Network and Spatial Analyses

4. March 2020

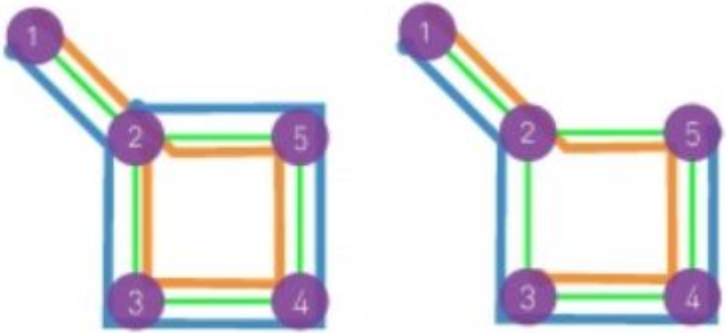
Lecture:
Introduction to Graph Theory II.

- Random Networks
- Small World Networks
- Scale-free and Barbasi Model

Seminar:
Network measures and plots in R

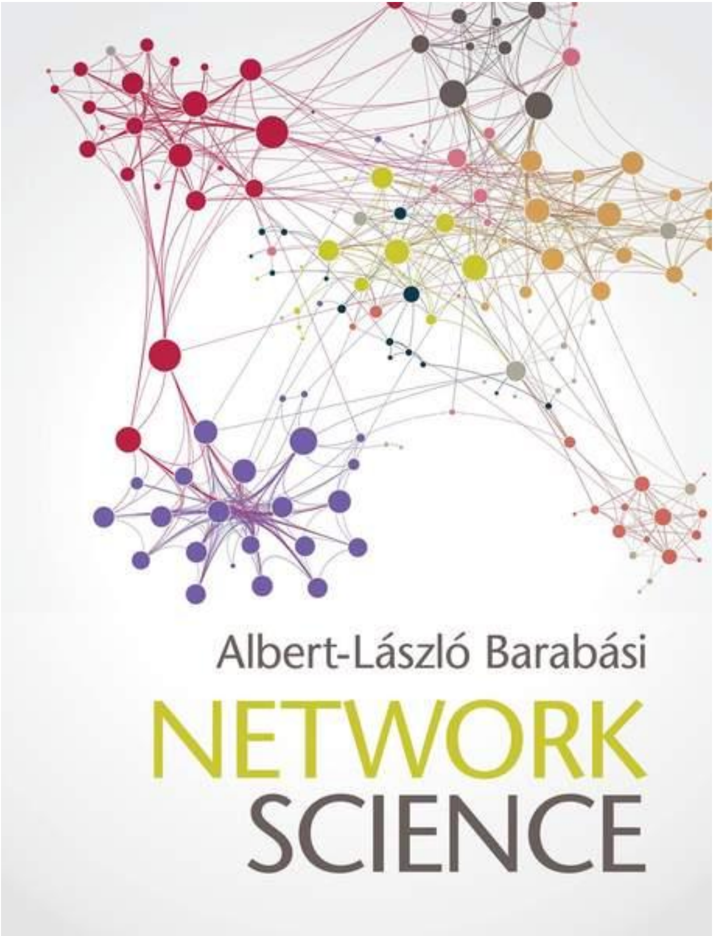
Course Github page: https://github.com/bokae/anet_course

Eulerian and Hamiltonian path



Arokhlo et al. (2011).

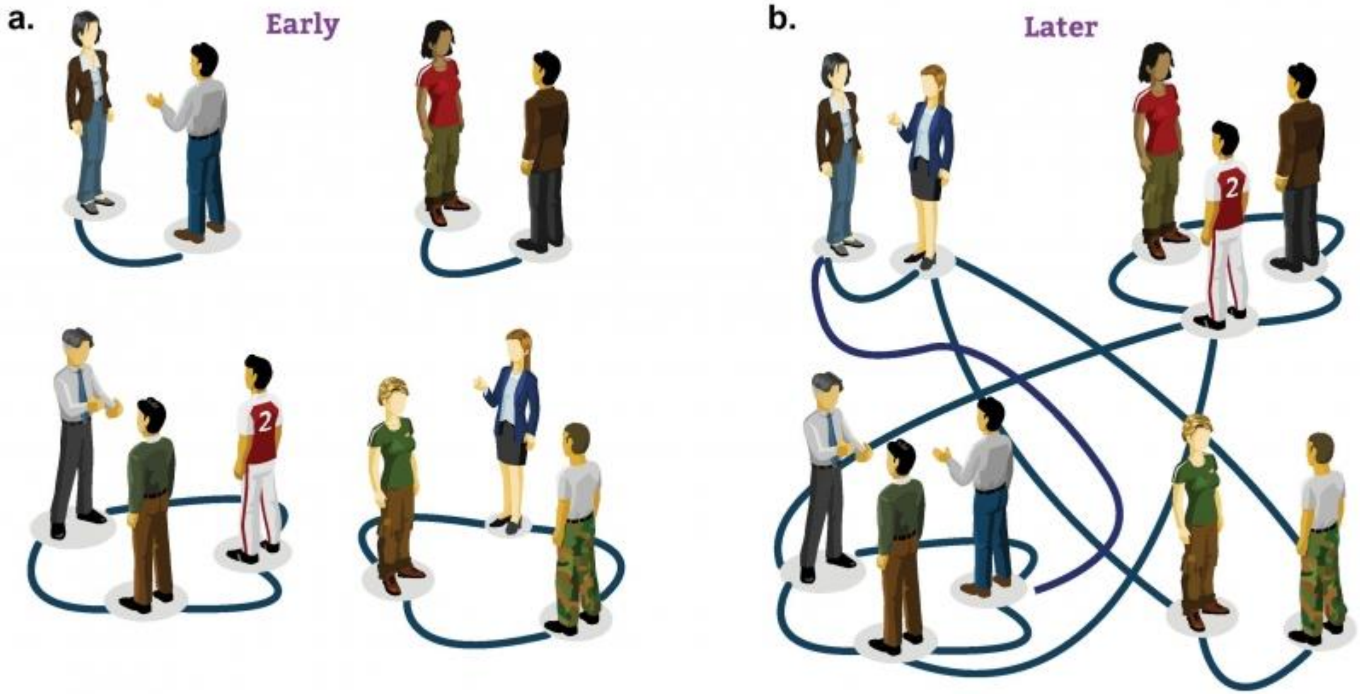
10.1007/978-3-642-23938-0_10



- Random Network Model
- Degree Distribution
- Small Worlds
- Power Laws and Scale-Free Networks
- Barabasi-Albert Model

Graphs, tables and pictures from Barabási, A. L. (2016). Network science. Cambridge university press. if we don't indicate differently!

Random Network



Random Network

G(N,L): have N number of nodes connected by L number of random connections.
Erdős-Rényi network

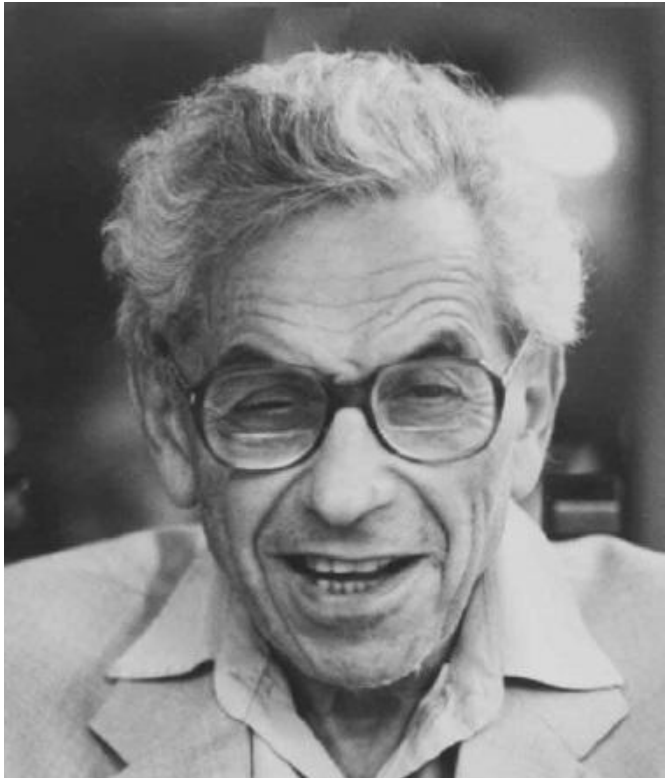
G(N,*p*): Between each pair of N_i and N_j there is a probability (*p*) of connection.
Gilbert

G(N,*p*) model fixes the probability of connections, while the G(N,L) fixes the total number of connections in the network.
In G(N,L) model the average degree easily calculable: $\langle k \rangle = \frac{2L}{N}$, but the number of connections are rarely fixed, so in other cases we use the G(N,*p*) model.

Pál Erdős

Alfréd Rényi

“A mathematician is a device for turning coffee into theorems”



Find the path between two authors:

Gergo Tóth | Paul Erdős

Gergo Tóth
co-authored 1 paper with
János Kertész
co-authored 3 papers with
András Kornai
co-authored 1 paper with
Zsolt Tuza
co-authored 7 papers with
Paul Erdős
distance = 4

https://www.csauthors.net/distance/gergo-toth/paul-erdos

How many handshakes away from Obama?

https://thepoliticalinsider.com/obama-handshake-fail/

Random Network

G(N,p): Between each pair of N_i and N_j there is a probability (p) of connection.

Each random network will be different regardless they have the same parameter N and p. Not only the visualization of the graph changes, but so does the number of connections L.

