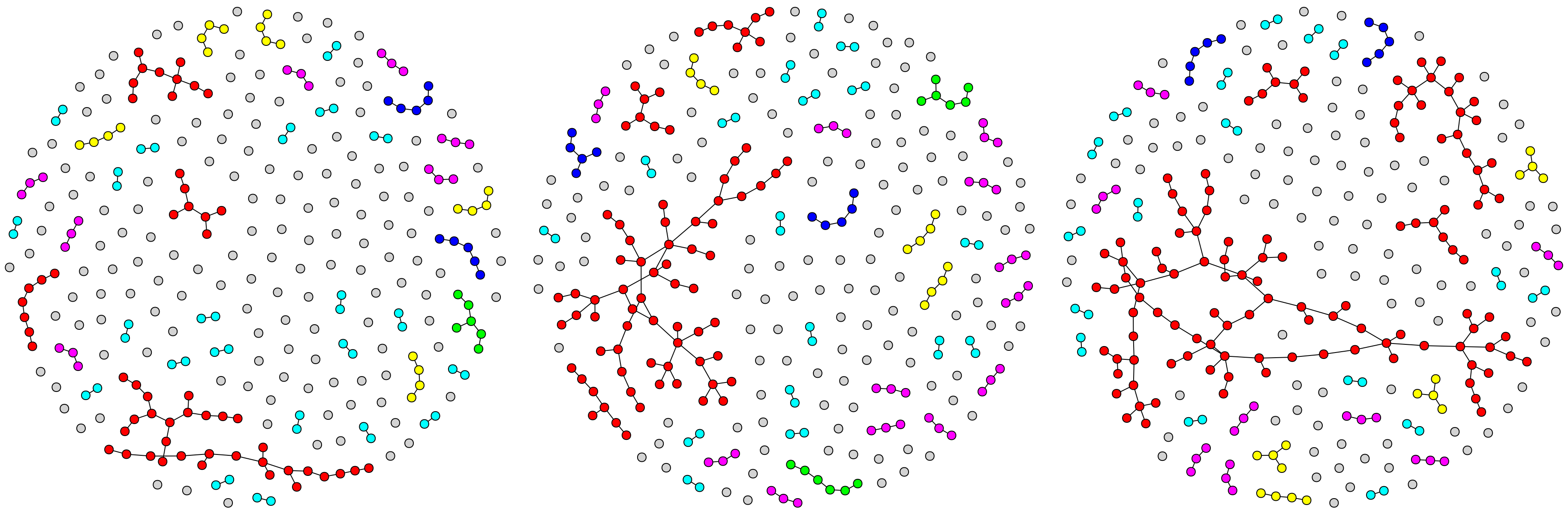
Complex Networks

(CR 15)

Fundamental network models

Class 7



Dr. Márton Karsai

ENS Lyon - 2016

Practical matters

Tutorials (TD):

- Samuel Unicomb (samuel.unicomb@inria.fr)

Course web page:

- perso.ens-lyon.fr/marton.karsai/Marton_Karsai/complexnet.html

Slides:

- http://perso.ens-lyon.fr/marton.karsai/protected/complexnets/
- login: complexnet
- psw: cnet123

Lectures (upcoming):

- 03/11/2016 - 10:15-12:15 - Amphi B
- 10/11/2016 - 10:15-12:15 - B1
- 17/11/2016 - No lecture (but TD)
- 24/11/2016 - 10:15-12:15 - Amphi B
- 01/12/2016 - 10:15-12:15 - Amphi B
- 08/12/2016 - No lecture (research schools)
- 15/12/2016 - 10:15-12:15 - Amphi B
- 05/01/2017 - 10:15-12:15 - Amphi B

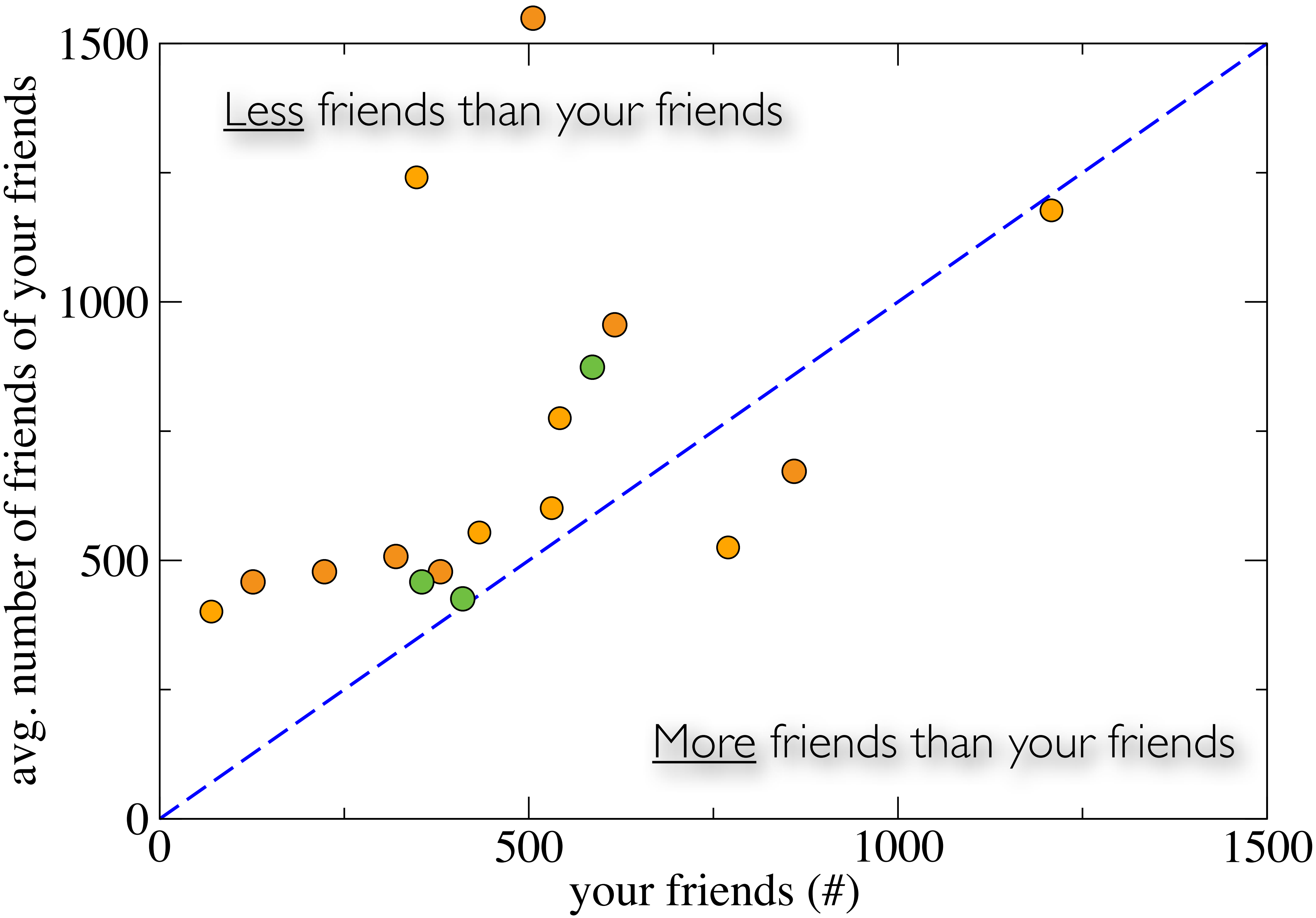
Tutorials (upcoming):

- 08/11/2016 - 15:45-17:45 - Salle 119
- 17/11/2016 - 10:15-12:15 - Amphi B

Exam:

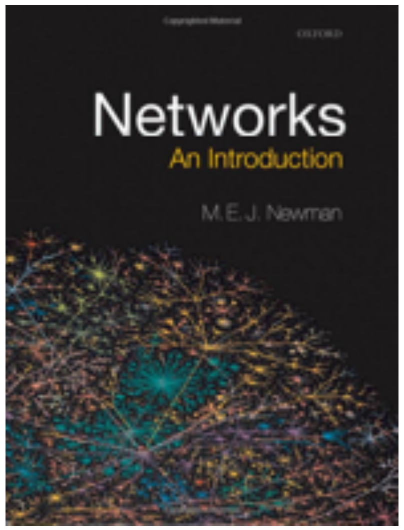
- 12/01/2017 - 10:15-12:15 - Amphi B (TBC)

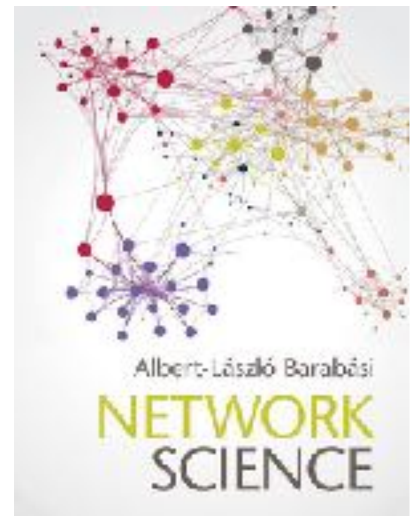
Degree-degree correlations in Facebook

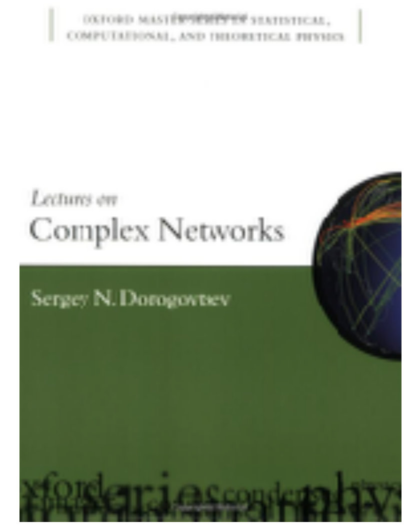


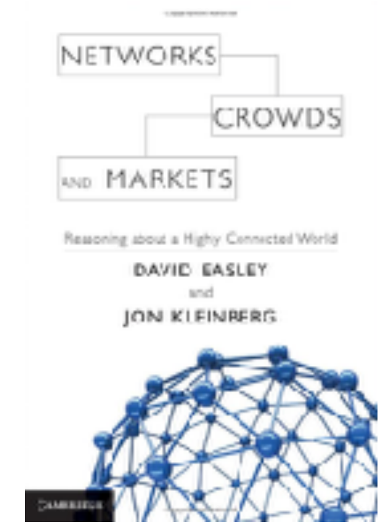
Materials

Lecture books









available free online

Reviews

SIAM REVIEW
Vol. 45, No. 2, pp. 167–256

© 2003 Society for Industrial and Applied Mathematics

The Structure and Function of Complex Networks*

M. E. J. Newman†

REVIEWS OF MODERN PHYSICS, VOLUME 74, JANUARY 2002

Statistical mechanics of complex networks

Réka Albert* and Albert-László Barabási

Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556

Characterization and Modeling of weighted networks

Marc Barthélemy[1], Alain Barrat[2], Romualdo Pastor-Satorras[3], and Alessandro Vespignani[2]

