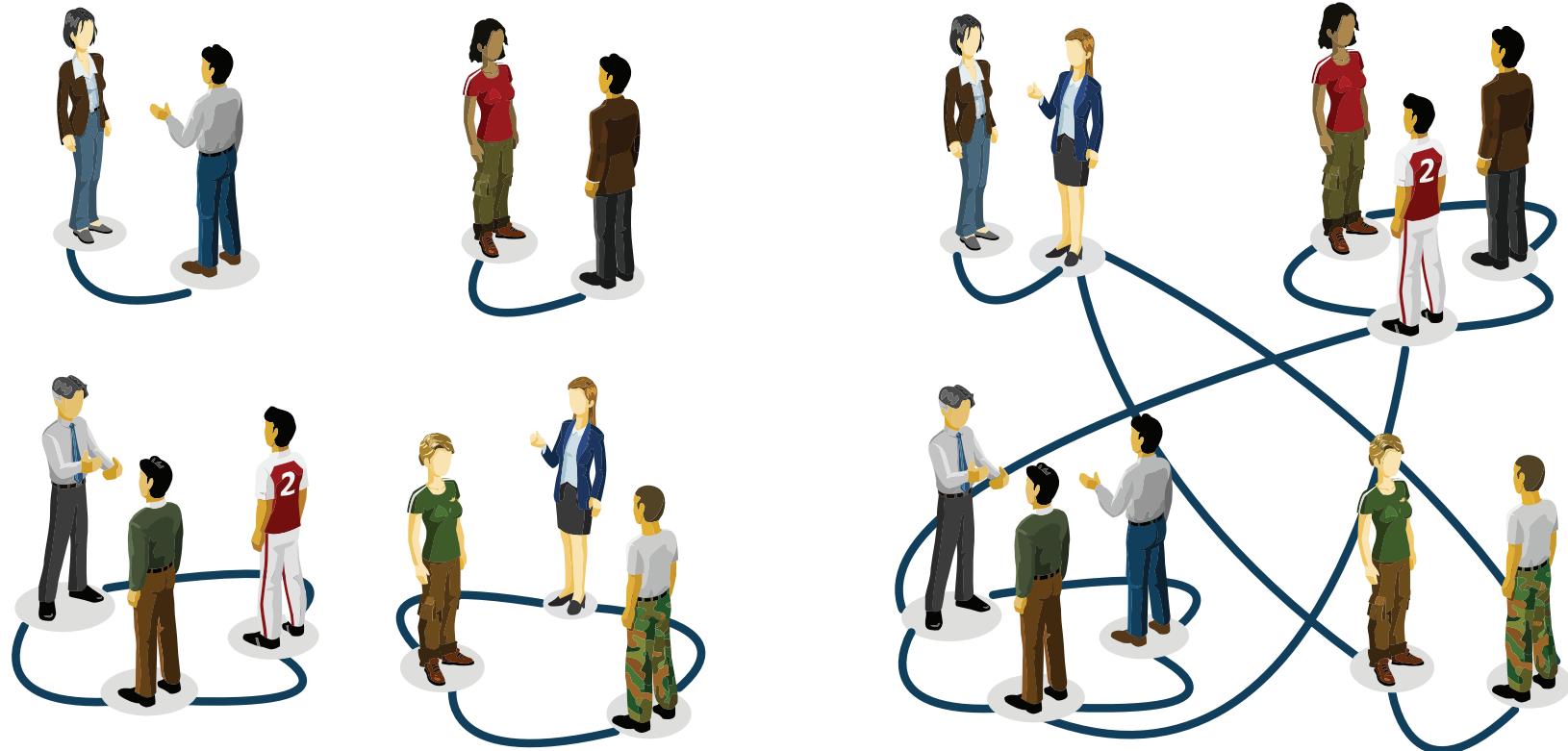


INTRODUCTION

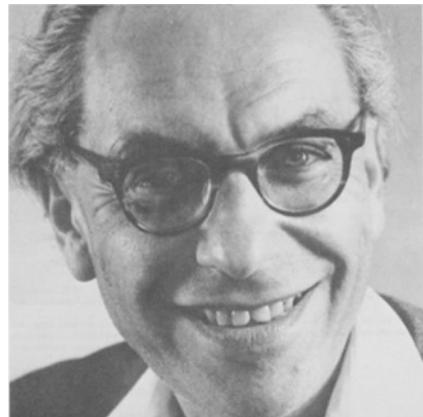
RANDOM NETWORK MODEL



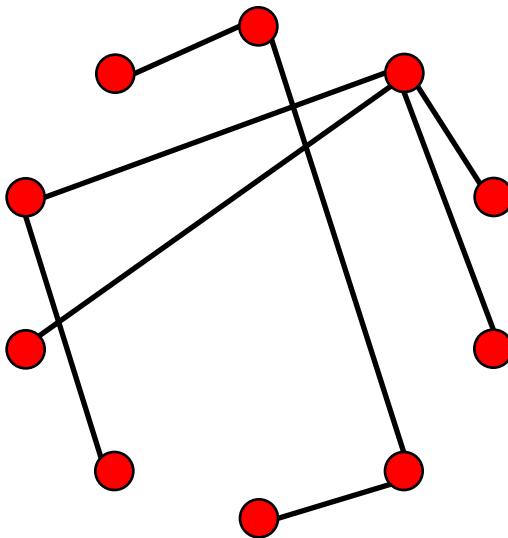
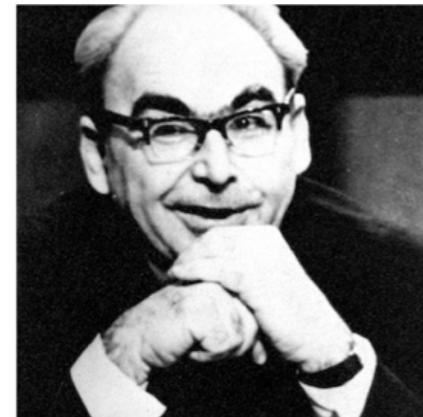
THE RANDOM NETWORK MODEL

RANDOM NETWORK MODEL

Pál Erdős
(1913-1996)



Alfréd Rényi
(1921-1970)



Erdős-Rényi model (1960)

Connect with probability p

$$p=1/6 \quad N=10$$

$$\langle k \rangle \sim 1.5$$

RANDOM NETWORK MODEL

Definition:

A **random graph** is a graph of N nodes where each pair of nodes is connected by probability p .

$G(N, L)$ Model

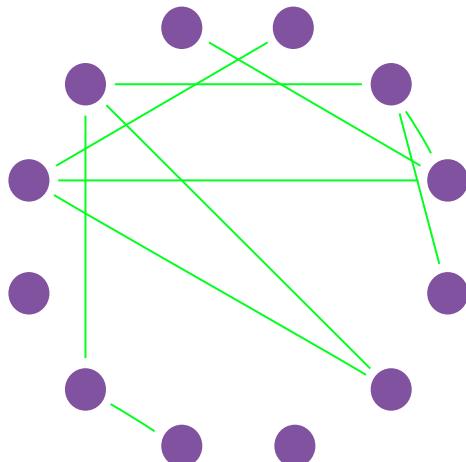
N labeled nodes are connected with L randomly placed links. Erdős and Rényi used this definition in their string of papers on random networks [2-9].

$G(N, p)$ Model

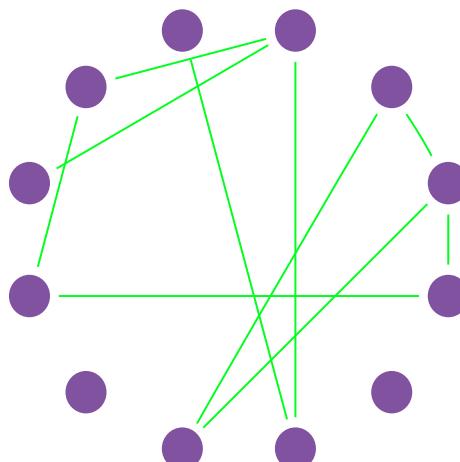
Each pair of N labeled nodes is connected with probability p , a model introduced by Gilbert [10].

RANDOM NETWORK MODEL

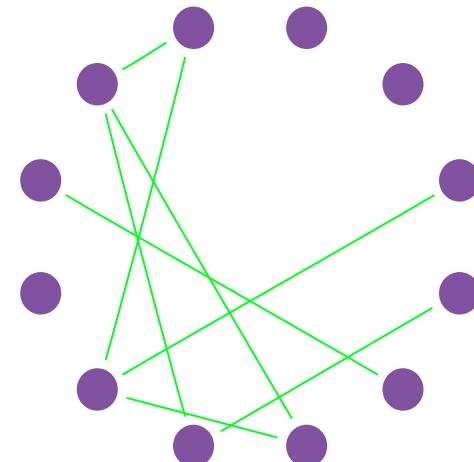
$p=1/6$
 $N=12$



$L=8$



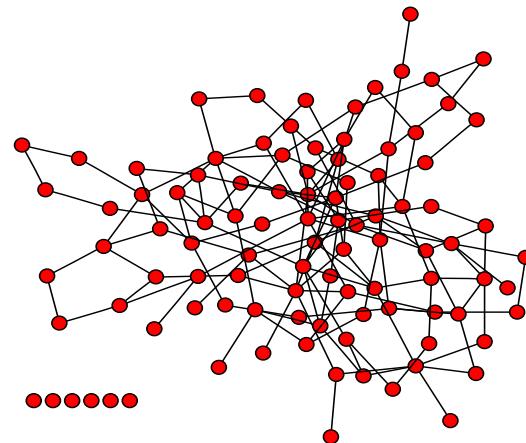
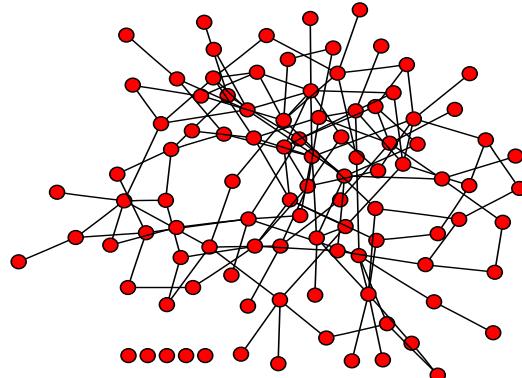
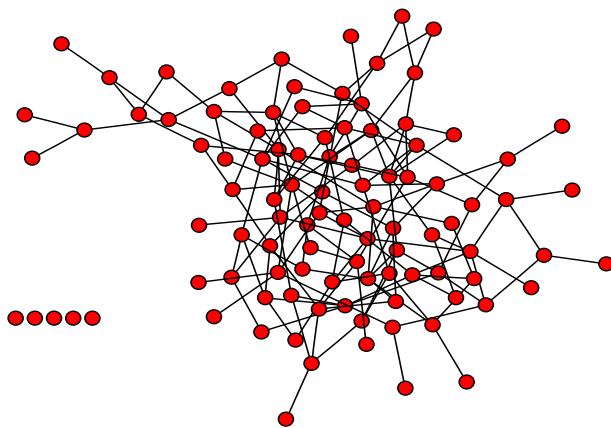
$L=10$



$L=7$

RANDOM NETWORK MODEL

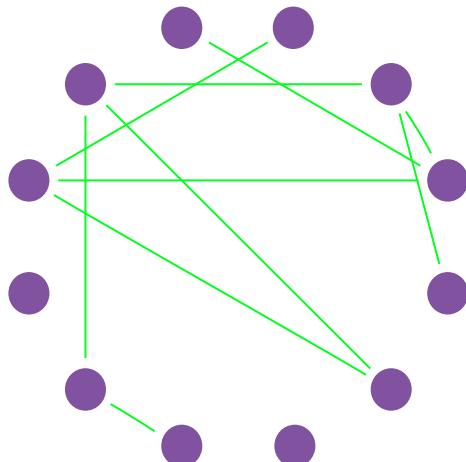
$p=0.03$
 $N=100$



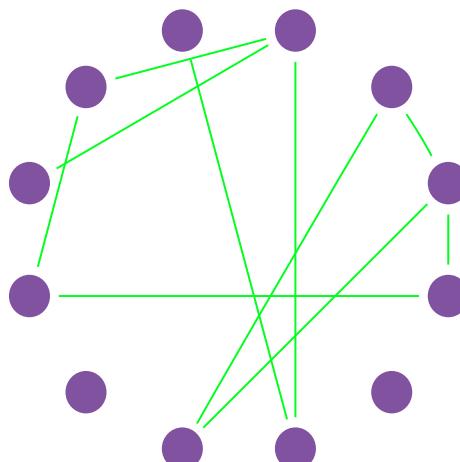
THE NUMBER OF LINKS IS
VARIABLE

RANDOM NETWORK MODEL

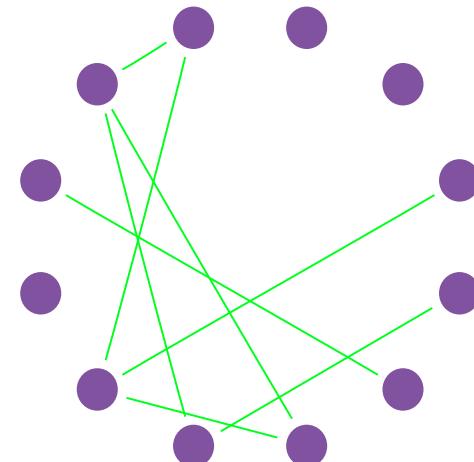
$p=1/6$
 $N=12$



$L=8$



$L=10$



$L=7$

Number of links in a random network

P(L): the probability to have exactly L links in a network of N nodes and probability p :

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

The maximum number of links
in a network of N nodes.

Binomial distribution...

Number of different ways we can choose
 L links among all potential links.

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

$$\langle x \rangle = Np$$

$$\langle x^2 \rangle = p(1-p)N + p^2N^2$$

$$\sigma_x = (\langle k^2 \rangle - \langle k \rangle^2)^{1/2} = [p(1-p)N]^{1/2}$$

RANDOM NETWORK MODEL

$P(L)$: the probability to have a network of exactly L links

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

- The average number of links $\langle L \rangle$ in a random graph

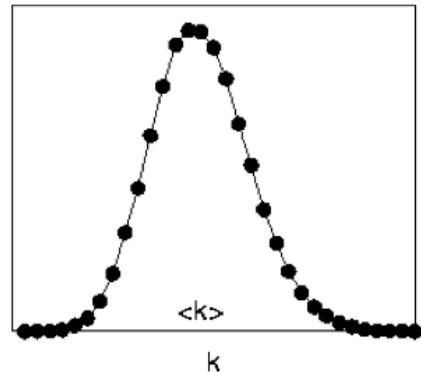
$$\langle L \rangle = \sum_{L=0}^{\frac{N(N-1)}{2}} L P(L) = p \frac{N(N-1)}{2} \qquad \qquad \langle k \rangle = 2L/N = p(N-1)$$

- The standard deviation

$$\sigma^2 = p(1-p) \frac{N(N-1)}{2}$$

DEGREE DISTRIBUTION

DEGREE DISTRIBUTION OF A RANDOM GRAPH



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

Select k nodes from N-1

probability of having k edges

probability of missing $N-1-k$ edges

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

$$\sigma_k^2 = p(1-p)(N-1)$$

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

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 $\langle k \rangle$.

POISSON DEGREE DISTRIBUTION

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

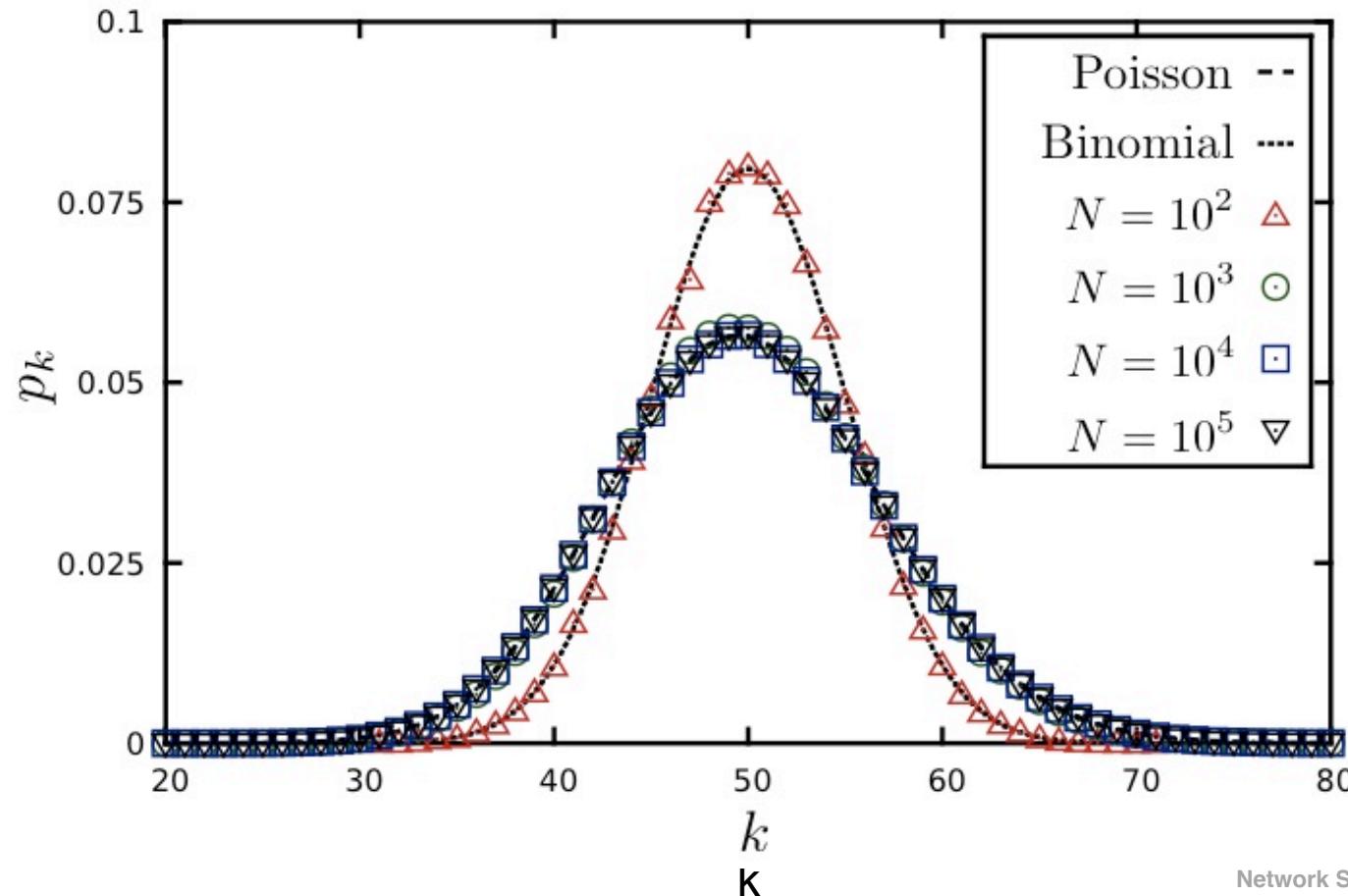
For large N and small k , we arrive to the Poisson distribution:

