NETWORK SCIENCE

Random Networks

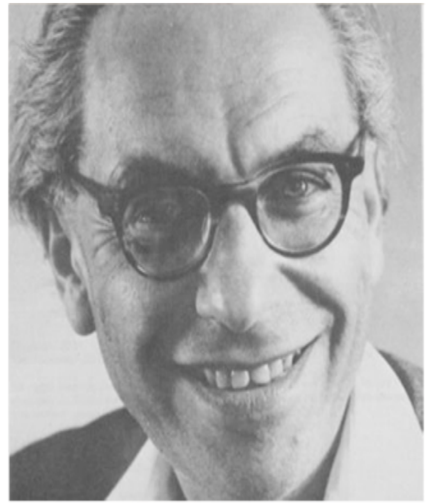
Prof. Marcello Pelillo
Ca' Foscari University of Venice

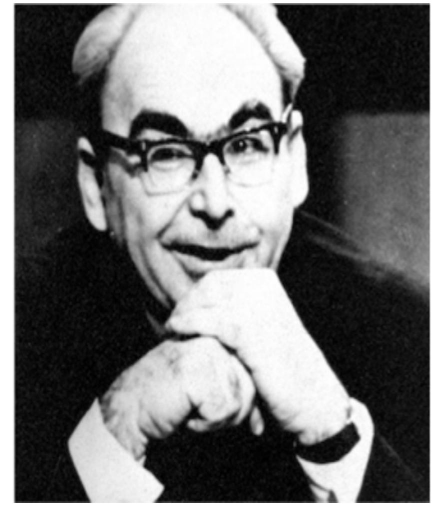
a.y. 2016/17

The random network model

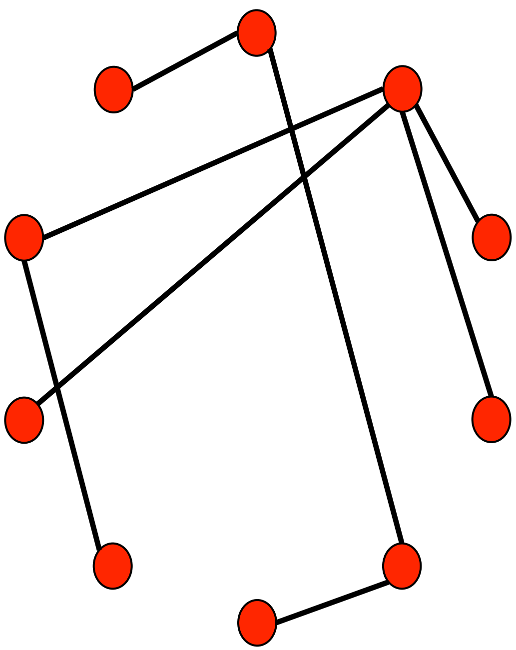
RANDOM NETWORK MODEL

Pál Erdös
(1913-1996)





Alfréd Rényi
(1921-1970)



Erdös-Rényi model (1960)

Connect with probability p

p=1/6 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**.

To construct a random network *G(N, p)*:

1) Start with *N* isolated nodes
2) Select a node pair, and generate a random number between 0 and 1. If the random number exceeds *p*, connect the selected node pair with a link, otherwise leave them disconnected
3) Repeat step (2) for each of the *N(N-1)/2* node pairs.

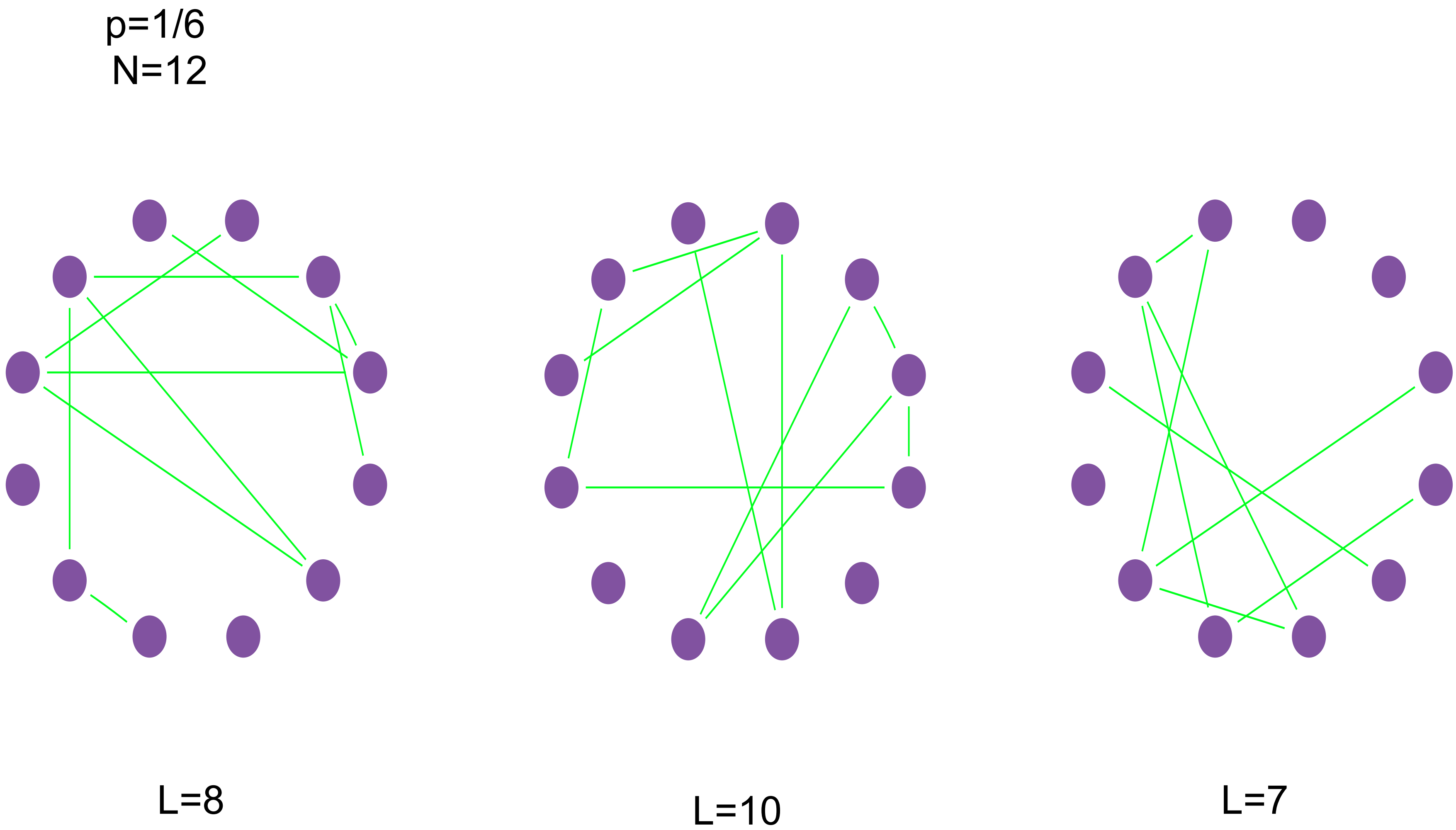
***G*(*N*, *L*) Model**

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**

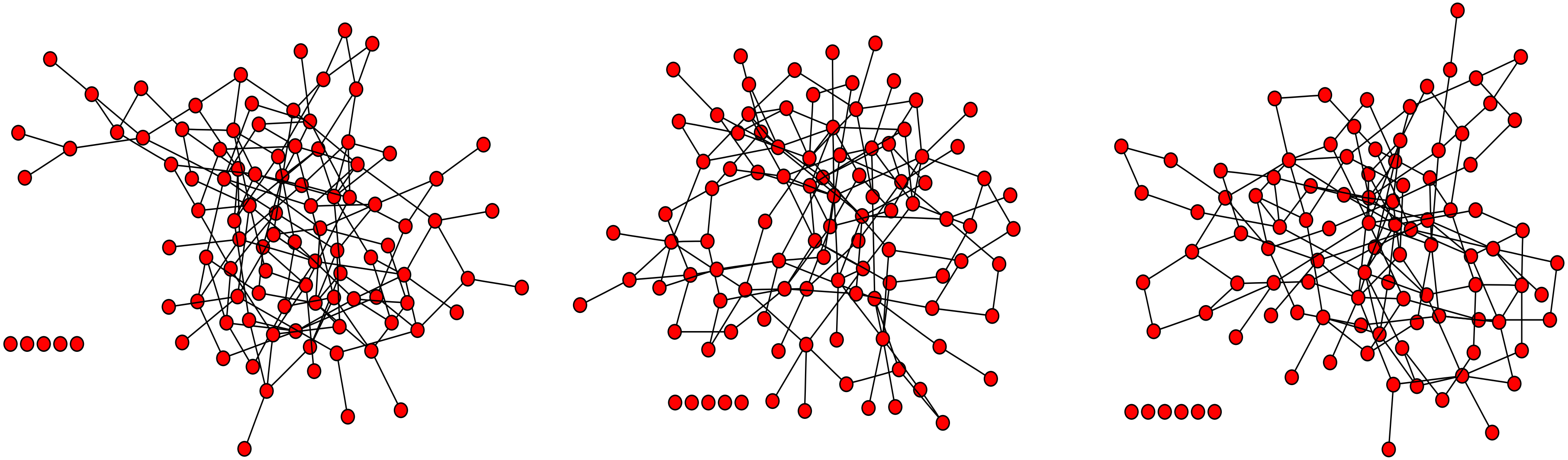
Each pair of N labeled nodes is connected with probability *p*, a model introduced by Gilbert [10].

RANDOM NETWORK MODEL



RANDOM NETWORK MODEL

p=0.03
N=100



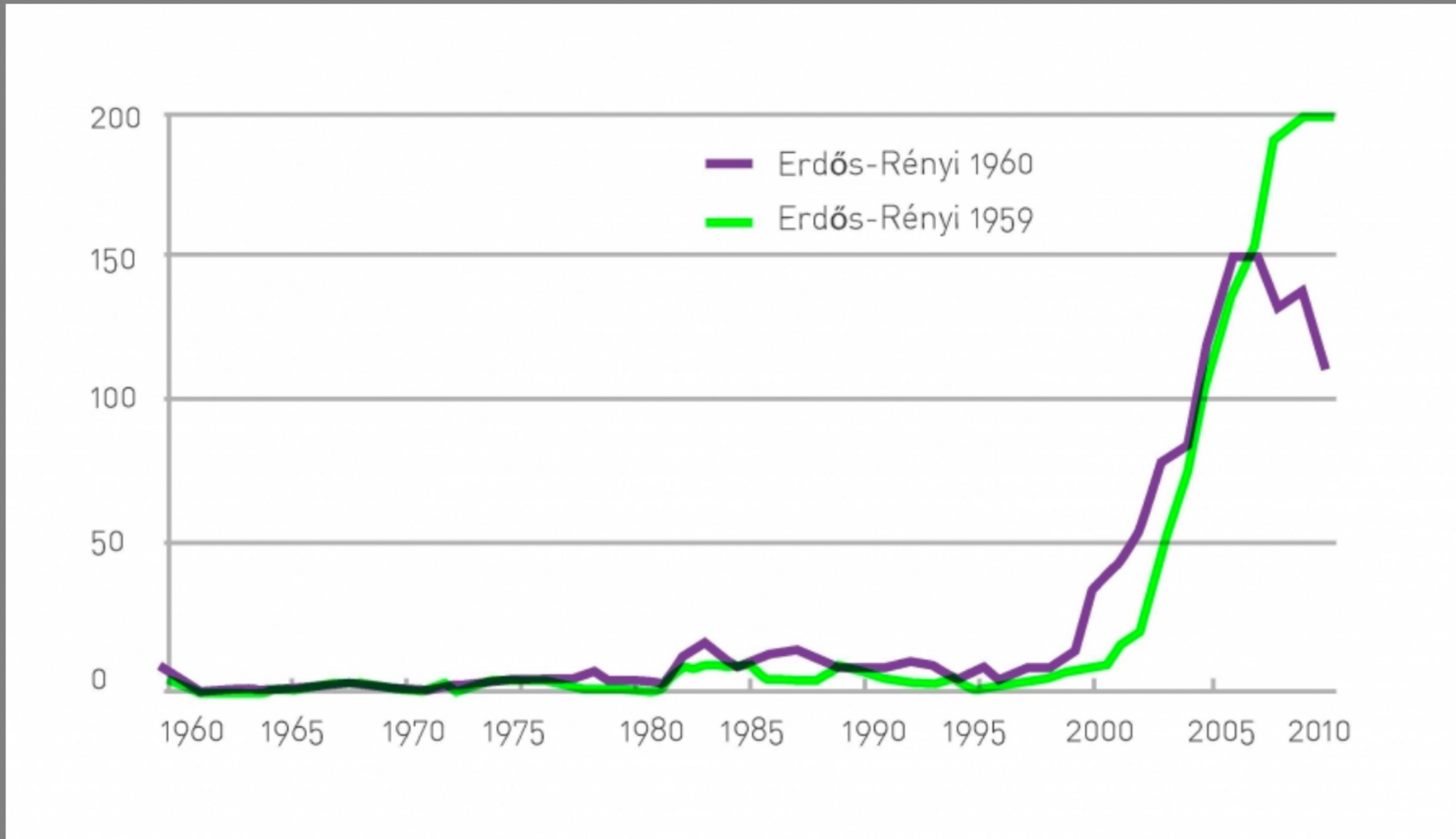


Image 3.15

Network Science and Random Networks

While today we perceive the Erdős-Rényi model as the cornerstone of network theory, the model was hardly known outside a small subfield of mathematics. This is illustrated by the yearly citations of the first two papers by Erdős and Rényi, published in 1959 and 1960 [2,3]. For four decades after their publication the papers gathered less than 10 citations each year. The number of citations exploded after the first papers on scale-free networks [21, 31, 32] have turned Erdős and Rényi's work into the reference model of network theory.

Binomial Distribution: The bottom line

Binomial Distribution: Mean and Variance

If we toss a fair coin N times, tails and heads occur with the same probability $p = 1/2$. The binomial distribution provides the probability p_x that we obtain exactly x heads in a sequence of N throws. In general, the binomial distribution describes the number of successes in N independent experiments with two possible outcomes, in which the probability of one outcome is p, and of the other is $1-p$.

The binomial distribution has the form

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

The mean of the distribution (first moment) is

$$\langle x \rangle = \sum_{x=0}^{N} x p_x = Np \qquad (3.4)$$

Its second moment is

$$\left\langle x^2 \right\rangle = \sum_{x=0}^{N} x^2 p_x = p(1-p)N + p^2 N^2 \qquad (3.5)$$

providing its standard deviation as

$$\sigma_x = \left(\left\langle x^2 \right\rangle - \langle x \rangle^2 \right)^{\frac{1}{2}} = [p(1-p)N]^{\frac{1}{2}} \qquad (3.6)$$

Equations (3.4) - (3.6) are used repeatedly as we characterize random networks.

