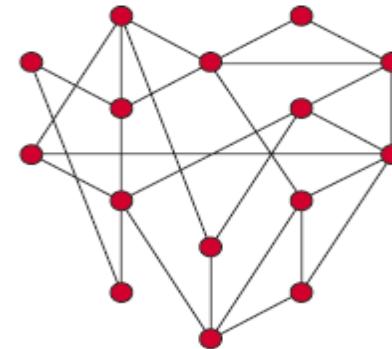




Random graphs

The null model of network science



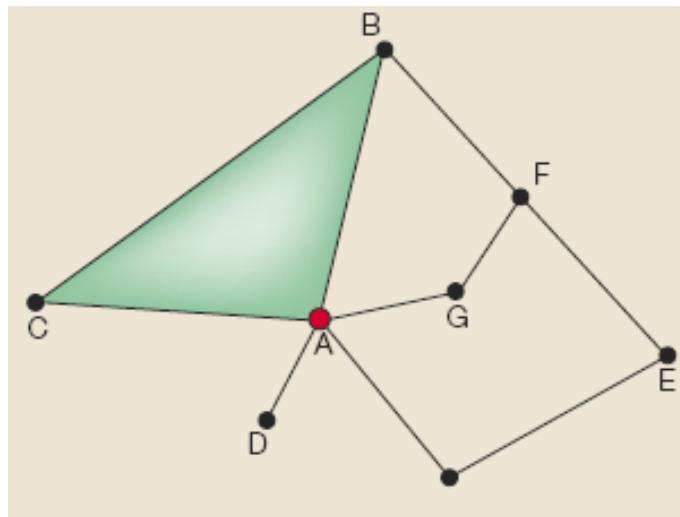
Practical class (this week)

Pick from the following options:

1. Try the exercises listed @ Problem set 1. You may take a look at the algorithms component (ex. 1) and check if you got the main ideas.
2. Try out Gephi (ex. 3 @ Problem Set 1) OR...
3. Start exploring a (more serious) network package associated with your favorite language. This may take a bit. Start as soon as possible (ex. 4 @ Problem Set 1).
4. Try to choose your Project and use it as test bed for (3). Let us know if you need any help.

Last class: characterization of networks

Average distances, diameter, shortest paths, transitivity, clustering, etc, etc.



Among those, We discussed 3:

- Degree distribution, $P(k)$
- Average path length, $\langle L \rangle$
- Clustering Coefficient, C

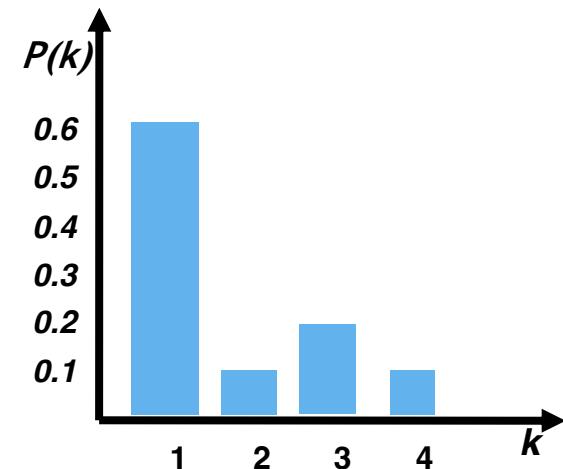
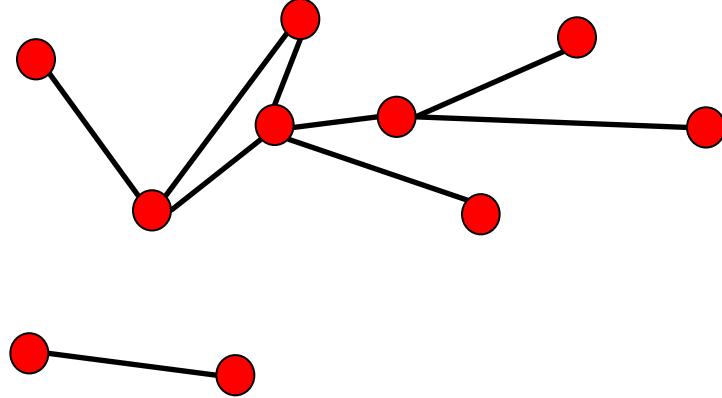
Degree distribution

We have a sample of values x_1, \dots, x_N (in our case degrees of nodes)

Distribution of x (a.k.a. PDF): probability that a randomly chosen value is x

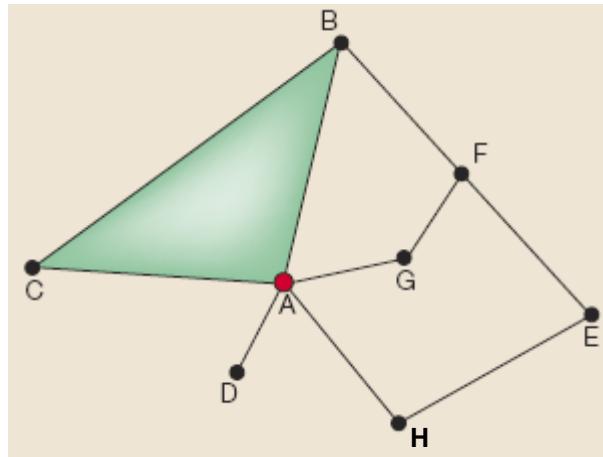
$$P(x) = (\# \text{ values } x) / N$$

$$\sum_i P(x_i) = 1 \text{ always!}$$



Average Path Length (APL), diameter, etc

shortest path & average path length (APL) :



in the example above, there are various alternatives to join C & E; the shortest path d_{CE} is either [CBFE] or [CAHE], which have the same “length”, that is, the number of hops is 3.

the average path length $\langle d \rangle$ is the average over all shortest paths between all pairs of n's of the graph :

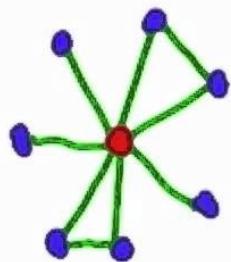
$$\langle d \rangle = \frac{1}{N(N-1)} \sum_{ik} d_{ik}$$

mind your friends' friends

Clustering coeff.

People tend to have friends who are also friends with each other.
How do we measure this?

$$C_i = \frac{\text{\# edges among neigs.}}{\max \text{\# edges among neigs.}} = \frac{e_i}{k_i(k_i - 1)/2}$$



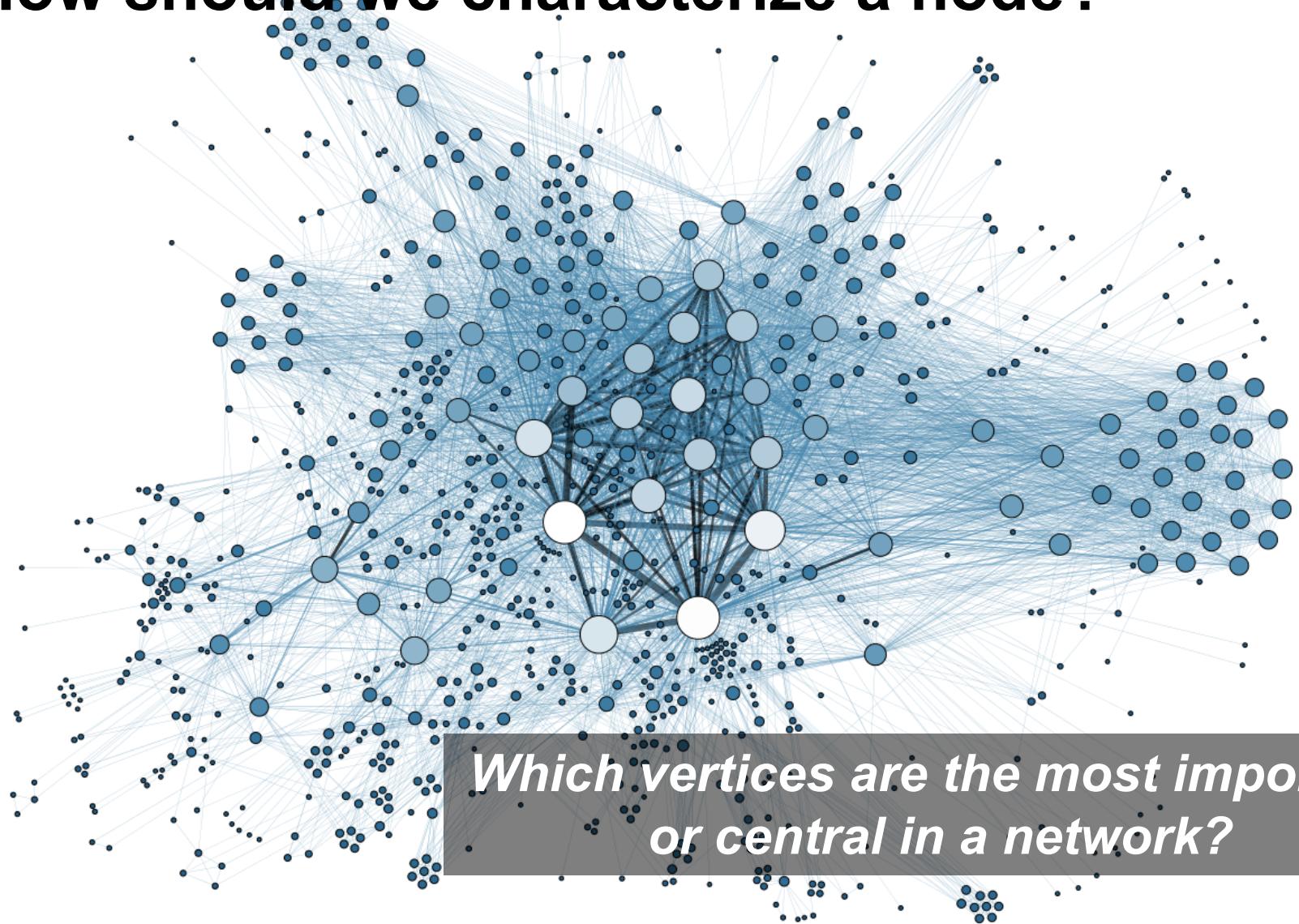
$$k_i=7, e_i=2, \text{ thus}$$

$$C_i = \frac{2}{7(7-1)/2} = 0.095$$

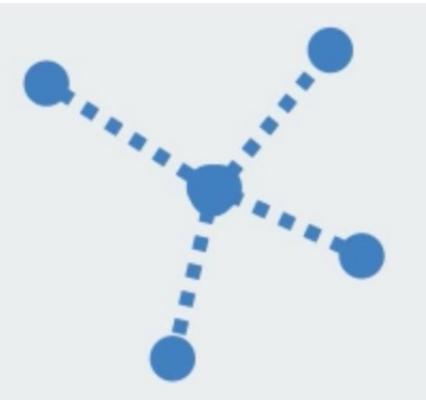
The network clustering coefficient is the average of the clustering coefficients for all nodes

$$\langle C \rangle = \frac{1}{N} \sum_i C_i$$

How should we characterize a node?



Centrality or Importance should depend on the context



How popular you are



How prestigious your friends are

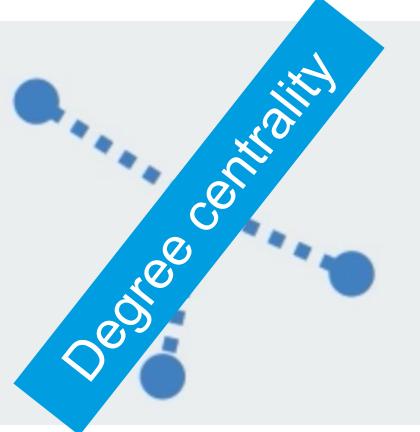


How close you are

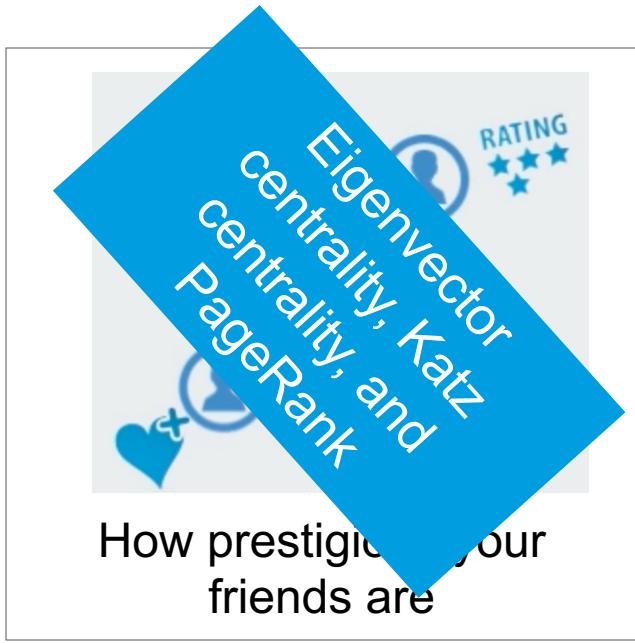


How do you influence the flow of information,
or bridge different parts of the network

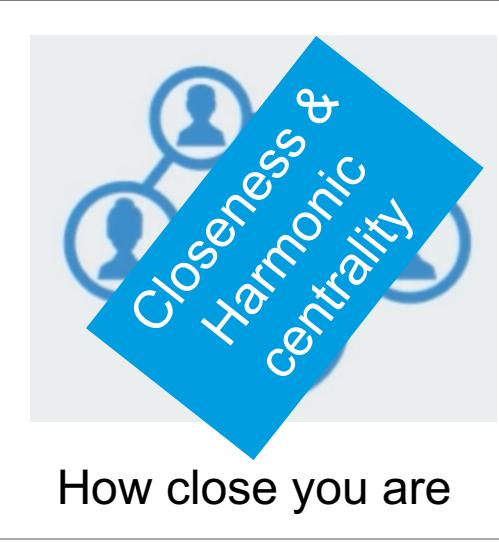
Centrality or Importance should depend on the context



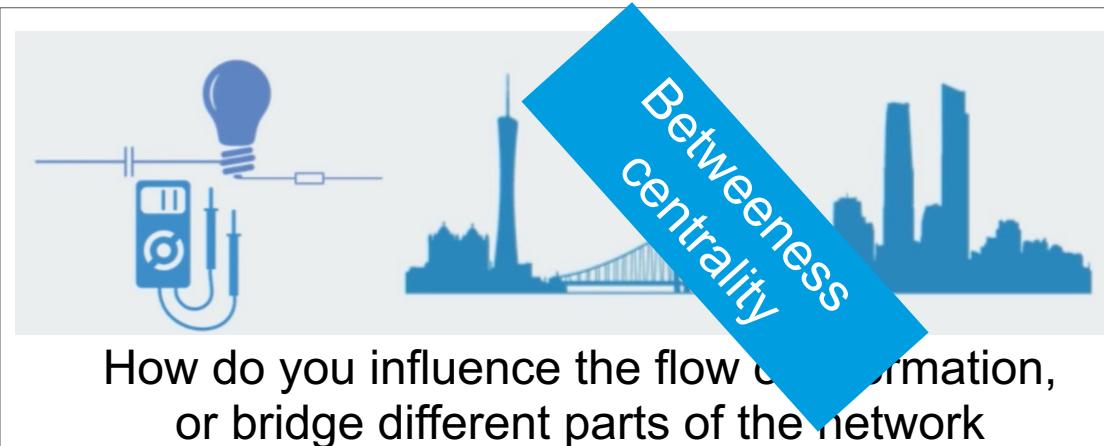
How popular you are



How prestigious your friends are

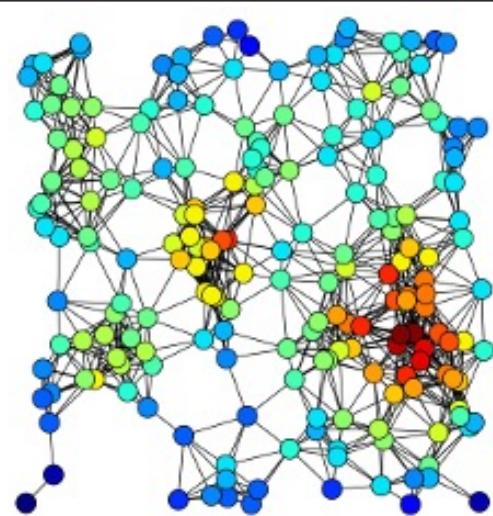


How close you are

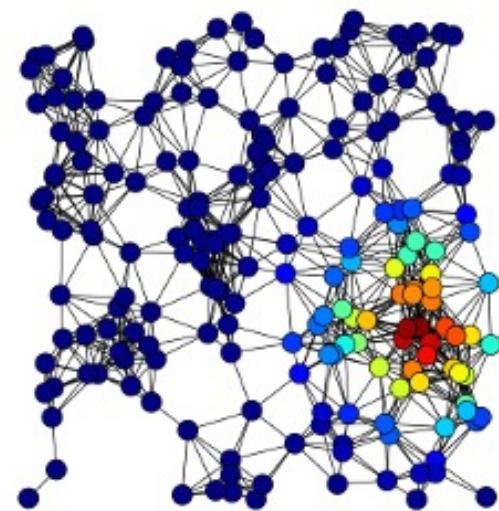


How do you influence the flow of information, or bridge different parts of the network

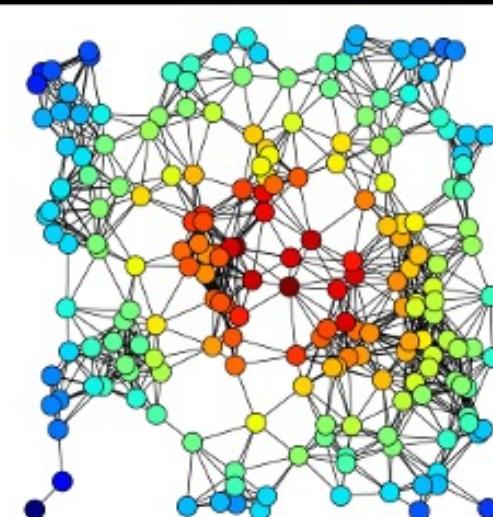
Same network, different metrics, different concepts.



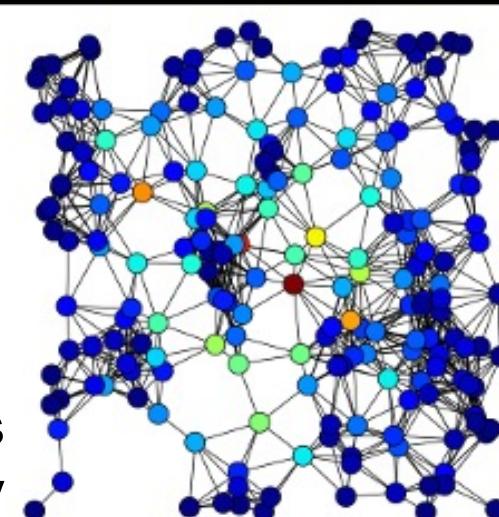
Degree
centrality



Eigenvector,
centrality

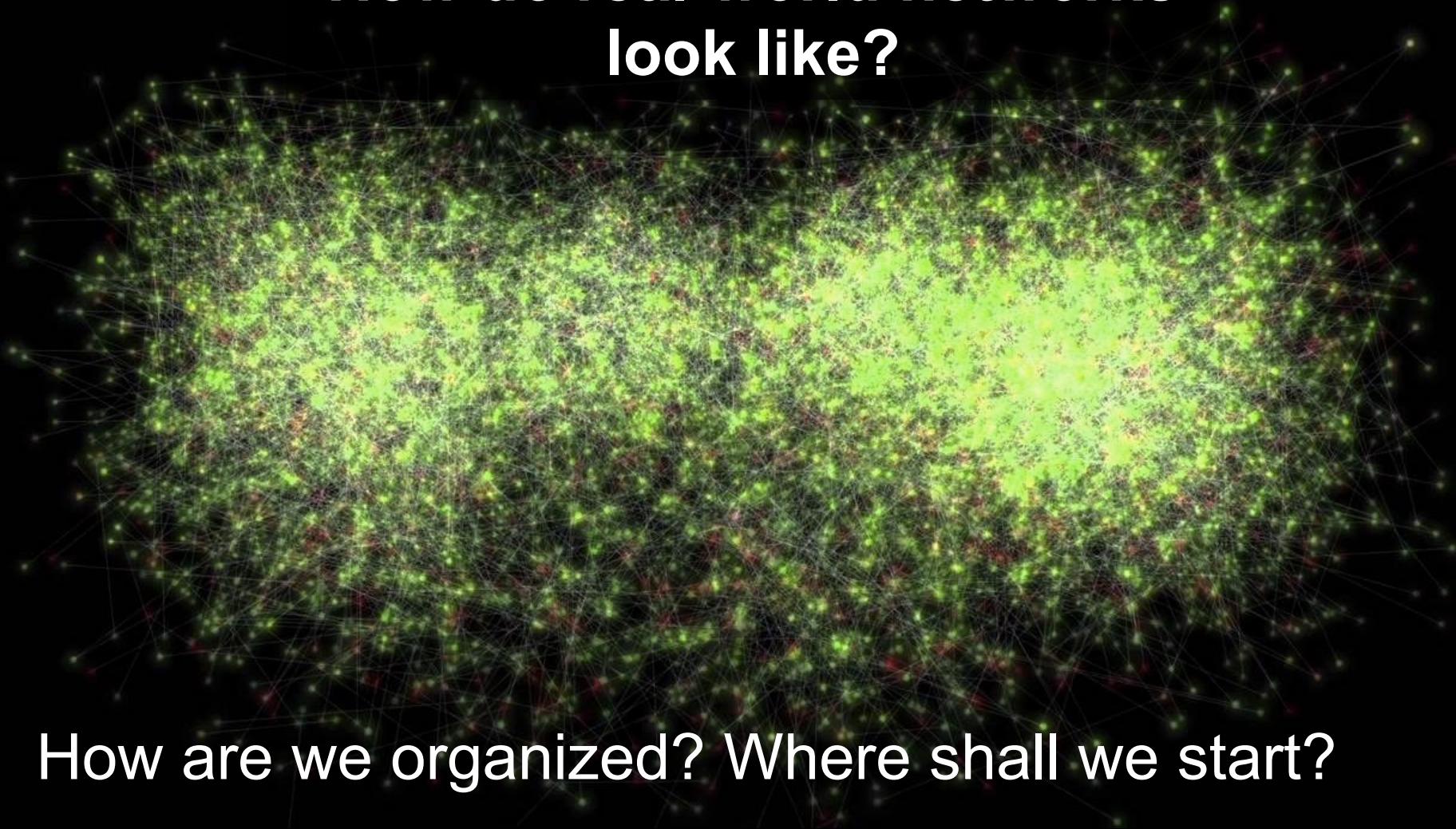


Closeness
Centrality



Betweenness
centrality

How do real-world networks look like?



How are we organized? Where shall we start?

The small world problem

Stanley Milgram (Harvard & Yale, Dep. Psychology, 1967)



Experimenter, 2015

The small-world experiment

Stanley Milgram (Harvard & Yale, Dep. Psychology, 1967)

- **Sources:** 296 letters originating from 296 different people from Omaha, Nebraska, were sent; the instructions in these letters were that they should be sent to whoever the recipient thought might be in better position to forward them to a given target person.
- **Targets:** Milgram chose a stock broker in Boston and a student in Sharon, Massachusetts as targets.



The small-world experiment

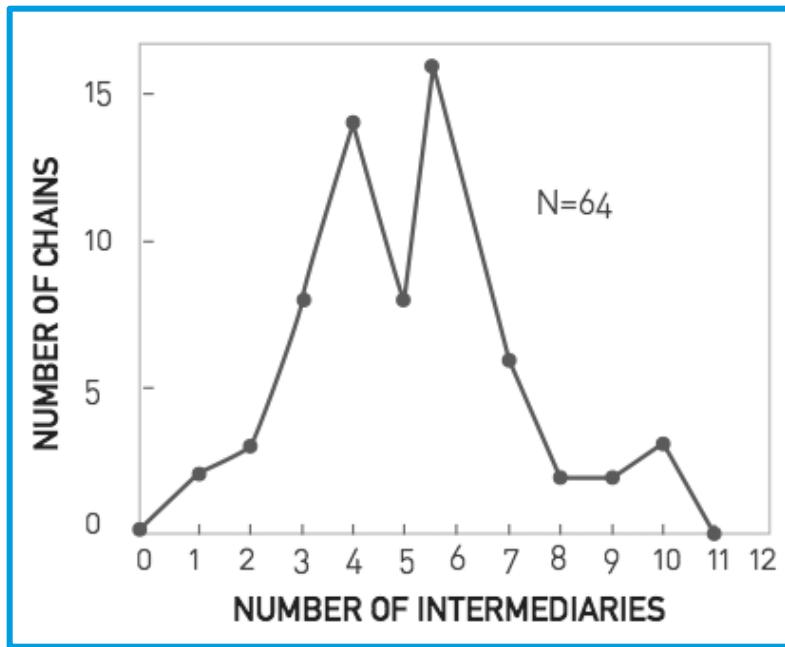
Stanley Milgram (Harvard & Yale, Dep. Psychology, 1967)

- Only 64 letters actually arrived; on average, *each letter hopped 6 times before reaching the designated address*. Hence the designation “*6 degrees of separation*”.
- This is the pioneering experiment on the famous issue of the small-world effect.

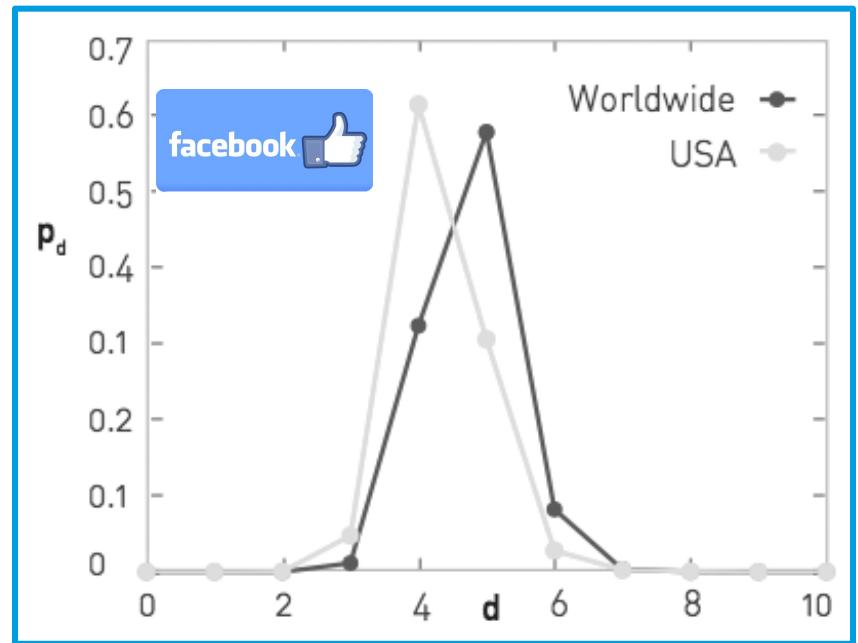


Small Worlds & the 6-degrees of separation

Milgram experiment



Facebook



Modeling small-world properties? Where shall we start?



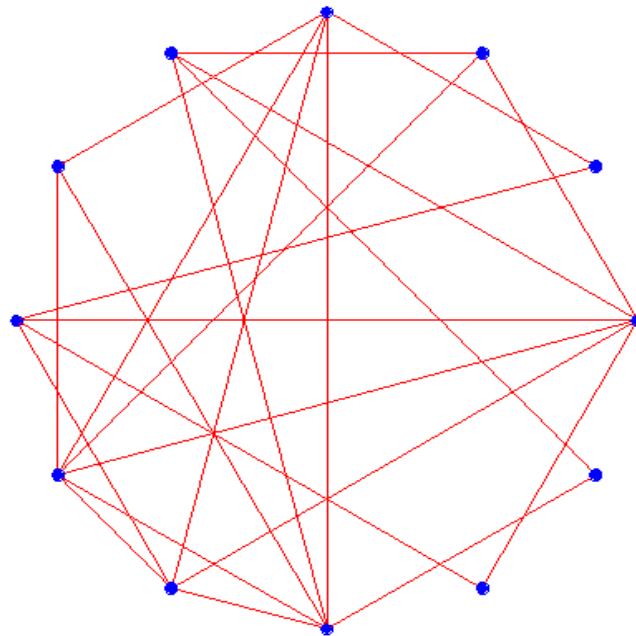
Modeling small-world properties? Where shall we start?



The simplest model one can think...

P. Erdős and A. Rényi, On Random Graphs, Publ. Math. 6, 290 (1959).

- A random network consists of N nodes where each pair of nodes is connected with probability p .

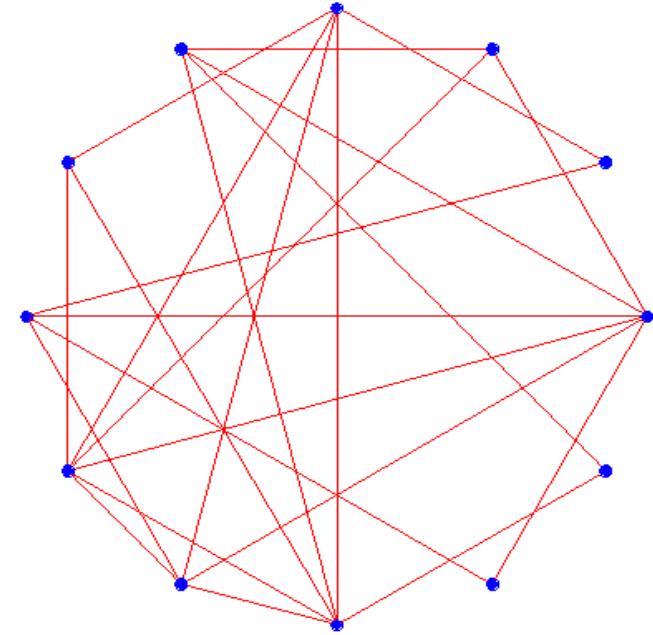


Random network model or the Erdős and Renyi (ER) Model
The null model of network science

The simplest model one can think...

P. Erdős and A. Rényi, On Random Graphs, Publ. Math. 6, 290 (1959).

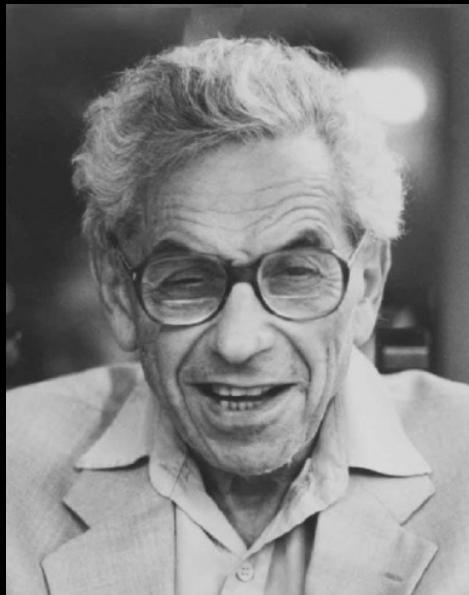
- Number of links?
- Average degree?
- Degree distribution?
- Clustering Coefficient?
- Small-world effect?
- Maximum degree?
- Giant component?



Random network model or the Erdős and Renyi (ER) Model
The null model of network science

Erdős & Rényi

“A persistent rumor has circulates in the USA: There are two intelligent races living on the surface of planet Earth: the standard people and the hungarians.”
– Isaac Asimov



Pál Erdős (1913-1996)



Alfréd Rényi (1921-1970)

Erdős & Rényi

“A persistent rumor has circulates in the USA: There are two intelligent races living on the surface of planet Earth: the standard people and the hungarians.”
– Isaac Asimov



John von Neumann, Paul Erdős, Eugene Wigner e Edward Teller, Leó Szilárd, Theodore von Kármán, Paul Halmos, George Polya e John G. Kemeny



Pál Erdős (1913–1996)

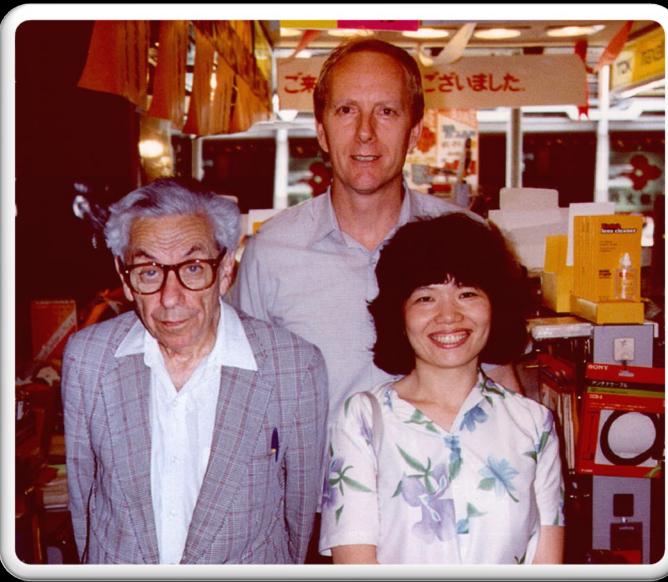
Erdős's hunt for interesting problems took him around the world.

He was a mathematical nomad, living from two half-full suitcases without a chequebook or credit card.

He would turn up at colleagues' houses and making mathematics into a social affair.

This led to him to an astonishing productivity: he published 1,525 papers during his life, with over 500 co-authors!

Pál Erdős (1913–1996)



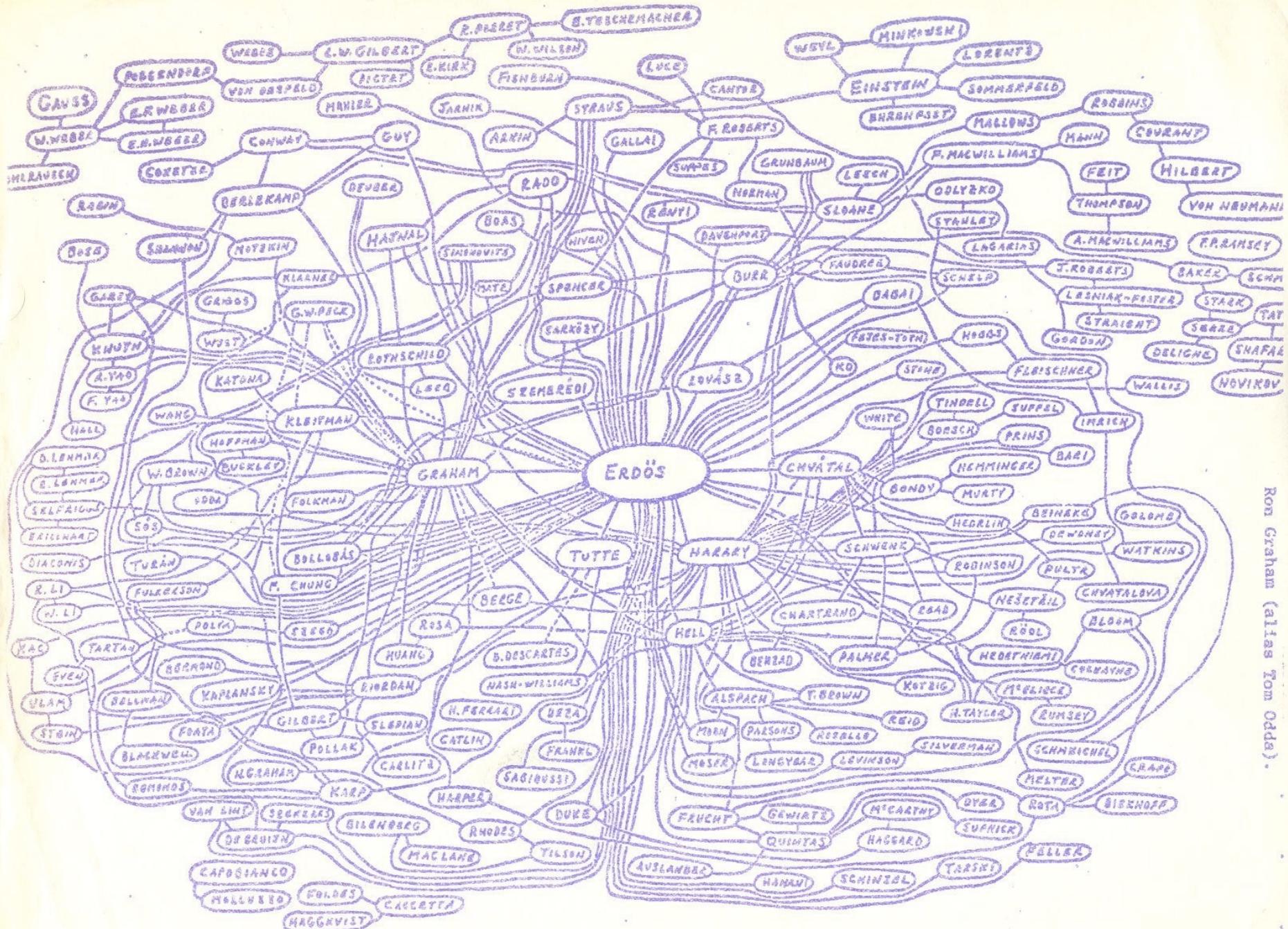
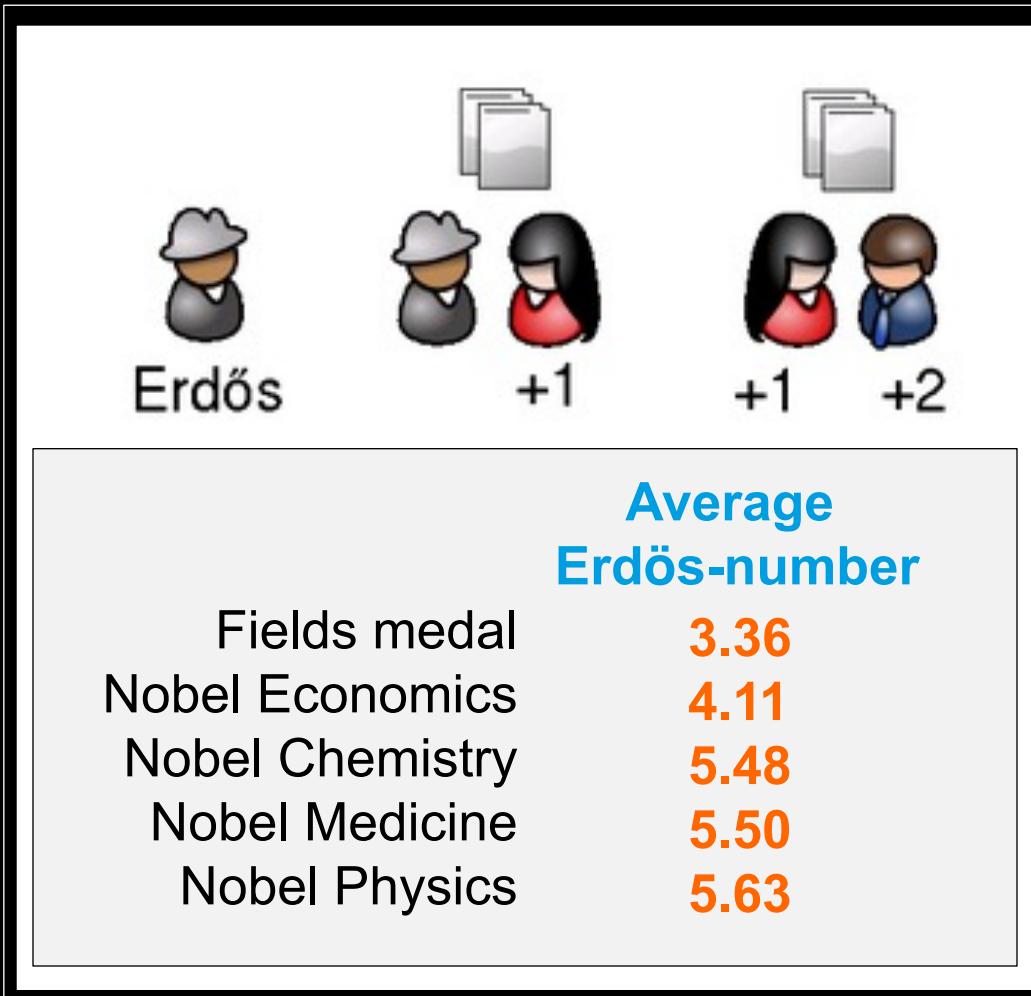


Figure 10

To appear in Topics in Graph Theory (F. Harary, ed.) New York Academy of Sciences (1979).

Science through Erdős



Back to random networks...

$G(N,p)$ model

Each pair of N nodes is connected with probability p .

$G(N,E)$ model

N nodes are connected by E randomly placed links.



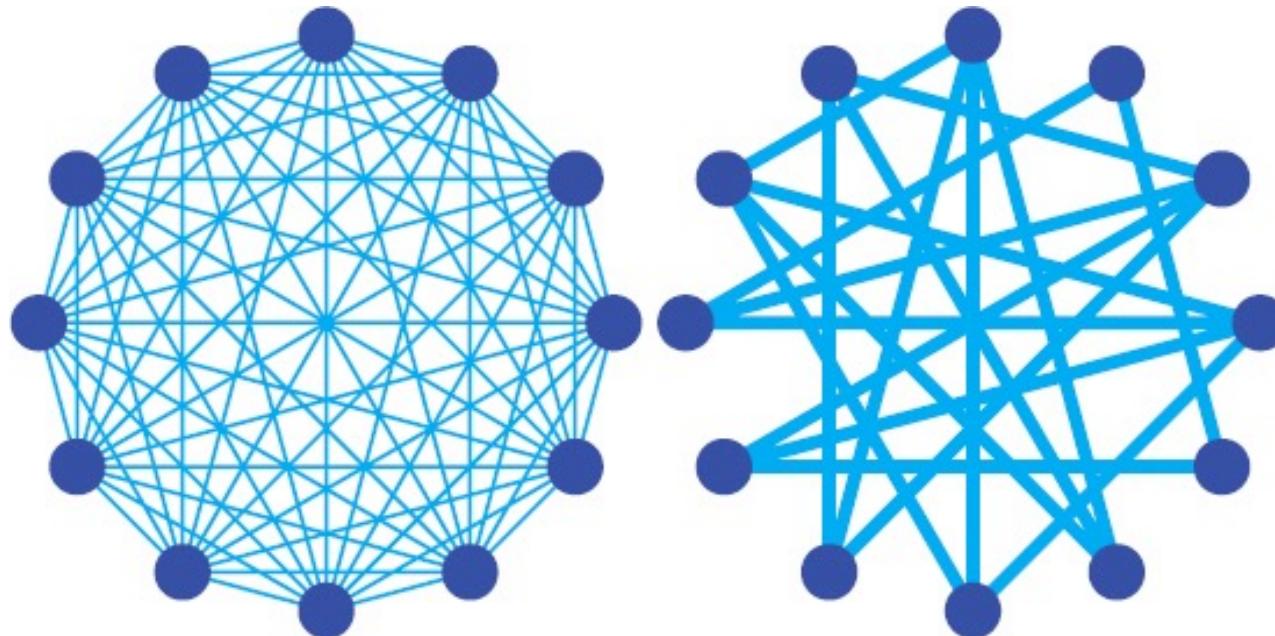
Are real-world networks random?