The probability of L attempts resulted in a connection is p^L.
The probability of remaining N(N-1)/2 - L attempts have not resulted in a connection:$(1-p)^{N(N-1)/2 - L}$

So, the number of different way we can replace L number of link between N(N-1)/2 number of node pairs is: $\binom{\frac{N(N-1)}{2}}{L}$

Random Network

So we can add the probability of the realization of the a random network gives L number of links by:

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

Then the expected number of links in a random graph is:

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

The binomial distribution has a form of:

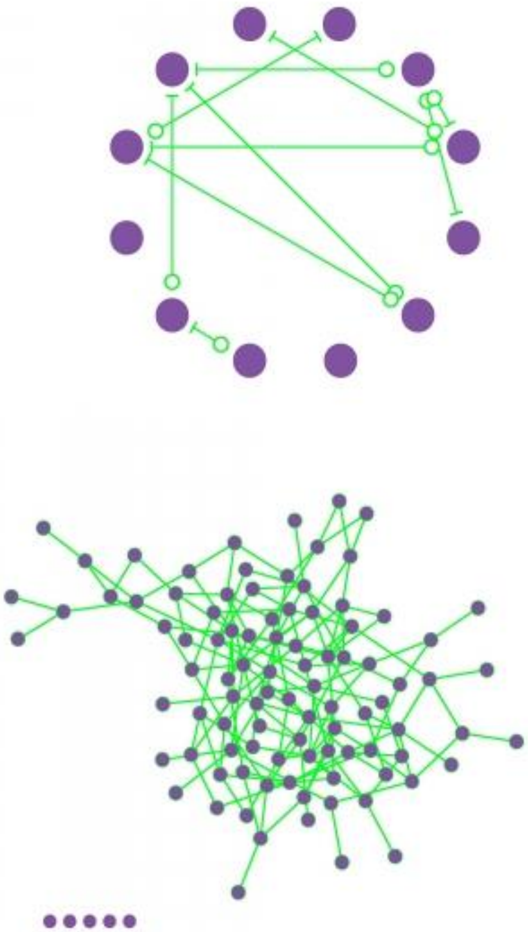
$$p_x = \binom{N}{x} p^x (1-p)^{N-x}$$

The first moment of the distribution:

$$\langle x \rangle = \sum_{x=0}^{N} x p_x = Np$$

Hence, the expected number of links is a produce of the probability and number of pairs we try to connect, which is L_{max}= N(N-1)/2.

Similarly, to get the average degree: $\langle k \rangle = \frac{2\langle L \rangle}{N} = p(N-1)$



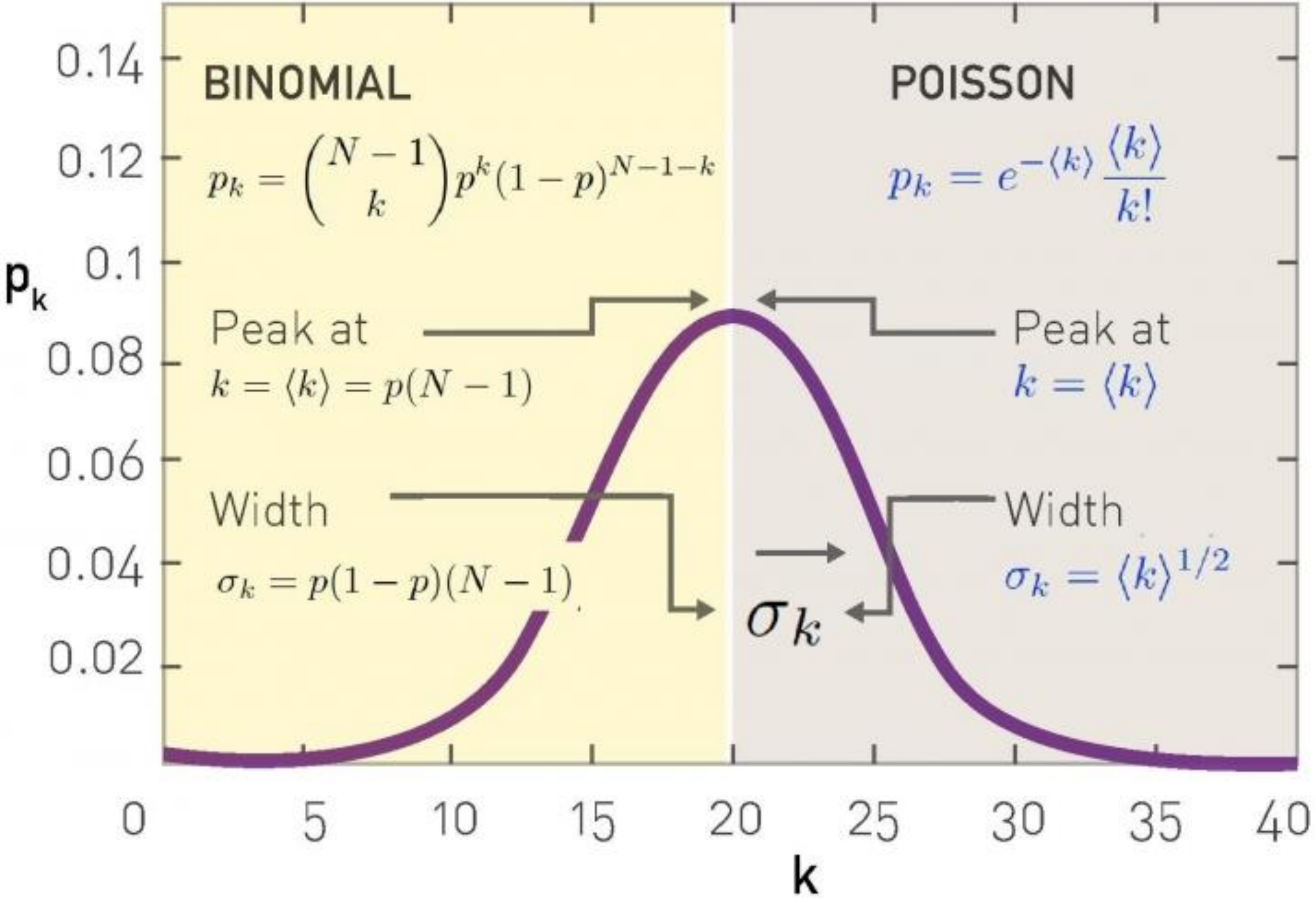
Random Network

How many different random graph can we describe with the same number of links?
Turn into a combinaturical question: how many ways can we put L number of connection on L_{max} places?

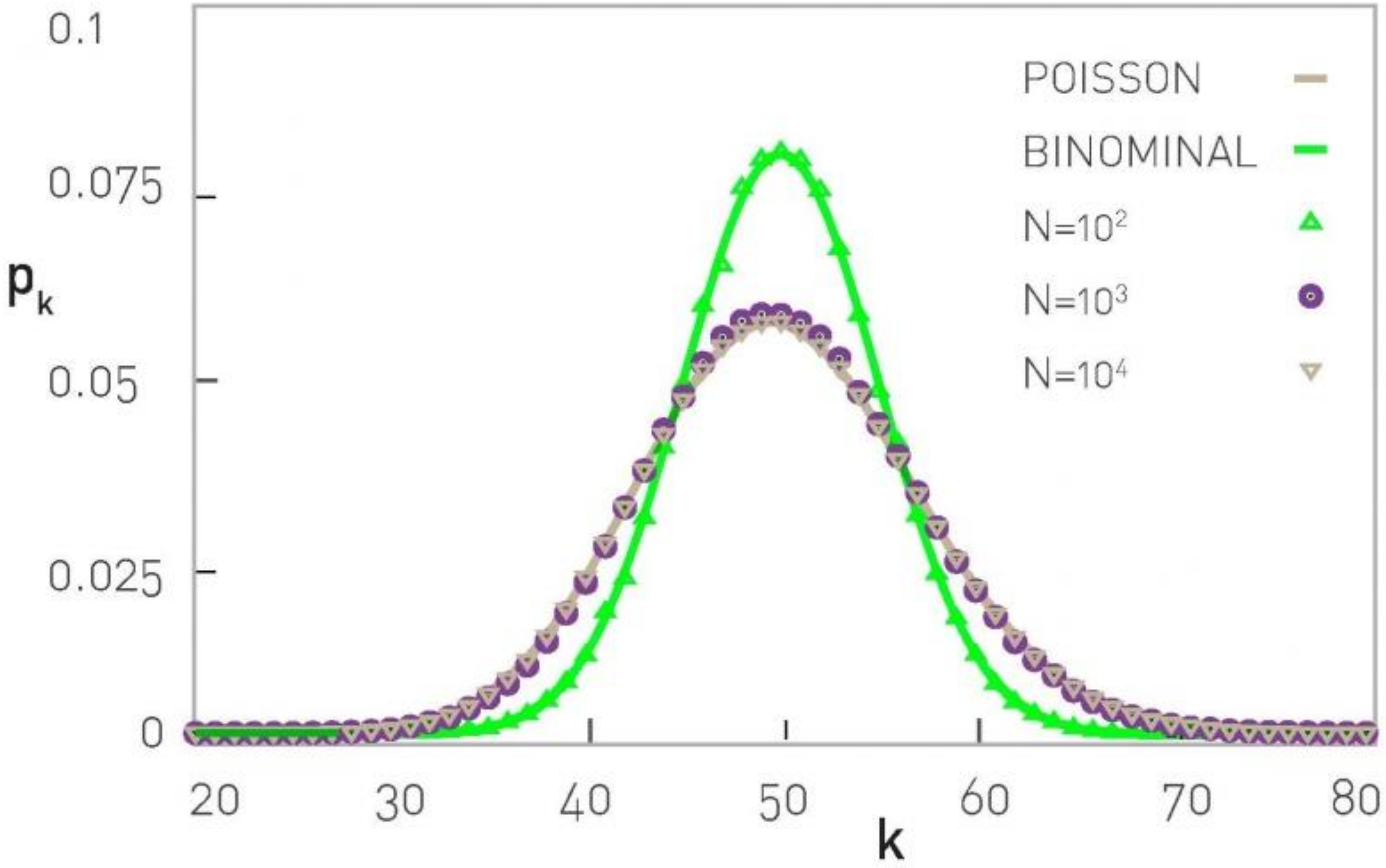
The first link of the sequence can be placed on L_{max} places, the second one L_{max} -1, the third one to L_{max}-2 places... $L_{max<}$ $(L_{max}-1)(L_{max}-2)...(L_{max}-L+1)$.