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

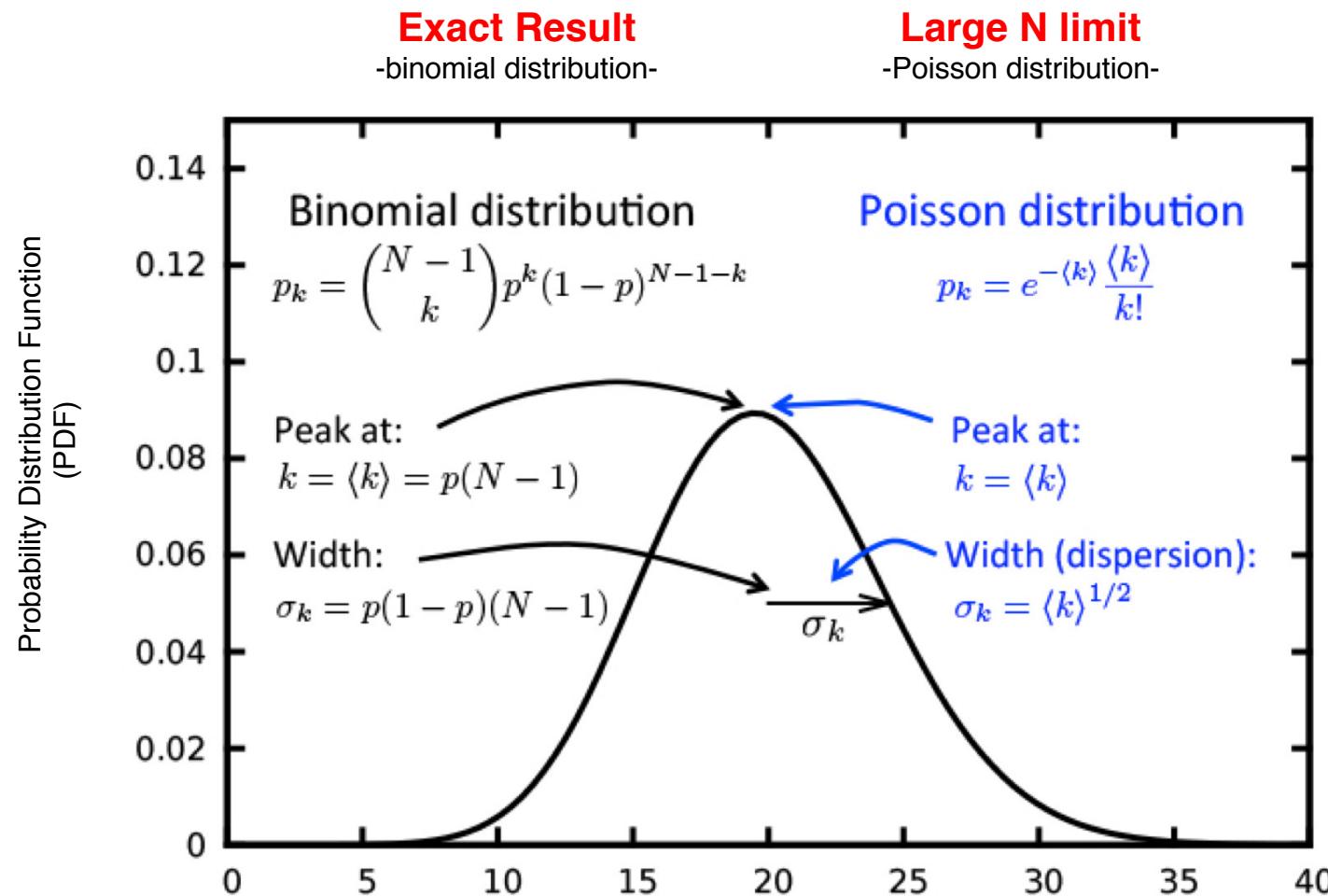
DEGREE DISTRIBUTION OF A RANDOM GRAPH

$\langle k \rangle = 50$

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



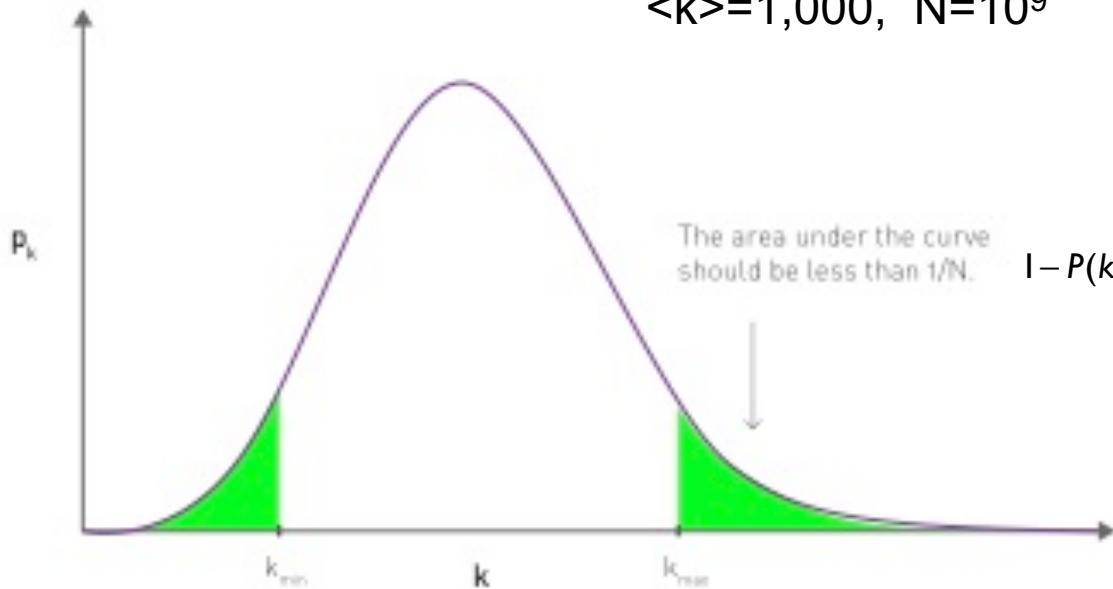
DEGREE DISTRIBUTION OF A RANDOM NETWORK



REAL NETWORKS ARE NOT
POISSON

MAXIMUM AND MINIMUM DEGREE

$$\langle k \rangle = 1,000, N = 10^9$$



$$NP(k_{min}) \approx 1.$$

$$P(k_{min}) = e^{-\langle k \rangle} \sum_{k=0}^{k_{min}} \frac{\langle k \rangle^k}{k!}. \quad k_{min} = 816$$

$$N[1 - P(k_{max})] \approx 1.$$
$$1 - P(k_{max}) = 1 - e^{-\langle k \rangle} \sum_{k=0}^{k_{max}} \frac{\langle k \rangle^k}{k!} = e^{-\langle k \rangle} \sum_{k=k_{max}+1}^{\infty} \frac{\langle k \rangle^k}{k!} \approx e^{-\langle k \rangle} \frac{\langle k \rangle^{k_{max}+1}}{(k_{max}+1)!},$$

$$\langle k \rangle = 1,000, N = 10^9$$

$$k_{max} = 1,185$$

$$\langle k \rangle \pm \sigma_k \quad \sigma_k = \langle k \rangle^{1/2}$$
$$\sigma_k = 31.62.$$

NO OUTLIERS IN A RANDOM SOCIETY

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

- The most connected individual has degree $k_{\max} \sim 1,185$
- The least connected individual has degree $k_{\min} \sim 816$

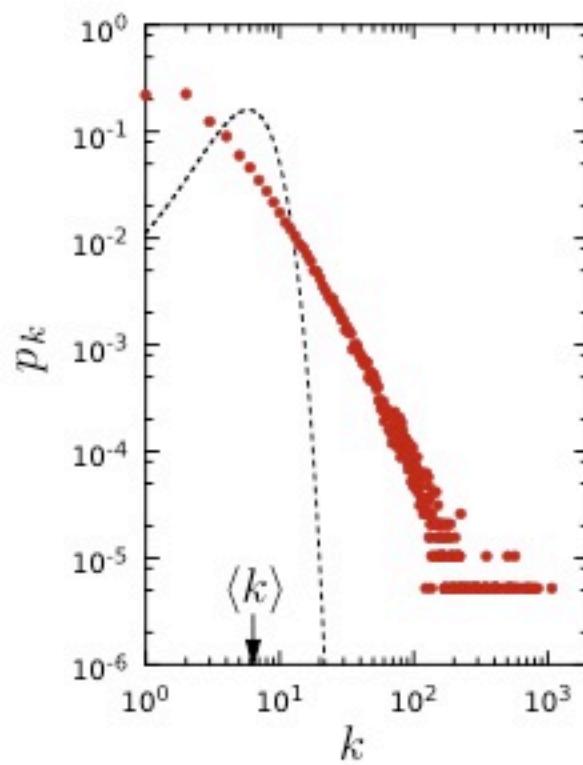
The probability to find an individual with degree $k > 2,000$ is 10^{-27} . Hence the chance of finding an individual with 2,000 acquaintances is so tiny that such nodes are virtually nonexistent in a random society.

- a random society would consist of mainly average individuals, with everyone with roughly the same number of friends.
- It would lack outliers, individuals that are either highly popular or reclusive.

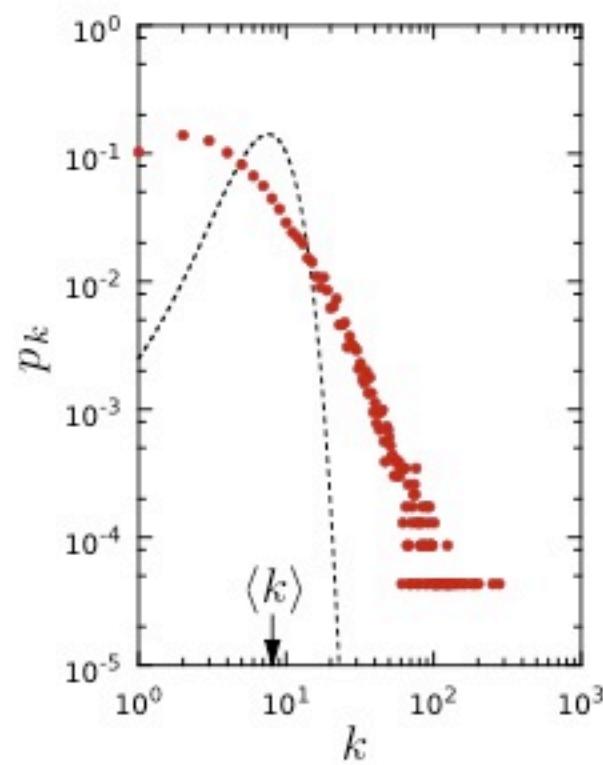
FACING REALITY: DEGREE DISTRIBUTION OF REAL NETWORKS

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

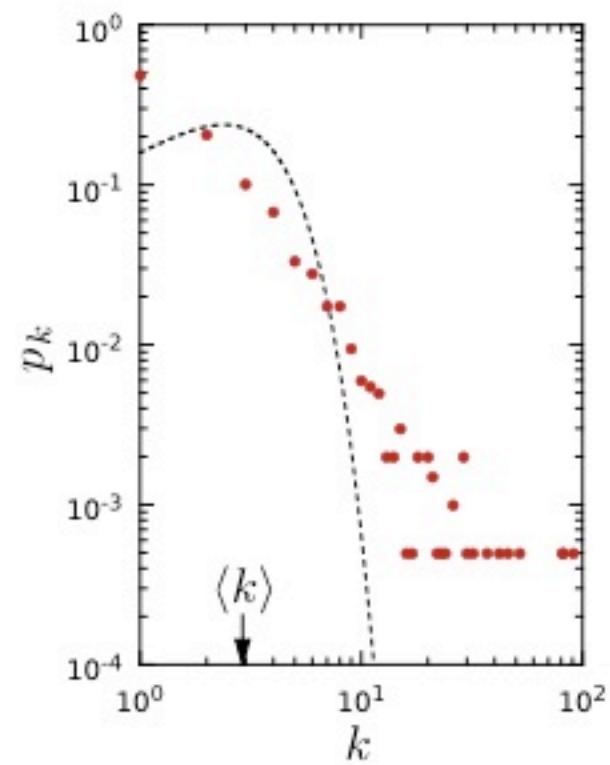
Internet



Science Collaboration



Protein Interactions



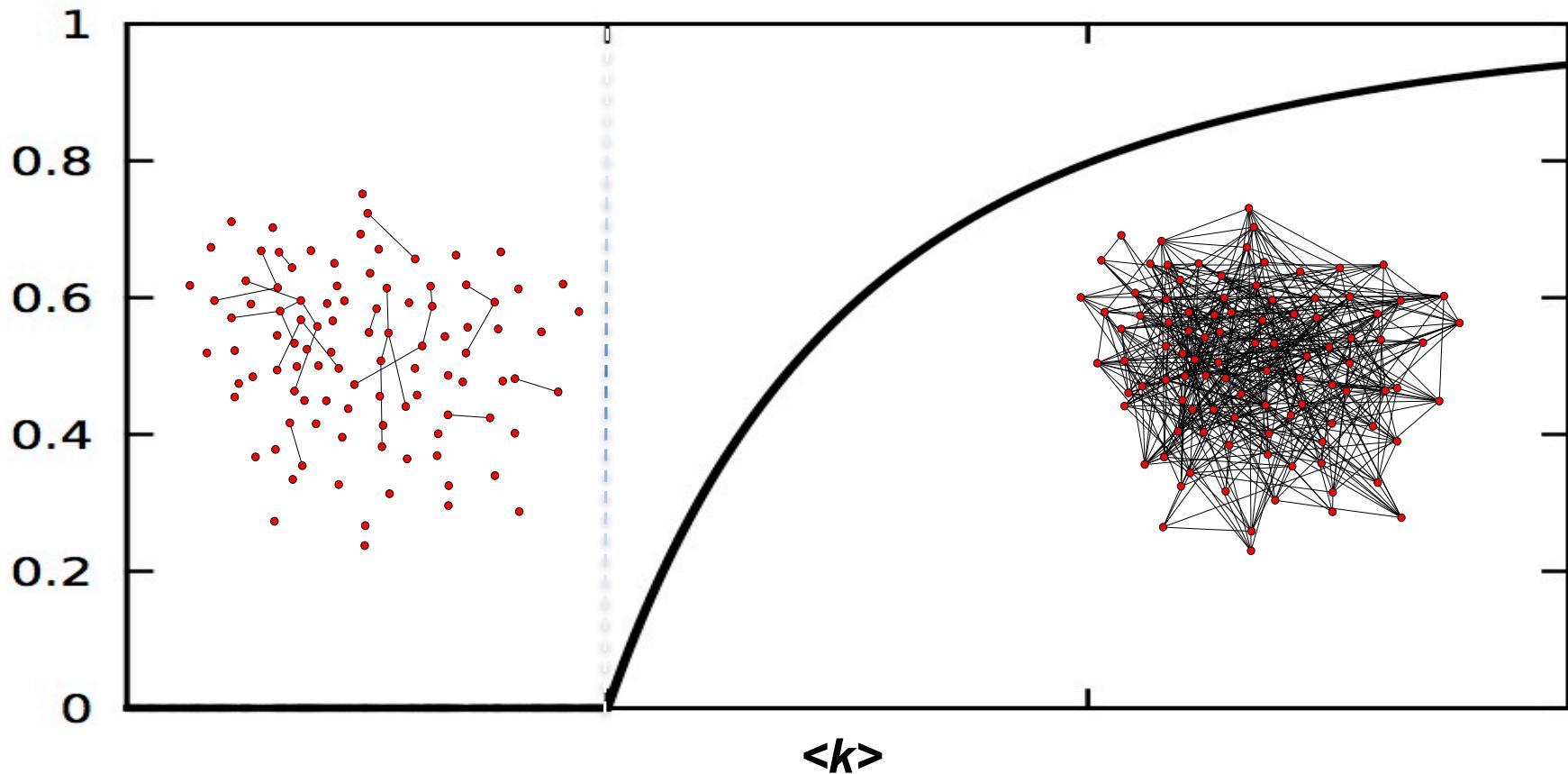
THE EVOLUTION OF A RANDOM NETWORK

EVOLUTION OF A RANDOM NETWORK

DISCONNECTED NODES



NETWORK.



How does this transition happen?

EVOLUTION OF A RANDOM NETWORK

DISCONNECTED NODES



NETWORK

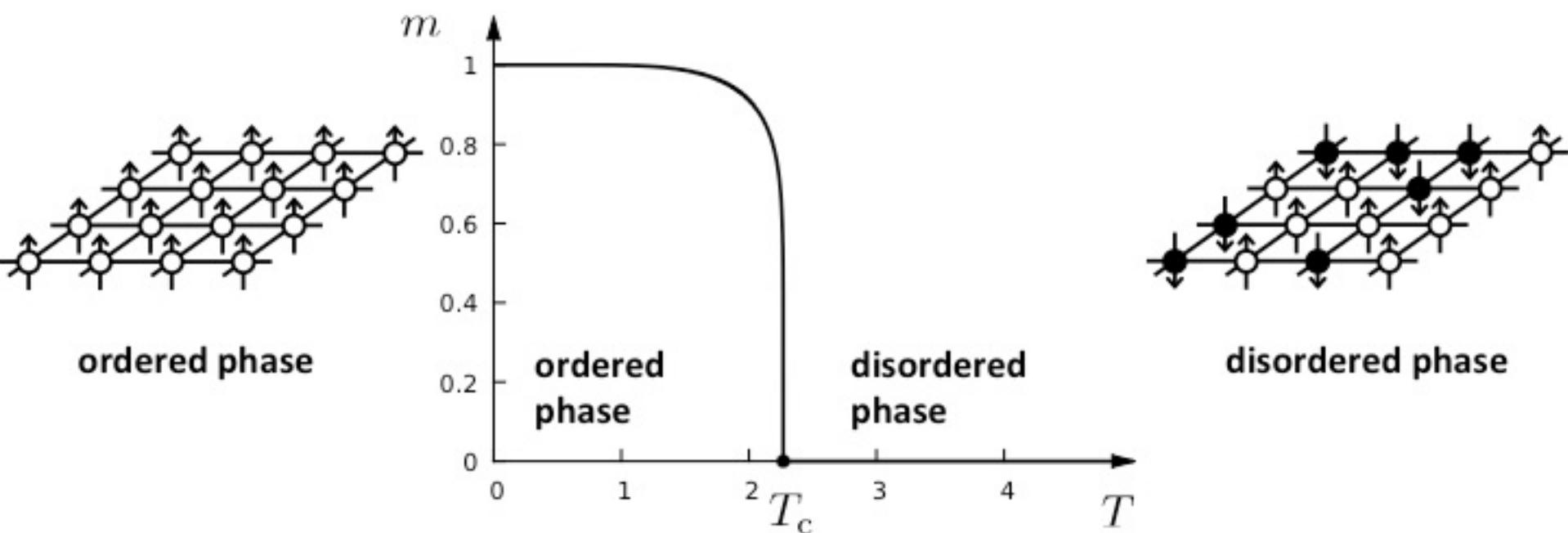
$$\langle k_c \rangle = 1 \quad (\text{Erdos and Renyi, 1959})$$

The fact that at least one link per node is *necessary* to have a giant component is not unexpected. Indeed, for a giant component to exist, each of its nodes must be linked to at least one other node.

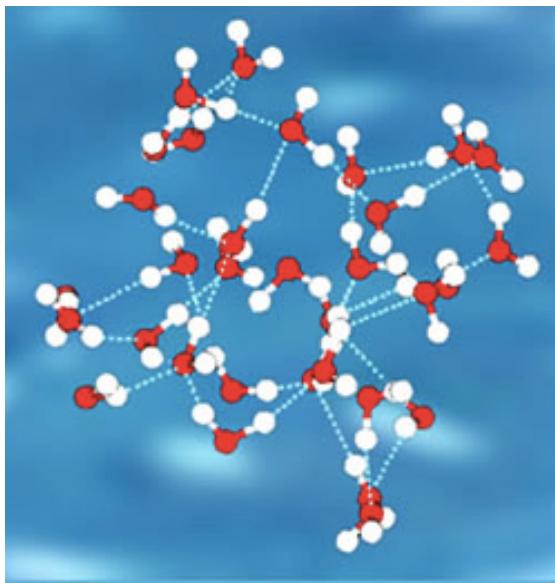
It is somewhat unexpected, however that one link is *sufficient* for the emergence of a giant component.

It is equally interesting that the emergence of the giant cluster is not gradual, but follows what physicists call a second order phase transition at $\langle k \rangle = 1$.

