Network Science

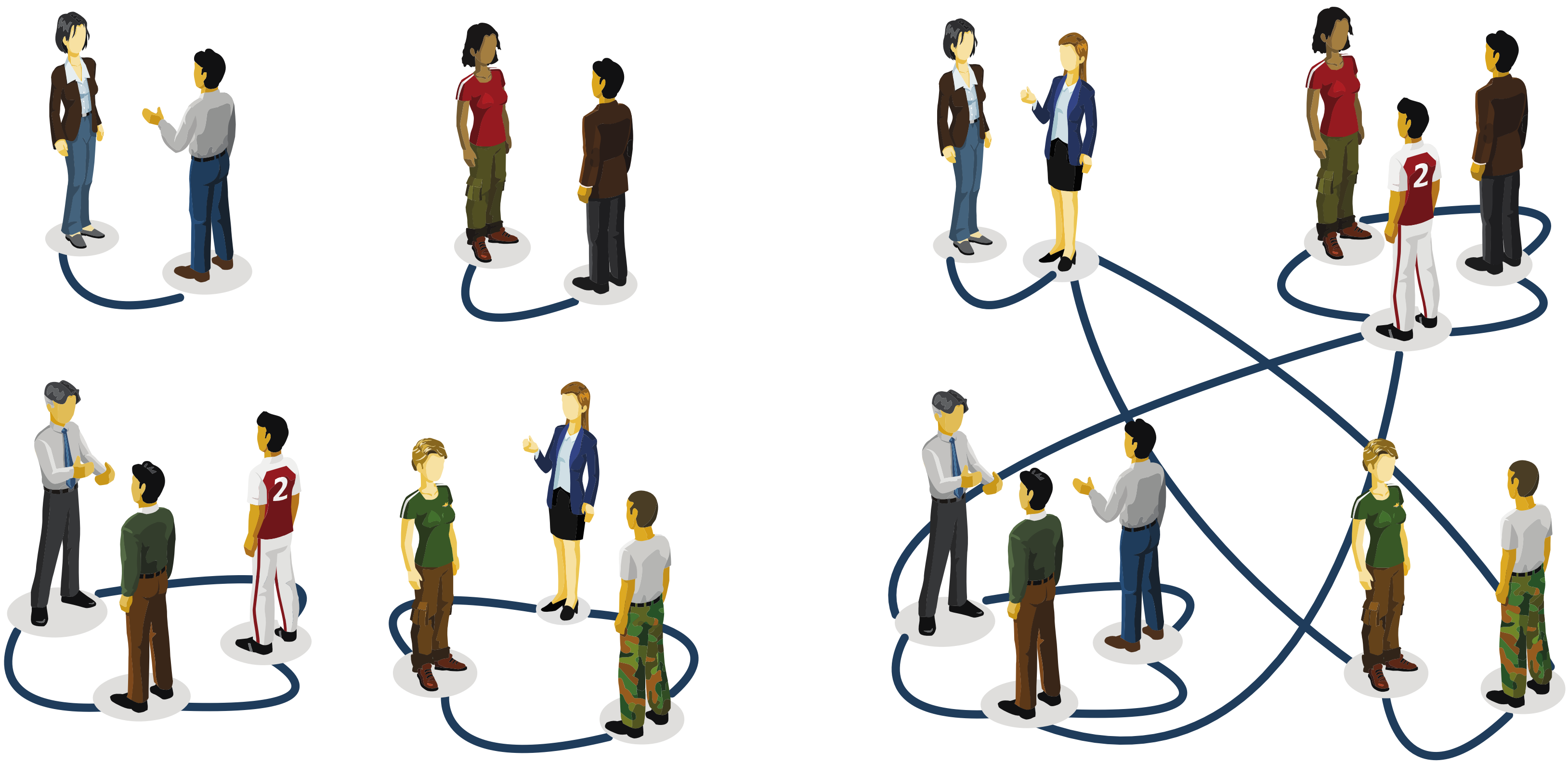
Class 3: Random Networks

Ganesh Bagler

— Adapted from —
Albert-László Barabási
(With Roberta Sinatra)

Introduction

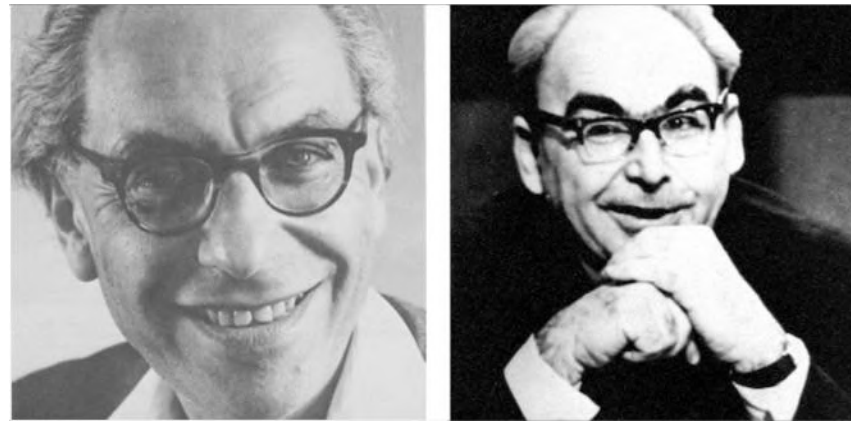
RANDOM NETWORK MODEL



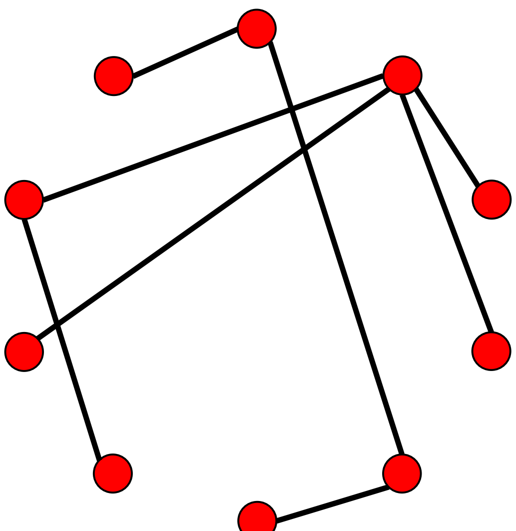
The random network model

RANDOM NETWORK MODEL

Pál Erdös
(1913-1996)



Alfréd Rényi
(1921-1970)



Erdös-Rényi model (1960)

Connect with probability p

p=1/6 N=10

<k> ~ 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**

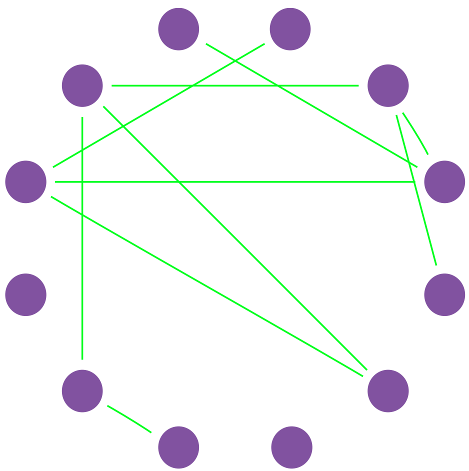
N labeled nodes are connected with L randomly placed links. Erdős and Rényi used this definition in their string of papers on random networks [2-9].

***G*(*N*, *p*) Model**

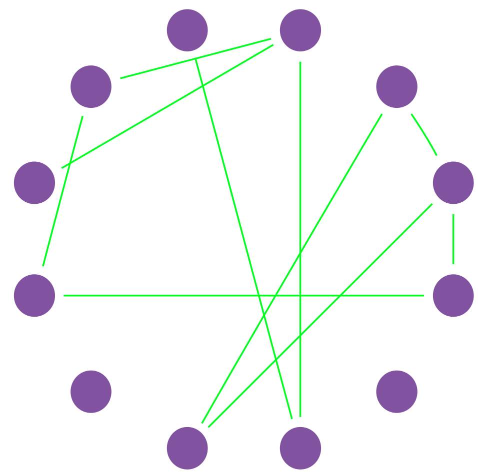
Each pair of N labeled nodes is connected with probability p, a model introduced by Gilbert [10].

RANDOM NETWORK MODEL

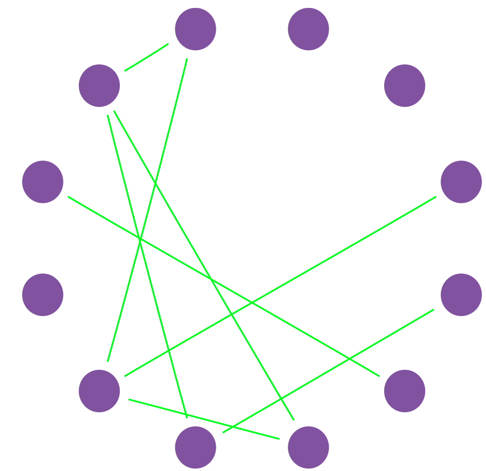
p=1/6
N=12



L=8



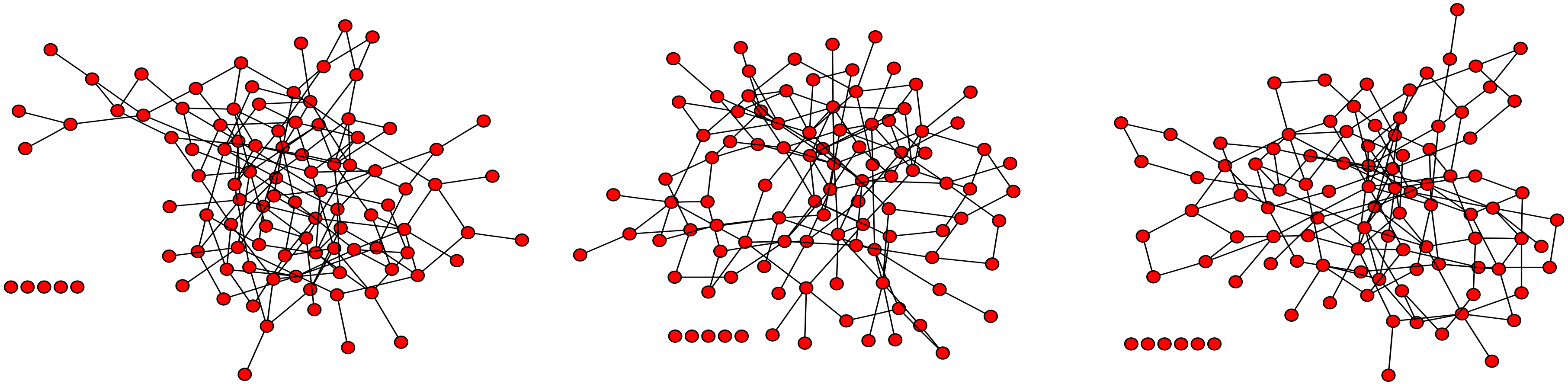
L=10



L=7

RANDOM NETWORK MODEL

p=0.03
N=100



The number of links is variable

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.

$$P(L) = \underbrace{\binom{\binom{N}{2}}{L}}_{\text{Number of different ways we can choose L links among all potential links.}} p^L (1-p)^{\frac{N(N-1)}{2}-L}$$

Binomial distribution...

MATH TUTORIAL | Binomial Distribution: The bottom line

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

$$< x > = Np$$

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

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

http://keral2008.blogspot.com/2008/10/derivation-of-mean-and-variance-of.html

The probability that a random network has exactly L links is the product of three terms;

1. The probability that L of the attempts to connect the $N(N-1)/2$ pairs of nodes have resulted in a link, which is p^L.

2. The probability that the remaining $N(N-1)/2 - L$ attempts have not resulted in a link, which is $(1-p)^{N(N-1)/2 - L}$.

3. A combinatorial factor

$$^{N(N-1)/2}C_L$$

counting the number of different ways we can place L links among $N(N-1)/2$ node pairs.

We can therefore write the probability that a particular realization of a random network has exactly L links as

$$\boxed{p_L = {}^{N(N-1)}C_L \cdot p^L \cdot (1-p)^{N(N-1)/2 - L}} \quad \text{— (3.1)}$$

RANDOM NETWORK MODEL

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

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

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

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

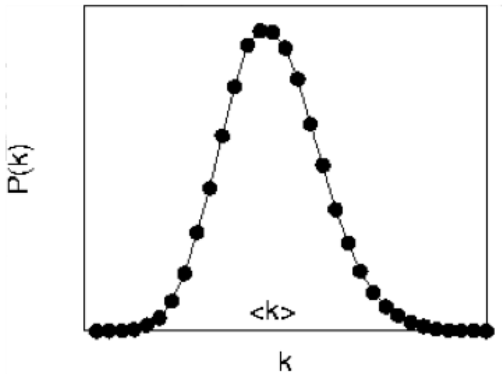
$$\langle k \rangle = 2L/N = p(N-1)$$

- 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

$$\langle k \rangle = p(N-1) \qquad \sigma_k^2 = p(1-p)(N-1)$$

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

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 <k>.