We have to divide the result by the number of different sequences, which is L!

$$\frac{L_{max}(L_{max}-1)(L_{max}-2)...(L_{max}-L+1)}{L!} = \frac{L_{max}!}{L!(L_{max}-L)!} = \binom{L_{max}}{L}$$



BINOMIAL
$p_k = \binom{N-1}{k} p^k (1-p)^{N-1-k}$
POISSON
$p_k = e^{-\langle k \rangle} \frac{\langle k \rangle}{k!}$
p_k
Peak at
$k = \langle k \rangle = p(N-1)$
Peak at
$k = \langle k \rangle$
Width
$\sigma_k = p(1-p)(N-1)$
σ_k
Width
$\sigma_k = \langle k \rangle^{1/2}$
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
5
10
15
20
25
30
35
40
k



0.1
0.075
0.05
0.025
0
p_k
POISSON
BINOMINAL
N=10^2
N=10^3
N=10^4
20
30
40
50
60
70
80
k

Real Networks are not Poisson

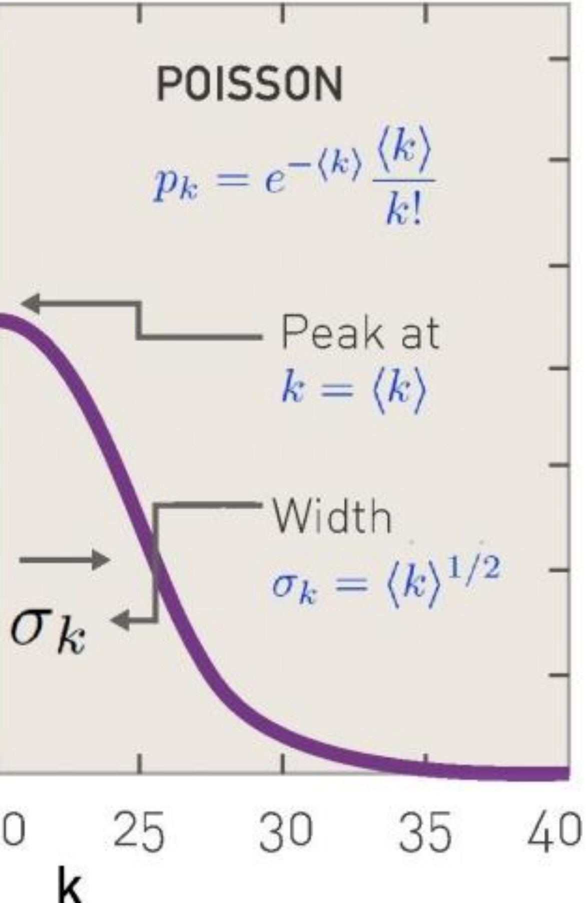
According to sociological estimations an average person knows ~1000 individuals on first name bases, so <k> = 1000.

By using the random network properties we can estimate that, in this random society the estimated k_{max} = 1185, while the k_{min} =816.

The dispersion of this random network is $\sigma_k = \langle k \rangle^{1/2}$, which for <k>=1000 is $\sigma_k = 31.62$.

So friendships should ranging from 968 to 1032. $\langle k \rangle \pm \sigma_k$

Real Networks are not Poisson

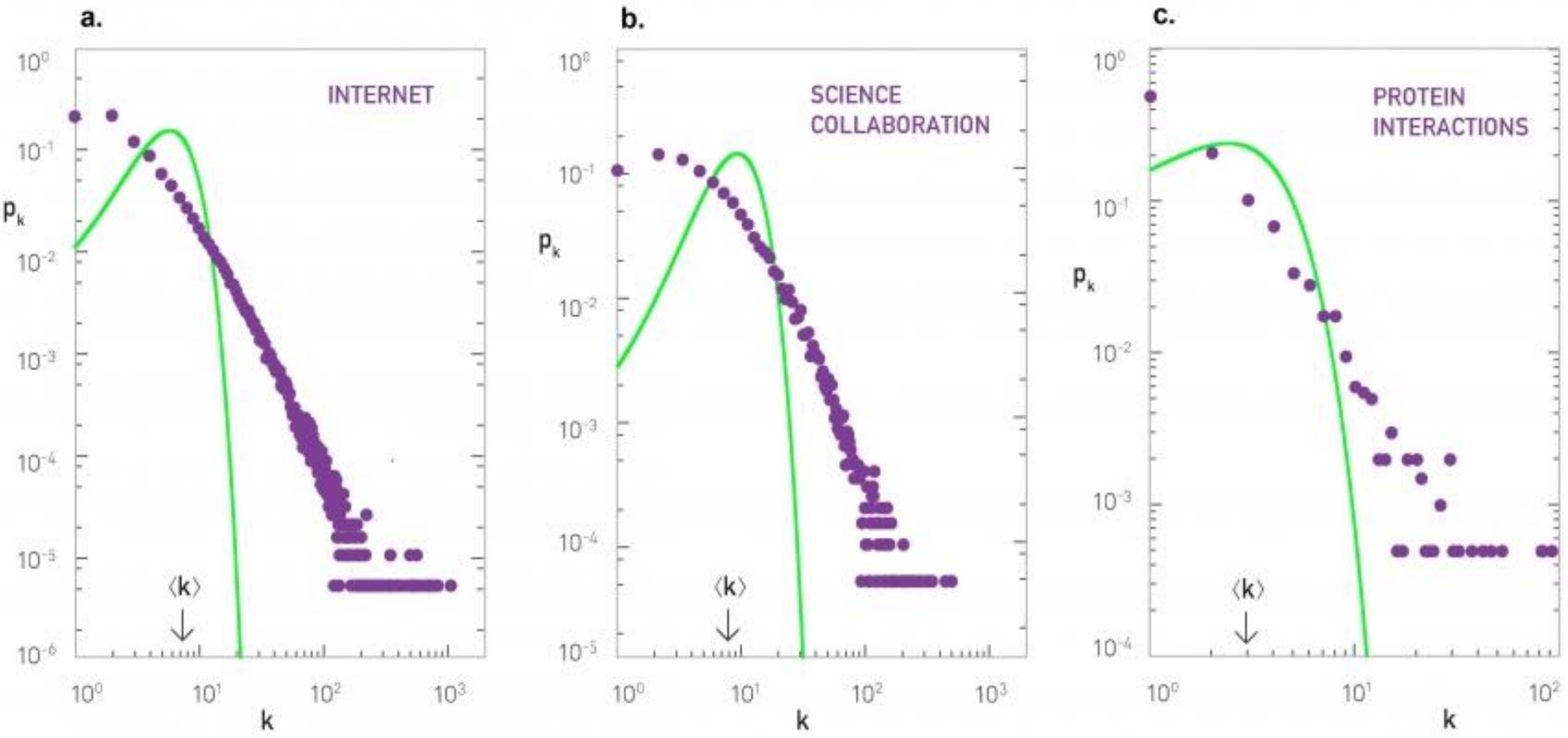


Why are Hubs missing?

$k! \approx \left[\sqrt{2\pi k}\right] \left(\frac{k}{e}\right)^k$,which allows us to rewrite the original approximation to:

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

, which predicts the chances of observing a node with high number connections decreases faster than exponentially!



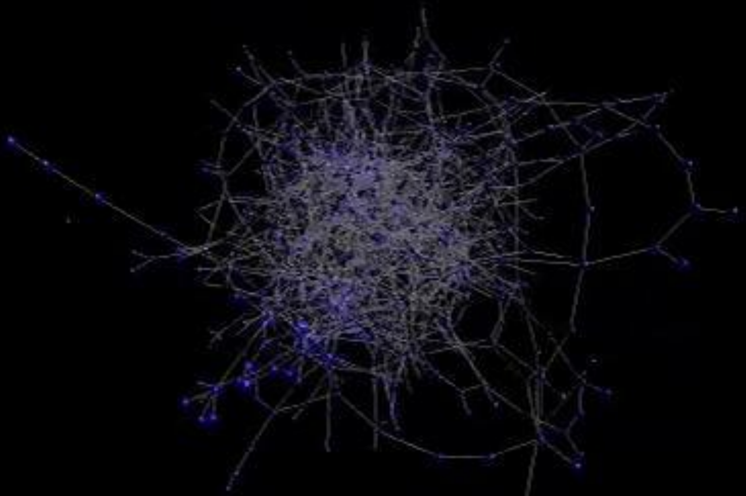
a.
INTERNET
b.
SCIENCE
COLLABORATION
c.
PROTEIN
INTERACTIONS
p_k
k
$\langle k \rangle$

Evolution of a Random Network

We have two extreme cases:

- $p = 0$, all the nodes are isolated, then the largest component $N_G = 1$ and N_G/N tends to 0.

- $p = 1$, therefore $\langle k \rangle = N-1$, then the network is a complete graph and all nodes belong to the giant component $N_G = N$, thus $N_G/N = 1$.



$p > n^{-1}$

The graph becomes a connected Giant Component

Evolution of a Random Network

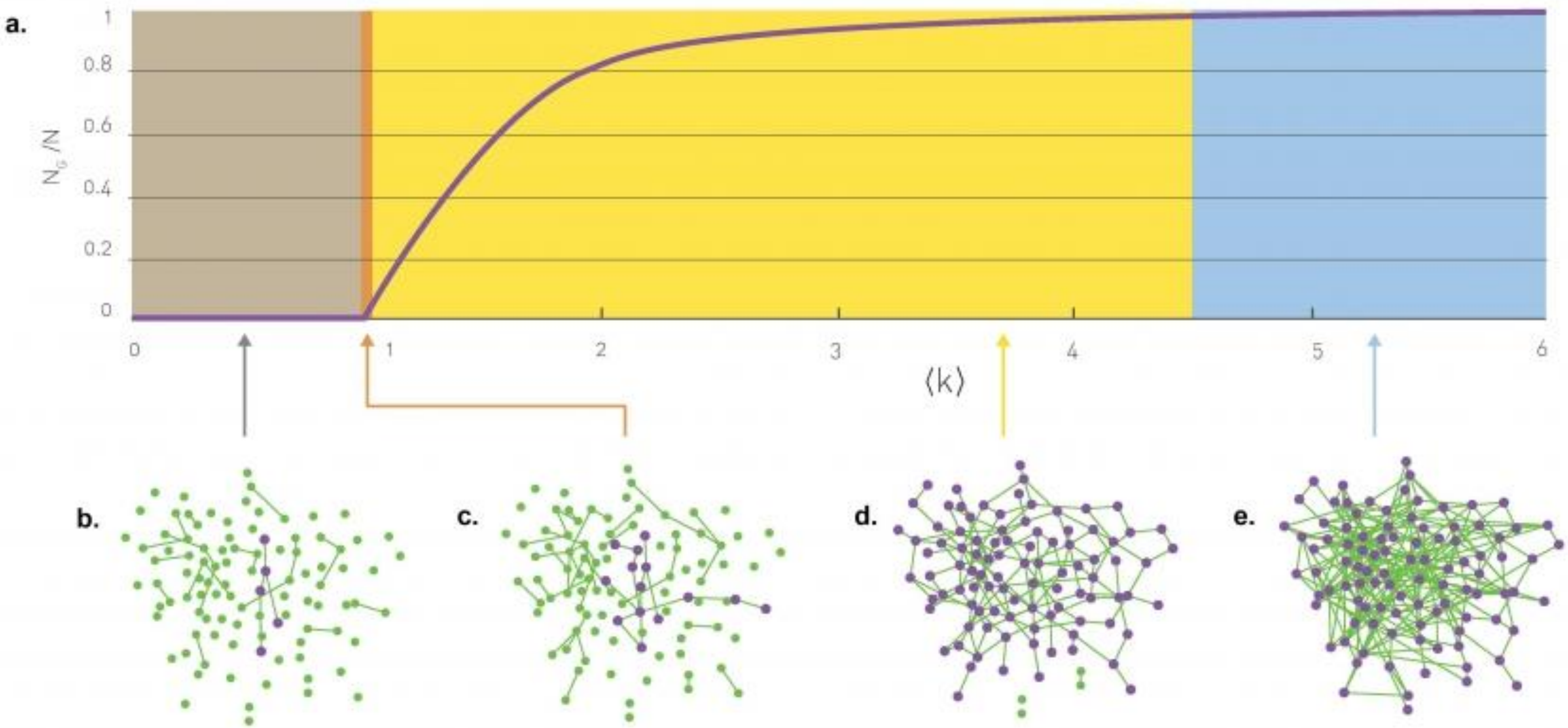
We have two extreme cases:

- $p = 0$, all the nodes are isolated, then the largest component $N_G = 1$ and N_G/N tends to 0.