Model fits empirical data?

- Small world effect ??
- Degree distribution ??
- Maximum degree ??
- Clustering coeff. ??

Some simple and fun math

- What's the average number of links of this network?
- ...or, if you prefer, what's the prob. of having E edges?

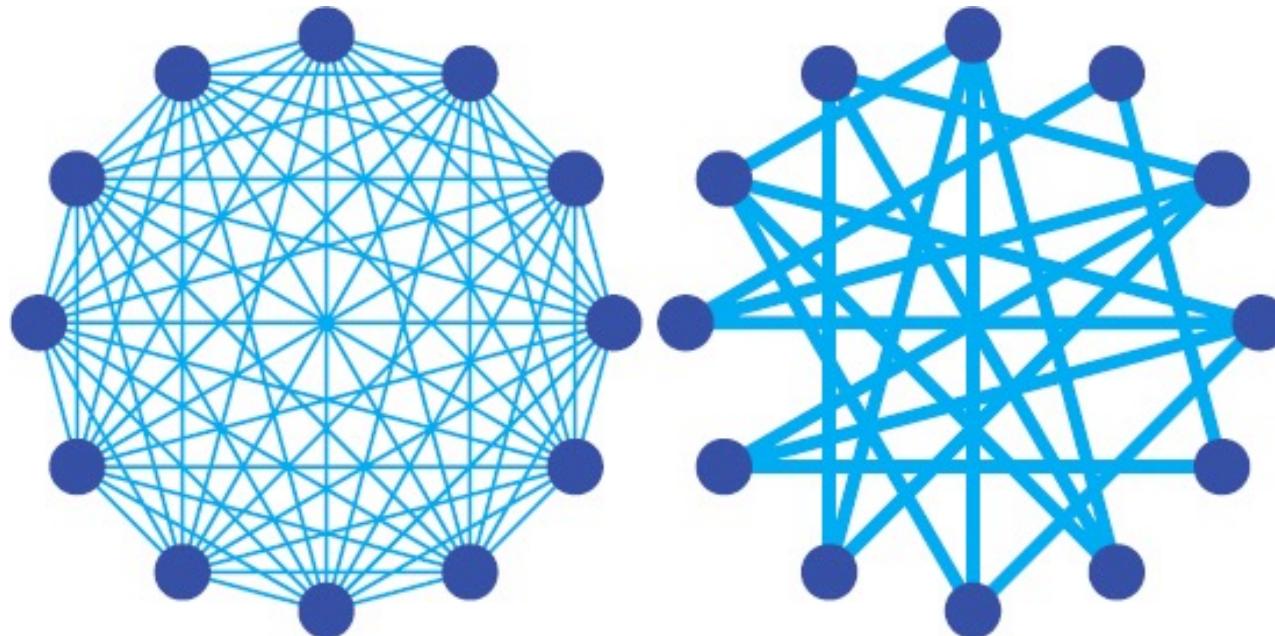


$$p = 1$$

$$0 < p < 1$$

Some simple and fun math

- What's the average number of links of this network?
- ...or, if you prefer, what's the prob. of having E edges?



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

$$E = ??$$
$$0 < p < 1$$

Binomial distribution

- The binomial distribution describes the number of successes in *n independent experiments* with 2 possible outcomes, in which one occurs with prob. p and the other with 1-p.
- The probability of having x successes is given by:

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

- With an average value of

$$\langle x \rangle = \sum_{x=0}^n x P(x) = np$$

Some simple and fun math

- What's the prob. of having E edges?

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

- **Average number of edges** of a random graph

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

Some simple and fun math

- What's the prob. of having E edges?

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

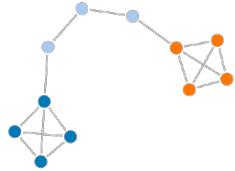
- Average degree of a random graph

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

E_{max}

Average degree increases linearly with p

Try it!

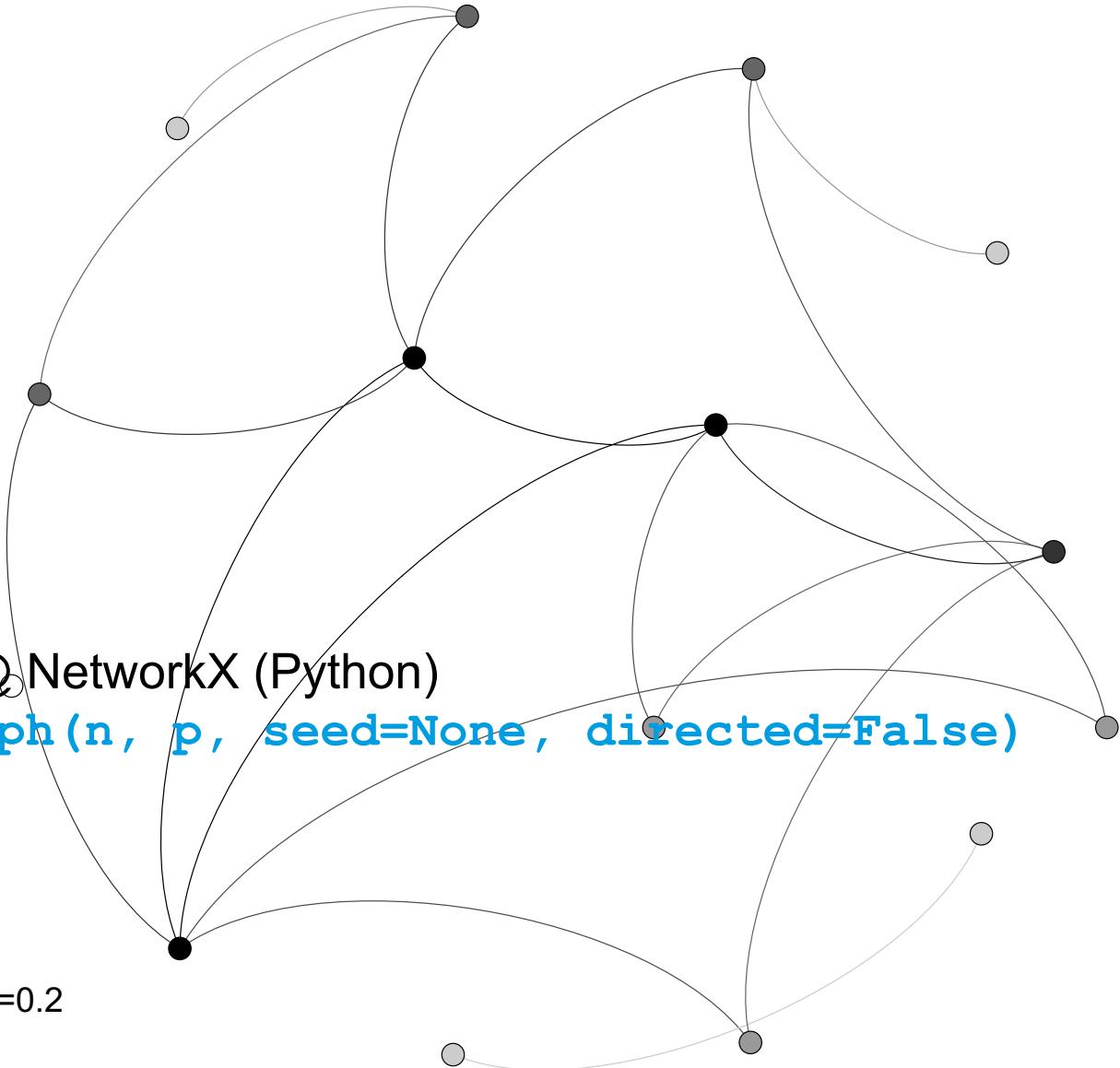


NetworkX

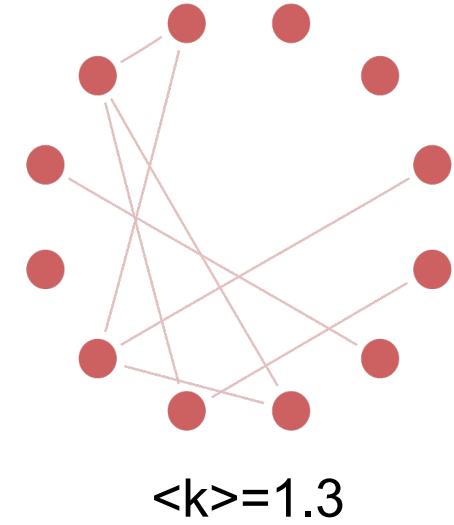
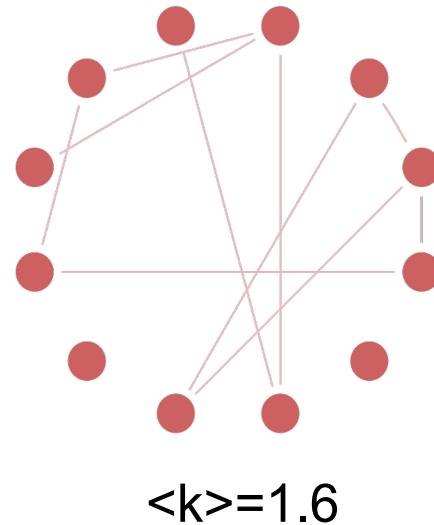
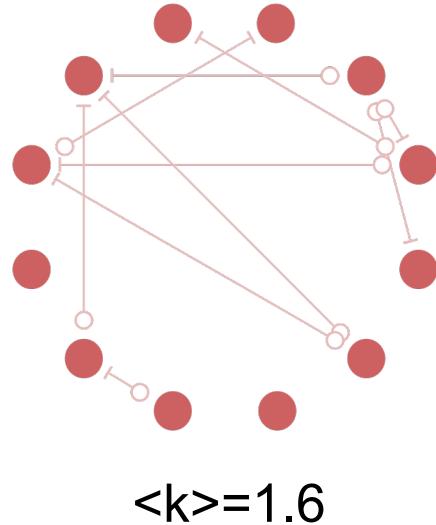
Random networks @ NetworkX (Python)

```
erdos_renyi_graph(n, p, seed=None, directed=False)
```

N=15, p=0.2

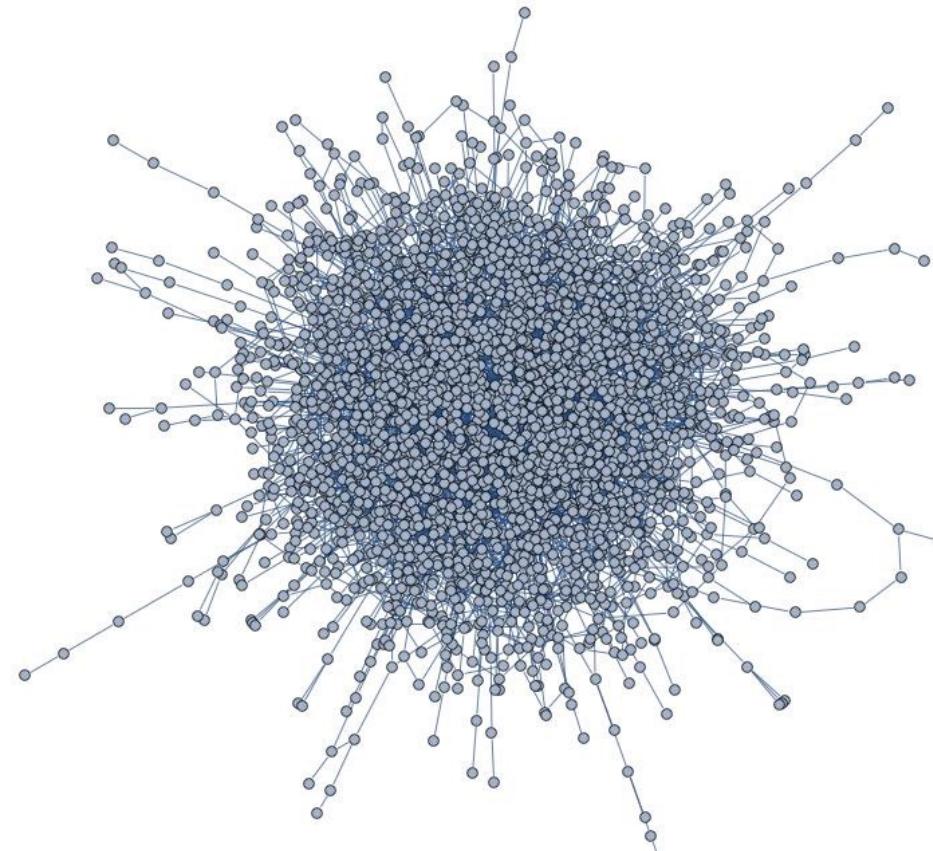


Random graphs are truly “random”. Try it!



$N=12, p=1/6, \text{Expected } \langle k \rangle = 1.833$

Nice! What's the degree distribution of a Erdos-Renyi (ER) random network?



Degree distribution of a random graph

- The probability that a node i has exactly k partners is given by

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

Probability that k links are present \times Probability that $N-1-k$ links are NOT present \times Number of ways one can arrange k links out of $N-1$ possible

Max degree = $N-1$

Degree distribution of a random graph

- The probability that a node i has exactly k partners is given by **a binomial distribution**

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

Number of ways we can arrange k links out of $N-1$ possible

Probability that $N-1-k$ links are NOT present

Probability that k links are present

Max number of links = $N-1$

Degree distribution of a random graph

- The probability that a node i has exactly k partners is given by **a binomial distribution**

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

- Real networks are large and sparse (*i.e.*, $\langle k \rangle \ll N$)
- For large/sparse networks the binomial dist. is well approximated by **a Poisson distribution**:

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

Degree distribution of a random graph

Small nets

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

N>>1


Large nets

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

How would you test the validity of these analytical insights?

Degree distribution of a random graph

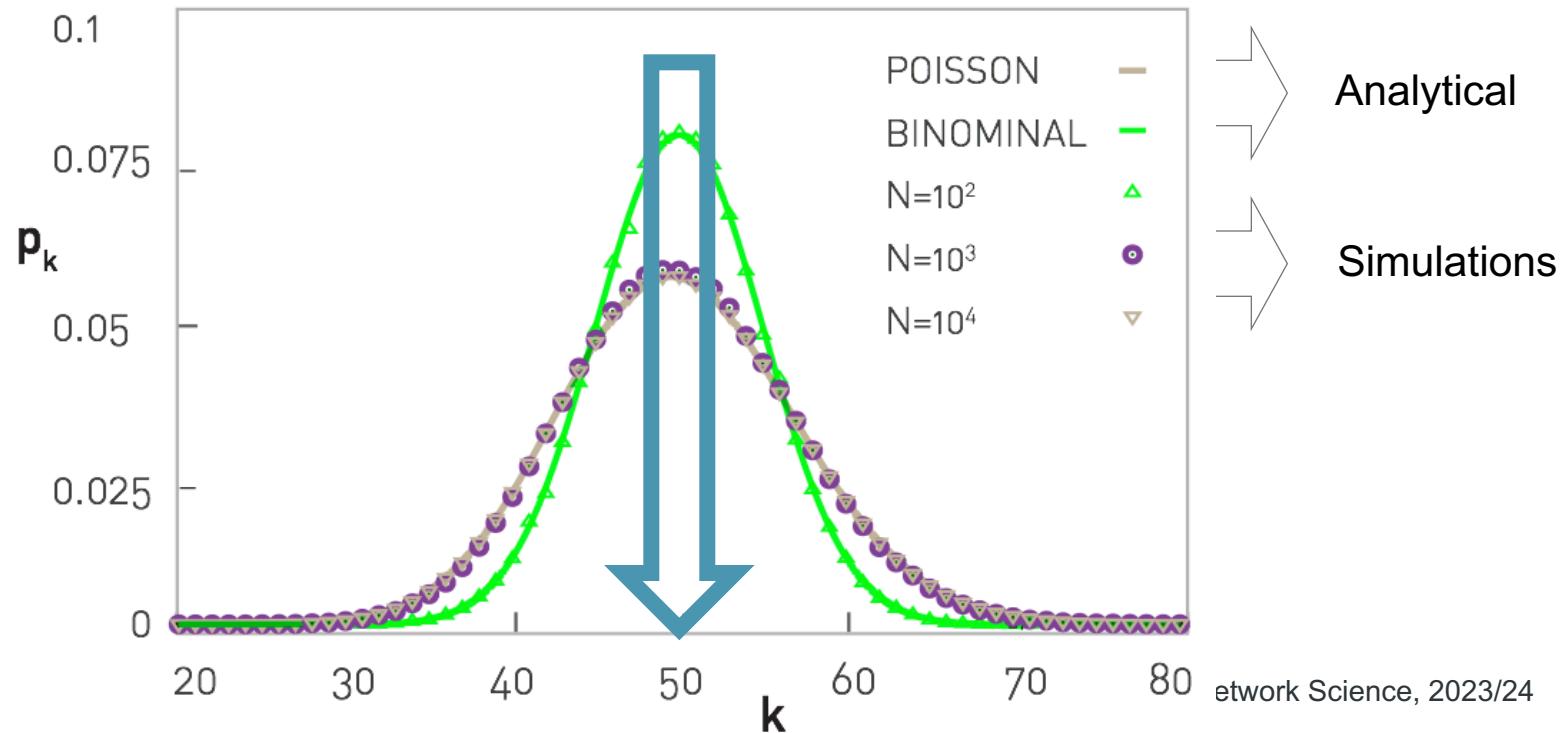
Small nets

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

$N \gg 1$

Large nets

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



Is a random graph a small world?

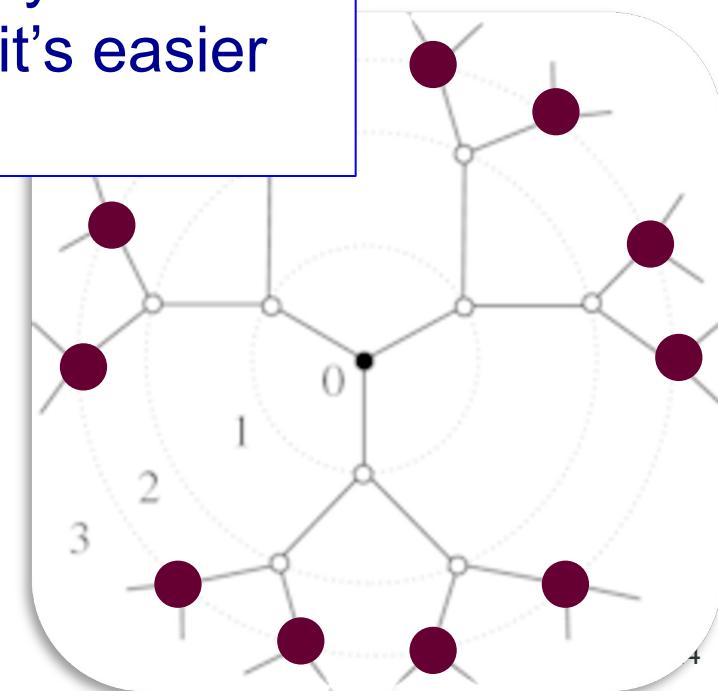
- Can we estimate the **average path length (APL, $\langle d \rangle$)** or the **diameter ($\langle d_{max} \rangle$)** for random nets? Or, at least, how it scales with N?
- Assuming a local interaction radius r , one may say that
 - 1 individual at distance 0
 - k^1 individuals at $d=1$
 - k^2 individuals at $d=2$
 - ...
 - k^n individuals at $d=n$

$$n(d) \sim k^d$$

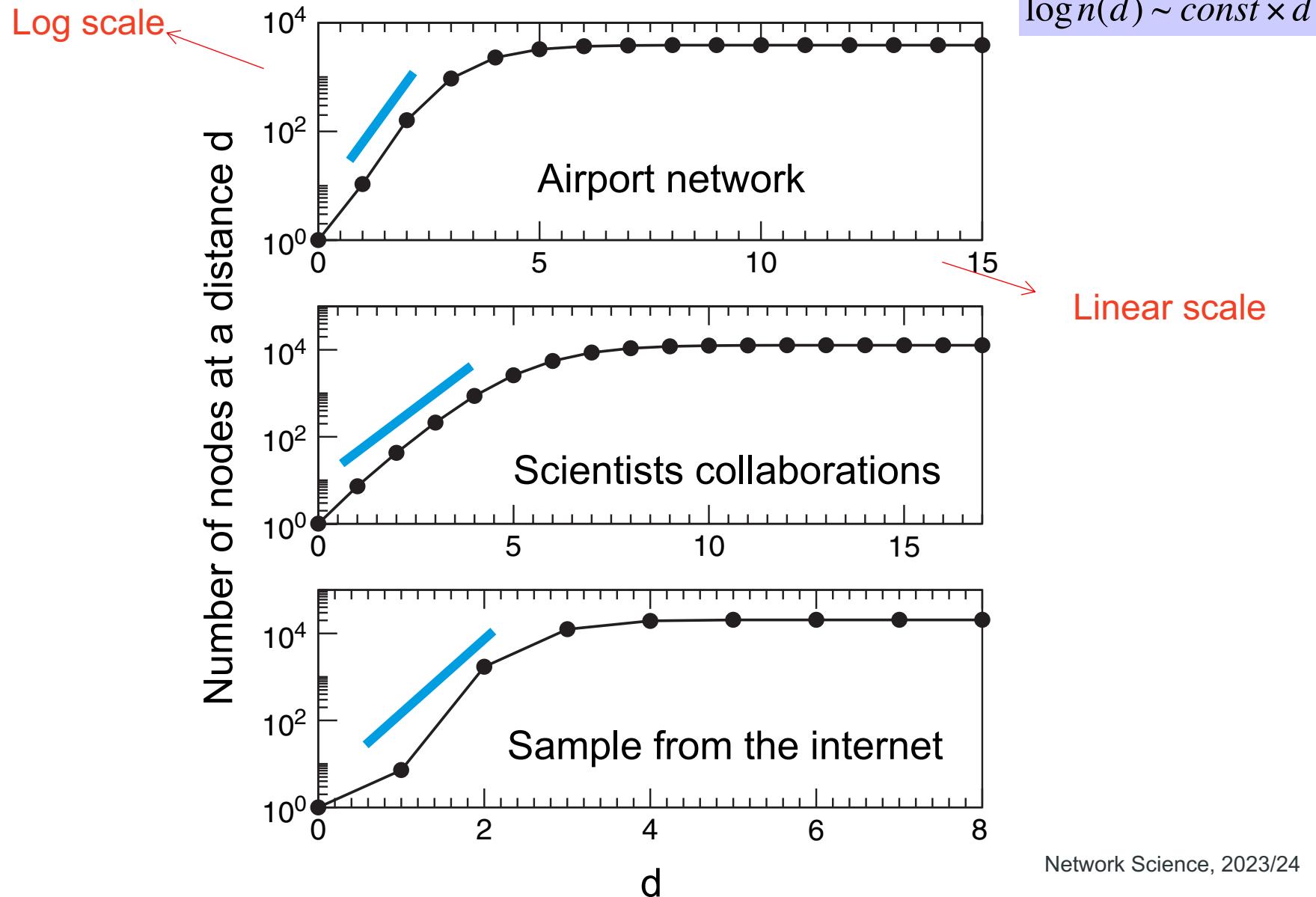
$$\log n(d) \sim \text{const} \times d$$

The APL is trickier to compute analytically... Yet, the diameter usually scales with the APL, and it's easier to grasp!

e, i.e., one



Is it a good approximation? YES!!



Small world?

- Can we estimate the APL or the diameter for random nets?

YES. Let's compute the number of nodes @ $d \leq d_{\max}$

- Thus, the number of individuals at $d \leq n$ is given by

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

- Networks are finite, thus $N(d_{\max}) = N$

- Assuming $N(d_{\max}) = \frac{\langle k \rangle^{d_{\max}+1} - 1}{\langle k \rangle - 1} \approx \frac{\langle k \rangle^{d_{\max}+1}}{\langle k \rangle} \Rightarrow N = \langle k \rangle^{d_{\max}}$

Small world?

- Can we estimate the APL or the diameter for random nets?
YES. Let's compute the number of nodes @ $d \leq d_{\max}$
- Thus, the number of individuals at $d \leq n$ is given by

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

- Networks are finite, thus $N(d_{\max}) = N$

- We get

$$d_{\max} \approx \frac{\ln N}{\ln \langle k \rangle}$$

Small world?

We reached to the diameter of a random network

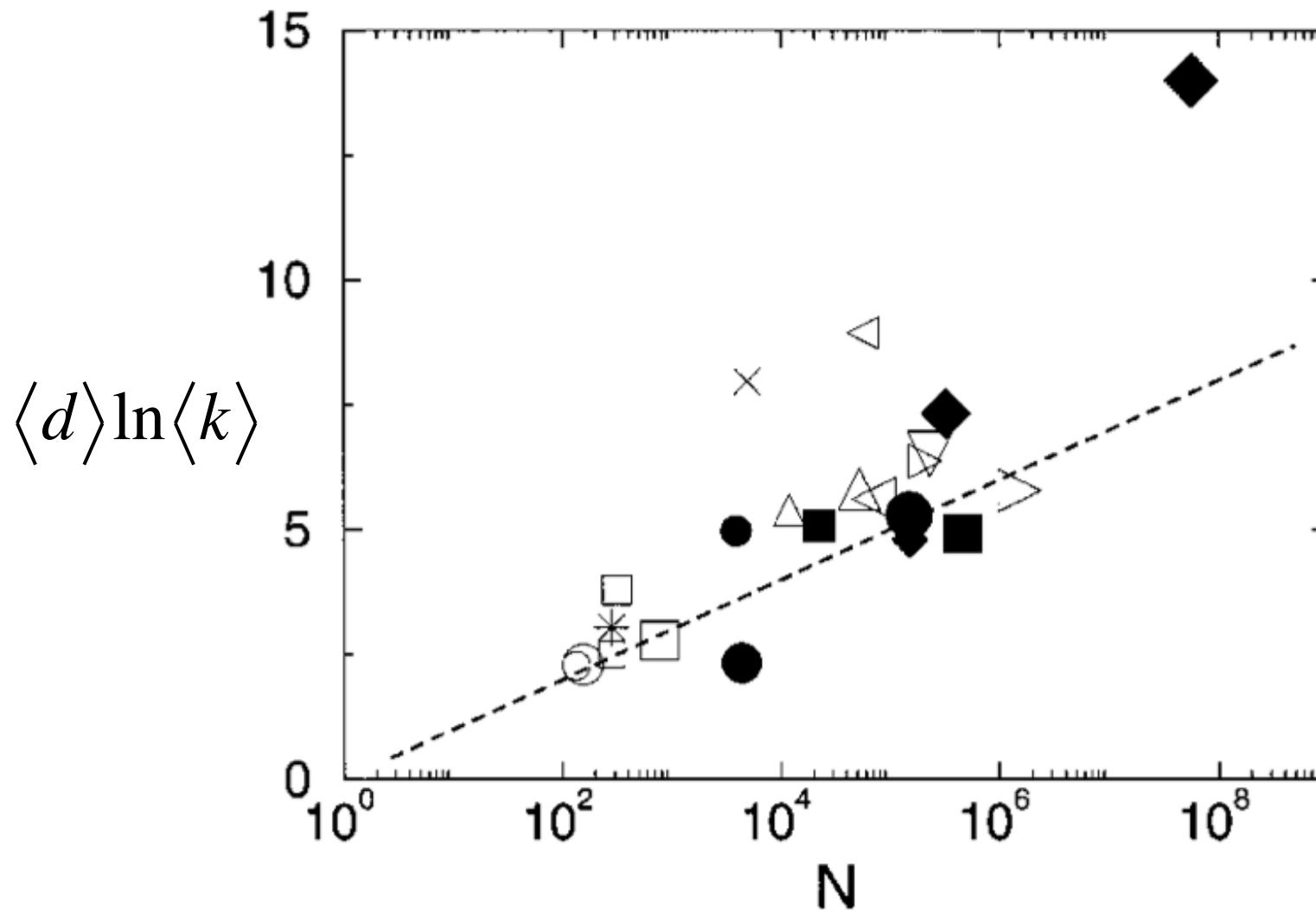
$$d_{\max} \approx \frac{\ln N}{\ln \langle k \rangle}$$

d_{\max} is often dominated by the extreme parts, being subject to large fluctuations. The average path length $\langle d \rangle$ will not suffer from this, and is often well approximated by this scaling

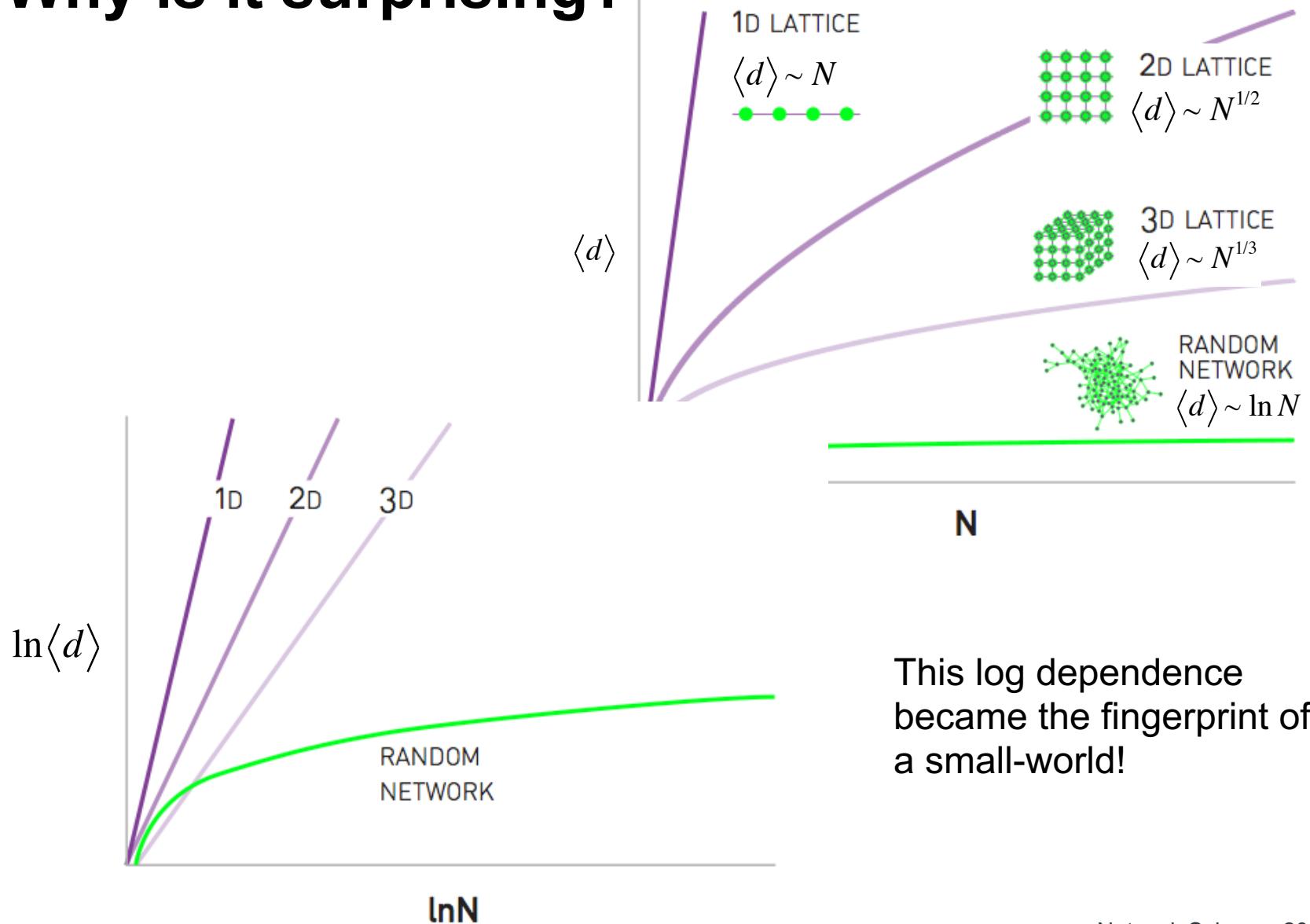
$$APL = \langle d \rangle \approx \frac{\ln N}{\ln \langle k \rangle}$$

constant

How does the APL scale with N? It works!!!



Why is it surprising?



It works!!!

Empirical data

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

Random networks: Overview (1)

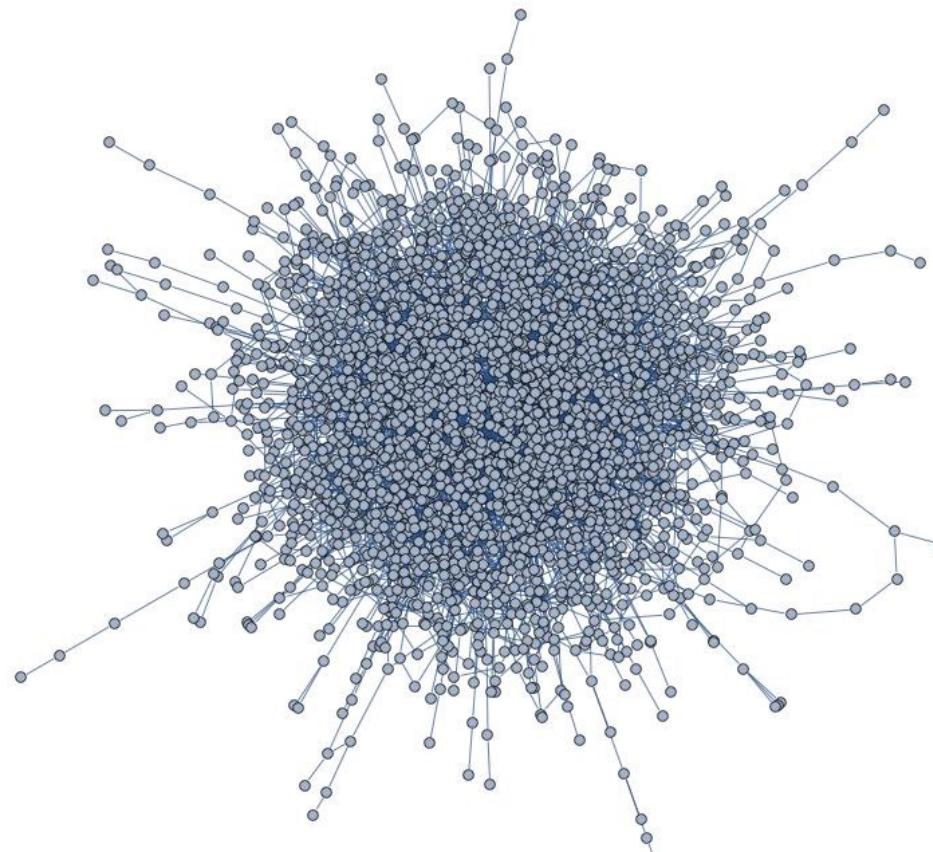
Model fits empirical data?

- Small world effect **YES!!**

Nice! Let's continue to compare the ER model with real networks...

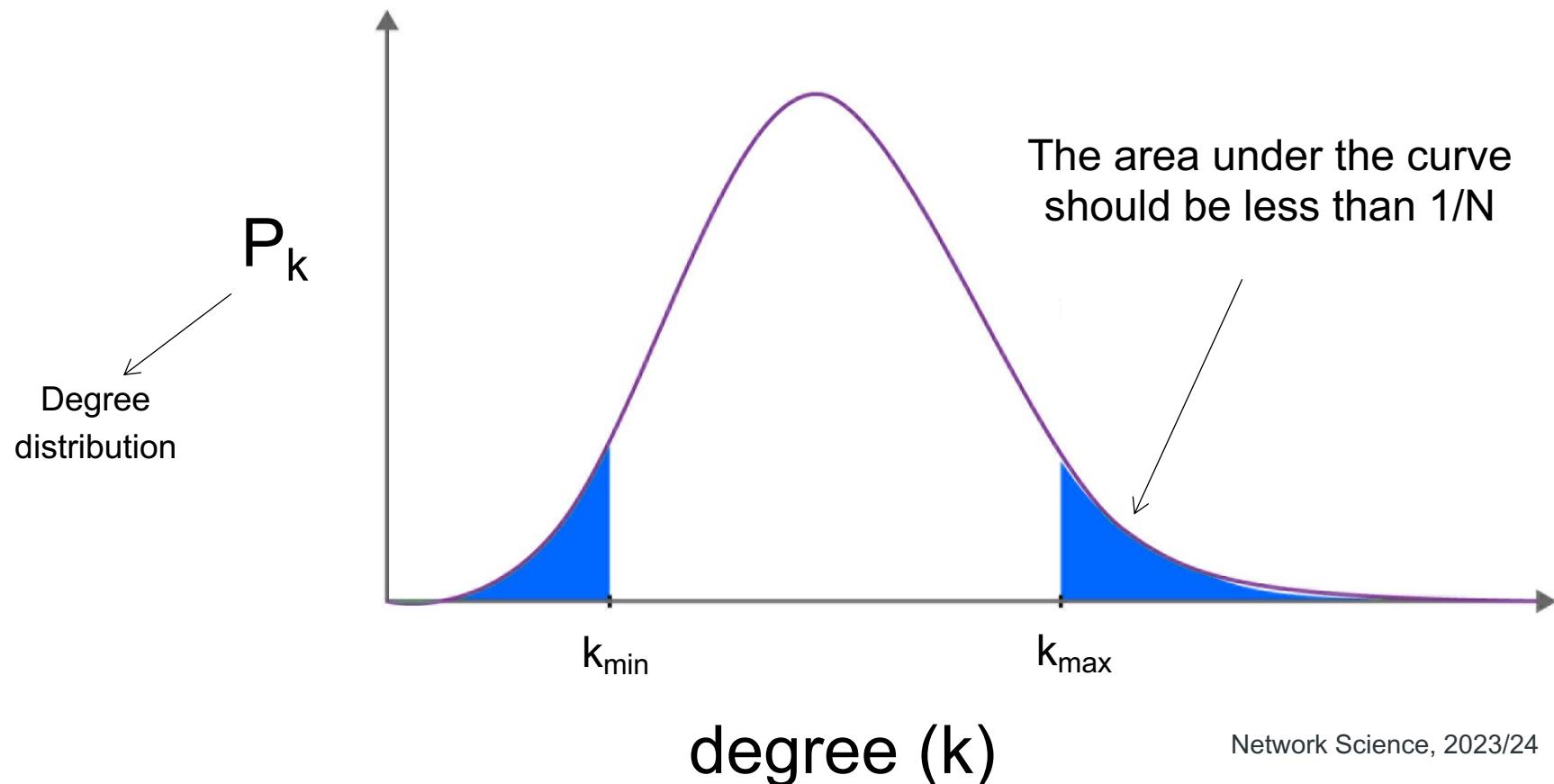
Nice! Let's continue to compare ER networks with real networks...

- Can we estimate the maximum and minimum degree of a random network?



Nice! Let's compare with real networks...

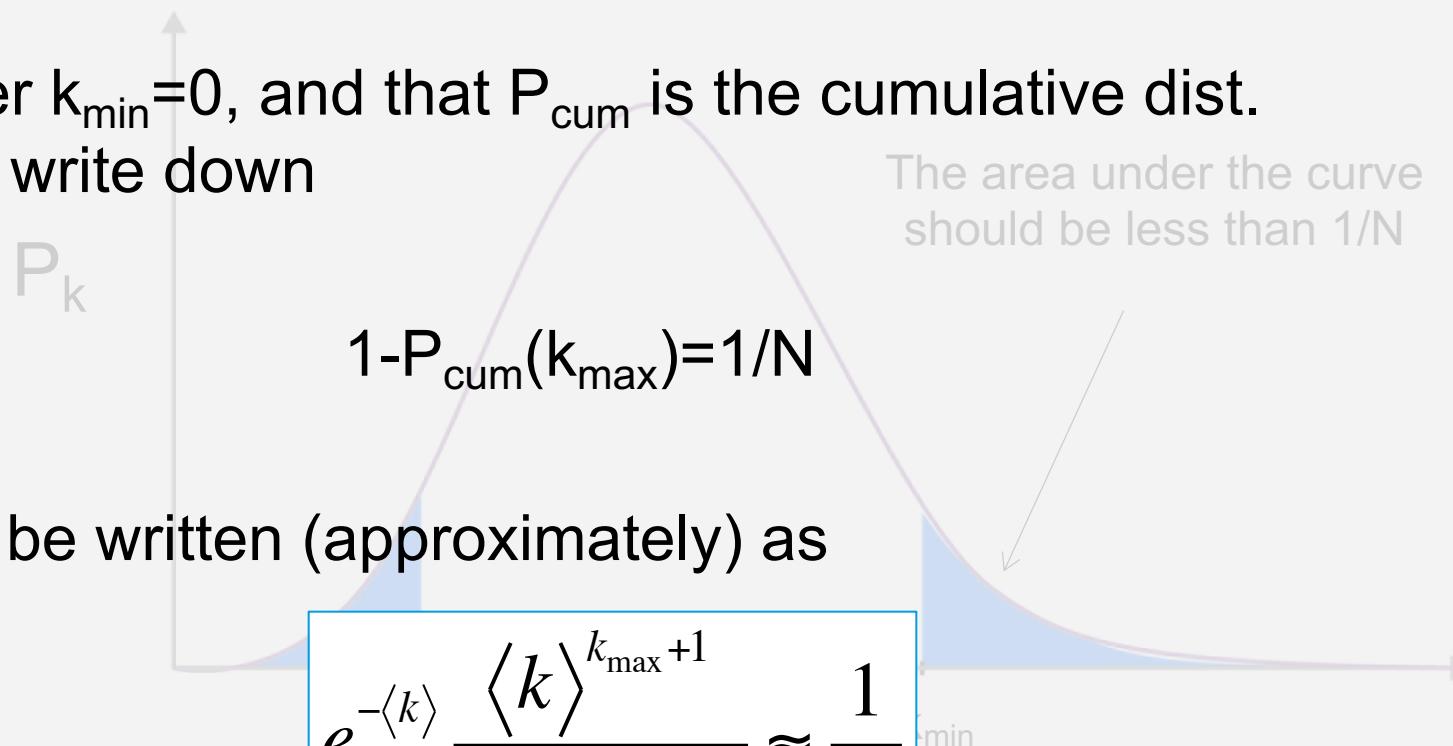
- Can we estimate the maximum and minimum degree of a random network?



Nice! Let's compare with real networks...

- Can we estimate the maximum and minimum degree of a random network?