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*:

The maximum number of links in a network of N nodes = number of pairs of distinct nodes.

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

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

Binomial distribution...

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

RANDOM NETWORK MODEL

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

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

- The average number of links *<L>* in a random graph

$$< L >= \sum_{L=0}^{\frac{N(N-1)}{2}} LP(L) = p\frac{N(N-1)}{2}$$

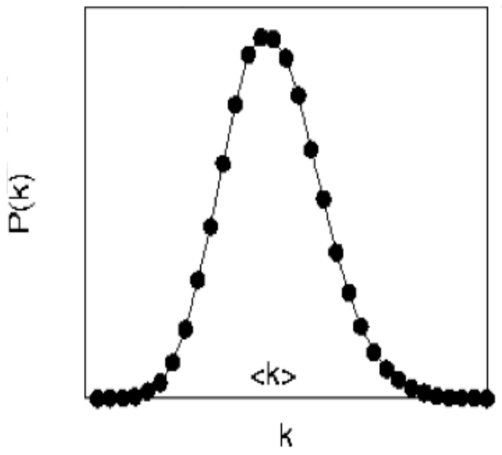
$$< k >= 2\frac{< L >}{N} = p(N-1)$$

- 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

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

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

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 OF A RANDOM GRAPH

$$P(k) = \binom{N-1}{k} p^k (1-p)^{(N-1)-k} \qquad <k> = p(N-1) \qquad p = \frac{<k>}{(N-1)}$$

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

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

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

$$(1-p)^{(N-1)-k} = e^{-<k>} \qquad \ln(1+x) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} x^n = x - \frac{x^2}{2} + \frac{x^3}{3} - \ldots \quad \text{for} \quad |x| \leq 1$$

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

POISSON DEGREE DISTRIBUTION

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

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

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

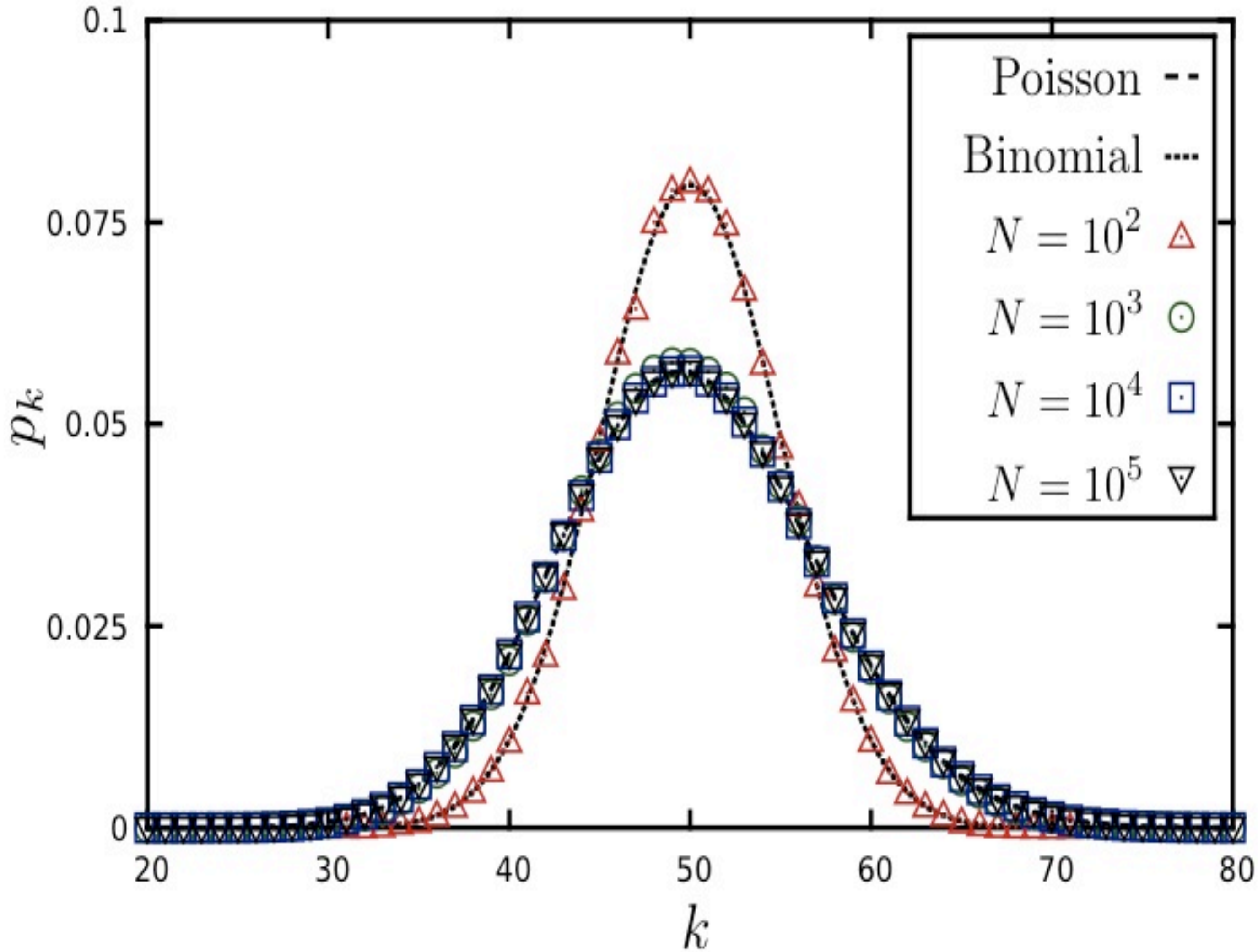
For large *N* and small *k*, we arrive to the Poisson distribution:

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

DEGREE DISTRIBUTION OF A RANDOM GRAPH

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

<k>=50

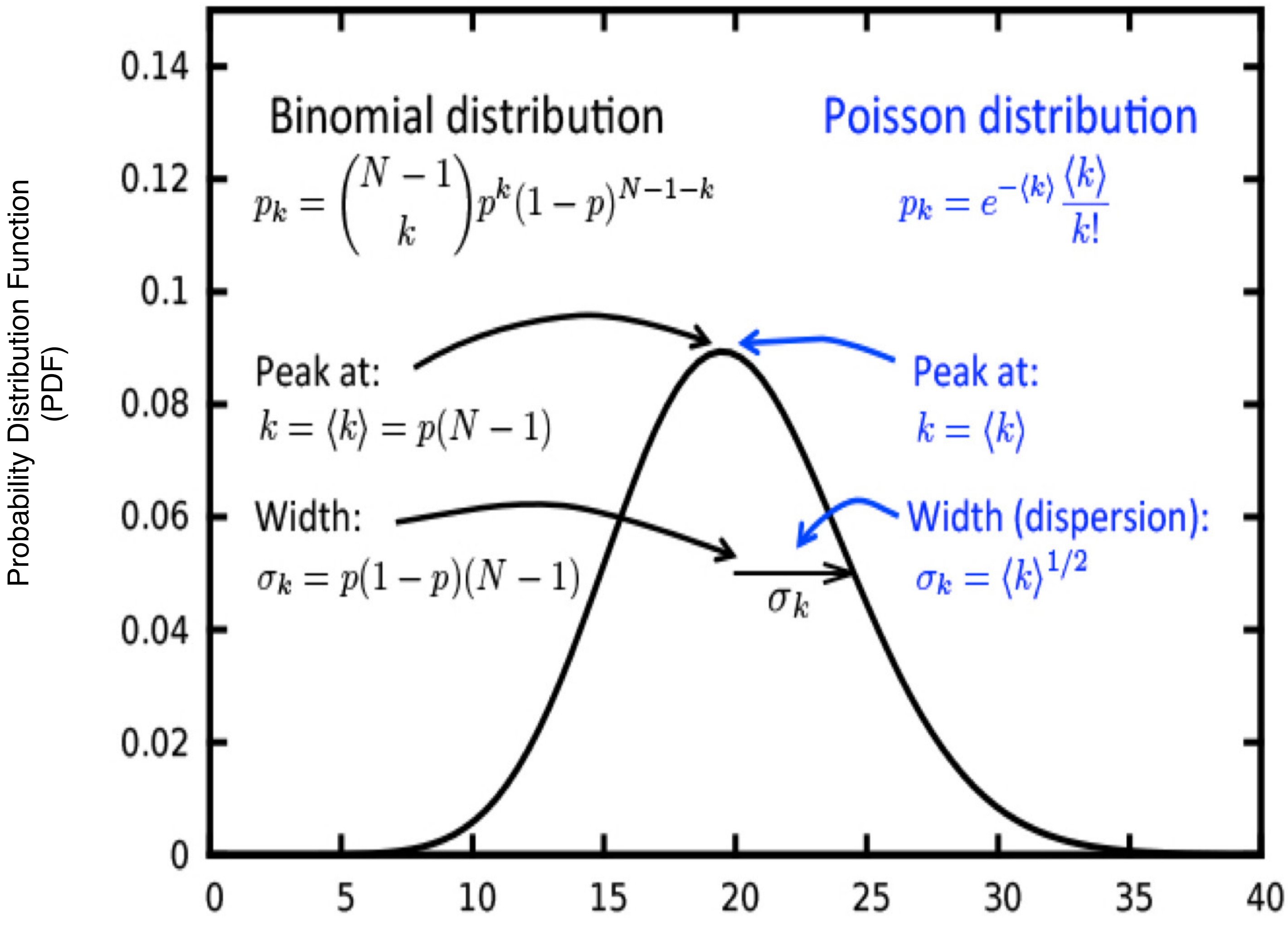


K

DEGREE DISTRIBUTION OF A RANDOM NETWORK

Exact Result
-binomial distribution-

Large N limit
-Poisson distribution-



Real Networks are not Poisson

NO OUTLIERS IN A RANDOM SOCIETY

Sociologists estimate that a typical person knows about 1,000 individuals on a first name basis, prompting us to assume that ‹k› ≈ 1,000.

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

→The most connected individual has degree k_{max}~1,185

→The least connected individual has degree k_{min} ~ 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 inexistent 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 recluse.

This suprising conclusion is a consequence of an important property of random networks:
In a large random network the degree of most nodes is in the narrow vicinity of ‹k›

$$P(k) = e^{-<k>} \frac{<k>^k}{k!} \qquad (3.8)$$

Box 3.4

Why are Hubs Missing?

We first note that the $1/k!$ term in (3.8) significantly decreases the chances of observing large degree nodes. Indeed, the Stirling approximation

$$k! \sim \left[\sqrt{2\pi k}\right] \left(\frac{k}{e}\right)^k$$

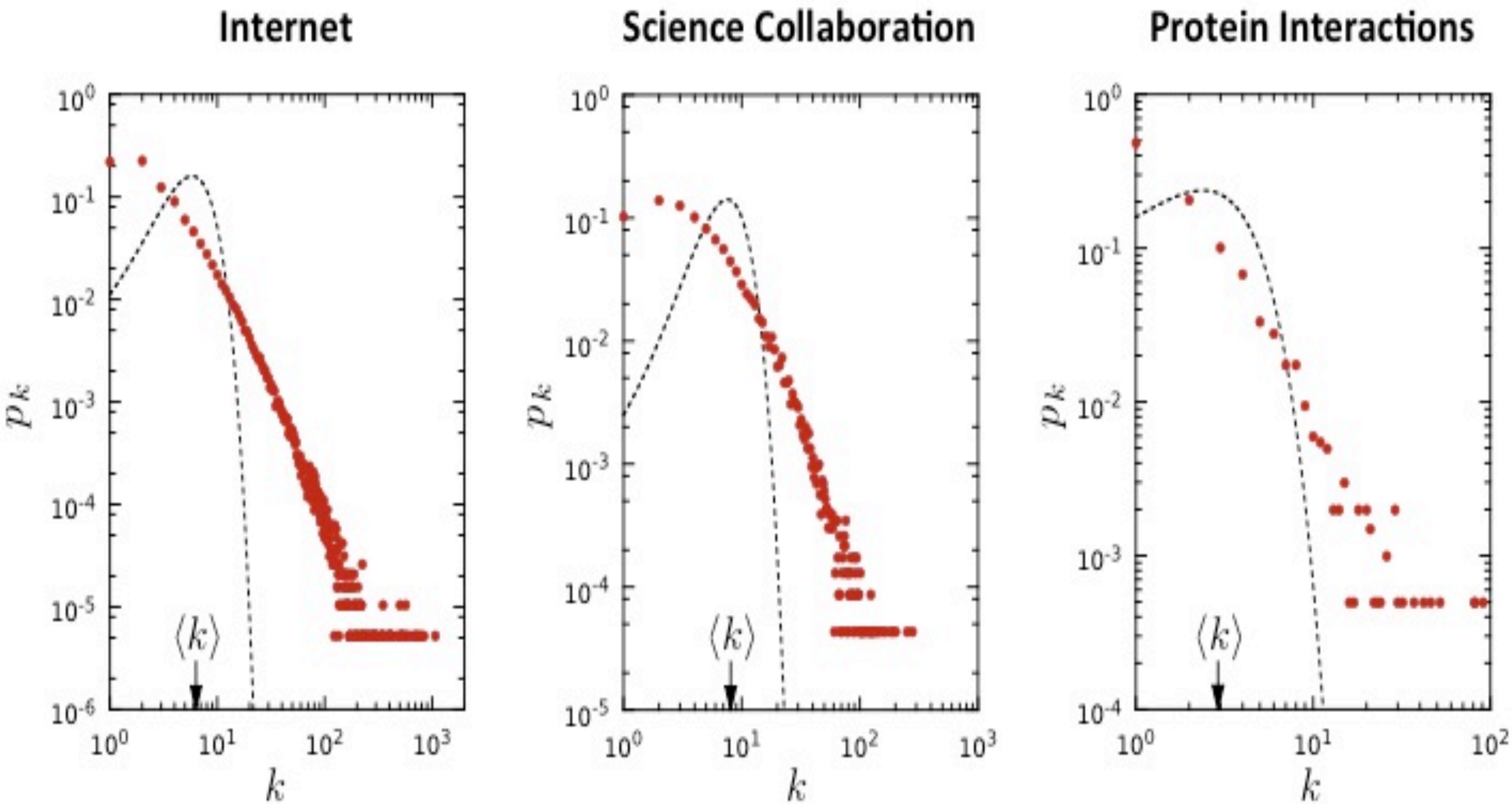
allows us rewrite (3.8) as

$$p_k = \frac{e^{-\langle k \rangle}}{\sqrt{2\pi k}} \left(\frac{e\langle k \rangle}{k}\right)^k \qquad (3.9)$$

For degrees $k > e<k>$ the term in the parenthesis is smaller than one, hence for large k both k-dependent terms in (3.9), i.e. $1/\sqrt{k}$ and $(e<k>/k)^k$ decrease rapidly with increasing k. Overall (3.9) predicts that in a random network the chance of observing a hub decreases faster than exponentially.