- $p = 1$, therefore $\langle k \rangle = N-1$, then the network is a complete graph and all nodes belong to the giant component $N_G = N$, thus $N_G/N = 1$.

Once $\langle k \rangle$ exceeds a critical value N_G/N increases, which means that a large cluster - what we call as a Giant Component - emerges.

Critical point: $\langle k \rangle = 1$
Subcritical regime : $0 < \langle k \rangle < 1$
Supercritical: $\langle k \rangle > 1$
Connected Regime: $\langle k \rangle > \ln N$

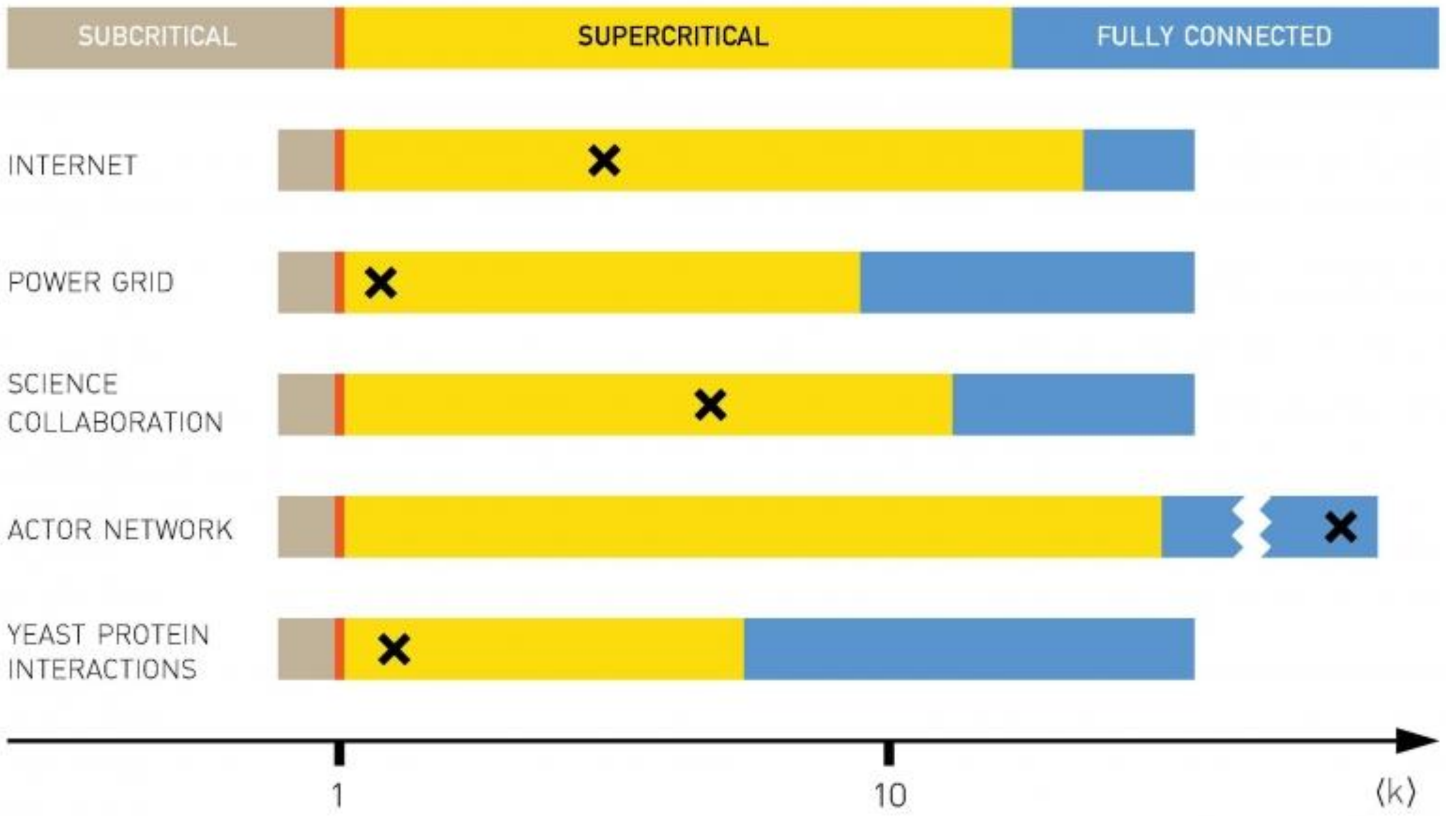

a.
1
0.8
0.6
0.4
0.2
0
N_G/N
0
1
2
3
4
5
6
$\langle k \rangle$
b.
c.
d.
e.

Real Networks are Supercritical

Critical point: $<k> = 1$
Subcritical regime : $0 < \; <k> \; < 1$
Supercritical: $<k> > 1$
Connected Regime: $<k> > \ln N$

Network	N	L	$\langle K \rangle$	lnN
Internet	192,244	609,066	6.34	12.17
Power Grid	4,941	6,594	2.67	8.51
Science Collaboration	23,133	94,437	8.08	10.05
Actor Network	702,388	29,397,908	83.71	13.46
Protein Interactions	2,018	2,930	2.90	7.61

Real Networks are Supercritical



Small World Networks

Six degrees of separation.

The distance between two randomly chosen node is short.

- What does short mean?
- Short compare to what?

Consider a random network with average degree of $<k>$, then a node has on average:

- $<k>$ nodes at distance $d=1$,
- $<k>^2$ nodes at distance $d=2$

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

So if we assume that a social network has a $<k> \sim 1000$, then we know 10^6 people at $d=2$, and the whole earth population is at $d=3$.

Small World Networks

The expected number of nodes in distance *d* from a randomly chosen node is:

$$N(d) = 1 + \langle k \rangle + \langle k^2 \rangle + \dots + \langle k^d \rangle = \frac{\langle k \rangle^{d+1}}{\langle k \rangle - 1}$$

N(*d*) must be within the range of total number of nodes. Therefore we can identity the maximum distance in the network, which is the network's diameter:

$$N(d_{max}) \approx N$$

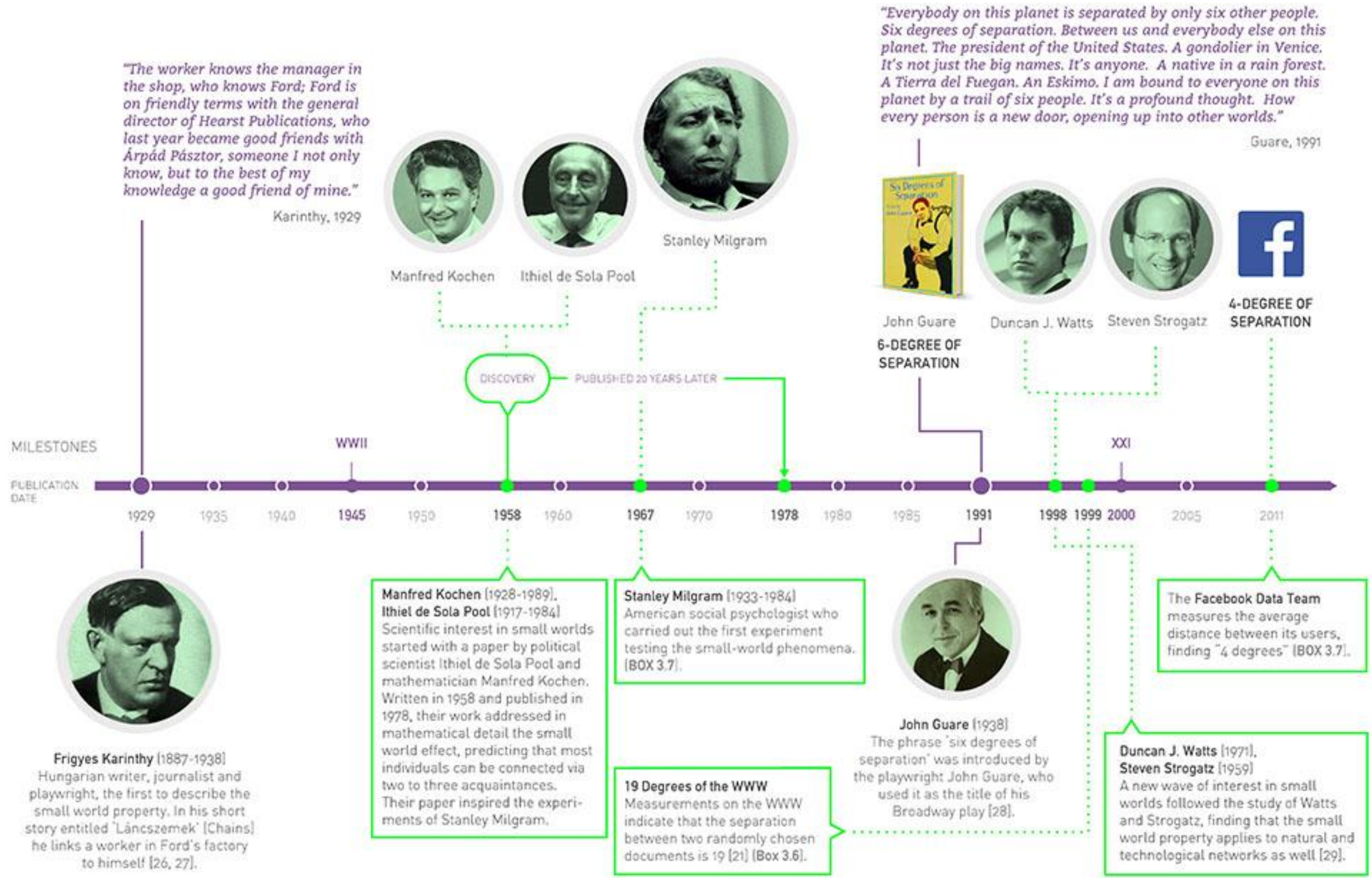
If we assume $\langle k \rangle >> 1$ then we can neglect the +/- 1 in the fraction, obtaining:

$$\langle k \rangle^{d_{max}} \approx N$$

Therefore, the diameter of a random network can be approximated by:

$$d_{max} \approx \frac{lnN}{ln\langle k \rangle} \longmapsto \langle d \rangle \approx \frac{lnN}{ln\langle k \rangle}$$

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



"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
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
1999
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 [26, 27].
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 3.7).
19 Degrees of the WWW
Measurements on the WWW indicate that the separation between two randomly chosen documents is 19 [21] (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 [28].
The Facebook Data Team measures the average distance between its users, finding "4 degrees" (BOX 3.7).
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 [29].