Pop science book

Outline

- Introduction and general network characteristics
- ER model (percolation) and WS models (recap)
- Scale-free networks, Barabási-Albert model
- Motifs and communities
- Temporal networks
- Multiplex and inter-dependent networks
- Spreading processes on networks
- Network sampling and visualisation

Slides are available:
http://perso.ens-lyon.fr/marton.karsai/protected/
login: complexnet
password: cnet123

Fundamental network models

Random Graphs

Central quantities in network analysis

- Degree distribution: $P(k)$
- Clustering coefficient: C
- Average path length: l

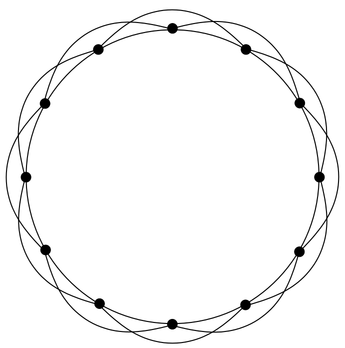
Real networks

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

Regular lattices

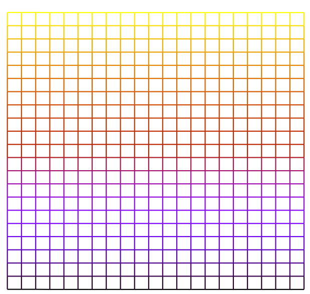
- Graphs where each node has the same degree k
- Translational symmetry in n directions

1D - Rings

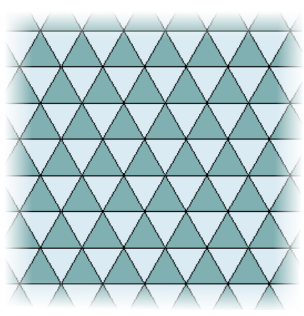


k=4

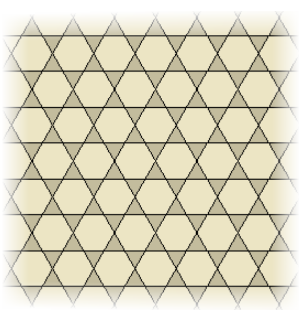
2D lattices



k=4

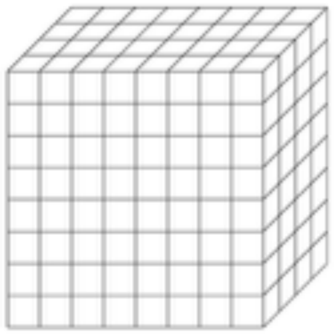


k=6



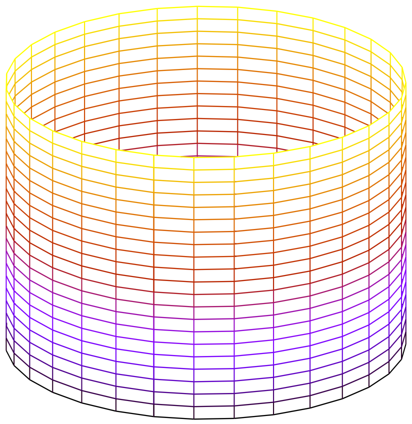
k=4

3D lattices

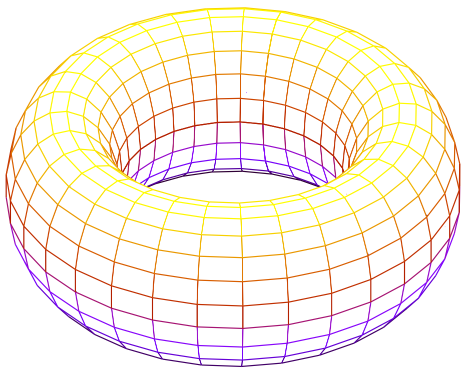


k=6

- Periodic boundary conditions



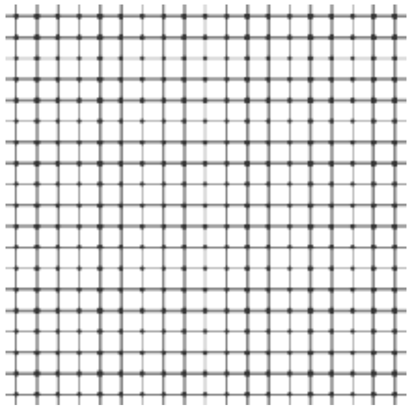
half-periodic



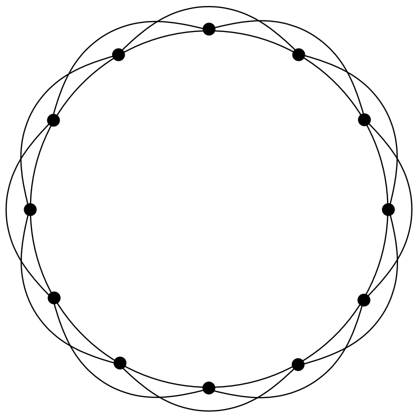
full-periodic

Regular lattices

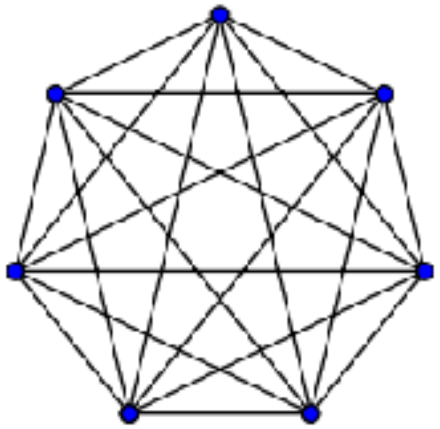
Clustering coefficient



C=0



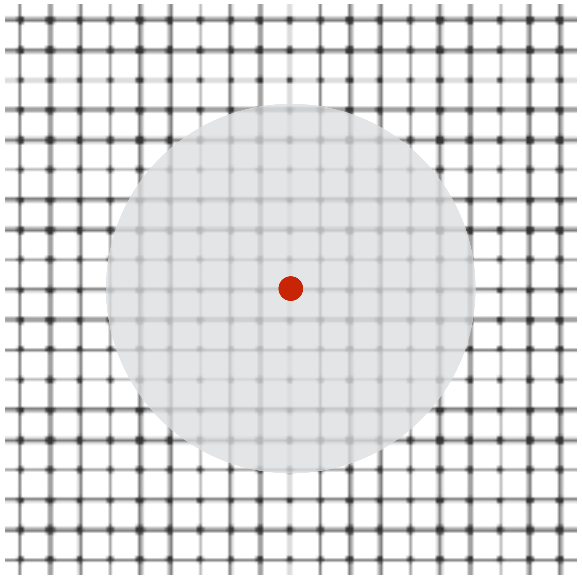
C=3/4



C=1

- Clustering coefficient depends on the structure (can be large)
- It is constant for each node

Path length



$$l_{max} \sim \langle l \rangle \sim N^{1/D}$$

Average path length is large if N>>1

- *D - dimension of the lattice*
- *d - distance from central node*

Regular lattices

Network	Degree distribution	Path length	Clustering coefficient
Real world networks	broad	short	large
Regular lattices	constant	long	can be large

ER Random Networks

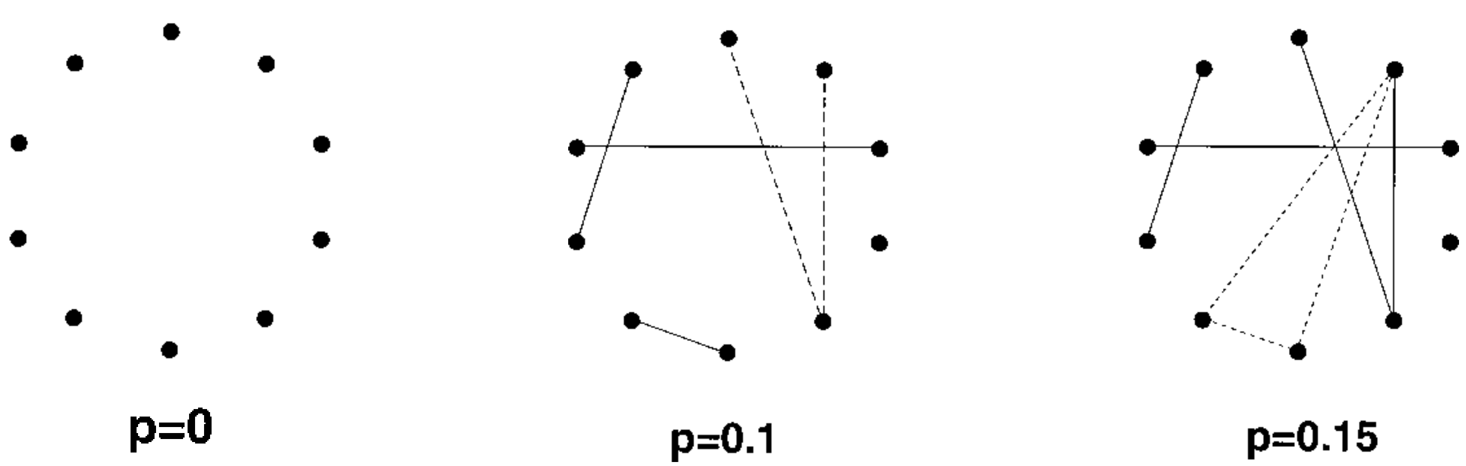
Erdős-Rényi model: simple way to generate random graphs