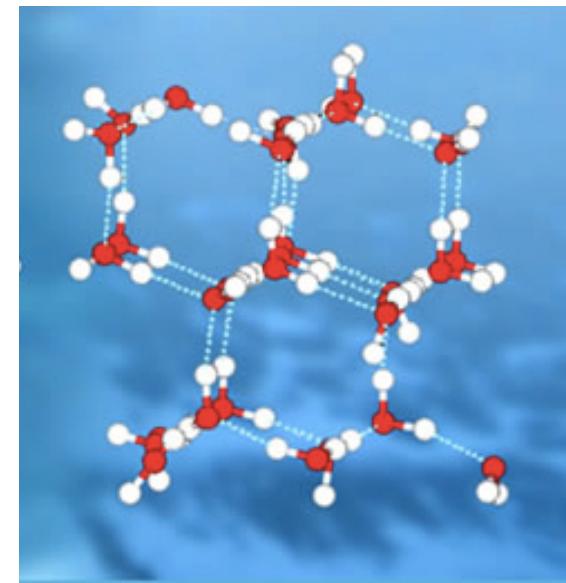
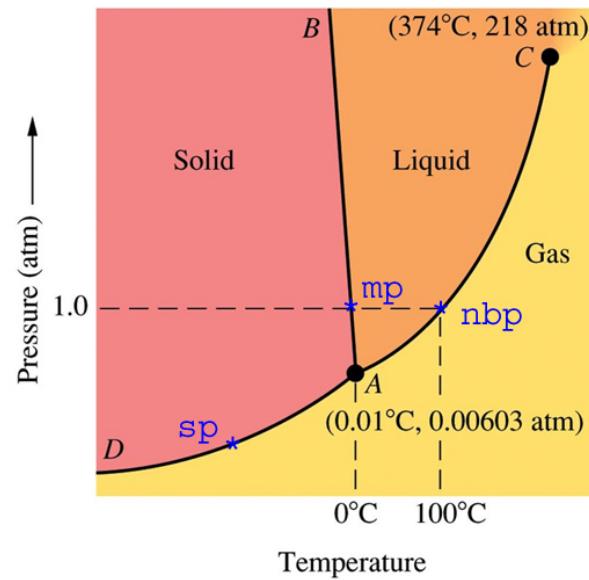
PHASE TRANSITIONS IN COMPLEX SYSTEMS I: MAGNETISM



PHASE TRANSITIONS IN COMPLEX SYSTEMS I: LIQUIDS



Water



Ice

CLUSTER SIZE DISTRIBUTION

Probability that a randomly selected node belongs to a cluster of size s :

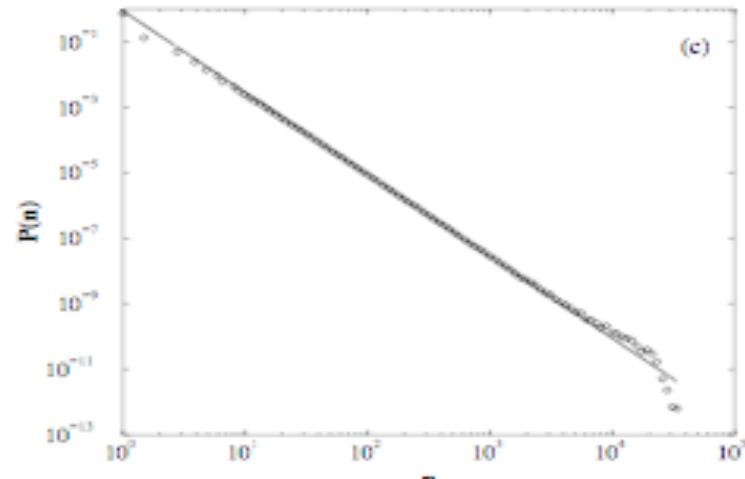
$$p(s) = \frac{e^{-\langle k \rangle s} (\langle k \rangle s)^{s-1}}{s!} \quad \langle k \rangle^{s-1} = \exp[(s-1)\ln\langle k \rangle]$$

$$p(s) = \frac{s^{s-1}}{s!} e^{-\langle k \rangle s + (s-1)\ln\langle k \rangle}$$

$$p(s) \sim s^{-3/2} e^{-(\langle k \rangle - 1)s + (s-1)\ln\langle k \rangle}$$

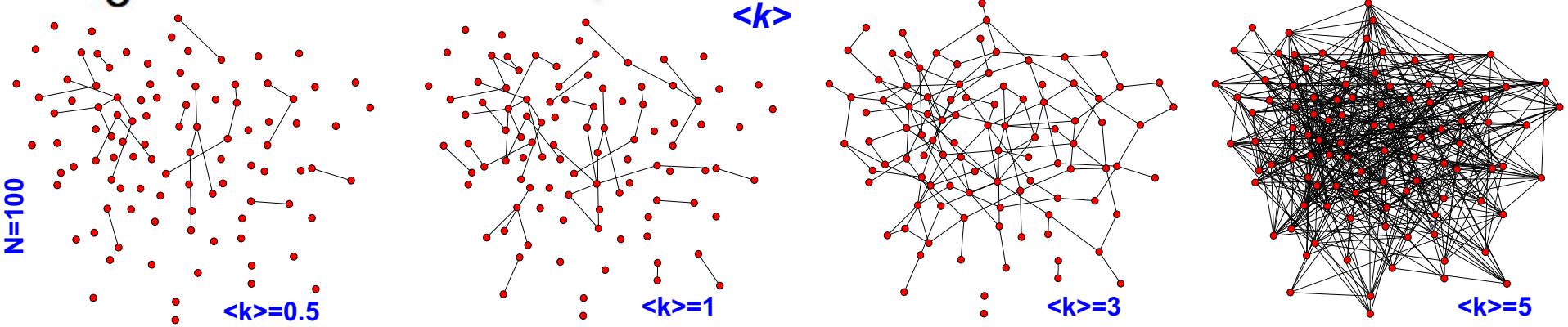
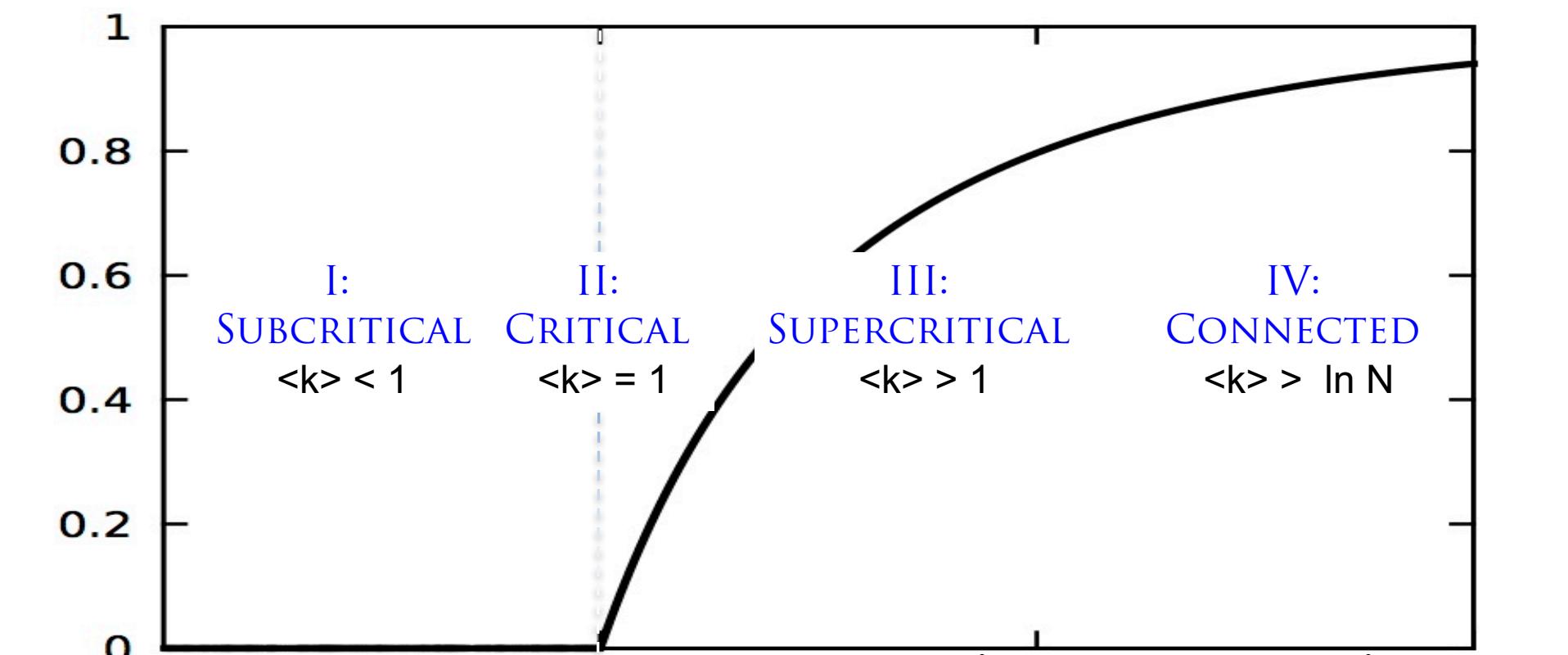
At the critical point $\langle k \rangle = 1$

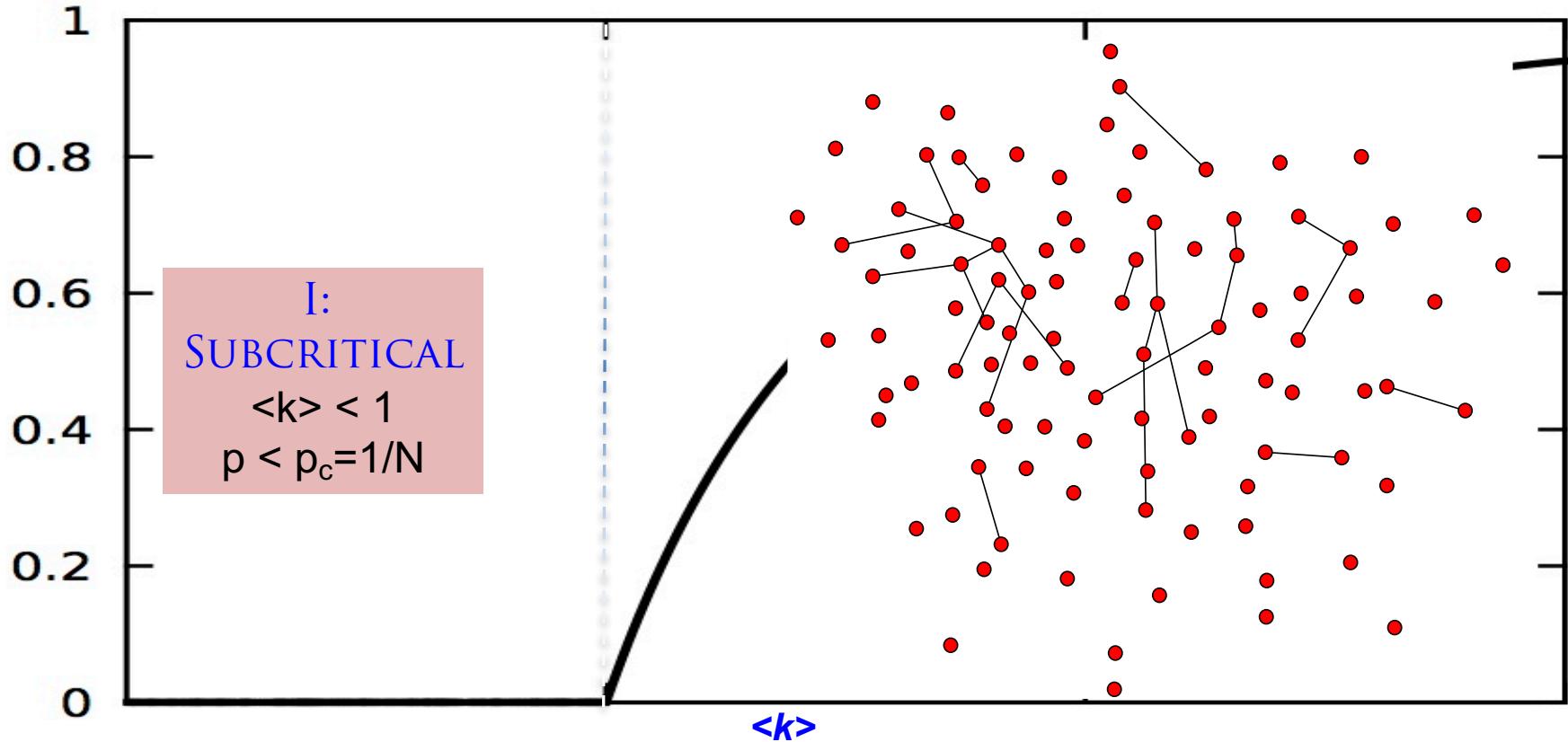
$$p(s) \sim s^{-3/2}$$



The distribution of cluster sizes at the critical point, displayed in a log-log plot. The data represent an average over 1000 systems of sizes. The dashed line has a slope of

$$-\tau_n = -2.5$$

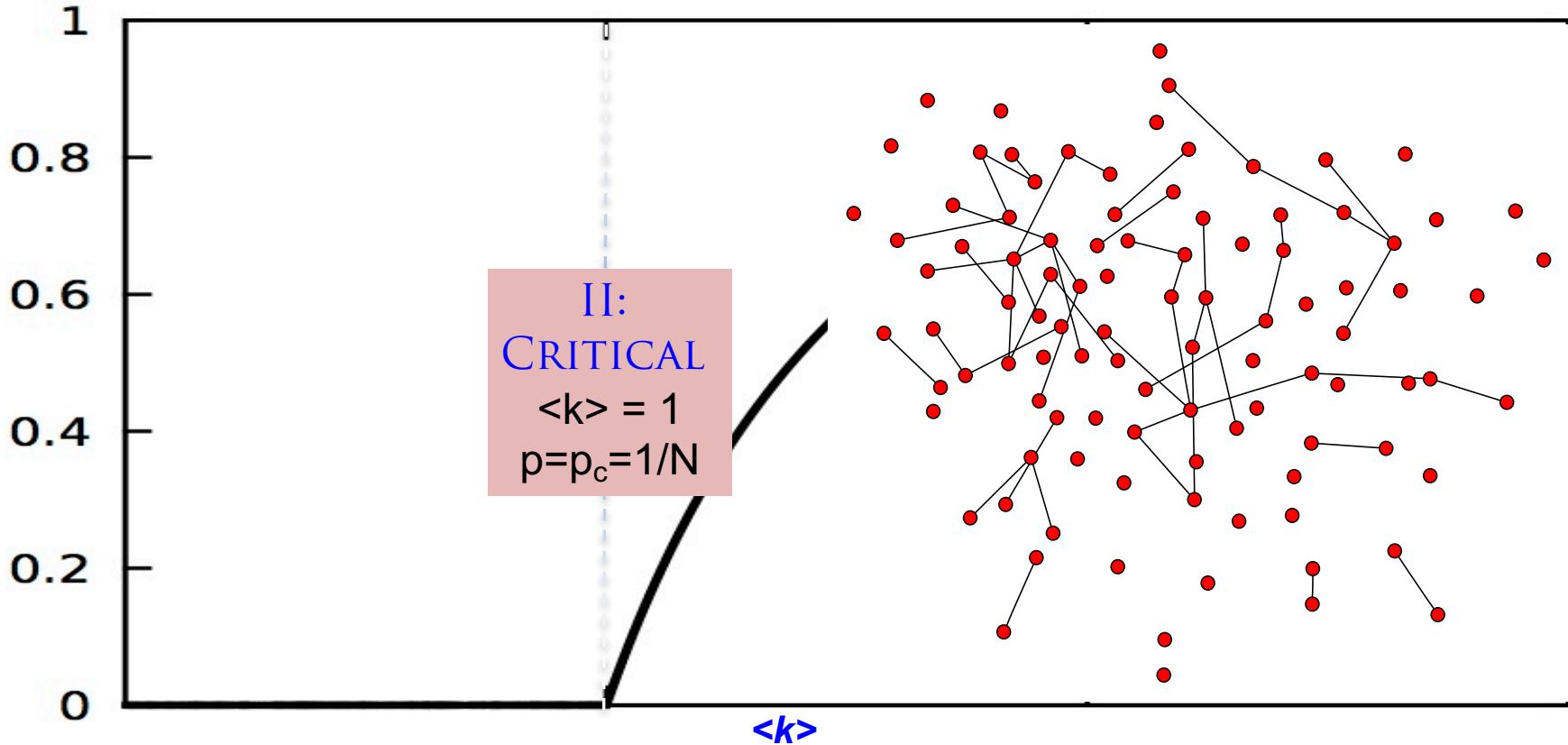




No giant component.

N - L isolated clusters, cluster size distribution is exponential $p(s) \sim s^{-3/2} e^{-(\langle k \rangle - 1)s + (s-1)\ln \langle k \rangle}$

The largest cluster is a tree, its size $\sim \ln N$



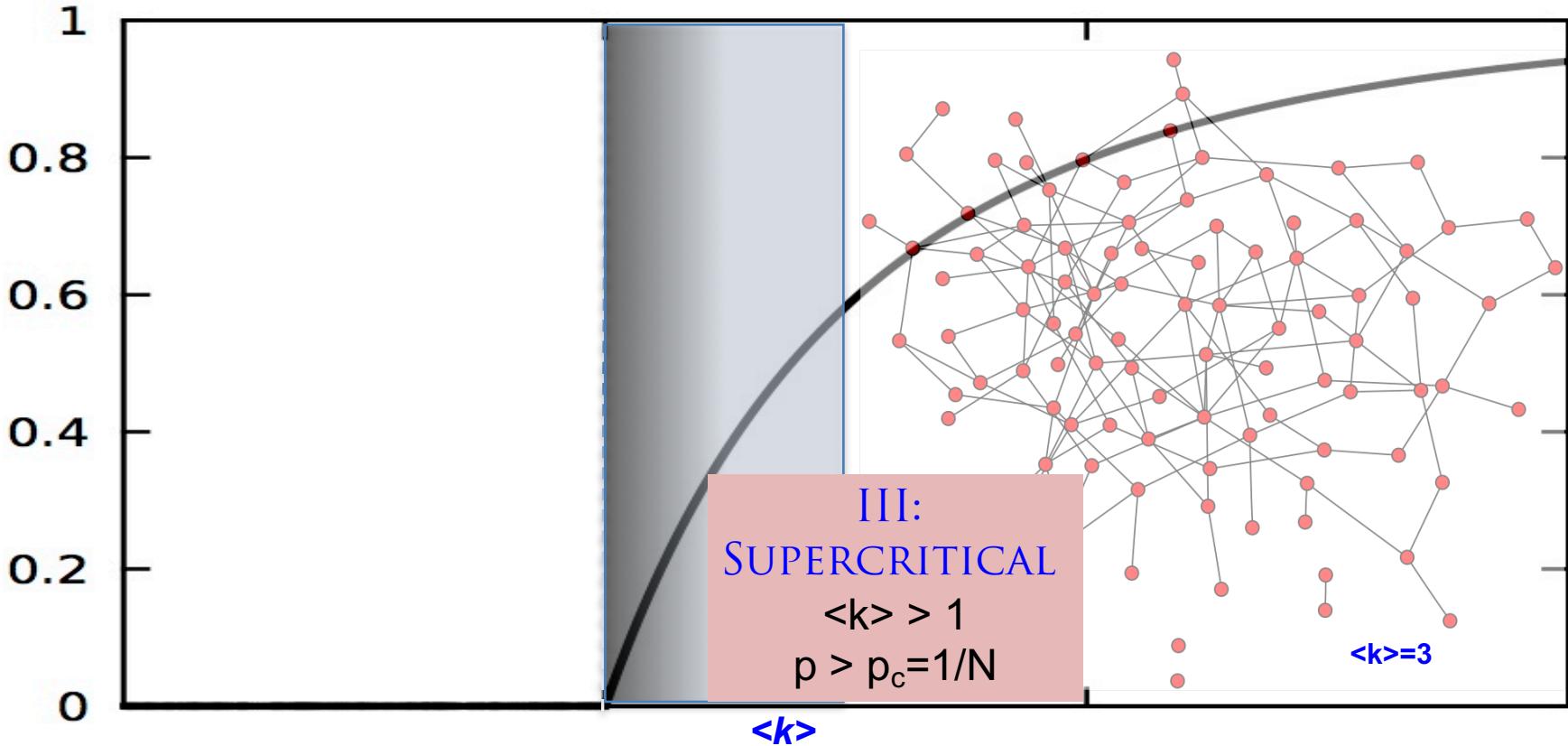
Unique giant component: $N_G \sim N^{2/3}$

→ contains a vanishing fraction of all nodes, $N_G/N \sim N^{-1/3}$

→ Small components are trees, GC has loops.