- Consider $k_{\min}=0$, and that P_{cum} is the cumulative dist. We can write down



which can be written (approximately) as

$$e^{-\langle k \rangle} \frac{\langle k \rangle^{k_{\max}+1}}{(k_{\max} + 1)!} \approx \frac{1}{N}$$

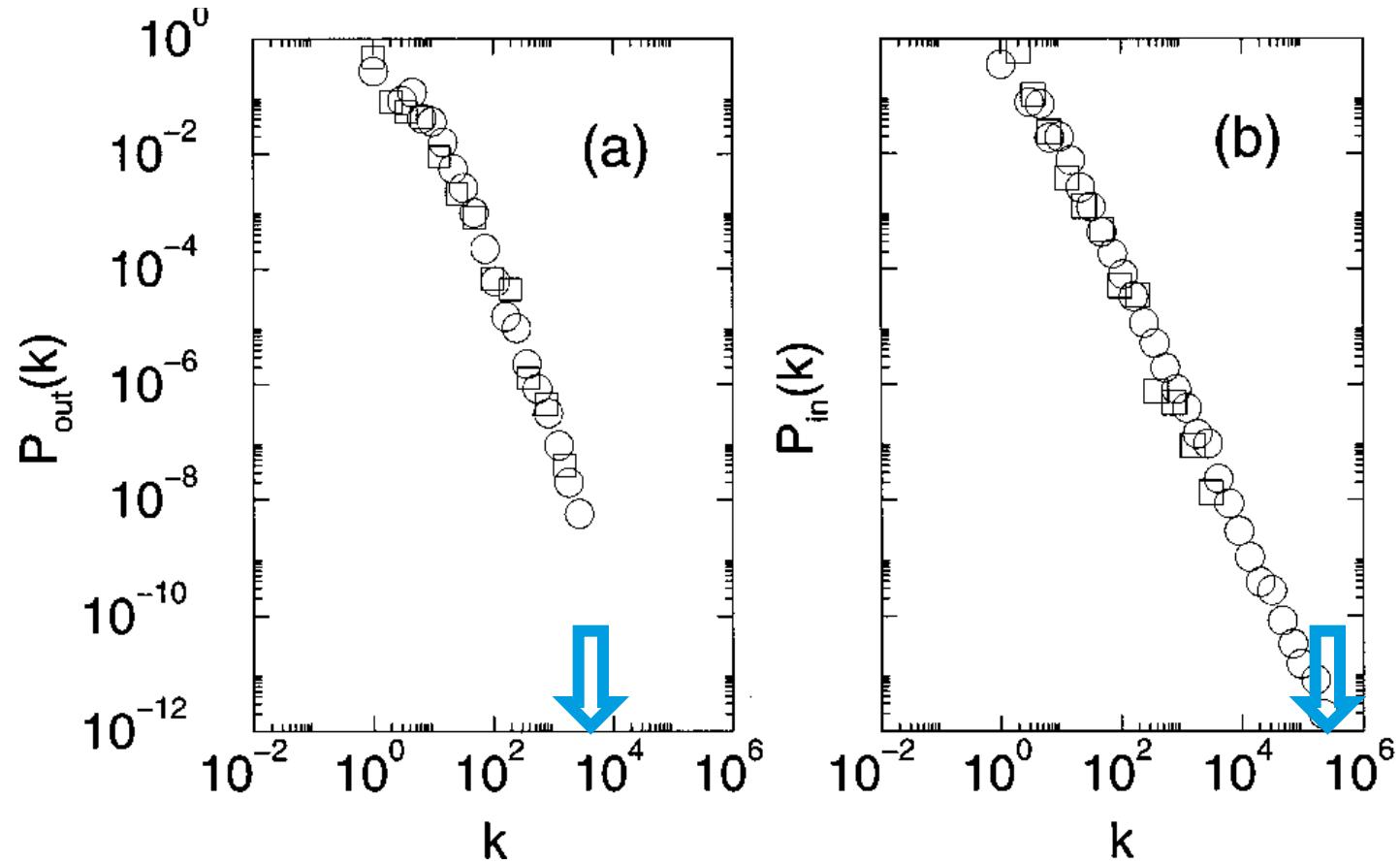
Nice! Let's compare with real networks...

- What's the maximum and minimum degree of a random network?
- Let's say we have $N=2 \times 10^8$ documents in the WWW (estimates indicate $\langle k \rangle \sim 4.60$).
- What would be the **maximum degree of the WWW** if we had a random graph?

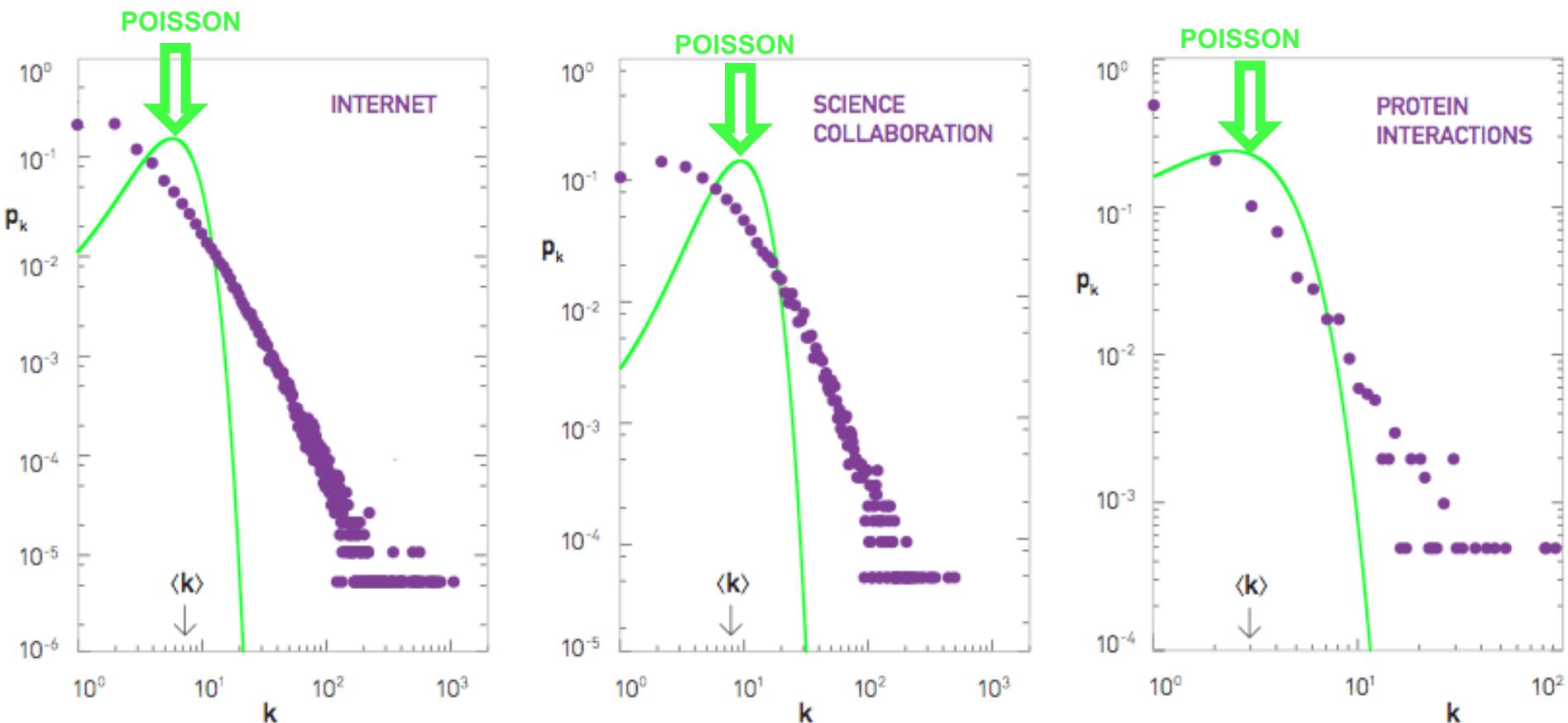
$k_{\max} \sim 21 \dots$ Weird!!

We are lacking hubs in ER networks!

Degree distribution of the World-Wide-Web

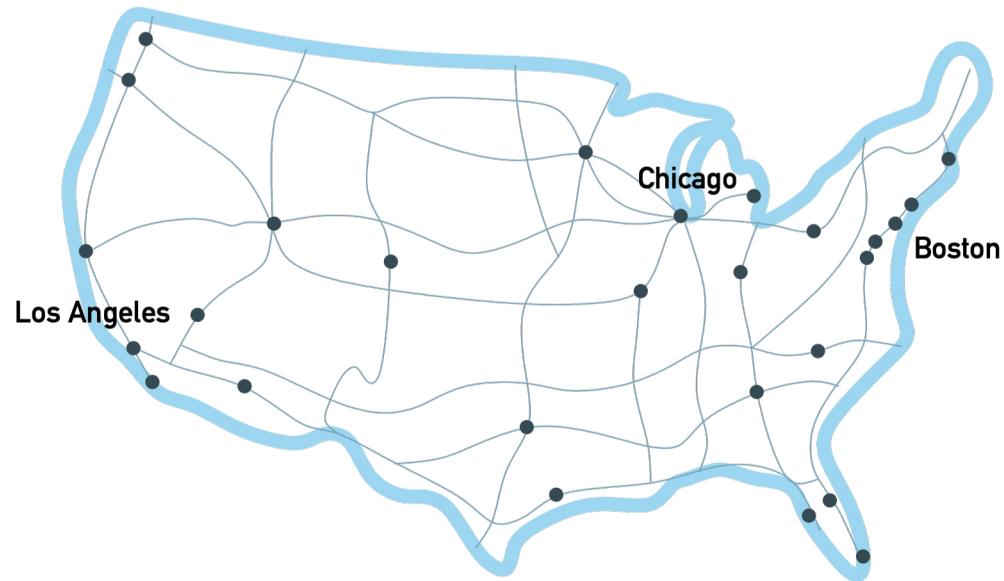
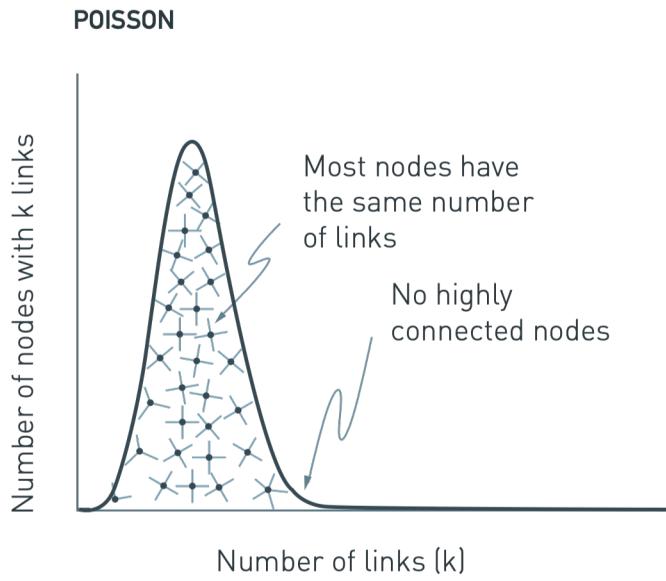


Real degree distributions



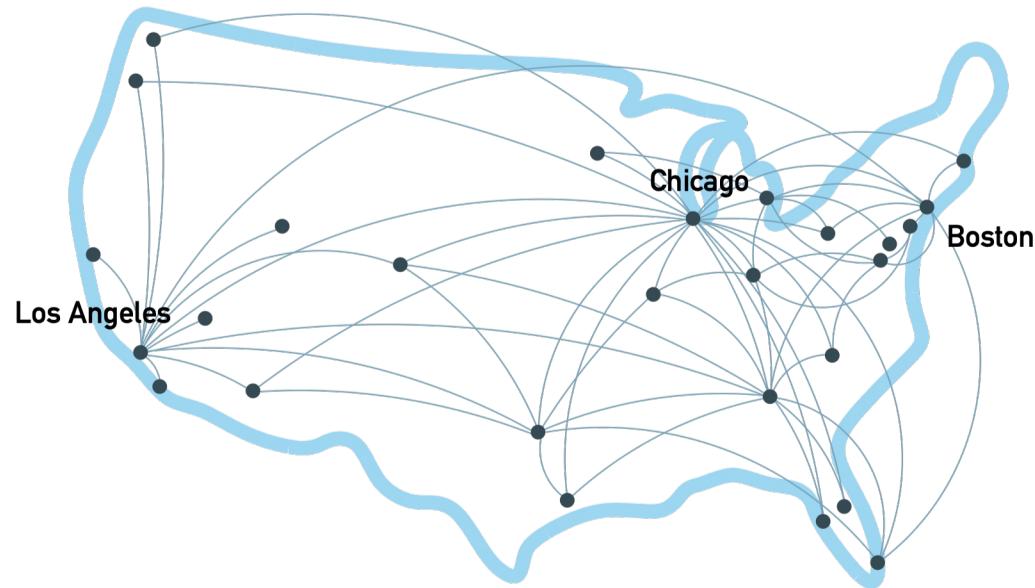
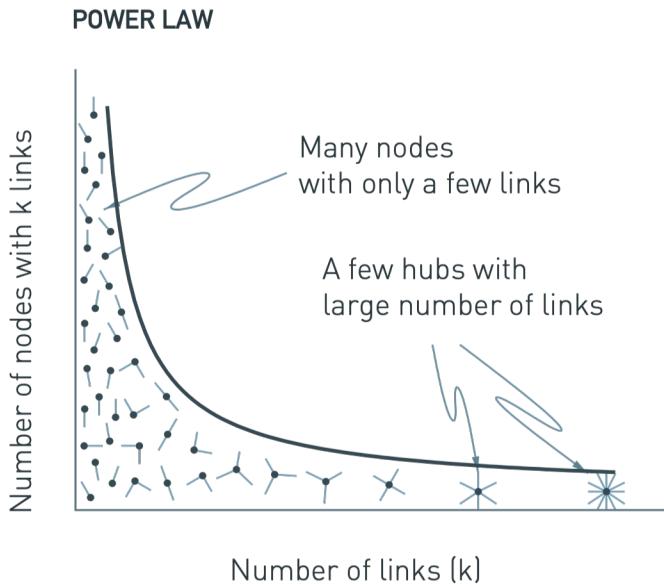
Real degree distributions

In some cases, random networks provide an accurate description



Real degree distributions

...but in many other cases it fails.



Random networks: Overview (2)

Model fits empirical data?

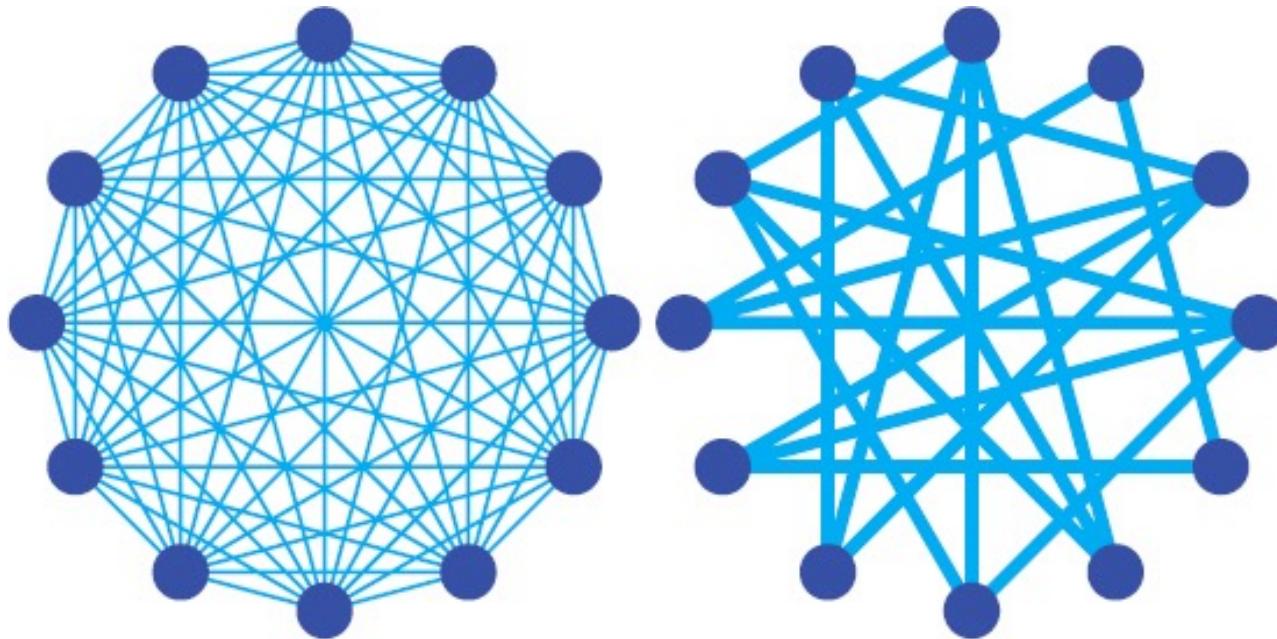
- Small-world effect **YES!**

- Degree distribution **Sometimes. Mostly NO!**

- Maximum degree **Sometimes. Mostly NO!**

Mind your friend's friends

- Can we compute the clustering coefficient for random nets?



$$C_i = \frac{\text{\# edges among neigs.}}{\max \text{\# edges among neigs.}} = \frac{e_i}{k_i(k_i - 1) / 2}$$

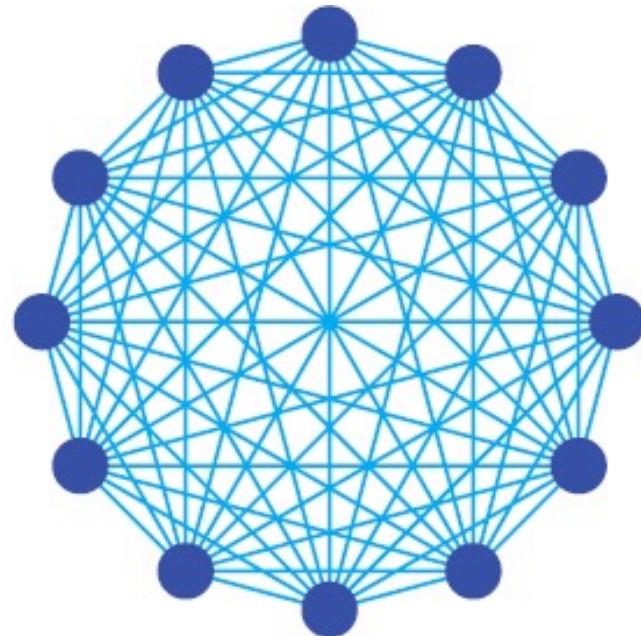
Mind your friend's friends

- Can we compute the clustering coefficient for random nets?

Average number of links e_i in the neighborhood of a node i with k_i partners?

$$\frac{k_i(k_i - 1)}{2}$$

complete graph

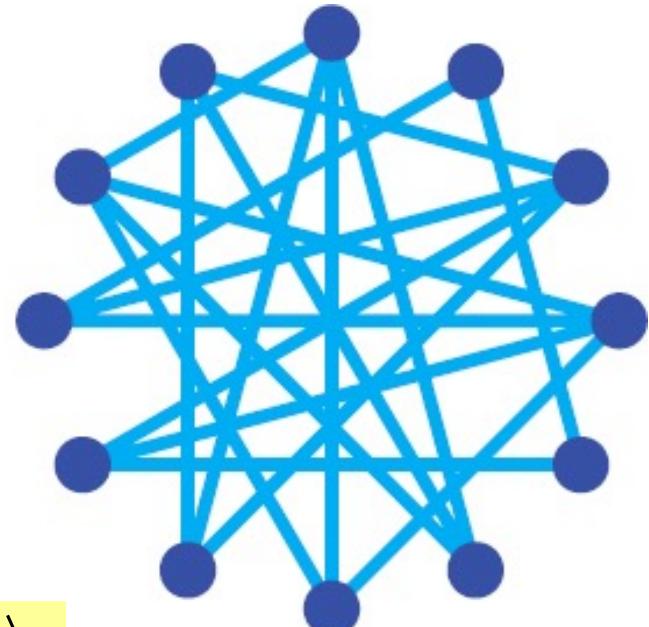


Mind your friend's friends

- Can we compute the clustering coefficient for random nets?

Average number of links e_i in the neighborhood of a node i with k_i partners?

$$p \times \frac{k_i(k_i - 1)}{2}$$



Thus, the clustering coefficient reads

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

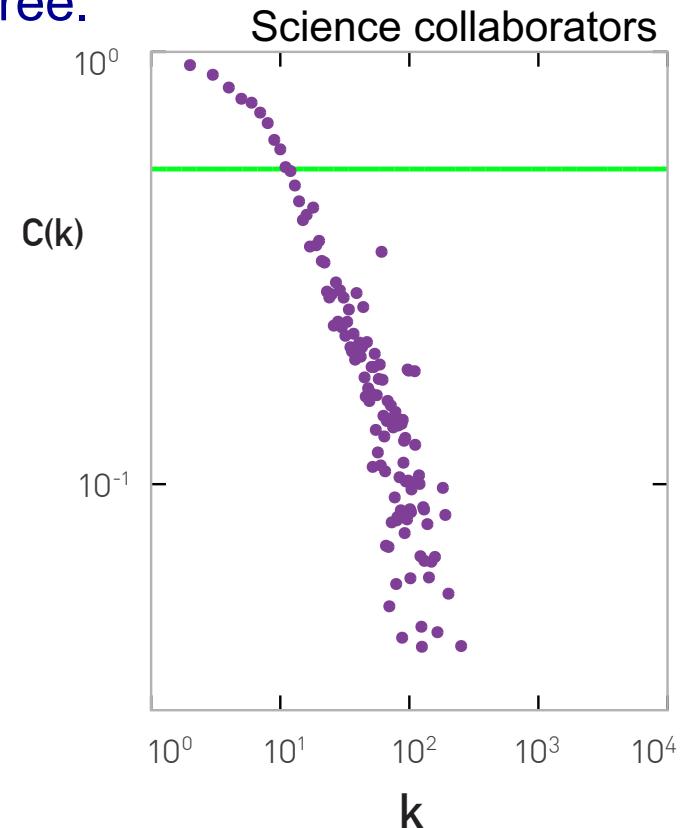
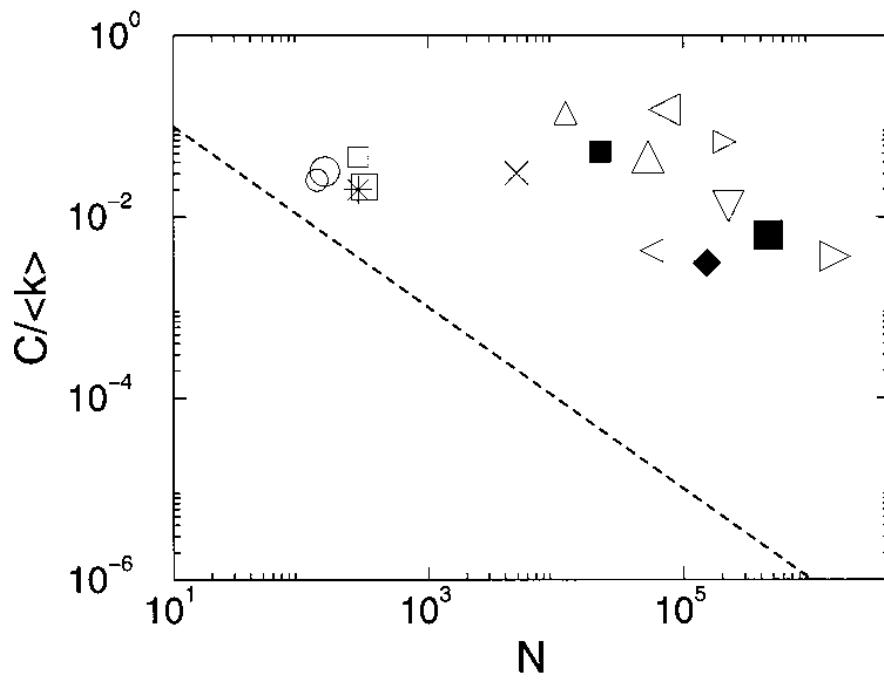
Mind your friend's friends

$$C_i \approx \frac{\langle k \rangle}{N}$$



Predictions:

1. For a fixed $\langle k \rangle$, larger the network, smaller the clustering coefficient.
2. Clustering coefficient is independent of the node's degree.



Clustering coeff. of a random graph

Real world networks often show large Clustering Coeff.

| | |
|-----------------------|----------|
| Film actors | → C=0.2 |
| Company directors | → C=0.59 |
| Math co-authorship | → C=0.15 |
| Physics co-authorship | → C=0.45 |
| Physics co-authorship | → C=0.45 |
| WWW | → C=0.11 |
| Neural networks | → C=0.18 |

$$C = \frac{\langle k \rangle}{N} \quad ???$$

Conclusion:

Are real-world networks random?

(Watts & Strogatz, Nature 1998)

- Degree distribution

Model fits empirical data?

Sometimes. Mostly NO!

- Maximum degree

Sometimes. Mostly NO!

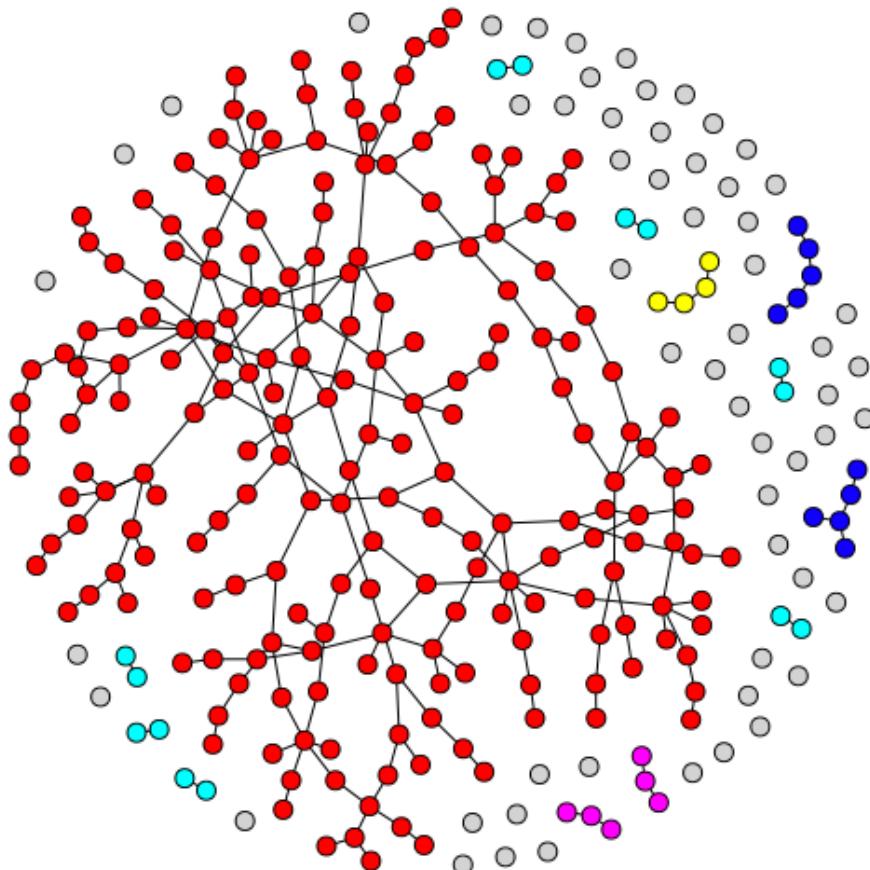
- Small world effect

YES!!

- Clustering coeff.

NO! CC is often high,
scaling with the degree
and not with N.

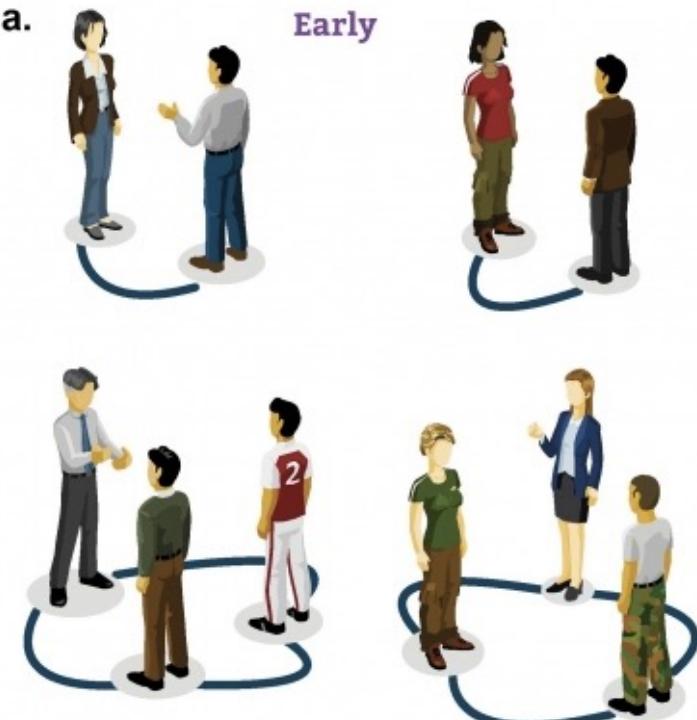
A last (& funny) message on random networks:
The bizarre world of the largest component



Last message on random networks:

The bizarre world of the largest component

Early on the guests form isolated groups.



As individuals mingle, changing groups, a connected network emerges



The bizarre world of the largest component

For $p=0$ all nodes
are isolated

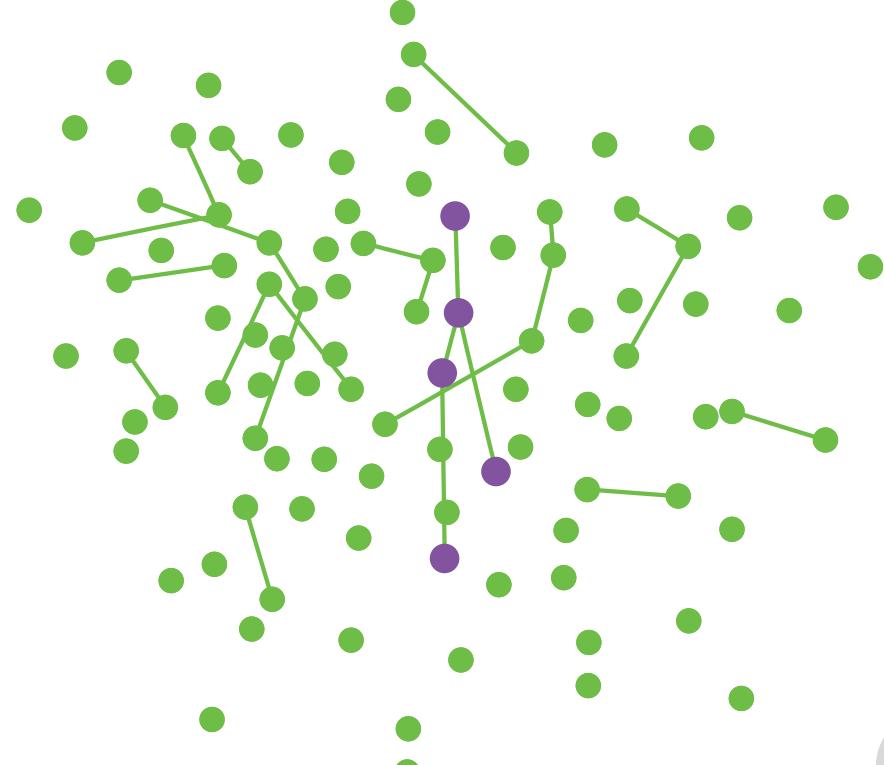
For $p=1$ we have a
fully-connected graph



*Average degree
increases linearly
with p*

*What happens with
the size of
the giant component?*

The bizarre world of giant components

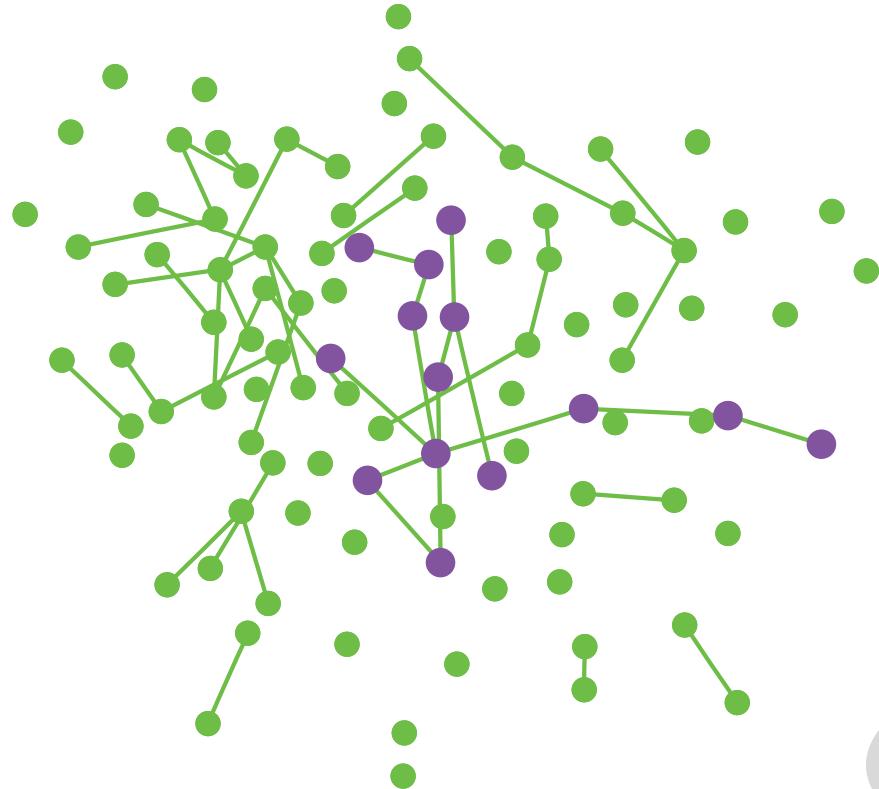


$p < 1/N$

*Subcritical
Regime*

$$\langle k \rangle < 1$$

The bizarre world of giant components

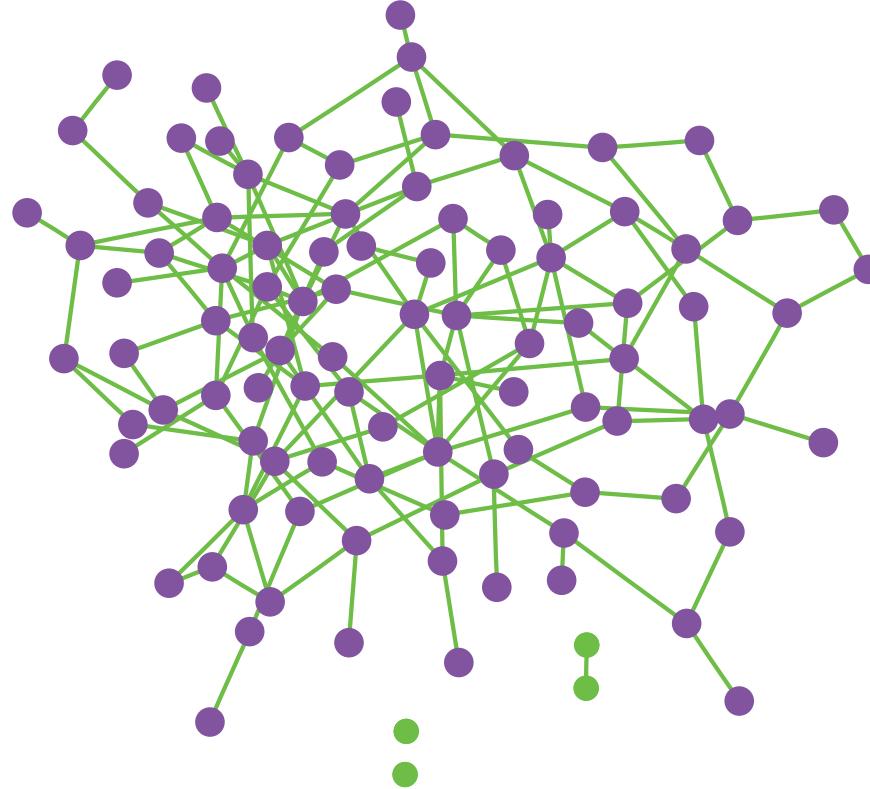


$p=1/N$

Critical Point

$$\langle k \rangle = 1$$

The bizarre world of giant components



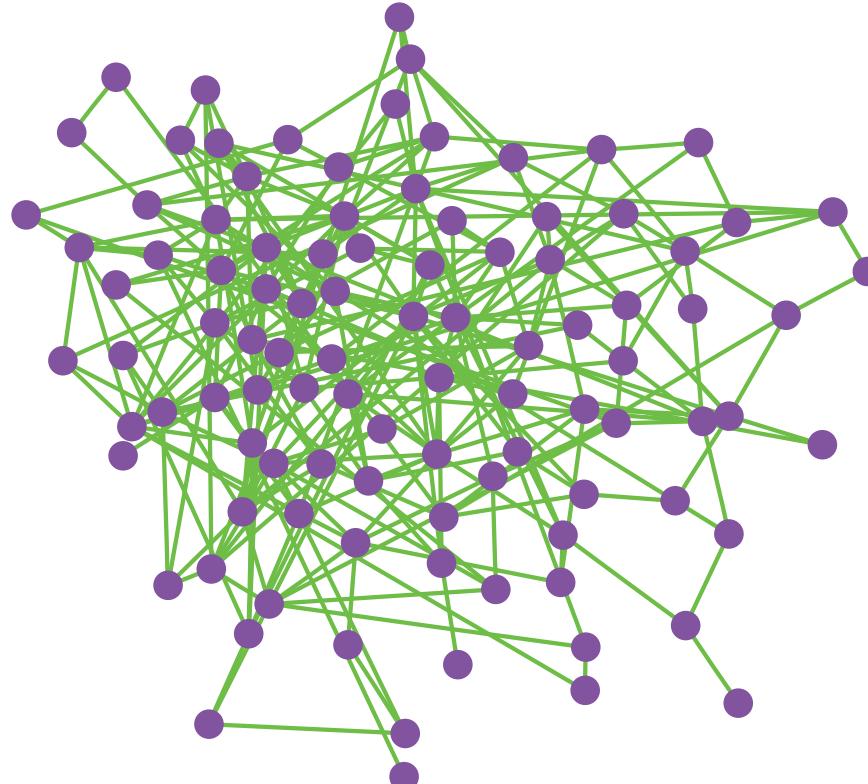
$p > 1/N$

*Supercritical
Regime*

$$\langle k \rangle > 1$$

giant component coexists with many disconnected components

The bizarre world of giant components



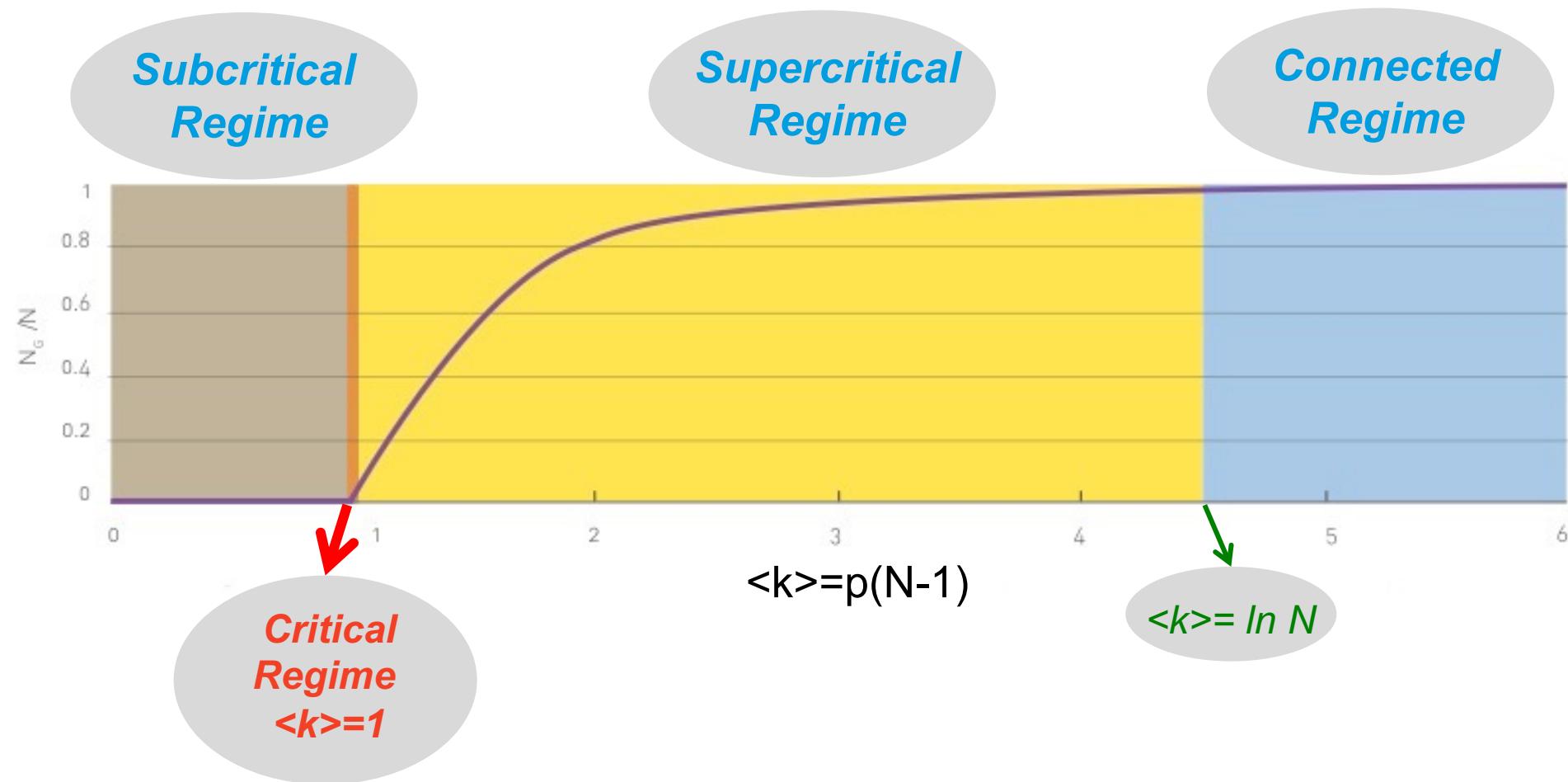
$$Np > \ln N$$

*Connected
Regime*

$$\langle k \rangle \gg \ln N$$

Connected network!

The bizarre world of giant components



Two predictions

Two predictions of random network theory are of direct importance for real networks:

- Once the average degree exceeds $\langle k \rangle = 1$, a giant component should emerge, with plenty of disconnected components. Hence only for $\langle k \rangle > 1$ the nodes organize themselves into a recognizable network.
- For $\langle k \rangle > \ln N$ all components are absorbed by the giant component, resulting in a single connected network.