FACING REALITY: Degree distribution of real networks

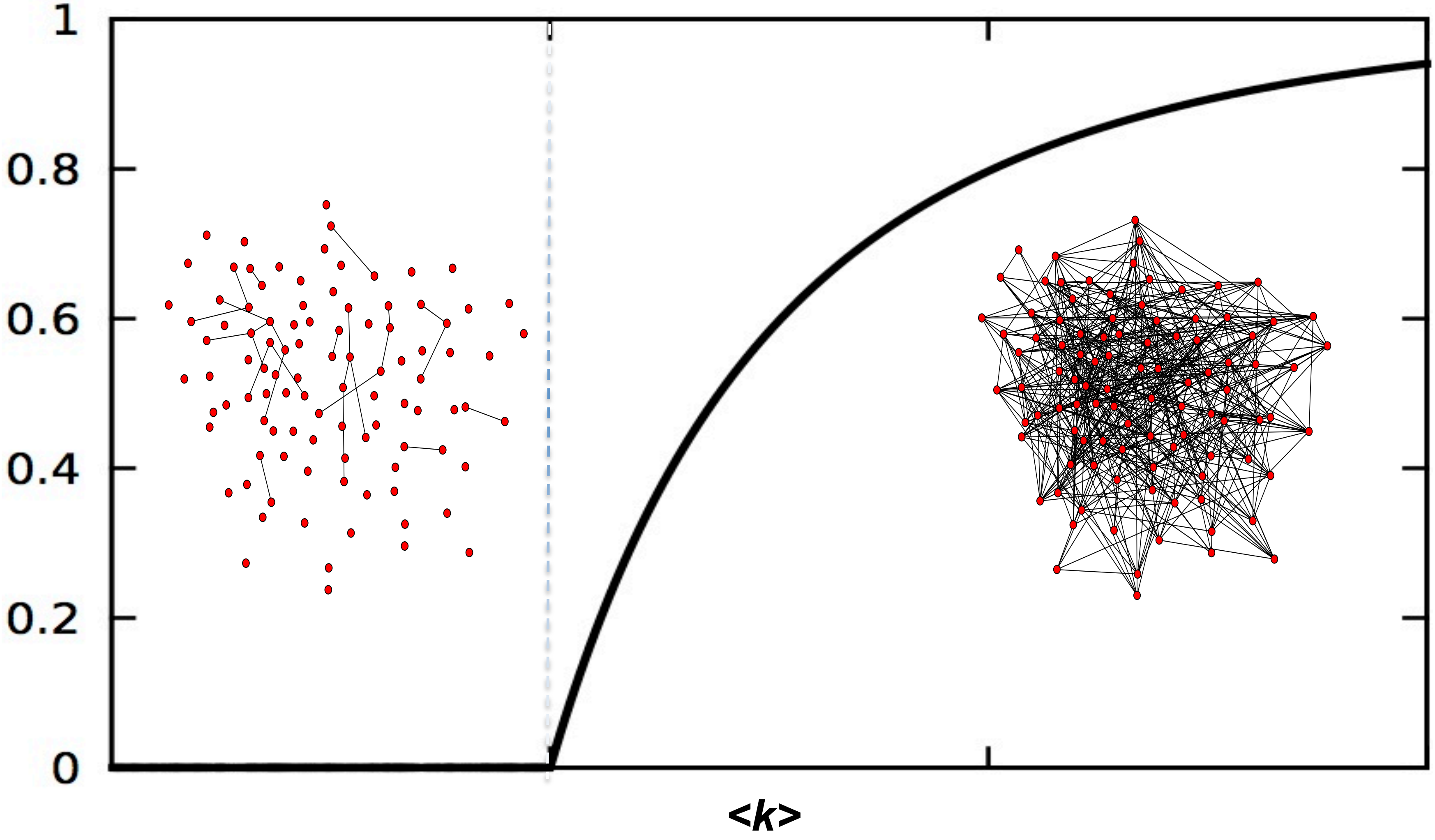
$$P(k) = e^{-<k>} \frac{<k>^k}{k!}$$



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.**

$<k_c>=1$ (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 $<k>=1$.

Let us denote with $u = 1 - N_G/N$ the fraction of nodes that are not in the giant component (*GC*), whose size we take to be N_G. If node *i* is part of the *GC*, it must link to another node *j*, which must also be part of the *GC*. Hence if *i* is *not* part of the *GC*, that could happen for two reasons:

- There is no link between *i* and *j* (probability for this is 1- *p*).

- There is a link between *i* and *j*, but *j* is not part of the *GC* (probability for this is *pu*).

Therefore the 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 therefore $(1 - p + pu)^{N-1}$, as there are $N - 1$ nodes that could serve as potential links to the *GC* for node *i*. As *u* is the fraction of nodes that do not belong to the *GC*, for any *p* and *N* the solution of the equation

$$u = (1 - p + pu)^{N-1} \qquad (3.30)$$

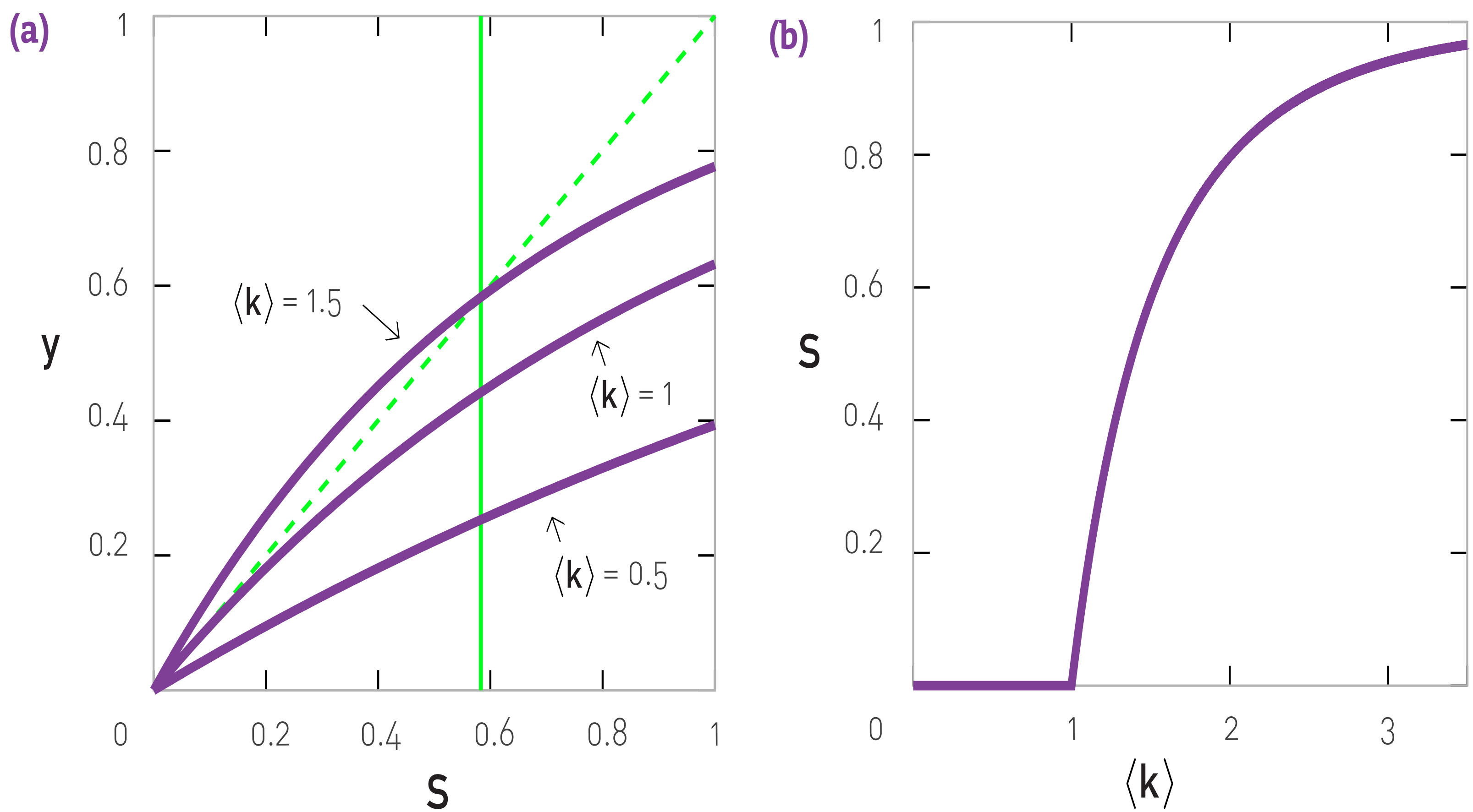
provides the size of the giant component via $N_G = N(1 - u)$. Using $p = \langle k \rangle / (N - 1)$ and taking the log of both sides, for $\langle k \rangle \ll N$ we obtain

$$\ln u \simeq (N-1)\ln\left[1 - \frac{\langle k \rangle}{N-1}(1-u)\right]. \qquad (3.31)$$

Taking an exponential of both sides leads to $u = \exp[-\langle k \rangle(1 - u)]$. If we denote with *S* the fraction of nodes in the giant component, $S = N_G / N$, then $S = 1 - u$ and (3.31) results in