Cluster size distribution: $p(s) \sim s^{-3/2}$

A jump in the cluster size:
 $N=1,000 \rightarrow \ln N \sim 6.9; N^{2/3} \sim 95$
 $N=7 \cdot 10^9 \rightarrow \ln N \sim 22; N^{2/3} \sim 3,659,250$

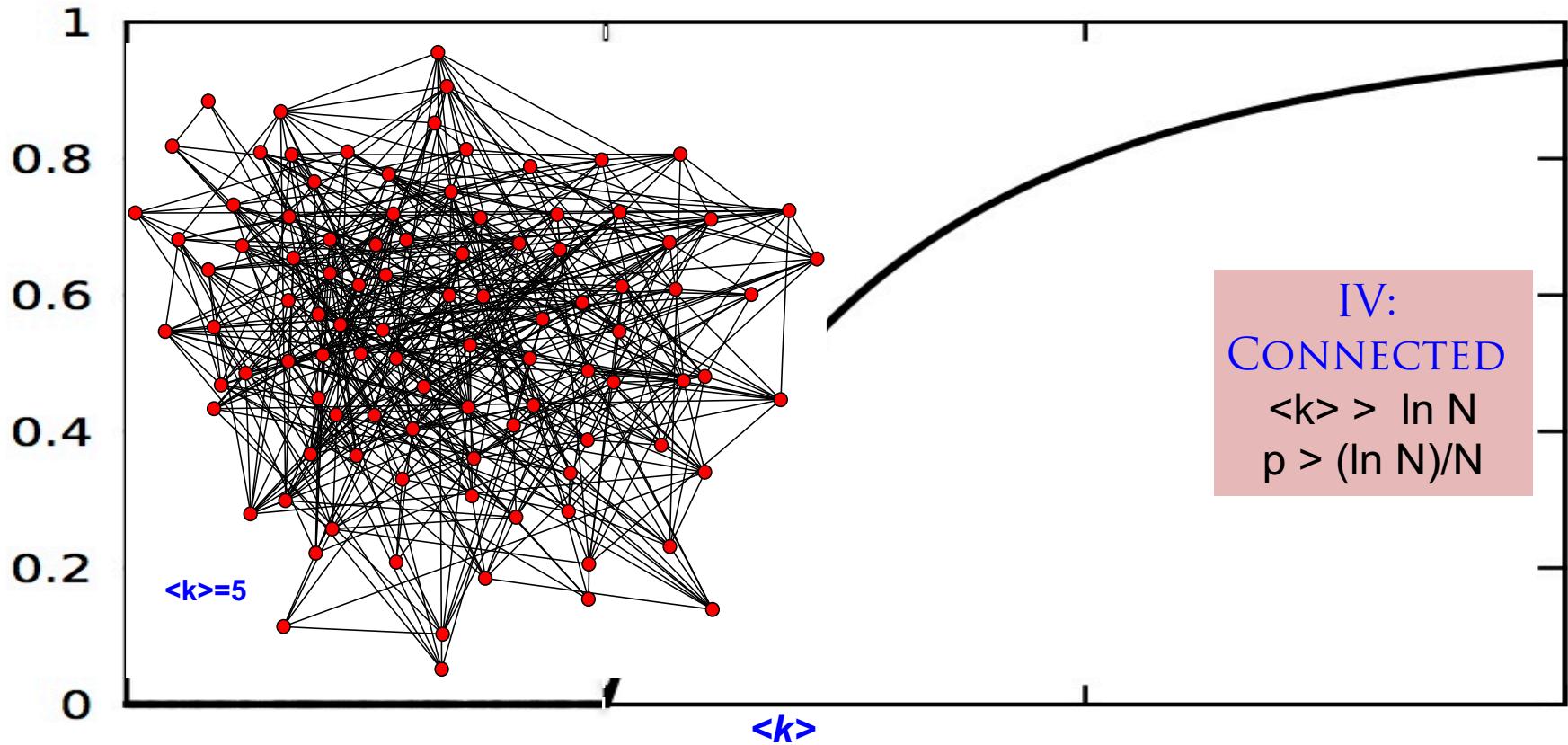


Unique giant component: $N_G \sim (p-p_c)N$

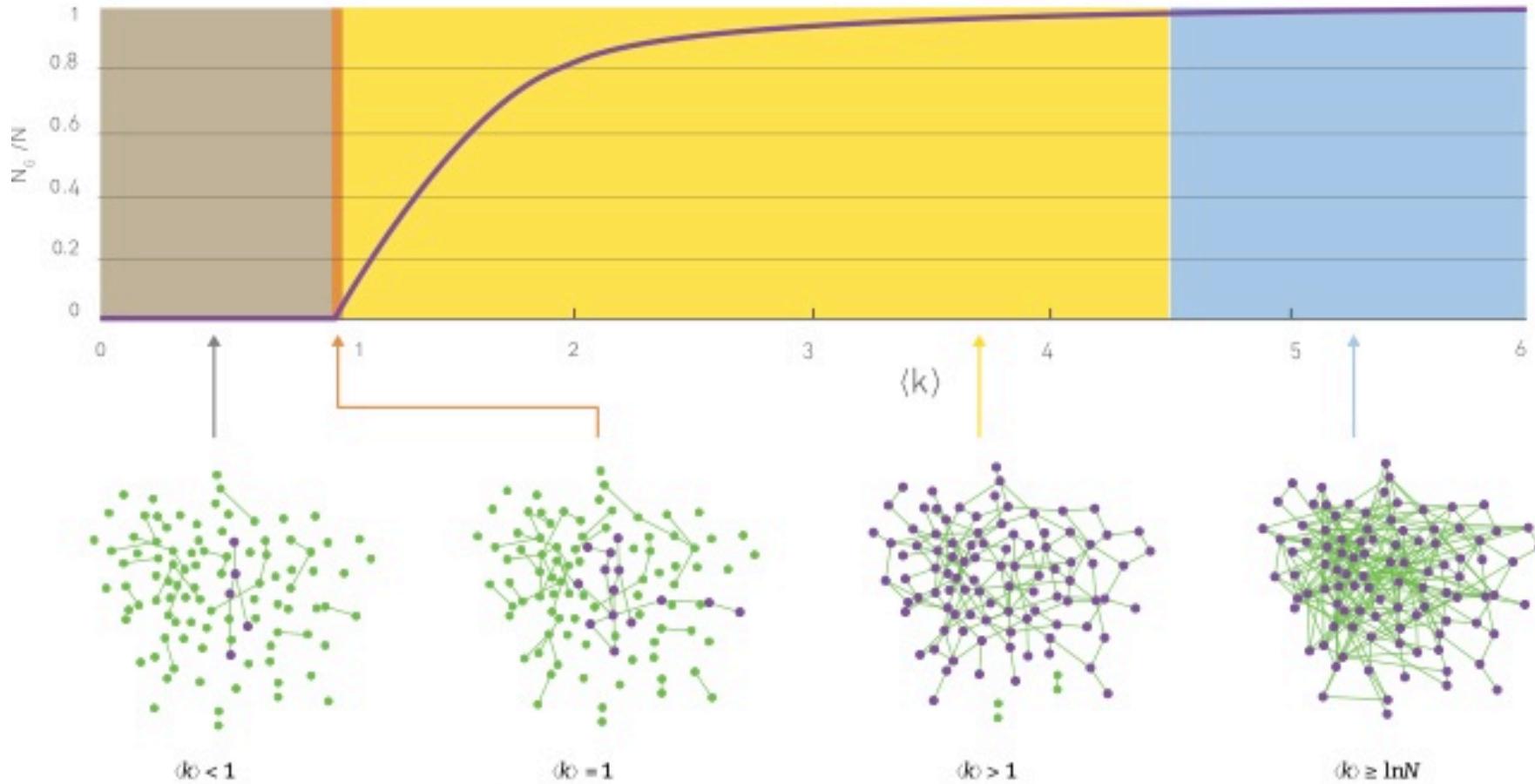
→ GC has loops.

Cluster size distribution: exponential

$$p(s) \sim s^{-3/2} e^{-(\langle k \rangle - 1)s + (s-1)\ln \langle k \rangle}$$



Only one cluster: $N_G=N$
→ GC is dense.
Cluster size distribution: None



(b) Subcritical Regime

- No giant component
- Cluster size distribution: $p_i \sim s^{-3/2} e^{-si}$
- Size of the largest cluster: $N_g \sim \ln N$
- The clusters are trees

(c) Critical Point

- No giant component
- Cluster size distribution: $p_i \sim s^{-3/2}$
- Size of the largest cluster: $N_g \sim N^{1/2}$
- The clusters may contain loops

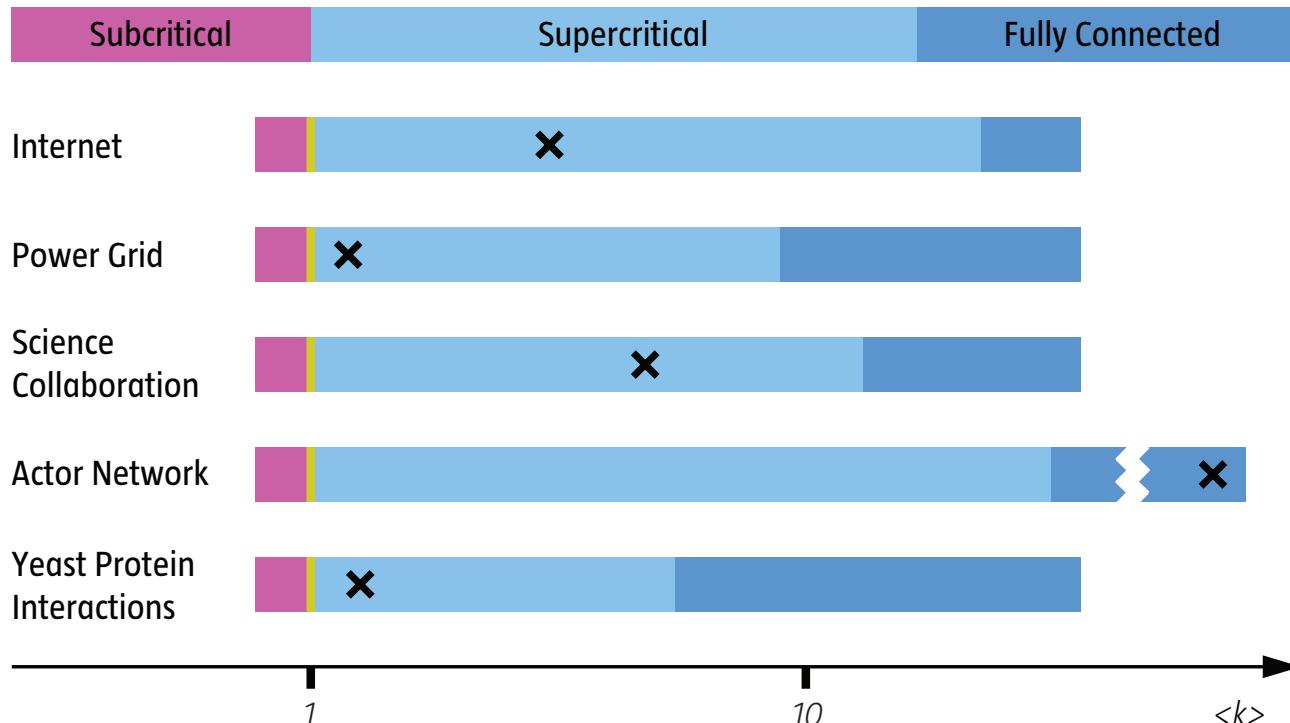
(d) Supercritical Regime

- Single giant component
- Cluster size distribution: $p_i \sim s^{-3/2} e^{-si}$
- Size of the giant component: $N_g \sim (\rho - \rho_c)N$
- The small clusters are trees
- Giant component has loops

(e) Connected Regime

- Single giant component
- No isolated nodes or clusters
- Size of the giant component: $N_g = N$
- Giant component has loops

REAL NETWORKS ARE
SUPERCRITICAL

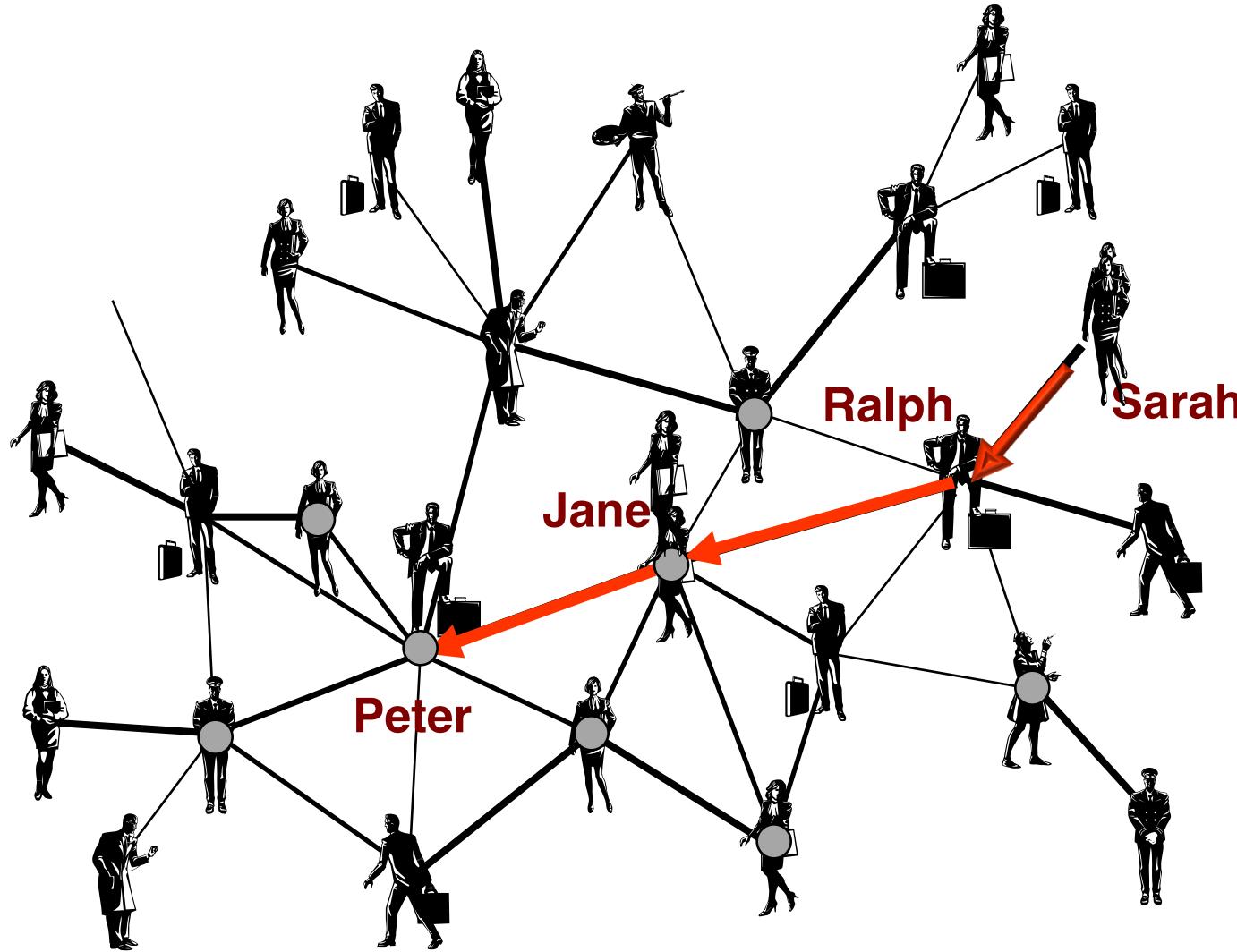


Network	N	L	$\langle k \rangle$	$\ln N$
Internet	192,244	609,066	6.34	12.17
Power Grid	4,941	6,594	2.67	8.51
Science Collaboration	23,133	186,936	8.08	10.04
Actor Network	212,250	3,054,278	28.78	12.27
Yeast Protein Interactions	2,018	2,930	2.90	7.61

SMALL WORLDS

SIX DEGREES

small worlds



Frigyes Karinthy, 1929
Stanley Milgram, 1967



Frigyes Karinthy (1887-1938)
Hungarian Writer

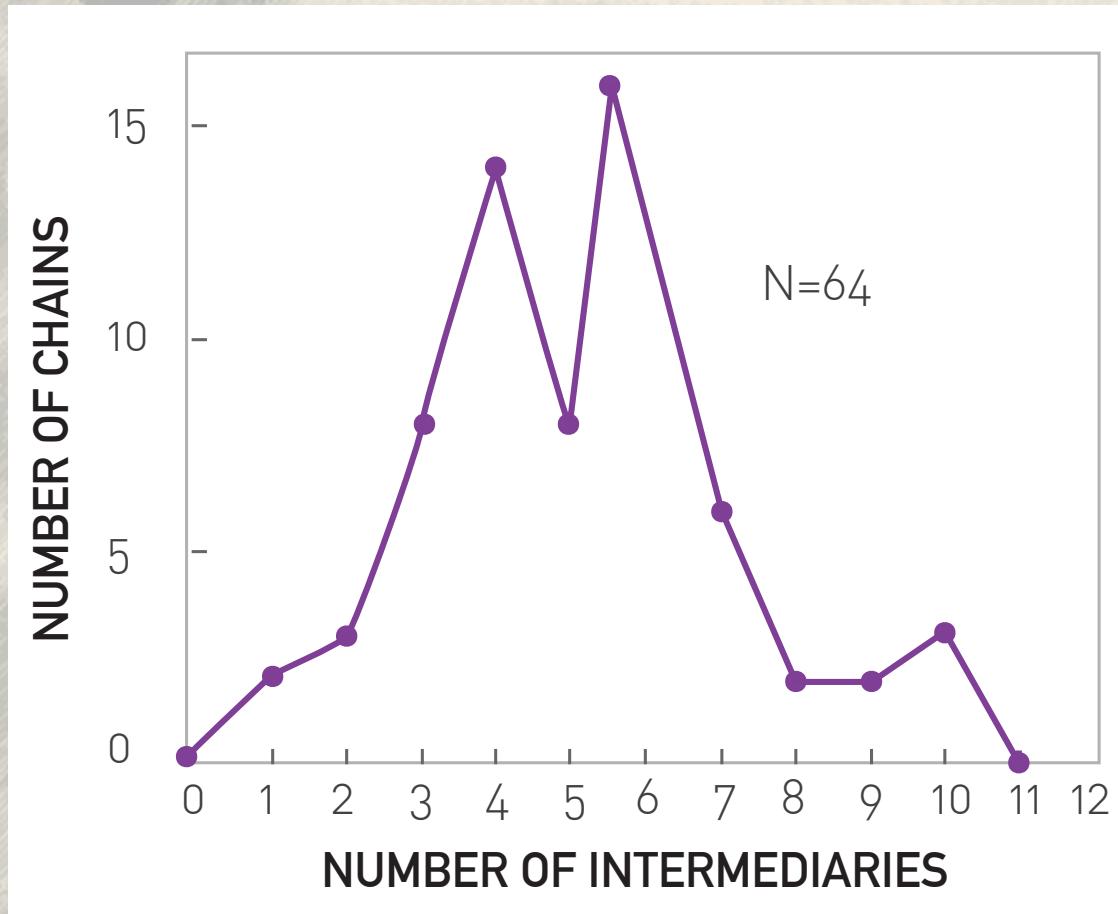
1929: *Minden másképpen van* (Everything is Different)
Láncszemek (Chains)

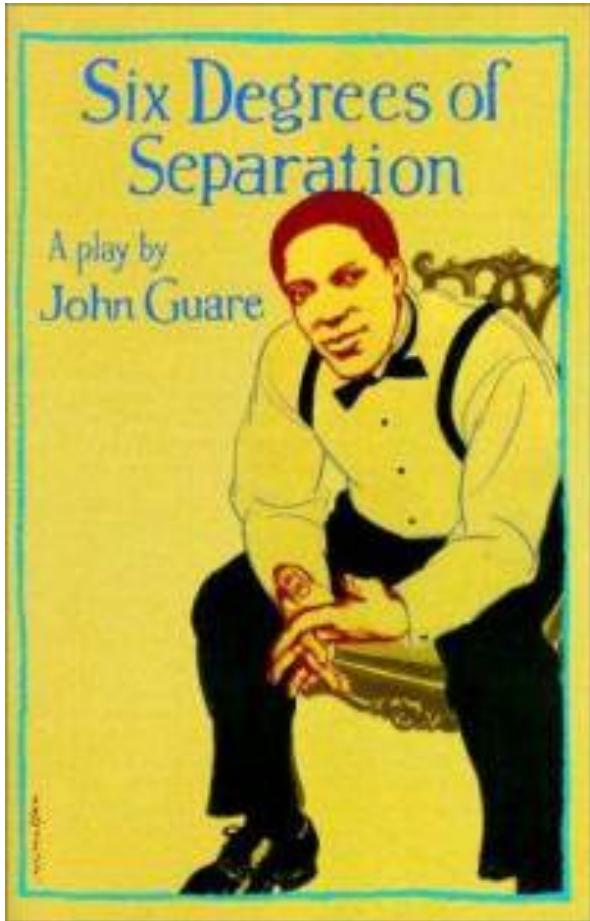
"Look, Selma Lagerlöf just won the Nobel Prize for Literature, thus she is bound to know King Gustav of Sweden, after all he is the one who handed her the Prize, as required by tradition. King Gustav, to be sure, is a passionate tennis player, who always participates in international tournaments. He is known to have played Mr. Kehrling, whom he must therefore know for sure, and as it happens I myself know Mr. Kehrling quite well."

"The worker knows the manager in the shop, who knows Ford; Ford is on friendly terms with the general director of Hearst Publications, who last year became good friends with Arpad Pasztor, someone I not only know, but to the best of my knowledge a good friend of mine. So I could easily ask him to send a telegram via the general director telling Ford that he should talk to the manager and have the worker in the shop quickly hammer together a car for me, as I happen to need one."