- 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 m links

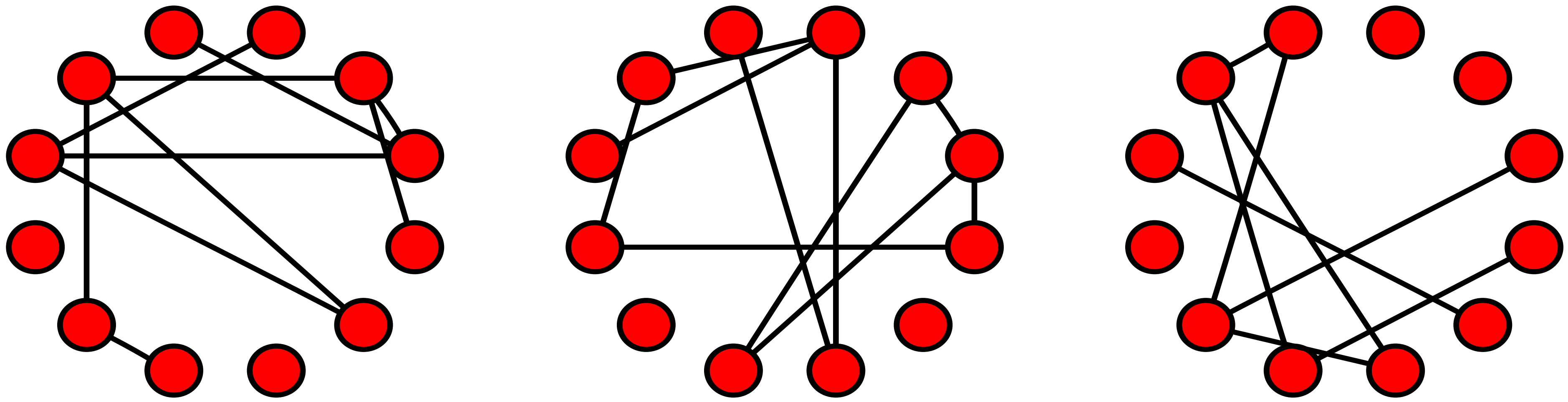
- The $G(N,p)$ definition
  1. Take N disconnected nodes
  2. Add an edge between any of the nodes independently with probability p

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

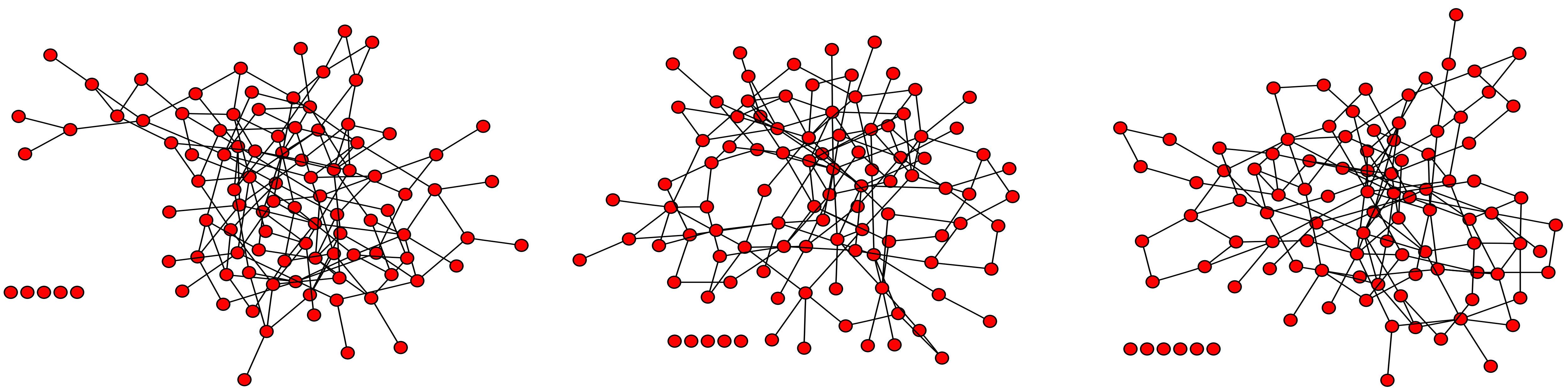


ER Random Networks

p=1/6
N=12

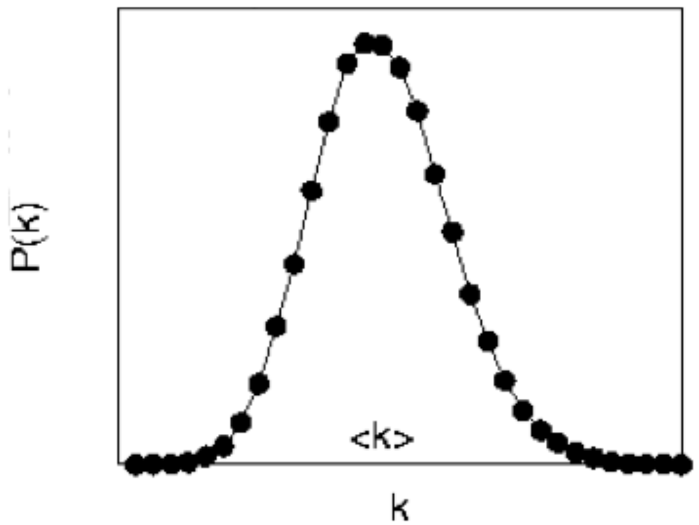


p=0.03
N=100



Slide from CCNR course, A. L. Barabasi (2012)

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

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

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

$$p = \frac{<k>}{(N-1)}$$

For large ***N*** and small ***k***, we can use the following approximations:

$$\binom{N-1}{k} = \frac{(N-1)!}{k!(N-1-k)!} = \frac{(N-1)(N-1-1)(N-1-2)\ldots(N-1-k+1)(N-1-k)!}{k!(N-1-k)!} = \frac{(N-1)^k}{k!}$$

$$\ln[(1-p)^{(N-1)-k}] = (N-1-k)\ln\left(1-\frac{<k>}{N-1}\right) = -(N-1-k)\frac{<k>}{N-1} = -<k>\left(1-\frac{k}{N-1}\right) \cong -<k>$$

ln(1-x) series expansion

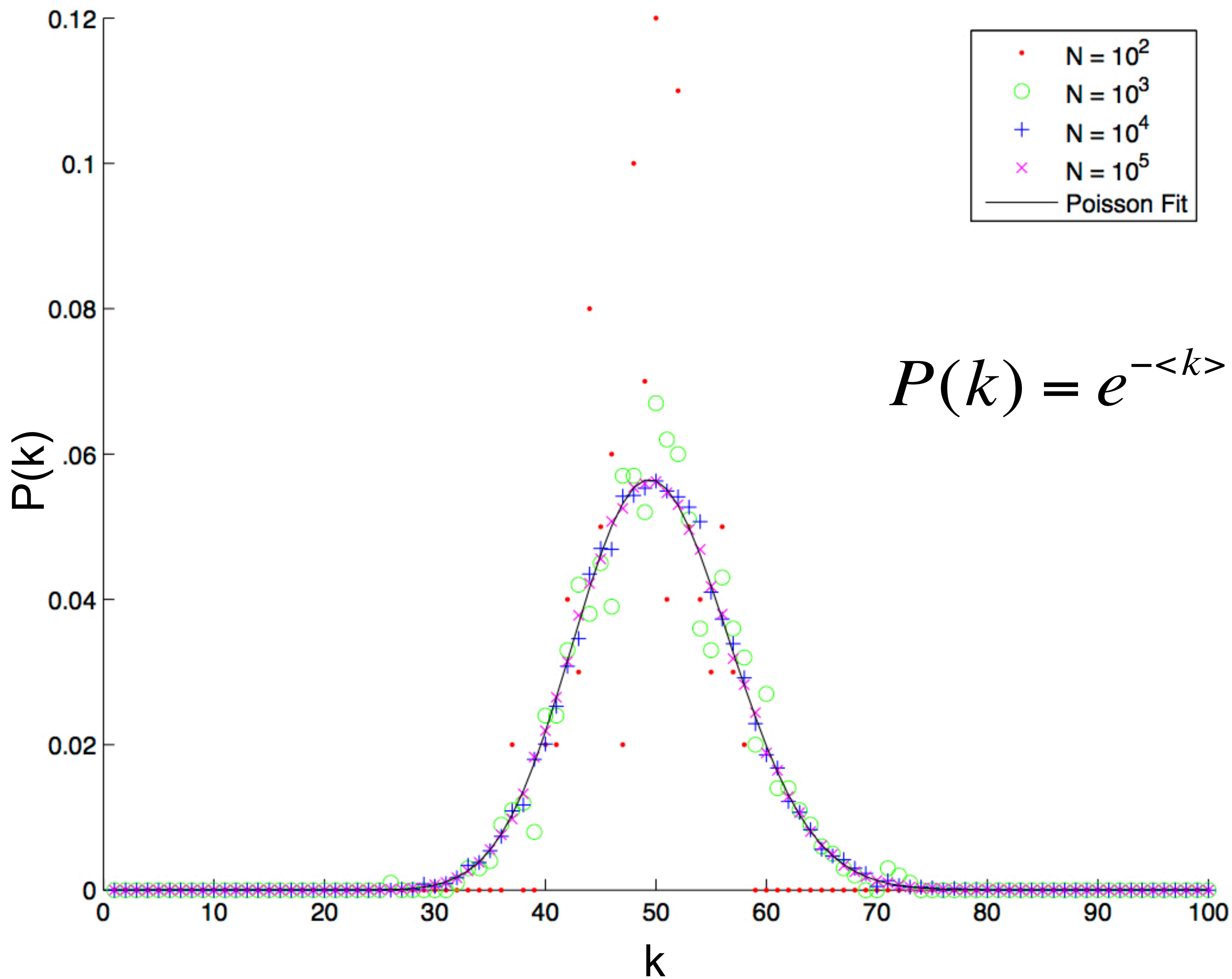
$$(1-p)^{(N-1)-k} = e^{-<k>}$$

$$\ln(1+x) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} x^n = x - \frac{x^2}{2} + \frac{x^3}{3} - \ldots \quad \text{for} \quad |x| \le 1$$

$$P(k) = \binom{N-1}{k} p^k (1-p)^{(N-1)-k} = \frac{(N-1)^k}{k!} p^k e^{-<k>} = \frac{(N-1)^k}{k!} \left(\frac{<k>}{N-1}\right)^k e^{-<k>} = e^{-<k>} \frac{<k>^k}{k!}$$