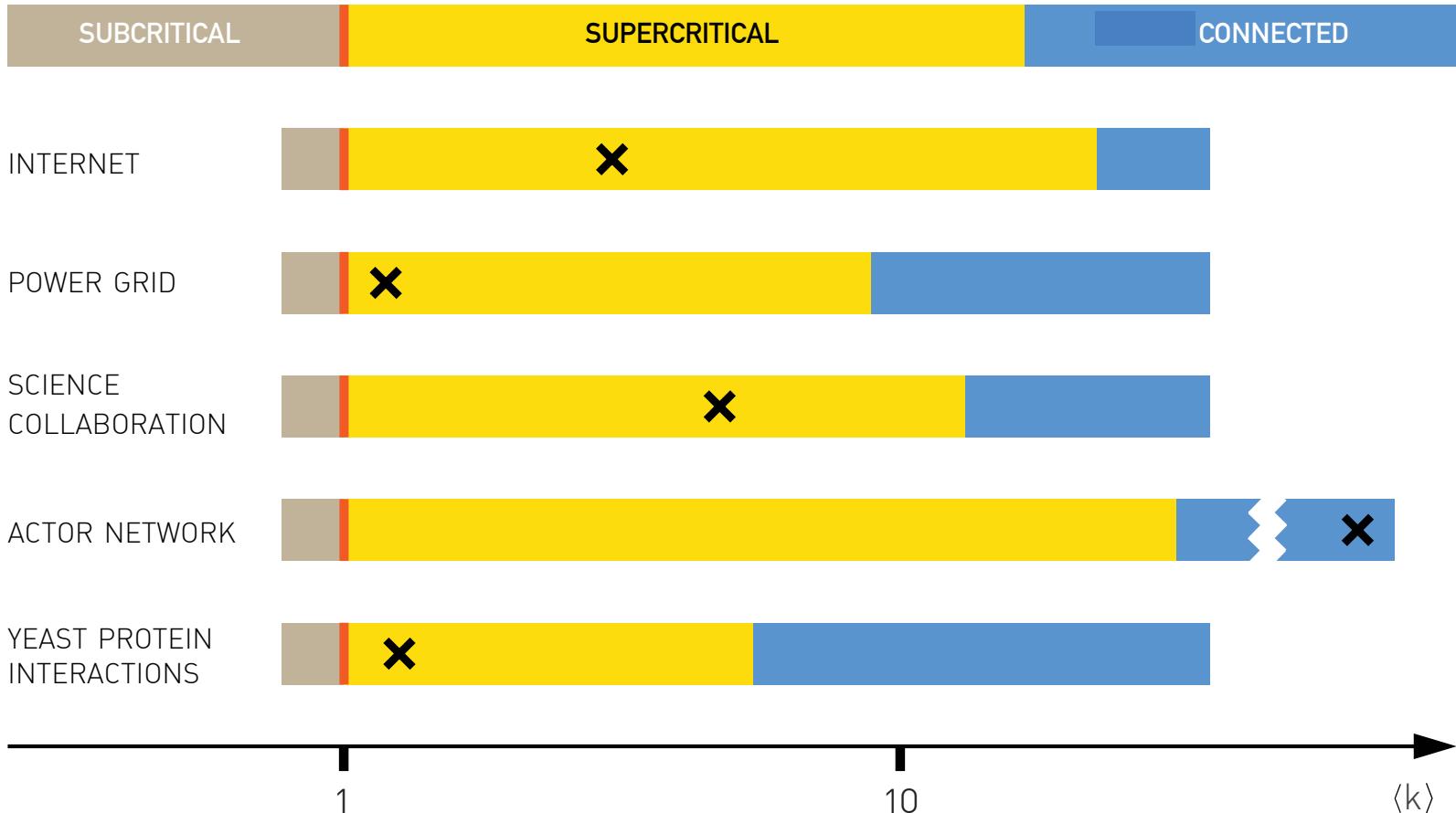
We have a problem!! Most networks show a single connected component. However, $\langle k \rangle$ is lower than $\ln N$

Where should we place real networks?

| Network | N | L | $\langle k \rangle$ | InN |
|-----------------------|---------|------------|---------------------|---------|
| 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 |

*Supercritical
Regime*

Where should we place real networks?



Where should we place real networks?

- Most real networks fall into the supercritical regime and, therefore, assuming that the world follows a ER graph, we should expect
 - A giant component (**correct!!**)
 - Plenty of disconnected components (**not necessarily!**)

In real-world networks, we can have connected networks for very low average degrees... We are still missing something!

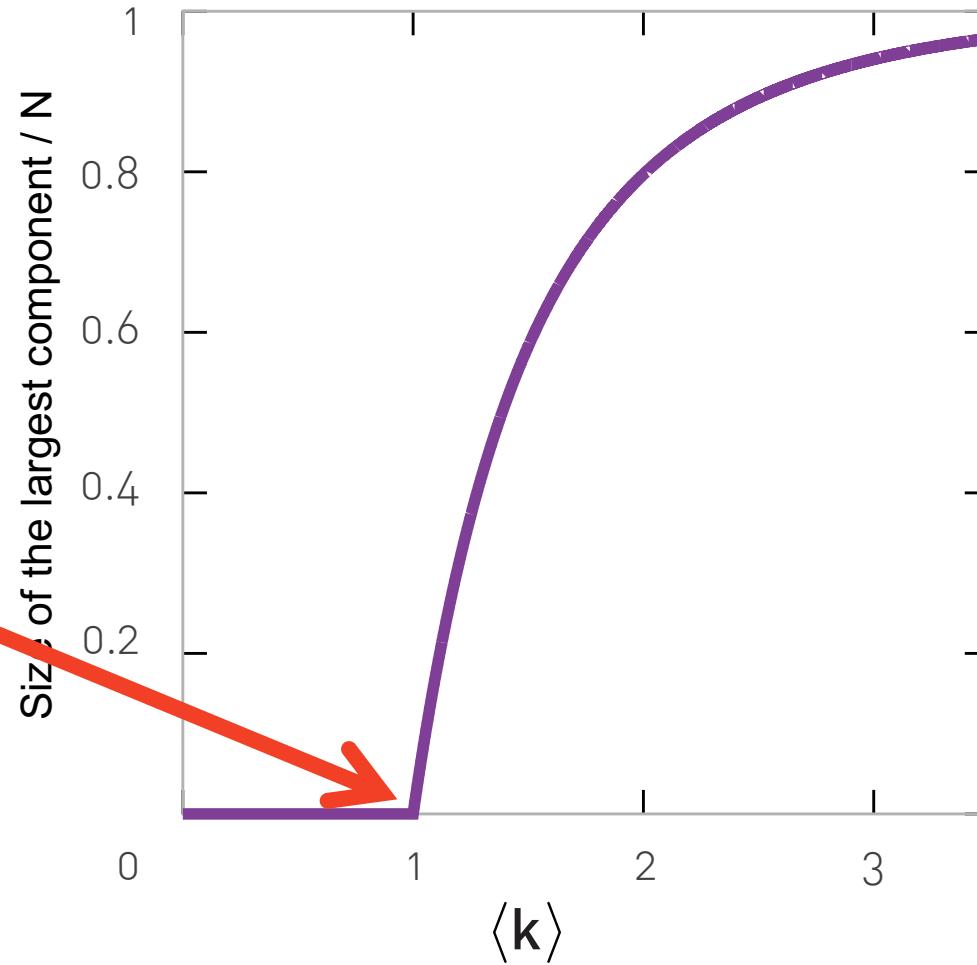
Today's conclusions on random graphs

- Degree distribution **Model fits empirical data?**
Sometimes. Mostly NO!
- Maximum degree **Sometimes. Mostly NO!**
- Clustering coeff. **NO! CC is often high,
scaling with the degree
and not with N.**
- Small world effect **YES!!**
- We need an additional mechanism to guarantee a single connected component for sparse graphs

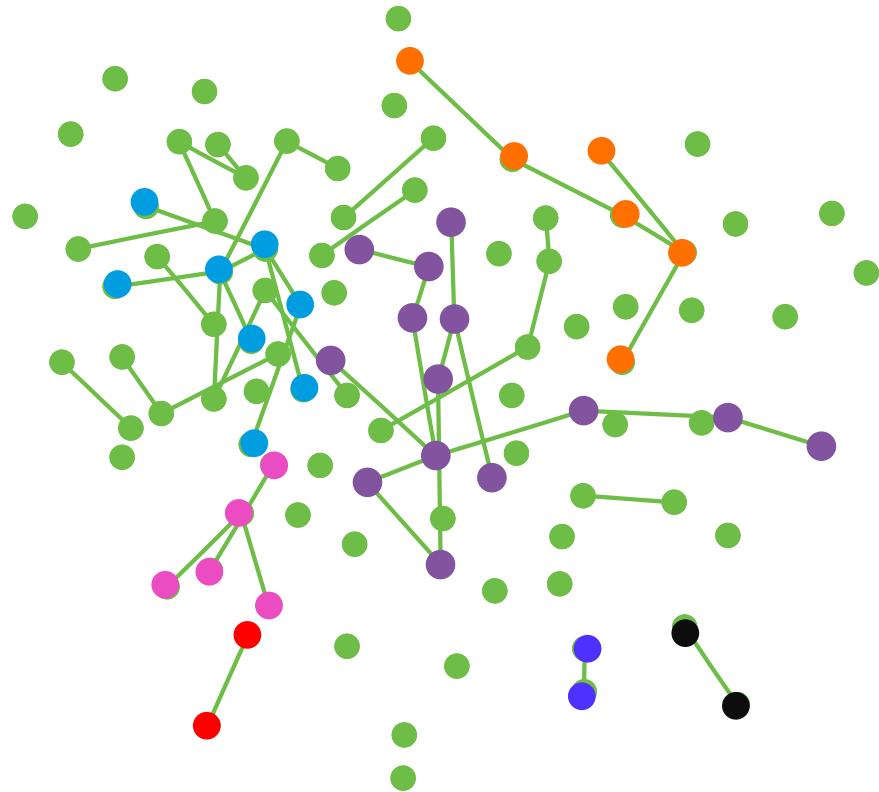
**Let me return
to such "critical" transition...**

Critical transitions?

What's happening here?

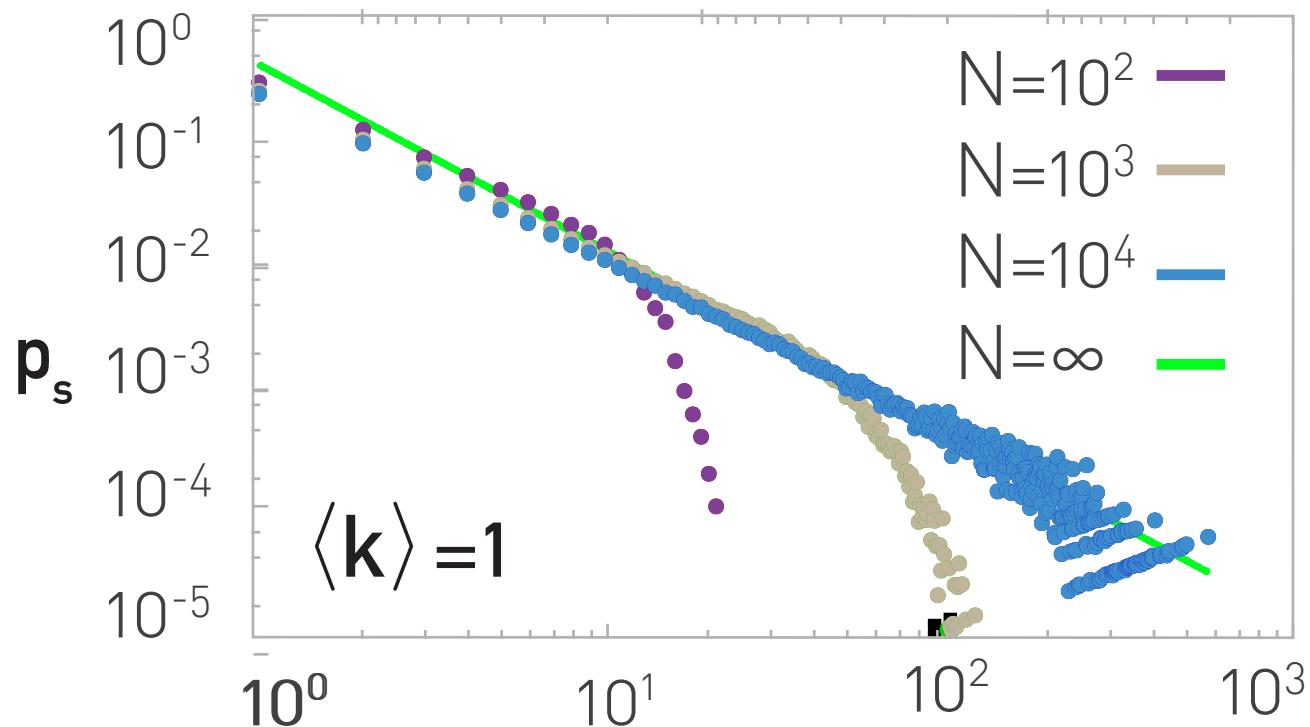


Critical transitions?



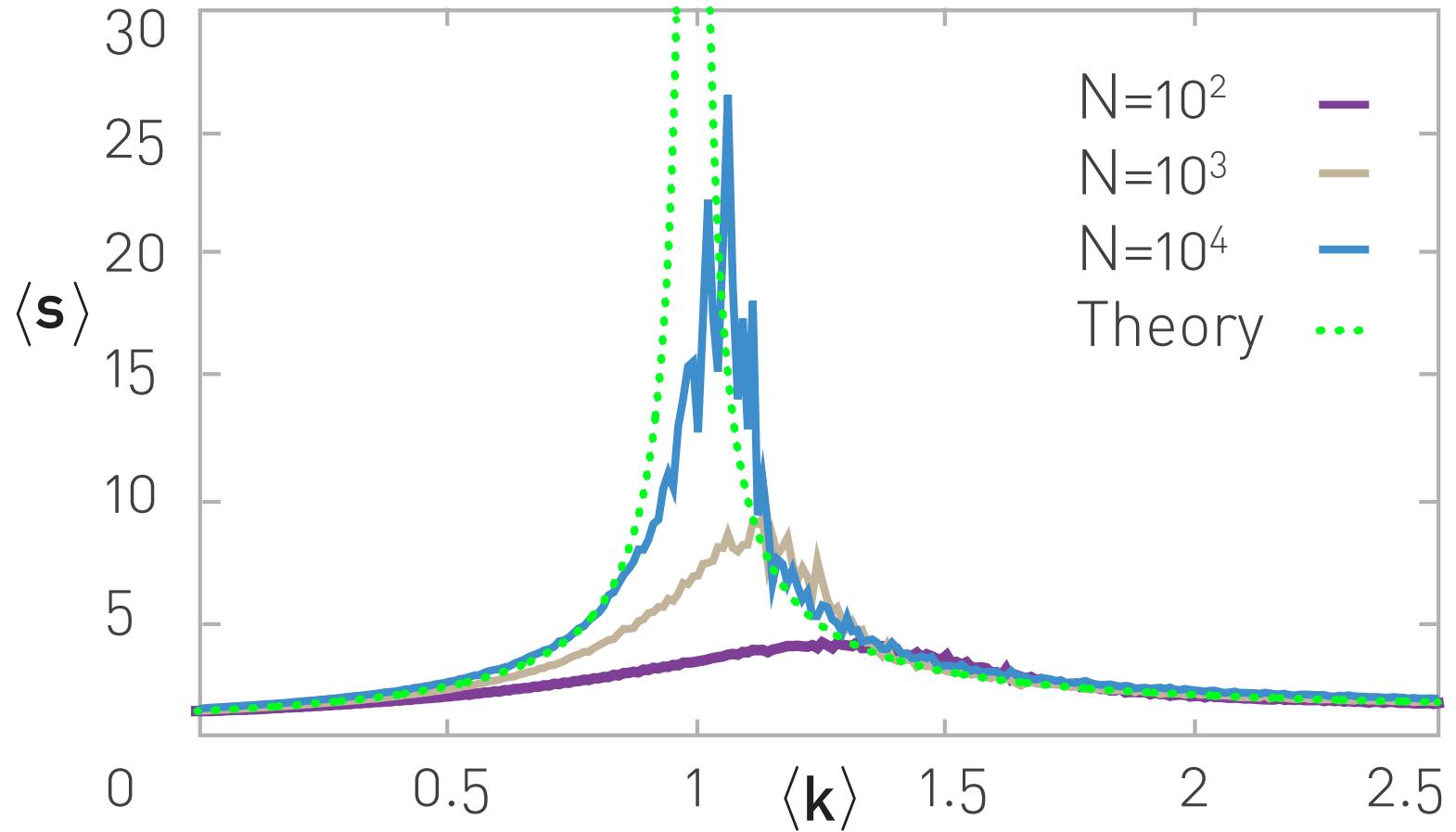
$$\langle k \rangle = 1$$

Component size dist. @ transition



$$p_s \sim s^{-3/2}$$

Average size of components @ transition

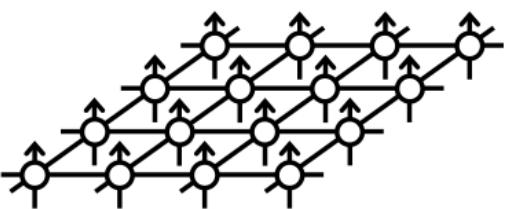


$$\langle s \rangle = \frac{1}{1 - \langle k \rangle + \langle k \rangle N_G / N}$$

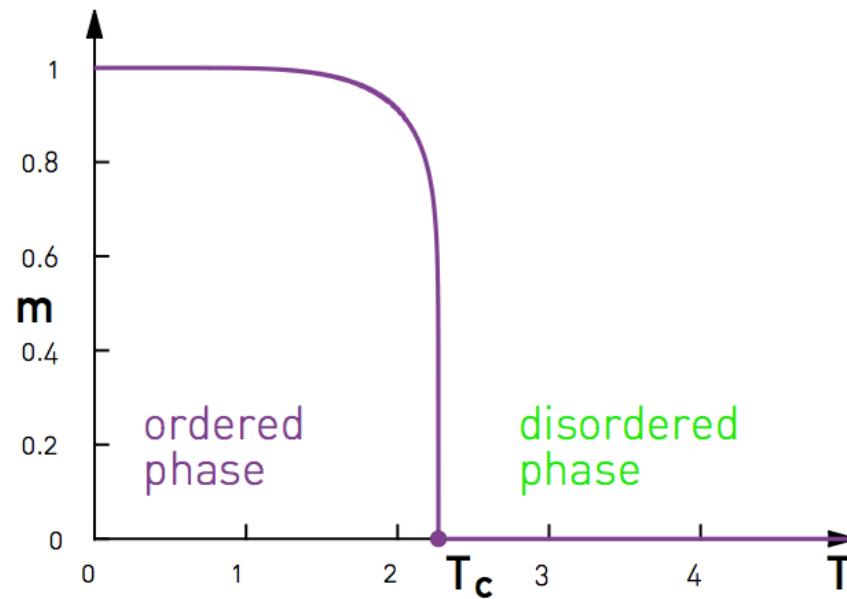
Phase/critical transitions everywhere

In ferromagnetic materials the magnetic moments of the individual atoms (spins) can point in two different directions.

At low temperatures all spins point in the same direction

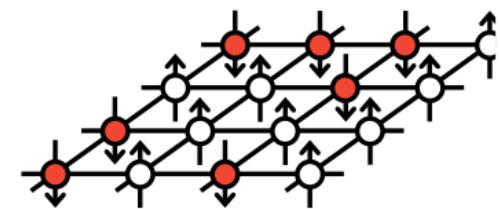


ordered phase



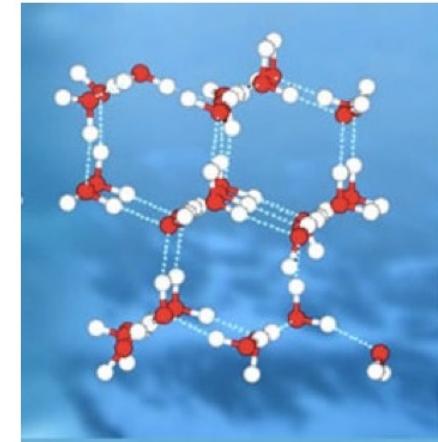
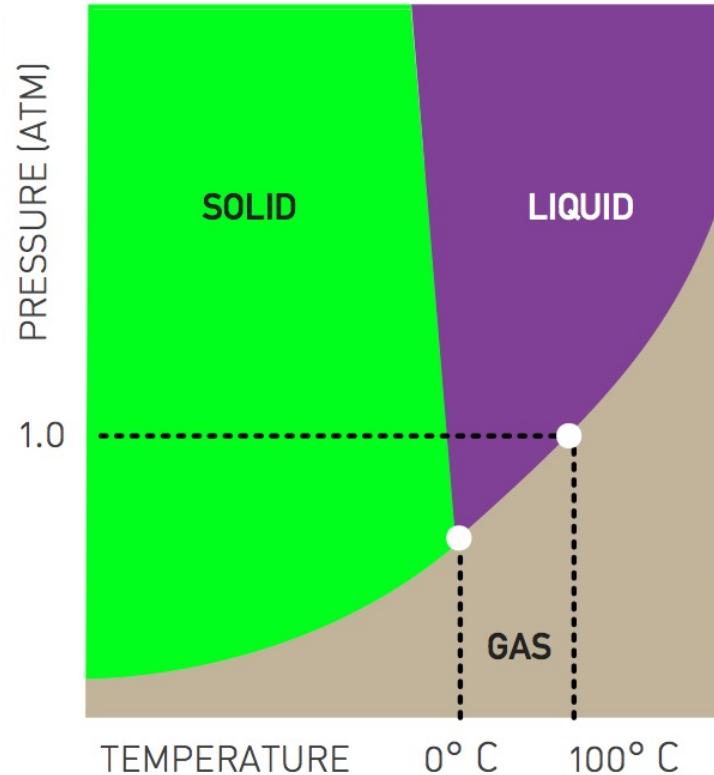
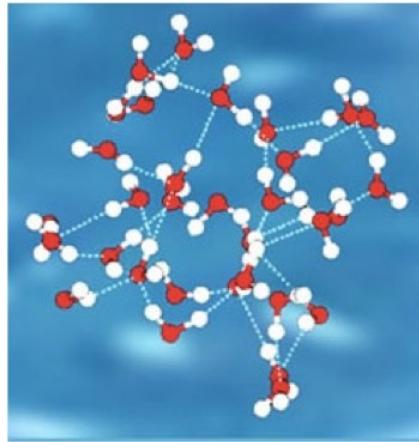
At a critical temperature T_c the system undergoes a phase transition just like we had in random graphs!

At high temperatures they choose randomly their direction. Here m (magnetization = sum of ups – downs) is zero.

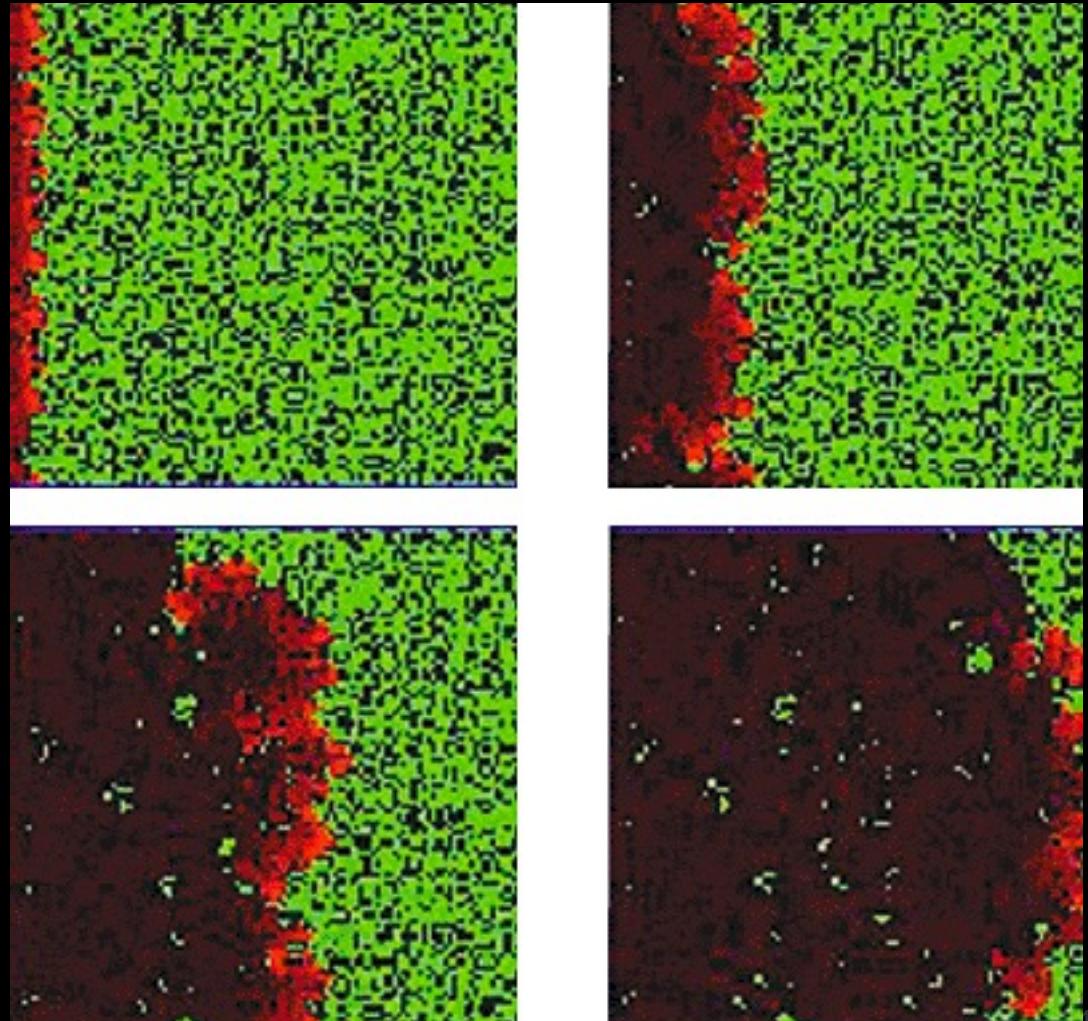


disordered phase

Phase/critical transitions everywhere



Phase/critical transitions everywhere

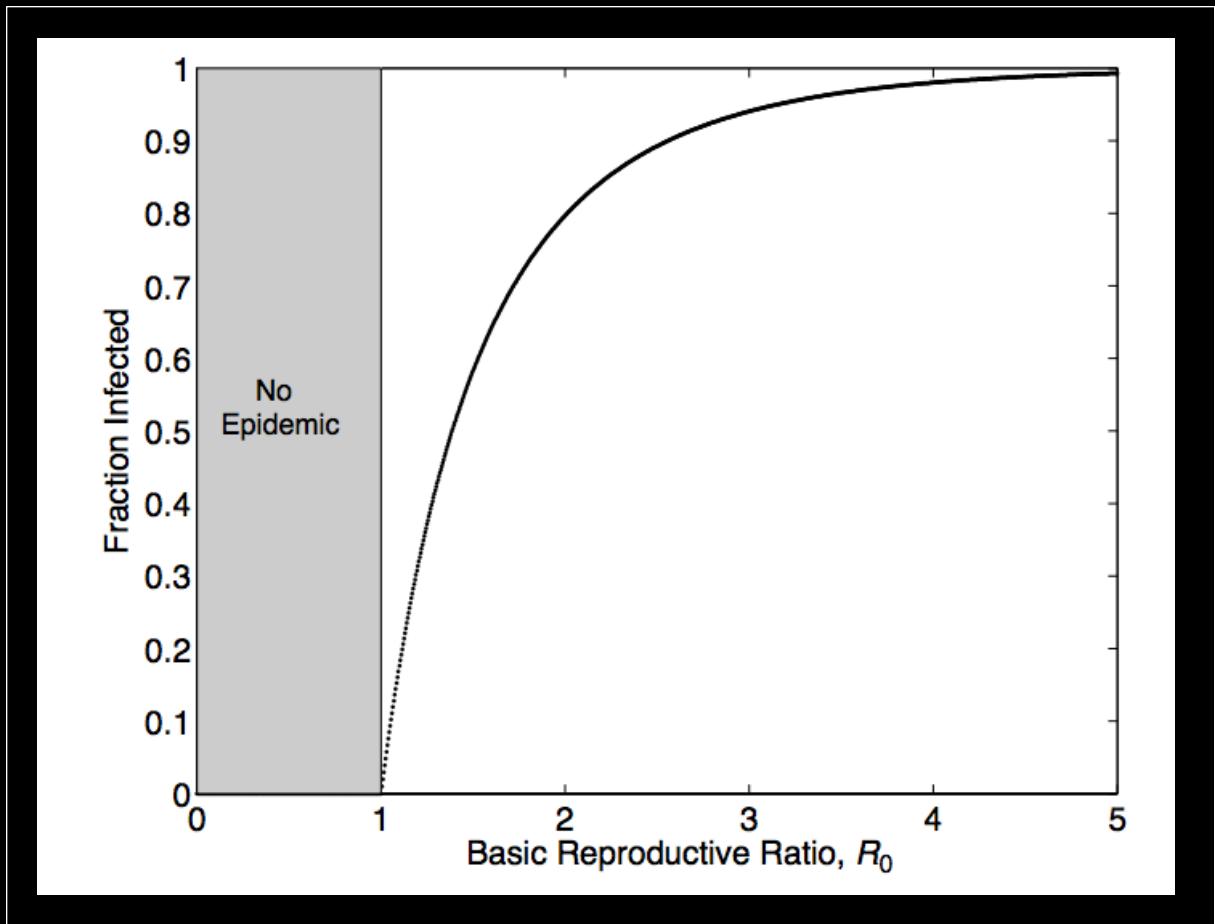


Phase/critical transitions everywhere

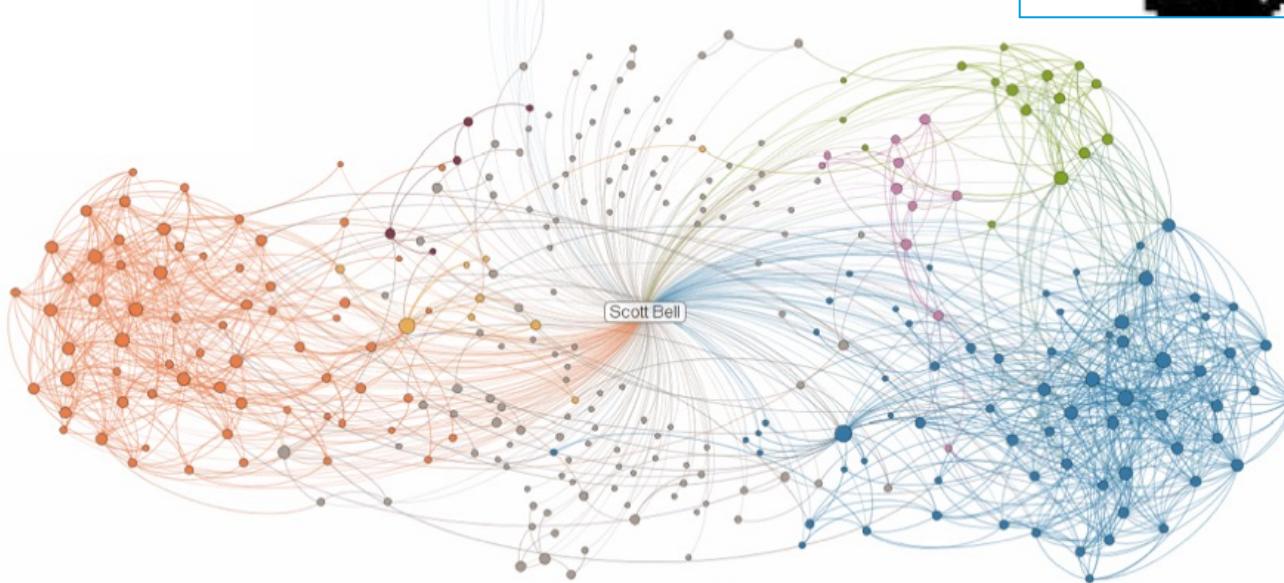
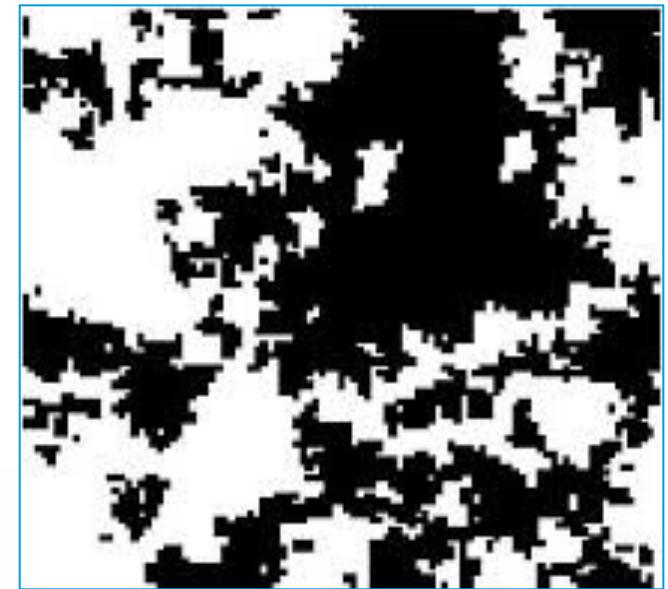


| Disease | Transmission | R_0 |
|----------------------------------|------------------|-------|
| Measles | Airborne | 12-18 |
| Pertussis | Airborne droplet | 12-17 |
| Diphtheria | Saliva | 6-7 |
| Smallpox | Social contact | 5-7 |
| Polio | Fecal-oral route | 5-7 |
| Rubella | Airborne droplet | 5-7 |
| Mumps | Airborne droplet | 4-7 |
| HIV/AIDS | Sexual contact | 2-5 |
| SARS | Airborne droplet | 2-5 |
| Influenza (1918 pandemic strain) | Airborne droplet | 2-3 |

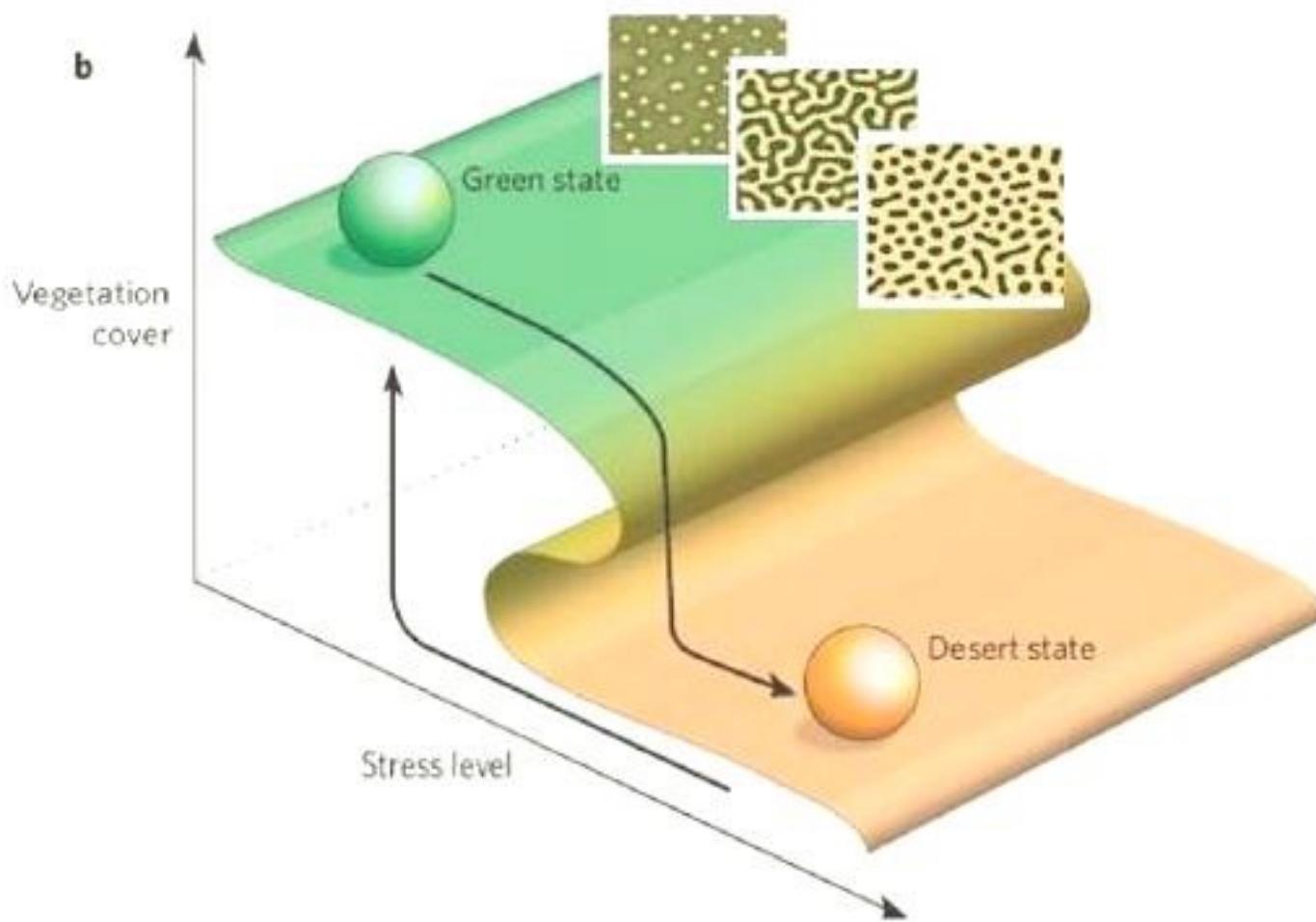
Phase/critical transitions everywhere



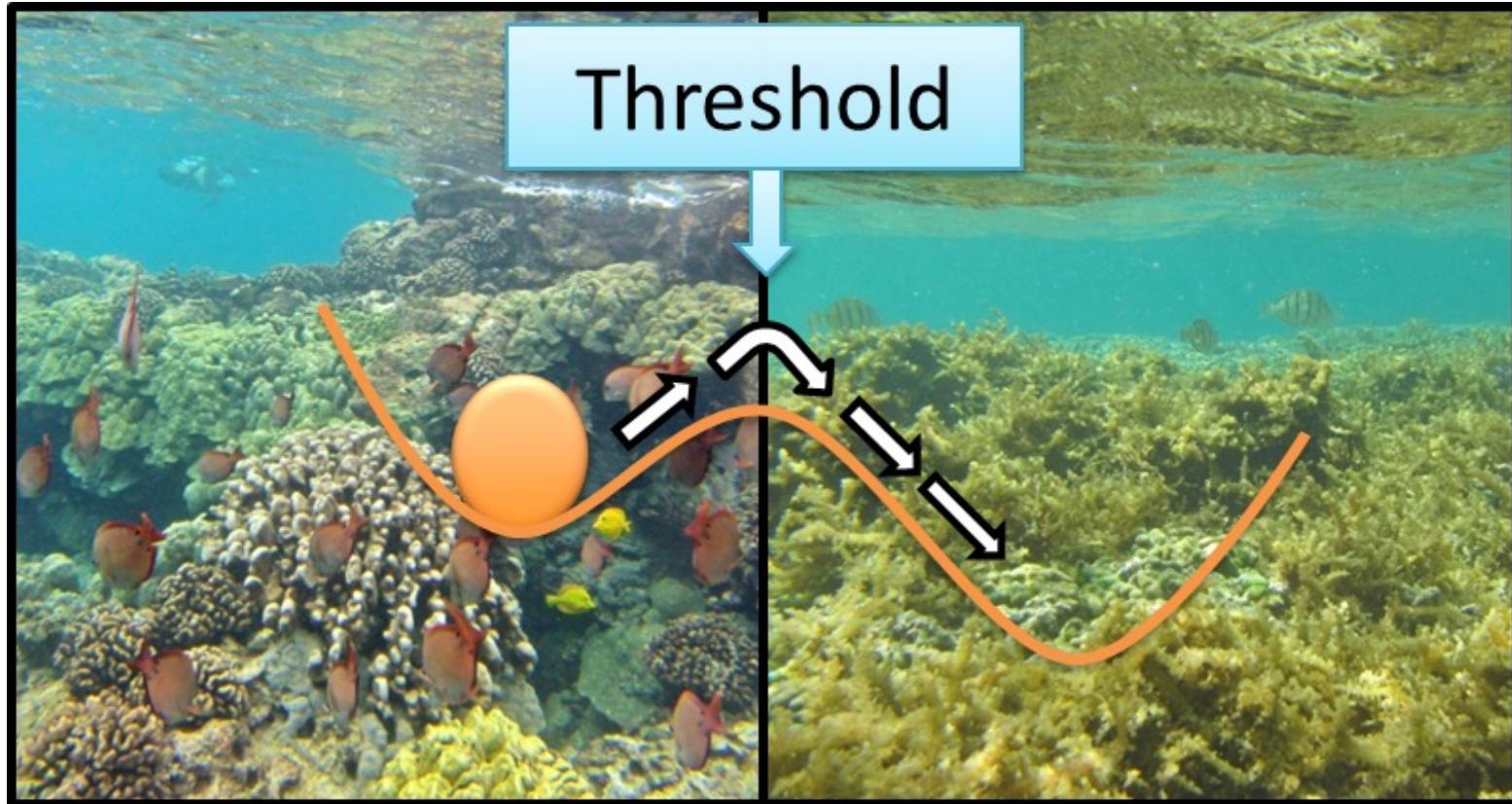
Phase/critical transitions everywhere



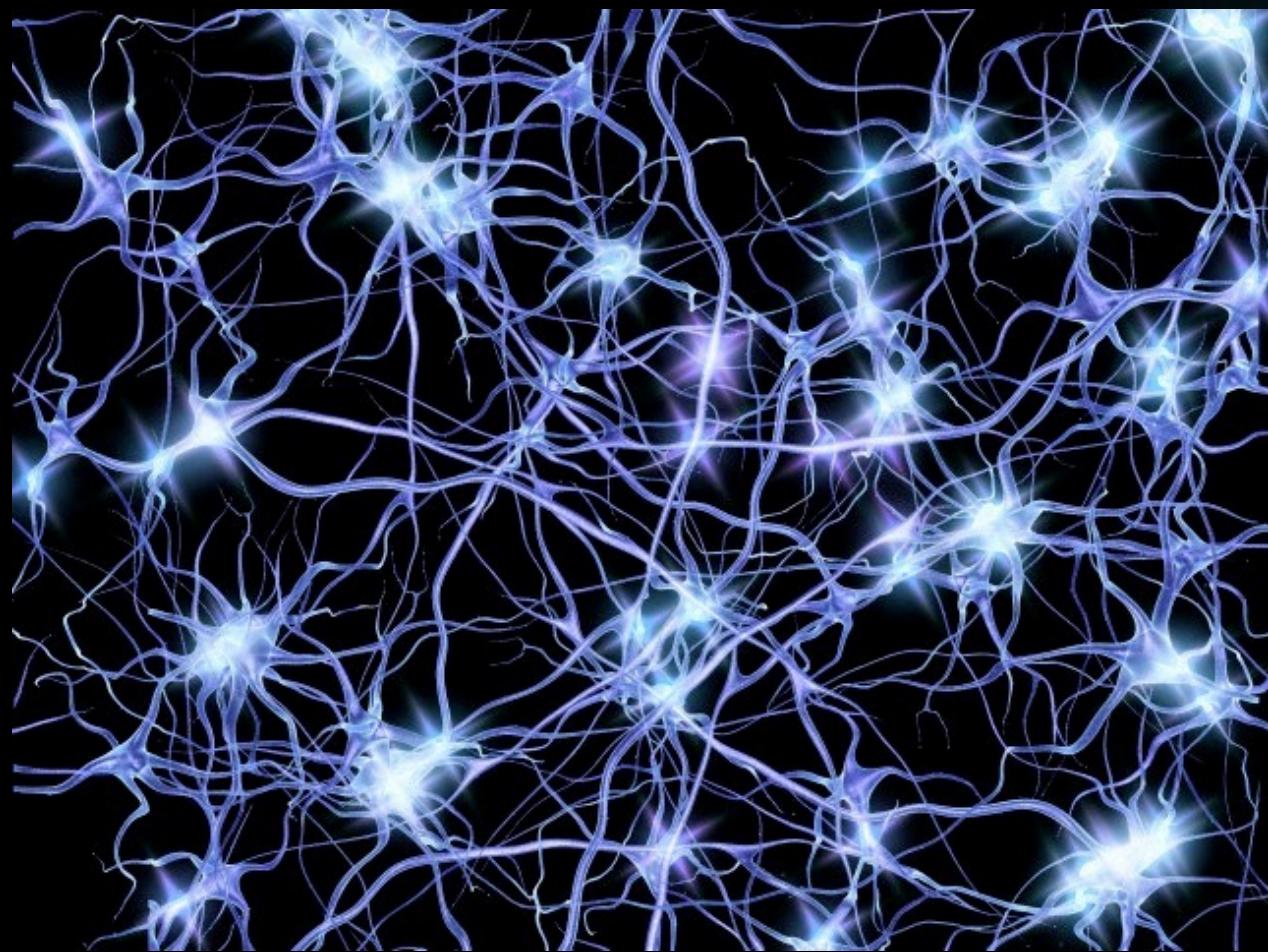
Regime shifts and tipping points



Regime shifts and tipping points



Phase transitions in the brain



Percolation in technological networks



The northeast
blackout, 2003

ISAT Geostar 45
23:15 EST 14 Aug. 2003



What's next: Networks' Revolutions

Watts & Strogatz, Nature 98

Network Science, 2023/2024

Conclusions so far:

Are real-world networks random?