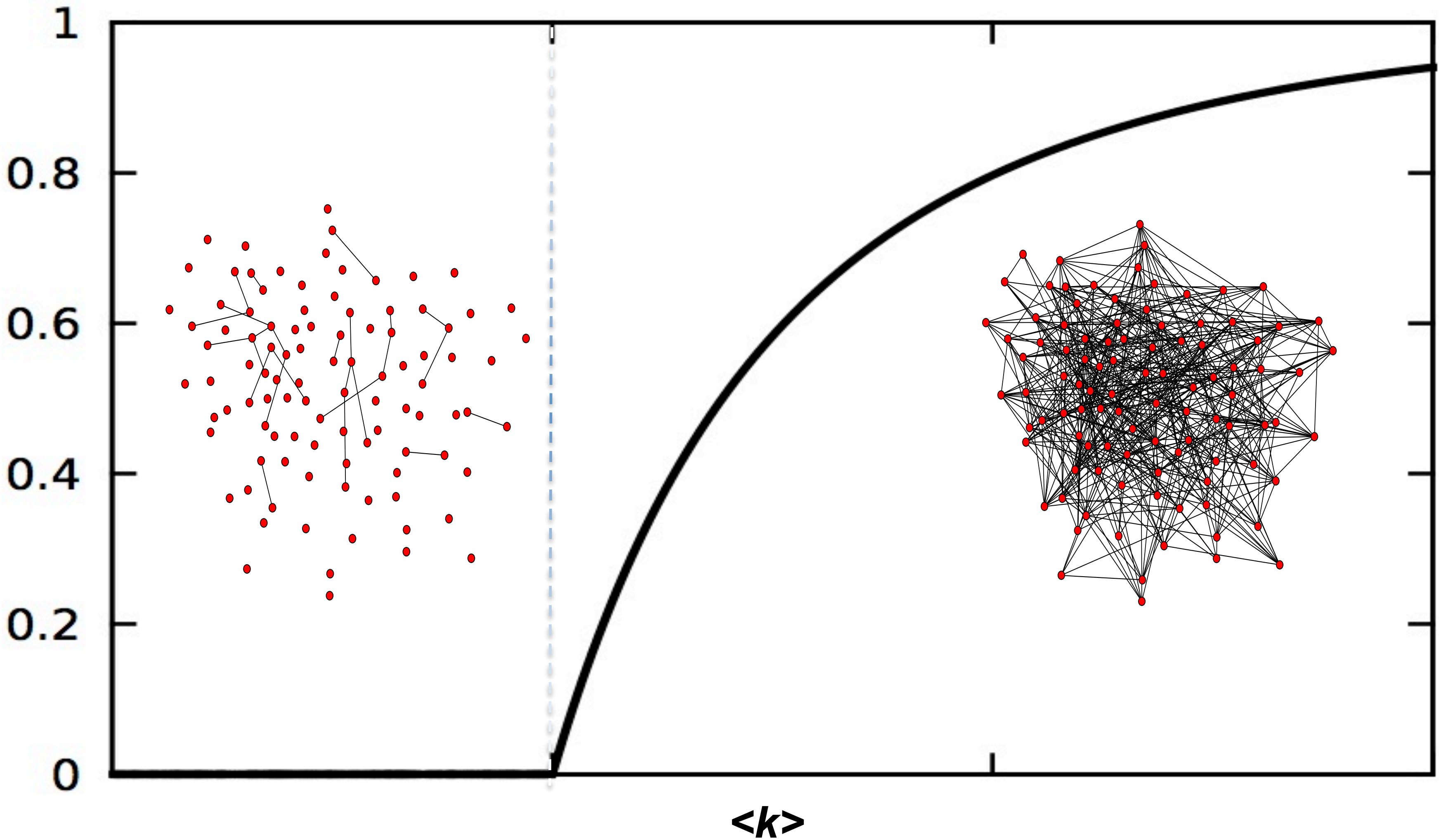
$$S = 1 - e^{-\langle k \rangle S}.$$

$$S = 1 - e^{-\langle k \rangle S}. \quad (3.32)$$



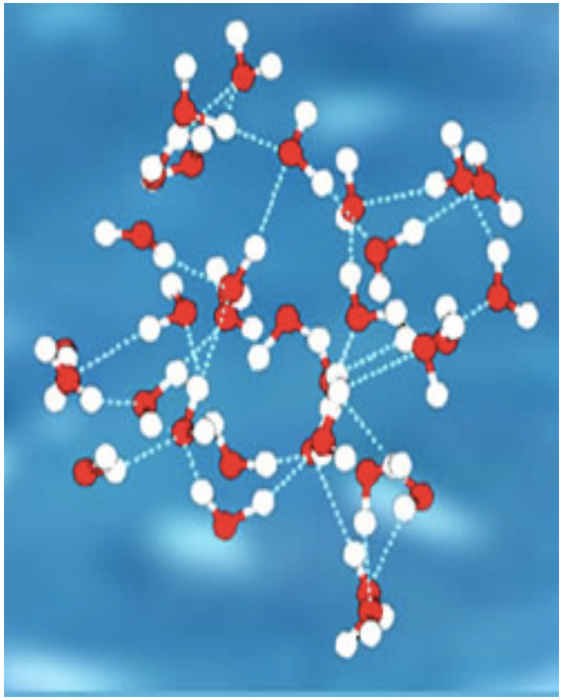
EVOLUTION OF A RANDOM NETWORK

disconnected nodes ➔ **NETWORK.**

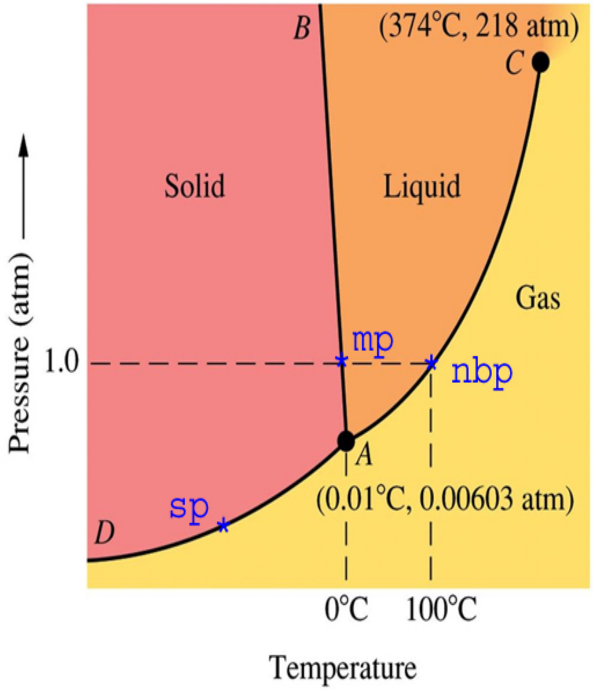


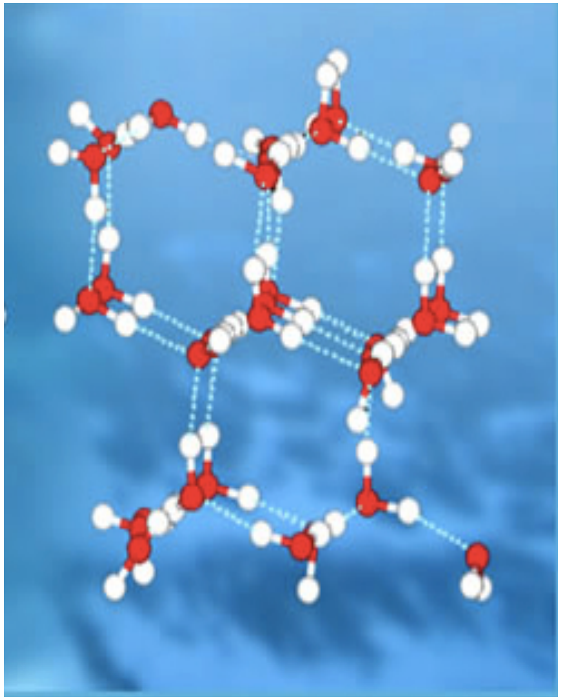
How does this transition happen?

Phase transitions in complex systems: liquids

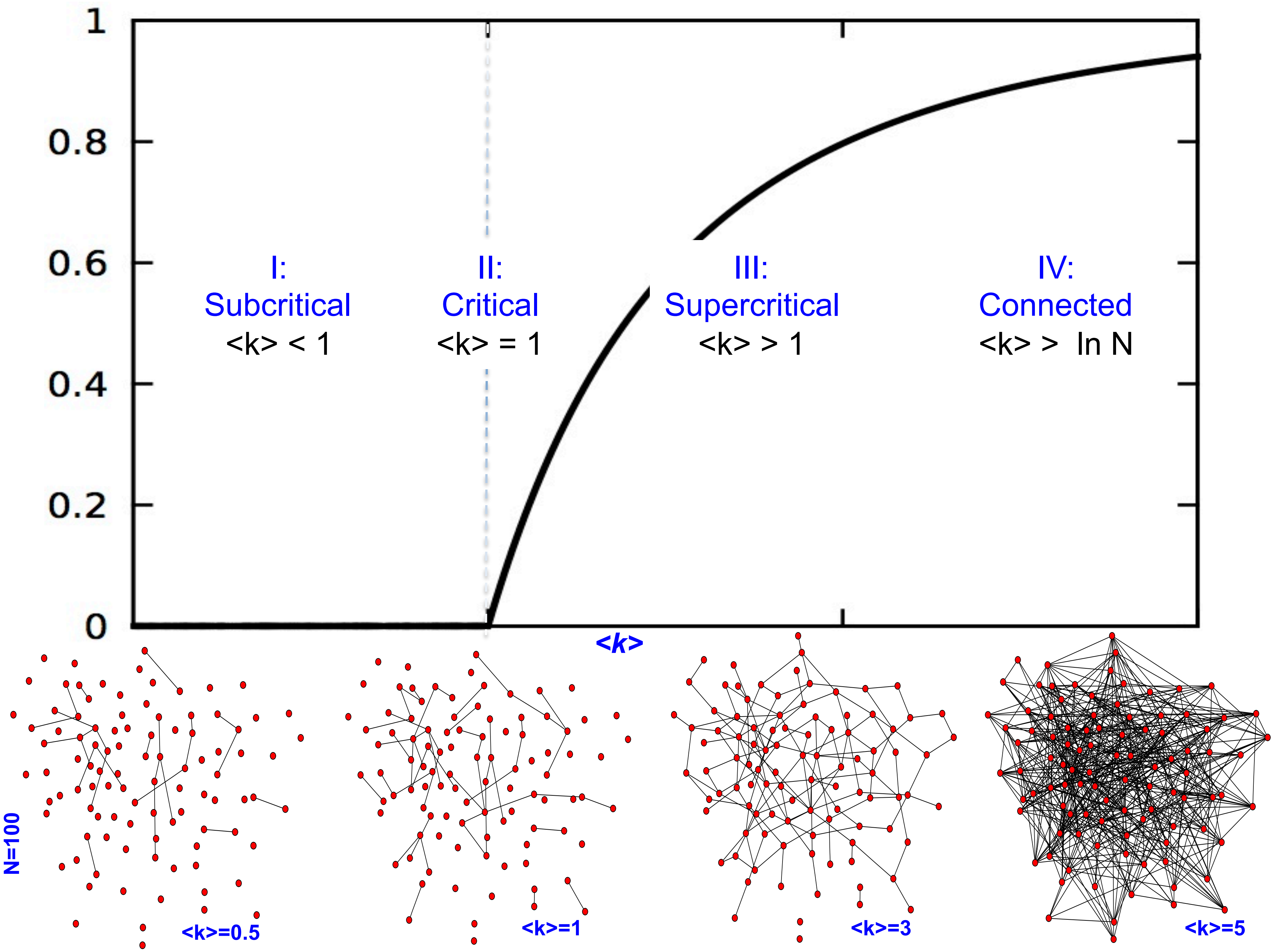


Water

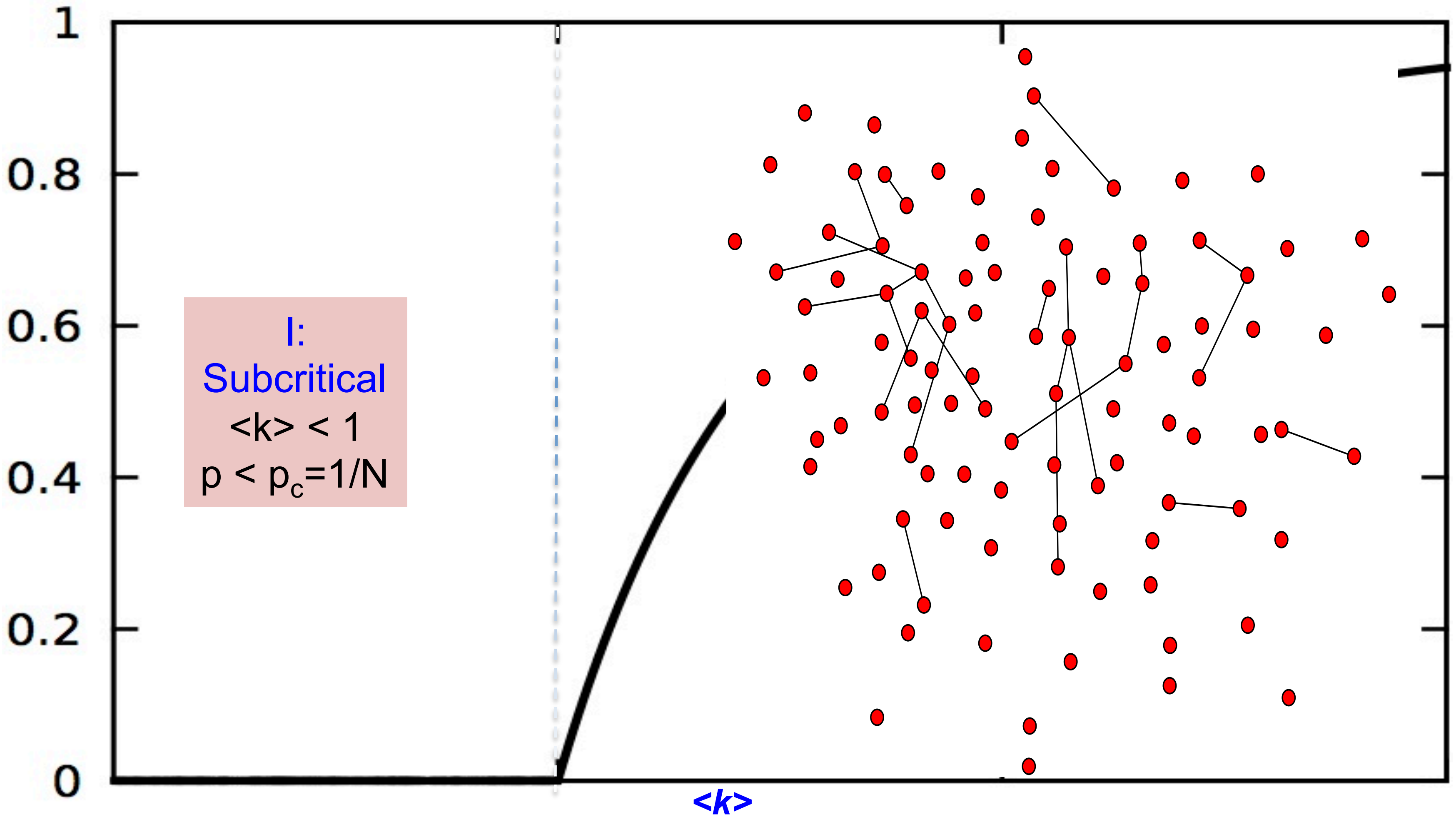




Ice



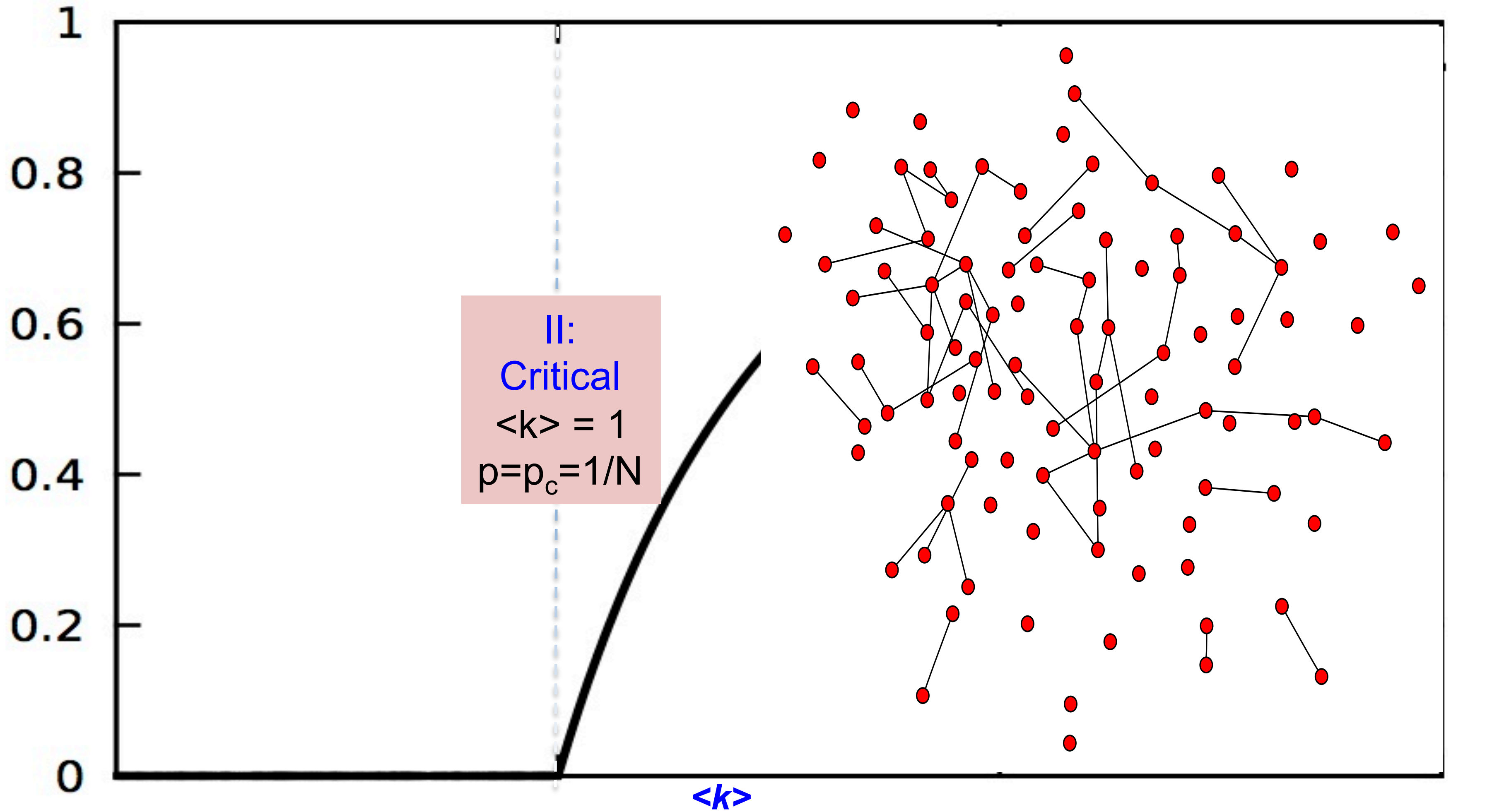
1
0.8
0.6
0.4
0.2
0
I:
Subcritical
<k> < 1
II:
Critical
<k> = 1
III:
Supercritical
<k> > 1
IV:
Connected
<k> > ln N
<k>
N=100
<k>=0.5
<k>=1
<k>=3
<k>=5