DEGREE DISTRIBUTION OF A RANDOM GRAPH

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

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

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

$$\ln[(1-p)^{(N-1)-k}]=(N-1-k)\ln(1-\frac{\langle k\rangle}{N-1})=-(N-1-k)\frac{\langle k\rangle}{N-1}=-\langle k\rangle(1-\frac{k}{N-1})\approx-\langle k\rangle$$

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

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

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:

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

DERIVING THE POISSON DISTRIBUTION

In a random network, the probability that node i has exactly k links is given by:

$$p_k = {}^{N-1}C_k \; p^k \; (1-p)^{N-1-k} \qquad \text{— (3.7)/(3.22)}$$

The number of ways we can select k links from $N-1$ potential links that a node can have.

The probability that k of its links are present.

The probability that the remaining $(N-1-k)$ links are missing.

FIRST TERM: We can rewrite the first term as,

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

$$\approx \frac{(N-1)^k}{k!} \qquad \text{— (3.23)}$$

LAST TERM: Since, $\ln\left[(1-p)^{(N-1)-k}\right] = (N-1-k)\ln\left(1-\frac{\langle k\rangle}{N-1}\right)$

Using series expansion,

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} \ldots \;\forall\; |x| \leq 1,$$

we obtain,

$$\ln\left[(1-p)^{N-1-k}\right] \approx (N-1-k)\,\frac{\langle k\rangle}{N-1}$$

$$= -\langle k\rangle\left[1-\frac{k}{N-1}\right] \approx -\langle k\rangle$$

$$\therefore\; (1-p)^{N-1-k} = e^{-\langle k\rangle} \qquad \text{— (3.24)}$$

Therefore, the expected number of links in a random graph is

$$\langle L\rangle = \sum_{L=0}^{\frac{N(N-1)}{2}} L \cdot p_L = p \cdot \frac{N(N-1)}{2} \qquad \text{— (3.2)}$$

L_{max} is given by $\langle L\rangle$ for maximum value of p,

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

Average degree of a random graph is given by,

$$\langle k\rangle = \frac{2\langle L\rangle}{N} = \frac{2}{N}\cdot\frac{N(N-1)}{2}\cdot p \quad \text{(using 3.2)}$$

$$\langle k\rangle = p\cdot(N-1) \qquad \text{— (3.3)}$$

Substituting (3.23) & (3.24) in (3.22), we get,

$$p_k = {}^{N-1}C_k \cdot p^k \cdot (1-p)^{N-1-k}$$

$$p_k = \frac{(N-1)}{k!} \cdot p^k \cdot e^{-\langle k\rangle}$$

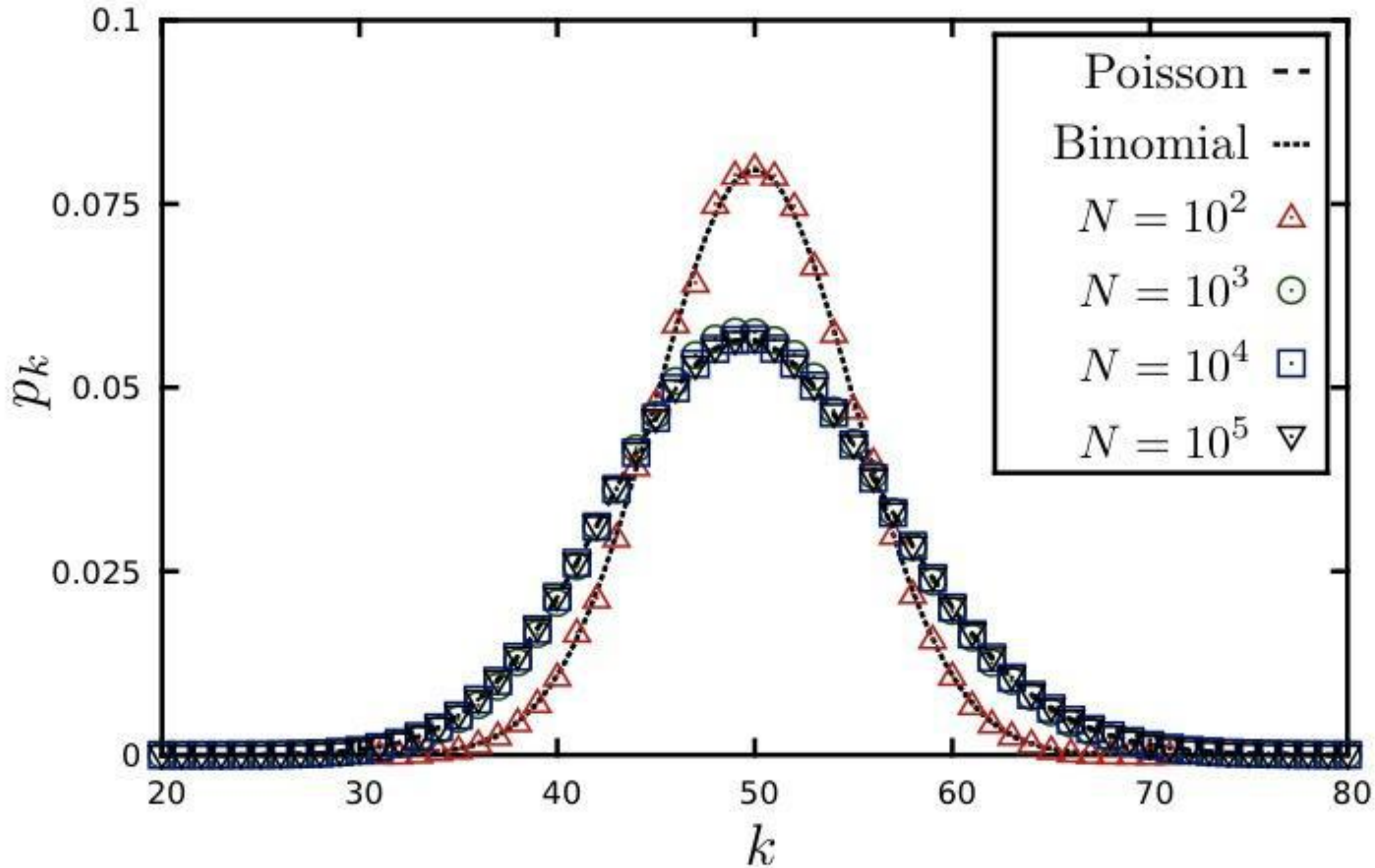
$$p_k = \frac{(N-1)^k}{k!}\left(\frac{\langle k\rangle}{N-1}\right)^k e^{-\langle k\rangle}$$

$$\boxed{p_k = e^{-\langle k\rangle}\,\frac{\langle k\rangle^k}{k!}} \qquad \text{— (3.25)}$$