- Degree distribution

Model fits empirical data?

Sometimes. Mostly NO!

- Maximum degree

Sometimes. Mostly NO!

- Small world effect

YES!!

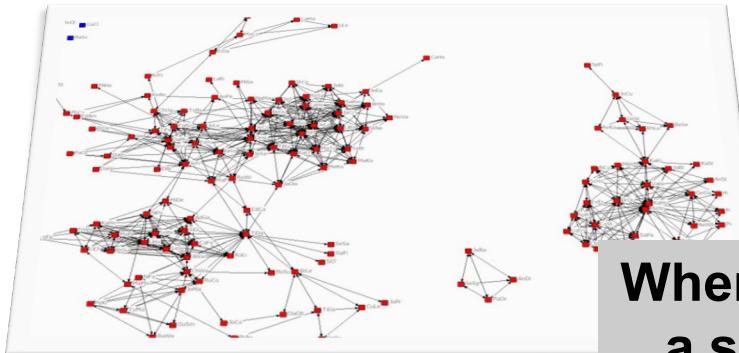
- Clustering coeff.

**NO! CC is often high,
scaling with the degree
and not with N.**

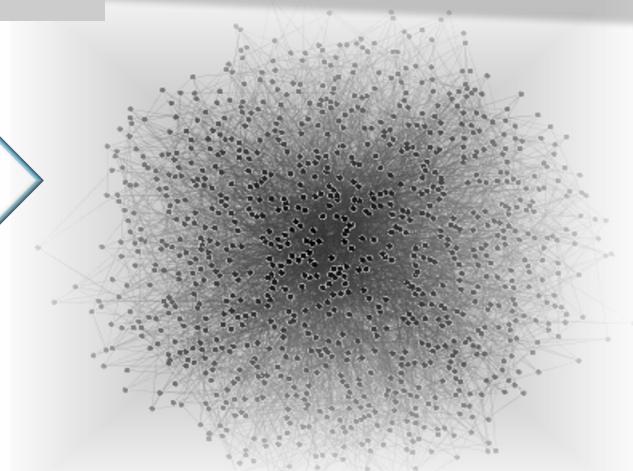
Clustering vs Randomness

Clustering implies locality

Randomness enables shortcuts



Where should we put
a social network?



Could a network which is so strongly locally structured be at the same time a small world?

Merging structure and randomness

Nature 393 (1998) 440

Collective dynamics of 'small-world' networks

Duncan J. Watts* & Steven H. Strogatz

Department of Theoretical and Applied Mechanics, Kimball Hall,
Cornell University, Ithaca, New York 14853, USA

Networks of coupled dynamical systems have been used to model biological oscillators^{1–4}, Josephson junction arrays^{5,6}, excitable media⁷, neural networks^{8–10}, spatial games¹¹, genetic control networks¹² and many other self-organizing systems. Ordinarily, the connection topology is assumed to be either completely regular or completely random. But many biological, technological and social networks lie somewhere between these two extremes.

Nature 393 (1998) 409

It's a small world

James J. Collins and Carson C. Chow

The concept of Six Degrees of Separation has been formalized in so-called 'small-world networks'. The principles involved could be of use in settings as diverse as improving networks of cellular phones and understanding the spread of infections.

news and views

length is short, scaling logarithmically with the size of the network.

What Watts and Strogatz⁵ do is to shift gradually from a regular network to a random network by increasing the probability of making random connections from 0 to 1 (see Fig. 1, page 441). They then measure the characteristic path length and the amount of clustering of the network as a function of the amount of randomness. They find that path length and clustering depend differently on the amount of randomness in the network. The characteristic path length drops quickly,

works that can be tuned to be rewired to introduce find that these systems have small lattices, yet have small m graphs. We call them with the small-world x degrees of separation¹⁵. *C. elegans*, the es, and the collaboration e small-world networks. ll-world coupling display imputational power, and bus diseases spread more easily in small-world networks than in regular lattices.

Merging structure and randomness

The Watts-Strogatz Small World model (Nature '98)

Watts & Strogatz invented a very simple model (**1 parameter!**) which interpolates between regular and random graphs.



Duncan J. Watts
Microsoft Research

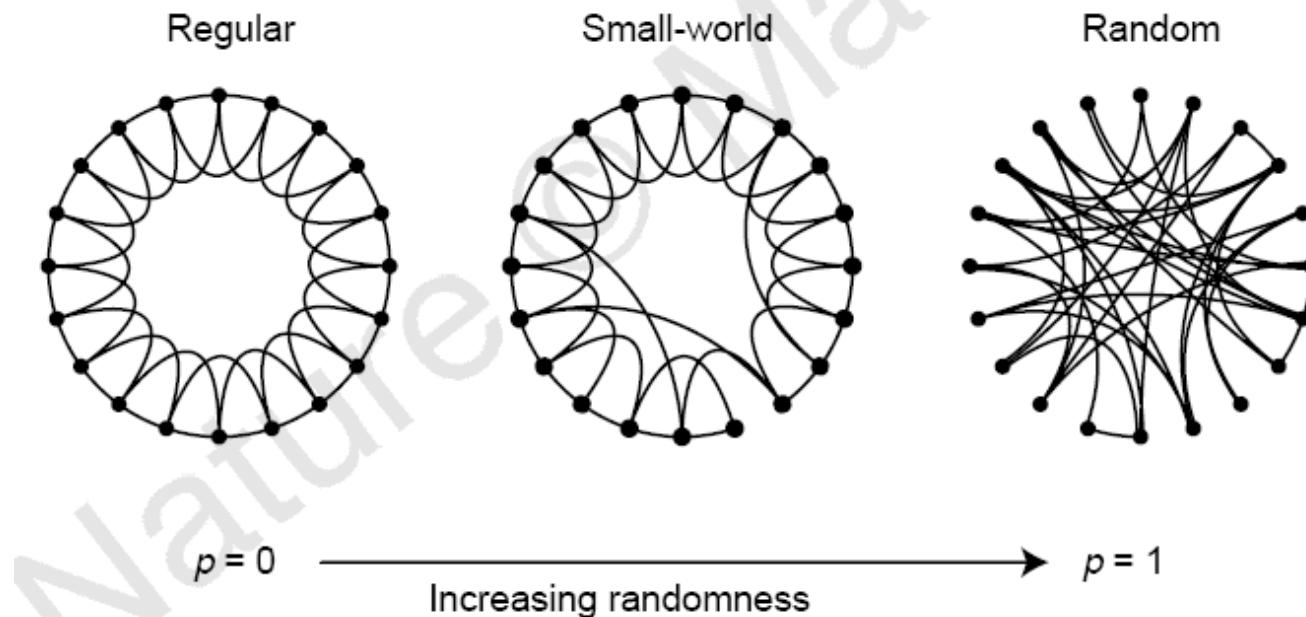


Steven H. Strogatz
Cornell Univ.

Merging structure and randomness

The Watts-Strogatz Small World model (Nature '98)

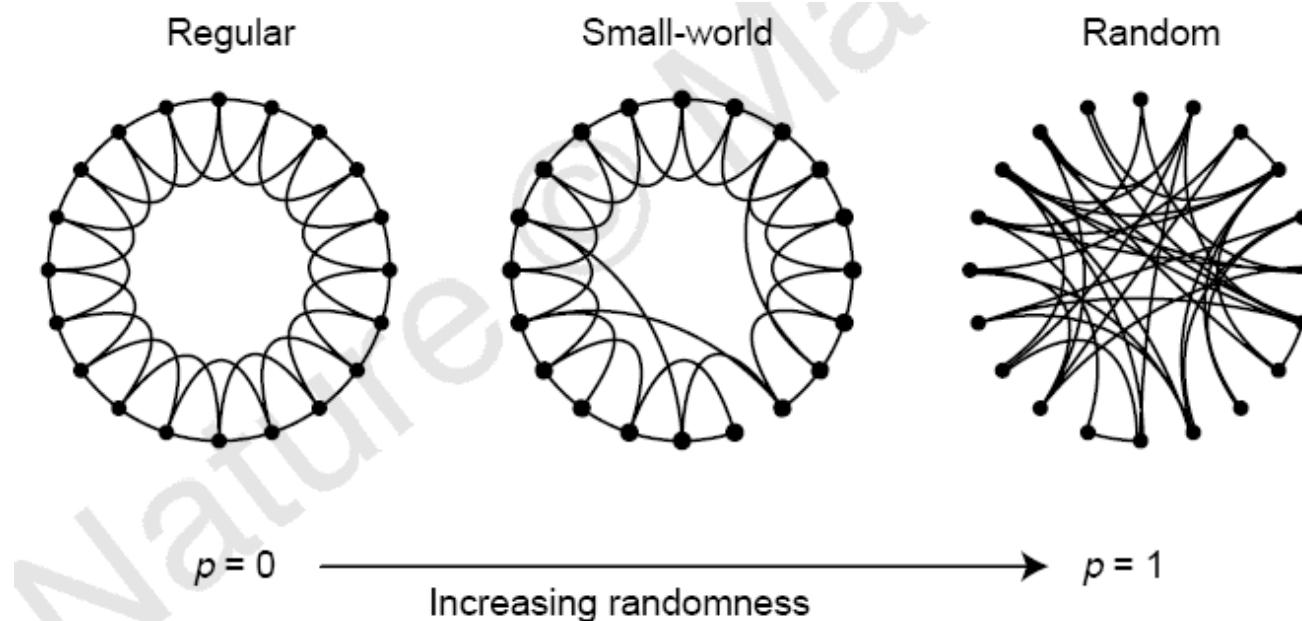
Watts & Strogatz invented a very simple model (**1 parameter!**) which interpolates between regular and random graphs.



Merging structure and randomness

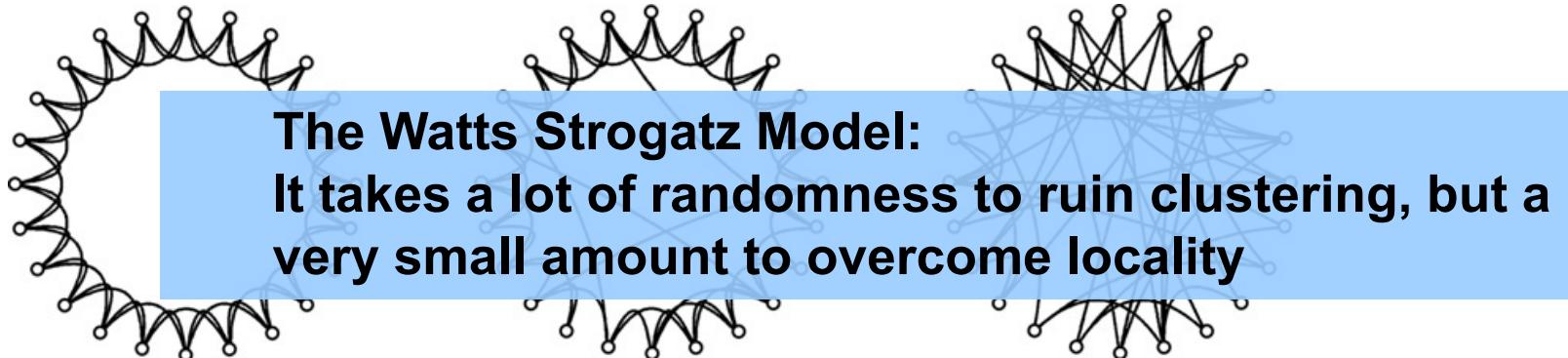
The Watts-Strogatz Small World model (Nature '98)

recipe : start from a regular graph (left); choose a circulating direction (say, clockwise); each edge one encounters is randomly re-directed with a probability p ; no repeated edges are allowed; stop when reaching the starting point;



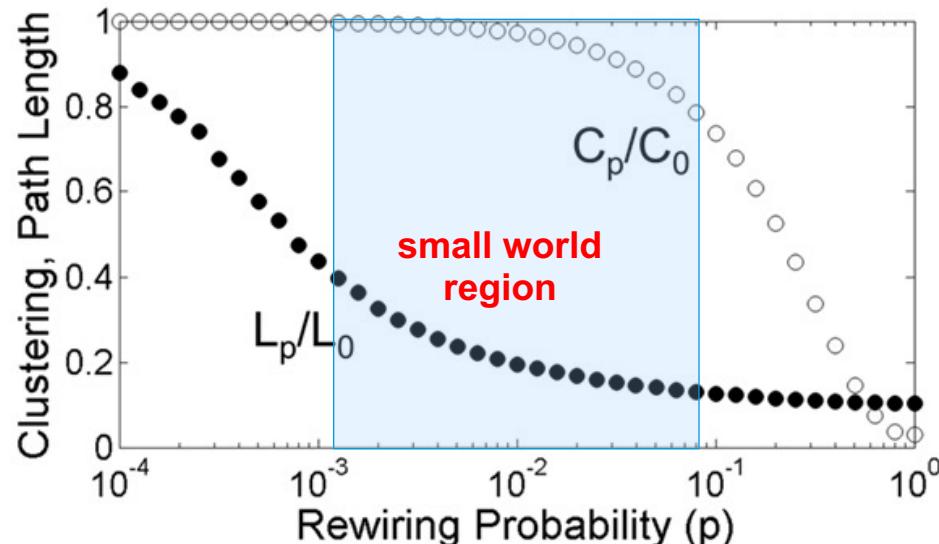
Merging structure and randomness

The Watts-Strogatz Small World model (Nature '98)



The Watts Strogatz Model:
It takes a lot of randomness to ruin clustering, but a
very small amount to overcome locality

$P = 0$ ————— increasing randomness ————— $P = 1$



1st Challenge:
Can you reproduce
this result?

Merging structure and randomness

NetworkX

Search docs

- Overview
- Download
- Installing
- Tutorial
- Reference
- Testing
- Developer Guide
- History
- Bibliography
- NetworkX Examples

watts_strogatz_graph

`watts_strogatz_graph(n, k, p, seed=None)` [\[source\]](#)

Return a Watts-Strogatz small-world graph.

Parameters:

`n : int`
The number of nodes

`k : int`

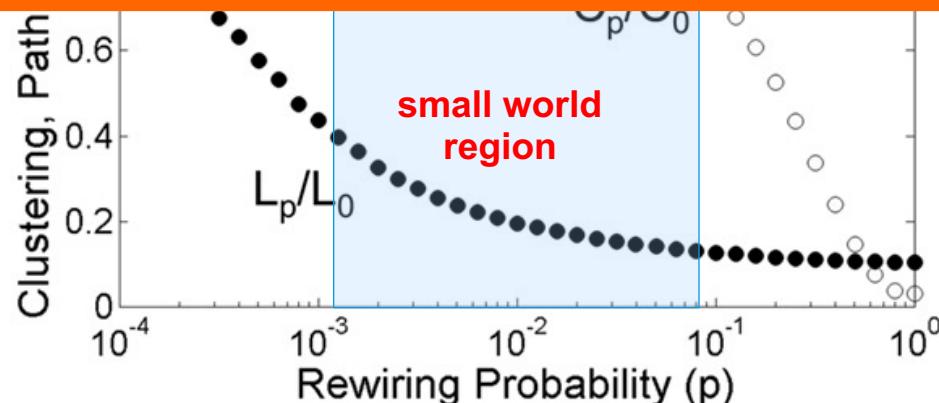
Each node is connected to `k` nearest neighbors in ring topology

`p : float`

The probability of rewiring each edge

`seed : int, optional`

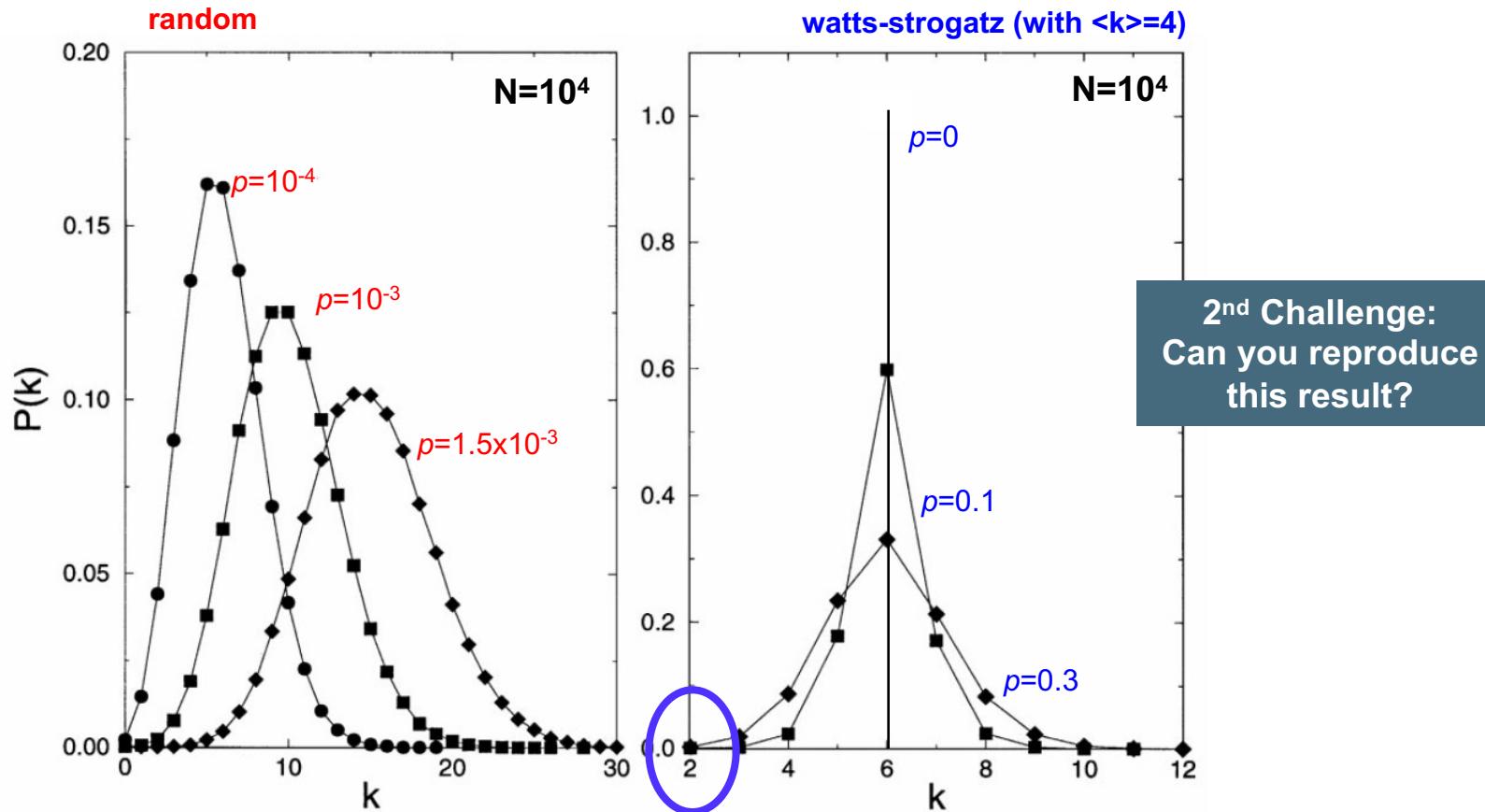
Seed for random number generator (default=None)



1st Challenge:
Can you reproduce
this result?

Merging structure and randomness

The Watts-Strogatz Small World model (Nature '98)



Homogeneous small worlds

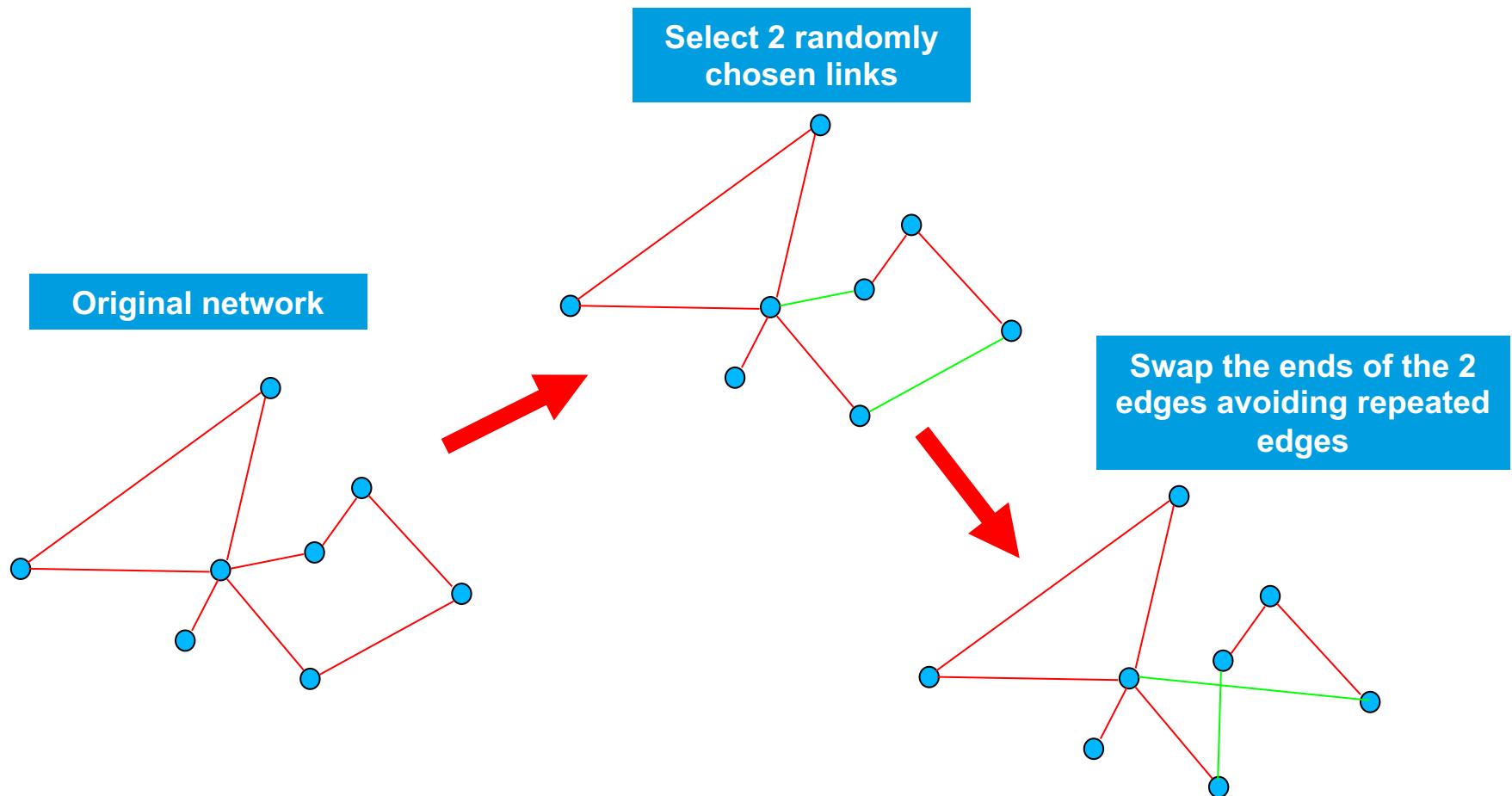
Challenge:

Imagine that you would like to test Small-world effects, yet keeping the degree distribution constant.

How can we adapt the WS model such we randomize a network without altering the degree of each node?

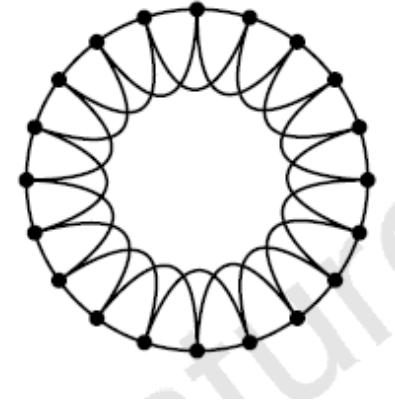


Degree-preserving randomization



Homogeneous small worlds

The following model generates small-world graphs with $P(k) = \delta(k - z)$



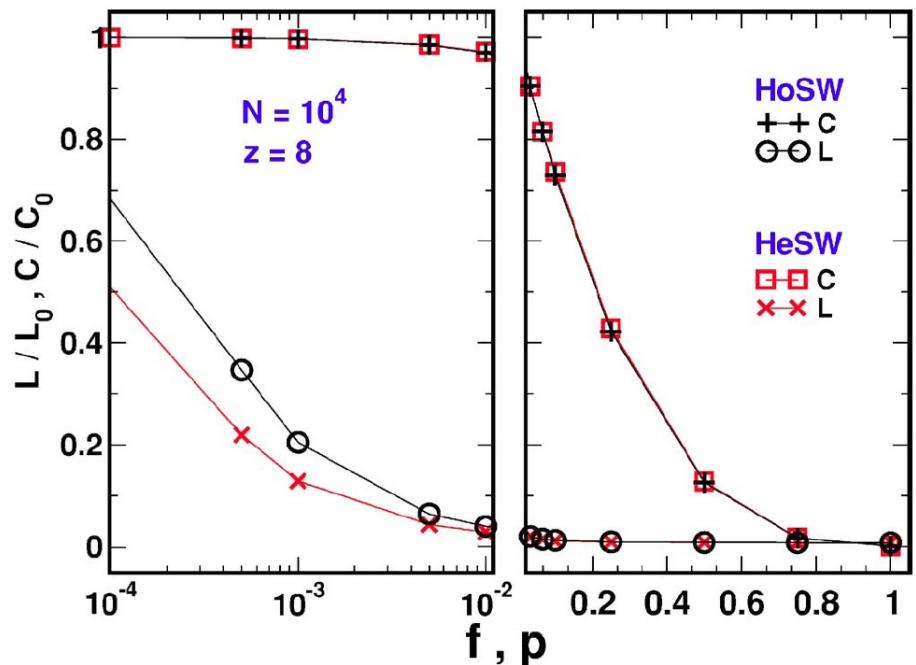
recipe : start with a regular graph;
repeat the following 2-step procedure until you have fE edges are rewired:

- i) chose 2 e's at random not used before in ii)
- ii) swap the ends of the two e's avoiding repeated e's;

similar to WS graphs, a 1-parameter model - f : we generate a $\langle C(f) \rangle$ & a $\langle L(f) \rangle$; what's their behavior ?



we can introduce small-world effects without introducing heterogeneity in the graph



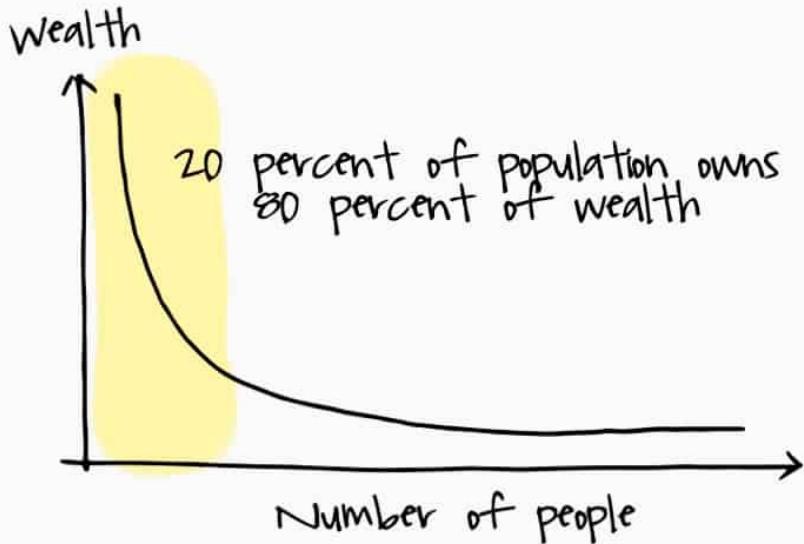


2nd network revolution

Barabási, Albert, Jeong, Dorogovtsev, Mendes, Havlin, Cohen...
(>1999)

Scale-free networks

- Watts and Strogratz explained a key point in real-world networks... but missed a key one.



Emergence of Scaling in Random Networks
Albert-László Barabási, et al.
Science **286**, 509 (1999);
DOI: 10.1126/science.286.5439.509

Emergence of Scaling in Random Networks

Albert-László Barabási* and Réka Albert

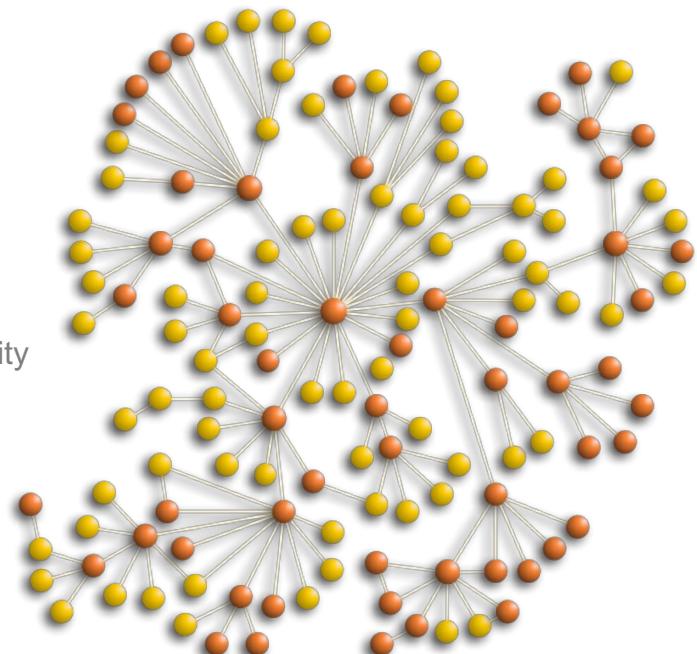
Systems as diverse as genetic networks or the World Wide Web are best described as networks with complex topology. A common property of many large networks is that the vertex connectivities follow a scale-free power-law distribution. This feature was found to be a consequence of two generic mechanisms: (i) networks expand continuously by the addition of new vertices, and (ii) new vertices attach preferentially to sites that are already well connected. A model based on these two ingredients reproduces the observed stationary scale-free distributions, which indicates that the development of large networks is governed by robust self-organizing phenomena that go beyond the particulars of the individual systems.



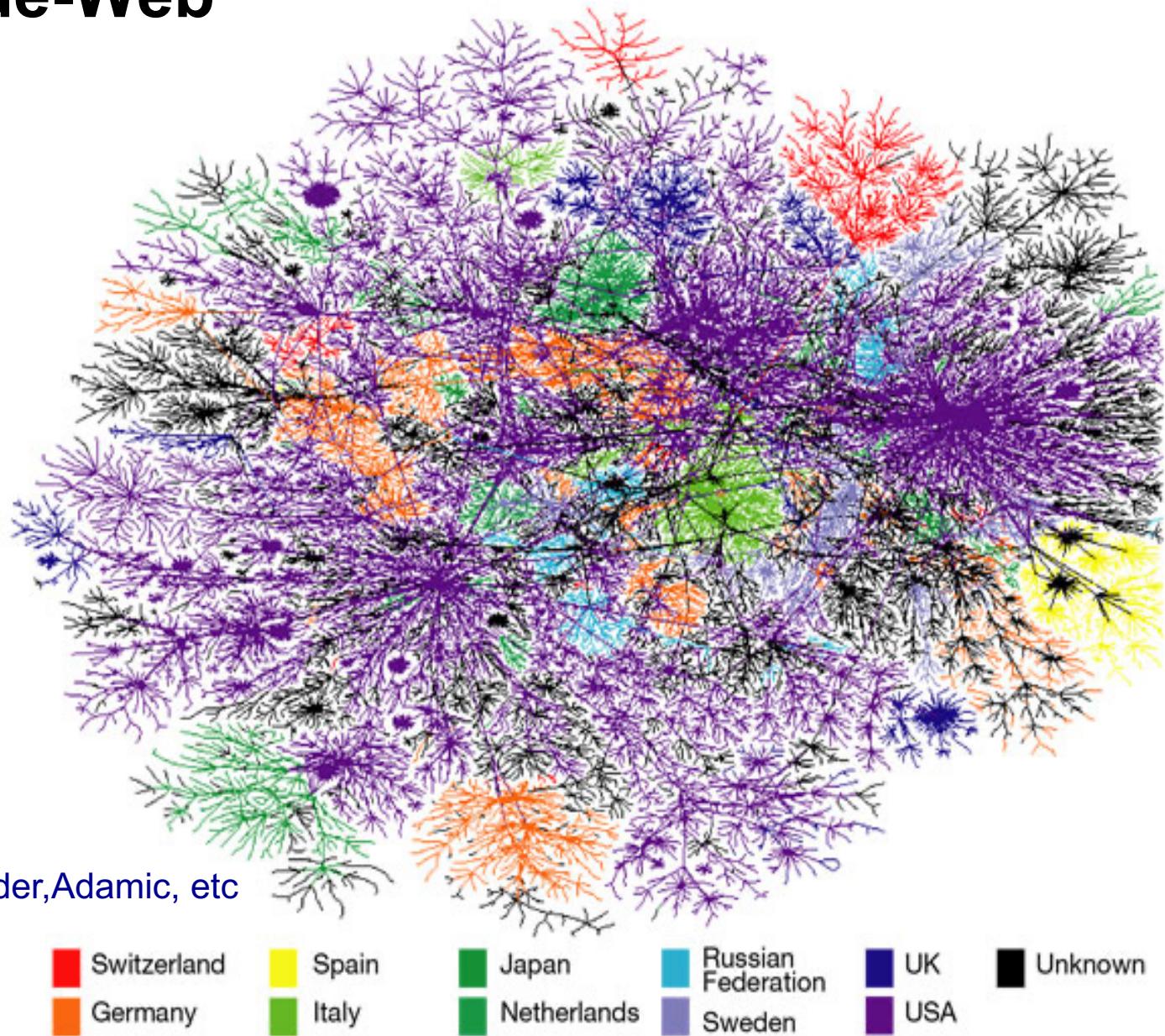
Reka Albert
Penn State University



Albert-László Barabási
Northeastern University

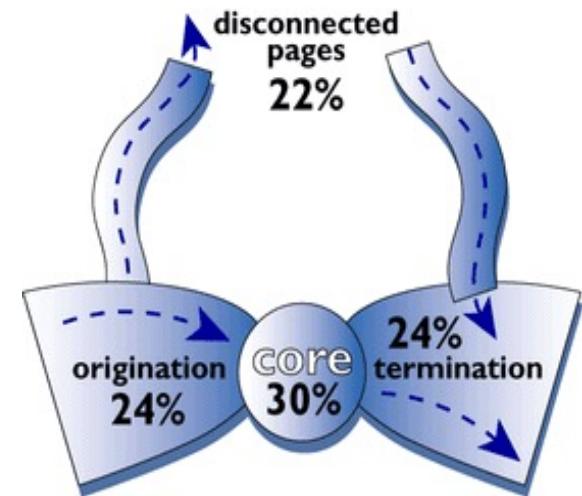
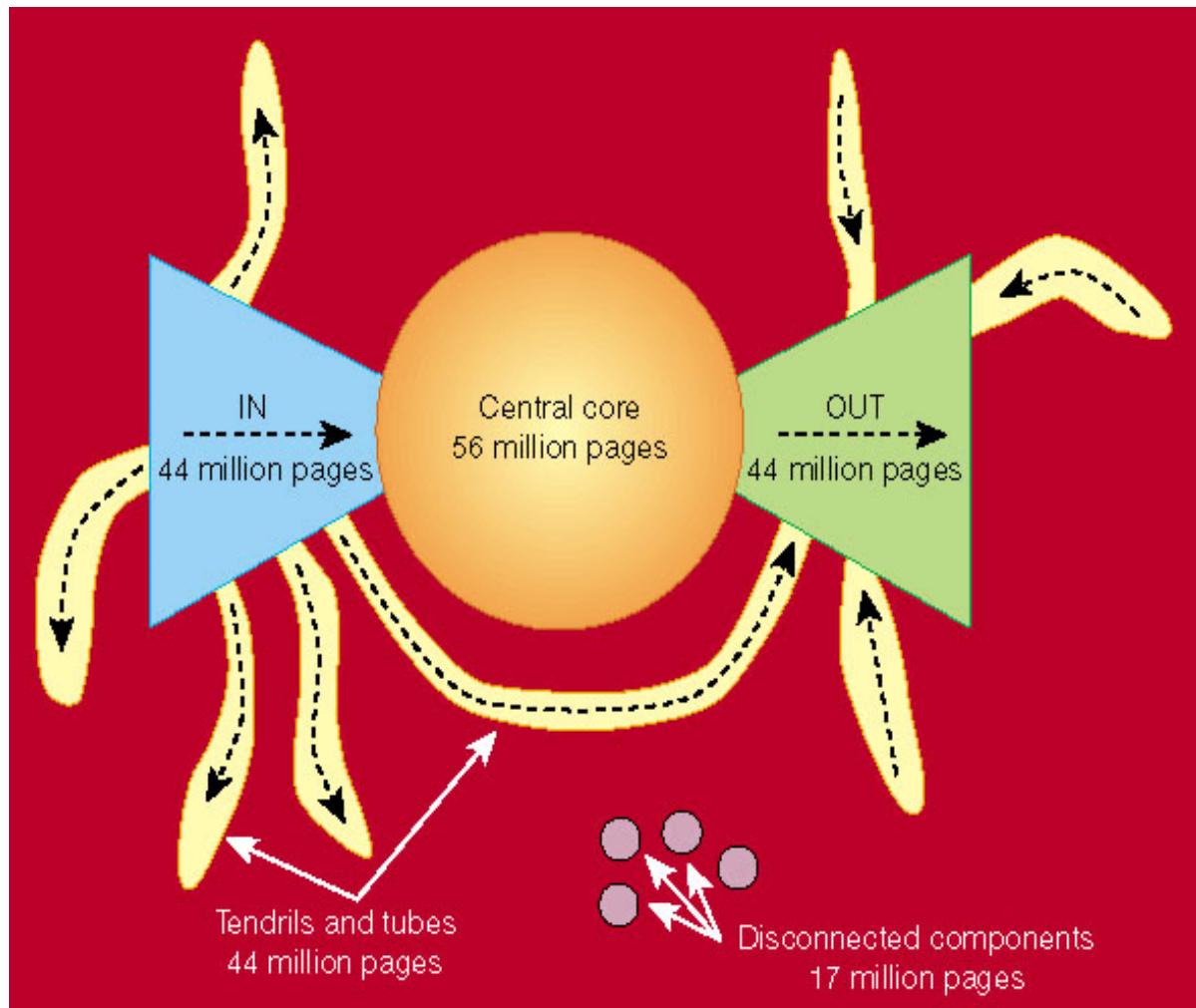


World-Wide-Web



World-Wide-Web is a “bow-tie”

200 million pages / 1500 million hyperlinks



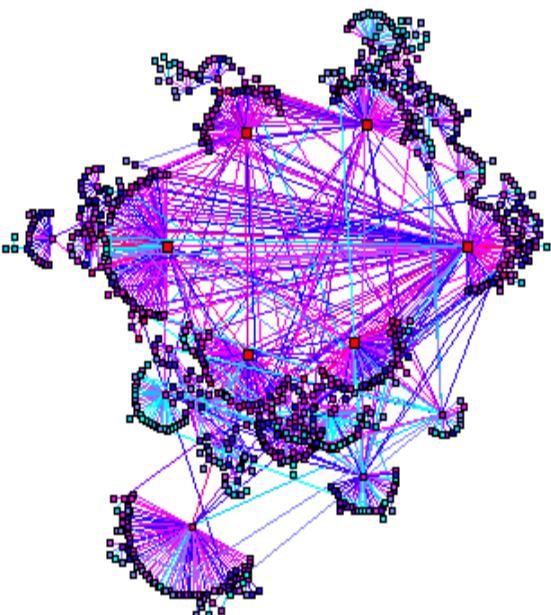
World-Wide-Web

Nodes: **WWW documents**

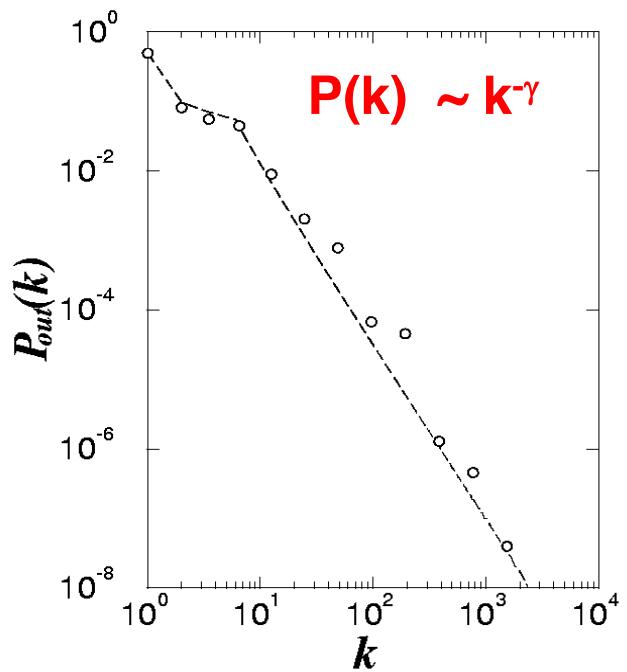
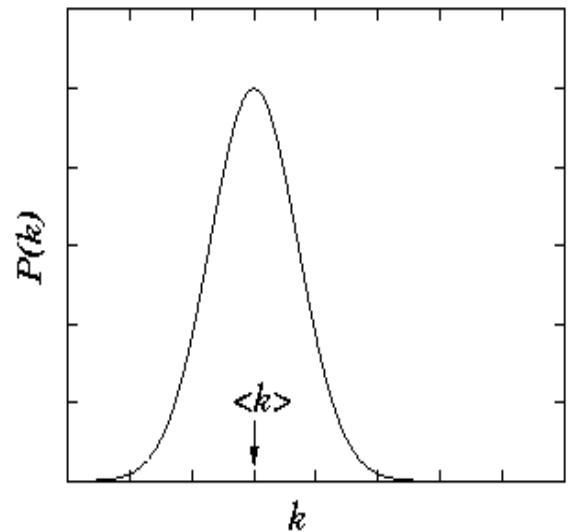
Links: **URL links**