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 ~ *ln N*



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

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

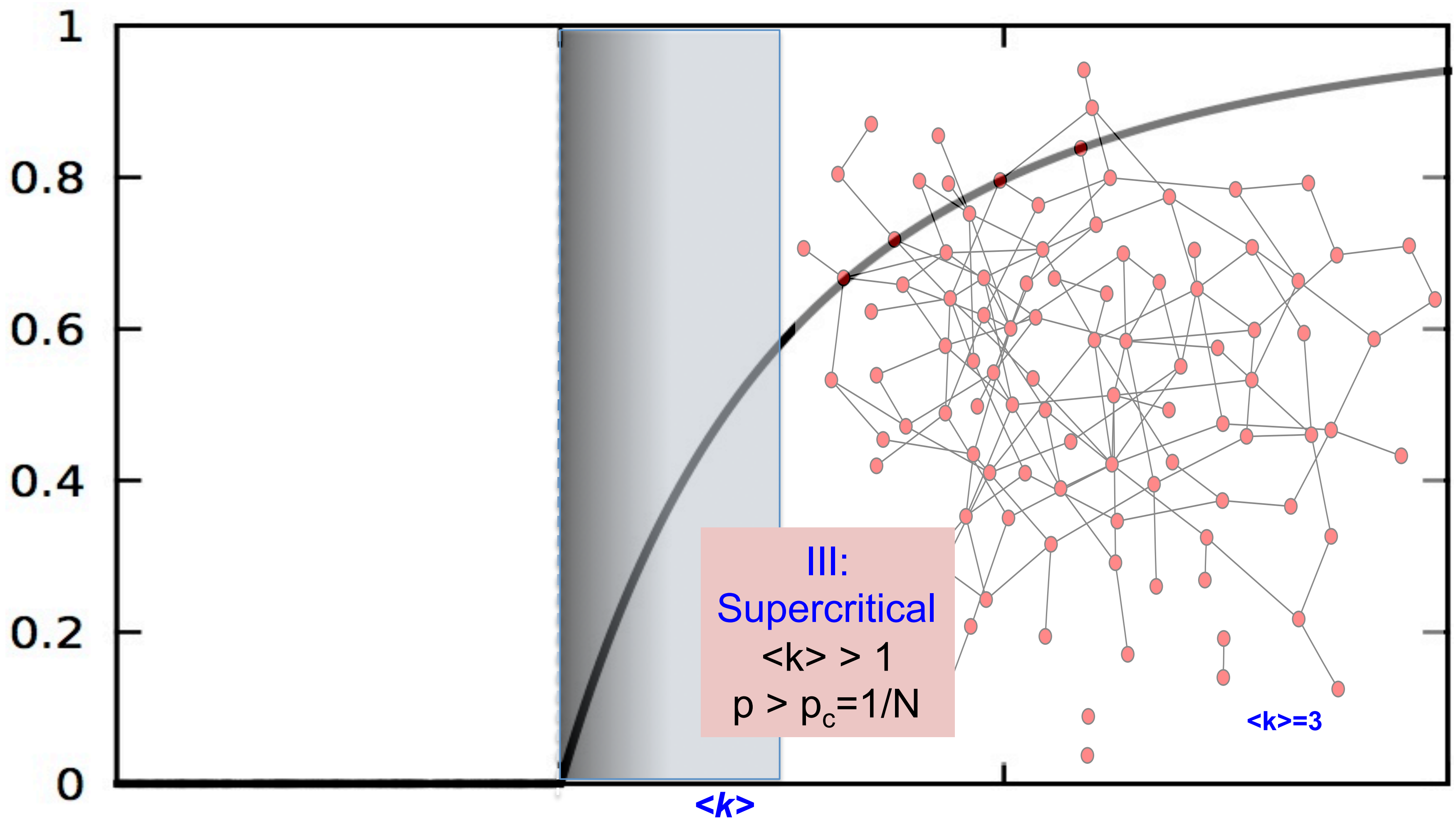
→Small components are trees, GC has loops.

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

A jump in the cluster size:

$N=1{,}000$ → ln N~ 6.9; $N^{2/3}$~95

$N=7\ 10^9$ → ln N~ 22; $N^{2/3}$~3,659,250

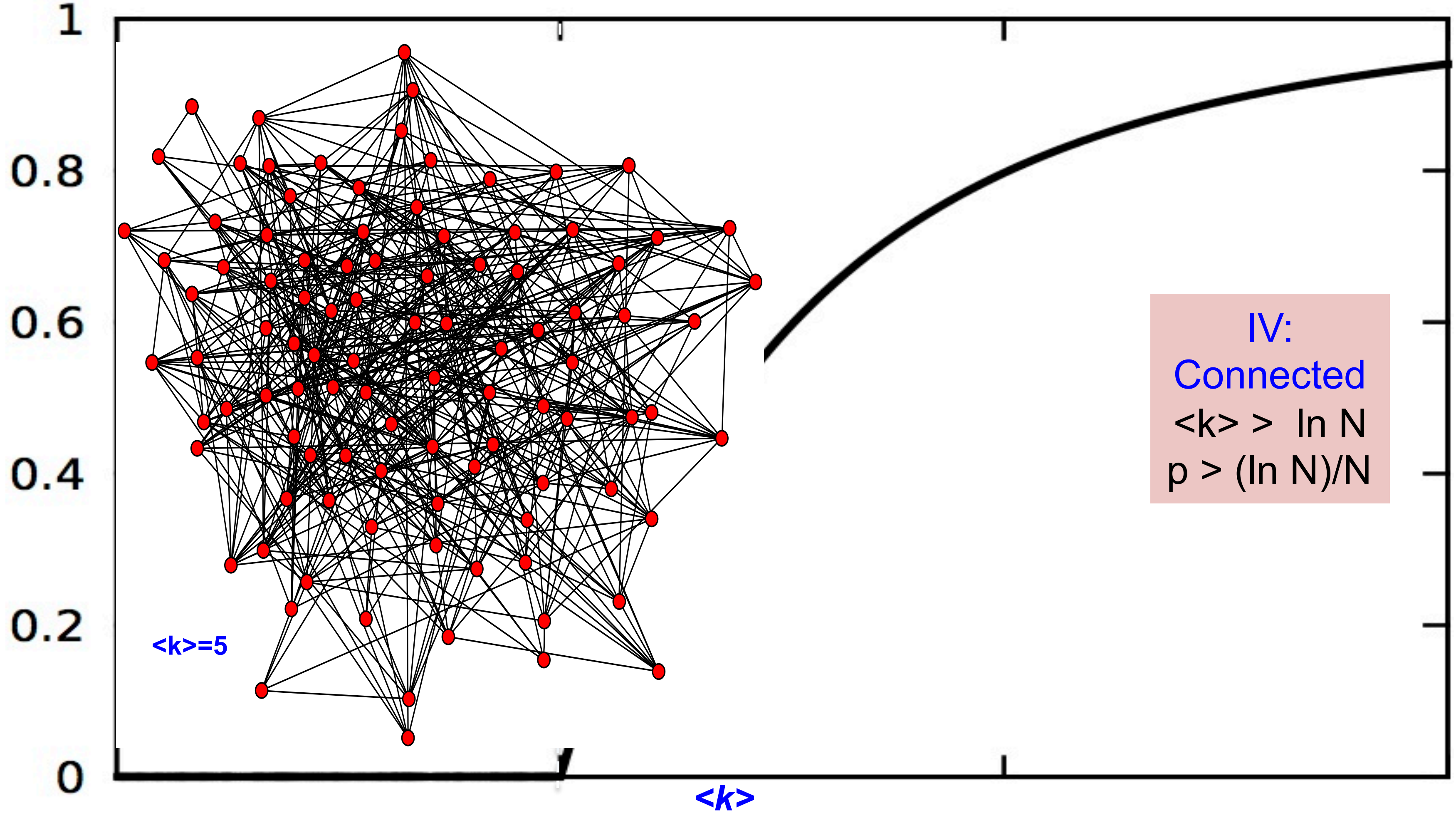


Unique giant component: $N_G \sim (p-p_c)N$

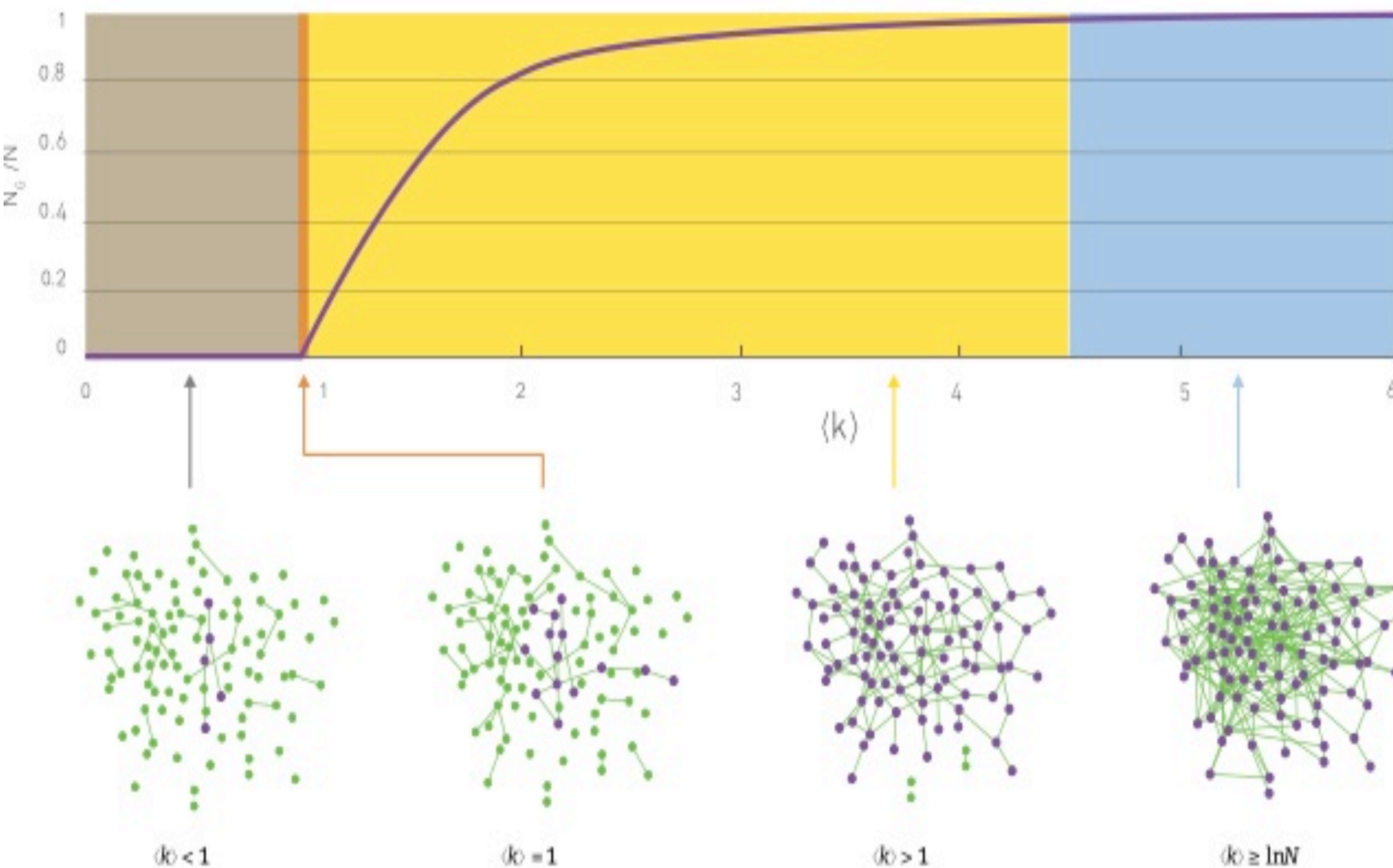
→GC has loops.

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



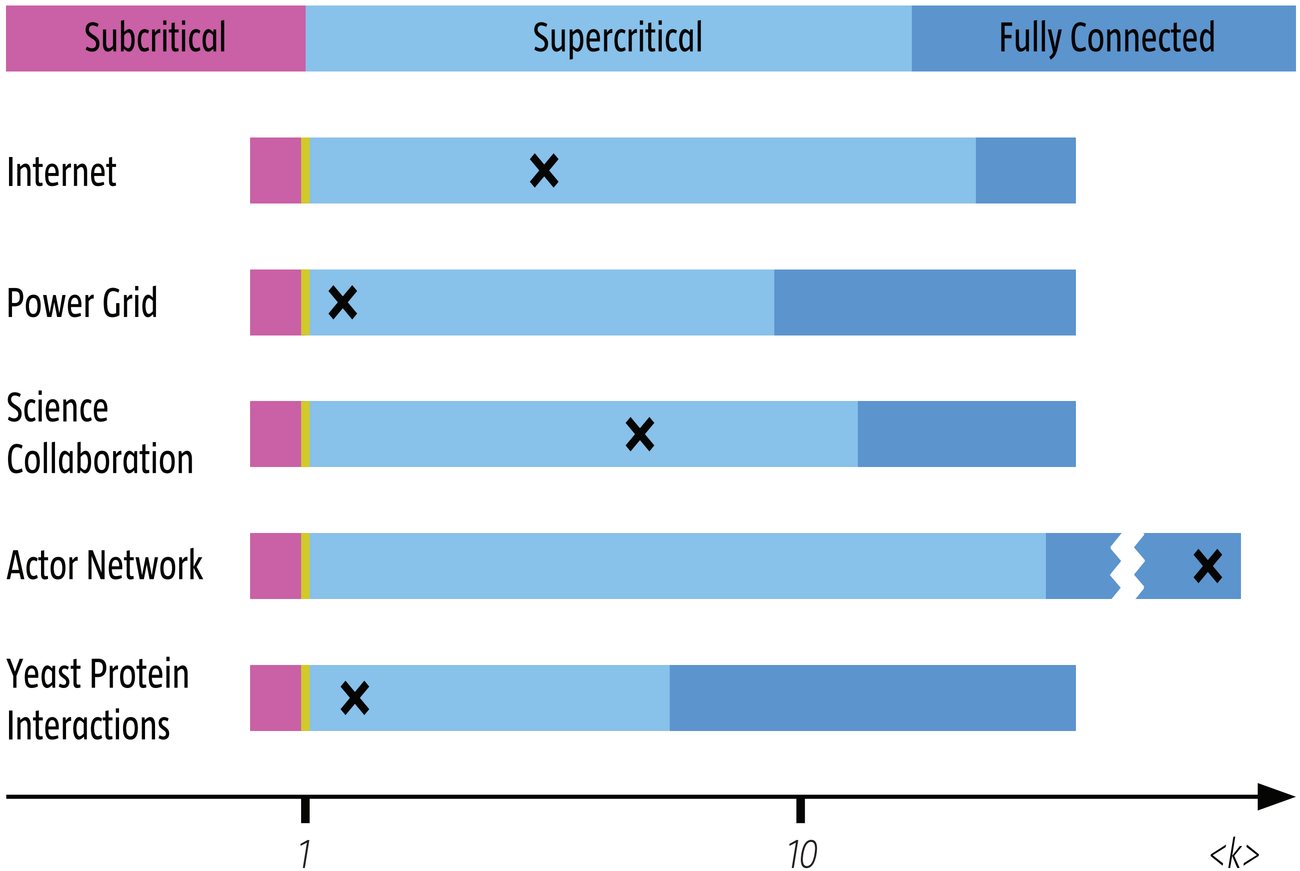
(b) Subcritical Regime
- No giant component
- Cluster size distribution: $p_s \sim s^{-3/2} e^{-\alpha s}$
- Size of the largest cluster: $N_G \sim \ln N$
- The clusters are trees