DEGREE DISTRIBUTION OF A RANDOM GRAPH

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

<k>=50

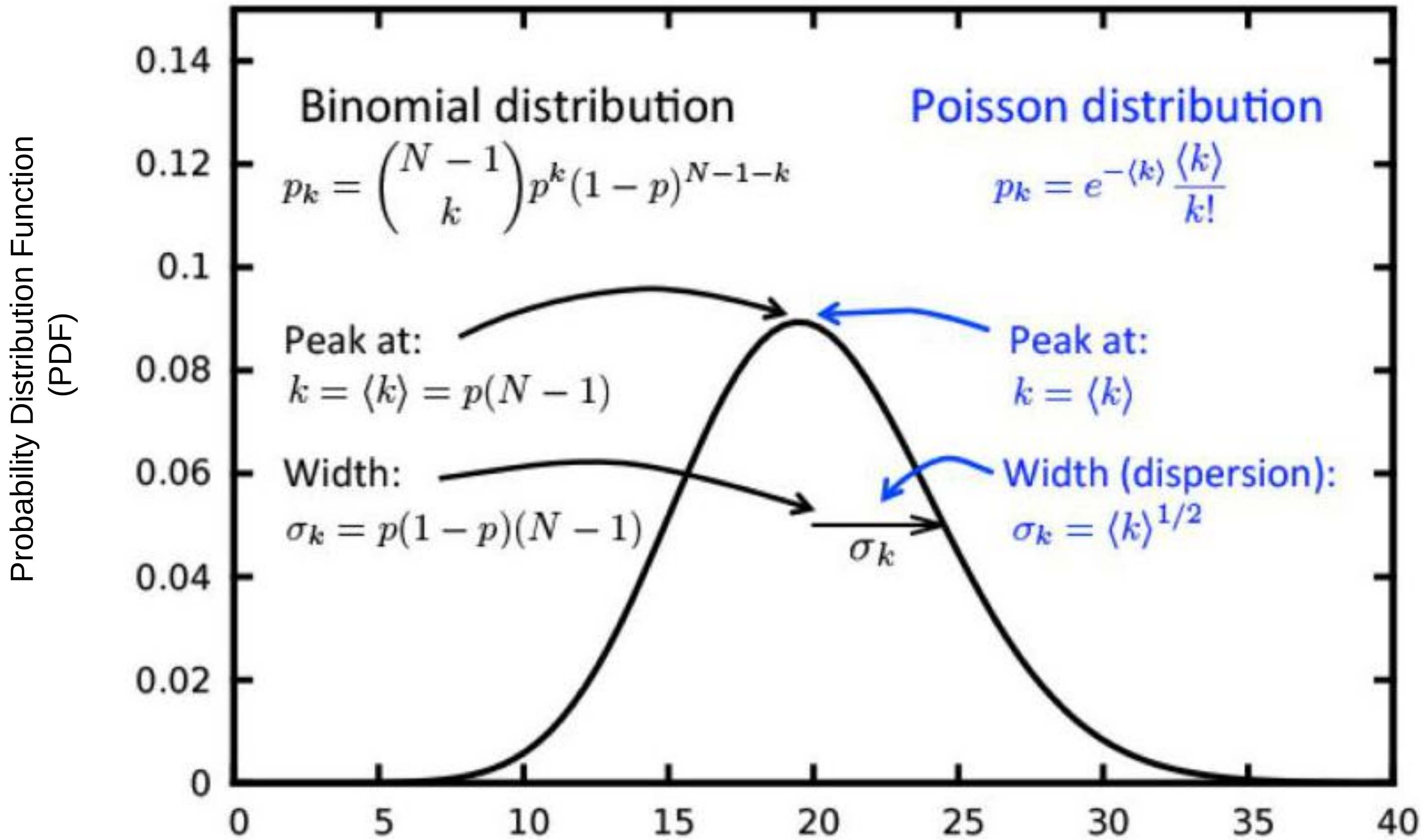


K

DEGREE DISTRIBUTION OF A RANDOM NETWORK

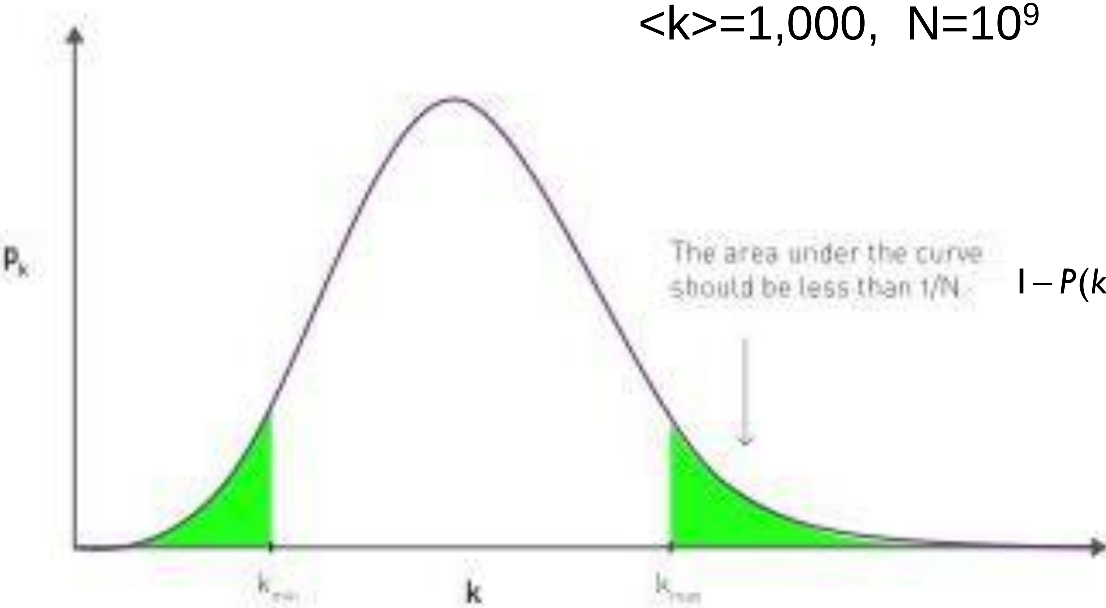
Exact Result -binomial distribution-

Large N limit -Poisson distribution-



Real Networks are not Poisson

Section 3.5 Maximum and minimum degree



$<k>=1,000, \quad N=10^9$

$$N\left[1-P(k_{max})\right] \approx 1.$$

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

$<k>=1,000, \quad N=10^9$

$k_{max}=1,185$

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

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

$k_{min}=816$

$$<k> \pm \sigma_k \qquad \sigma_k = <k>^{1/2}$$

$$\sigma_k = 31.62.$$

NO OUTLIERS IN A RANDOM SOCIETY

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

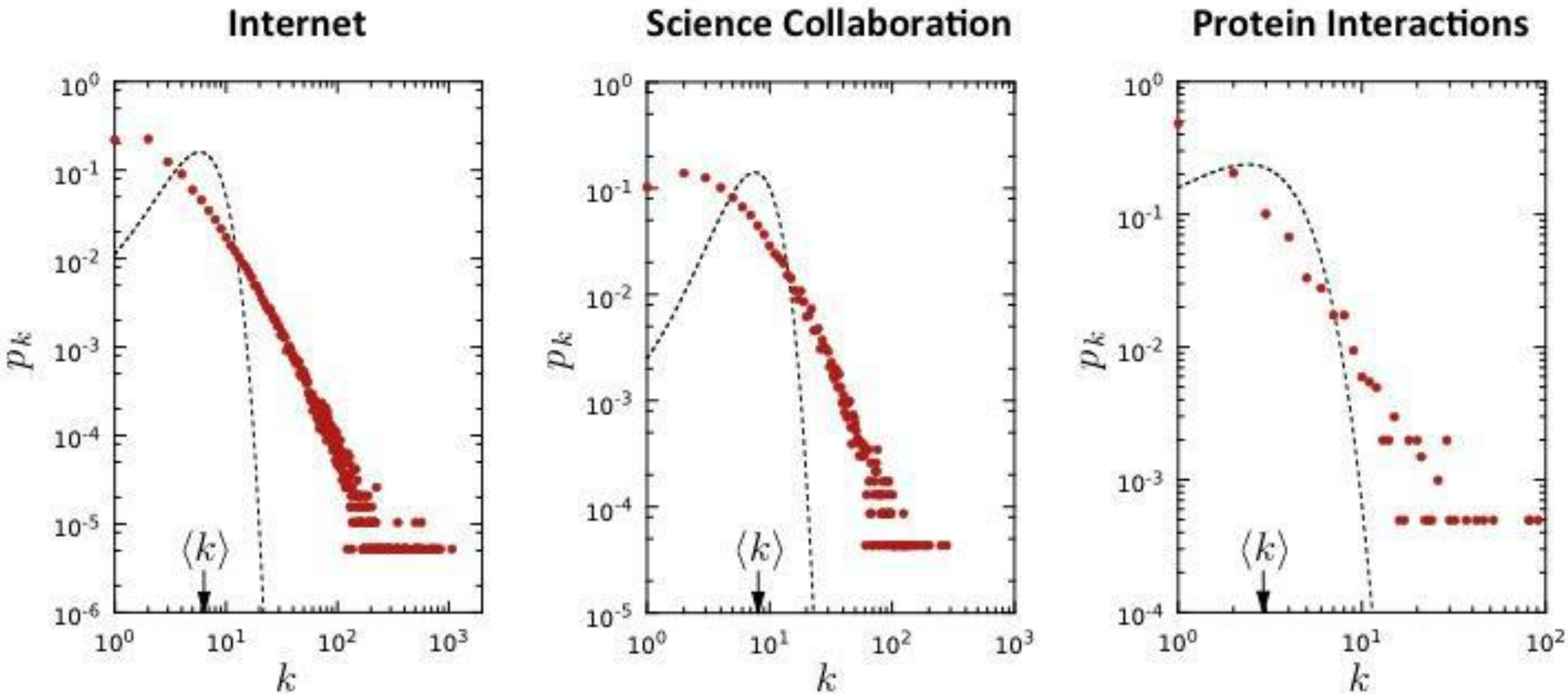
→The most connected individual has degree k_{max}~1,185
→The least connected individual has degree k_{min} ~ 816

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.

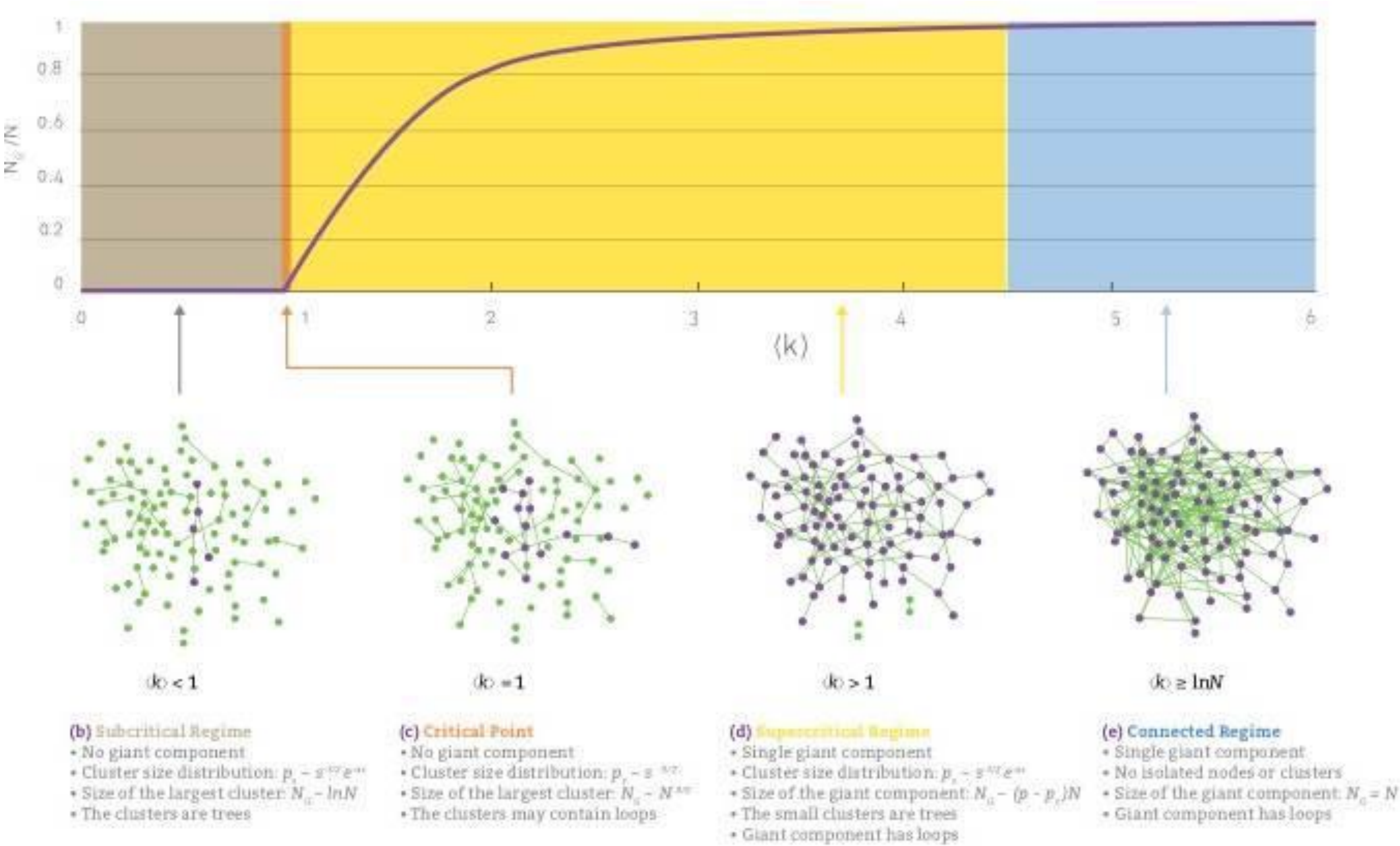
→a random society would consist of mainly average individuals, with everyone with roughly the same number of friends.
→It would lack outliers, individuals that are either highly popular or recluse.

FACING REALITY: Degree distribution of real networks

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



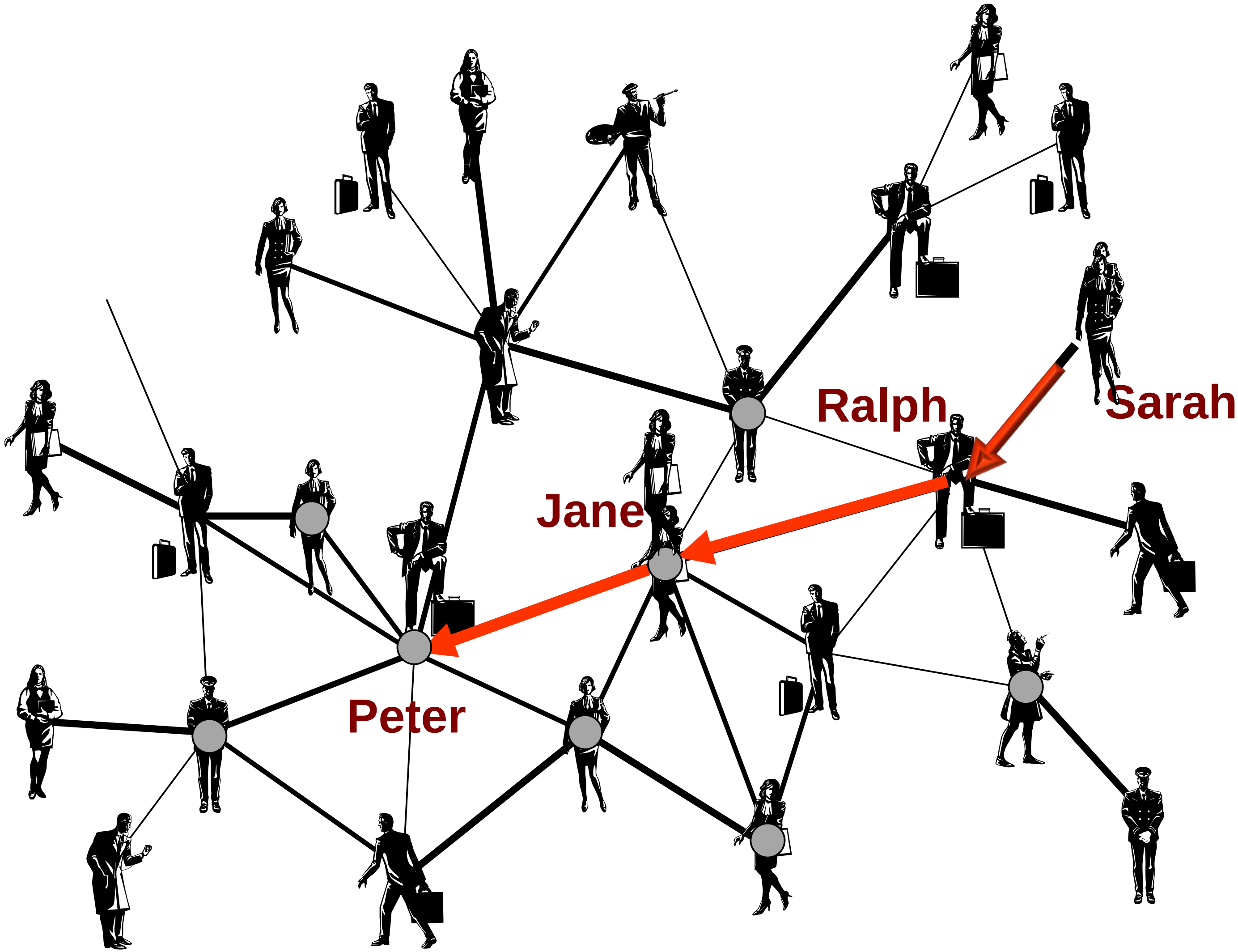
The evolution of a random network



1
0.8
0.6
0.4
0.2
0
N_G/N
0
1
2
3
4
5
6
$\langle k \rangle$
$\langle k \rangle < 1$
$\langle k \rangle = 1$
$\langle k \rangle > 1$
$\langle k \rangle \geq \ln N$
(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

Small worlds



Frigyes Karinthy, 1929
Stanley Milgram, 1967

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."