Over 3 billion documents

ROBOT: collects all URL's found in a document and follows them recursively

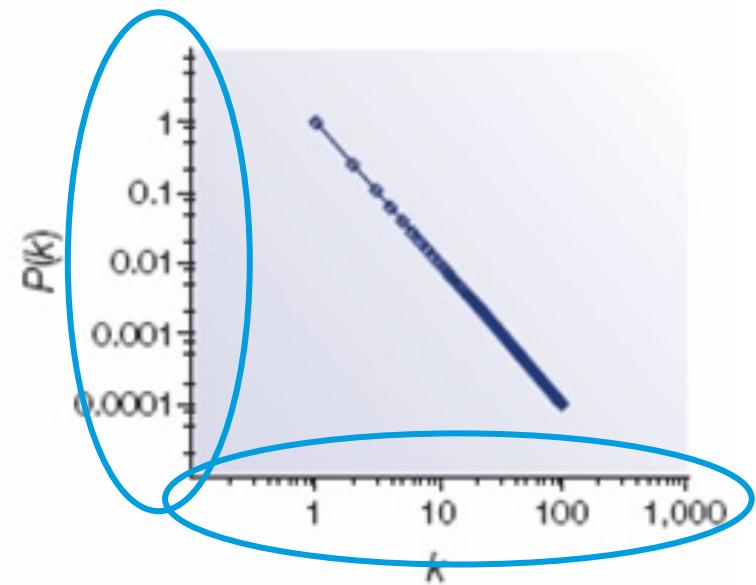


R. Albert, H. Jeong, A-L Barabasi, *Nature*, 401 130 (1999).



Power-law degree distributions

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



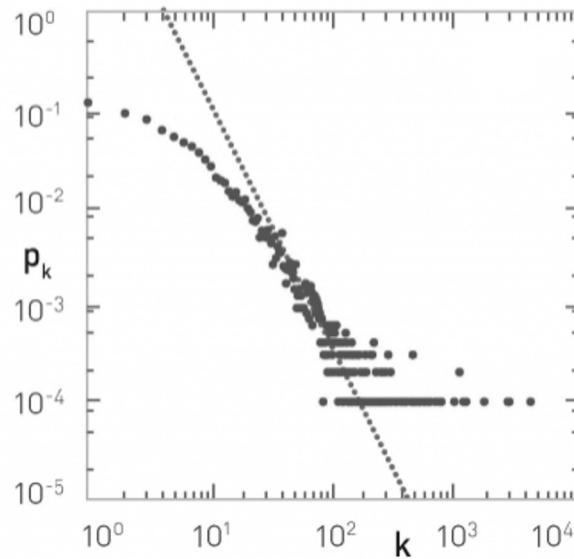
$$\ln P_k \sim \ln k^{-\gamma}$$



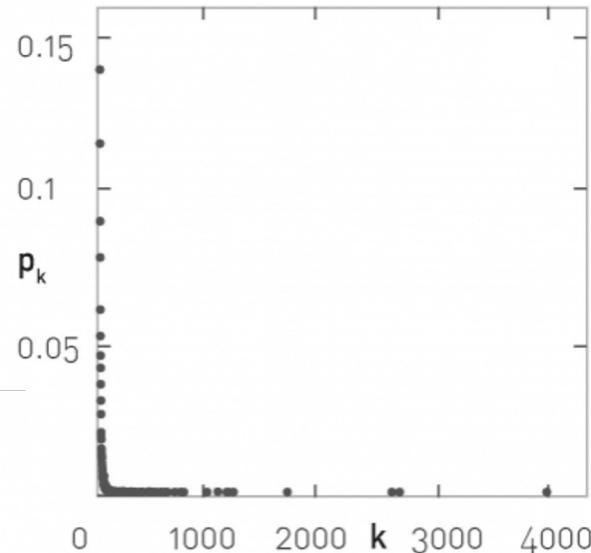
$$\ln P_k \sim -\gamma \ln k$$

Plotting degree distributions

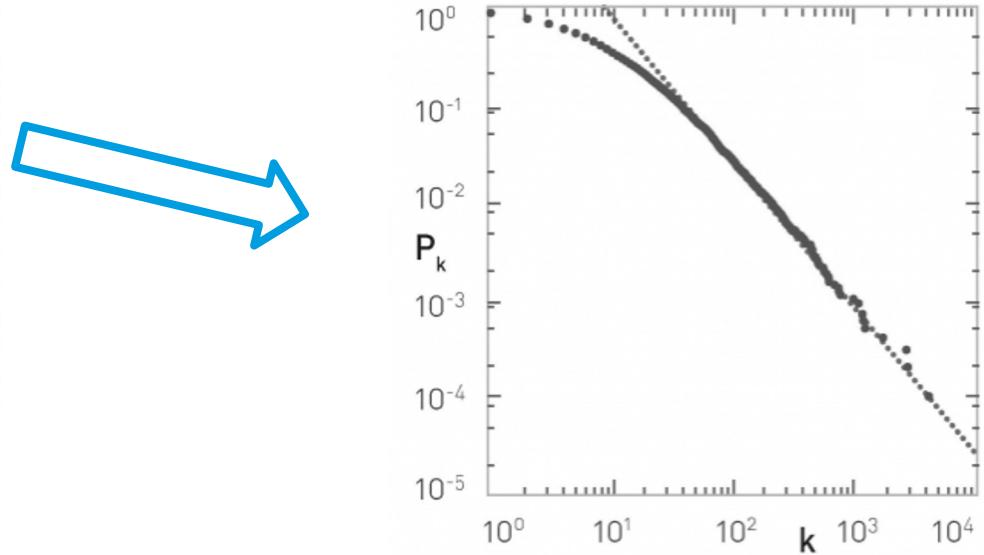
log-log scale



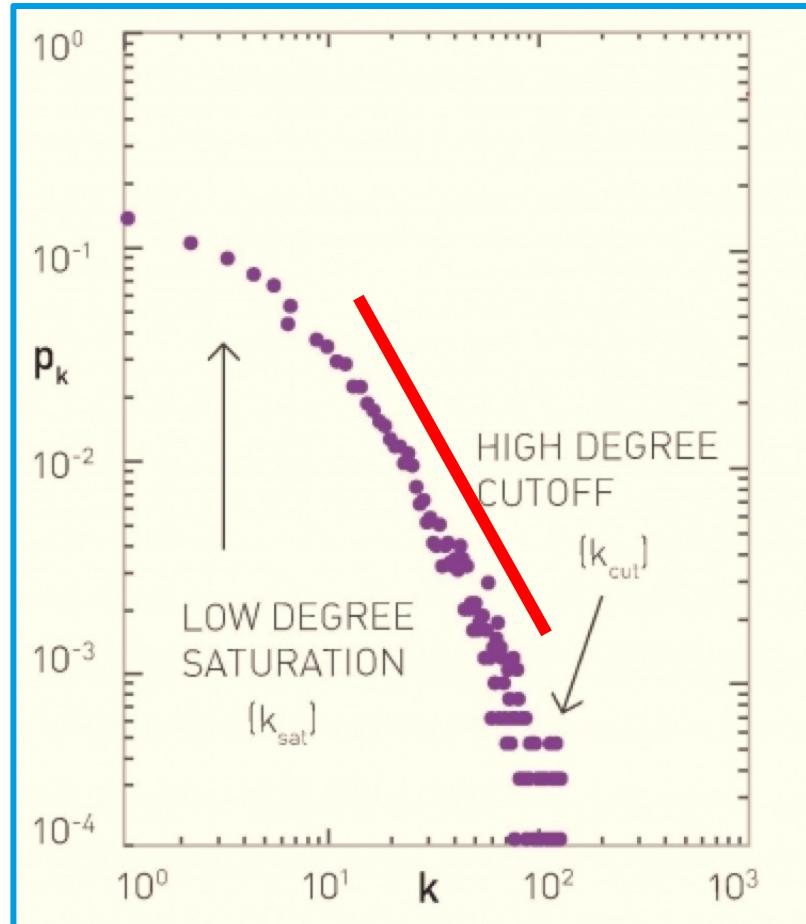
Linear scale



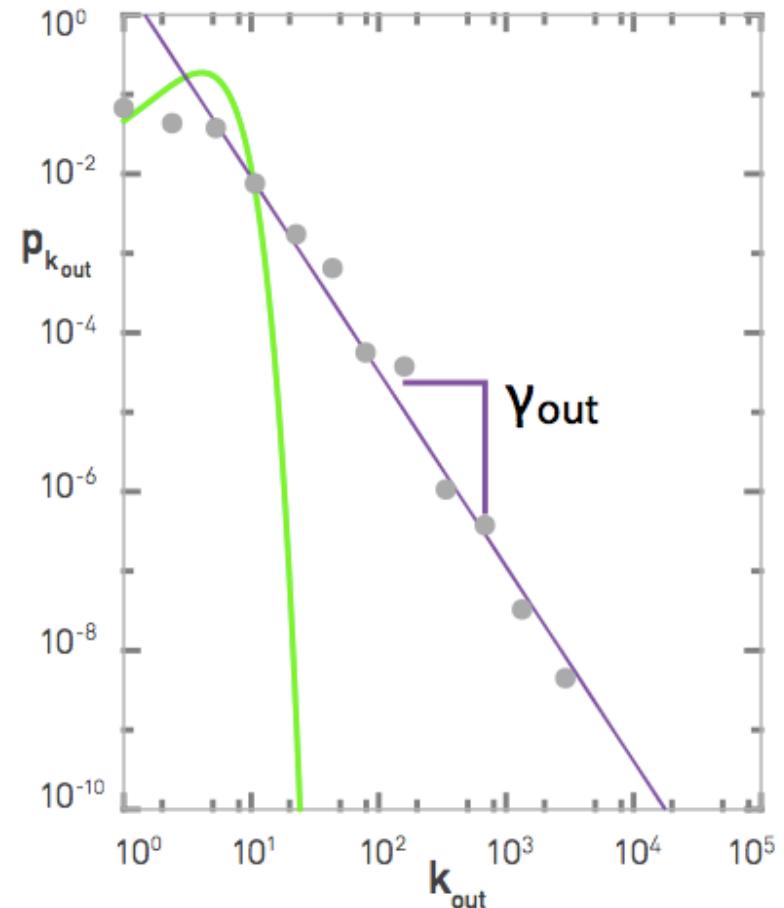
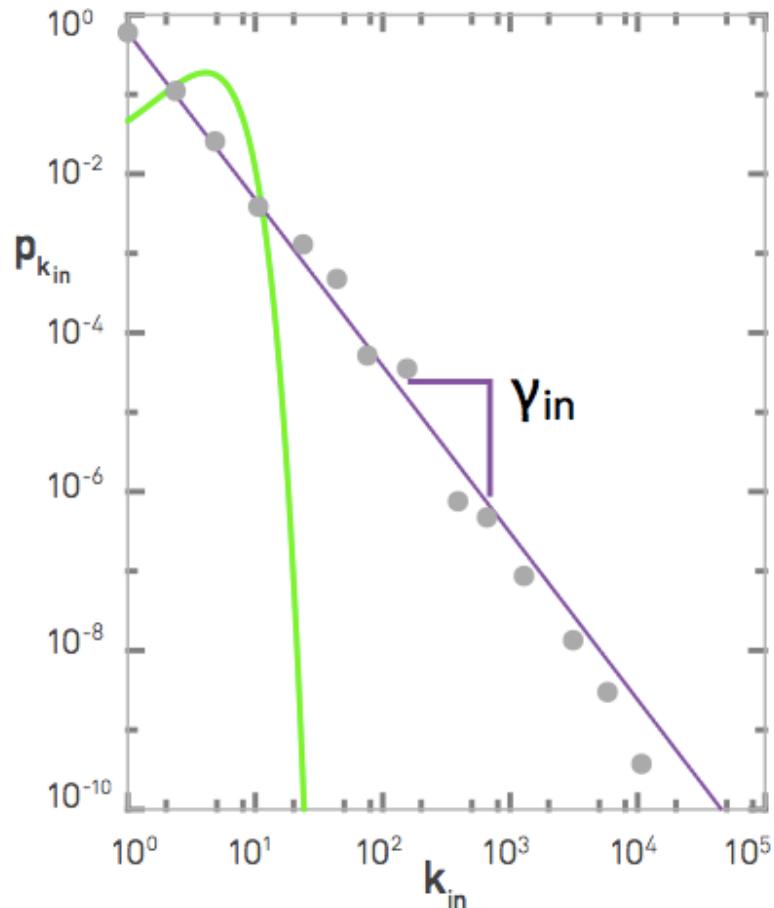
cumulative



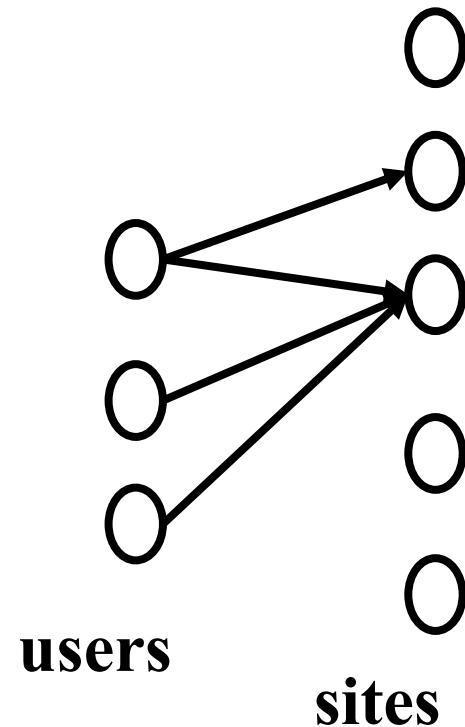
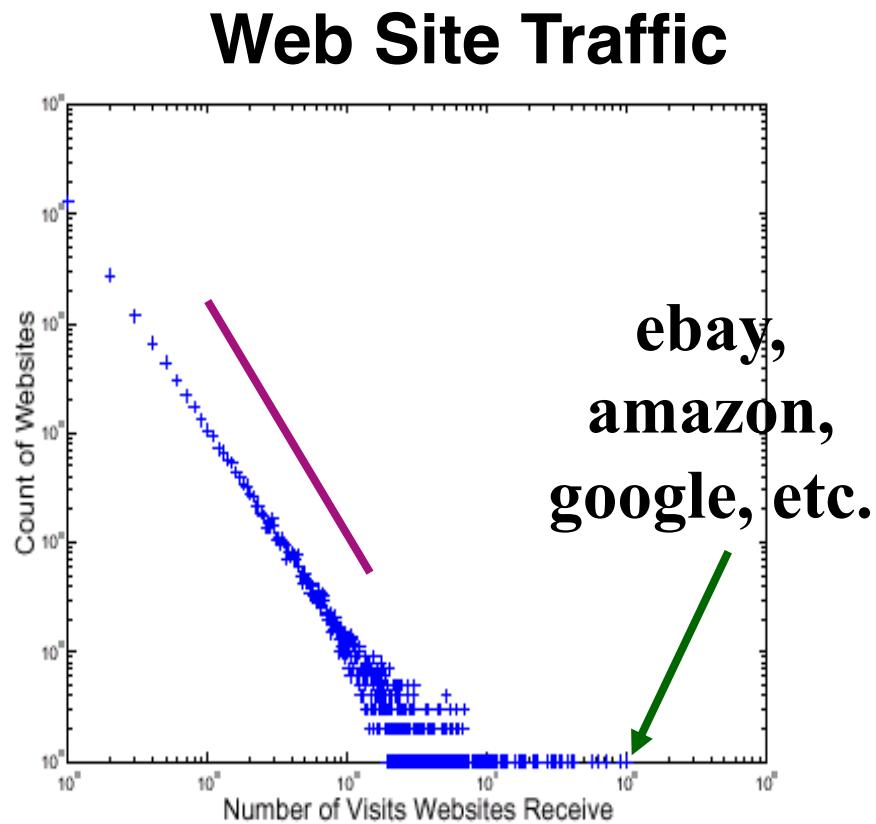
Plotting degree distributions



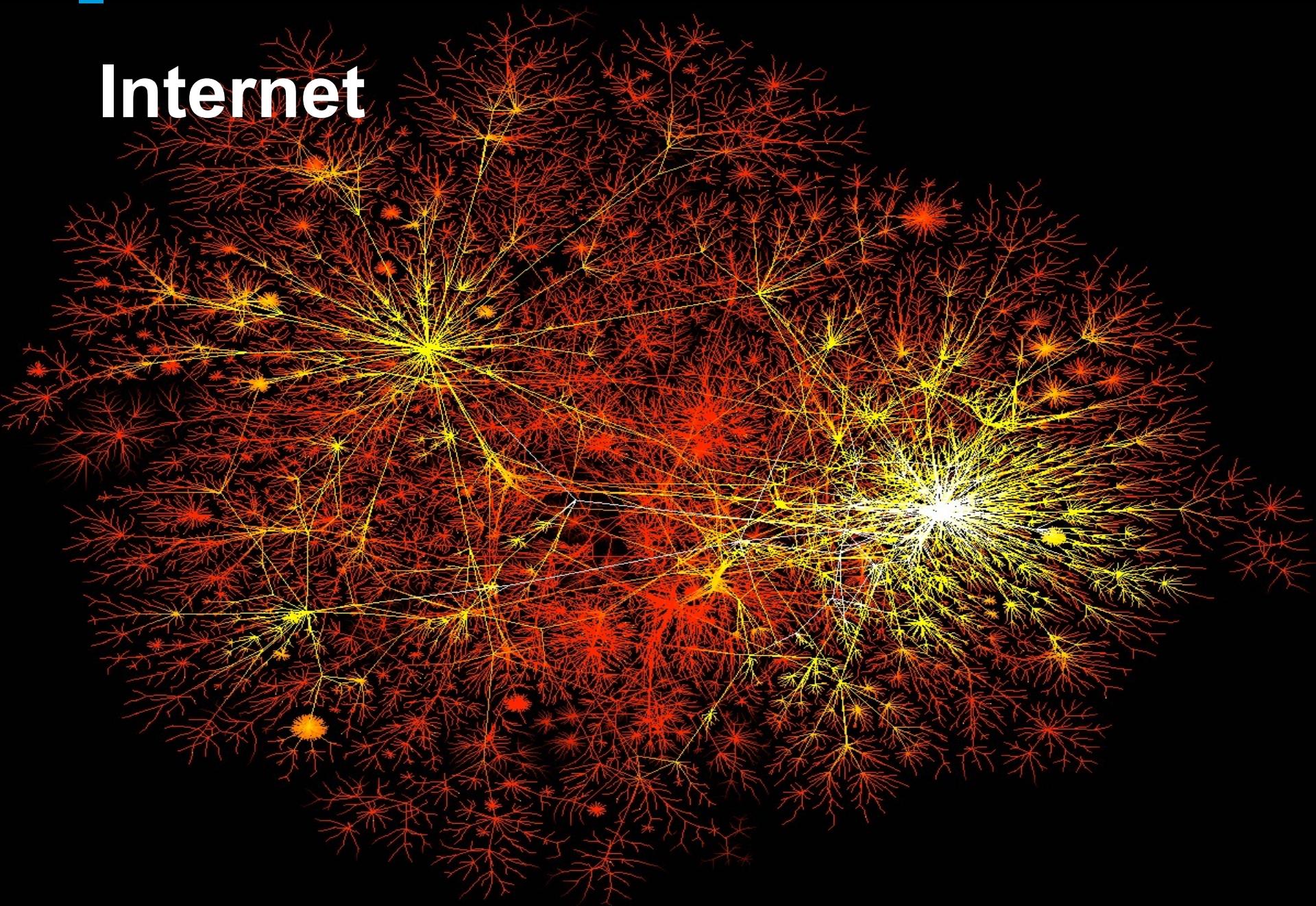
World-Wide-Web as a directed graph: *in & out degree distributions*



Power-laws are everywhere in the WWW



Internet



Internet

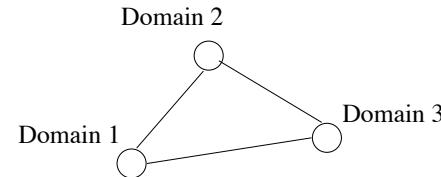
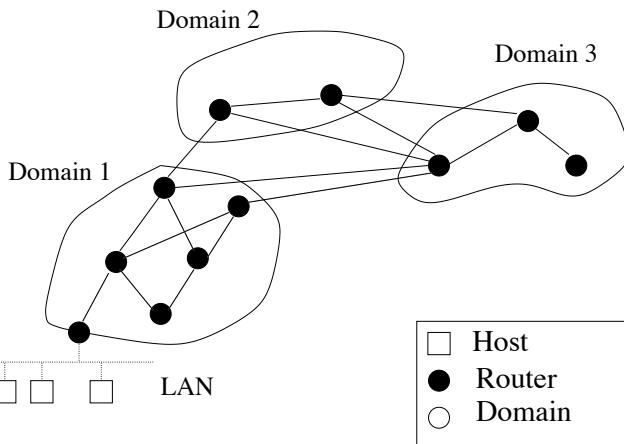
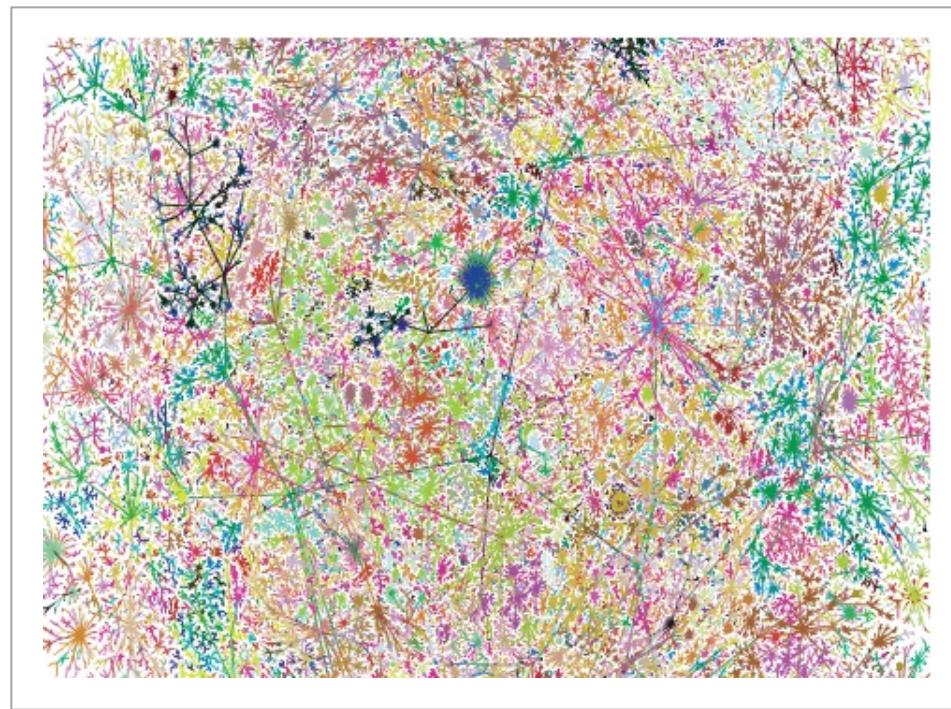
Internet (Faloutsos et al, 1999)

Nodes = computers, and edges
physical connections among them
... quite difficult



2 levels

Nodes = routers or
Nodes = domains

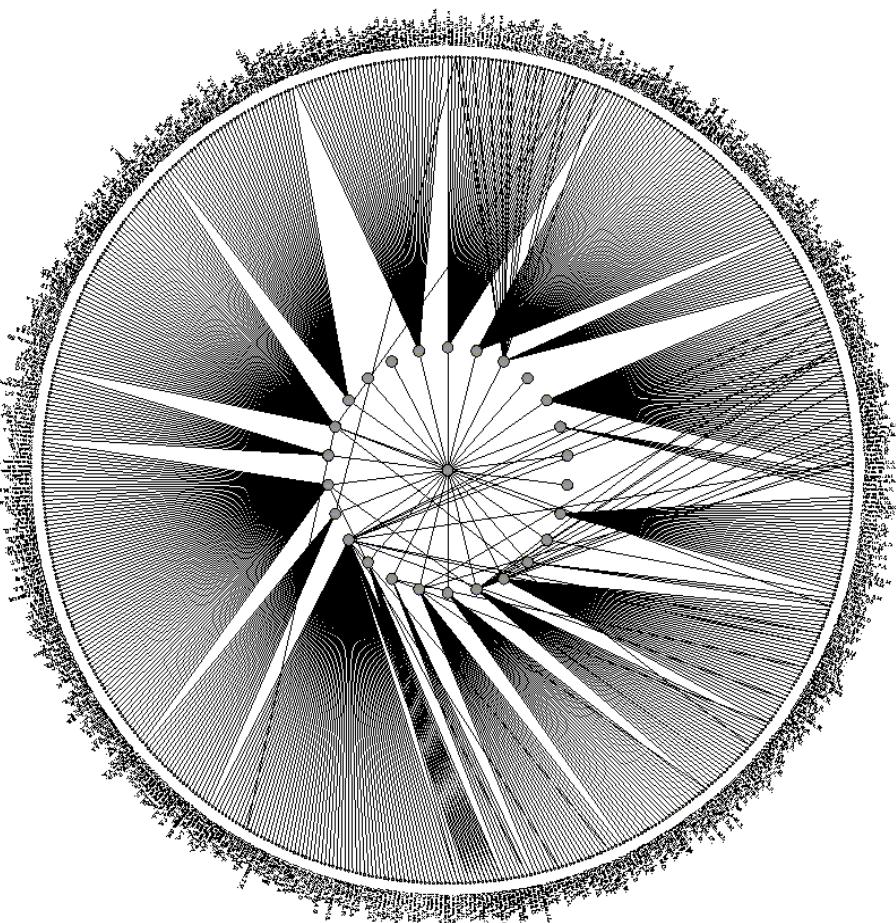


Both showed a scale-free degree dist.

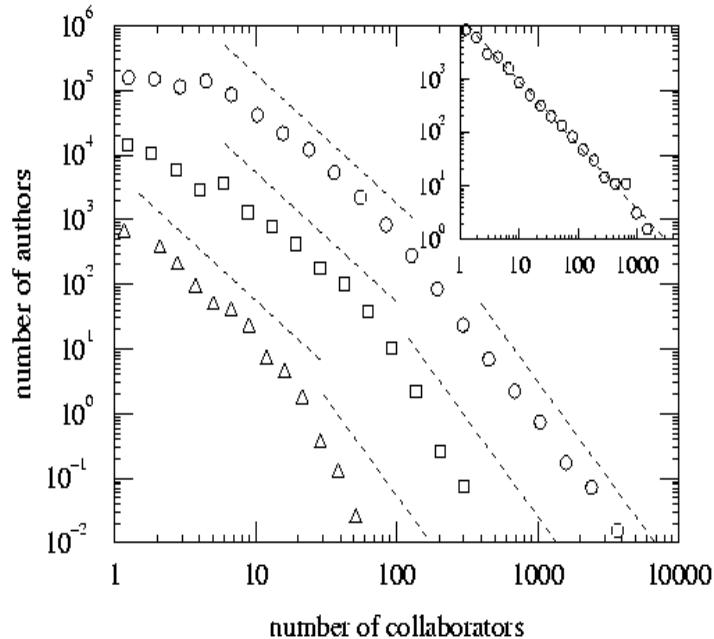
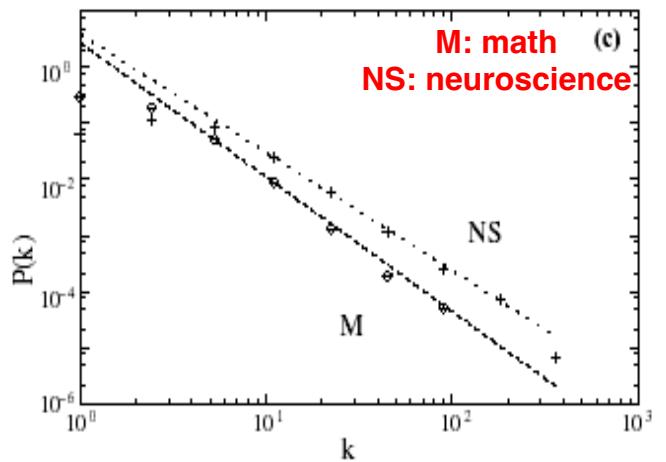
Science collaborations

Nodes: scientist (authors)

Links: joint publication



(Newman, 2000, Barabasi et al 2001)

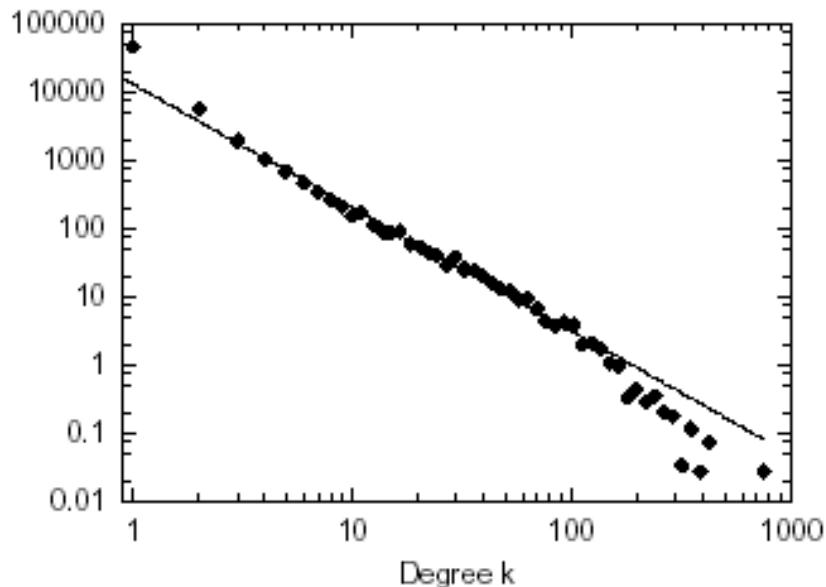


Online communities

Nodes: online user

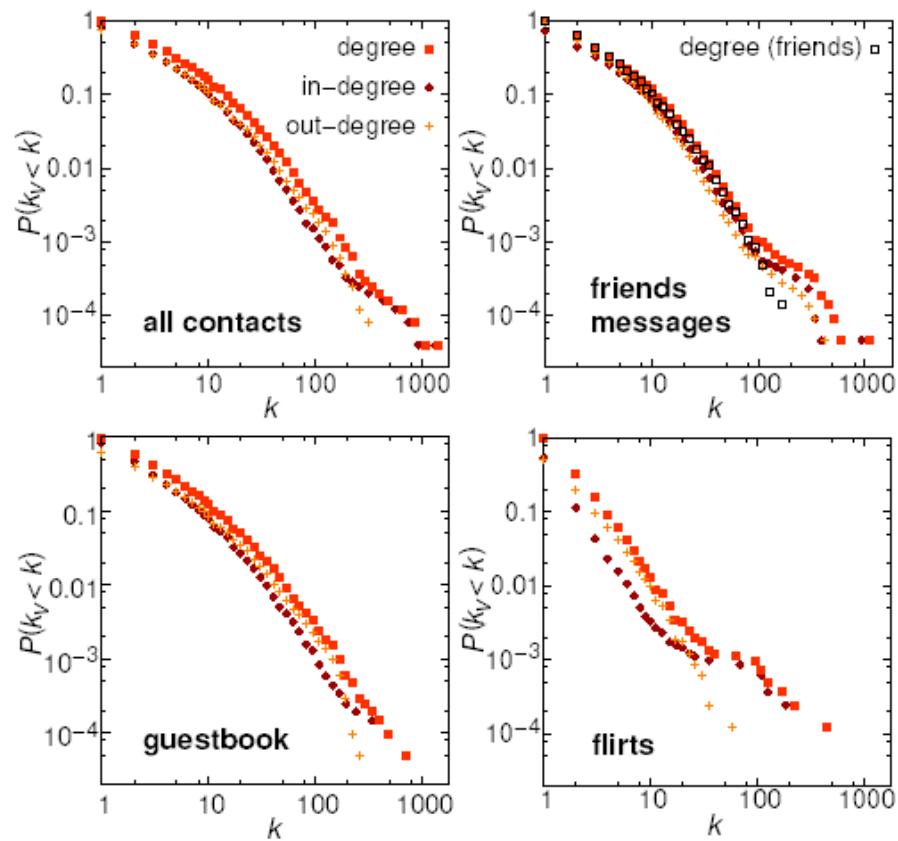
Links: email contact

Kiel University log files
112 days, N=59,912 nodes



Ebel, Mielsch, Bornholdtz, PRE 2002.

Pussokram.com online community;
512 days, 25,000 users.

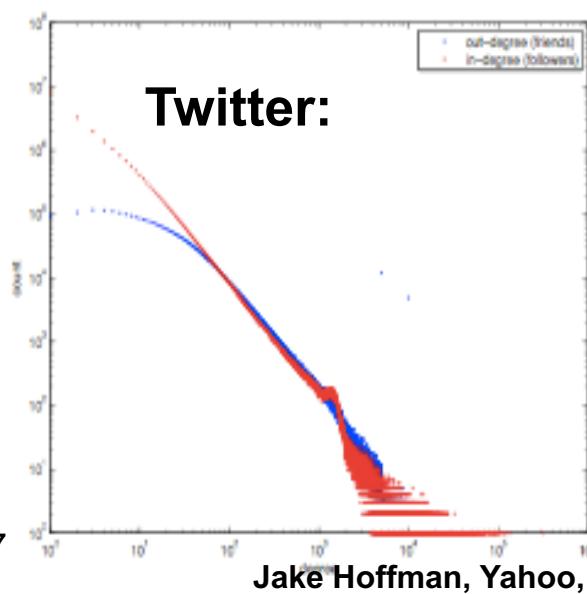


Holme, Edling, Liljeros, 2002.

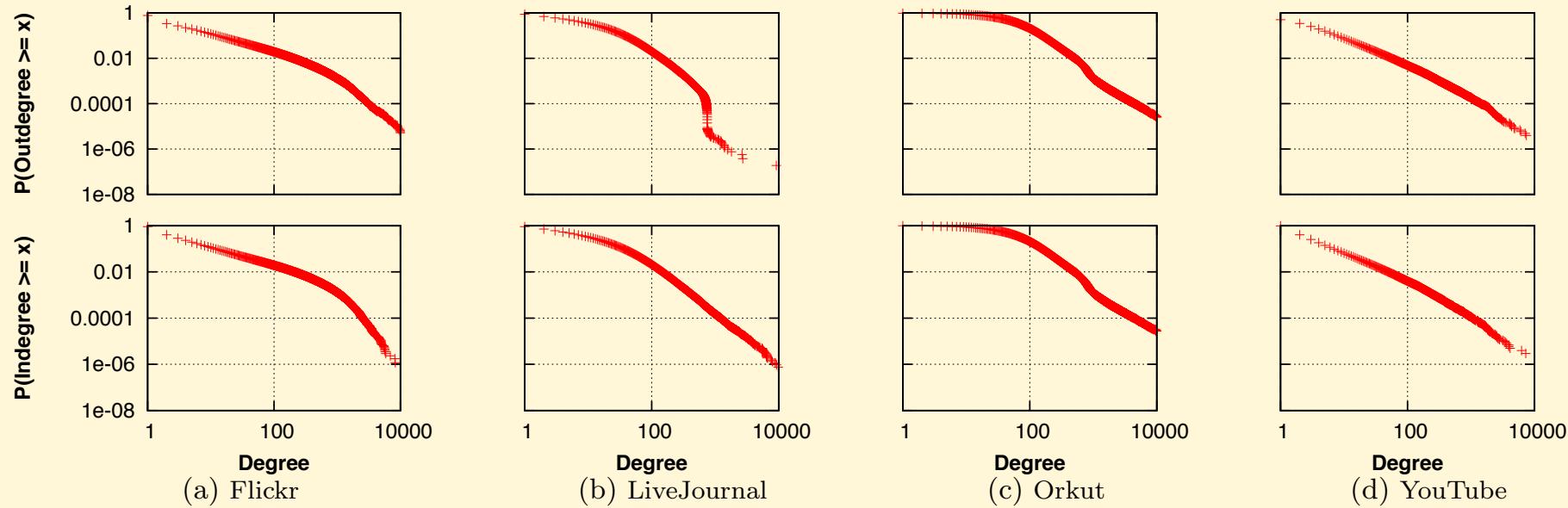
Network Science, 2023/24

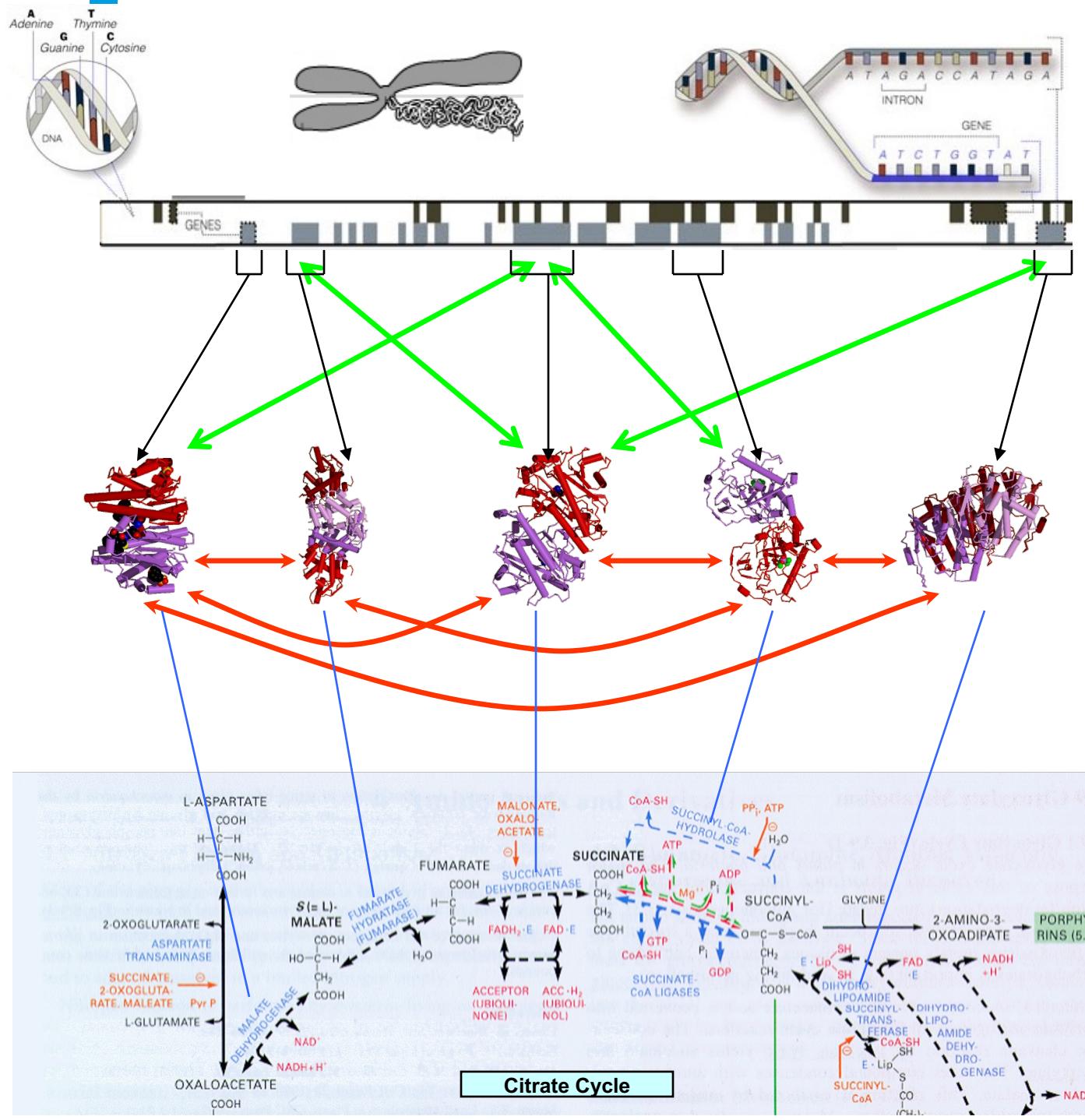
Online communities

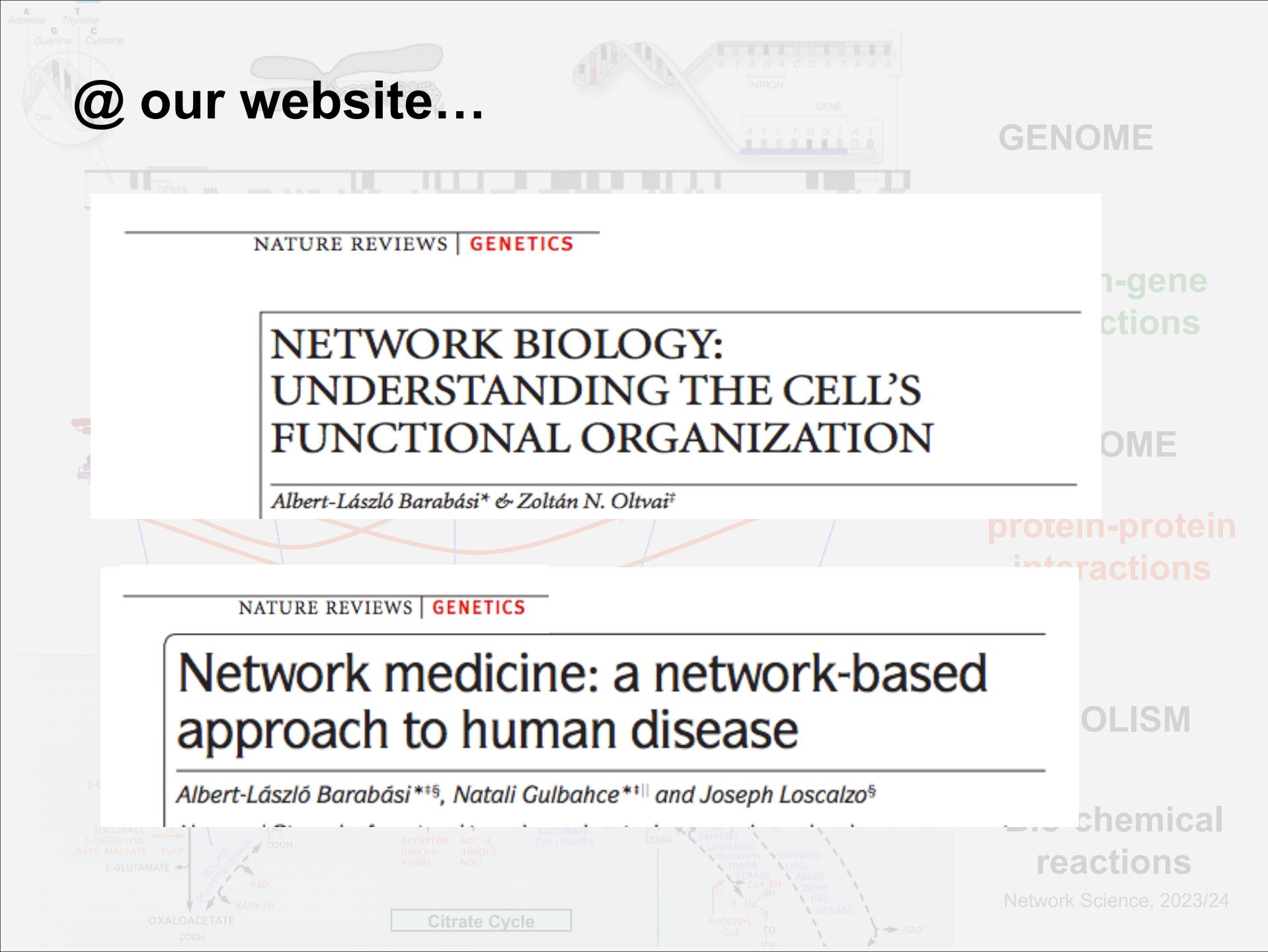
All distributions show a fat-tail behavior:
you have degrees spreading several
orders of magnitude.



