate networks, which have pre-determined degrees or degree distribution, but they are maximally random otherwise**

Why?

- They can be used as **null models** to decide whether some empirically observed pattern is interesting

Null models for networks

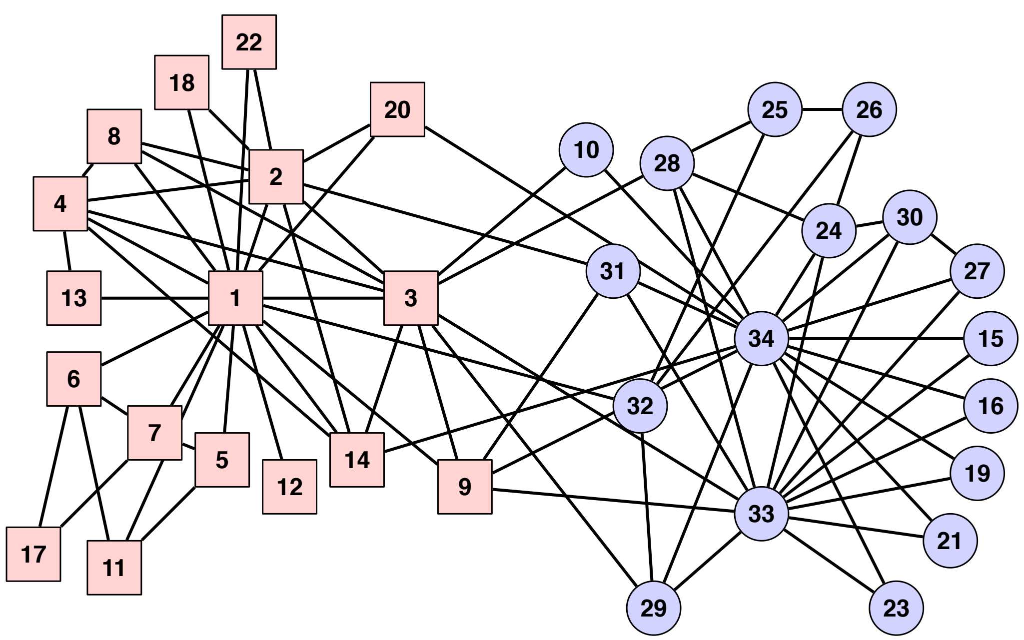
- "*…in the study of statistical properties of graphs, the* ***null model*** *is a graph which matches one specific graph in some of its structural features, but which is otherwise taken to be an instance of a random graph. The null model is used as a term of comparison, to verify whether the graph in question displays some feature, such as community structure, or not.*" (Wikipedia)

- Null models of a graph are sampled from an ensemble of graphs exhibiting the same structural constraints but maximally random otherwise