Small World Networks

Clustering Coefficient

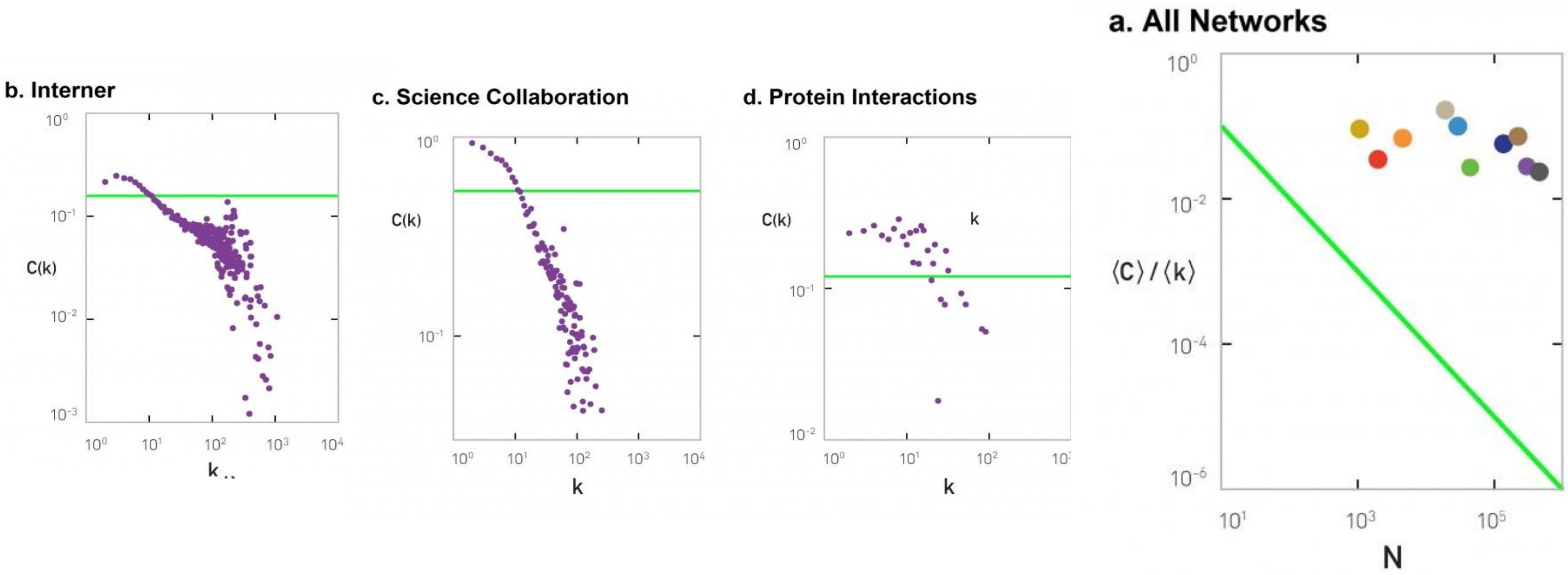
Degree of a node does not include information about the neighbors of the node. Local clustering measures the density of connections between the immediate neighbors of the ego. In a random network to calculate C_i first we need to know the expected number of a links of the ego network of *i*:

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

Thus, the local clustering coefficient in a random random network can be expressed by:

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

- If we fix <k>, the larger the network, the smaller is the node's local clustering.
- Local clustering coefficient of a node is independent from the number of connections of the node.



-Looking on real life networks, <C>/<k> does not decrease as N^{-1}

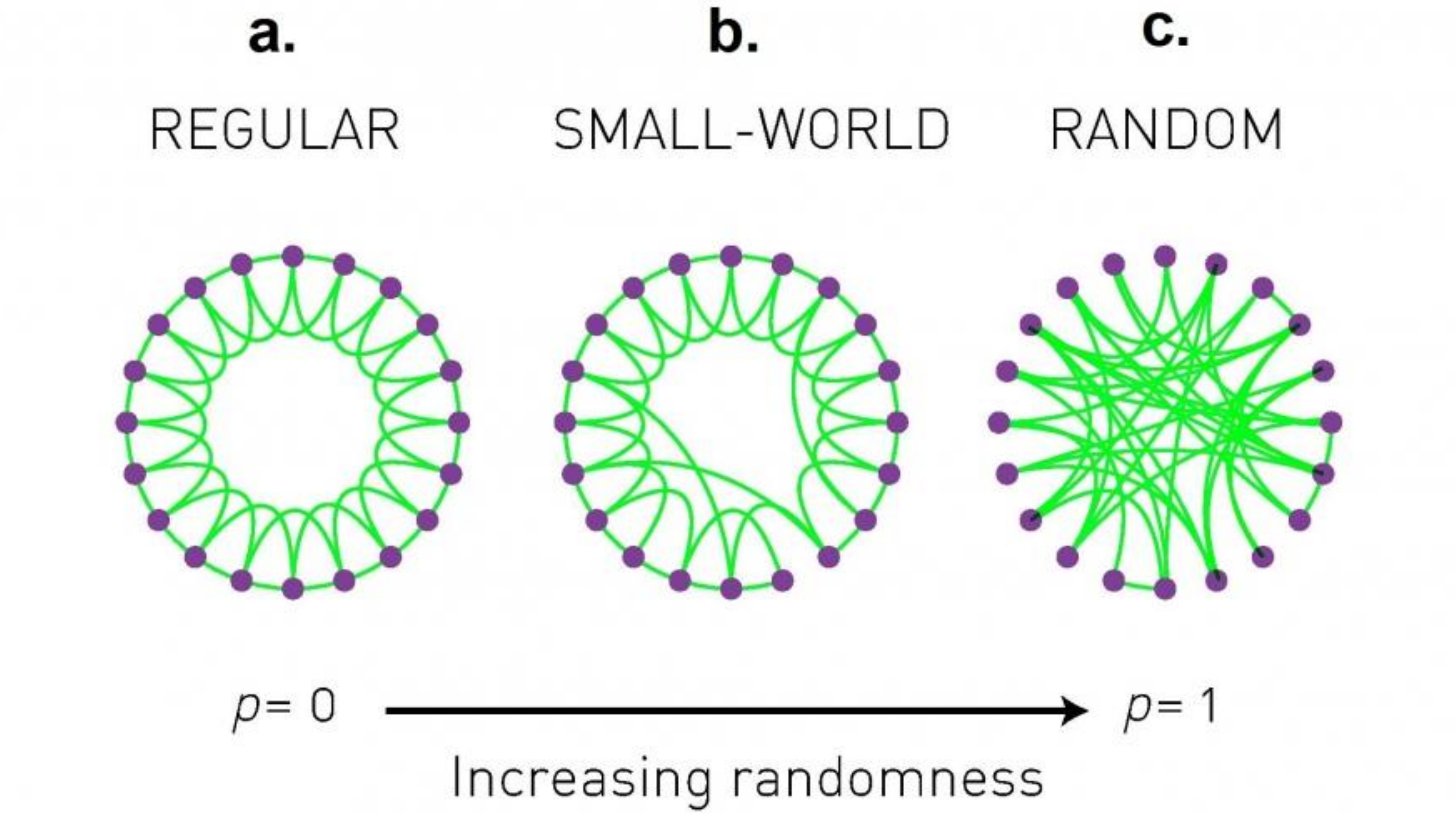
- Real network has a much higher clustering coefficient than corresponding random G(N,L)