Mislove et al., Measurement and Analysis of Online Social Networks, IMC'2007







@ our website...

GENOME

n-gene
ctions

OME

protein-protein
interactions

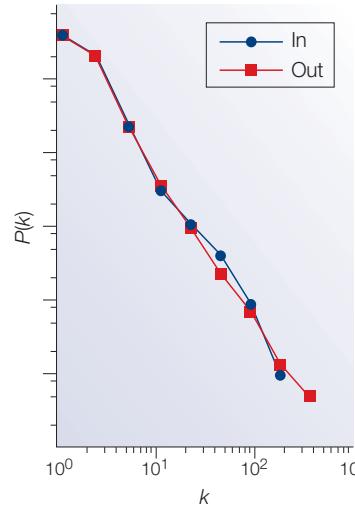
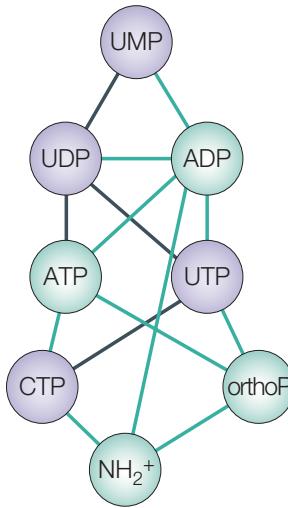
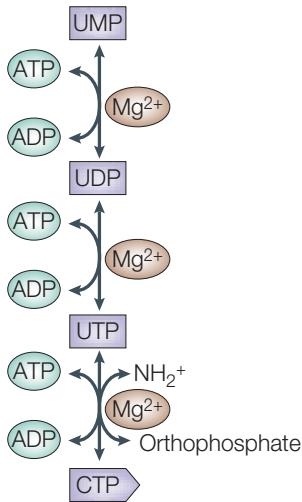
OLISM

chemical
reactions

Network Science, 2023/24

Biological networks: Metabolic networks

Jeong, Albert, Oltvai, Barabási, Fell, Wagner, etc (after 2000).



Nodes : metabolites

Edges : enzyme-catalyzed biochemical reactions

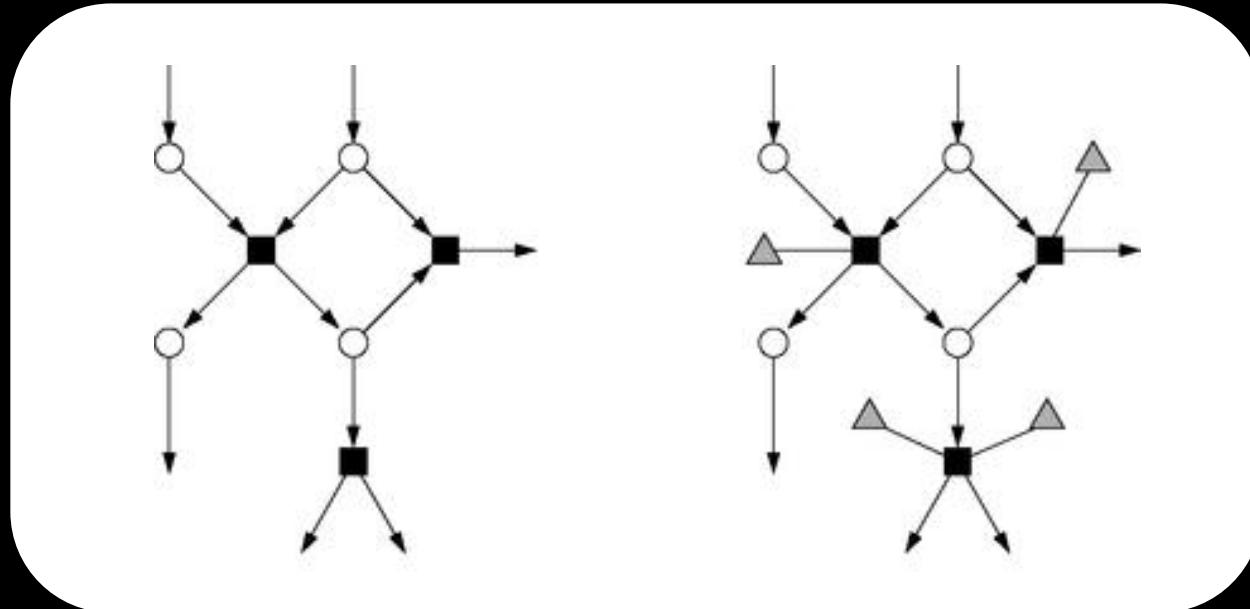
Again, this is a directed graph, as many reactions are irreversible.

The **in** (**out**) degree denotes the number of reactions that produce (consume) a given metabolite.

The analysis of these nets for 43 organisms indicates that “cellular biochemical nets” are **scale-free** (hubs = water, ATP, etc).

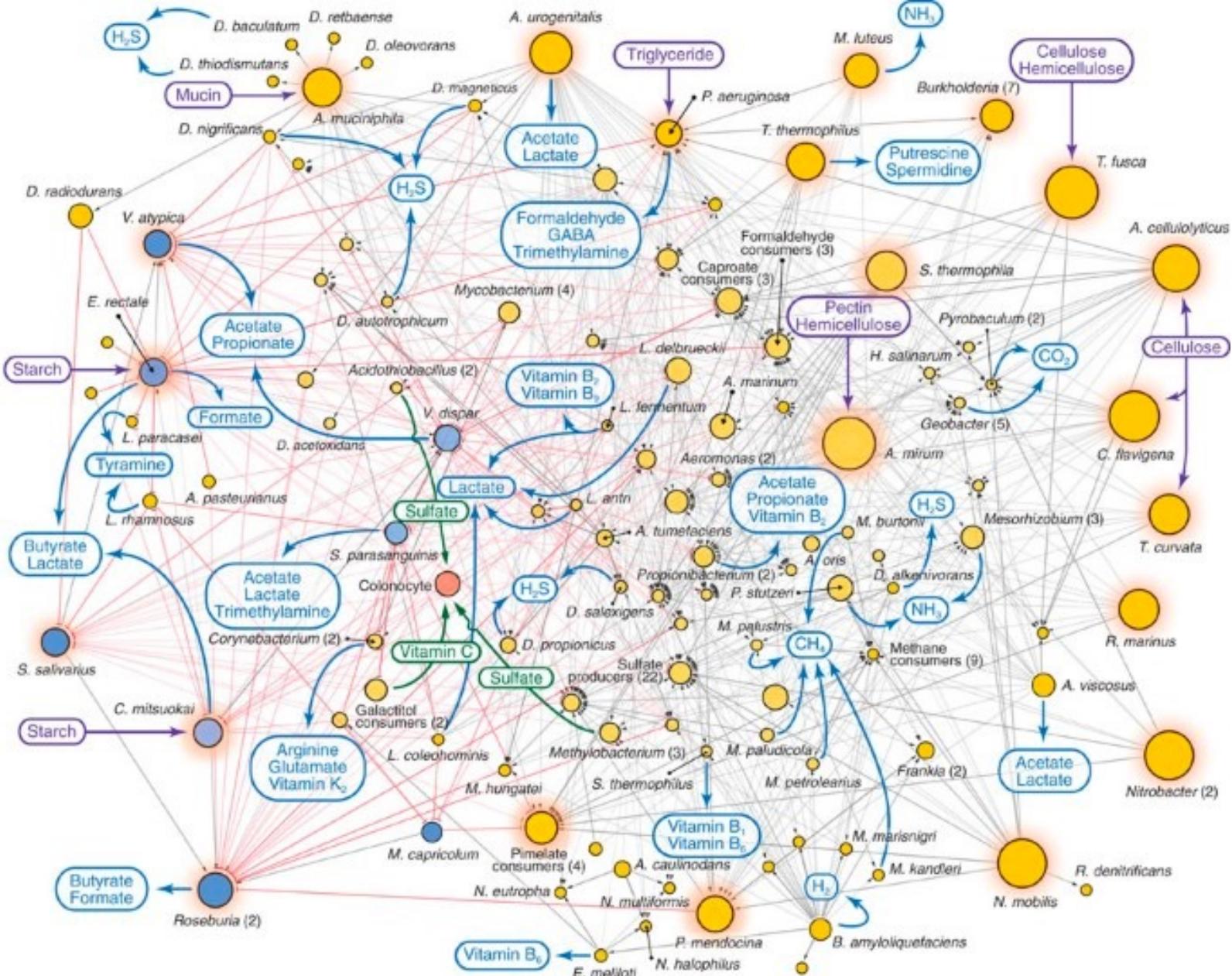
Biological networks: Metabolic networks

Jeong, Albert, Oltvai, Barabási, Fell, Wagner, etc (after 2000).



Bipartite representation (LEFT): metabolites (circles) and reactions (squares) and directed edges indicating which metabolites are substrates (inputs) and products (outputs) of which reactions.

Tripartite representation (RIGHT): A third type of vertex (triangles) can be introduced to represent enzymes, with undirected edges linking them to the reactions they catalyze. The resulting network is a mixed directed/undirected tripartite network.



Control (influencer)

Control (non-influencer)

→ Pos. metabolic influence ($W_{ii} > 0$)

→ Macromolecule degradation

T2D (influencer)

T2D (non-influencer)

→ Neg. metabolic influence ($W_{\text{m}} < 0$)

Metabolite export

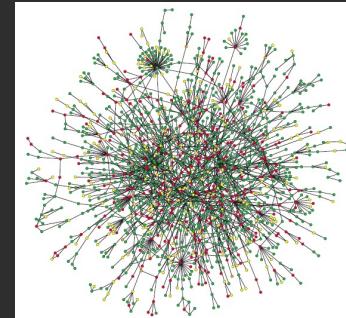
Host (colonocyte)

Biological networks: Protein-interaction networks

Uetz et al, Ito et al, Giot et al, Li et al, Jeong et al, Sneppen, Wagner, etc (after 2000).

Nodes : proteins

Edges : experimental evidence that they bind
creating protein complexes



Biological networks: genetic regulatory nets

Agrawal 2002. Shaw 2003. Provero 2002, Farkas 2003.

Ex:

Nodes : DNA segments

Edges : indirect interaction through RNA and protein
expression products

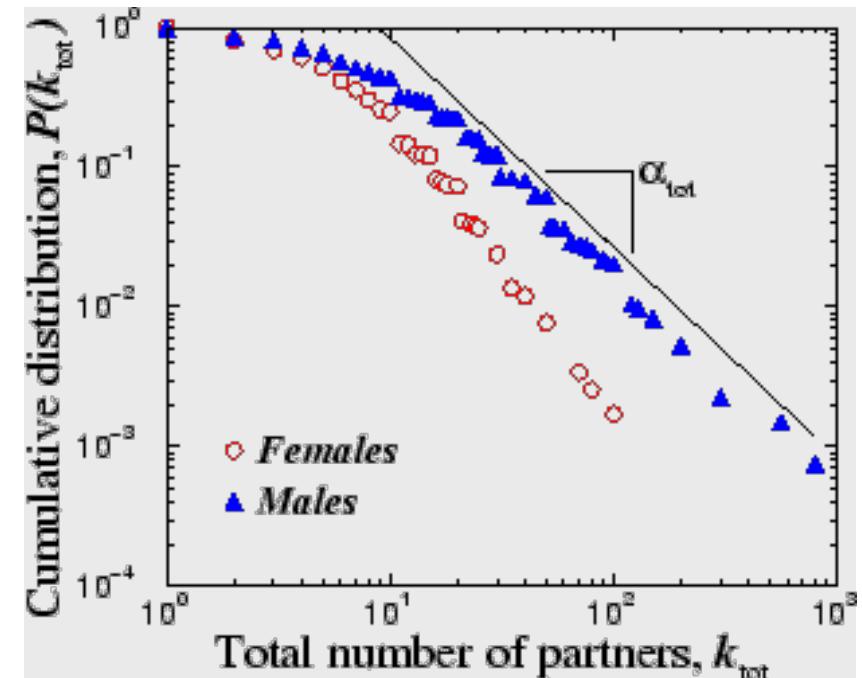


Plethora of Modelling challenges (ex: Boolean nets, evolutionary
networks) and application of “bioinformatics” techniques, DB, etc.

Swedish sexual interaction network



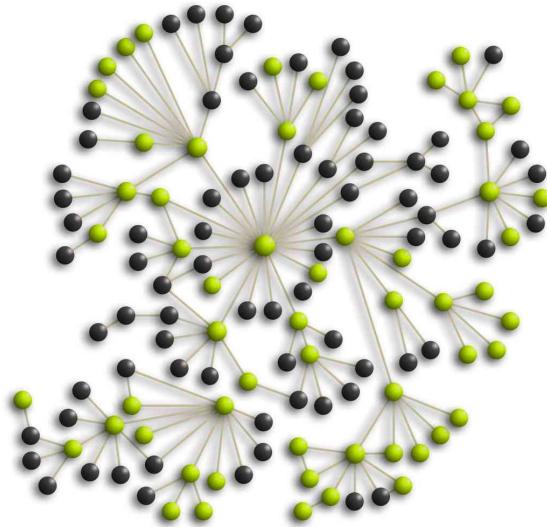
Nodes: people (Females; Males)
Links: sexual relationships



4781 Swedes; 18-74;
59% response rate.

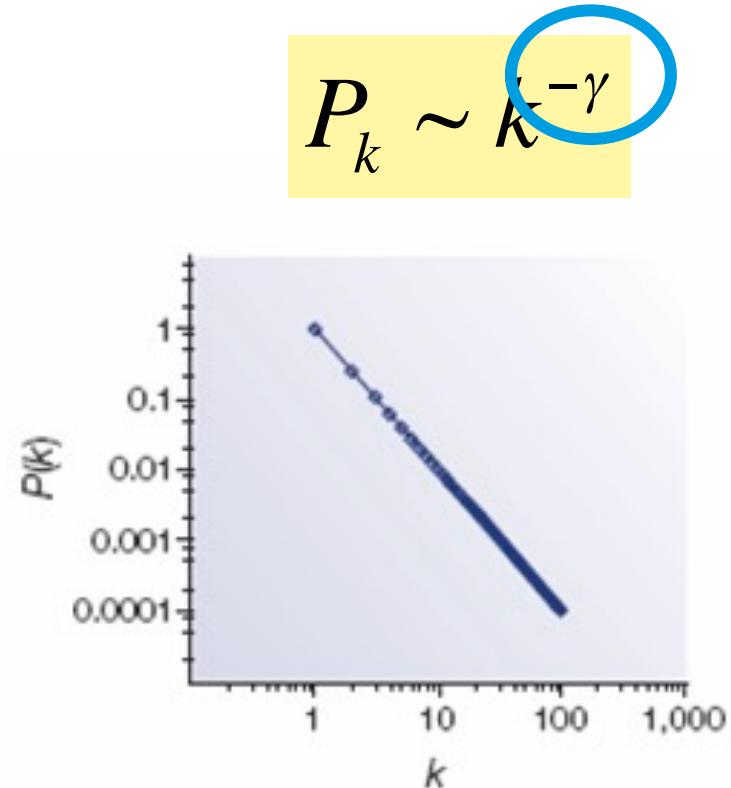
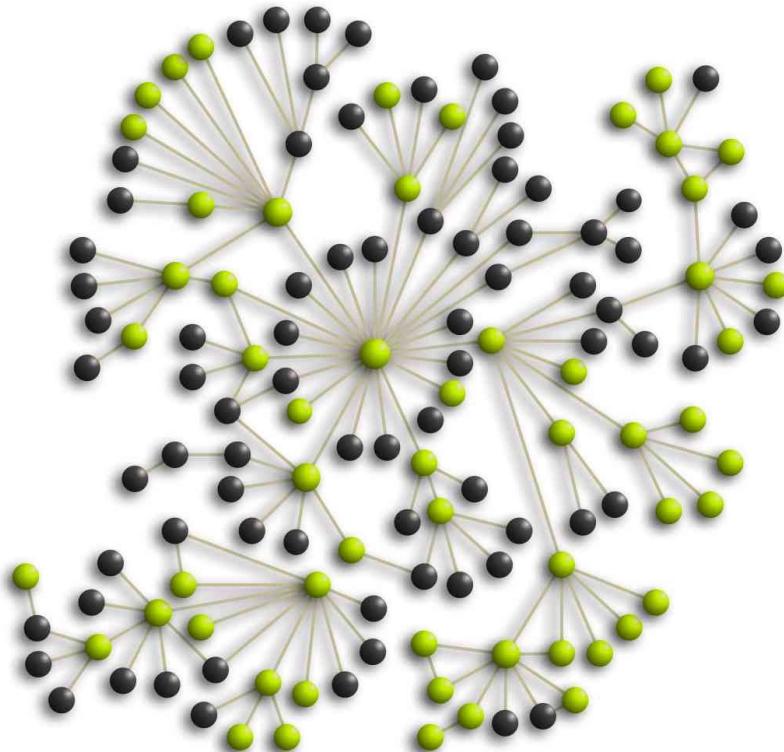
Liljeros et al. Nature 2001

Many real-world nets are scale-free



WWW, Internet (routers and domains), electronic circuits, computer software, movie actors, coauthorship networks, sexual web, instant messaging, email web, citations, phone calls, metabolic, protein interaction, protein domains, brain function web, linguistic networks, comic book characters, international trade, bank system, encryption trust net, energy landscapes, astrophysical network...

Universality? I will get back to this point...



| WWW | actors | citations | sex | cellular | phones | linguistics |
|----------------|----------------|--------------|----------------|----------------|----------------|----------------|
| $\gamma = 2.1$ | $\gamma = 2.3$ | $\gamma = 3$ | $\gamma = 3.5$ | $\gamma = 2.1$ | $\gamma = 2.1$ | $\gamma = 2.8$ |

A closer look at power-law distributions

discrete representation: $P_k = Ck^{-\gamma}$ with C given by

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

Riemann-zeta
function

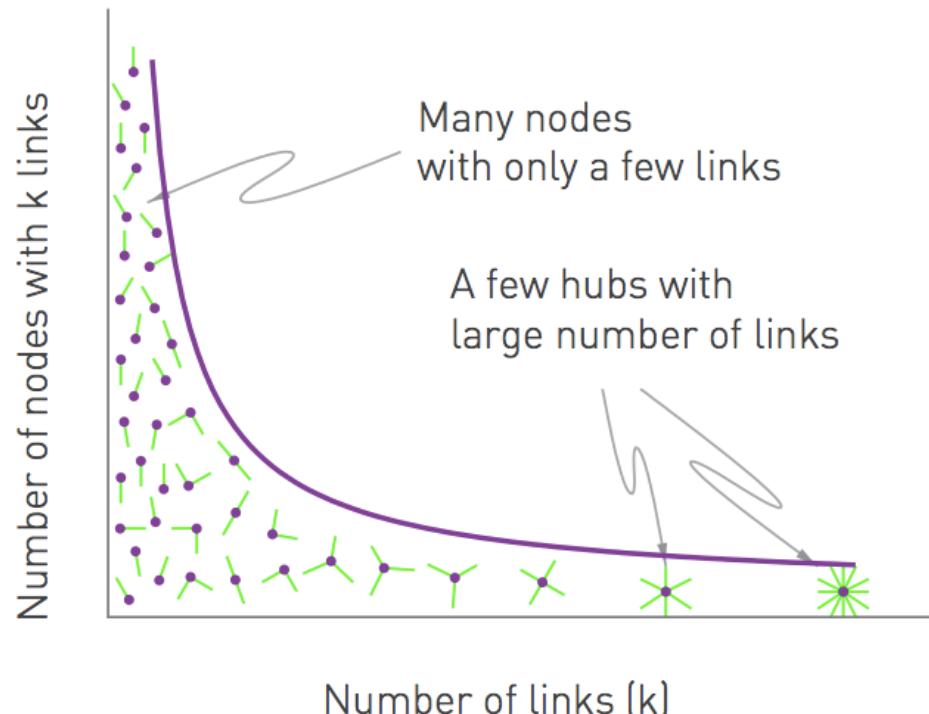
continuum description: $P_k = Ck^{-\gamma}$ with C given by

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

Largest hub

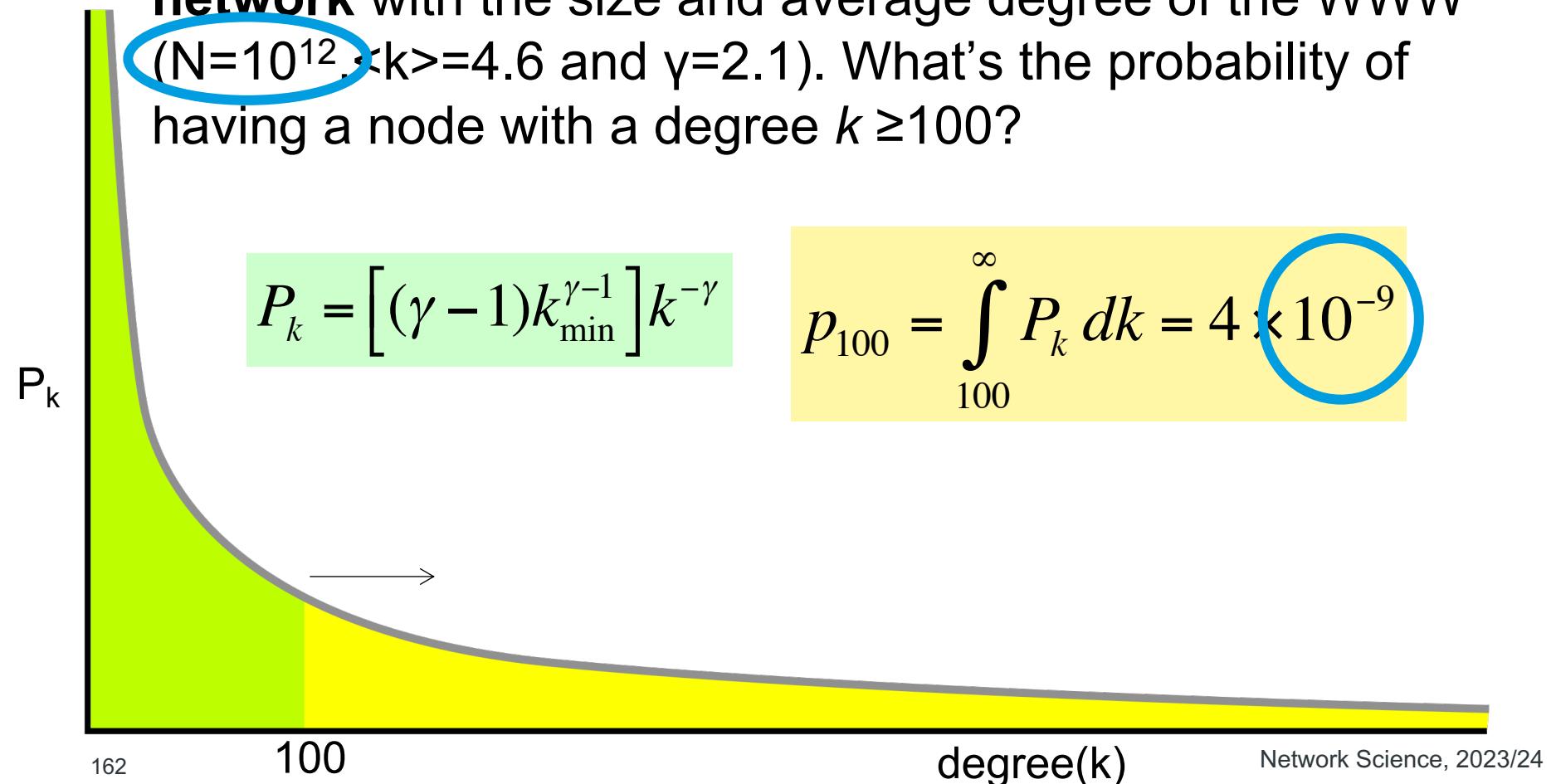
- Just for the fun of it: Imagine that we have a **scale-free network** with the size and average degree of the WWW ($N=10^{12}$, $\langle k \rangle = 4.6$ and $\gamma = 2.1$). What's the probability of having a node with a degree $k \geq 100$?

$$P_k = [(\gamma - 1)k_{\min}^{\gamma-1}]k^{-\gamma}$$



Largest hub

- Just for the fun of it: Imagine that we have a **scale-free network** with the size and average degree of the WWW ($N=10^{12}$, $\langle k \rangle = 4.6$ and $\gamma=2.1$). What's the probability of having a node with a degree $k \geq 100$?

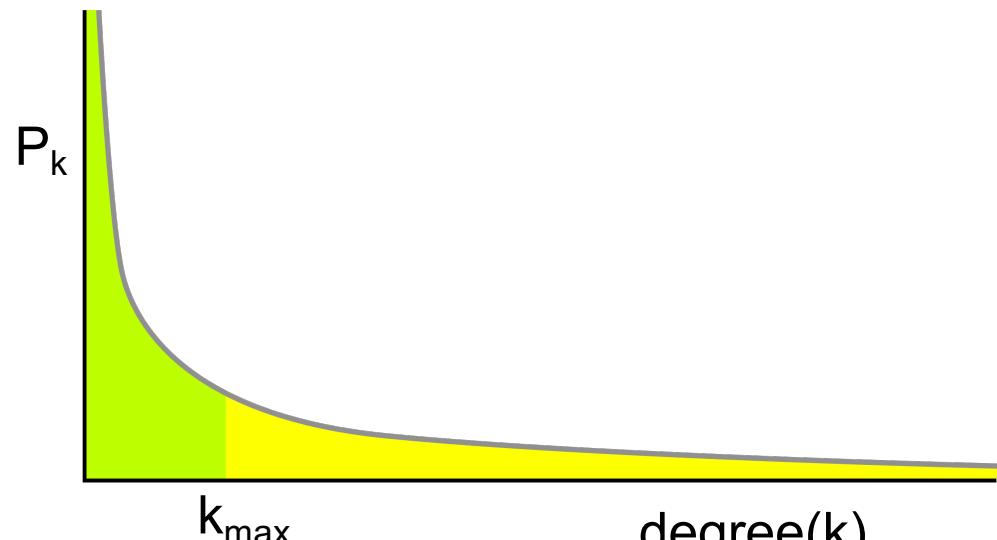


Can we estimate the largest hub?

- We have the degree dist:
- k_{\max} will be given by

$$P_k = [(\gamma - 1)k_{\min}^{\gamma-1}]k^{-\gamma}$$

$$\int_{k_{\max}}^{\infty} P_k dk < \frac{1}{N}$$

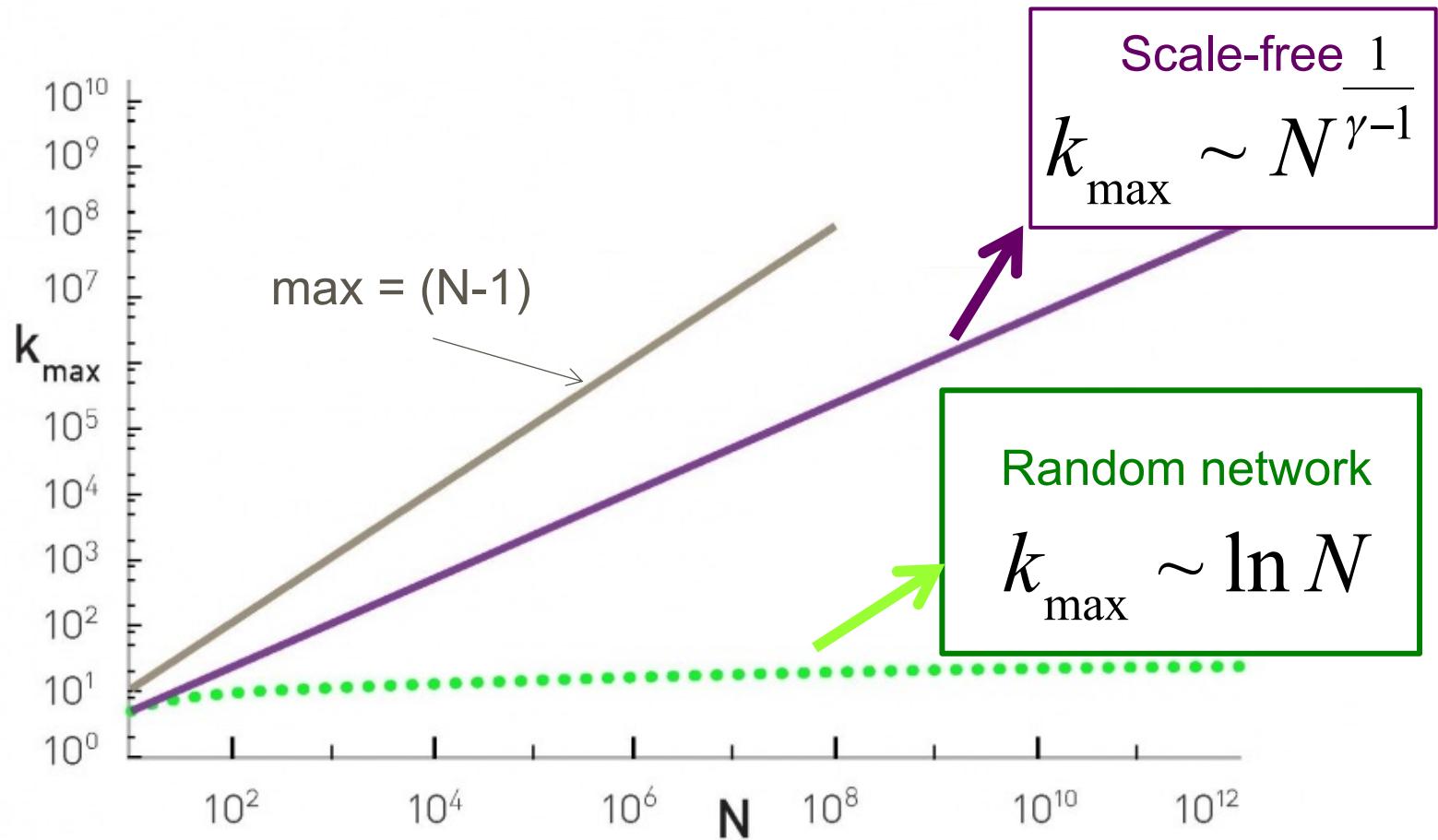


- Which yields

$$k_{\max} = k_{\min} N^{\frac{1}{\gamma-1}}$$

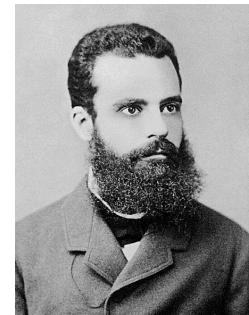
- Example: WWW gives $k_{\max}=95000$

Can we estimate the largest hub?



Revisiting Pareto

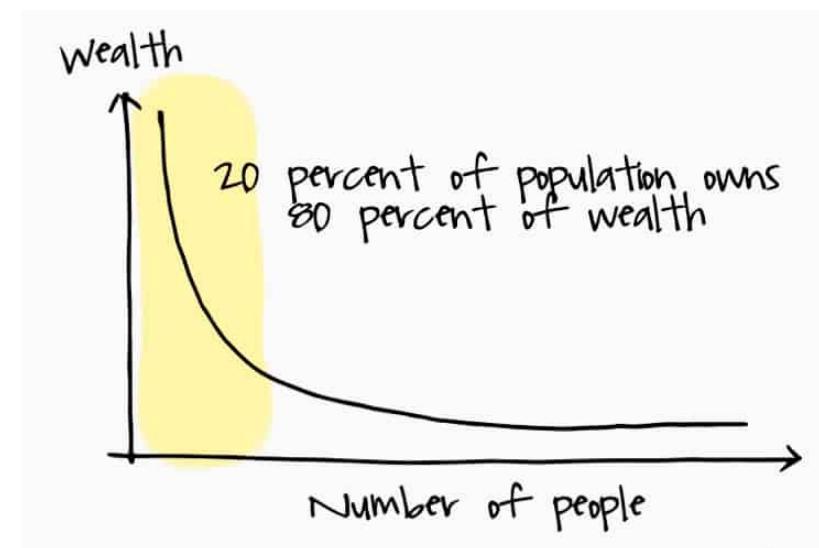
Vilfred Pareto
(1848-1923)



- Power-law distributions are also called Pareto distributions.
- Power-laws is the principle behind the 80/20 rule:
Roughly 80 percent of money is earned by only 20 percent of the population.

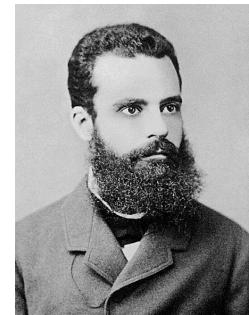
Distribution of world GDP

| Quintile of population | Income |
|------------------------|--------|
| Richest 20% | 82.70% |
| Second 20% | 11.75% |
| Third 20% | 2.30% |
| Fourth 20% | 1.85% |
| Poorest 20% | 1.40% |



Revisiting Pareto

Vilfred Pareto
(1848-1923)

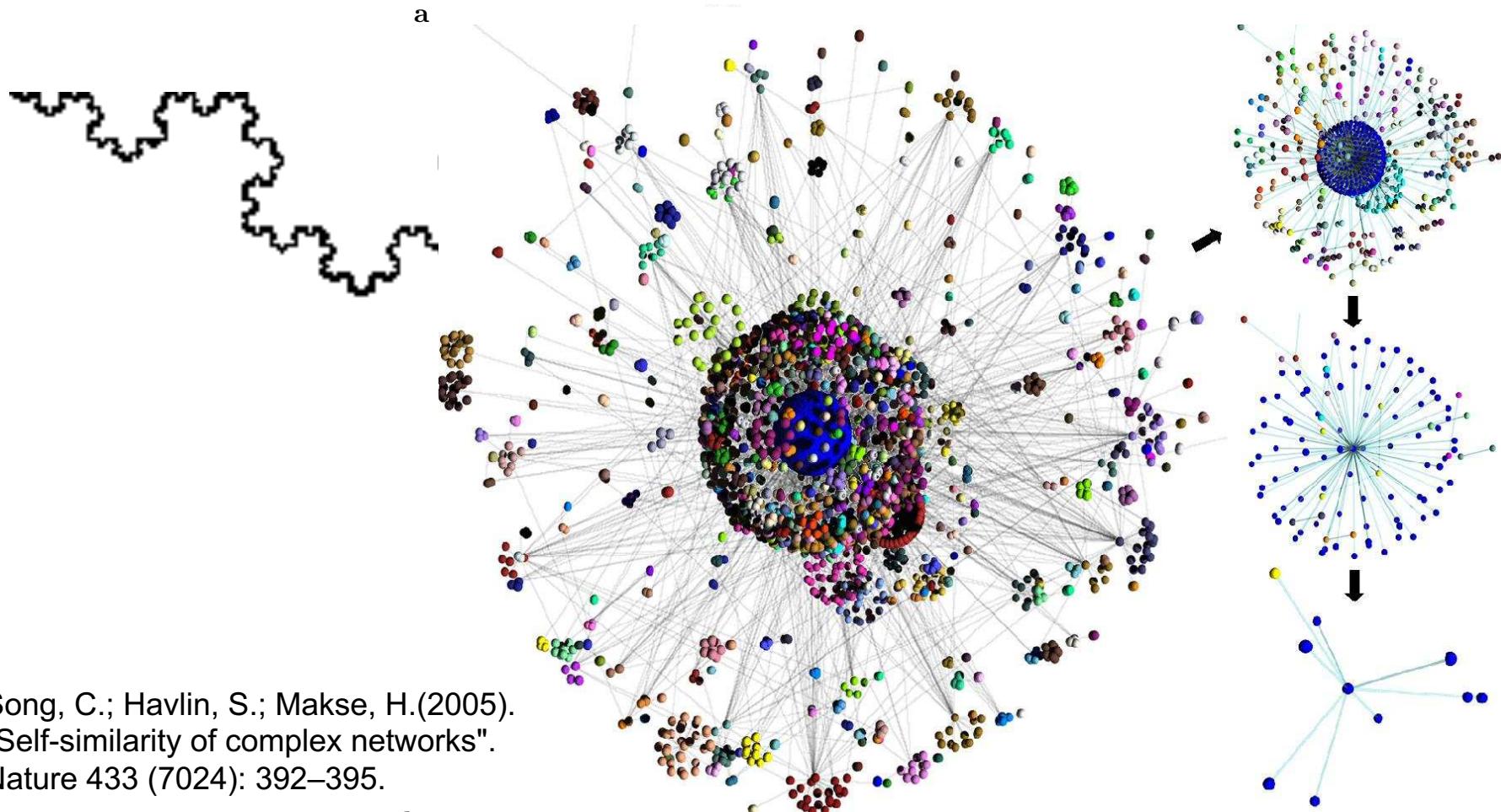


- Power-law distributions are also called Pareto distributions.
- Power-laws is the principle behind the 80/20 rule:
Roughly 80 percent of money is earned by only 20 percent of the population.
 - WWW: 80% of the links point to 15% of the pages.
 - Citations: 80% of all citations point to 38% of scientists
 - Hollywood: 80% of all links connect 30% of actors.
 - USA: 1% of the population earns 15% of the total US income.

Scale-free? What does it mean?

1st explanation: Scale invariance

Scale-invariance, self-similarity, etc.



Song, C.; Havlin, S.; Makse, H.(2005).
"Self-similarity of complex networks".
Nature 433 (7024): 392–395.

Scale-free? What does it mean?

1st explanation: Scale invariance

Scale-invariance?

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

$$P(ck) = (ck)^{-\gamma} = \text{Const.} P(k) \propto P(k)$$

- Scaling the argument k by a constant factor c causes only a proportionate scaling of the function itself.
- In other words, all power laws (with a given exponent) are equivalent up to constant factors, since each is simply a scaled version of the others.

Scale-free? What does it mean?

2nd explanation: Lack of well-defined average value

- *nth moments of a distribution*

$$\langle k^n \rangle = \sum_{k_{\min}}^{\infty} k^n P_k \approx \int_{k_{\min}}^{\infty} k^n P_k dk$$

- n=0 sums to one.
- n=1 gives the **average** degree
- n=2 helps us to calculate the **variance**
- n=3 determines the **skewness**...

$$\sigma^2 = \langle k^2 \rangle - \langle k \rangle^2$$

Scale-free? What does it mean?

2nd explanation: Lack of well-defined average value

- For Scale-free networks we have

$$\langle k^n \rangle \approx \int_{k_{\min}}^{k_{\max}} k^n P_k dk = C \frac{k_{\max}^{n-\gamma+1} - k_{\min}^{n-\gamma+1}}{n - \gamma + 1}$$

- n=0 sums to one.
- n=1 gives the **average** degree
- n=2 helps us to calculate the **variance**
- n=3 determines the **skewness**

$$k_{\max} = k_{\min} N^{\frac{1}{\gamma-1}}$$



Scales with N

Scale-free? What does it mean?

2nd explanation: Lack of well-defined average value

- For Scale-free networks we have

$$\langle k^n \rangle \approx \int_{k_{\min}}^{k_{\max}} k^n P_k dk = C \frac{k_{\max}^{n-\gamma+1} - k_{\min}^{n-\gamma+1}}{n - \gamma + 1}$$

Do not scale with N

- n=0 sums to one.
- n=1 gives the **average** degree
- n=2 helps us to calculate the **variance**
- n=3 determines the **skewness**

$$k_{\max} = k_{\min} N^{\frac{1}{\gamma-1}}$$

Scales with N

Scale-free? What does it mean?

2nd explanation: Lack of well-defined average value

- For Scale-free networks we have (for large N)

$$\langle k^n \rangle \xrightarrow{N \rightarrow \infty} N^{\frac{n-\gamma+1}{\gamma-1}}$$

For large N,
it diverges for

$\gamma < 2$

$$\langle k \rangle \rightarrow N^{\frac{2-\gamma}{\gamma-1}}$$

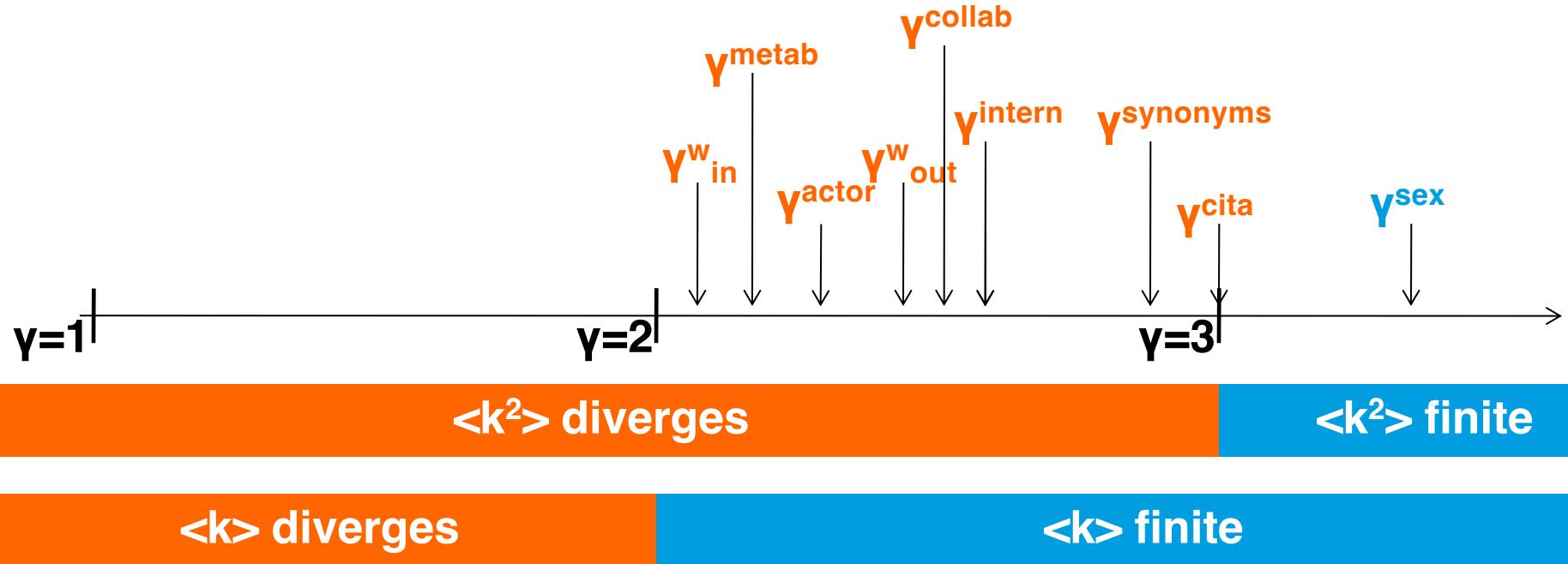
$$\langle k^2 \rangle \rightarrow N^{\frac{3-\gamma}{\gamma-1}}$$

- $n=0$ sums to one.
- $n=1$ gives the **average** degree
- $n=2$ helps us to calculate the **variance**
- $n=3$ determines the **skewness**

Diverges for

$\gamma < 3$

Scale-free? What does it mean?