(c) Critical Point
- No giant component
- Cluster size distribution: $p_s \sim s^{-3/2}$
- Size of the largest cluster: $N_G \sim N^{2/3}$
- The clusters may contain loops

(d) Supercritical Regime
- Single giant component
- Cluster size distribution: $p_s \sim s^{-3/2} e^{-\alpha s}$
- Size of the giant component: $N_G \sim (p - p_c)N$
- 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

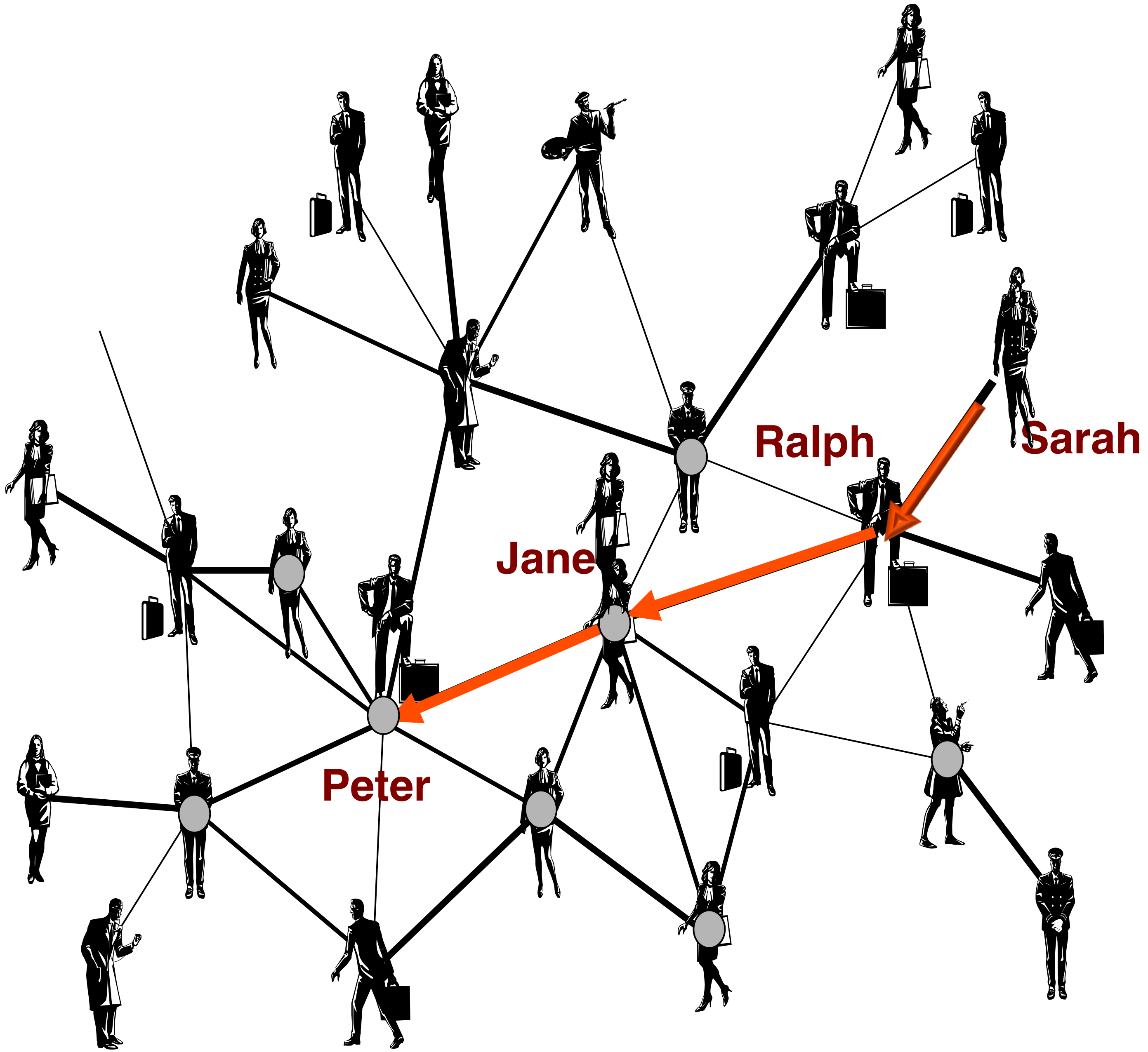
Real networks are supercritical



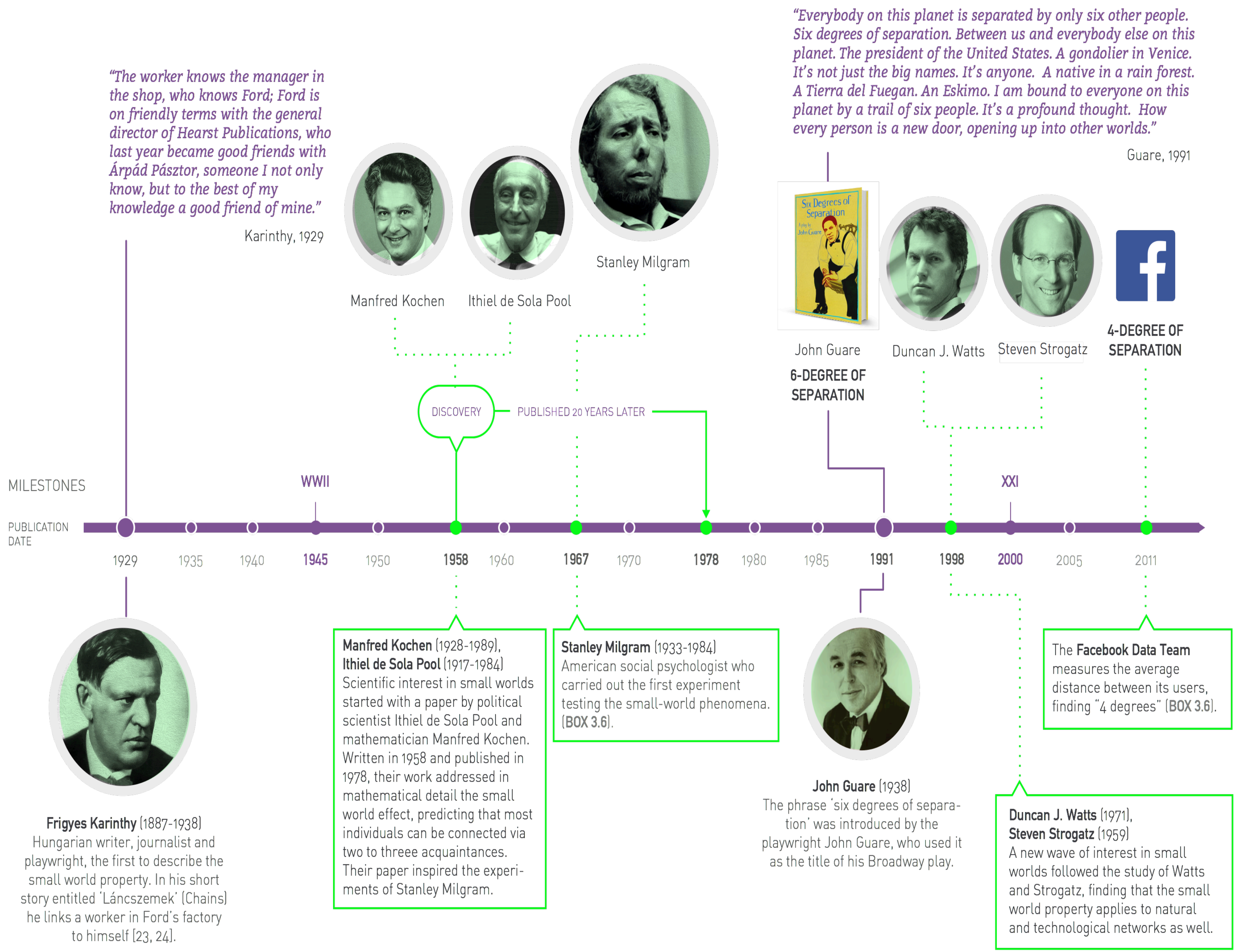
Network	N	L	$\langle k \rangle$	$\ln N$
Internet	192,244	609,066	6.34	12.17
Power Grid	4,941	6,594	2.67	8.51
Science Collaboration	23,133	186,936	8.08	10.04
Actor Network	212,250	3,054,278	28.78	12.27
Yeast Protein Interactions	2,018	2,930	2.90	7.61

Small worlds

SIX DEGREES small worlds



Frigyes Karinthy, 1929
Stanley Milgram, 1967



"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
Manfred Kochen
Ithiel de Sola Pool
Stanley Milgram
"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
Six Degrees of Separation
John Guare
6-DEGREE OF SEPARATION
Duncan J. Watts
Steven Strogatz
4-DEGREE OF SEPARATION
DISCOVERY
PUBLISHED 20 YEARS LATER
MILESTONES
WWII
XXI
PUBLICATION DATE
1929
1935
1940
1945
1950
1958
1960
1967
1970
1978
1980
1985
1991
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 threee 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**).
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.

1929: Frigyes Kartinthy

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."

1967: Stanley Milgram

HOW TO TAKE PART IN THIS STUDY

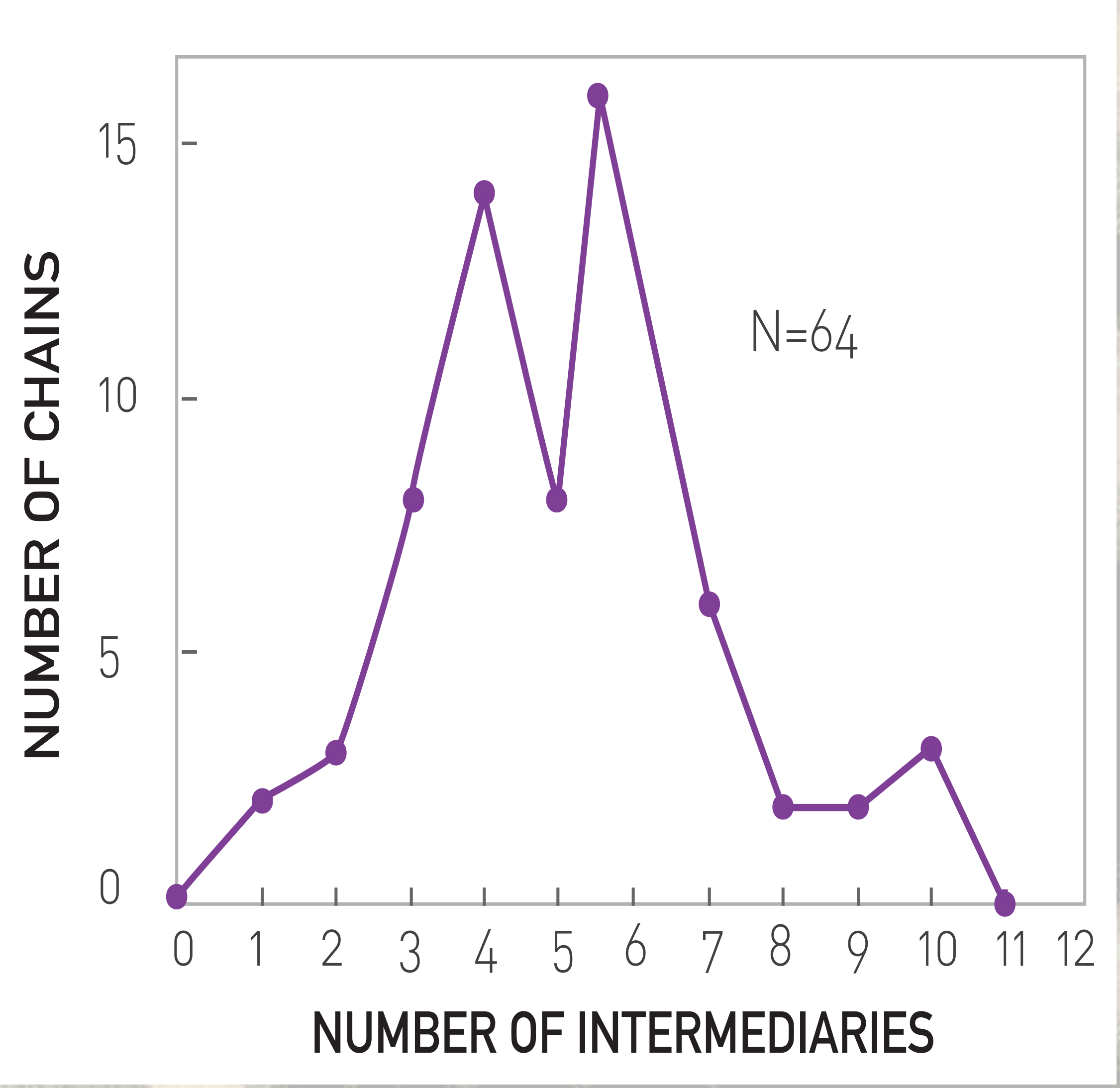
1. ADD YOUR NAME TO THE ROSTER AT THE BOTTOM OF THIS SHEET, so that the next person who receives this letter will know who it came from.

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.