Poisson distribution

Degree distribution - Random Graphs



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

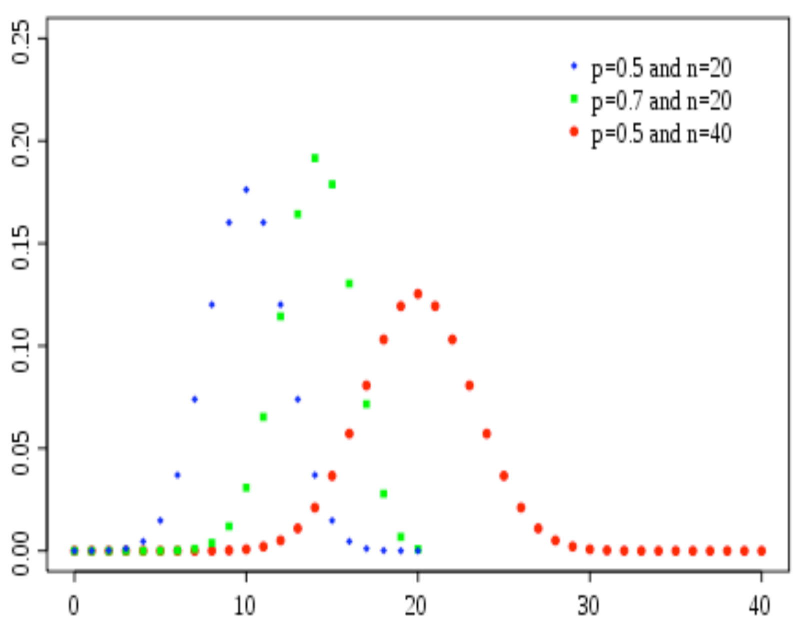
Degree distribution - Random Graphs

Probability Distribution Function (PDF)

Exact Result

-binomial distribution-

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



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

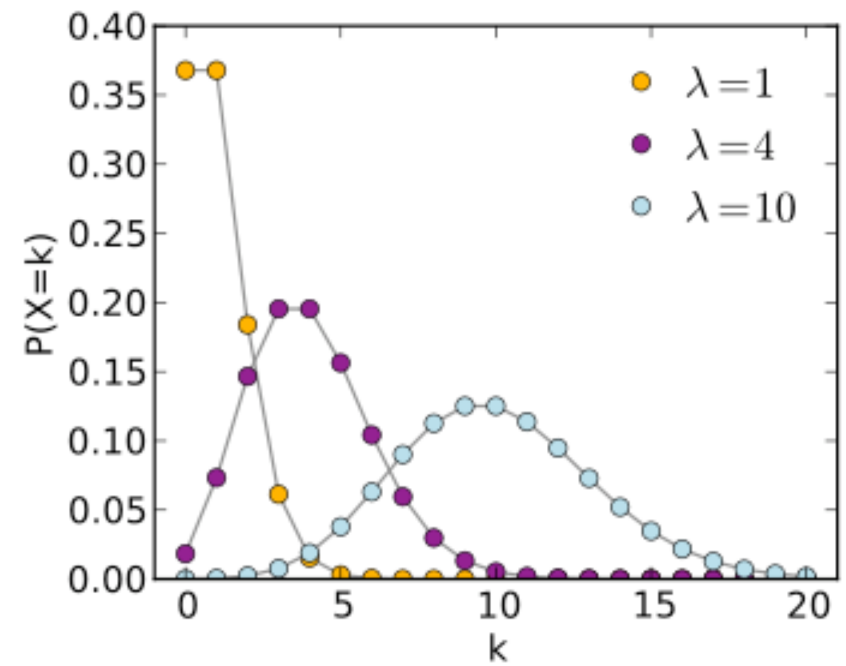
$$<k^2> = p(1-p)(N-1) + p^2(N-1)^2$$

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

Large N limit

-Poisson distribution-

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



$$<k> = <k>$$

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

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

Degree distribution - Random Graphs

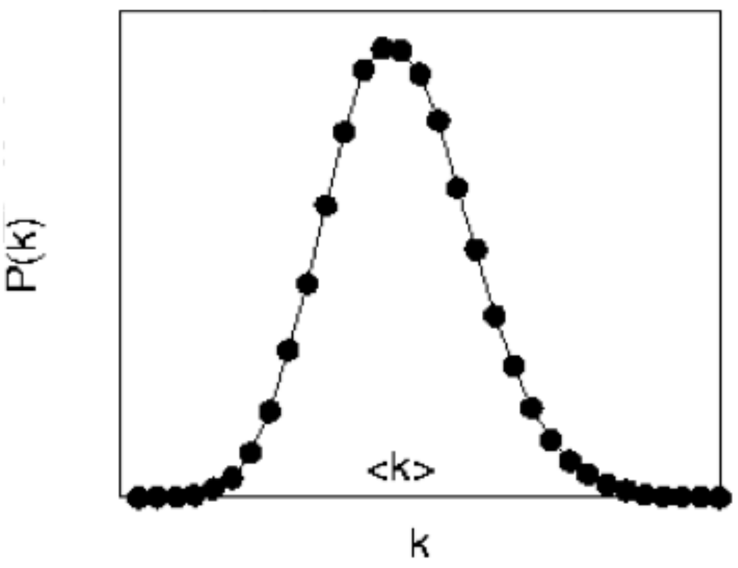
What does it mean? Continuum formalism:

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

If we consider a network with average degree <k> then the probability to have a node whose degree exceeds a degree k_0 is:

$$P(k > k_0) = \int_{k_0}^{\infty} e^{-<k>} \frac{<k>^k}{k!} dk$$

For example, with <k>=10,
•the probability to find a node whose degree is at least twice the average degree is 0.00158826.
•the probability to find a node whose degree is at least ten times the average degree is $1.79967152 \times 10^{-13}$
•the probability to find a node whose degree is less than a tenth of the average degree is 0.00049
See http://www.stud.feec.vutbr.cz/~xvapen02/vypocty/po.php



What does it mean? Discrete formalism:

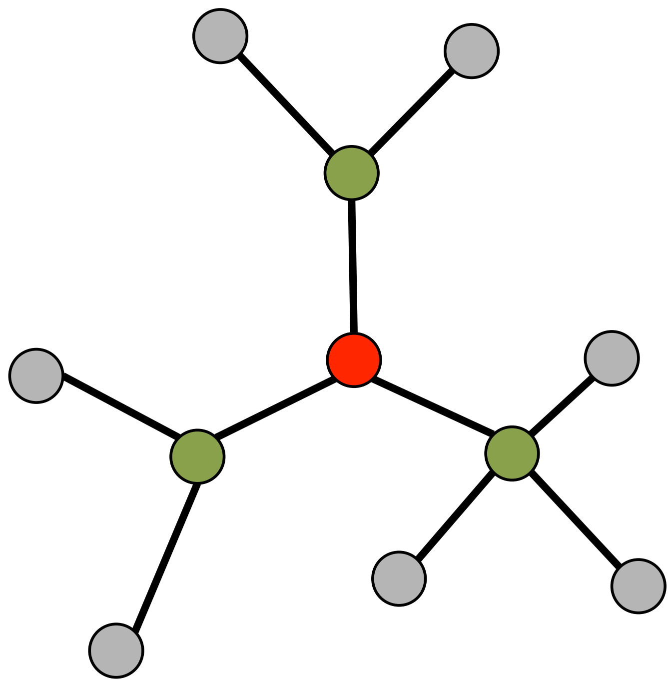
$$P(k) = \binom{N-1}{k} p^k (1-p)^{(N-1)-k}$$

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

•The probability of seeing a node with very high of very low degree is exponentially small.
•Most nodes have comparable degrees.
•The larger the size of a random network, the more similar are the node degrees

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:

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

geometric series

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 $\quad p = \frac{<k>}{N\text{-}1}$