HOW TO TAKE PART IN THIS STUDY

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.





"Everybody on this planet is separated by only six other people. Six degrees of separation. Between us and everybody else on this planet. The president of the United States. A gondolier in Venice.... It's not just the big names. It's anyone. A native in a rain forest. A Tierra del Fuegan. An Eskimo. I am bound to everyone on this planet by a trail of six people. It's a profound thought. How every person is a new door, opening up into other worlds."

WWW: 19 DEGREES OF SEPARATION

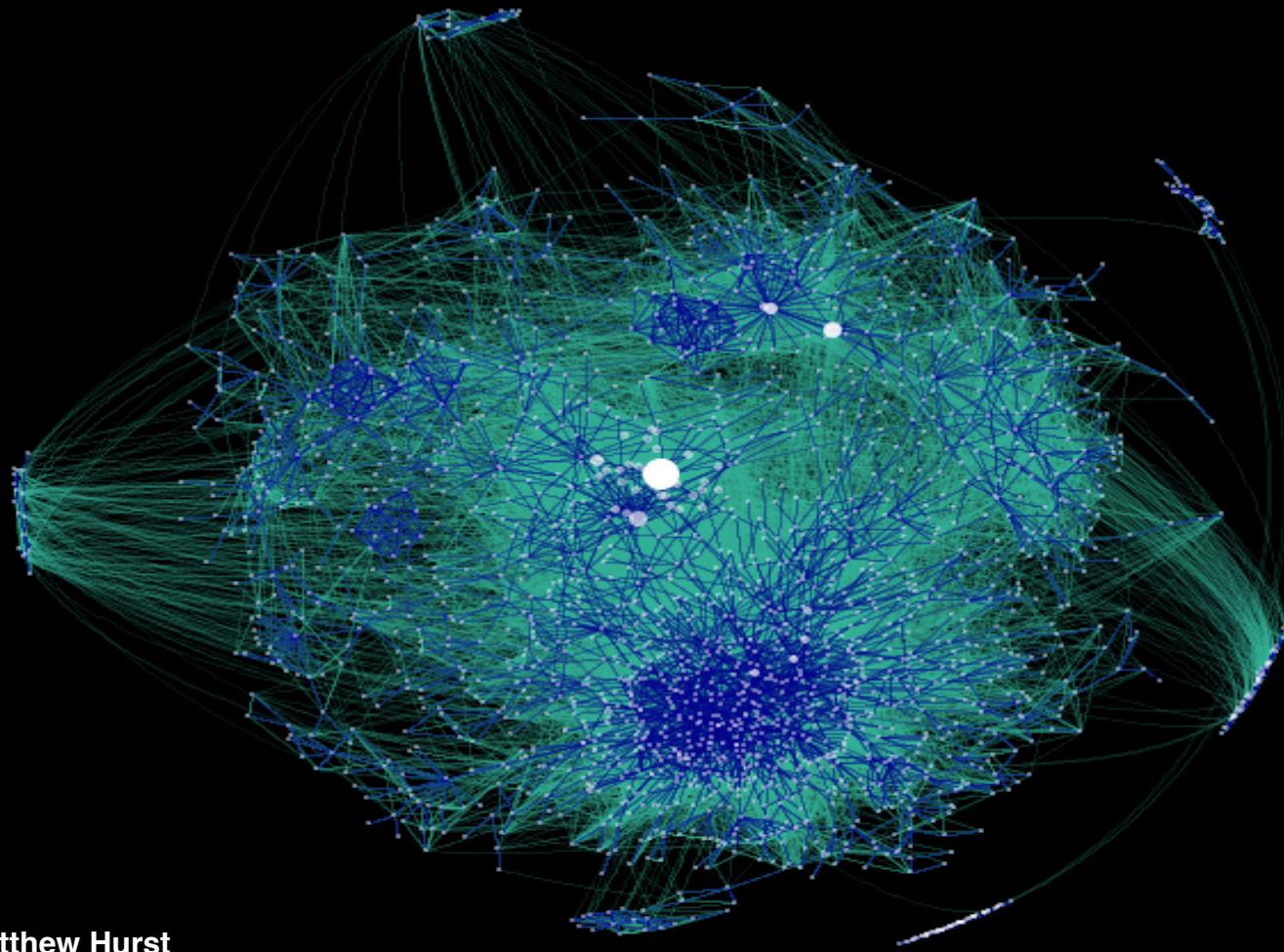
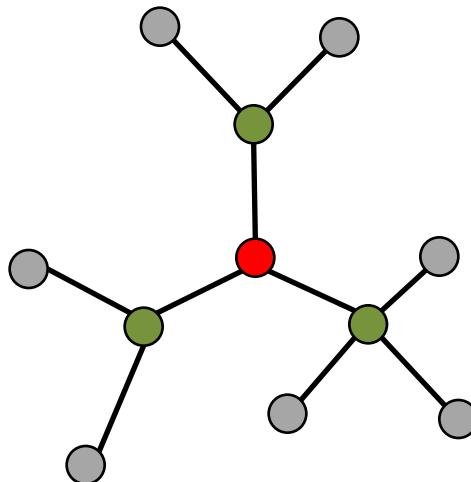


Image by **Matthew Hurst**
Blogosphere

Network Science: Random Graphs

DISTANCES IN RANDOM GRAPHS

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



$\langle k \rangle$ nodes at distance one ($d=1$).

$\langle k \rangle^2$ nodes at distance two ($d=2$).

$\langle k \rangle^3$ nodes at distance three ($d =3$).

...

$\langle k \rangle^d$ nodes at distance d .

$$N = 1 + \langle k \rangle + \langle k \rangle^2 + \dots + \langle k \rangle^{d_{\max}} = \frac{\langle k \rangle^{d_{\max}+1} - 1}{\langle k \rangle - 1} \approx \langle k \rangle^{d_{\max}}$$

➡

$$d_{\max} = \frac{\log N}{\log \langle k \rangle}$$

DISTANCES IN RANDOM GRAPHS

$$d_{\max} = \frac{\log N}{\log \langle k \rangle}$$

In most networks this offers a better approximation to the average distance between two randomly chosen nodes, $\langle d \rangle$, than to d_{\max} .

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

We will call the *small world phenomena* the property that the average path length or the diameter depends logarithmically on the system size.

Hence, "small" means that $\langle d \rangle$ is proportional to $\log N$, rather than N .

The $1/\log \langle k \rangle$ term implies that denser the network, the smaller will be the distance between the nodes.

DISTANCES IN RANDOM GRAPHS

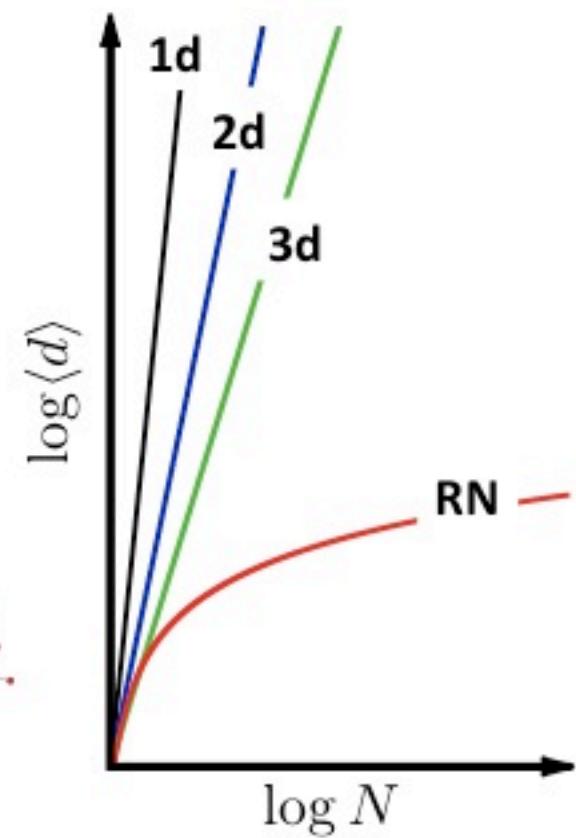
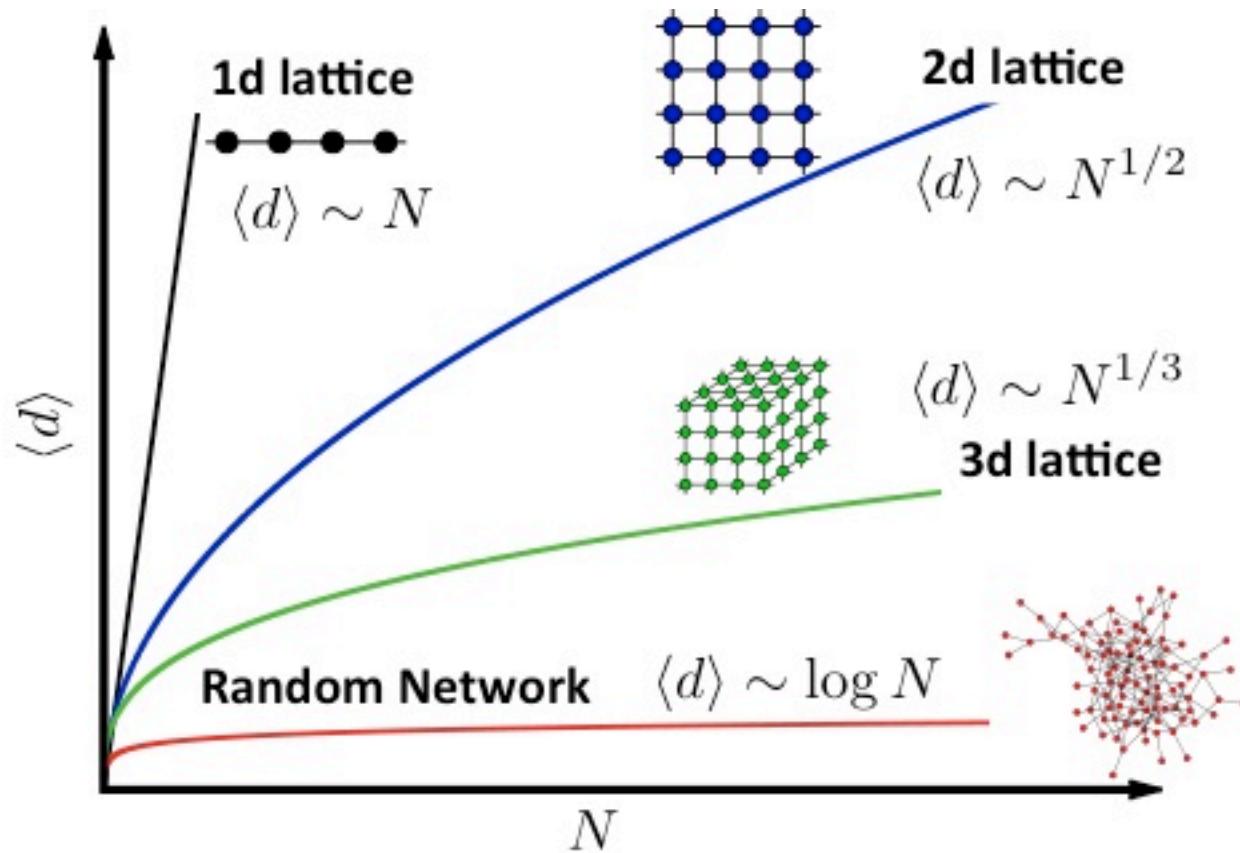
compare with real data

NETWORK	N	L	$\langle k \rangle$	$\langle d \rangle$	d_{max}	$\frac{\ln N}{\ln \langle k \rangle}$
Internet	192,244	609,066	6.34	6.98	26	6.58
WWW	325,729	1,497,134	4.60	11.27	93	8.31
Power Grid	4,941	6,594	2.67	18.99	46	8.66
Mobile Phone Calls	36,595	91,826	2.51	11.72	39	11.42
Email	57,194	103,731	1.81	5.88	18	18.4
Science Collaboration	23,133	93,439	8.08	5.35	15	4.81
Actor Network	702,388	29,397,908	83.71	3.91	14	3.04
Citation Network	449,673	4,707,958	10.43	11.21	42	5.55
E. Coli Metabolism	1,039	5,802	5.58	2.98	8	4.04
Protein Interactions	2,018	2,930	2.90	5.61	14	7.14

Given the huge differences in scope, size, and average degree, the agreement is excellent.

Why are small worlds surprising?

Surprising compared to what?



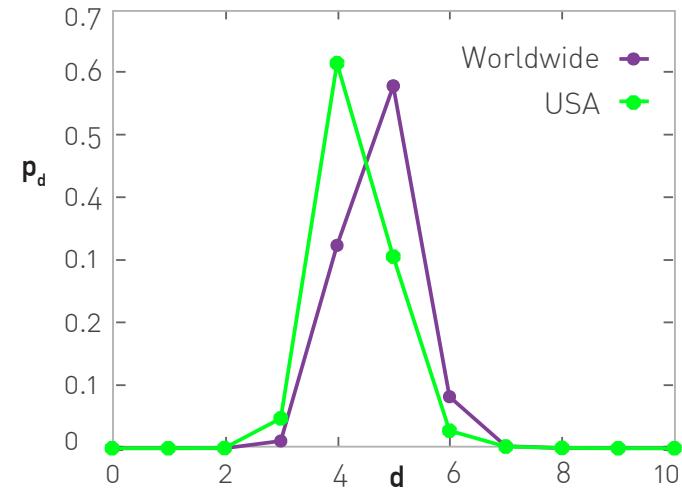
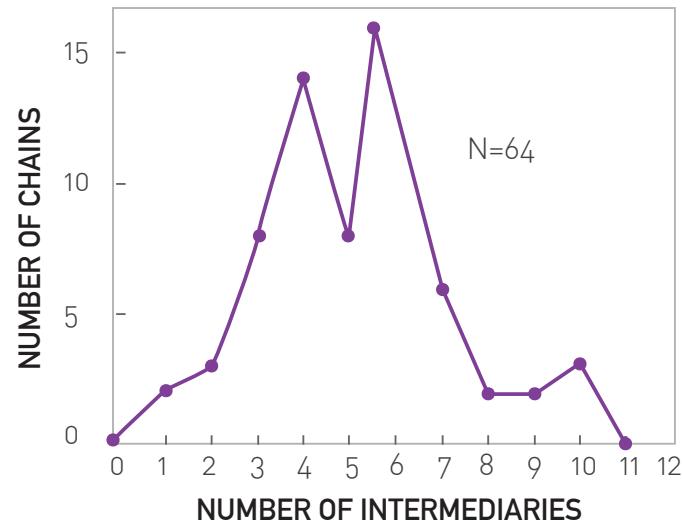
Three, Four or Six Degrees?

For the globe's social networks:

$$\langle k \rangle \approx 10^3$$

$N \approx 7 \times 10^9$ for the world's population.

$$\langle d \rangle = \frac{\ln(N)}{\ln \langle k \rangle} = 3.28$$



"The worker knows the manager in the shop, who knows Ford; Ford is on friendly terms with the general director of Hearst Publications, who last year became good friends with Árpád Pásztor, someone I not only know, but to the best of my knowledge a good friend of mine."

Karinthy, 1929



MILESTONES

PUBLICATION DATE

1929

1935

1940

WWII

1945

1950

1958

1960

1967

1970

1978

1980

1985

1991

XXI

1998

2000

2005

2011



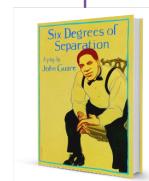
Frigyes Karinthy (1887-1938)
Hungarian writer, journalist and playwright, the first to describe the small world property. In his short story entitled 'Láncszemek' (Chains) he links a worker in Ford's factory to himself [23, 24].

Manfred Kochen (1928-1989),
Ithiel de Sola Pool (1917-1984)
Scientific interest in small worlds started with a paper by political scientist Ithiel de Sola Pool and mathematician Manfred Kochen. Written in 1958 and published in 1978, their work addressed in mathematical detail the small world effect, predicting that most individuals can be connected via two to three acquaintances. Their paper inspired the experiments of Stanley Milgram.

Stanley Milgram (1933-1984)
American social psychologist who carried out the first experiment testing the small-world phenomena. (BOX 3.6).

"Everybody on this planet is separated by only six other people. Six degrees of separation. Between us and everybody else on this planet. The president of the United States. A gondolier in Venice. It's not just the big names. It's anyone. A native in a rain forest. A Tierra del Fuegan. An Eskimo. I am bound to everyone on this planet by a trail of six people. It's a profound thought. How every person is a new door, opening up into other worlds."

Guare, 1991



4-DEGREE OF SEPARATION

6-DEGREE OF SEPARATION



John Guare (1938)
The phrase 'six degrees of separation' was introduced by the playwright John Guare, who used it as the title of his Broadway play.

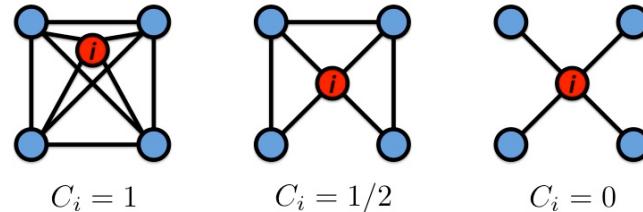
The Facebook Data Team measures the average distance between its users, finding "4 degrees" (BOX 3.6).

Duncan J. Watts (1971),
Steven Strogatz (1959)
A new wave of interest in small worlds followed the study of Watts and Strogatz, finding that the small world property applies to natural and technological networks as well.

CLUSTERING COEFFICIENT

CLUSTERING COEFFICIENT

$$C_i \equiv \frac{2 \langle L_i \rangle}{k_i(k_i - 1)}$$

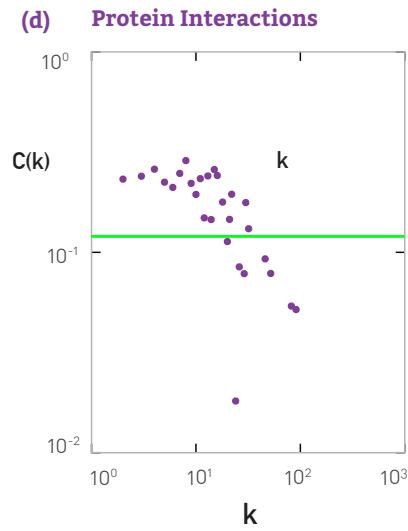
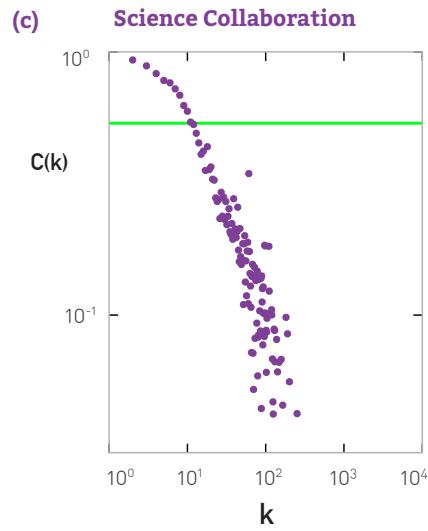
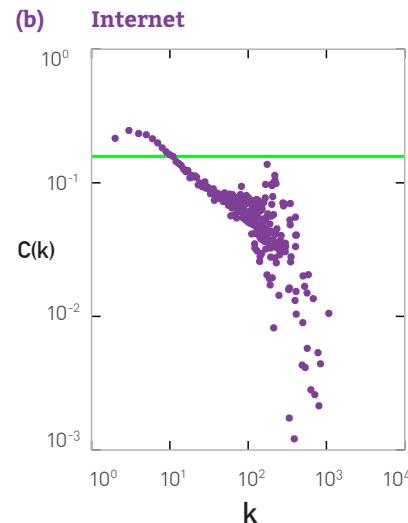
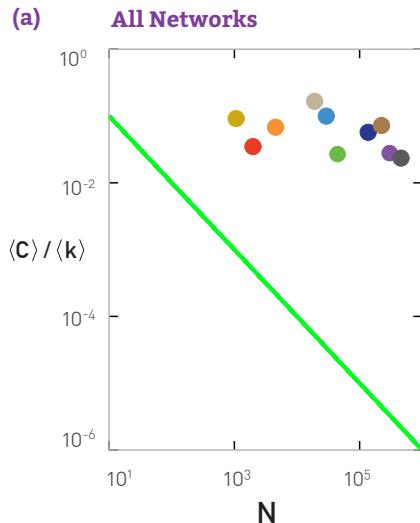


Since edges are independent and have the same probability p ,

$$\langle L_i \rangle \approx p \frac{k_i(k_i - 1)}{2} \quad \Rightarrow \quad C_i = \frac{2\langle L_i \rangle}{k_i(k_i - 1)} = p = \frac{\langle k \rangle}{N}.$$

- The clustering coefficient of random graphs is small.
- For fixed degree C decreases with the system size N .
- C is independent of a node's degree k .