SIX DEGREES | 1967: Stanley Milgram

HOW TO TAKE PART IN THIS STUDY

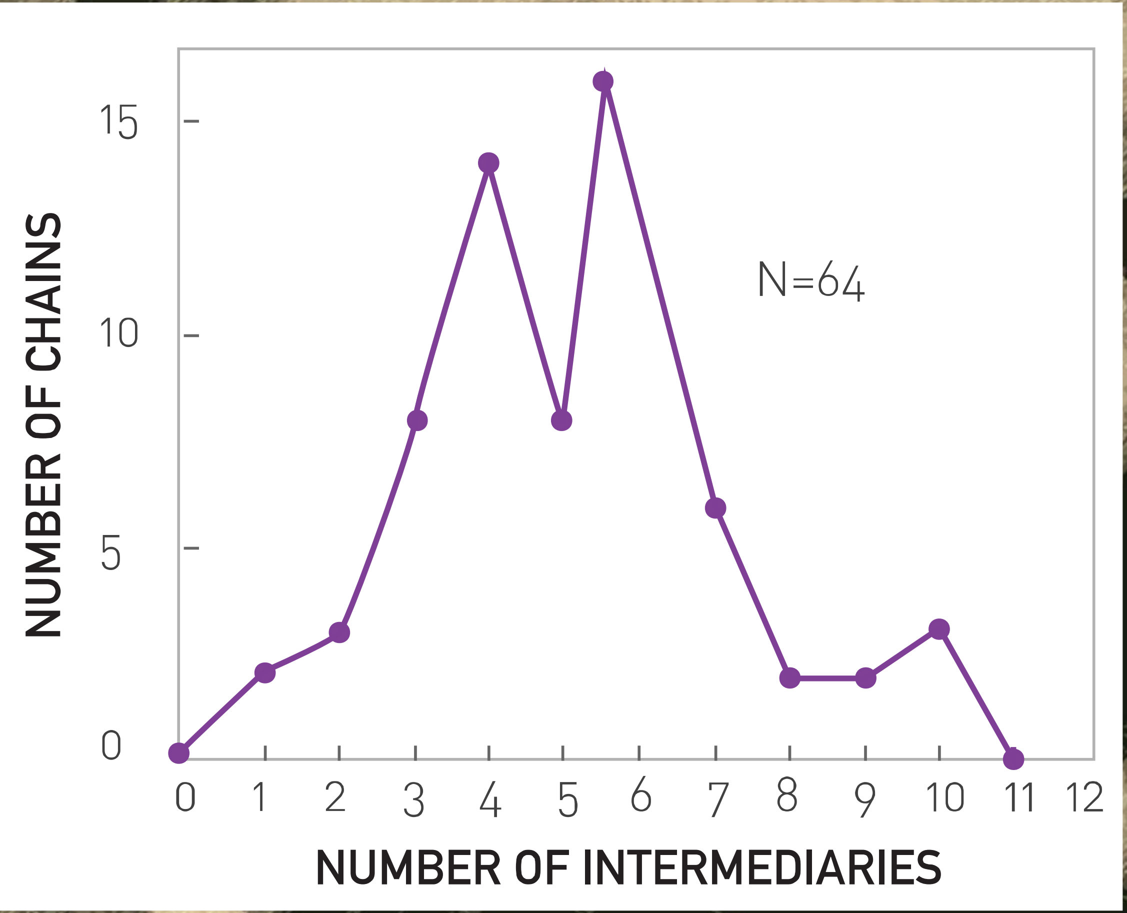
1. ADD YOUR NAME TO THE ROSTER AT THE BOTTOM OF THIS SHEET, so that the next person who receives this letter will know who it came from.

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.

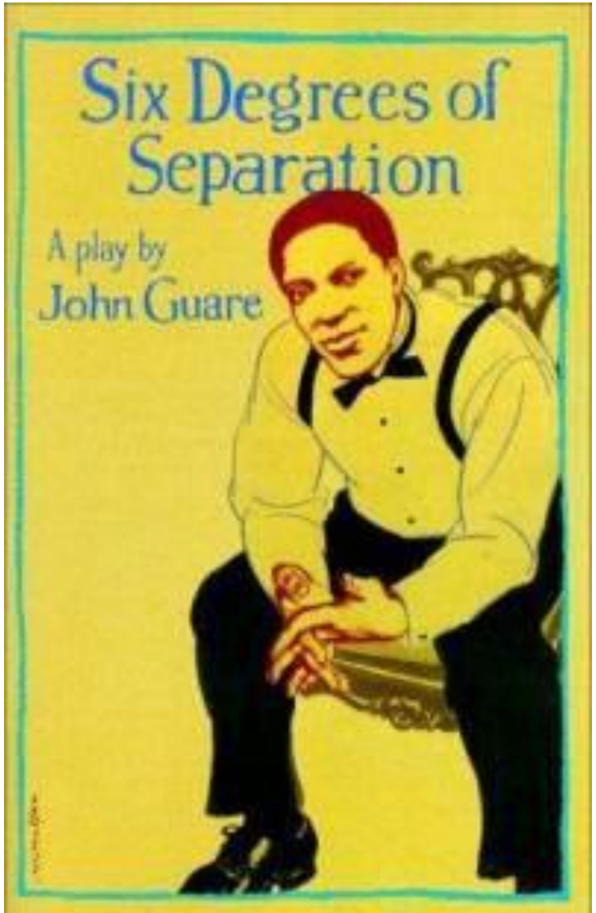
3. IF YOU KNOW THE TARGET PERSON ON A PERSONAL BASIS, MAIL THIS FOLDER DIRECTLY TO HIM (HER). Do this only if you have previously met the target person and know each other on a first name basis.

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.

SIX DEGREES | 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."

WWW: 19 DEGREES OF SEPARATION

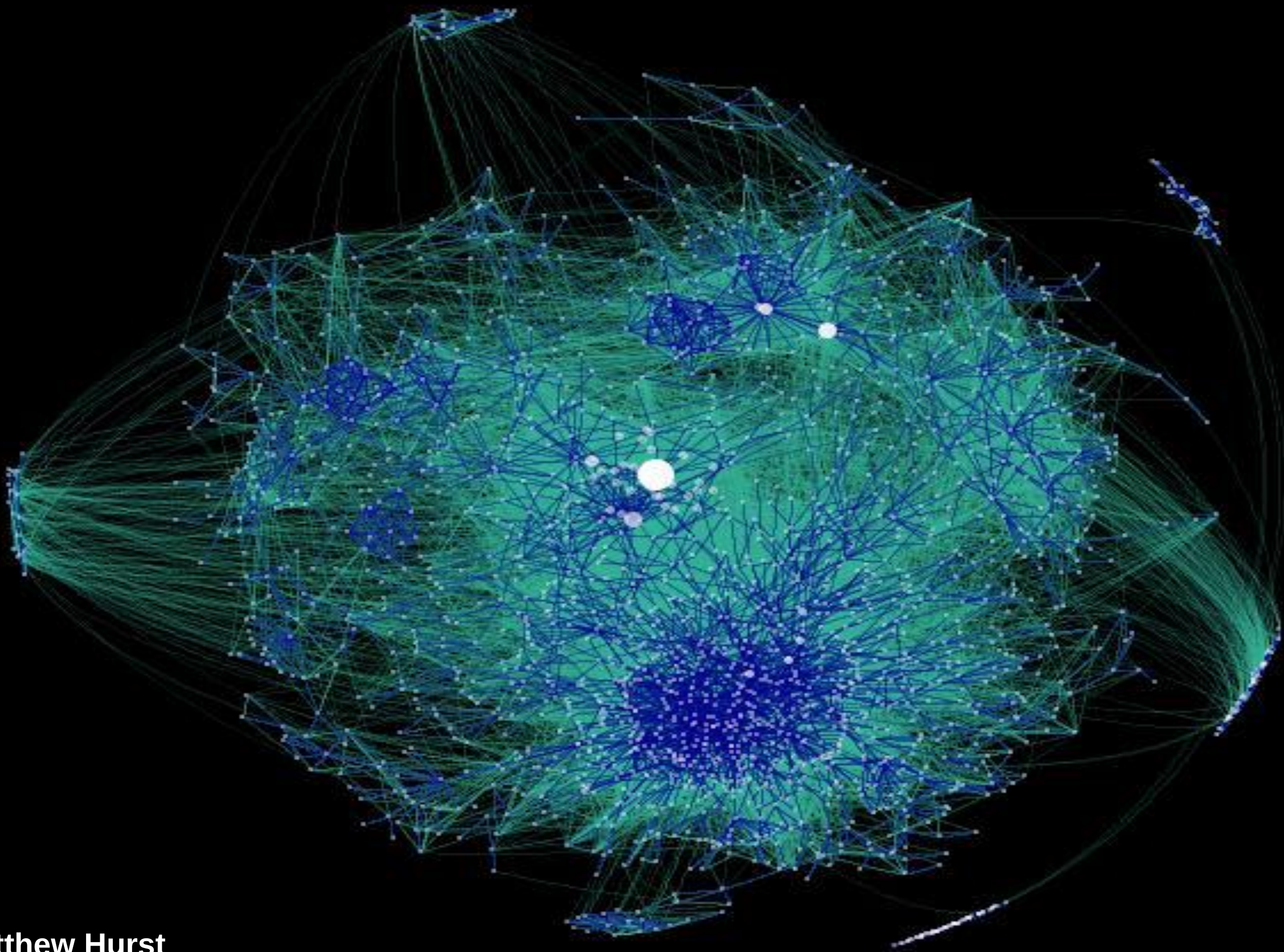
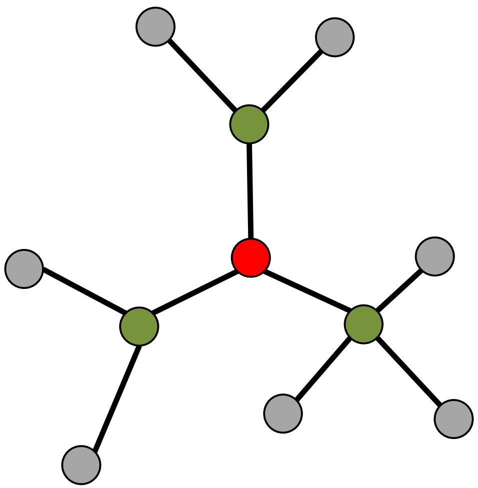


Image by **Matthew Hurst**
Blogosphere

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.

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

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, ⟨d⟩, 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 ⟨d⟩ is proportional to log N, rather than N.

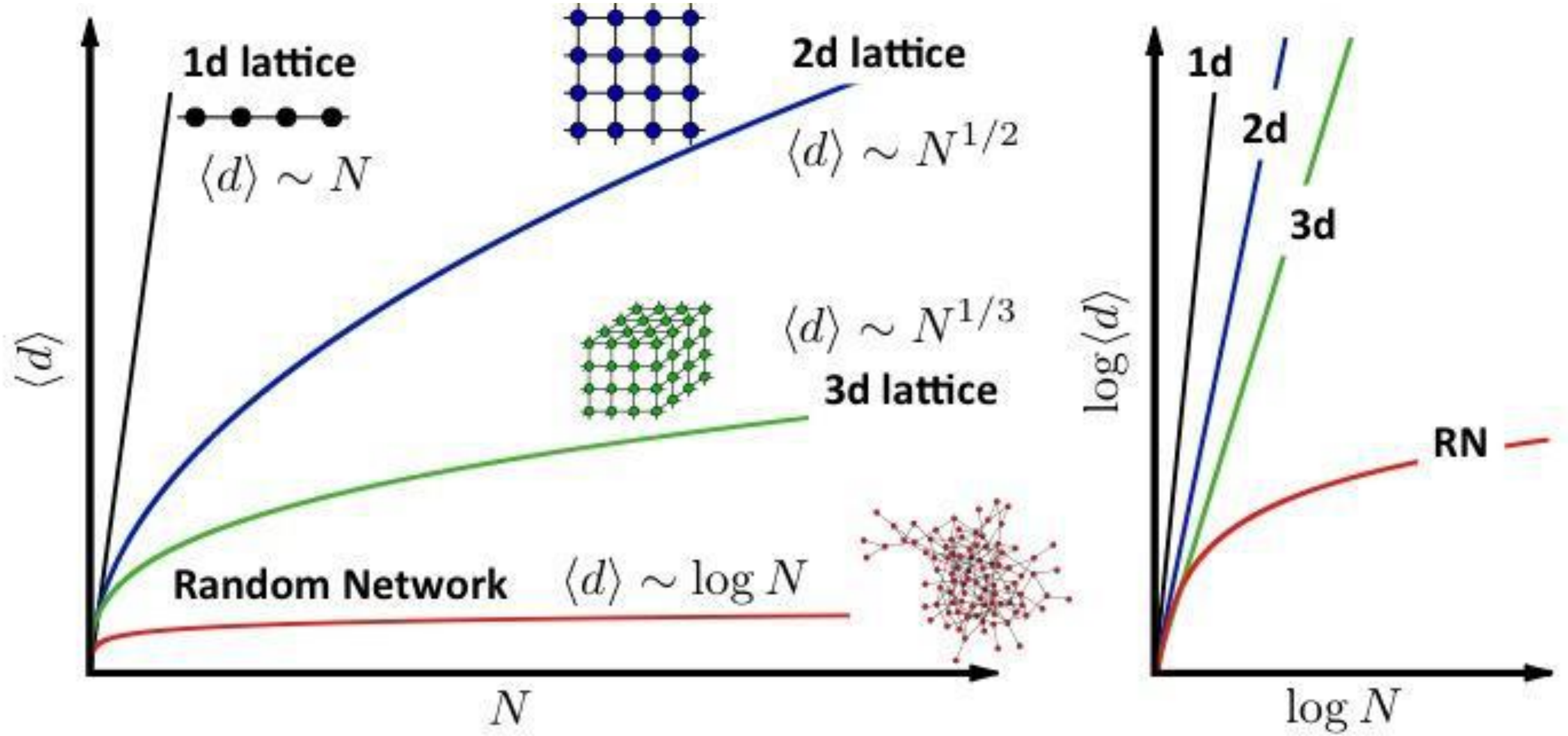
The 1/log⟨k⟩ term implies that denser the network, the smaller will be the distance between the nodes.