if N>>1 and <k>≈1

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

- 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

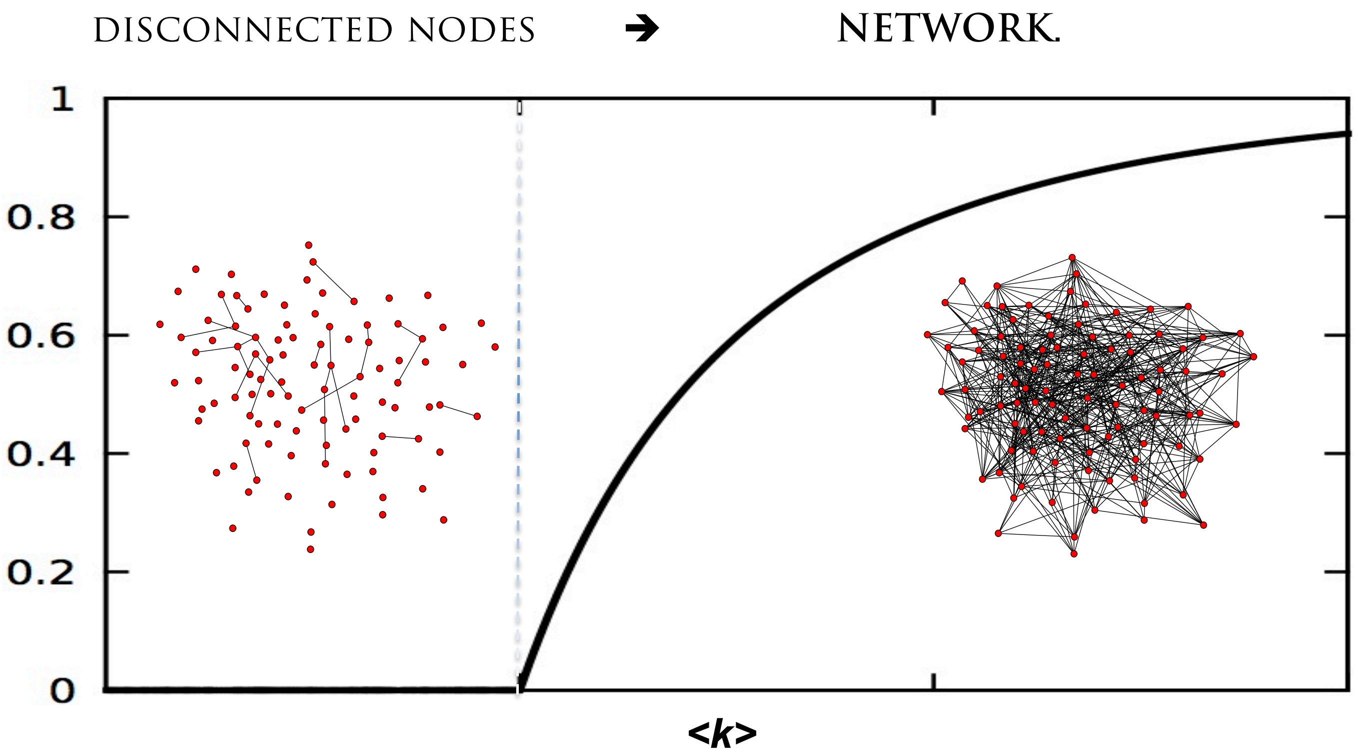
- **Small clustering coefficient** $C_i \equiv \frac{1}{N} < k >= p$

Real-world networks

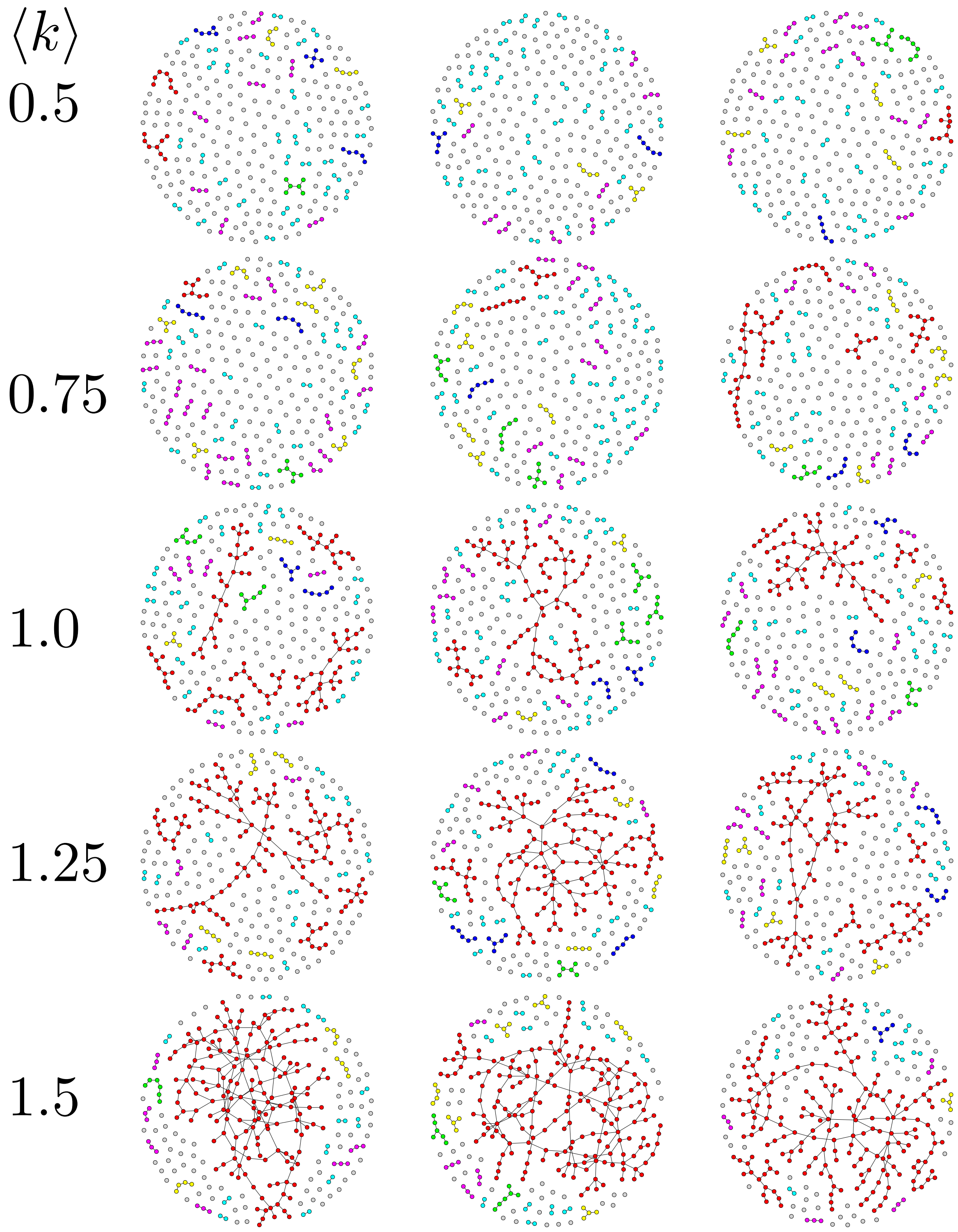
Network	Size	$\langle k \rangle$	ℓ	ℓ_{rand}	C	C_{rand}	Reference
WWW, site level, undir.	153 127	35.21	3.1	3.35	0.1078	0.00023	Adamic, 1999
Internet, domain level	3015–6209	3.52–4.11	3.7–3.76	6.36–6.18	0.18–0.3	0.001	Yook *et al.*, 2001a, Pastor-Satorras *et al.*, 2001
Movie actors	225 226	61	3.65	2.99	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
- Question: How and when does this transition happen



Evolution of Random Graphs

Let us denote with $u=1-N_g/N$, i.e. the fraction of nodes that are NOT part of the giant component (GC) N_g .

For a node ***i*** to be part of the GC, it needs to connect to it via another node ***j***. If ***i*** is NOT part of the GC, that could happen for two reasons:

Case A: node ***i*** does not connect to node ***j***, Probability: $1-p$

Case B: node ***i*** connects to ***j***, but ***j*** is not connected to the GC: Probability: pu

Total probability that ***i*** is not part of the GC via node ***j*** is: $1-p+pu$

The probability that ***i*** is not linked to the GC via ***any other node*** is $(1-p+pu)^{N-1}$

Hence:

$$u = (1 - p + pu)^{N-1}$$

For any ***p*** and ***N*** this equation provides the size of the giant component as $N_{GC}=N(1-u)$

Evolution of Random Graphs

$$u = (1 - p + pu)^{N-1}$$

Using p=<k>/(N-1) and taking the log of both sides and using ***<k><<N*** we obtain:

$$\ln u = (N-1)\ln\left[1 - \frac{<k>}{N-1}(1-u)\right] \approx -(N-1)\frac{<k>}{N-1}(1-u) = -<k>(1-u)$$

Taking an exponential of both sides we obtain

$$u = e^{-<k>(1-u)}$$

Or, if we denote with S the fraction of nodes in the giant component, $S=N_{GC}/N$, i.e. S=1-u