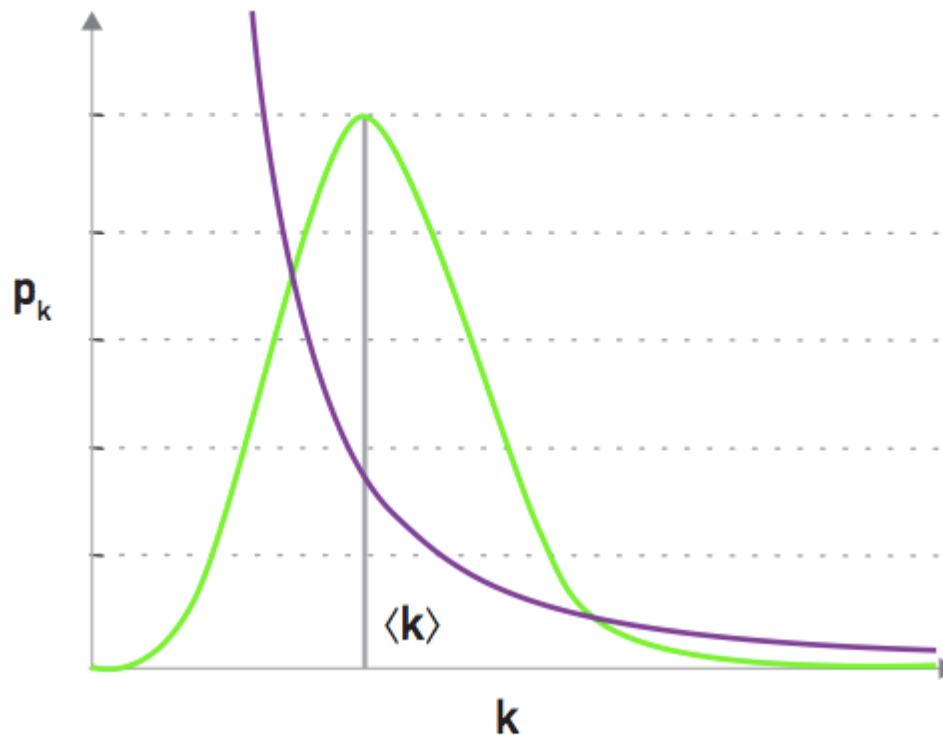


Most networks are in a regime in which the variance diverges for large N ☺

$$k = \langle k \rangle \pm \infty$$

For large N, average values are not meaningful, as fluctuations are too large!

Scale-free? What does it mean?



Random Network

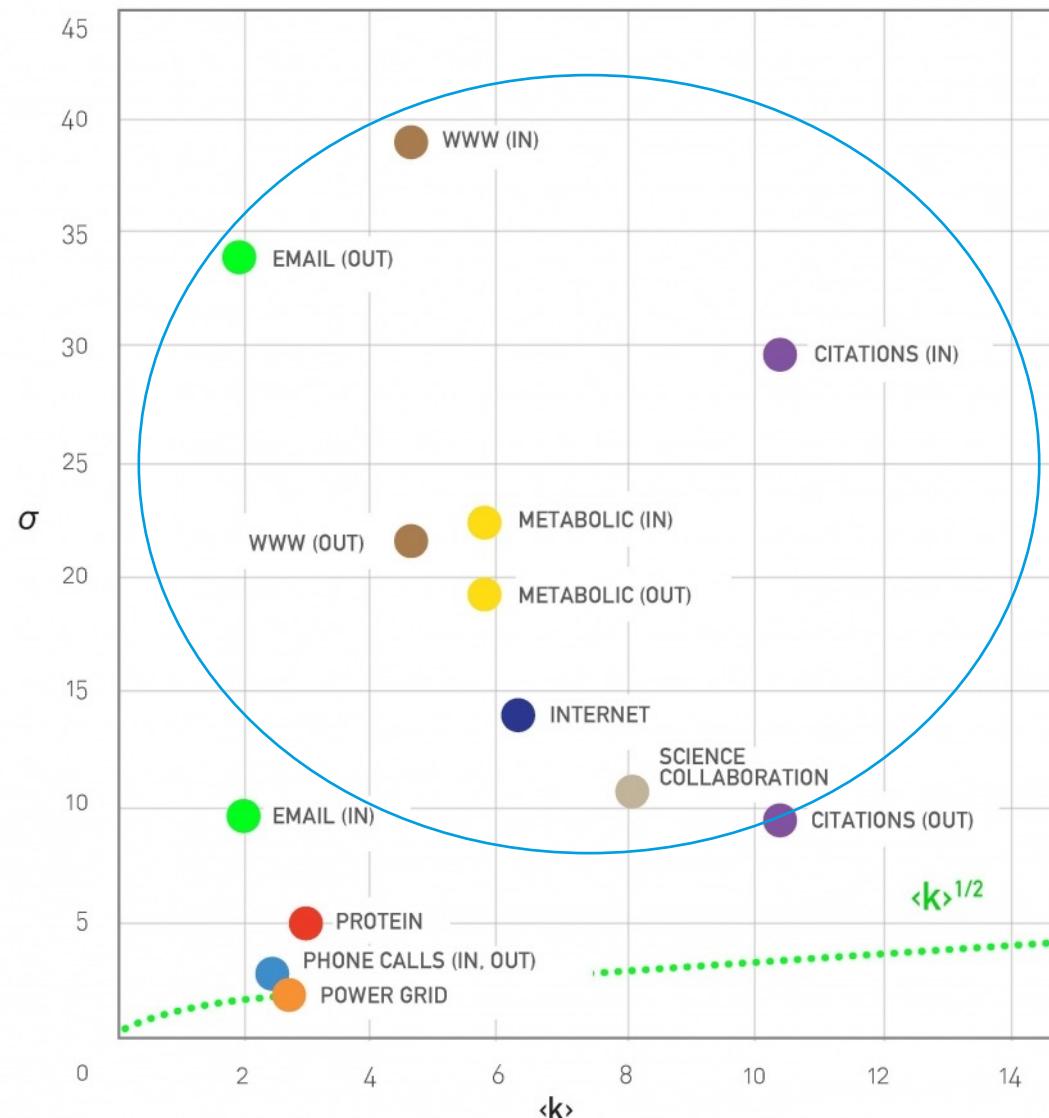
Randomly chosen node: $k = \langle k \rangle \pm \langle k \rangle^{1/2}$
Scale: $\langle k \rangle$

Scale-Free Network

Randomly chosen node: $k = \langle k \rangle \pm \infty$
Scale: none

Scale-free? Networks are finite, yet...

Standard deviation is very large in real networks



Do hubs affect the small world property?

Let's compute the average path length (APL) for a scale-free network...

Do we live in a ultra-small-world?

$$APL = \langle L \rangle \approx$$

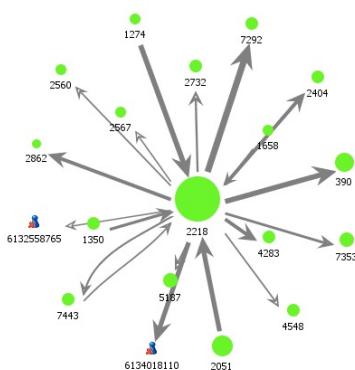
| | |
|---------------------------|------------------|
| const. | $\gamma = 2$ |
| $\ln \ln N$ | $2 < \gamma < 3$ |
| $\frac{\ln N}{\ln \ln N}$ | $\gamma = 3$ |
| $\ln N$ | $\gamma > 3$ |

Anomalous regime ($\gamma \leq 2$)

$$k_{\max} = k_{\min} N^{\frac{1}{\gamma-1}}$$



**K_{\max} grows
faster than N**



const.

$\gamma=2$

$\ln \ln N$

$2 < \gamma < 3$

$\frac{\ln N}{\ln \ln N}$

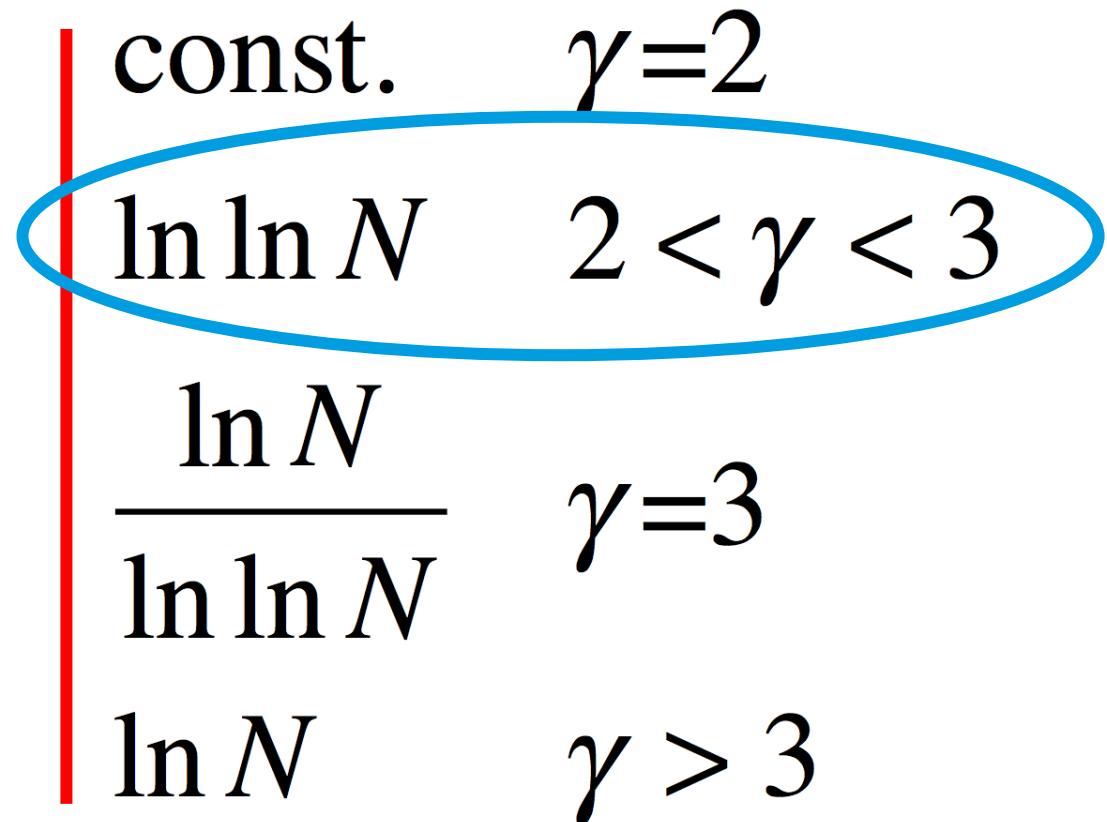
$\gamma=3$

$\ln N$

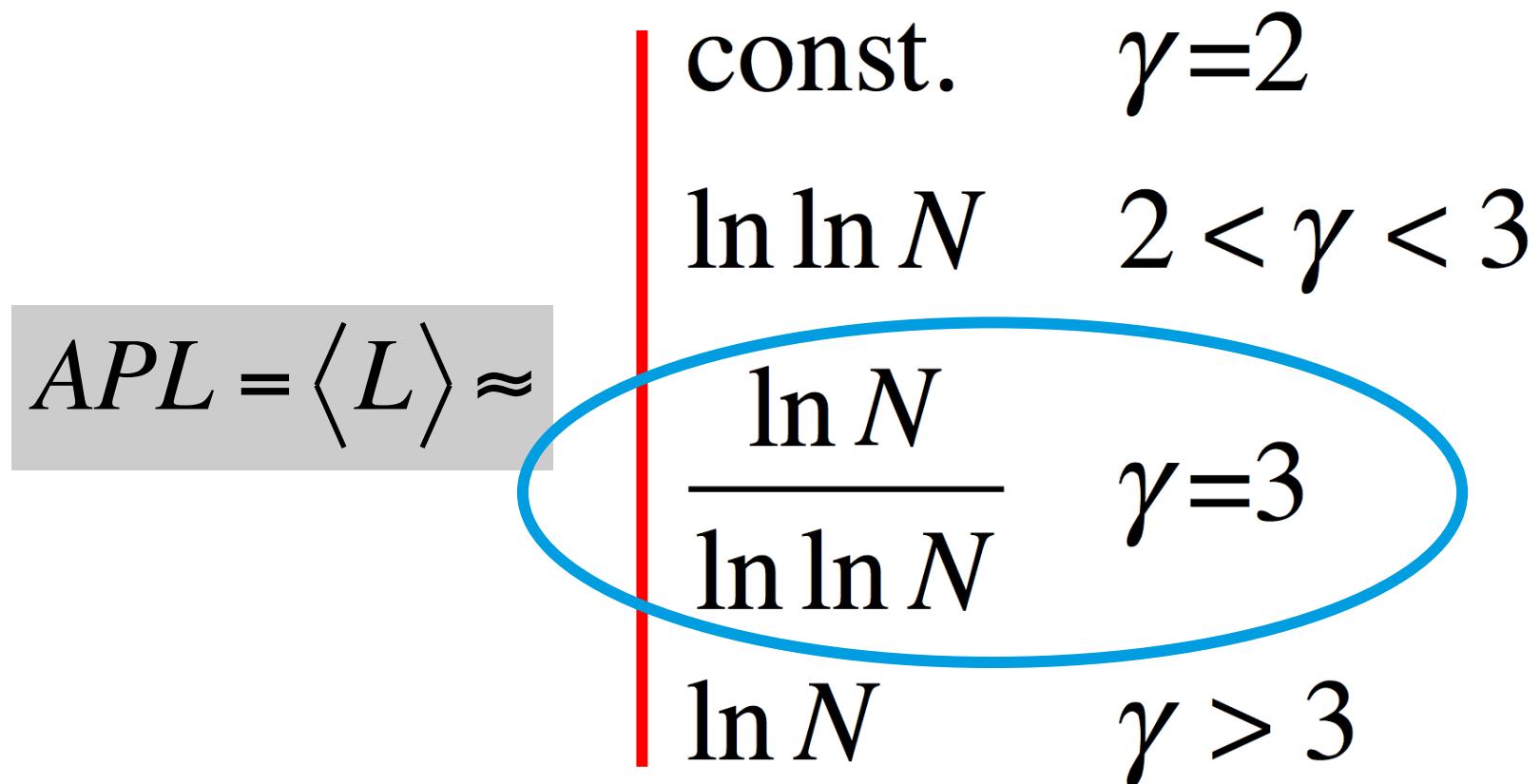
$\gamma > 3$

Ultra-small world regime

$$APL = \langle L \rangle \approx$$

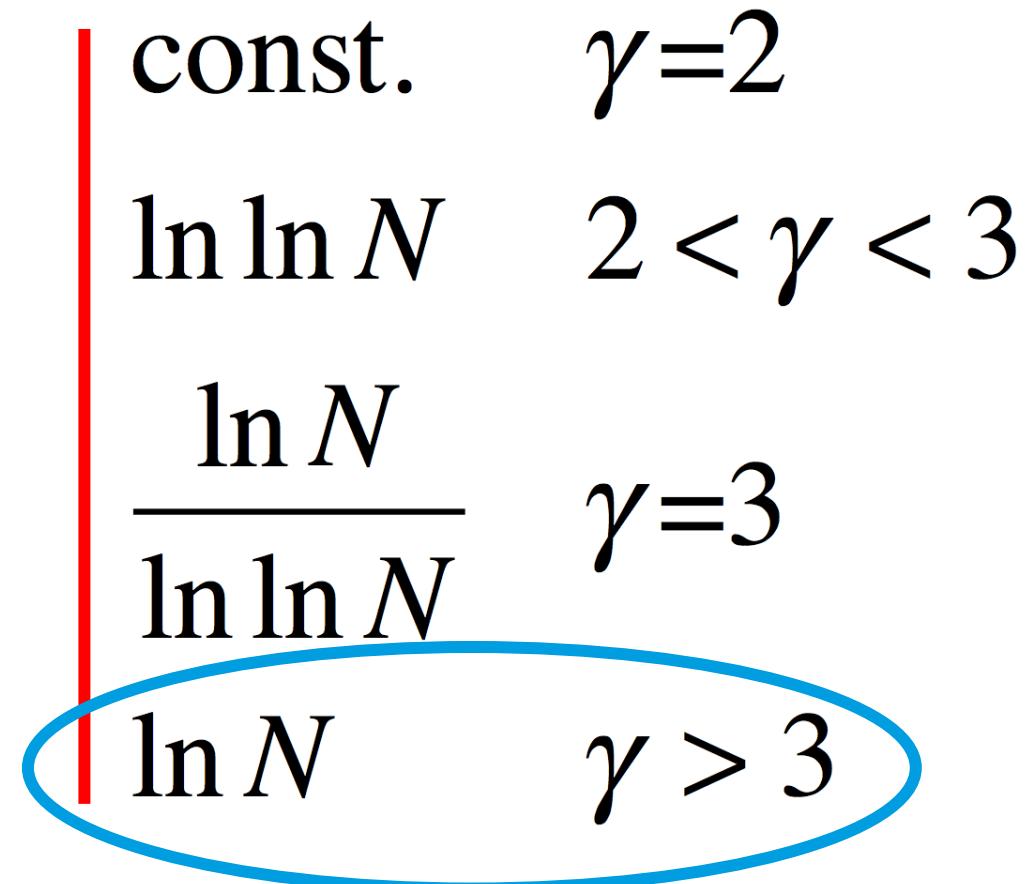


Critical regime

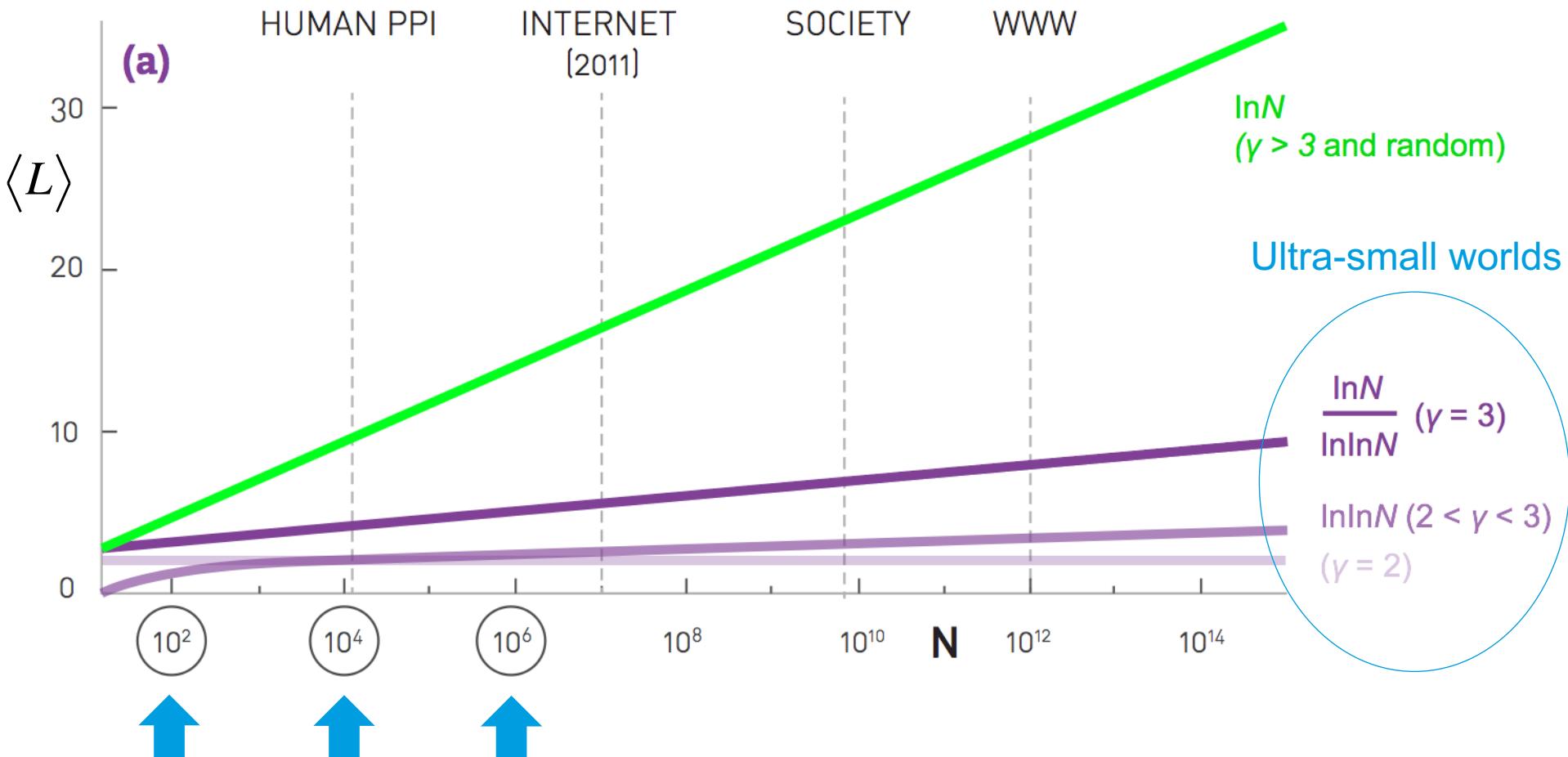


Small-world regime (same as random graphs)

$$APL = \langle L \rangle \approx$$

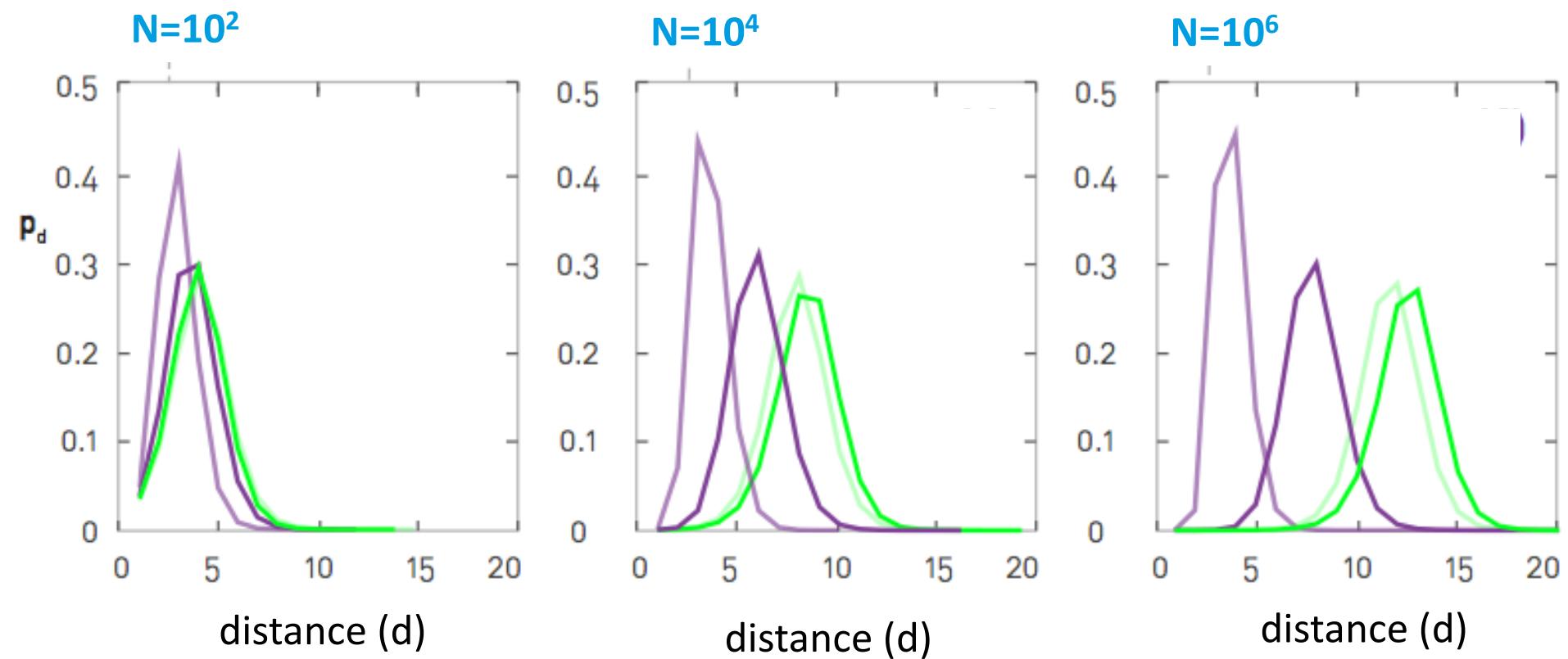


Do we live in a ultra-small-world?

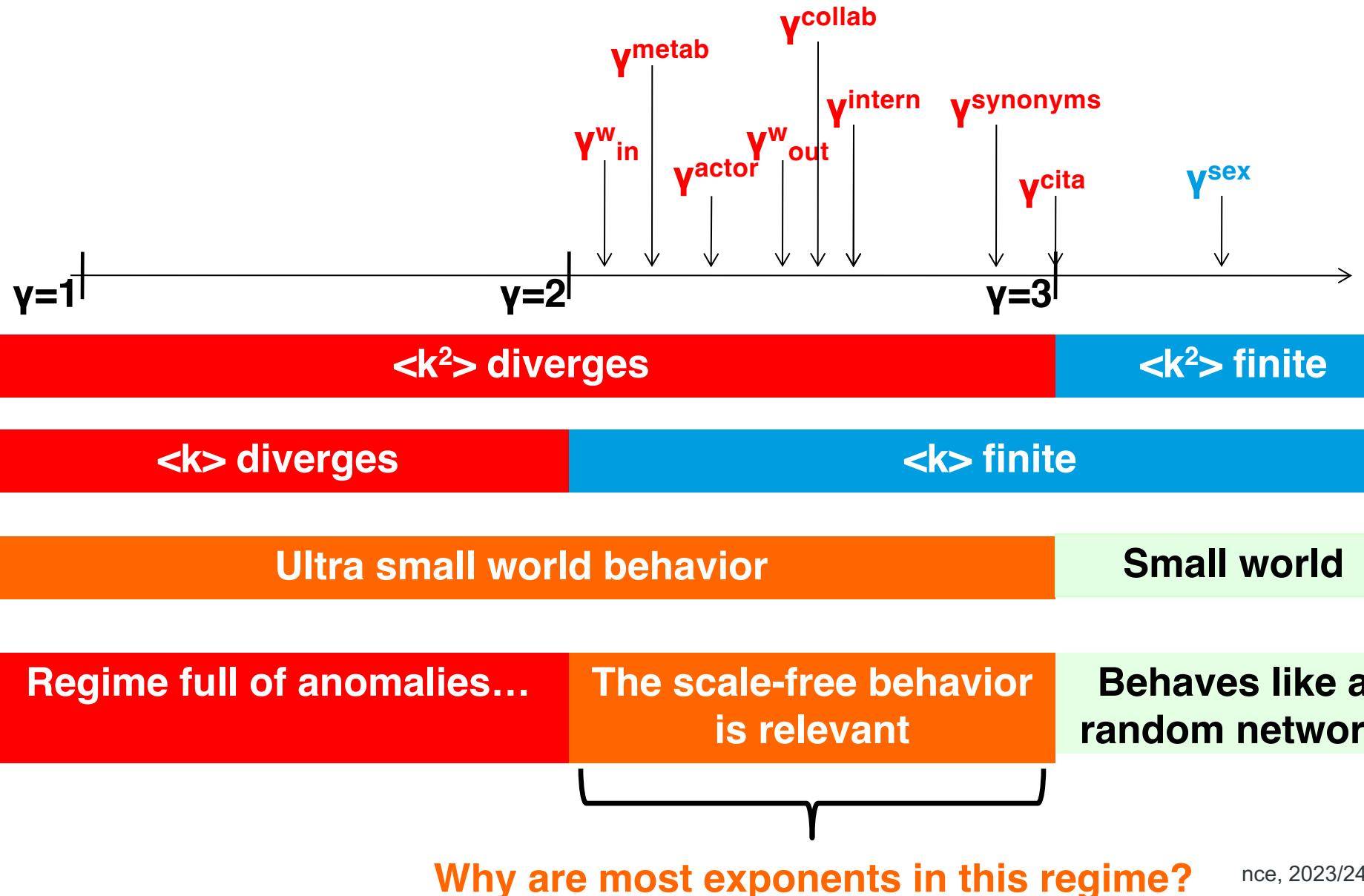


Do we live in a ultra-small-world?

● $\gamma=2.1$ ● $\gamma=3.0$ ● $\gamma=5.0$ ● Random net

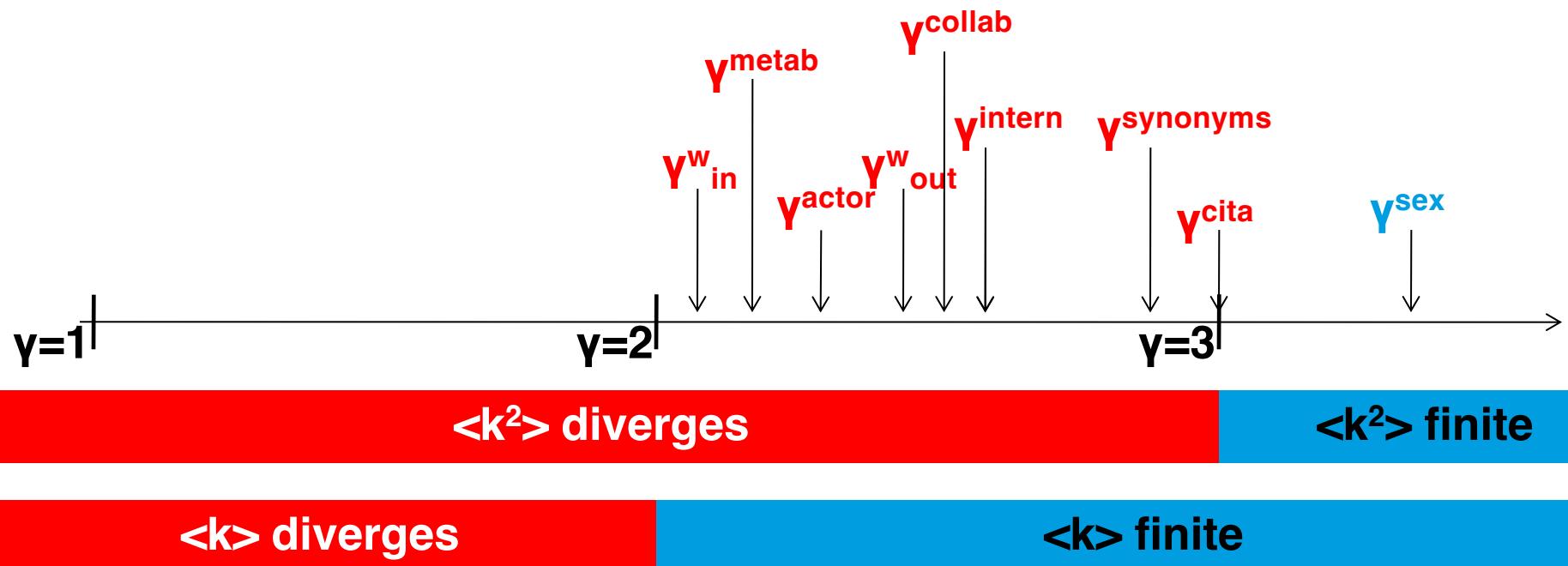


The universe of scale-free networks



A world of magic exponents

Can you find a good argument which justifies this picture?



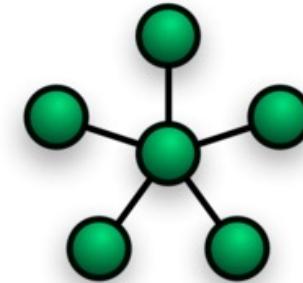
Magic exponents?

- Why is it hard to find networks with $\gamma \leq 2$?

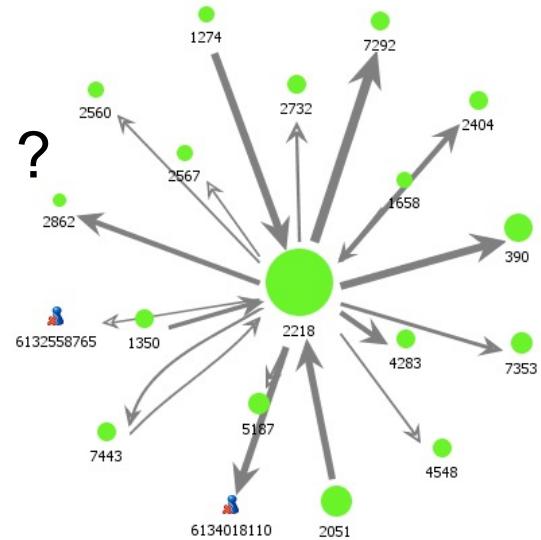
$$k_{\max} = k_{\min} N^{\frac{1}{\gamma-1}}$$



**K_{max} grows
faster than N**



—



- Why is it hard to find networks with $\gamma > 3$?

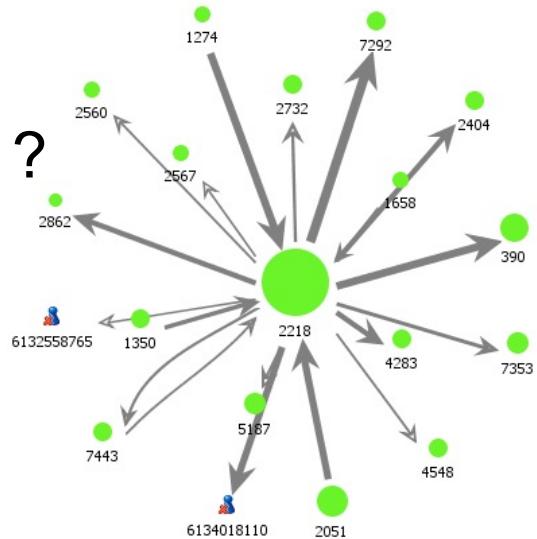
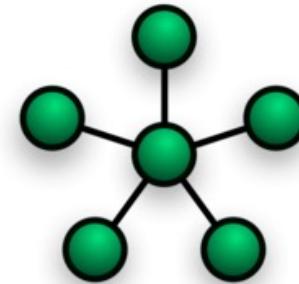
Magic exponents?

- Why is it hard to find networks with $\gamma \leq 2$?

$$k_{\max} = k_{\min} N^{\frac{1}{\gamma-1}}$$



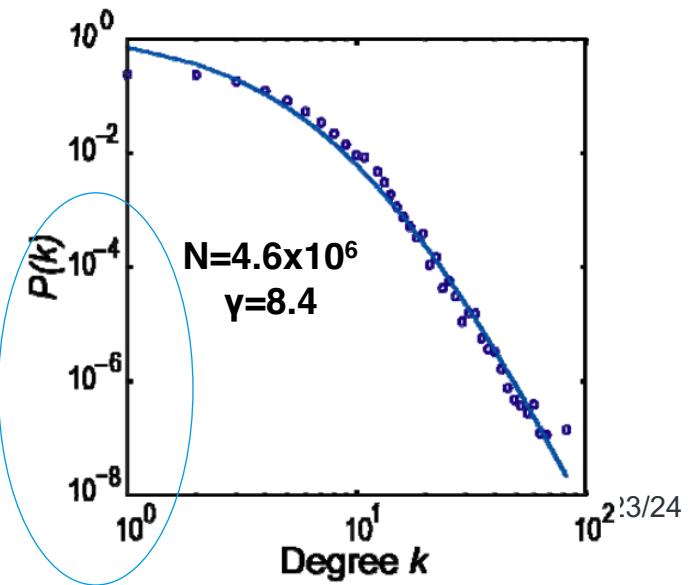
K_{max} grows
faster than N



- Why is it hard to find networks with $\gamma >> 3$?

Mobile Call Network

Onella et al. PNAS 2007



Let's return to one of the challenges discussed in the last class

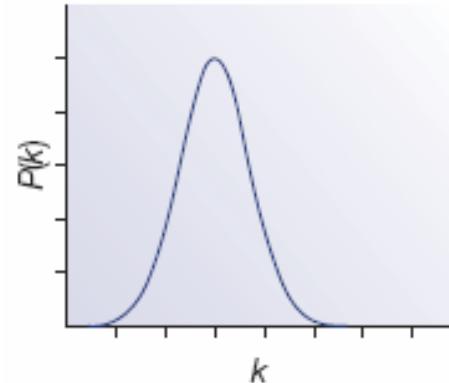
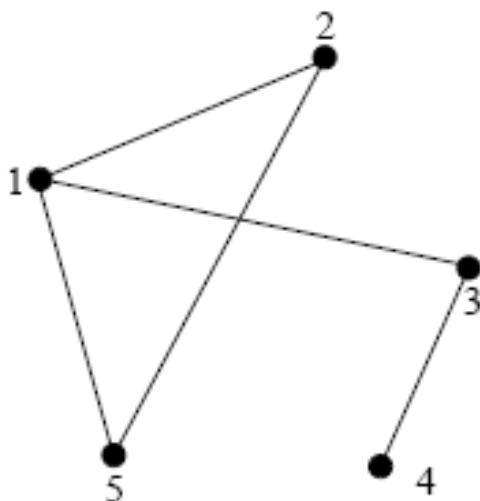
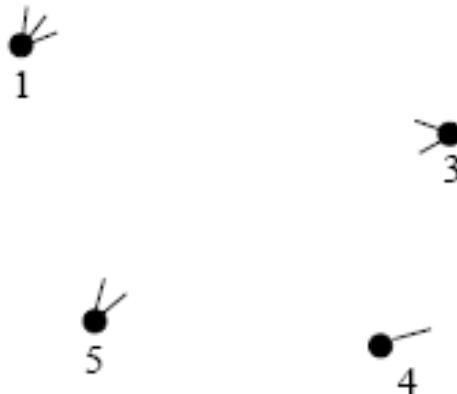
Challenge:

Can you imagine an algorithm capable of creating a random graph with an arbitrary degree distribution?

Configuration model



How to generate a graph compatible with a given $P(k)$



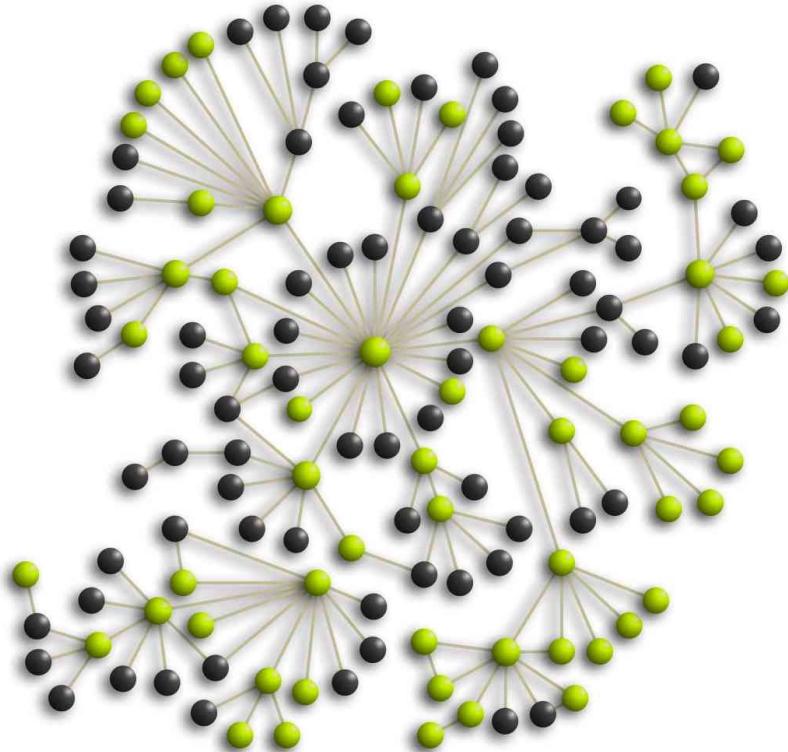
1. Create a histogram of $P(k)$ (discretization) the sum of which gives $N\langle k \rangle$
2. add stubs to n's according to histogram
3. connect the stubs at random.
4. this leads to a random graph with a pre-defined $P(k)$.

2nd challenge of the day

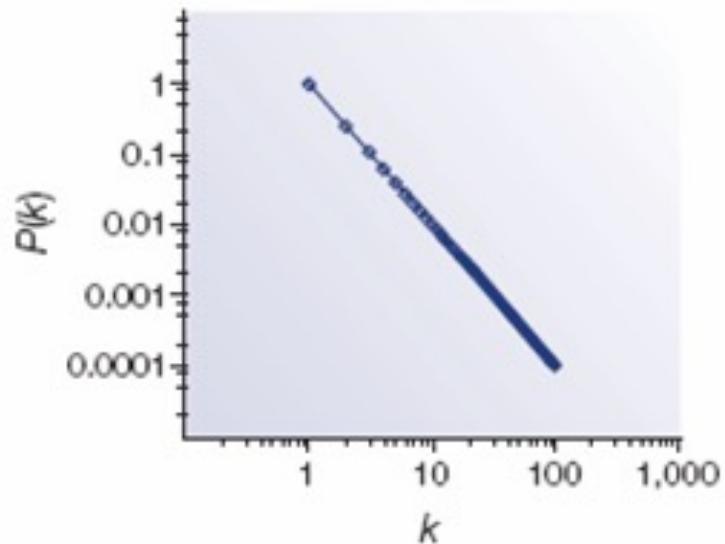
Challenge 5:

Can you create an algorithm capable of increasing its clustering coefficient without changing its degree distribution?

Next challenge: Universality?



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



| WWW | actors | citations | sex | cellular | phones | linguistics |
|----------------|----------------|--------------|----------------|----------------|----------------|----------------|
| $\gamma = 2.1$ | $\gamma = 2.3$ | $\gamma = 3$ | $\gamma = 3.5$ | $\gamma = 2.1$ | $\gamma = 2.1$ | $\gamma = 2.8$ |