DISTANCES IN RANDOM GRAPHS | compare with real data

Network	N	L	‹k›	‹d›	d_{max}	$\ln N/\ln\langle k\rangle$
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	93,437	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?



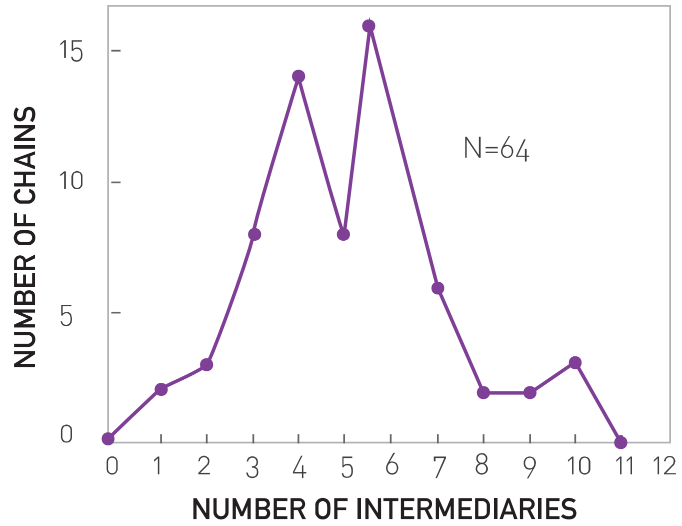
Three, Four or Six Degrees?

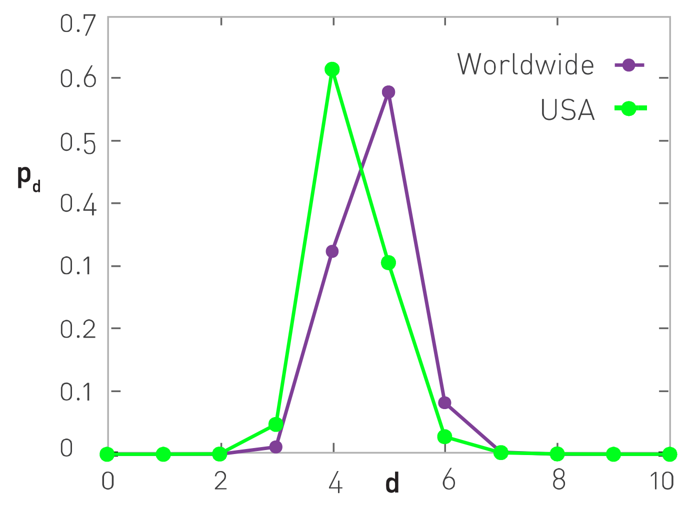
For the globe's social networks:

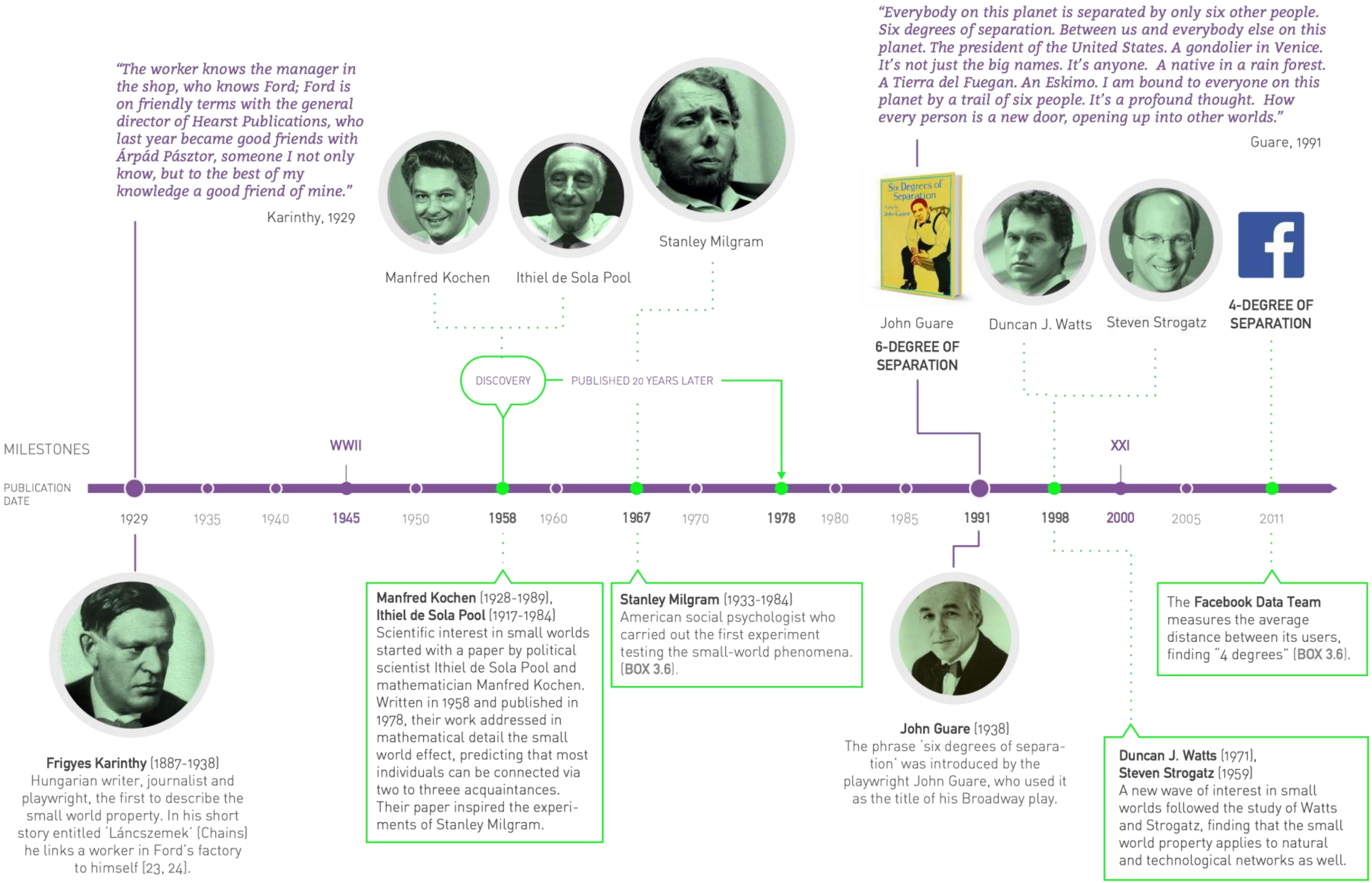
$\langle k \rangle \simeq 10^3$

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

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



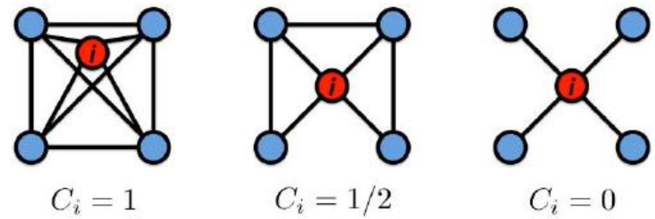



"The worker knows the manager in the shop, who knows Ford; Ford is on friendly terms with the general director of Hearst Publications, who last year became good friends with Árpád Pásztor, someone I not only know, but to the best of my knowledge a good friend of mine."
Karinthy, 1929
Manfred Kochen
Ithiel 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.
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.
The **Facebook Data Team** measures the average distance between its users, finding "4 degrees" (BOX 3.6).

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*,

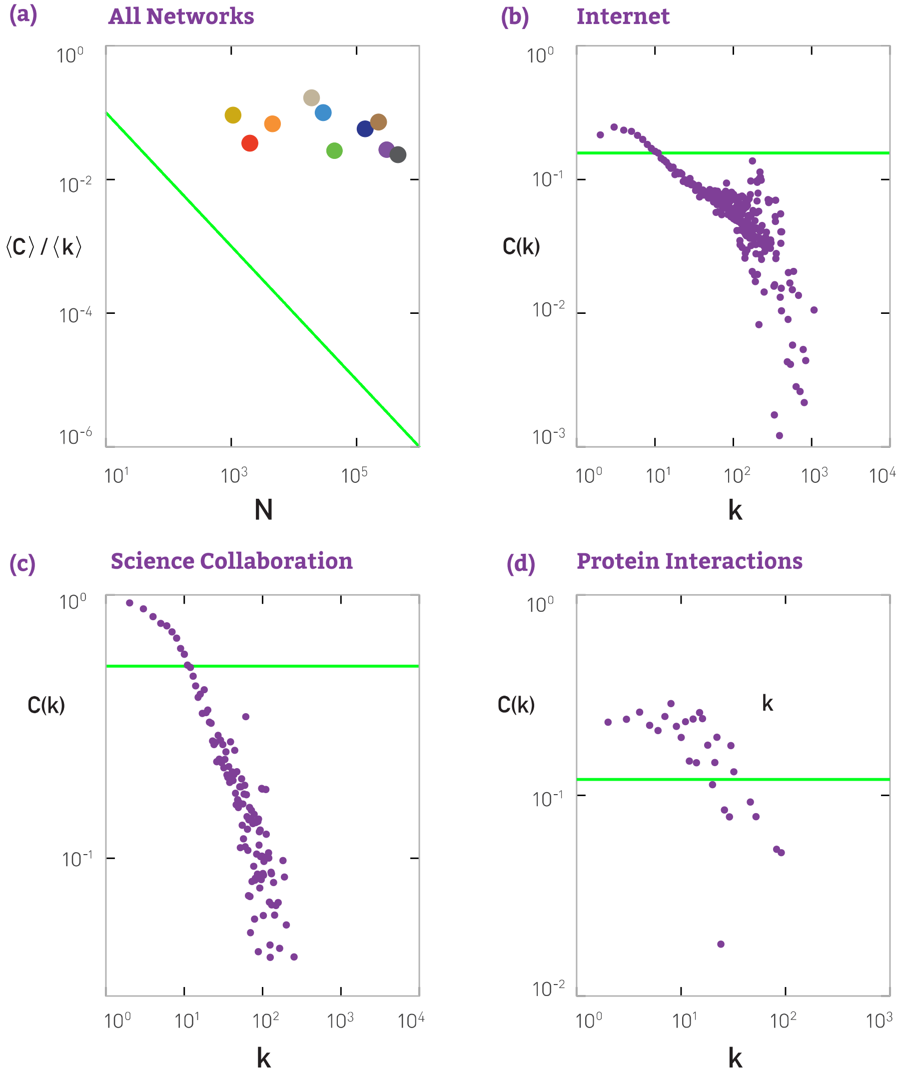
$$\langle L_i \rangle \cong p\,\frac{k_i(k_i - 1)}{2} \quad \Rightarrow \quad C_i = \frac{2\langle L_i \rangle}{k_i(k_i - 1)} = p = \frac{\langle k \rangle}{N}.$$

•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

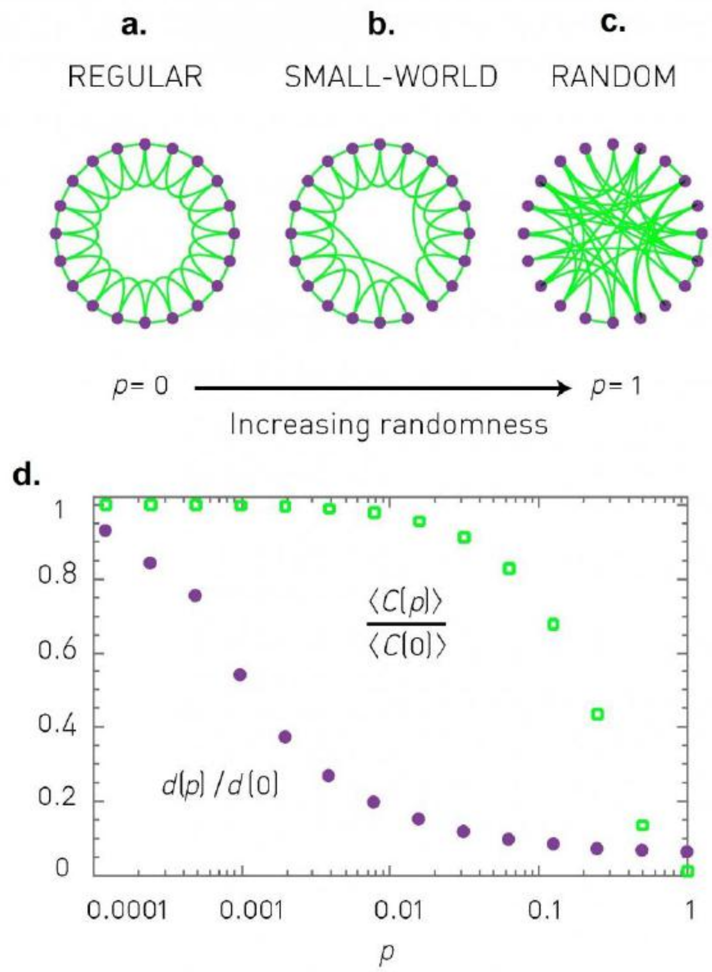


$$C_i = \frac{2\langle L_i \rangle}{k_i(k_i - 1)} = p = \frac{\langle k \rangle}{N}.$$

C decreases with the system size *N*.