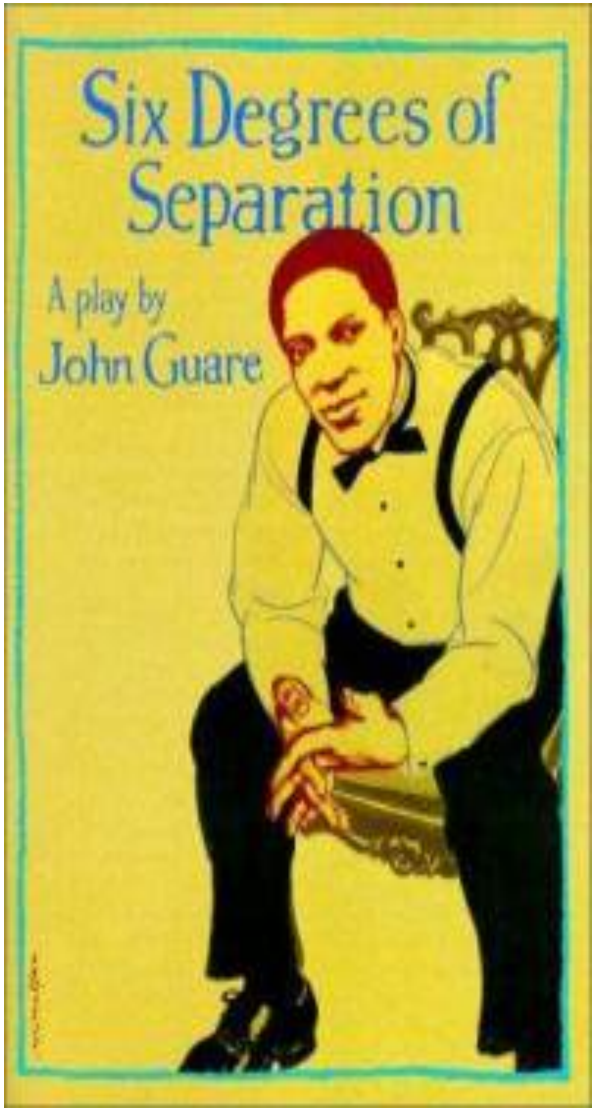
3. IF YOU KNOW THE TARGET PERSON ON A PERSONAL BASIS, MAIL THIS FOLDER DIRECTLY TO HIM (HER). Do this only if you have previously met the target person and know each other on a first name basis.

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.

1967: Stanley Milgram



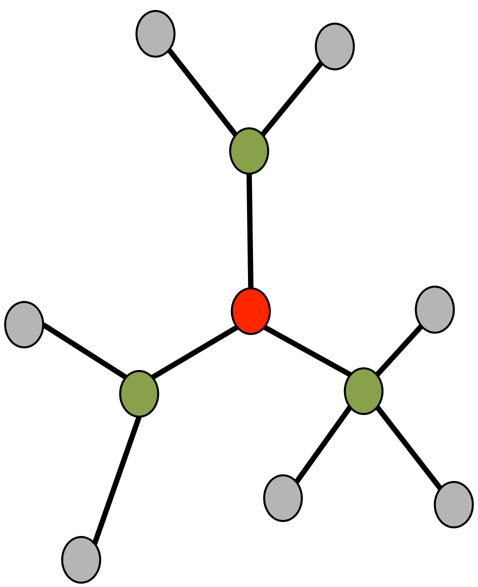
1991: John Guare



"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."

DISTANCES IN RANDOM GRAPHS

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



$<k>$ nodes at distance one (d=1).

$<k>^2$ nodes at distance two (d=2).

$<k>^3$ nodes at distance three (d =3).

...

$<k>^d$ nodes at distance d.

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

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} .

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

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

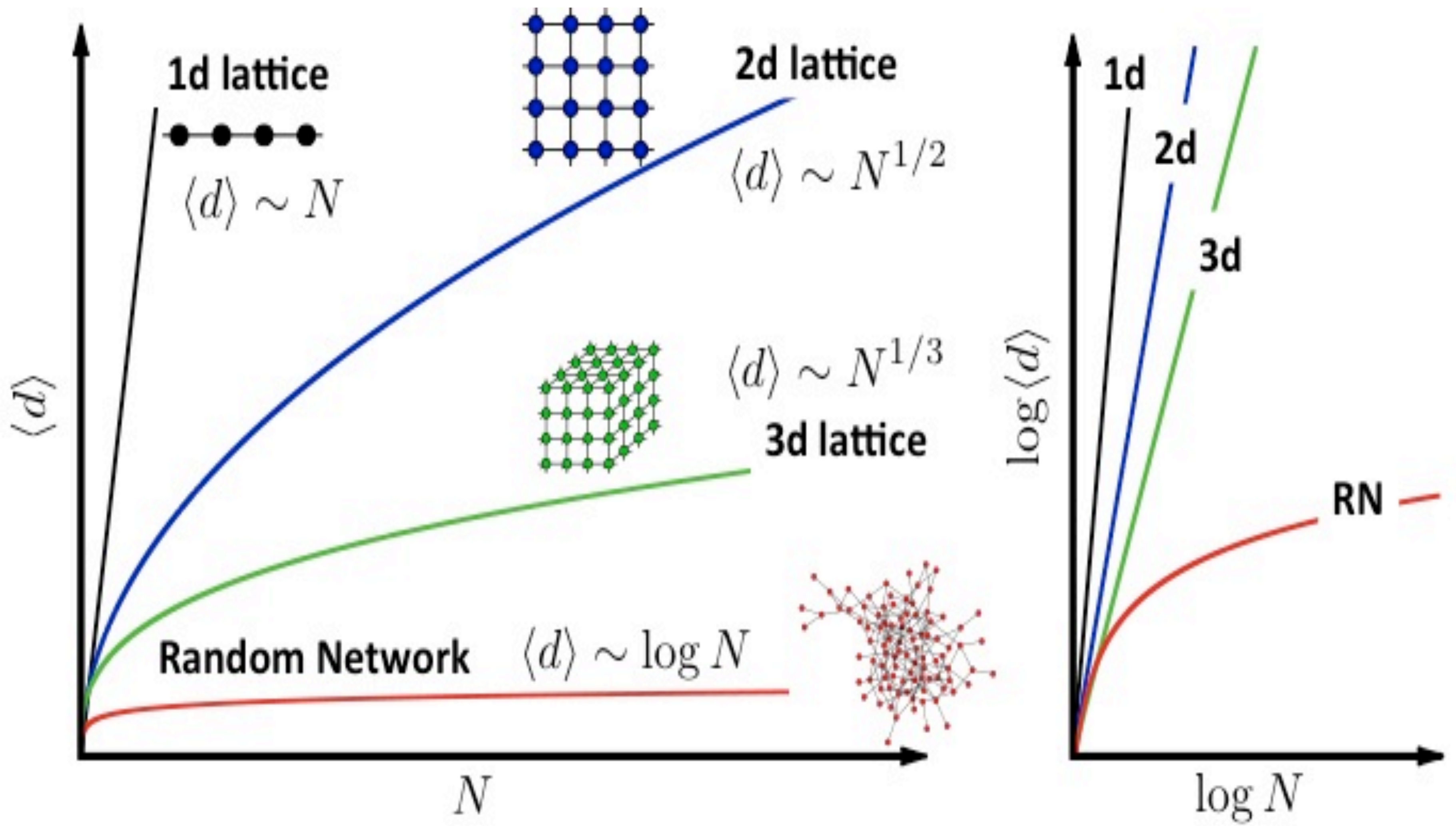
compare with real data

Network	N	L	‹k›	‹d›	d_{max}	ln*N*/ln‹*k*›
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	93,437	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?

Suprising compared to what?



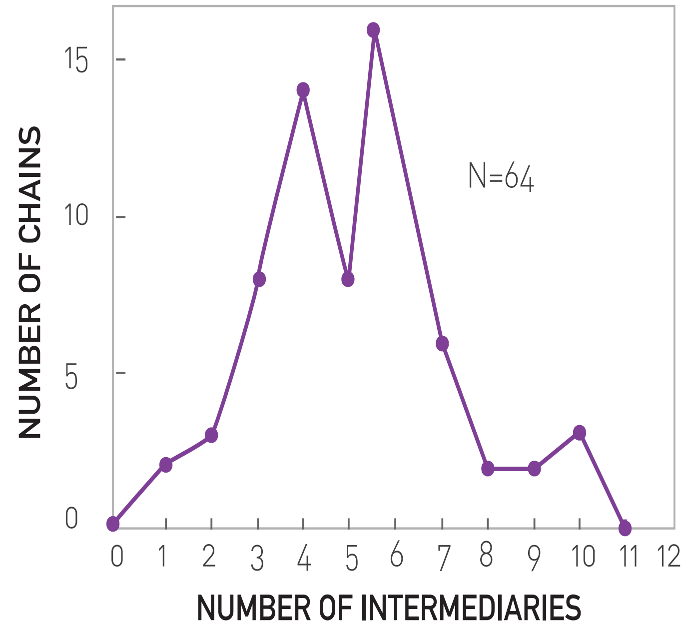
Three, Four or Six Degrees?

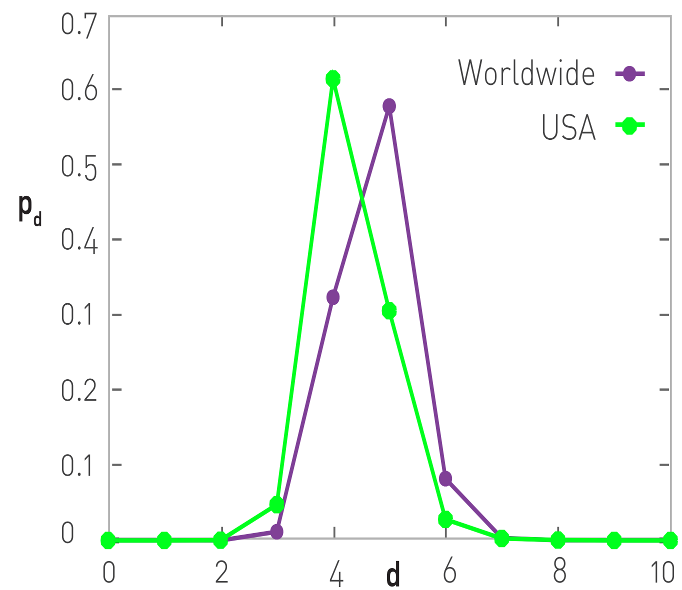
For the globe's social networks:

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

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

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

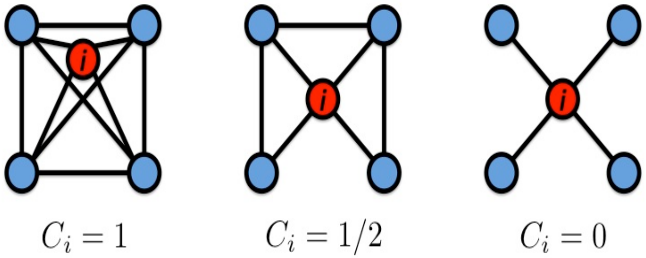




Clustering coefficient

CLUSTERING COEFFICIENT

$$C_i \equiv \frac{2 < L_i >}{k_i(k_i - 1)}$$

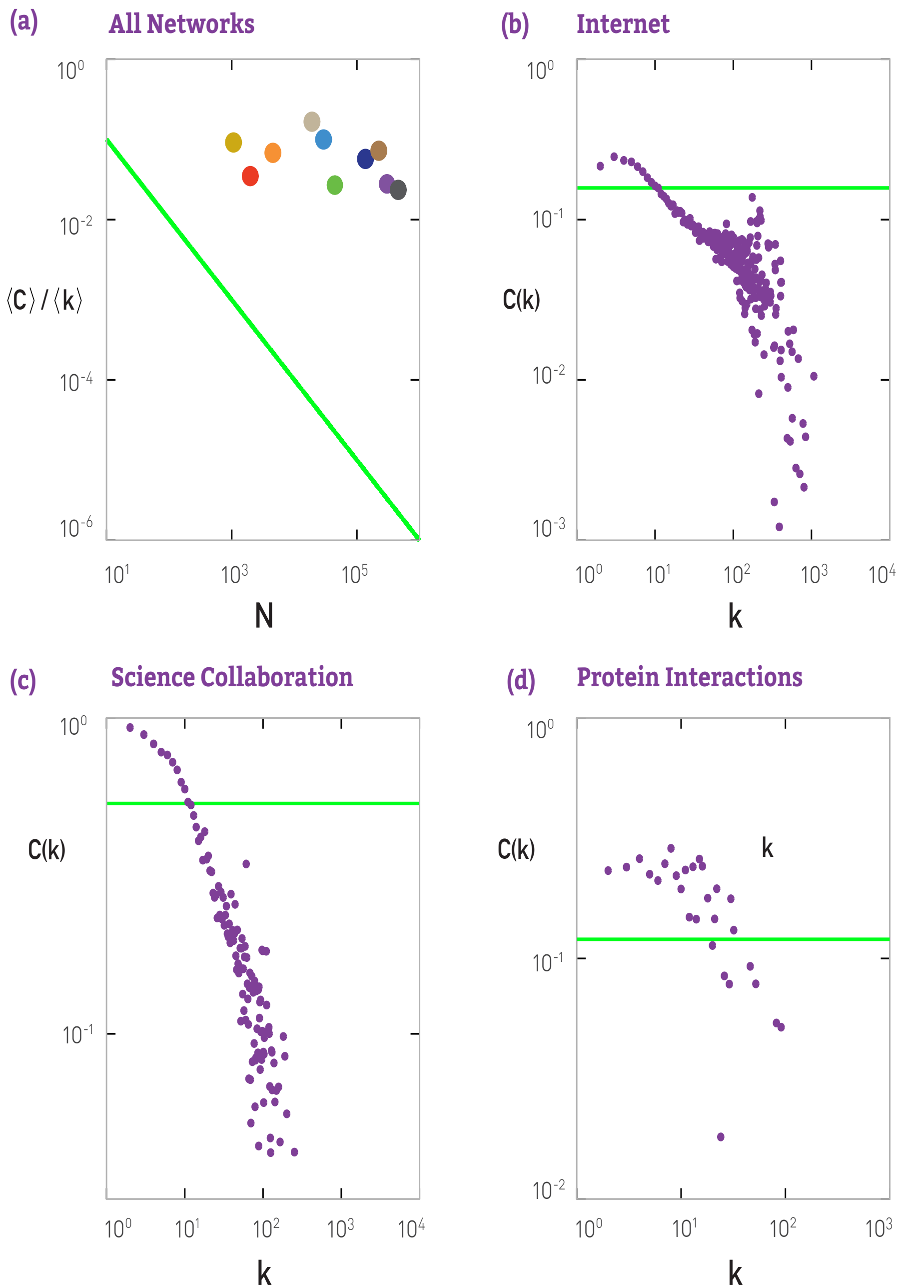


Since edges are independent and have the same probability *p*,

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

- 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*.