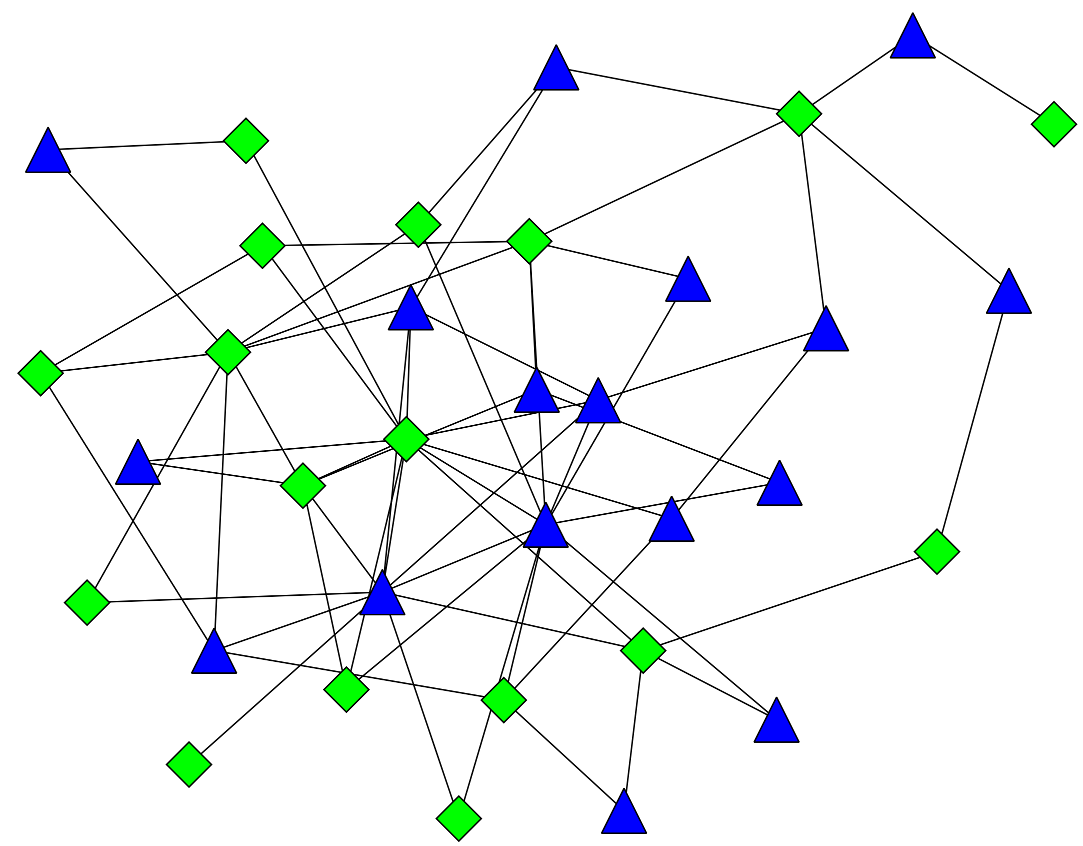
Random graphs with specified degrees

Configuration model

- **The degree sequence $\vec{k} = \{k_i\}$ can contain fixed degrees coming from an empirical network or sampled from a distribution**
  - provides a graph maximally random otherwise, e.g. with no community structure of degree correlations…



karate club



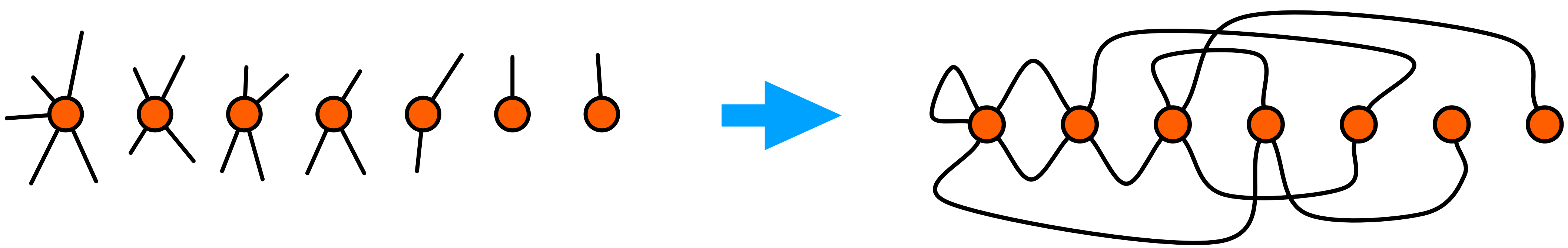
configuration model

Random graphs with specified degrees

Molloy-Reed model for configuration networks

Original idea:

1. Given a degree sequence $\vec{k} = \{k_1, k_2, \ldots, k_n\}$
2. Assign each node $i \in V$ with k_i number of stubs
3. Select random pairs of unmatched stubs and connect them
4. Repeat 3 until there are unmatched stubs



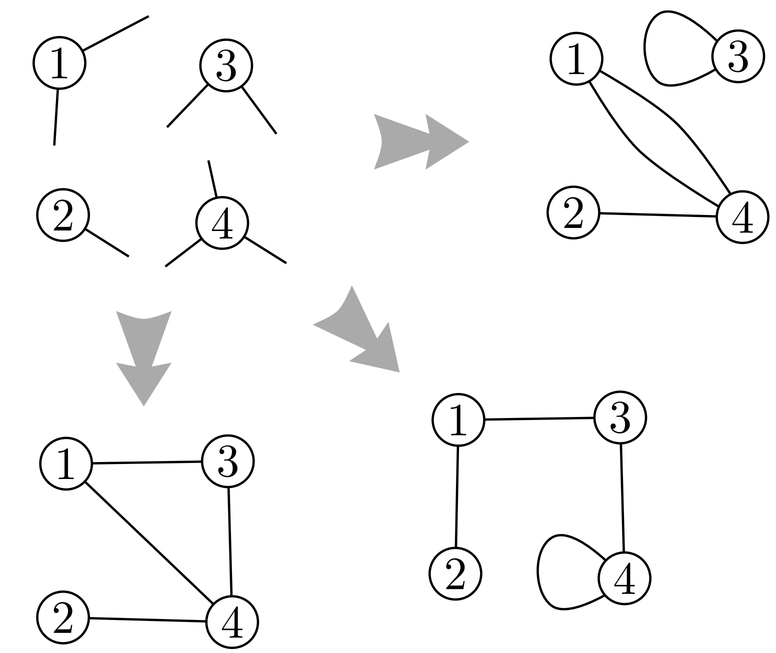
- This process will produce a configuration model with exact degree sequence
- Possible to select multiple times stubs of the same pair of nodes ➜ **Multilinks**
- Possible to select the stubs of the same node to connect ➜ **Self-links**

The obtained graph is not simple…but the density of multi and self-links ➜ 0 as N ➜ ∞

Configuration model - for real networks

Recipe 1:

- Give each vertex k_i “stubs”
- Pick two stubs uniformly at random and create an edge by joining them
- Repeat until no stubs remain
- Control of multiple and self edges is possible



Recipe 2:

- Start with any network structure with degree sequence $[k_1, k_2, \ldots, k_n]$
- Select two edges randomly and exchange their end points
- Repeat ad infinitum