C is independent of a node's degree *k*.

Watts-Strogatz Model



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

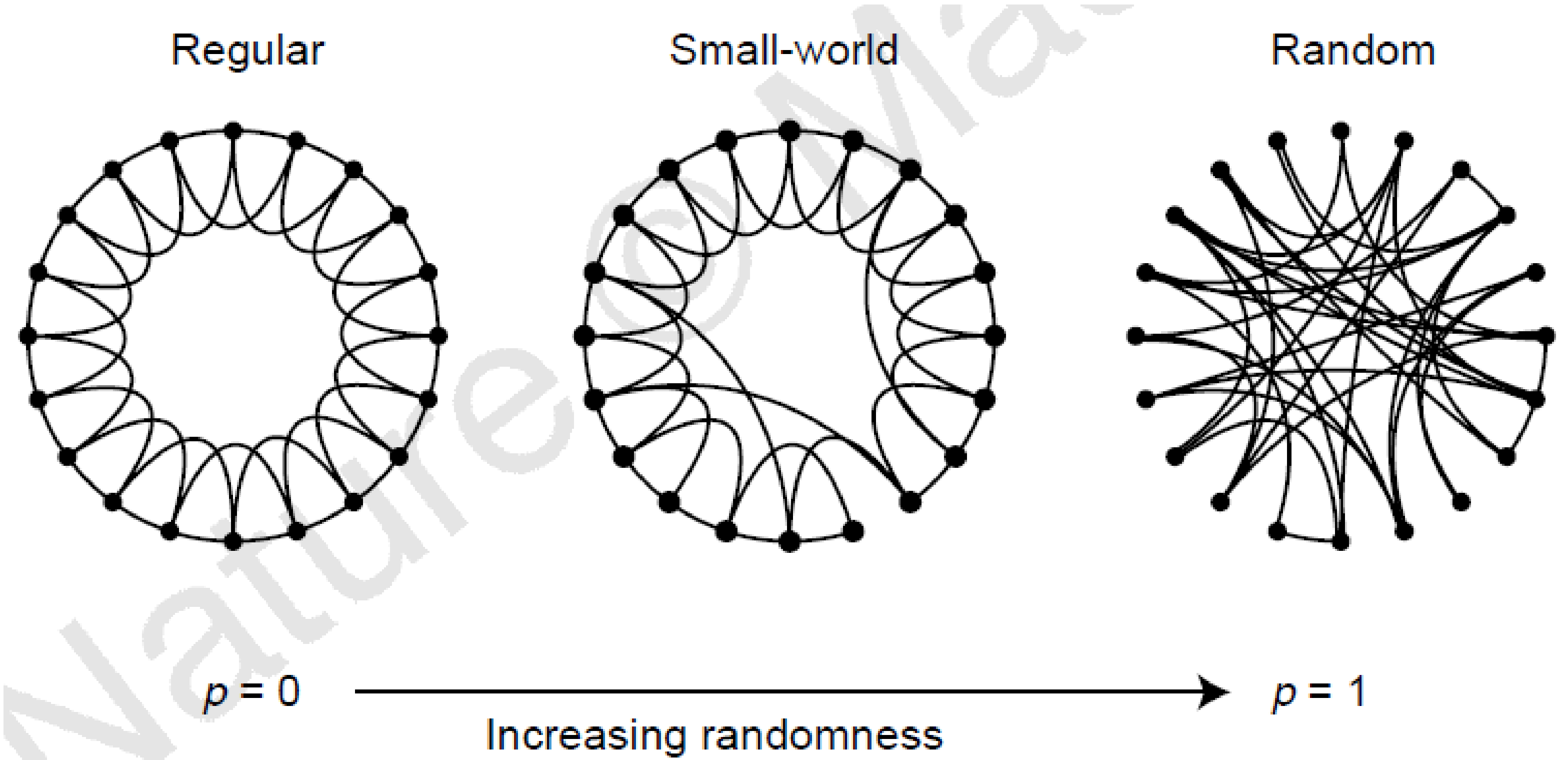
Small-World Networks

- Many self-organizing systems could be viewed as networks of coupled dynamical systems.
- In such 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.
- Simple models of networks that can be tuned through this middle ground: regular networks 'rewired' to introduce increasing amounts of disorder.
- These systems can be highly clustered, like regular lattices, yet have small characteristic path lengths, like random graphs.

→**'small-world' networks**

"Collective dynamics of small-world networks", Watts and Strogatz,

Small -World Networks

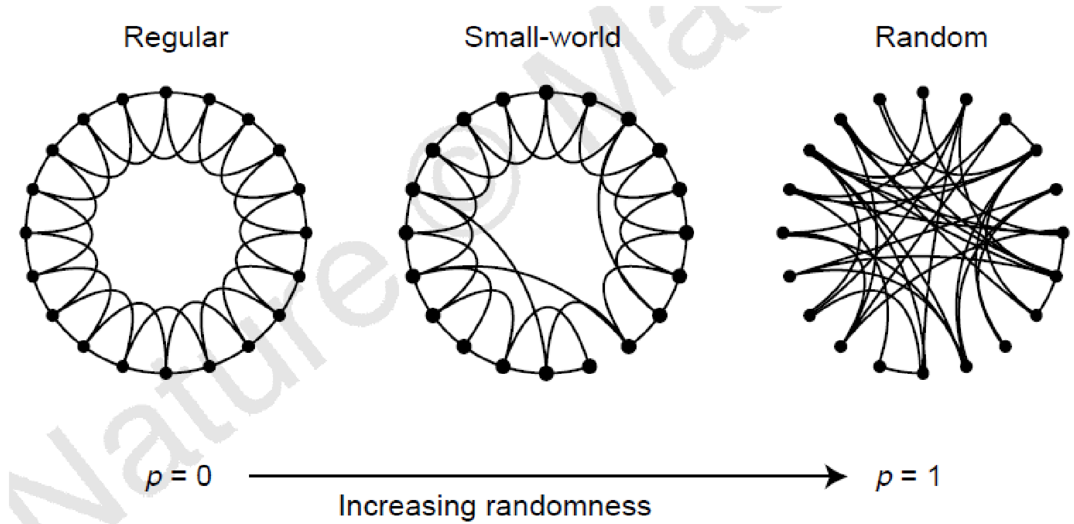


“Collective dynamics of small-world networks”, Watts and Strogatz,

Procedure used by Watts and Strogatz (Nature, 1998)

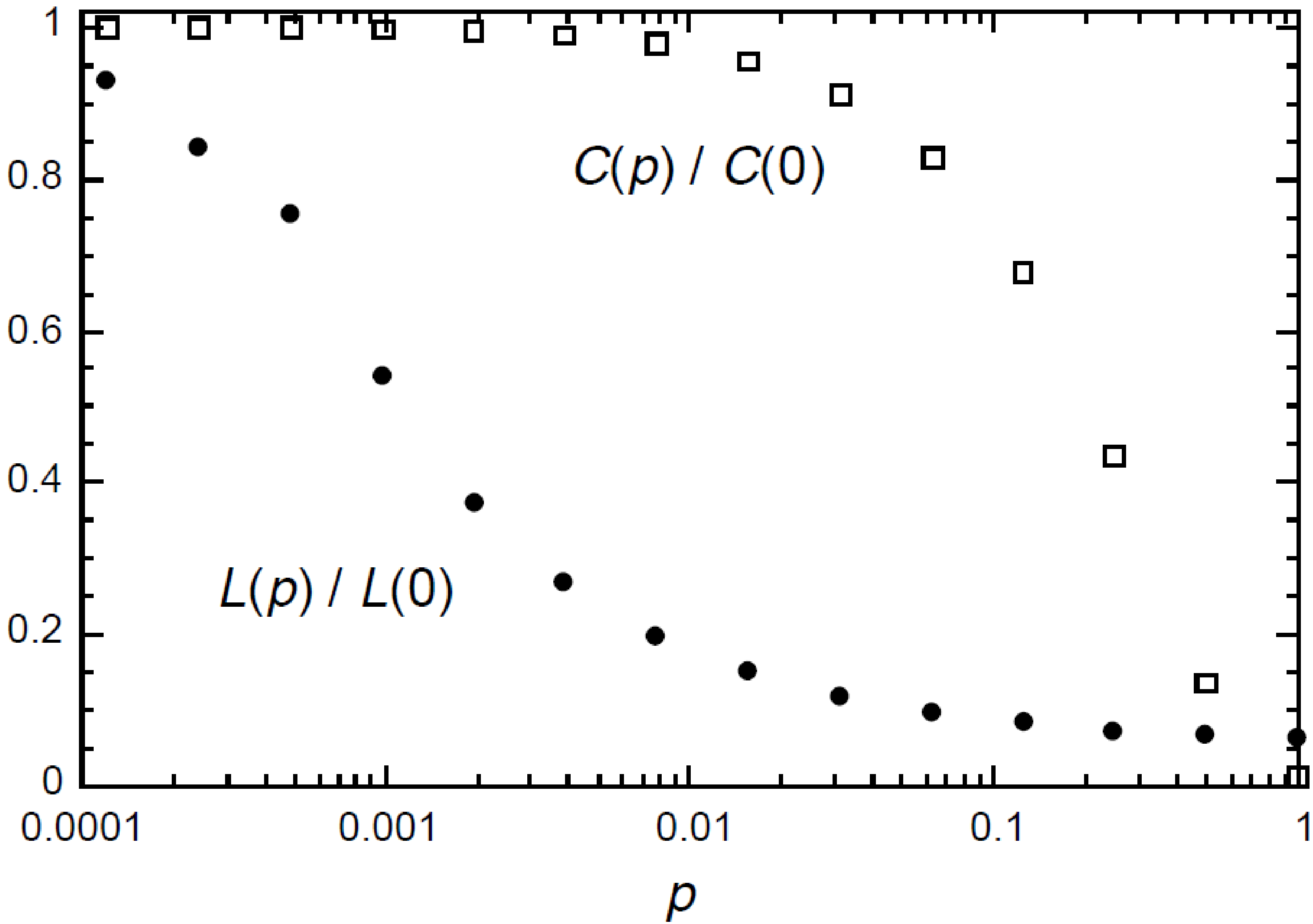
- Random rewiring procedure for interpolating between a regular ring lattice and a random network, without altering the number of vertices or edges in the graph.
- Start with a ring of n vertices, each connected to its k nearest neighbours by undirected edges. (For clarity, $n = 20$ and $k = 4$ is used.)
- Choose a vertex and the edge that connects it to its nearest neighbour in a clockwise sense. With probability p, reconnect this edge to a vertex chosen uniformly at random over the entire ring, with duplicate edges forbidden; otherwise leave the edge in place.
- Repeat this process by moving clockwise around the ring, considering each vertex in turn until one lap is completed.
- Next, consider the edges that connect vertices to their second-nearest neighbours clockwise. As before, randomly rewire each of these edges with probability p, and continue this process, circulating around the ring and proceeding outward to more distant neighbours after each lap, until each edge in the original lattice has been considered once.
- As there are $nk/2$ edges in the entire graph, the rewiring process stops after $k/2$ laps.