$$S = 1 - e^{-<k>S}$$

Erdos and Renyi, 1959

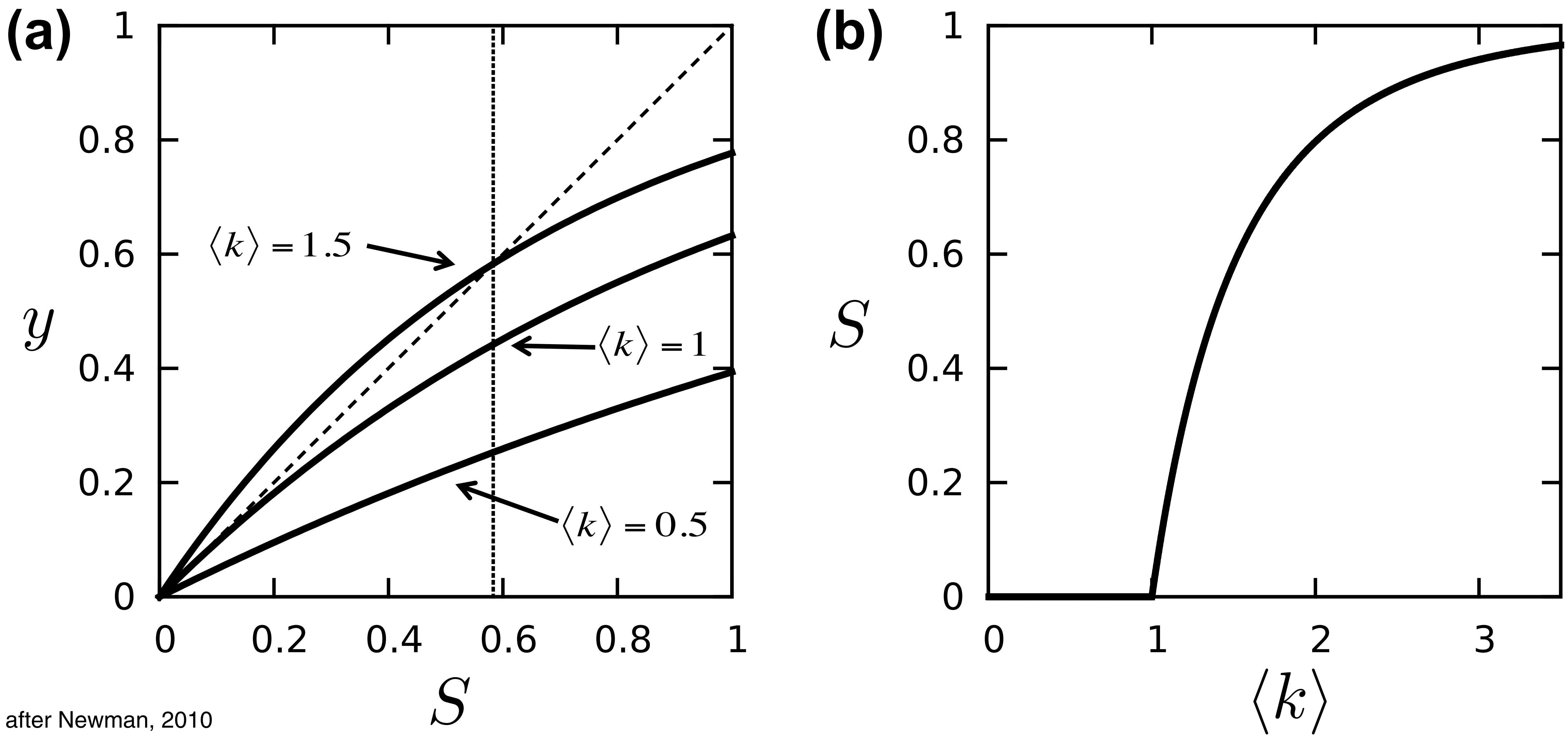
Evolution of Random Graphs

$$S = 1 - e^{-<k>S}$$

S: the fraction of nodes in the giant component, $S=N_g/N$

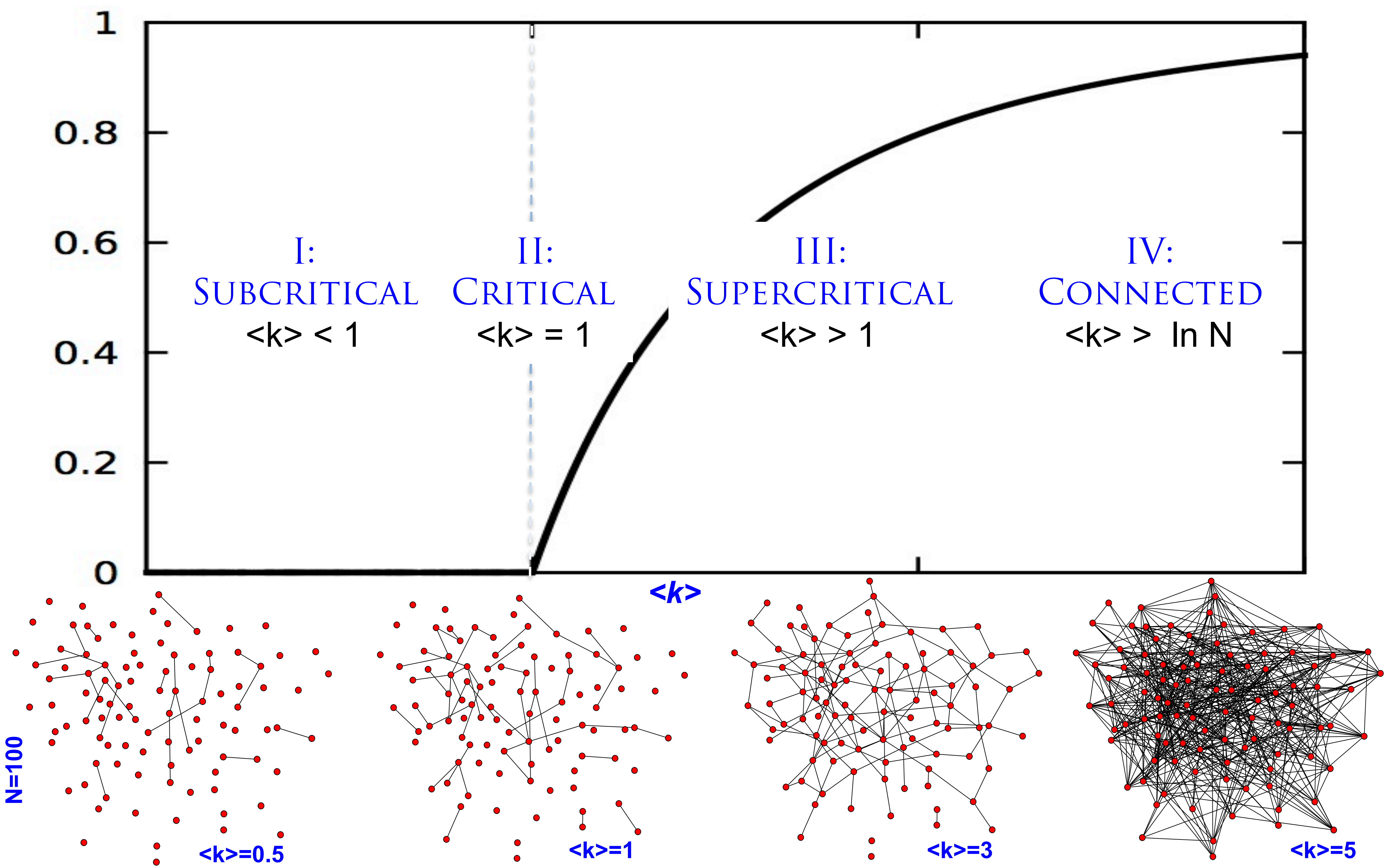
Phase transition point: $\frac{d}{dS}(1-e^{-<k>S}) = 1 \qquad <k> e^{-<k>S} = 1$

Set S=0, we obtain a phase transition at **<k>=1**



after Newman, 2010

ER Random Network - catch up



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
Regular lattices	constant	long	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

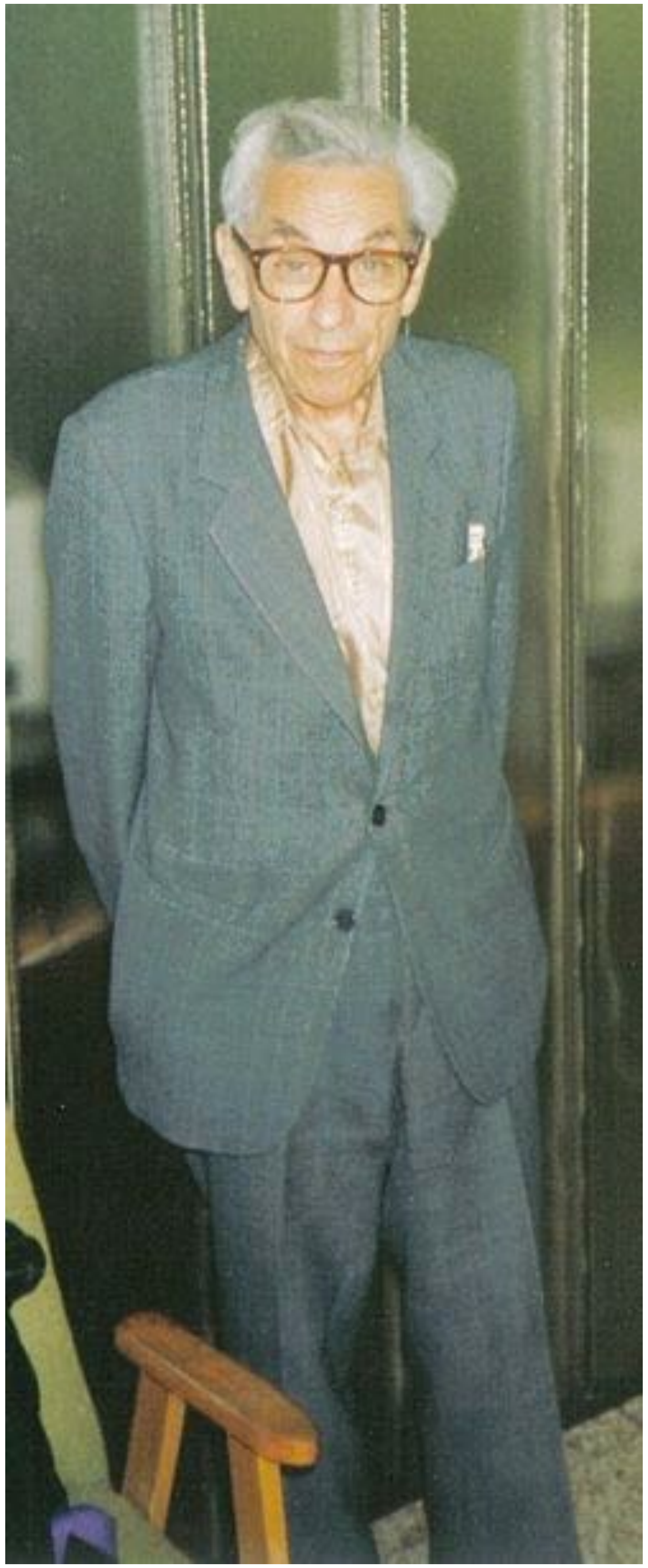
Milgram small-world experiment (1967)

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 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 $\sim 1\,\mathrm{km\,s^{-1}}$) 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 $d = \frac{log N}{log \langle k \rangle}$ ✓
- **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)