Small World Networks

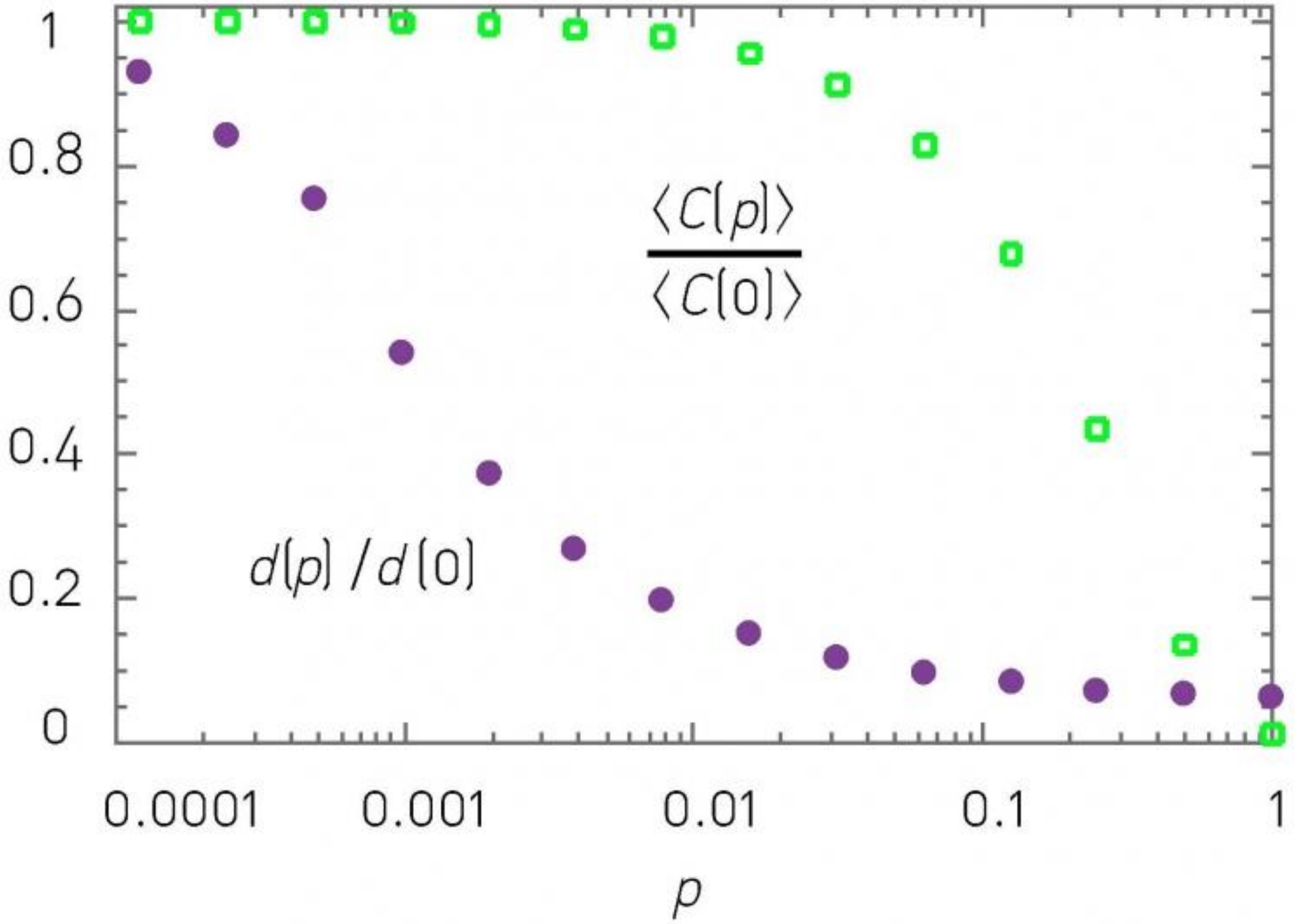
Watts-Strogatz Model

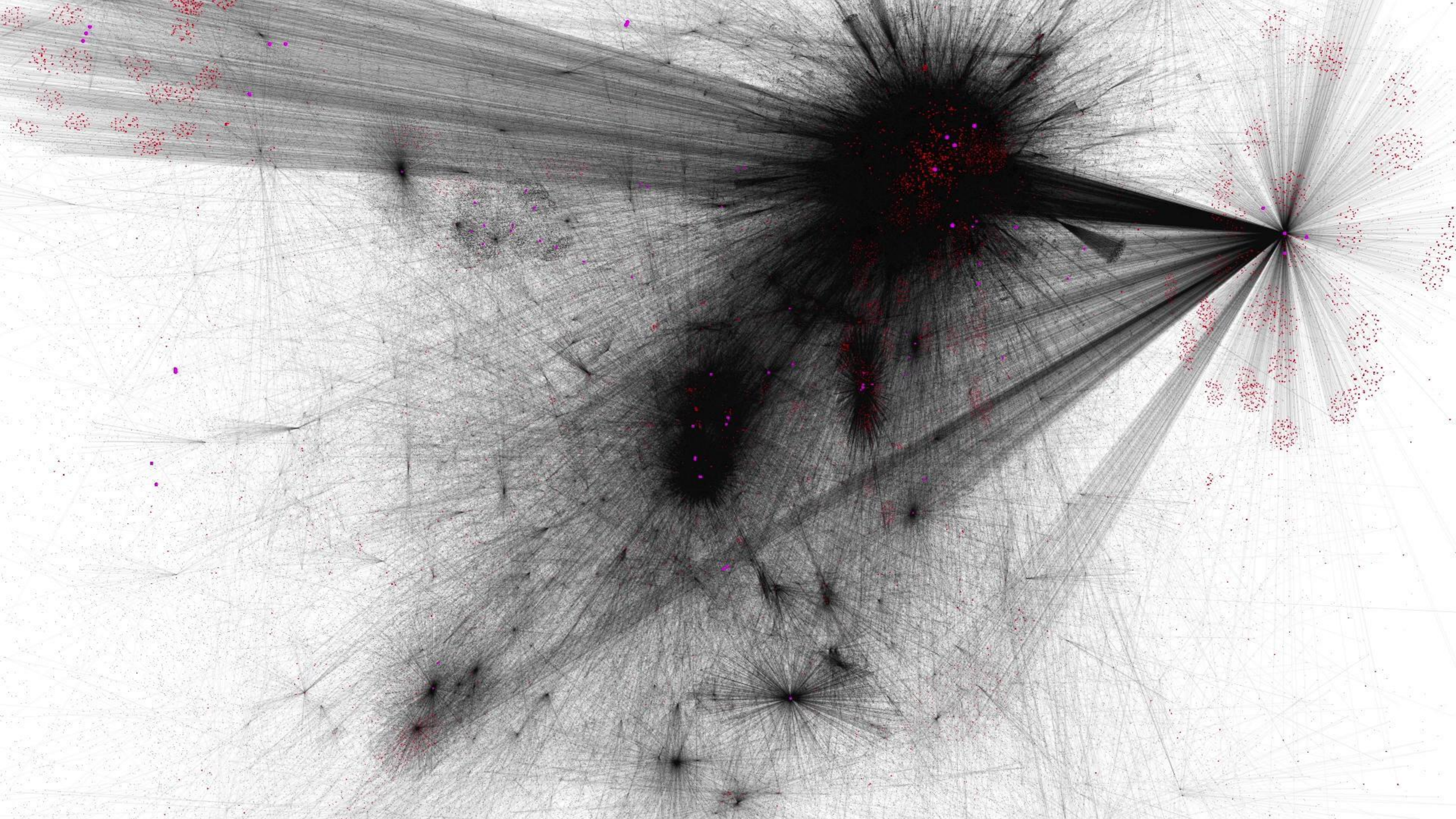
What gives the word *small* in Small World Networks?
In real networks the average distance between two points is a logarithmic function of N.

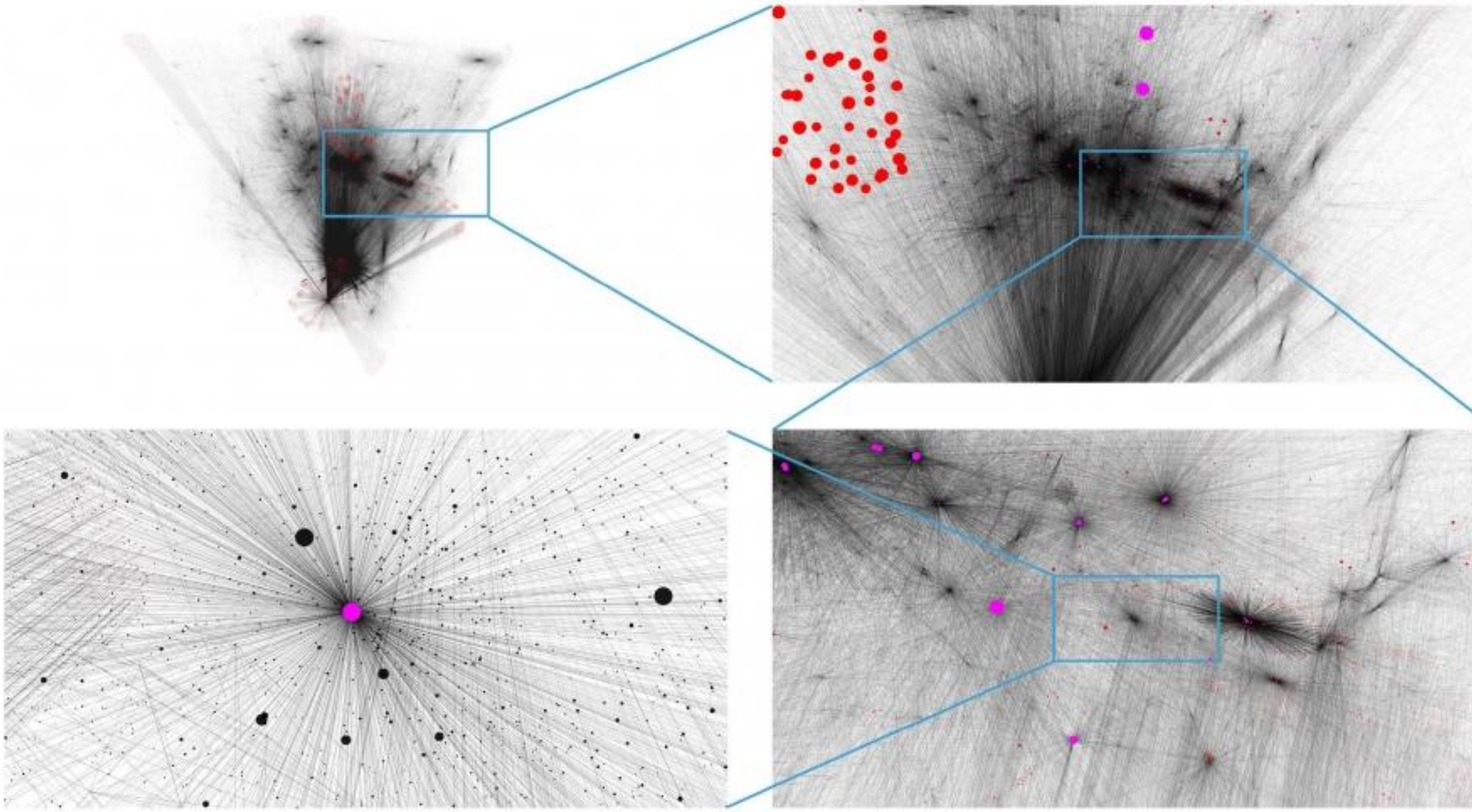
What makes it a real life network?
The average clustering coefficient of a real network is much higher than we would expect from a random network with the same properties of N and L.


a.
REGULAR
b.
SMALL-WORLD
c.
RANDOM
p= 0
p= 1
Increasing randomness

d.



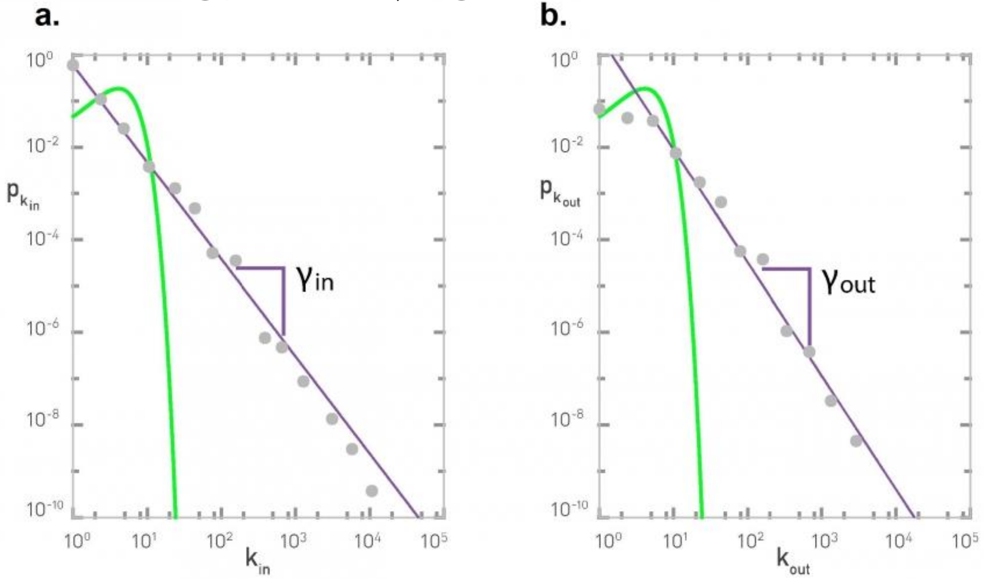




Scale-Free Networks

If the network of WWW was to be a random network than the degree distribution should follow Poisson distribution.