"Collective dynamics of small-world networks", Watts and Strogatz,

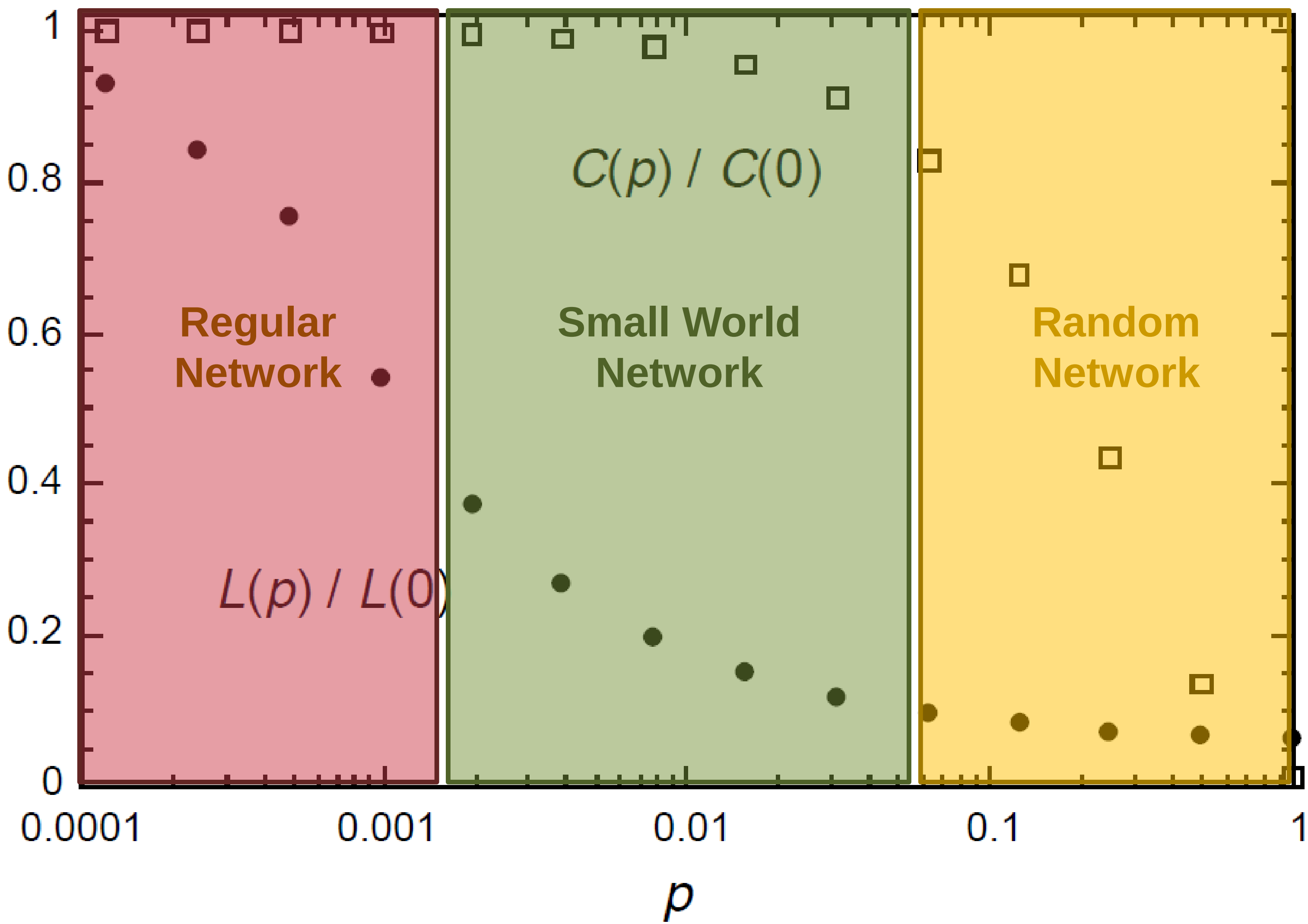


- Above figure shows three realizations of this process, for different values of *p*.
- For *p* = 0, the original ring is unchanged; as *p* increases, the graph becomes increasingly disordered until for *p* = 1, all edges are rewired randomly.
- One of the main results is that for intermediate values of *p*, the graph is a small-world network: highly clustered like a regular graph, yet with small characteristic path length, like a random graph.

“Collective dynamics of small-world networks”, Watts and Strogatz,



“Collective dynamics of small-world networks”, Watts and Strogatz,



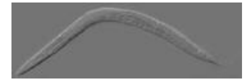
“Collective dynamics of small-world networks”, Watts and Strogatz,

- **The neural network of the worm *Caenorhabditis elegans***

- **The power grid of the western United States**

- **The collaboration graph of film actors**

C. elegans



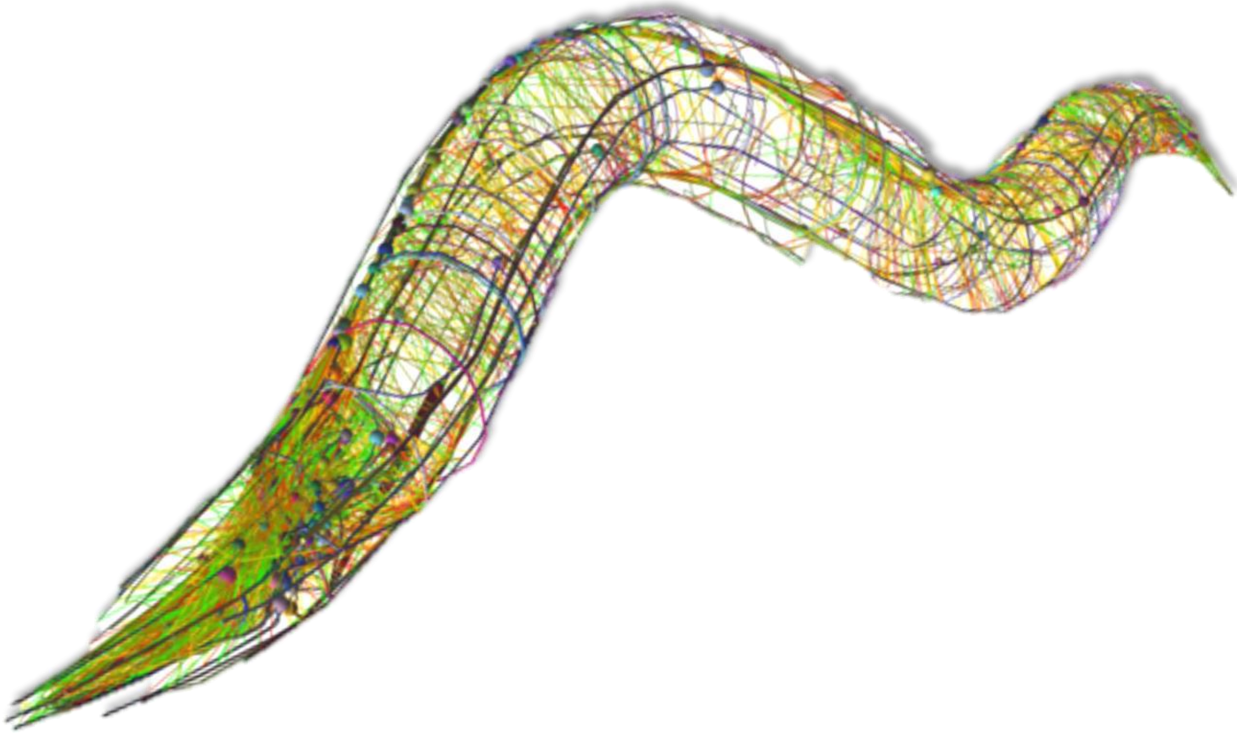


Table 1 Empirical examples of small-world networks

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

Characteristic path length L and clustering coefficient C for three real networks, compared to random graphs with the same number of vertices (n) and average number of edges per vertex (k). (Actors: $n = 225{,}226$, $k = 61$. Power grid: $n = 4{,}941$, $k = 2.67$. *C. elegans*: $n = 282$, $k = 14$.) The graphs are defined as follows. Two actors are joined by an edge if they have acted in a film together. We restrict attention to the giant connected component[16] of this graph, which includes ~90% of all actors listed in the Internet Movie Database (available at http://us.imdb.com), as of April 1997. For the power grid, vertices represent generators, transformers and substations, and edges represent high-voltage transmission lines between them. For *C. elegans*, an edge joins two neurons if they are connected by either a synapse or a gap junction. We treat all edges as undirected and unweighted, and all vertices as identical, recognizing that these are crude approximations. All three networks show the small-world phenomenon: $L \gtrsim L_{\text{random}}$ but $C \gg C_{\text{random}}$.

“Collective dynamics of small-world networks”, Watts and Strogatz,

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:

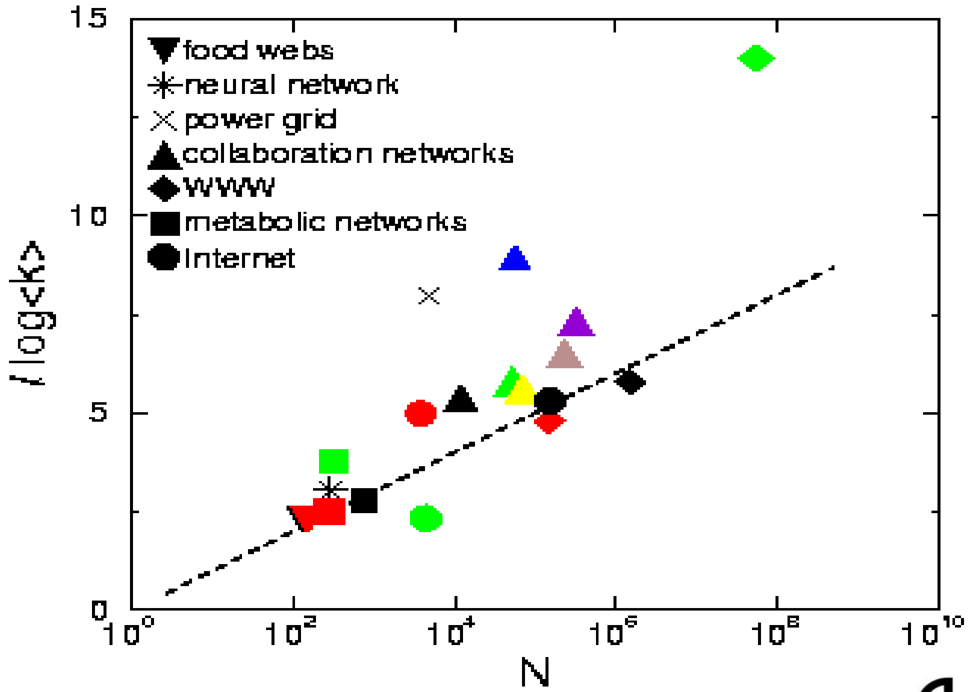
$$C_i = \frac{2\langle L_i \rangle}{k_i(k_i - 1)} = p = \frac{\langle k \rangle}{N}.$$

Degree Distribution:

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

Prediction:

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

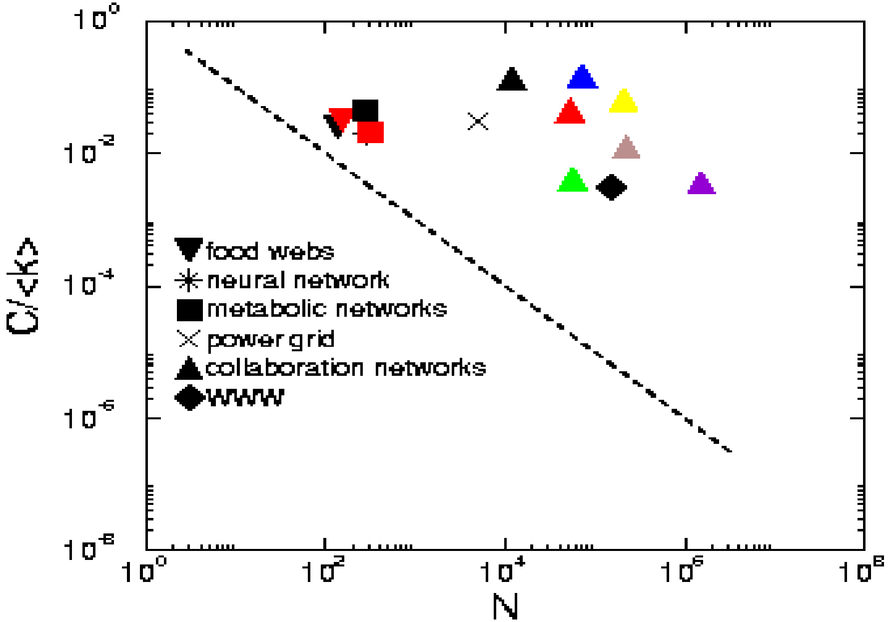


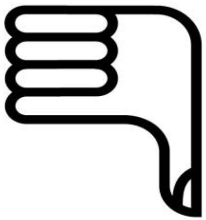
Real networks have short distances like random graphs.

CLUSTERING COEFFICIENT

Prediction:

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





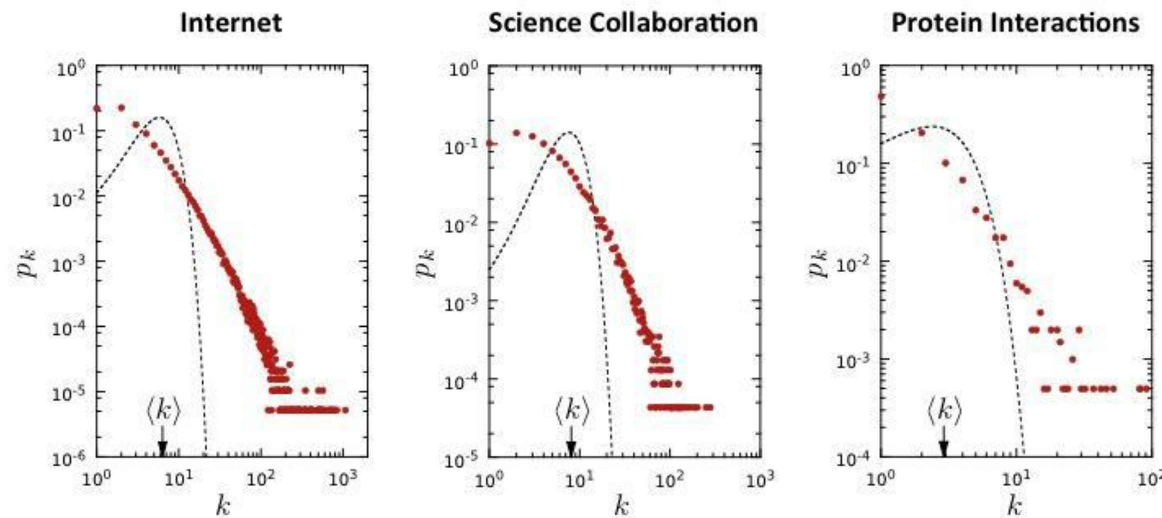
C_{rand} underestimates with orders of magnitudes the clustering coefficient of real networks.