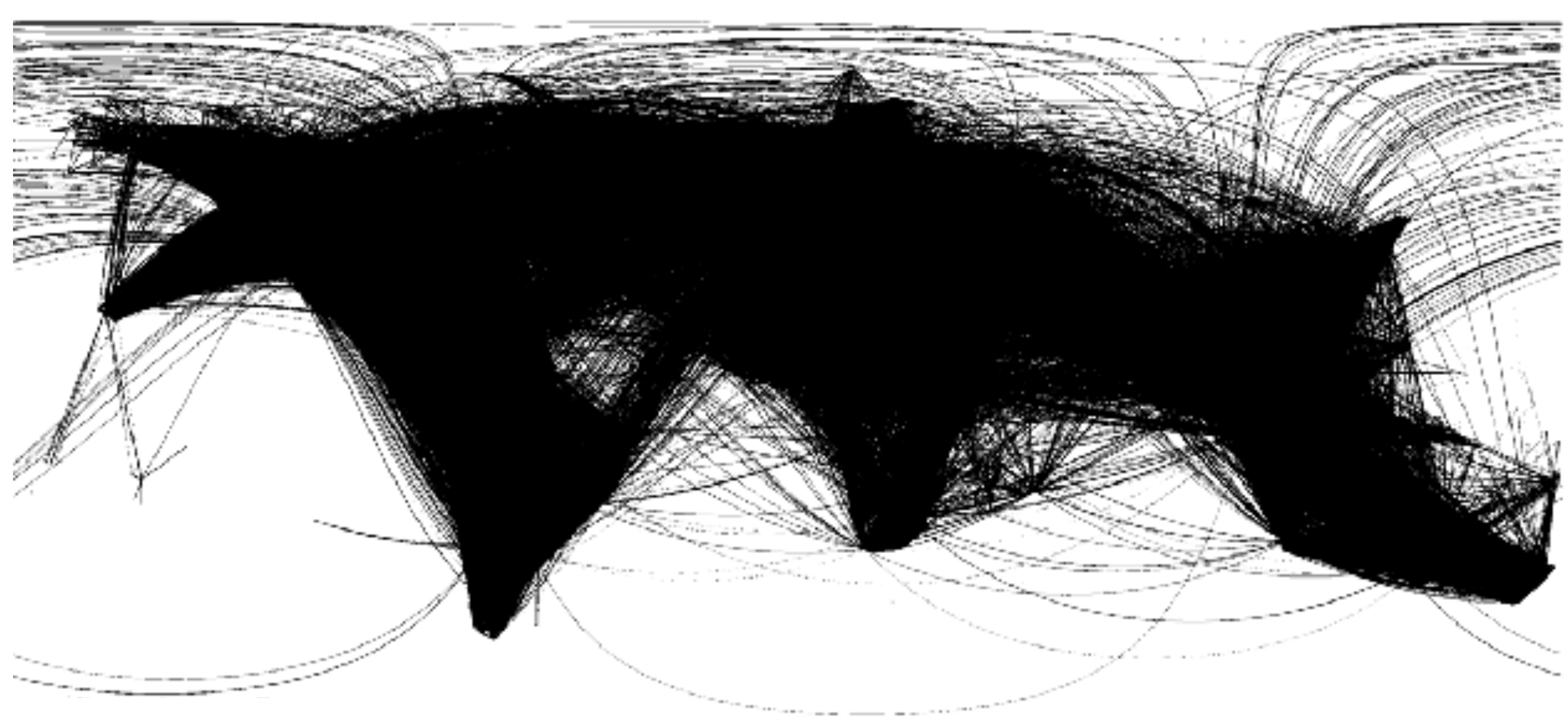
Clustering vs. Interconnectedness

High clustering

- Locally structured
- No connections between nodes apart

Random

- Globally interconnected
- Low clustering



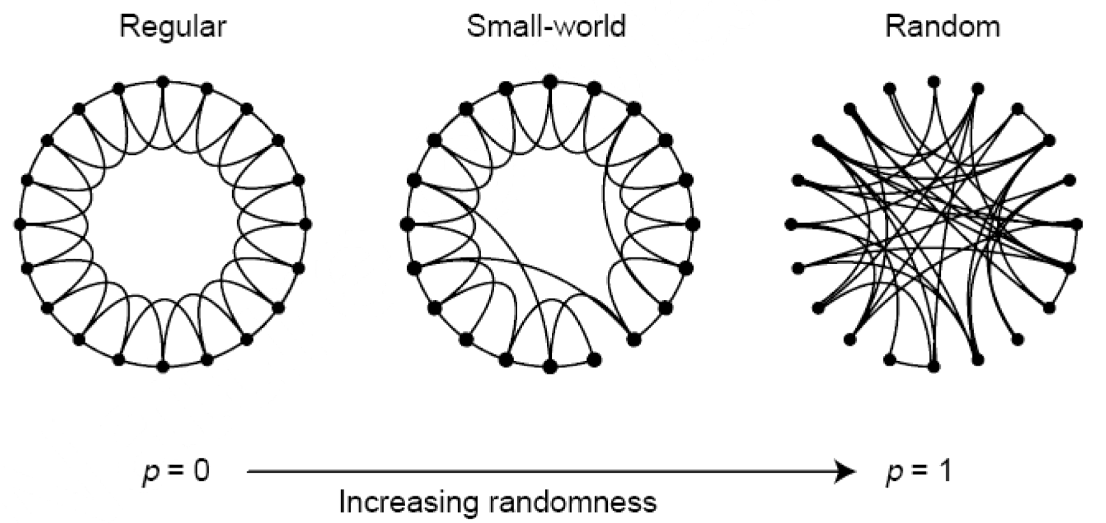
Clustering vs. Interconnectedness

Real networks have high clustering and short distances

The Watts-Strogatz model

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

- It interpolates between and 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.

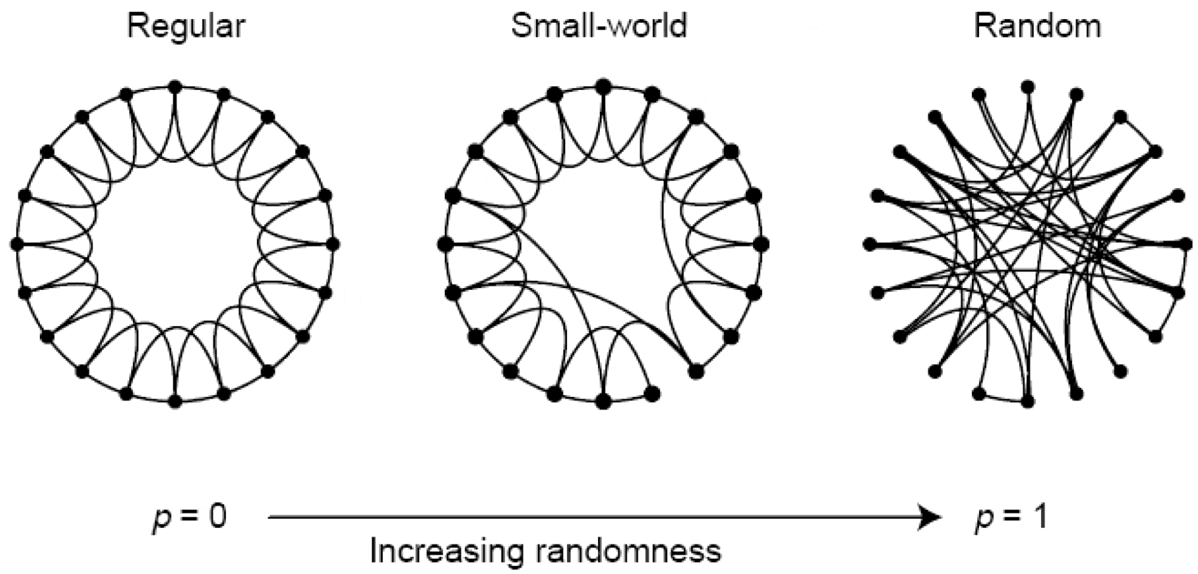
By varying p the network can be transformed from a completely ordered ($p=0$) to a completely random ($p=1$) structure