CLUSTERING COEFFICIENT



$$C_i = \frac{2\langle L_i \rangle}{k_i(k_i - 1)} = p = \frac{\langle k \rangle}{N}.$$

C decreases with the system size N .

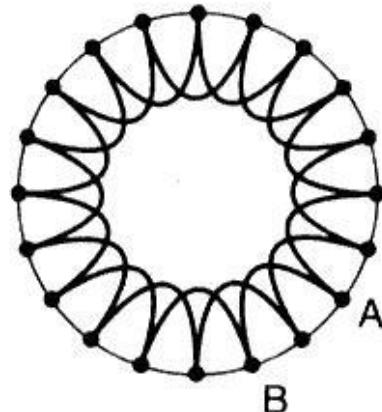
C is independent of a node's degree k .

Watts-Strogatz Model

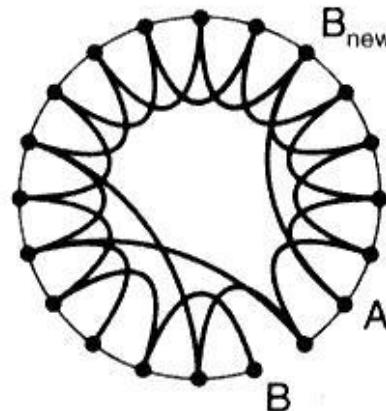
Reconciling two observations:

- **High clustering:** my friends' friends tend to be my friends
- **Short average paths**

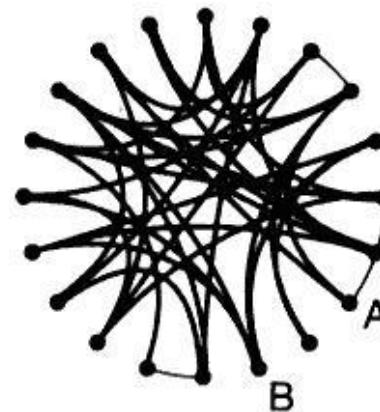
Regular



Small World



Random

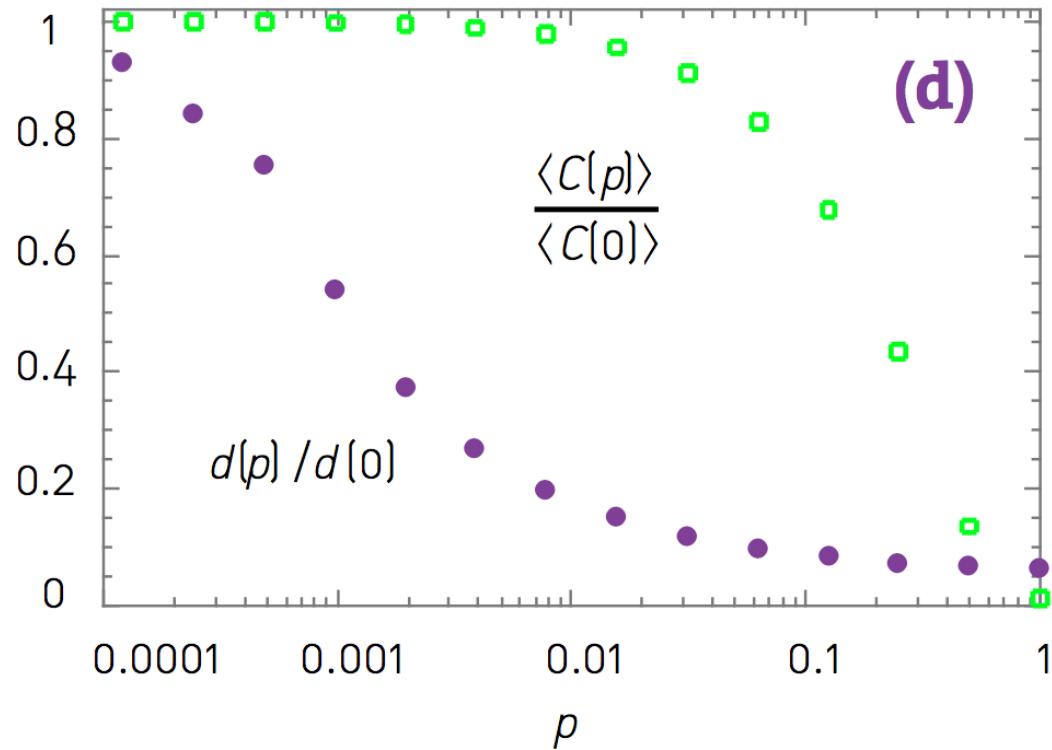


$p=0$

$p=1$

Increasing randomness

Watts-Strogatz Model



REAL NETWORKS ARE NOT
RANDOM

ARE REAL NETWORKS LIKE RANDOM GRAPHS?

As quantitative data about real networks became available, we can compare their topology with the predictions of random graph theory.

Note that once we have N and $\langle k \rangle$ for a random network, from it we can derive every measurable property. Indeed, we have:

Average path length:

$$\langle l_{rand} \rangle \approx \frac{\log N}{\log \langle k \rangle}$$

Clustering Coefficient:

$$C_i = \frac{2\langle L_i \rangle}{k_i(k_i - 1)} = p = \frac{\langle k \rangle}{N}.$$

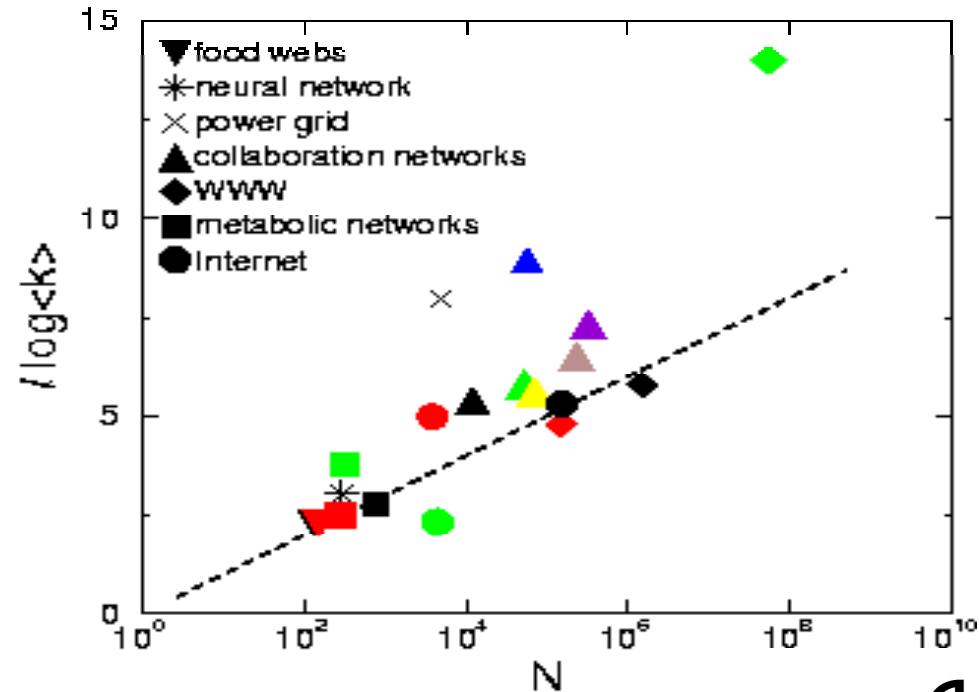
Degree Distribution:

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

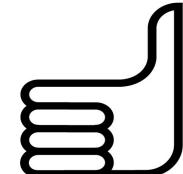
PATH LENGTHS IN REAL NETWORKS

Prediction:

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



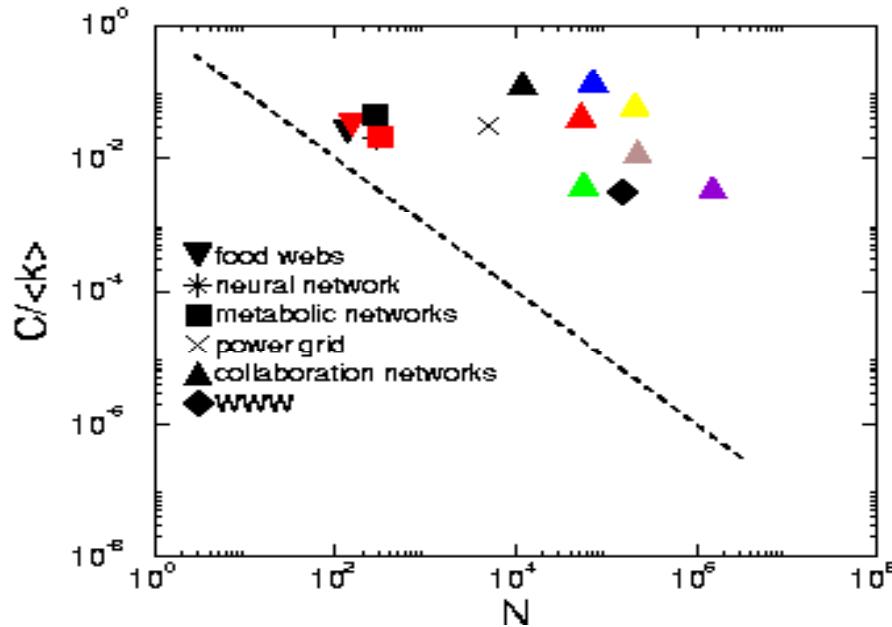
Real networks have short distances
like random graphs.



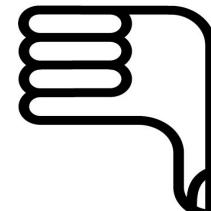
CLUSTERING COEFFICIENT

Prediction:

$$C_i = \frac{2\langle L_i \rangle}{k_i(k_i - 1)} = p = \frac{\langle k \rangle}{N}.$$



C_{rand} underestimates with orders of magnitudes the clustering coefficient of real networks.



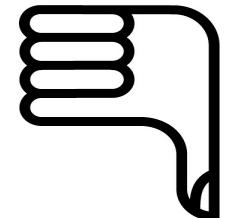
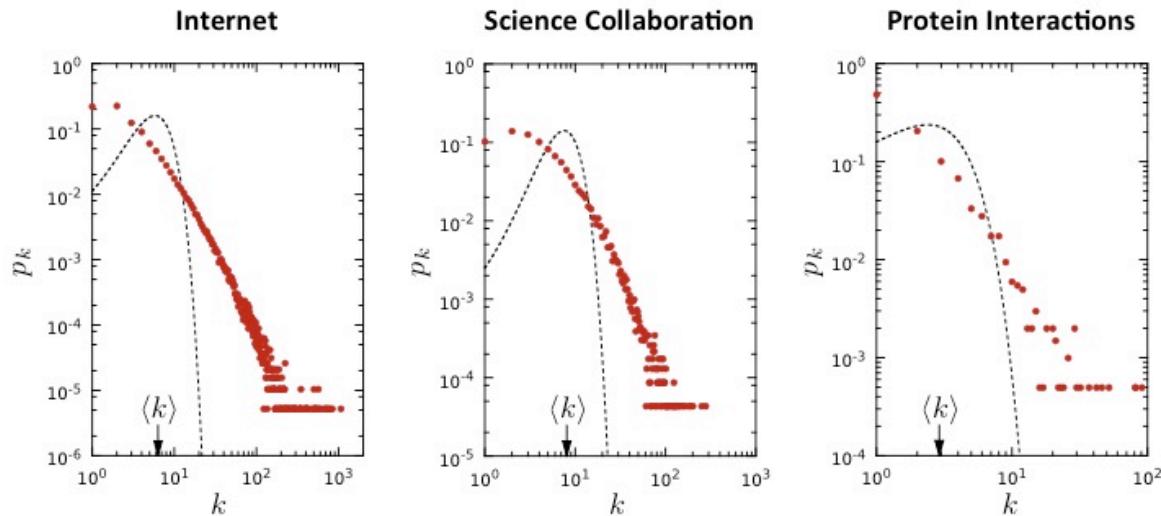
THE DEGREE DISTRIBUTION

Prediction:

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

Data:

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



ARE REAL NETWORKS LIKE RANDOM GRAPHS?

As quantitative data about real networks became available, we can compare their topology with the predictions of random graph theory.

Note that once we have N and $\langle k \rangle$ for a random network, from it we can derive every measurable property. Indeed, we have:

Average path length:

$$\langle l_{rand} \rangle \approx \frac{\log N}{\log \langle k \rangle}$$



Clustering Coefficient:

$$C_i = \frac{2\langle L_i \rangle}{k_i(k_i - 1)} = p = \frac{\langle k \rangle}{N}.$$



Degree Distribution:

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



IS THE RANDOM GRAPH MODEL RELEVANT TO REAL SYSTEMS?

(B) Most important: we need to ask ourselves, are real networks random?

The answer is simply: NO

THERE IS NO NETWORK IN NATURE THAT WE KNOW OF THAT
WOULD BE DESCRIBED BY THE RANDOM NETWORK MODEL.

IF IT IS WRONG AND IRRELEVANT, WHY DID WE DEVOT TO IT A FULL CLASS?

It is the reference model for the rest of the class.

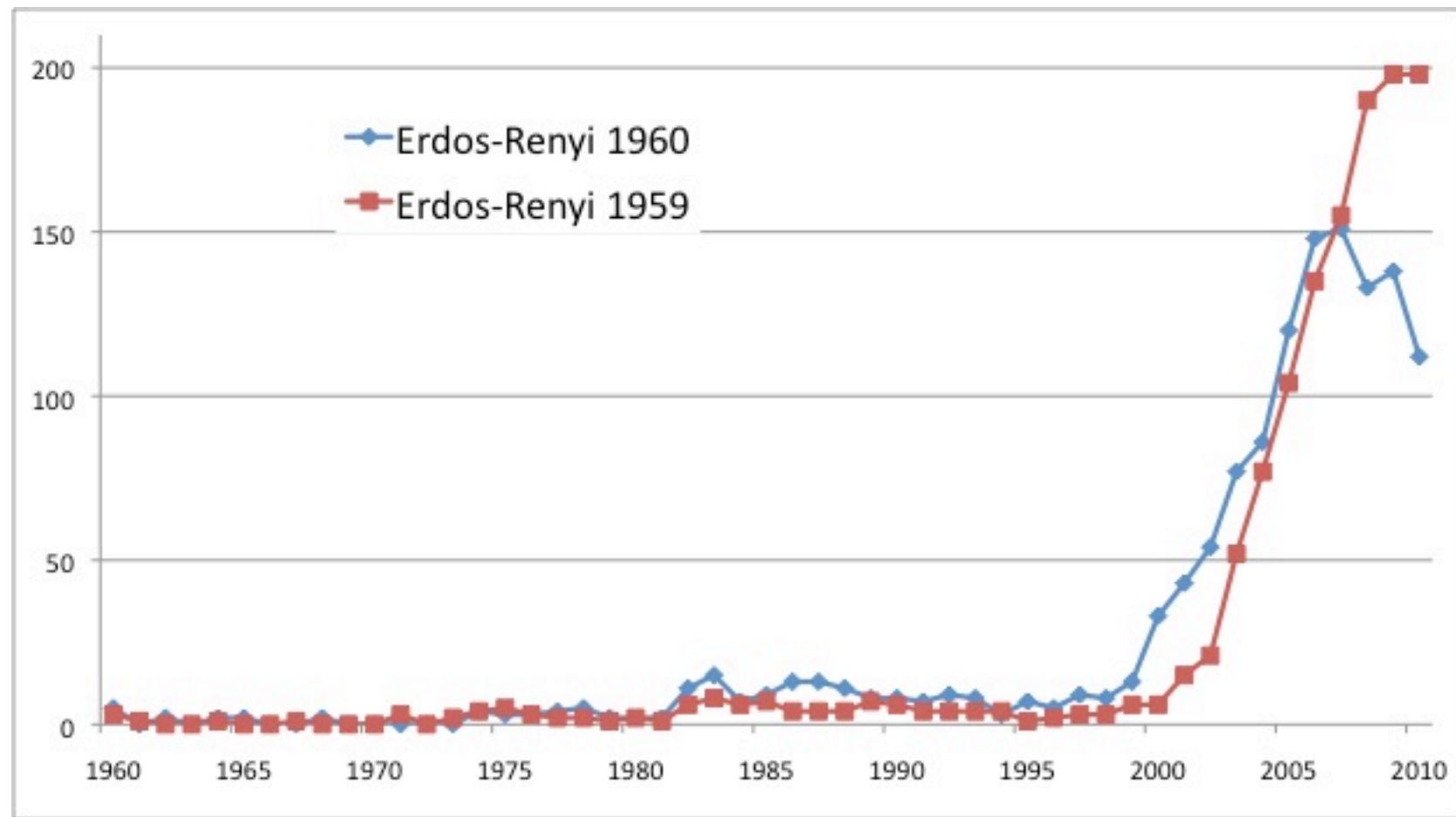
It will help us calculate many quantities, that can then be compared to the real data, understanding to what degree is a particular property the result of some random process.

Patterns in real networks that are shared by a large number of real networks, yet which deviate from the predictions of the random network model.

In order to identify these, we need to understand how would a particular property look like if it is driven entirely by random processes.

WHILE WRONG AND IRRELEVANT, IT WILL TURN OUT TO BE EXTREMELY USEFUL!

Erdös-Rényi MODEL (1960)



HISTORICAL NOTE

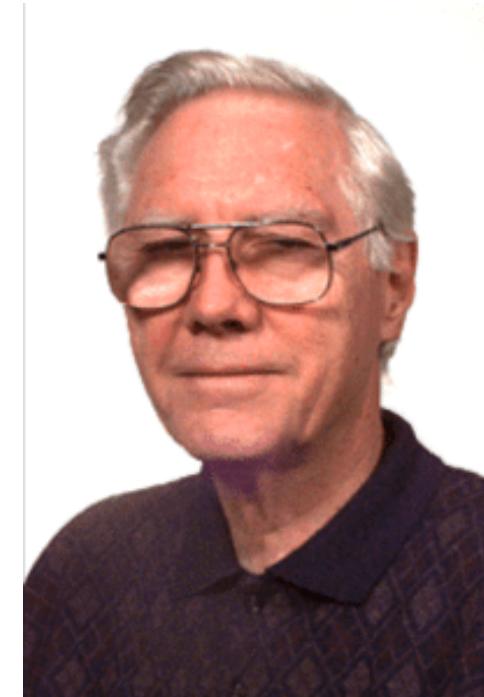


ANATOL RAPOPORT
1911- 2007

1951, Rapoport and Solomonoff:

- first systematic study of a random graph.
- demonstrates the phase transition.

→ natural systems: neural networks; the social networks of physical contacts (epidemics); genetics.