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 <k> for a random network, from it we can derive every measurable property. Indeed, we have:

Average path length:

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

Clustering Coefficient:

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

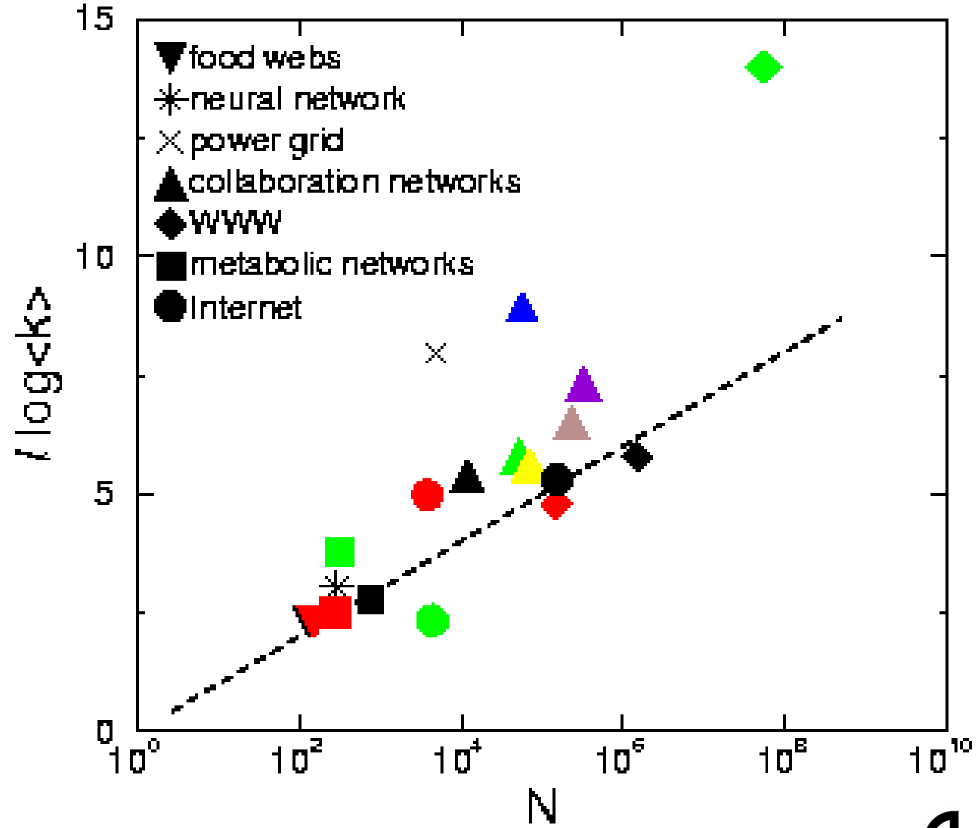
Degree Distribution:

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

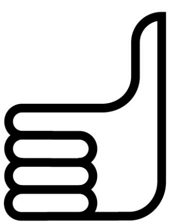
PATH LENGTHS IN REAL NETWORKS

Prediction:

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

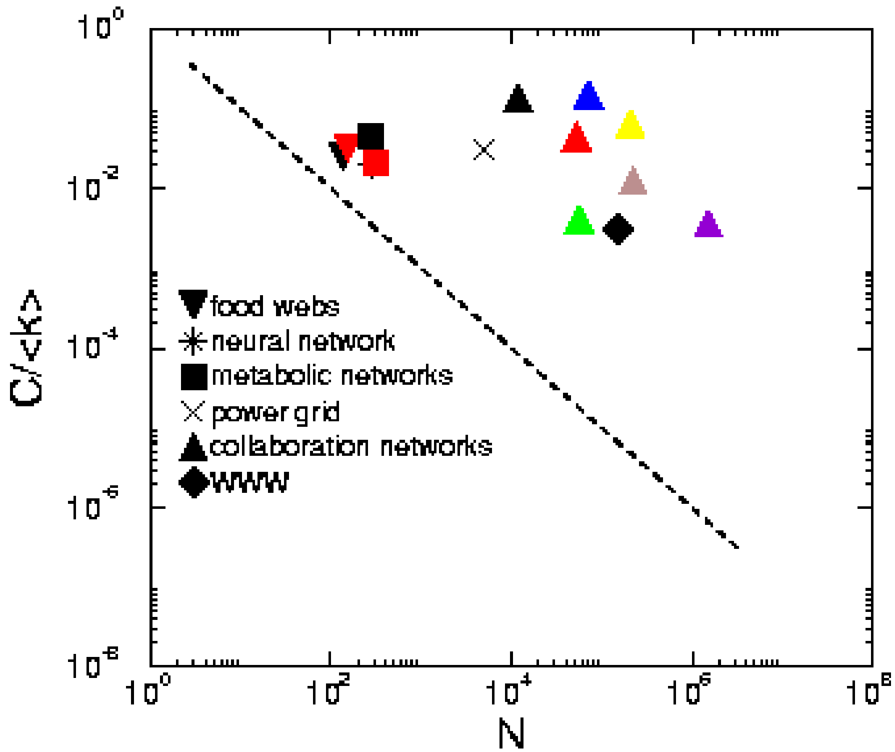


Real networks have short distances like random graphs.

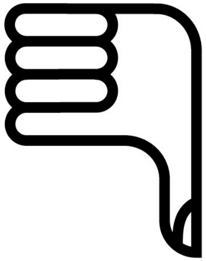


Prediction:

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



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



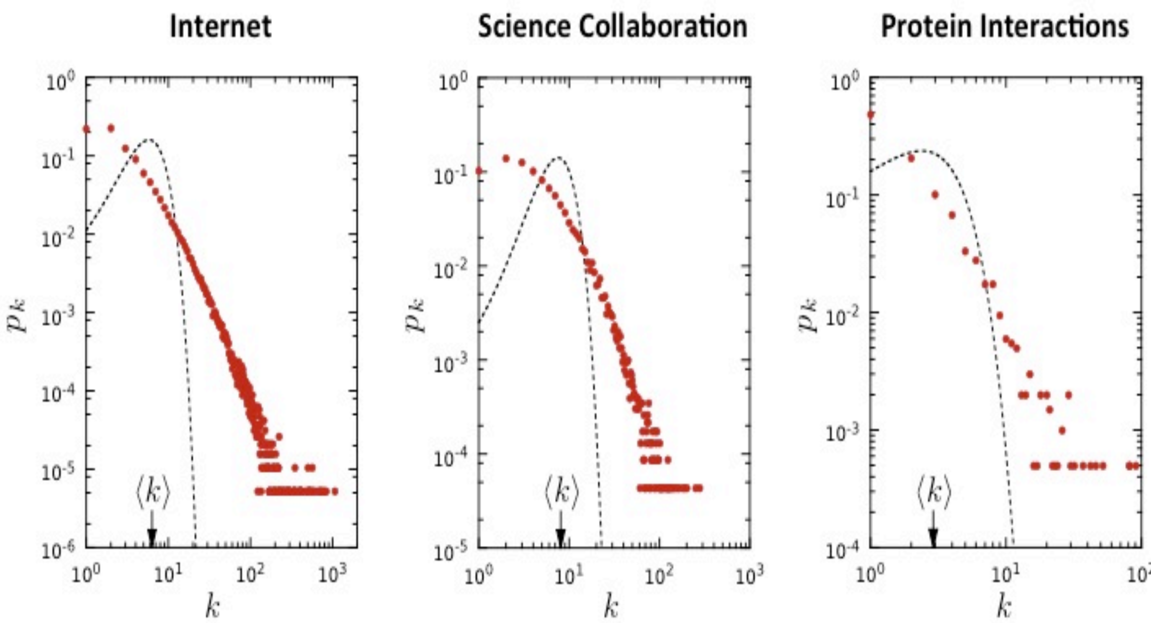
THE DEGREE DISTRIBUTION

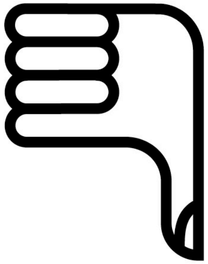
Prediction:

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

Data:

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





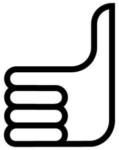
ARE REAL NETWORKS LIKE RANDOM GRAPHS?

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 <k> for a random network, from it we can derive every measurable property. Indeed, we have:

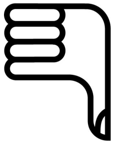
<u>Average path length:</u>

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



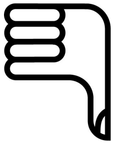
<u>Clustering Coefficient:</u>

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



<u>Degree Distribution:</u>

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



The Watts-Strogatz Model

We start from a ring of nodes, each node being connected to their immediate and next neighbors. Hence initially each node has ‹C› = 3/4 (p = 0).

With probability p each link is rewired to a randomly chosen node. For small p the network maintains high clustering but the random long-range links can drastically decrease the distances between the nodes.

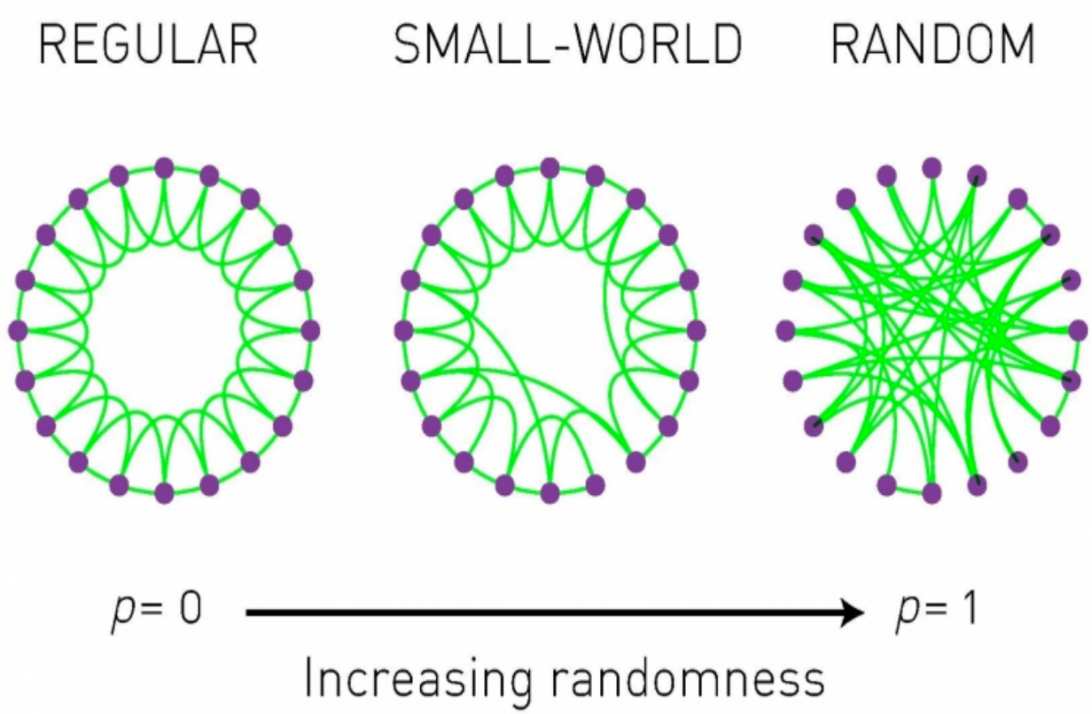
For p = 1 all links have been rewired, so the network turns into a random network.

Regular networks (p=0)

- large distances (bad)
- large clustering coefficients (good)

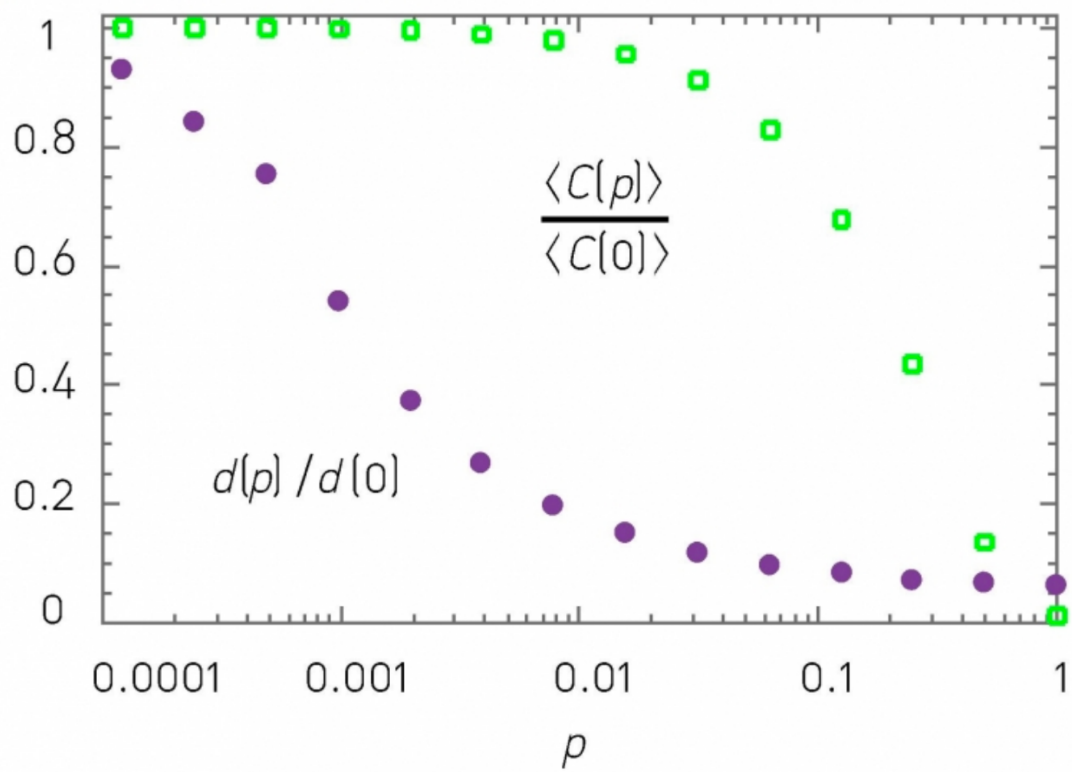
Random networks (p=1):

- small distances (good)
- small clustering coefficients (bad)



Watts-Strogatz Model

All graphs have N=1000 and ‹k›=10.



The dependence of the average path length d(p) and clustering coefficient ‹C(p)› on the rewiring parameter p. Note that d(p) and ‹C(p)› have been normalized by d(0) and ‹C(0)› obtained for a regular lattice (i.e. for p=0 in (a)). The rapid drop in d(p) signals the onset of the small-world phenomenon. During this drop, ‹C(p)› remains high.

Hence in the range 0.001‹p‹0.1 short path lengths and high clustering coexist.

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 extremly USEFUL!

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?