But it follows a power law distribution, with the exponent of gamma: $p_k = k^{-\gamma}$
Take the logarithm, we obtain: $\log p_k = -\gamma \log k$



Scale-Free Networks

In discrete formalization of the degree distribution we provide a probability the node has exactly *k* links: $p_k = Ck^{-\gamma}$

,where *C* is a normalizing condition. $\sum_{k=1}^{\infty} = 1 = C\sum_{k=1}^{\infty} k^{-\gamma}$

$$C = \frac{1}{\sum_{k=1}^{\infty} k^{-\gamma}} = \frac{1}{\zeta(\gamma)}$$

Thus, with $k > 0$ case discrete power-law distribution can be expressed by: $p_k = \frac{k^{-\gamma}}{\zeta(\gamma)}$

Scale-Free Networks

In discrete formalization of the degree distribution we provide a probability the node has exactly *k* links: $p_k = Ck^{-\gamma}$

,where *C* is a normalizing condition. $\sum_{k=1}^{\infty} = 1 = C\sum_{k=1}^{\infty} k^{-\gamma}$

$$C = \frac{1}{\sum_{k=1}^{\infty} k^{-\gamma}} = \frac{1}{\zeta(\gamma)}$$

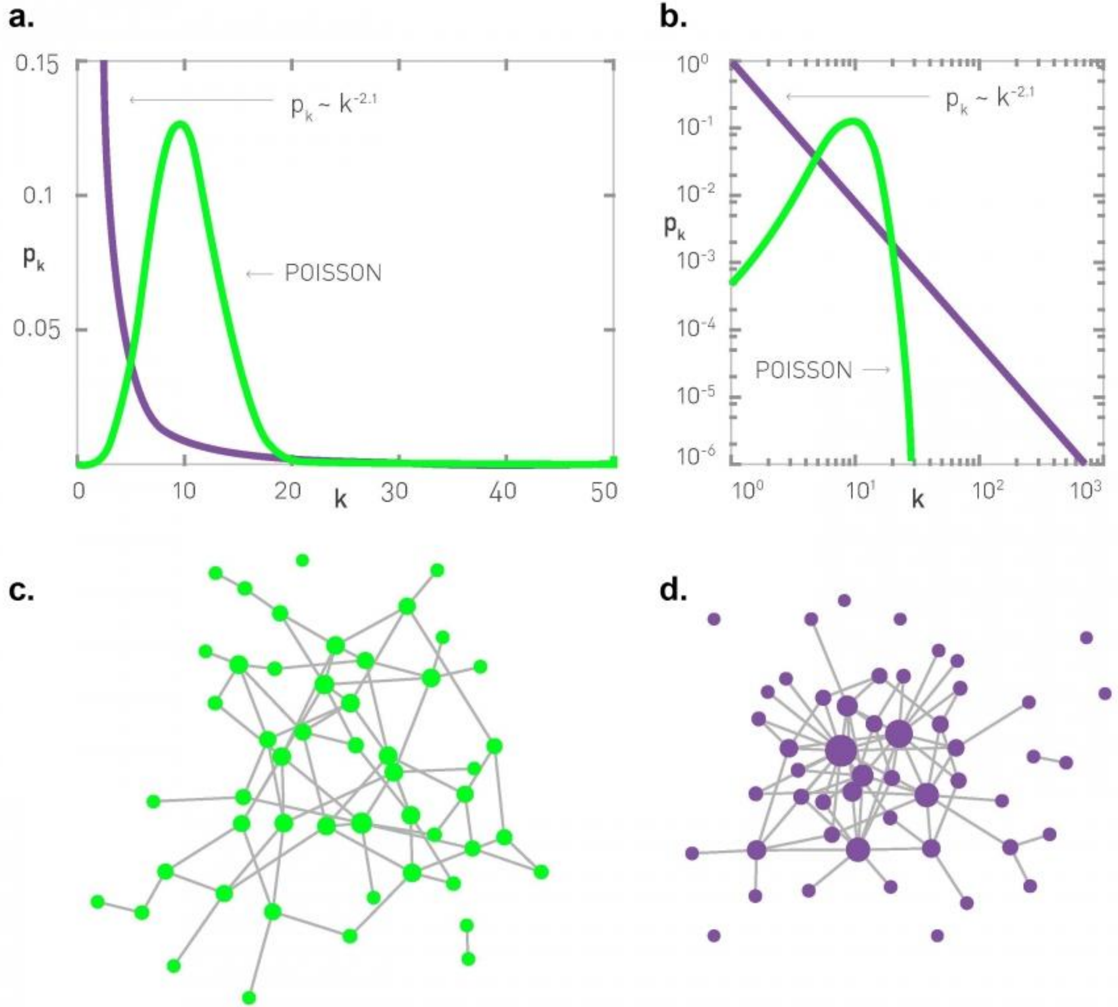
Thus, with $k > 0$ case discrete power-law distribution can be expressed by: $p_k = \frac{k^{-\gamma}}{\zeta(\gamma)}$

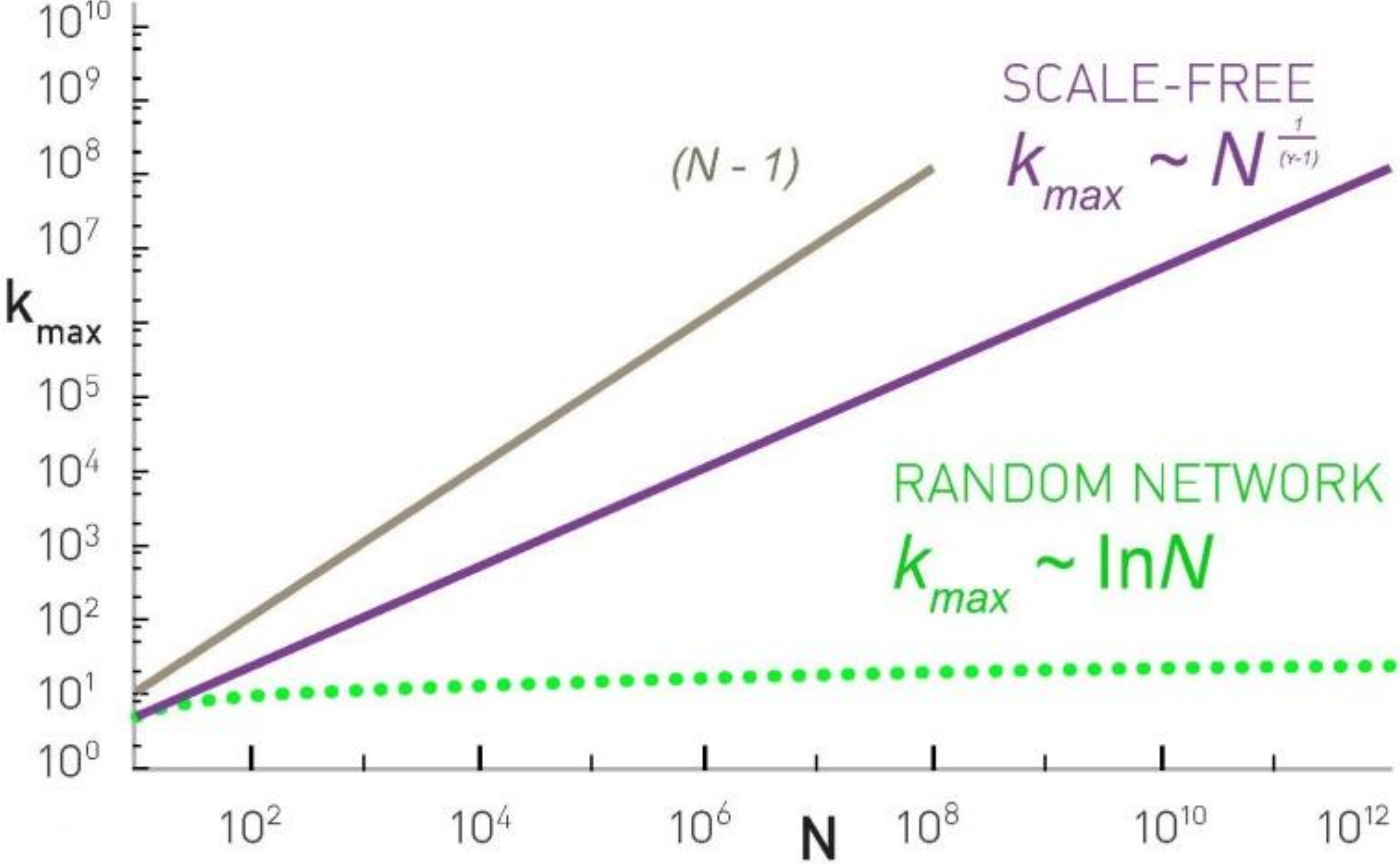
So let's assume that the degrees can form any positive real number, so we rewrite the power-law distribution as an integral: $p(k) = Ck^{-\gamma}$

by normalizing we obtain the integrate of a continuum: $\int_{k_{min}}^{\infty} p(k)dk = 1$