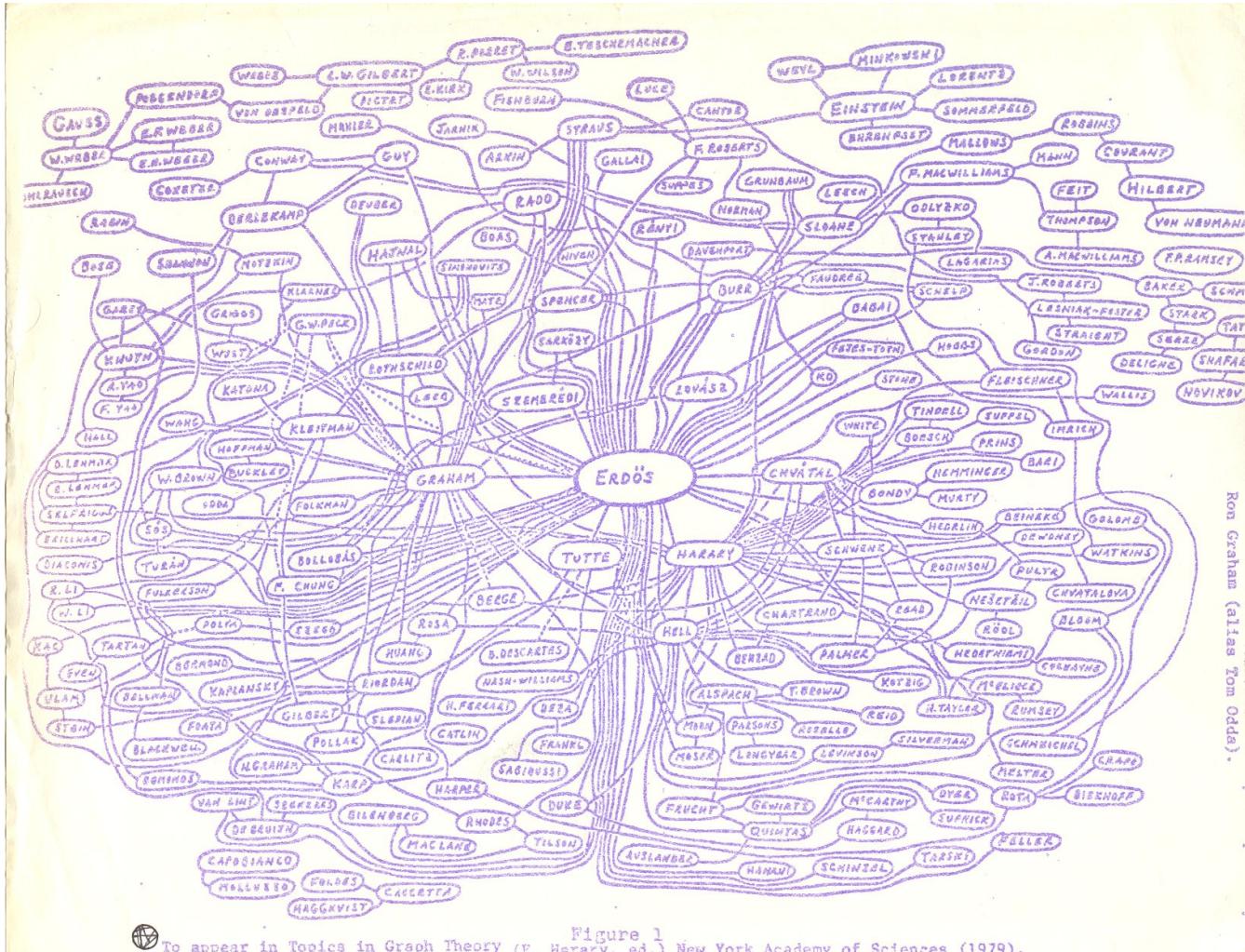


1959: $G(N,p)$

EDGAR N. GILBERT
(b.1923)

Why do we call it the Erdos-Renyi random model?

NETWORK DATA: SCIENCE COLLABORATION NETWORKS



Erdos:
1,400 papers
507 coauthors

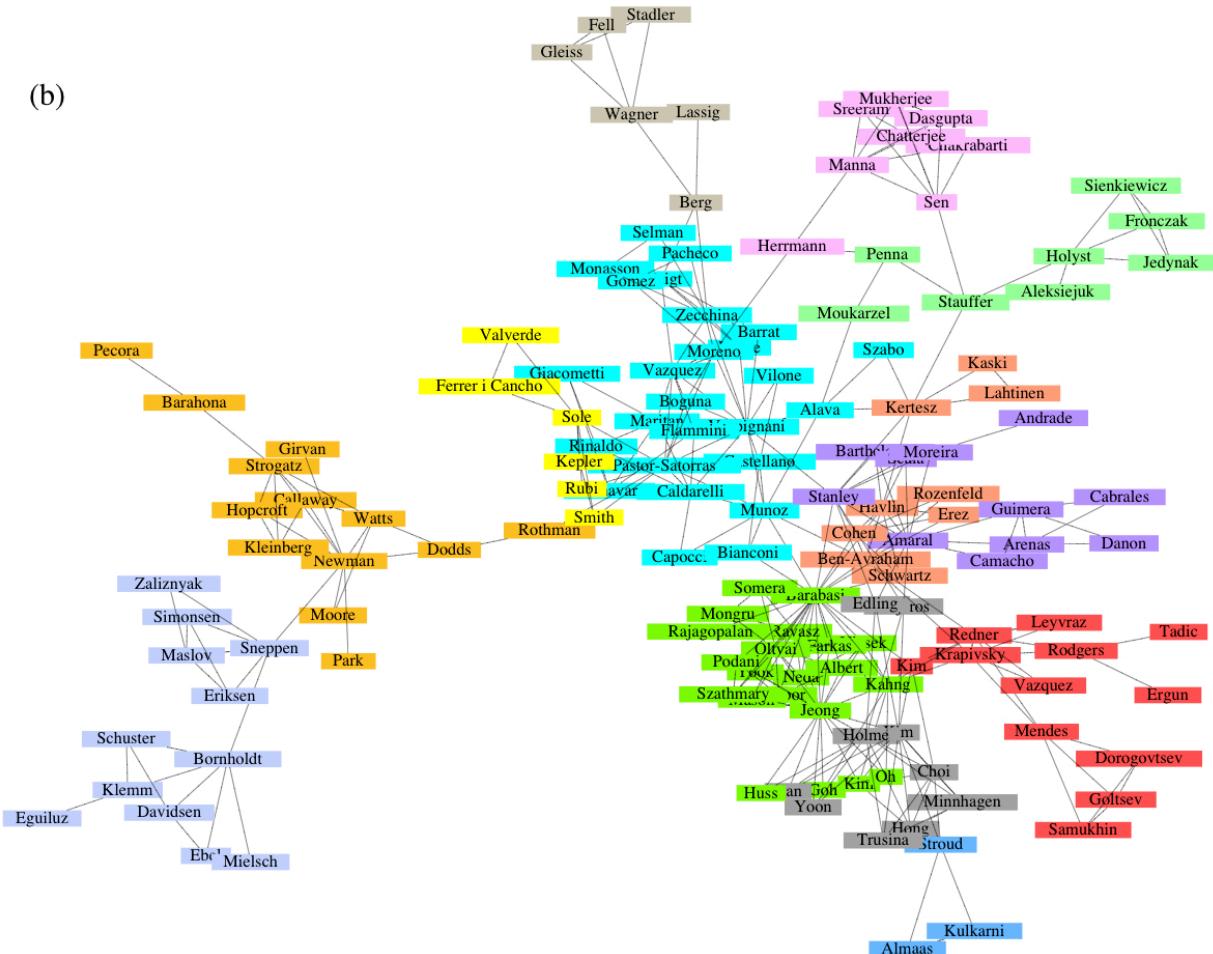
Einstein: EN=2
Paul Samuelson EN=5

ALB: EN: 3

Ron Graham (alias Tom Oada).

NETWORK DATA: SCIENCE COLLABORATION NETWORKS

(b)



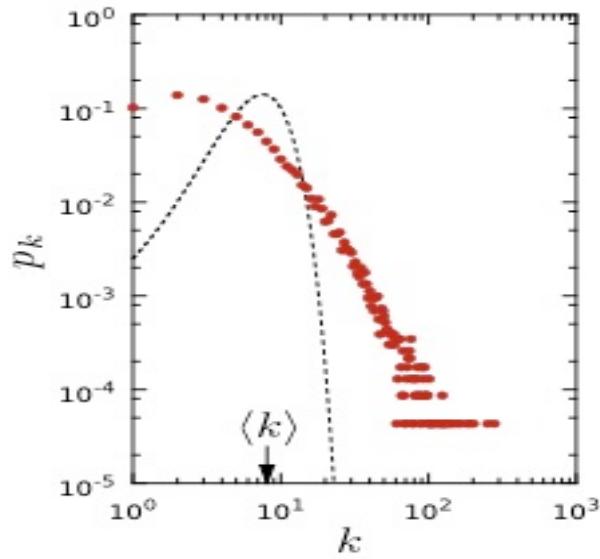
Collaboration Network:
Nodes: Scientists
Links: Joint publications

Physical Review:
1893 – 2009.

$N=449,673$
 $L=4,707,958$

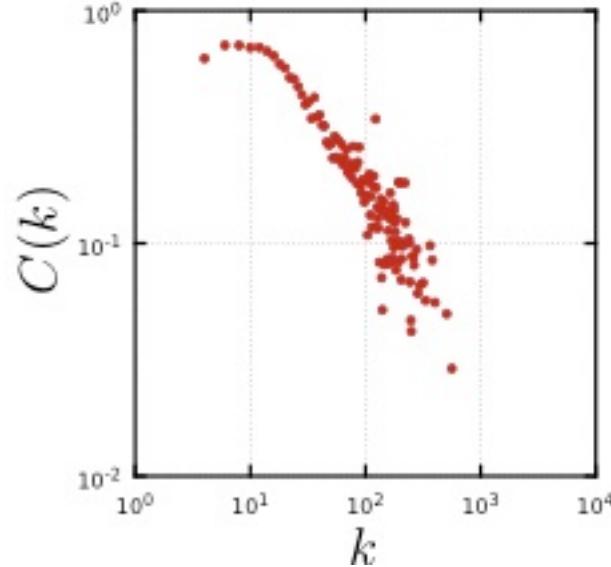
See also Stanford Large Network database
<http://snap.stanford.edu/data/#canets>.

Science Collaboration



Scale-free

Science Collaboration



Hierarchical

Hubs represent the most striking difference between a random and a scale-free network. Their emergence in many real systems raises several fundamental questions:

- Why does the random network model of Erdős and Rényi fail to reproduce the hubs and the power laws observed in many real networks?
- Why do so different systems as the WWW or the cell converge to a similar scale-free architecture?

GROWTH AND PREFERENTIAL ATTACHMENT

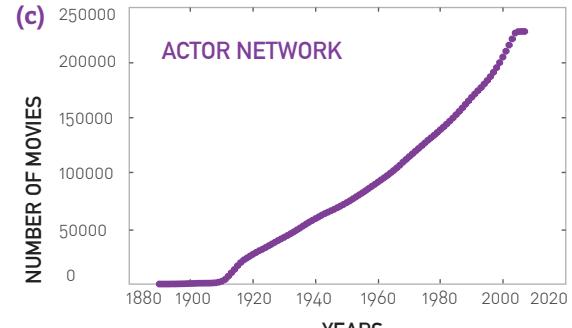
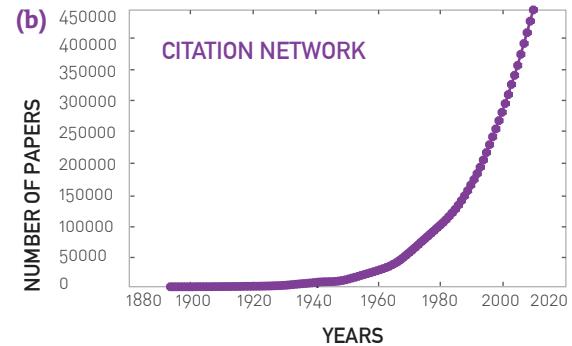
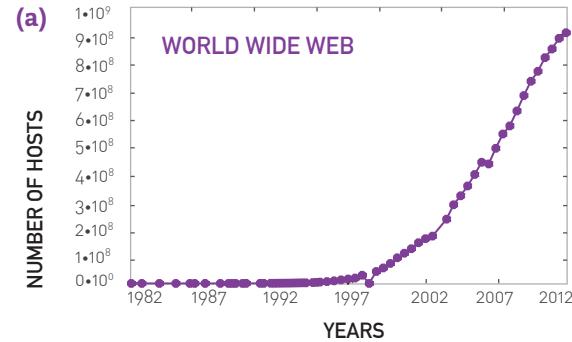
BA MODEL: Growth

ER model:

the number of nodes, N , is fixed (static models)

NETWORKS EXPAND THROUGH THE ADDITION OF NEW NODES

if we wish to model these networks, we cannot resort to a static model. Our modeling approach must instead acknowledge that networks are the product of a steady growth process.



BA MODEL: Preferential attachment

ER model: links are added randomly to the network

NEW NODES PREFER TO CONNECT TO THE MORE CONNECTED NODES

Growth and Preferential Attachment

The random network model differs from real networks in two important characteristics:

Growth: While the random network model assumes that the number of nodes is fixed (time invariant), real networks are the result of a growth process that continuously increases.

Preferential Attachment: While nodes in random networks randomly choose their interaction partner, in real networks new nodes prefer to link to the more connected nodes.

THE BARABÁSI-ALBERT MODEL

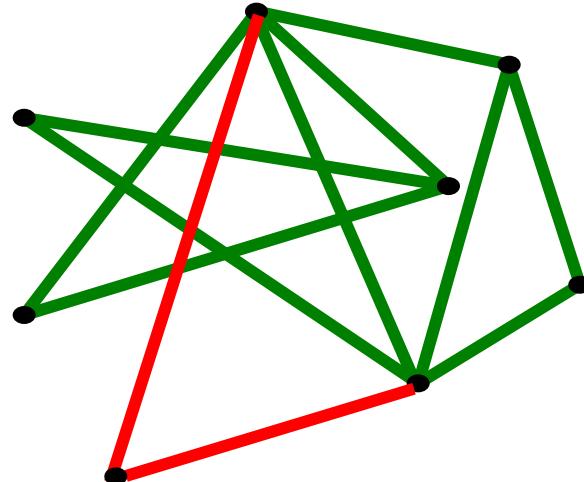
Origin of SF networks: Growth and preferential attachment

(1) Networks continuously expand by the addition of new nodes

WWW : addition of new documents

(2) New nodes prefer to link to highly connected nodes.

WWW : linking to well known sites



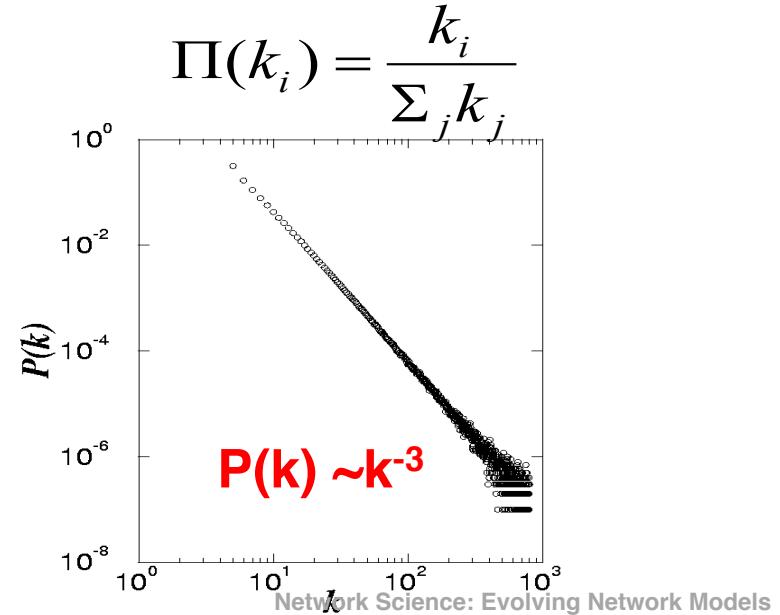
Barabási & Albert, *Science* **286**, 509 (1999)

GROWTH:

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.

PREFERENTIAL ATTACHMENT:

the probability that a node connects to a node with k links is proportional to k .



Origin of SF networks: Growth and preferential attachment

GROWTH:

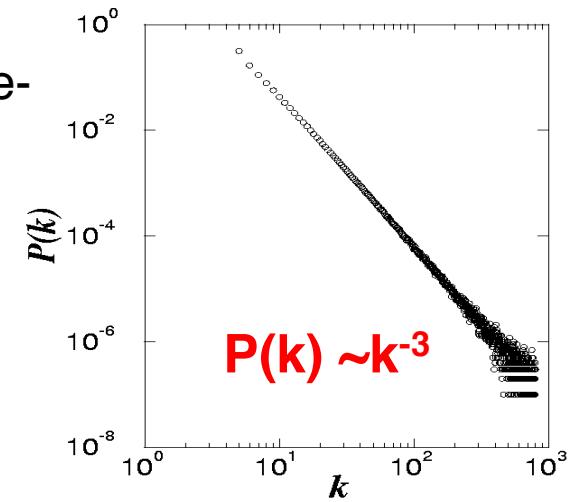
At each time step we add a new node with m ($\leq m_0$) links that connect the new node to m nodes already in the network.

PREFERENTIAL ATTACHMENT:

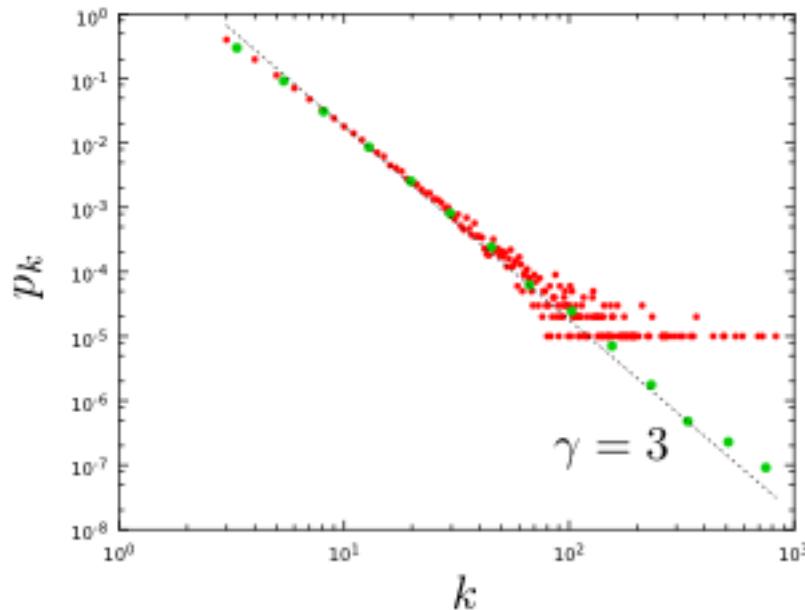
the probability that a node connects to a node with k links is proportional to k . $\Pi(k_i) = \frac{k_i}{\sum_j k_j}$

Preferential attachment is a probabilistic mechanism: A new node is free to connect to *any* node in the network, whether it is a hub or has a single link. Equation (5.1) implies, however, that if a new node has a choice between a degree-two and a degree-four node, it is twice as likely that it connects to the degree-four node.

After t timesteps the Barabási-Albert model generates a network with $N = t + m_0$ nodes and $m_0 + mt$ links. As Figure 5.4 shows, the obtained network has a power-law degree distribution with degree exponent $\gamma=3$.



Origin of SF networks: Growth and preferential attachment



Degree distribution of a network generated by the Barabási-Albert model. The plot shows for a single network of size $N=100,000$ and $m=3$. It shows both the linearly-binned (red symbols) as well as the log-binned version (green symbols). The straight line is added to guide the eye and has slope -3, corresponding to the resulting network's degree exponent.