The Watts-Strogatz model

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

$C \simeq 3/4$

$\langle l \rangle = N/2K$



$C \simeq K/N$

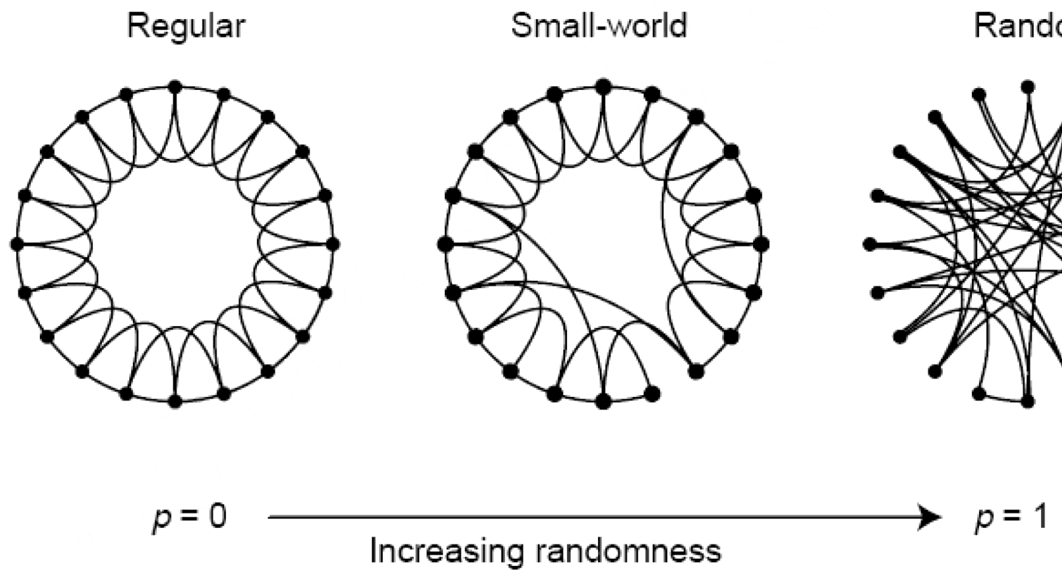
$\langle l \rangle = \ln(N)/\ln(k)$

The Watts-Strogatz model

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

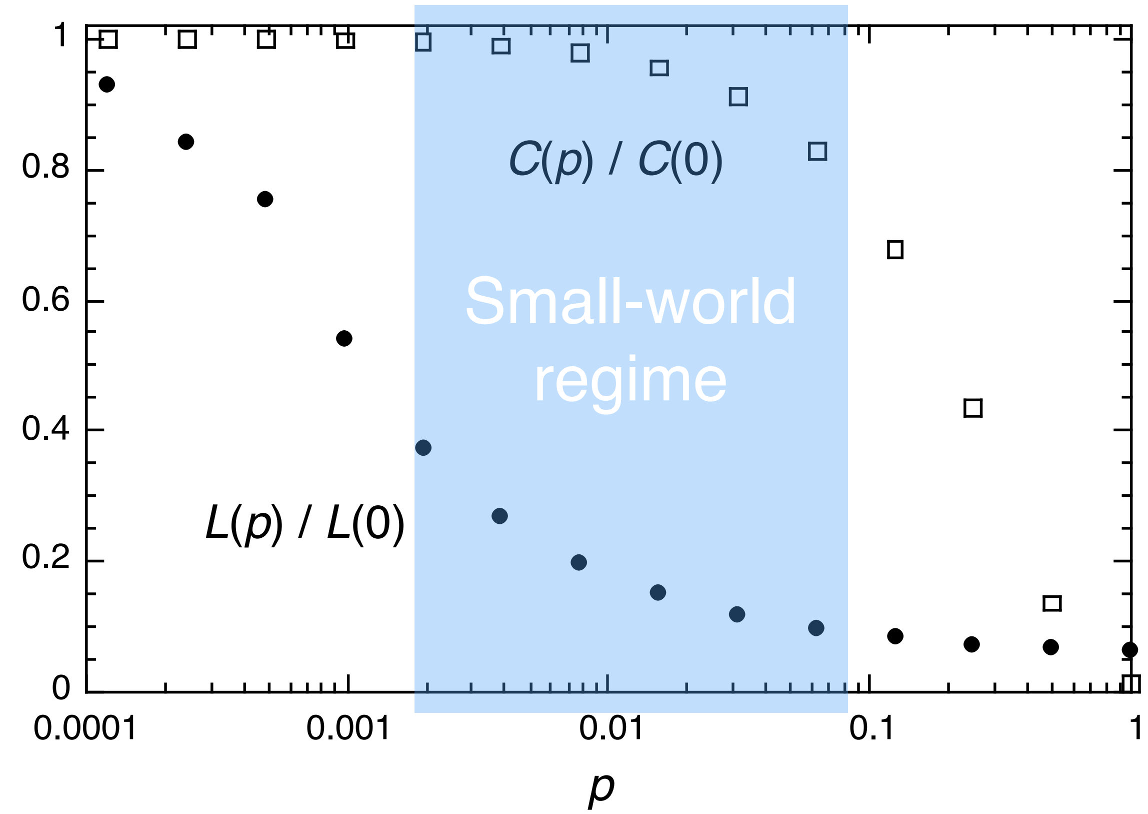
$C \simeq 3/4$

$\langle l \rangle = N/2K$



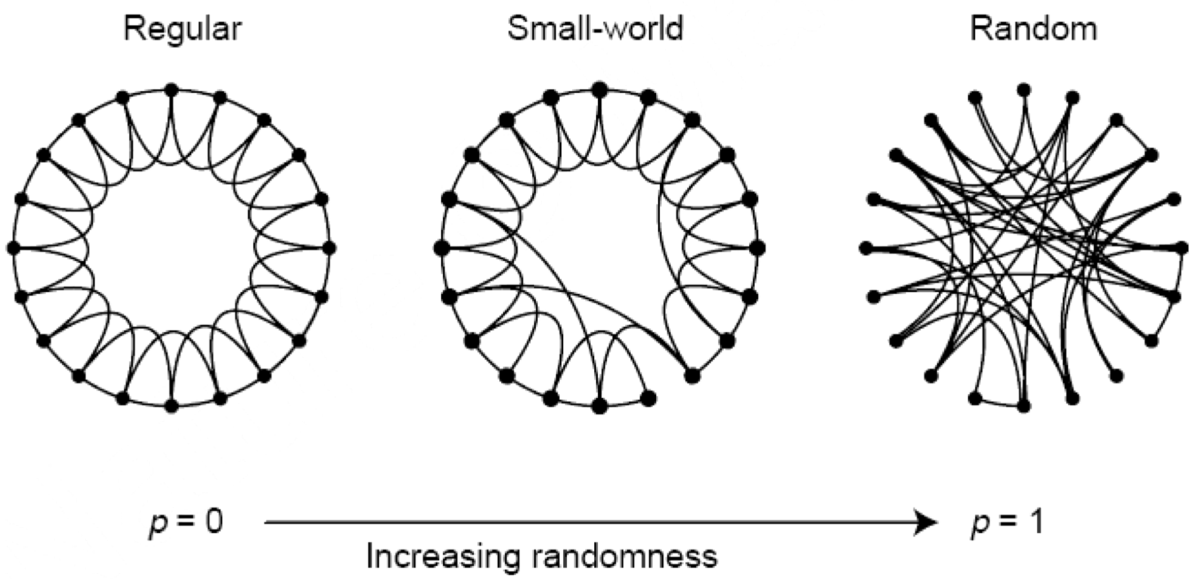
$C \simeq K/N$

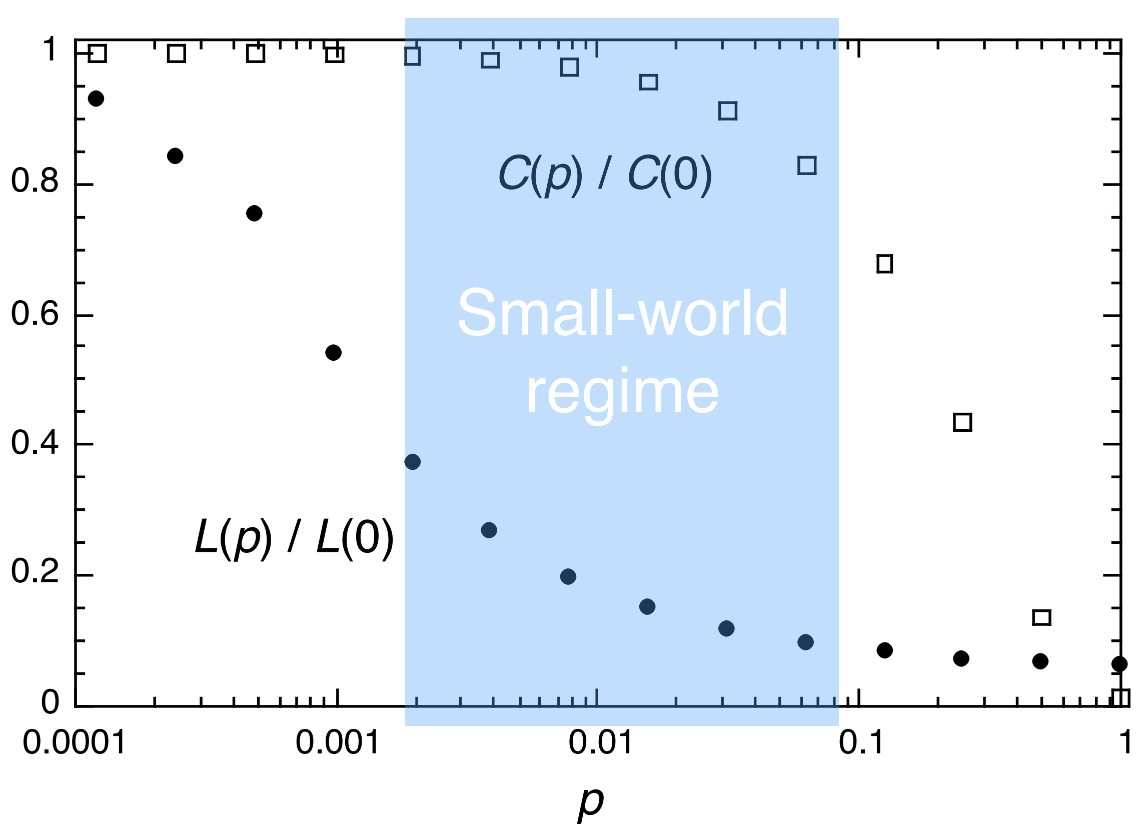
$\langle l \rangle = ln(N)/ln(k)$



The Watts-Strogatz model

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





Clustering (global):

Barrat & Weigt (2000)
Newman, Strogatz, and Watts (2001)

$$C'(p) = \frac{3 \times number\ of\ triangles}{number\ of\ connected\ triples} = \frac{3 \times \triangle}{\wedge} \simeq C(0)(1-p)^3$$

- Decreasing only after an extended initial regime for larger p values

Path length (average)

Barrat (1999), Barthalemy & Amaral (1999)

$$\ell(N,p) \sim \frac{N^{1/d}}{K} f(pKN)$$

$$f(u) = \begin{cases} const & if \quad u << 1 \\ ln(u)/u & if \quad u >> 1 \end{cases}$$

- Decreasing rapidly for small p values

Degree distribution

Barrat & Weigt (2000)

$$P(k) = \sum_{n=0}^{f(k,K)} C_{K/2}^{n} (1-p)^n p^{K/2-n} \frac{(pK/2)^{k-K/2-n}}{(k-K/2-n)!} e^{-pK/2}$$

- Very similar to random graphs
  - Well defined average
  - Exponential tail

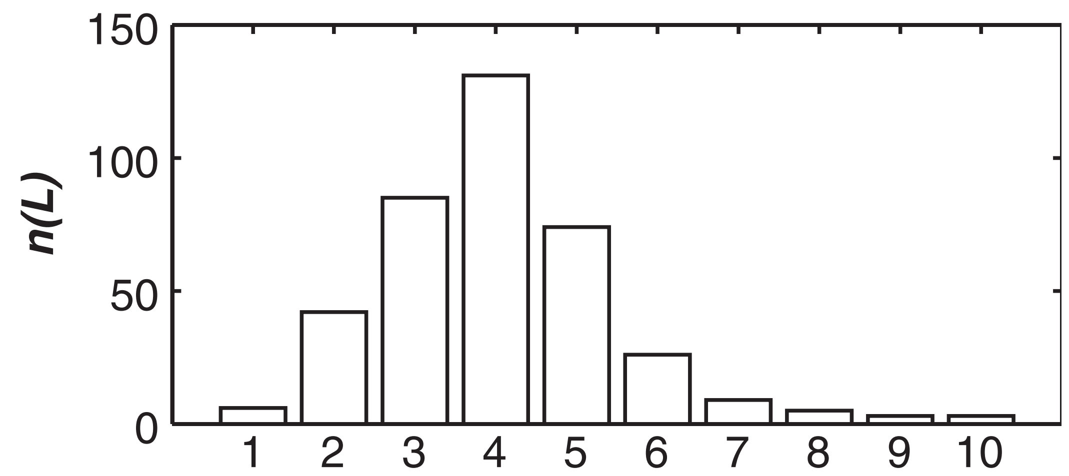
The Watts experiment

- 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)

ER Random Network - catch up

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