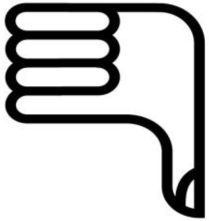
Prediction:

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

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>

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

<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>

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

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.

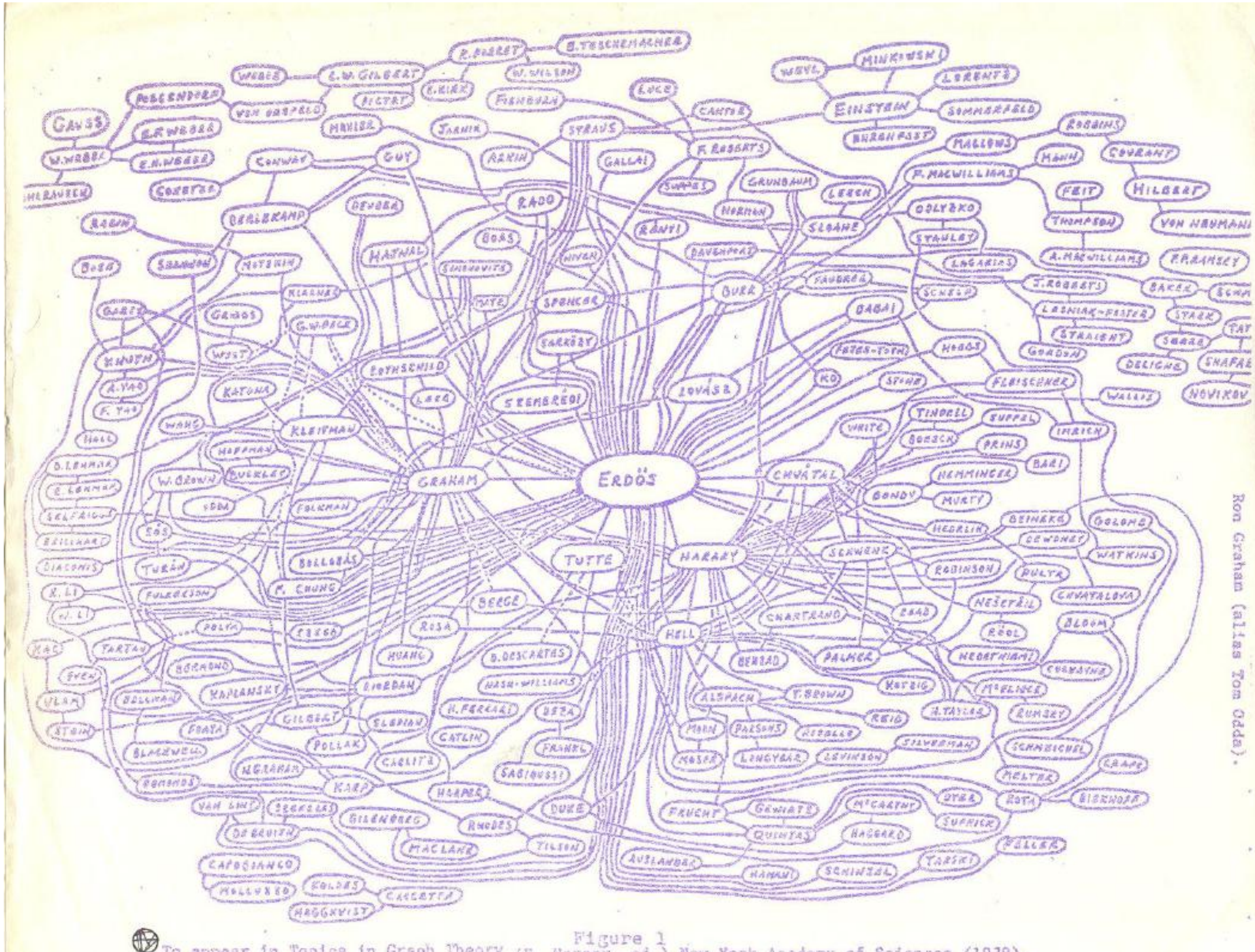
Patterns in real networks that are shared by a large number of real networks, yet which deviate from the predictions of the random network model.

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!

Summary

NETWORK DATA: SCIENCE COLLABORATION NETWORKS



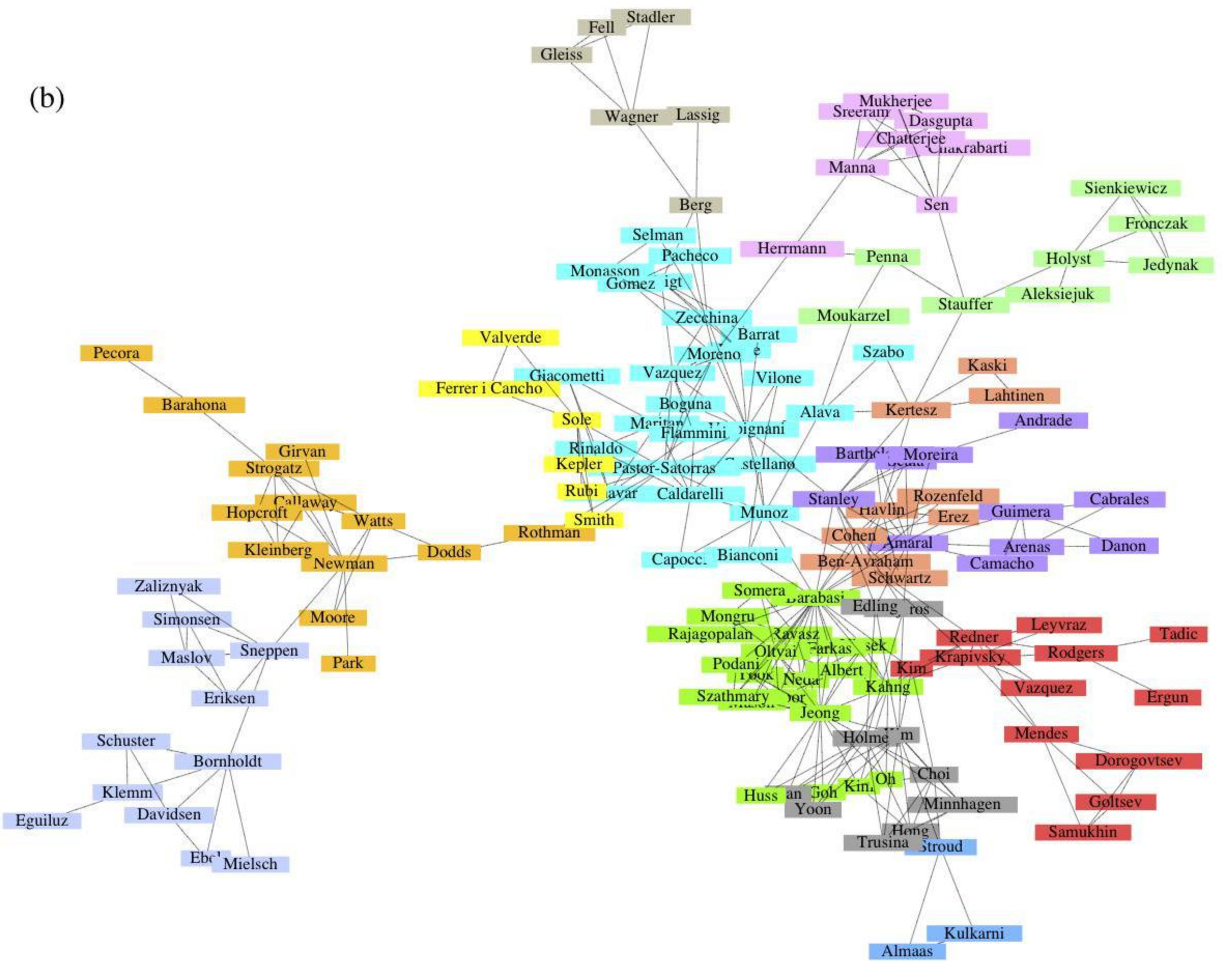
Erdos:
1,400 papers
507 coauthors

Einstein: EN=2
Paul Samuelson EN=5

….

ALB: EN: 3

NETWORK DATA: SCIENCE COLLABORATION NETWORKS

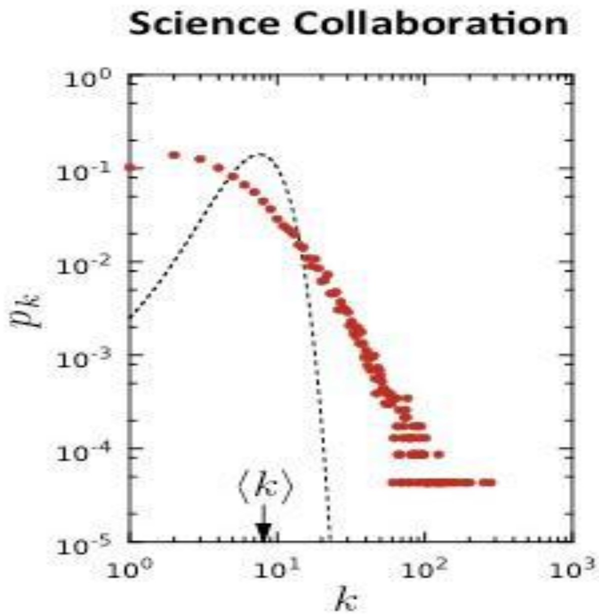


Collaboration Network:
Nodes: Scientists
Links: Joint publications

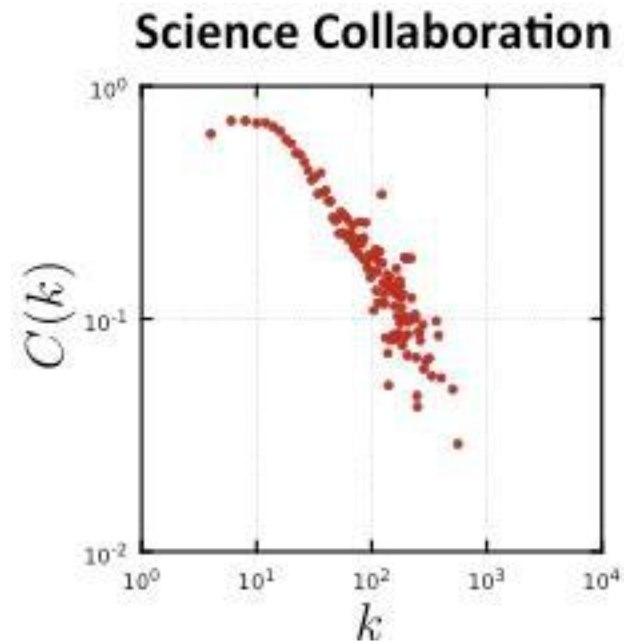
Physical Review:
1893 – 2009.

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

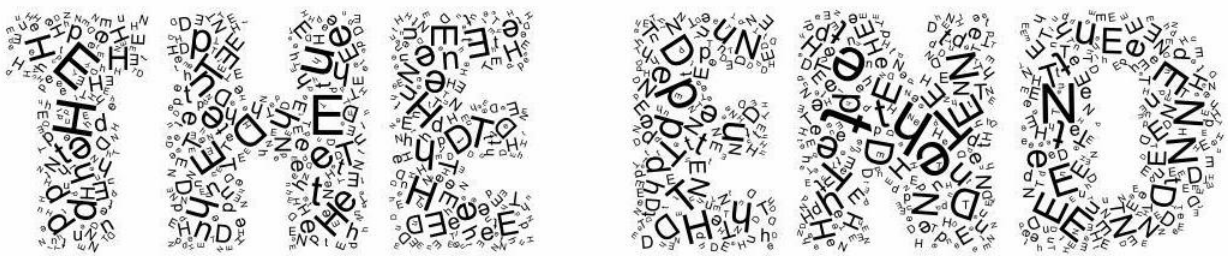
See also Stanford Large Network database
http://snap.stanford.edu/data/#canets.



Scale-free



Hierarchical


THE END