All nodes follow the same growth law

$$\frac{\partial k_i}{\partial t} \propto \Pi(k_i) = A \frac{k_i}{\sum_j k_j}$$

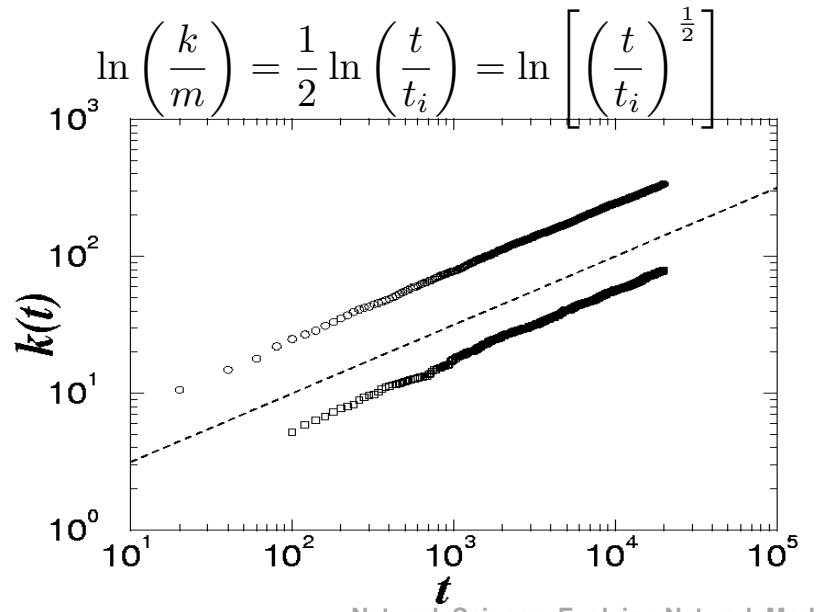
Use: $\sum_j k_j = 2mt$ During a unit time (time step): $\Delta k = m \rightarrow A = m$

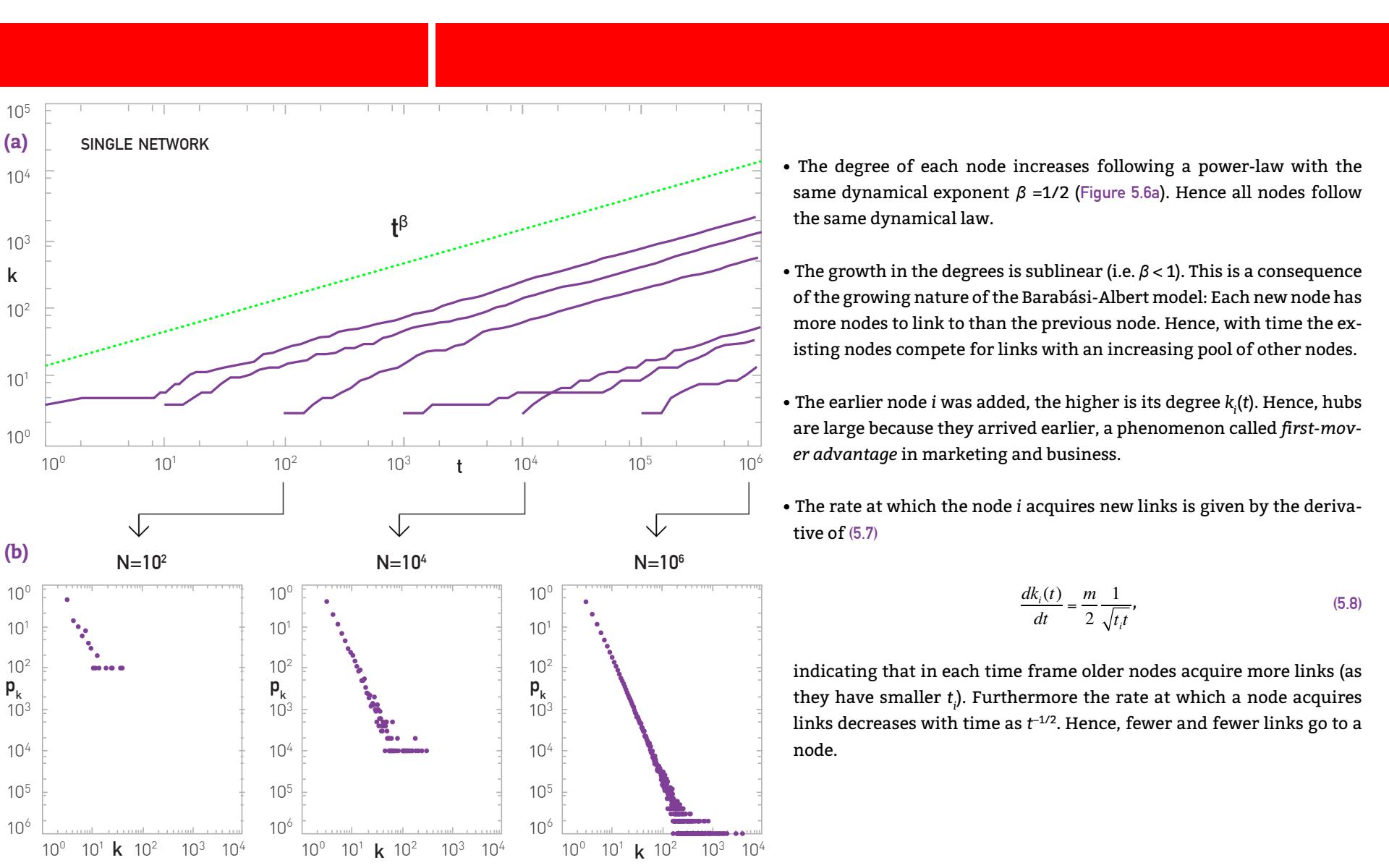
$$\frac{\partial k_i}{\partial t} = \frac{k_i}{2t} \quad \frac{\partial k_i}{k_i} = \frac{\partial t}{2t} \quad \int_m^k \frac{\partial k_i}{k_i} = \int_{t_i}^t \frac{\partial t}{2t}$$

$$k_i(t) = m \left(\frac{t}{t_i} \right)^\beta \quad \beta = \frac{1}{2}$$

β : dynamical exponent

A.-L.Barabási, R. Albert and H. Jeong, *Physica A* **272**, 173 (1999)

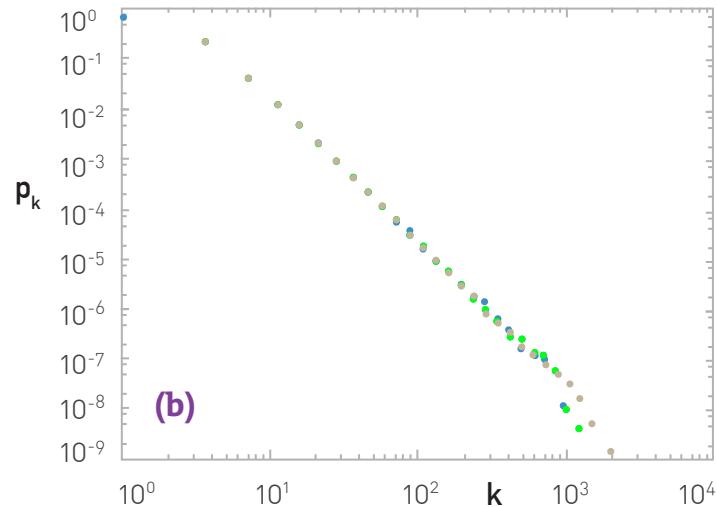
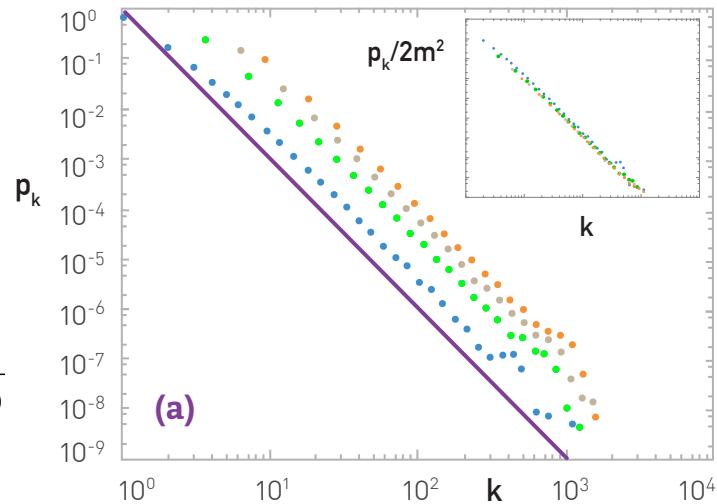




DEGREE DISTRIBUTION

NUMERICAL SIMULATION OF THE BA MODEL

$$P(k) = \frac{2m(m+1)}{k(k+1)(k+2)}$$



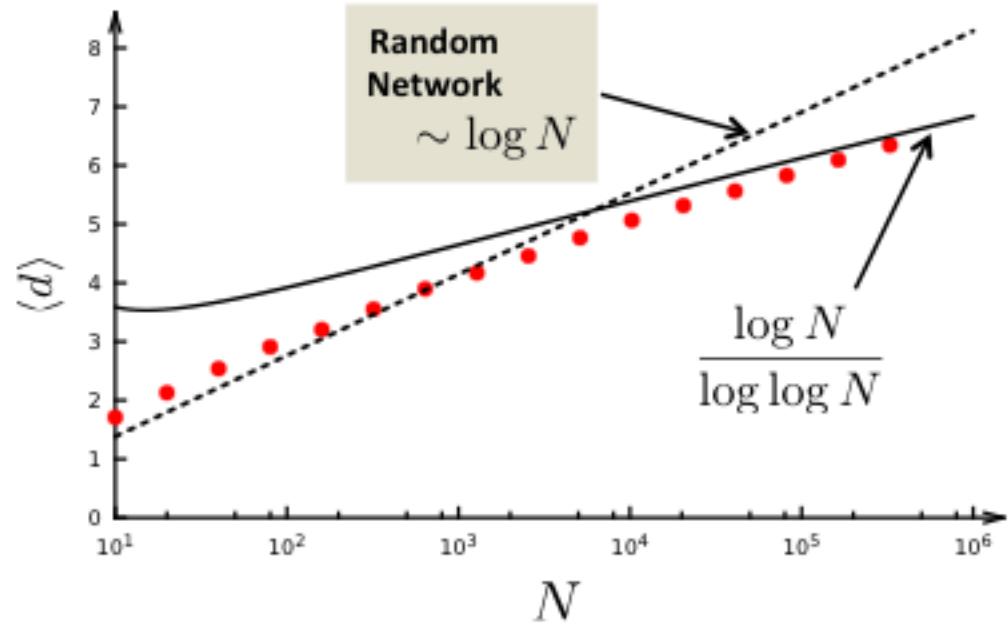
(a) We generated networks with $N=100,000$ and $m_0=m=1$ (blue), 3 (green), 5 (grey), and 7 (orange). The fact that the curves are parallel to each other indicates that γ is independent of m and m_0 . The slope of the purple line is -3 , corresponding to the predicted degree exponent $\gamma=3$. Inset: (5.11) predicts $p_k \sim 2m^2$, hence $p_k/2m^2$ should be independent of m . Indeed, by plotting $p_k/2m^2$ vs. k , the data points shown in the main plot collapse into a single curve.

(b) The Barabási-Albert model predicts that p_k is independent of N . To test this we plot p_k for $N = 50,000$ (blue), 100,000 (green), and 200,000 (grey), with $m_0=m=3$. The obtained p_k are practically indistinguishable, indicating that the degree distribution is stationary, i.e. independent of time and system size.

DIAMETER AND CLUSTERING COEFFICIENT

DIAMETER

$$D \sim \frac{\log N}{\log \log N}$$



Bollobas, Riordan, 2002

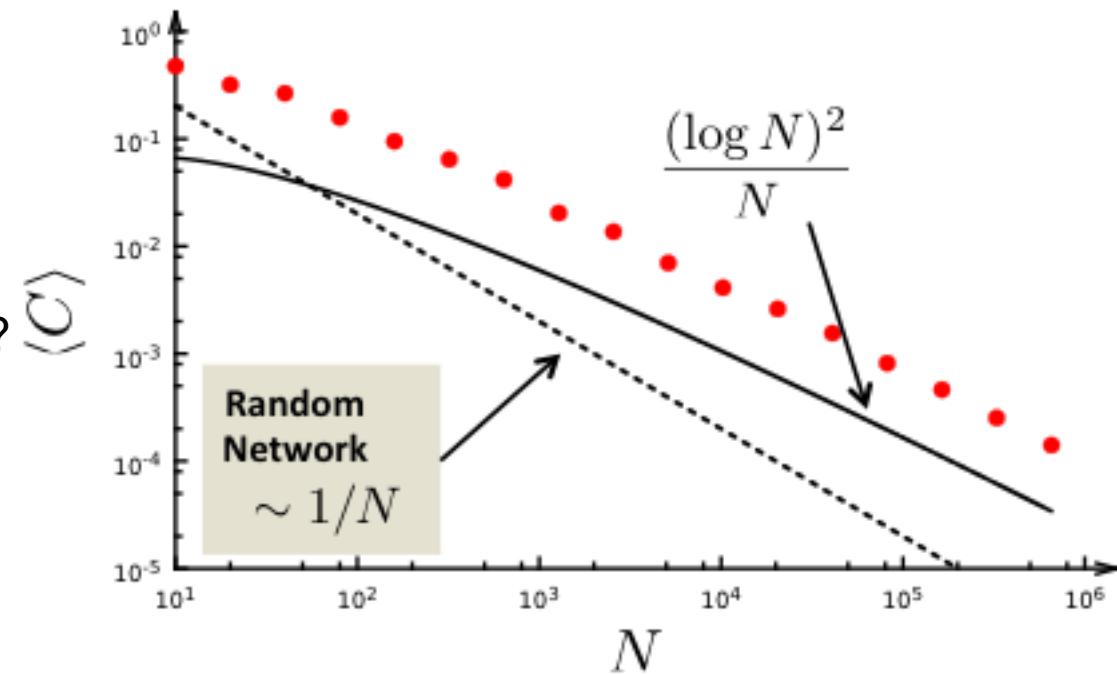
CLUSTERING COEFFICIENT

Reminder: for a random graph we have:

$$C_{rand} = \frac{\langle k \rangle}{N} \sim N^{-1}$$

What is the functional form of $C(N)$? 

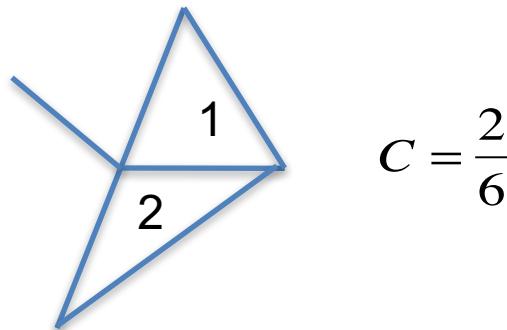
$$C_l = \frac{m}{4} \frac{(\ln N)^2}{N},$$



Konstantin Klemm, Victor M. Eguiluz,
Growing scale-free networks with small-world behavior,
Phys. Rev. E 65, 057102 (2002), cond-mat/0107607

CLUSTERING COEFFICIENT OF THE BA MODEL

$$C = \frac{\frac{Nr(\triangle)}{k(k-1)}}{2}$$



$$C = \frac{2}{6}$$

Denote the probability to have a link between node i and j with $P(i,j)$
The probability that three nodes i,j,l form a triangle is $P(i,j)P(i,l)P(j,l)$

The expected number of triangles in which a node i with degree k_i participates is thus:

$$Nr_i(\triangle) = \sum_{i=1}^N di \sum_{j=1}^N dj P(i,j) P(i,l) P(j,l)$$

We need to calculate $P(i,j)$.

CLUSTERING COEFFICIENT OF THE BA MODEL

Calculate $P(i,j)$.

Node j arrives at time $t_j=j$ and the probability that it will link to node i with degree k_i already in the network is determined by preferential attachment:

$$P(i,j) = m \prod_{l=1}^j k_l(j) = m \frac{k_i(j)}{\sum_{l=1}^j k_l} = m \frac{k_i(j)}{2mj}$$

$$k_i(t) = m \left(\frac{t}{t_i} \right)^{1/2} = m \left(\frac{j}{i} \right)^{1/2}$$

Where we used that the arrival time of node j is $t_j=j$ and the arrival time of node i is $t_i=i$

$$P(i,j) = \frac{m}{2} (ij)^{-\frac{1}{2}}$$

$$Nr_l(\triangleleft) = \int_{i=1}^N di \int_{j=1}^N dj P(i,j) P(i,l) P(j,l) = \frac{m^3}{8} \int_{i=1}^N di \int_{j=1}^N dj (ij)^{-\frac{1}{2}} (il)^{-\frac{1}{2}} (jl)^{-\frac{1}{2}} = \frac{m^3}{8l} \int_{i=1}^N \frac{di}{i} \int_{j=1}^N \frac{dj}{j} = \frac{m^3}{8l} (\ln N)^2$$

$$C = \frac{\frac{m^3}{8l} (\ln N)^2}{k_l(k_l - 1)/2}$$

$$k_l(t) = m \left(\frac{N}{l} \right)^{1/2}$$

Which is the degree of node l at current time, at time $t=N$

Let us approximate:

$$k_l(k_l - 1) \approx k_l^2 = m^2 \frac{N}{l}$$

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

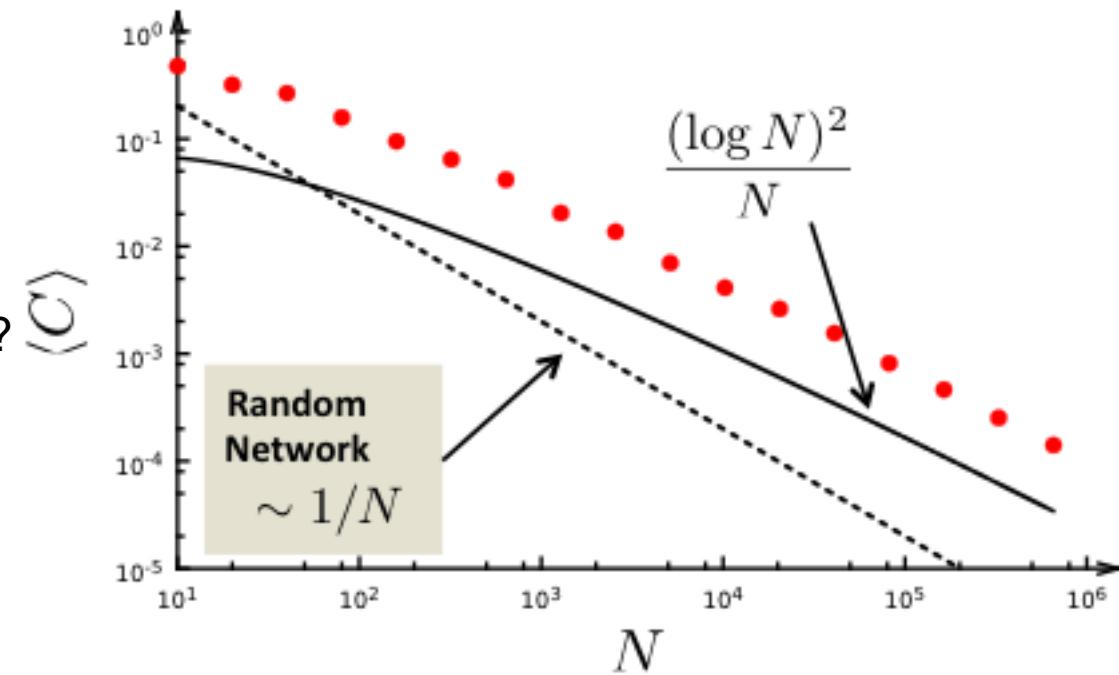
CLUSTERING COEFFICIENT

Reminder: for a random graph we have:

$$C_{rand} = \frac{\langle k \rangle}{N} \sim N^{-1}$$

What is the functional form of $C(N)$? 

$$C = \frac{m}{8} \frac{(\ln N)^2}{N}$$



Konstantin Klemm, Victor M. Eguiluz,
Growing scale-free networks with small-world behavior,
Phys. Rev. E 65, 057102 (2002), cond-mat/0107607

SUMMARY

Number of Nodes

$$N = t$$

Number of Links

$$N = mt$$

Average Degree

$$\langle k \rangle = 2m$$

Degree Dynamics

$$k_i(t) = m (t/t_i)^\beta$$

Dynamical Exponent

$$\beta = 1/2$$

Degree Distribution

$$p_k \sim k^{-\gamma}$$

Degree Exponent

$$\gamma = 3$$

Average Distance

$$\langle d \rangle \sim \log N / \log \log N$$

Clustering Coefficient

$$\langle C \rangle \sim (\ln N)^2 / N$$

The network grows, but the degree distribution is stationary.

SUMMARY

Number of Nodes

$$N = t$$

Number of Links

$$N = mt$$

Average Degree

$$\langle k \rangle = 2m$$

Degree Dynamics

$$k_i(t) = m (t/t_i)^\beta$$

Dynamical Exponent

$$\beta = 1/2$$

Degree Distribution

$$p_k \sim k^{-\gamma}$$

Degree Exponent

$$\gamma = 3$$

Average Distance

$$\langle d \rangle \sim \log N / \log \log N$$

Clustering Coefficient

$$\langle C \rangle \sim (\ln N)^2 / N$$

Consequently, the modeling philosophy behind the model is simple: *to understand the topology of a complex system, we need to describe how it came into being.*

The network grows, but the degree distribution is stationary.

SUMMARY

Number of Nodes

$$N = t$$

Number of Links

$$N = mt$$

Average Degree

$$\langle k \rangle = 2m$$

Degree Dynamics

$$k_i(t) = m (t/t_i)^\beta$$

Dynamical Exponent

$$\beta = 1/2$$

Degree Distribution

$$p_k \sim k^{-\gamma}$$

Degree Exponent

$$\gamma = 3$$

Average Distance

$$\langle d \rangle \sim \log N / \log \log N$$

Clustering Coefficient

$$\langle C \rangle \sim (\ln N)^2 / N$$

- The model predicts $\gamma=3$ while the degree exponent of real networks varies between 2 and 5 (Table 4.2).

- Many networks, like the WWW or citation networks, are directed, while the model generates undirected networks.

- Many processes observed in networks, from linking to already existing nodes to the disappearance of links and nodes, are absent from the model.

- The model does not allow us to distinguish between nodes based on some intrinsic characteristics, like the novelty of a research paper or the utility of a webpage.

- While the Barabási-Albert model is occasionally used as a model of the Internet or the cell, in reality it is not designed to capture the details of any particular real network. It is a minimal, proof of principle model whose main purpose is to capture the basic mechanisms responsible for the emergence of the scale-free property. Therefore, if we want to understand the evolution of systems like the Internet, the cell or the WWW, we need to incorporate the important details that contribute to the time evolution of these systems, like the directed nature of the WWW, the possibility of internal links and node and link removal.