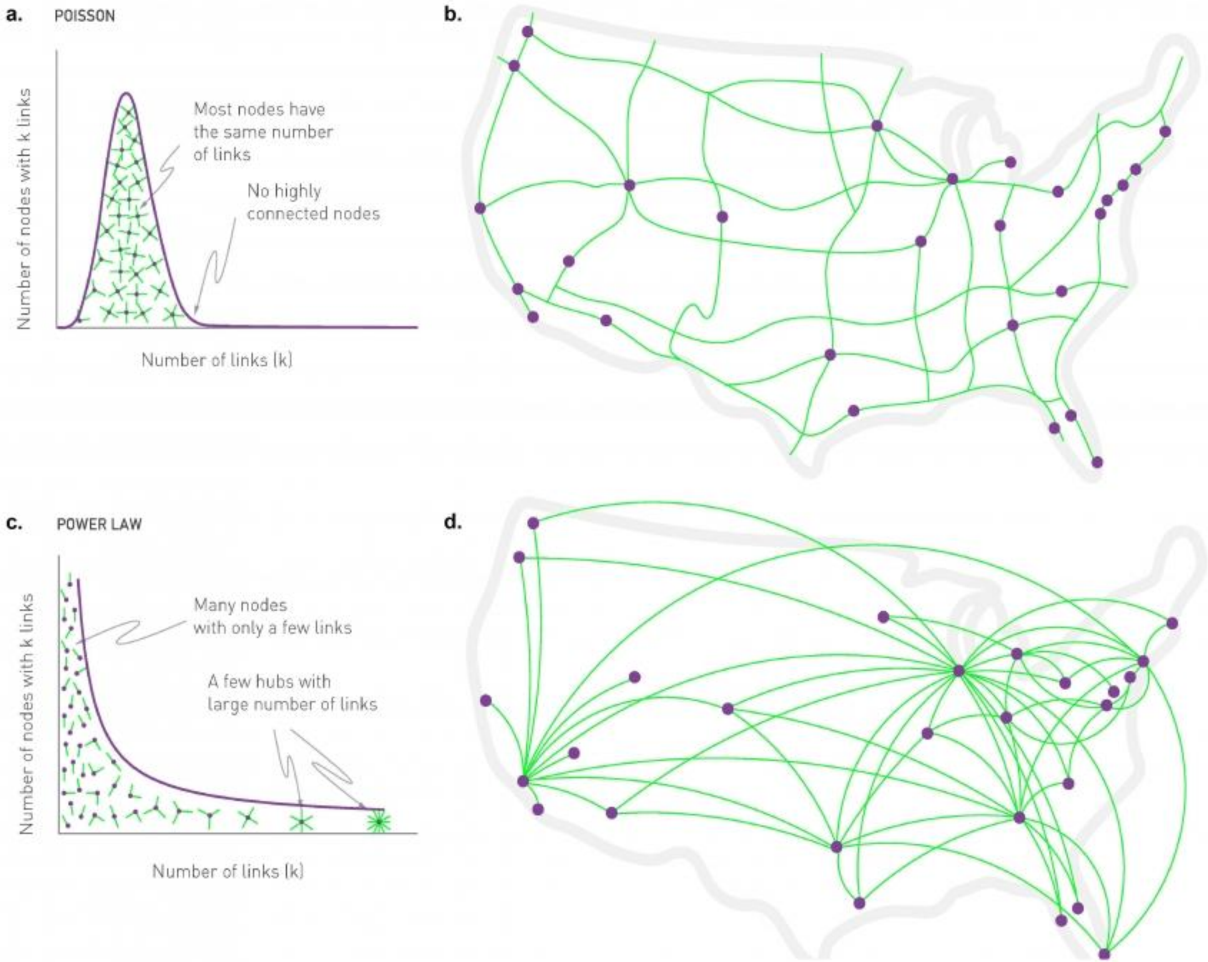
,where $C = \frac{1}{\int_{k_{min}}^{\infty} k^{-\gamma} dk} = (\gamma - 1)k_{min}^{\gamma-1}$

Thus, in continuous case the degree distribution can be expressed as: $\int_{k_1}^{k_2} p(k)dk$


a.
0.15
0.1
0.05
p_k
$p_k \sim k^{-2.1}$
POISSON
0
10
20
k
30
40
50
b.
10^0
10^{-1}
10^{-2}
p_k
10^{-3}
10^{-4}
10^{-5}
10^{-6}
$p_k \sim k^{-2.1}$
POISSON
10^0
10^1
k
10^2
10^3
c.
d.



SCALE-FREE
$k_{max} \sim N^{\frac{1}{(\gamma-1)}}$
(N - 1)
RANDOM NETWORK
$k_{max} \sim \ln N$
k_{max}
N
10^0
10^1
10^2
10^3
10^4
10^5
10^7
10^8
10^9
10^{10}
10^2
10^4
10^6
10^8
10^{10}
10^{12}


a.
POISSON
Number of nodes with k links
Most nodes have the same number of links
No highly connected nodes
Number of links (k)
b.
c.
POWER LAW
Number of nodes with k links
Many nodes with only a few links
A few hubs with large number of links
Number of links (k)
d.

Scale-Free Networks - Why do we have Hubs?

The Barabasi-Albert Model

For to every one who has will more be given, and
he will have abundance; but from him who has not,
even what he has will be taken away.
— Matthew 25:29, RSV.

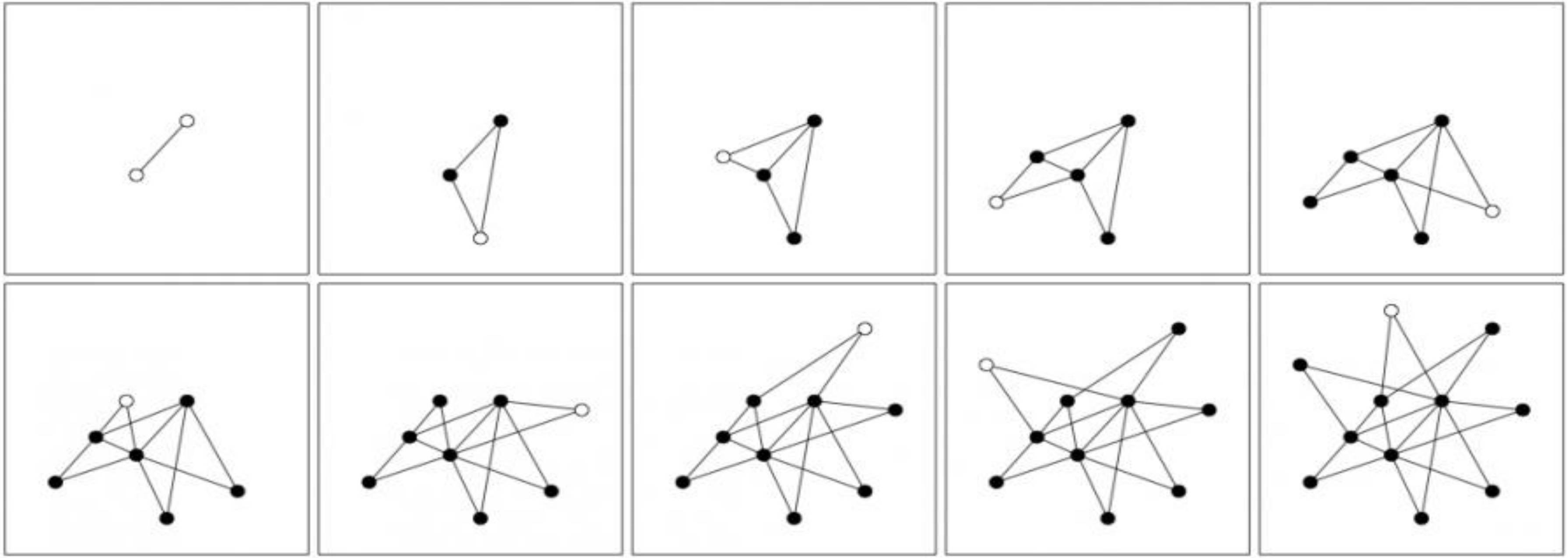
Growth
Real networks are getting larger and larger of a result of a growth process. While in random networks we usually have fix N.

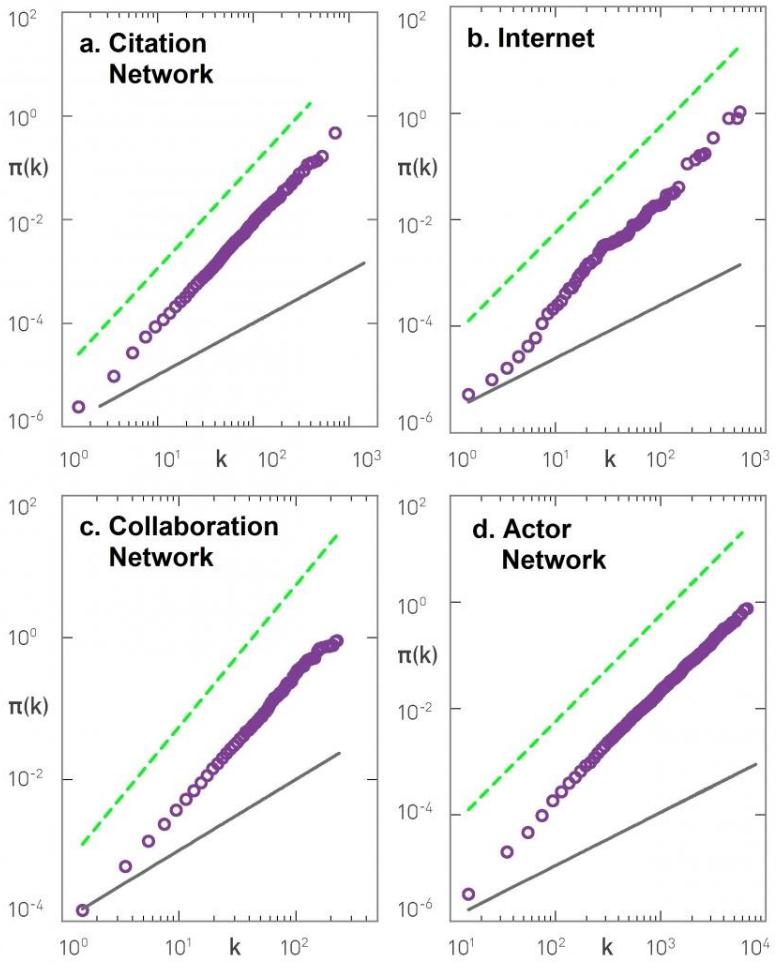
Preferential Attachment
In real networks nodes tend to connect to the more connected nodes.

Scale-Free Networks - Why do we have Hubs?

At each step we add new node to the network, with links connect to nodes already in the network.

The probability of that a link of a new node connected to a present node is depends on the degree of the node: $\prod(k_i) = \frac{k_i}{\sum_j k_j}$





The dashed line refers to the linear preferential attachment, while the solid lines shows the absence of preferential attachment.

Cumulative preferential attachment:

$$\pi(k) = \sum_{k_i=0}^{k} \prod(k_i)$$

Likelihood to connect to a node depends on the degree of the node!