Digital epidemiology
Lesson 8

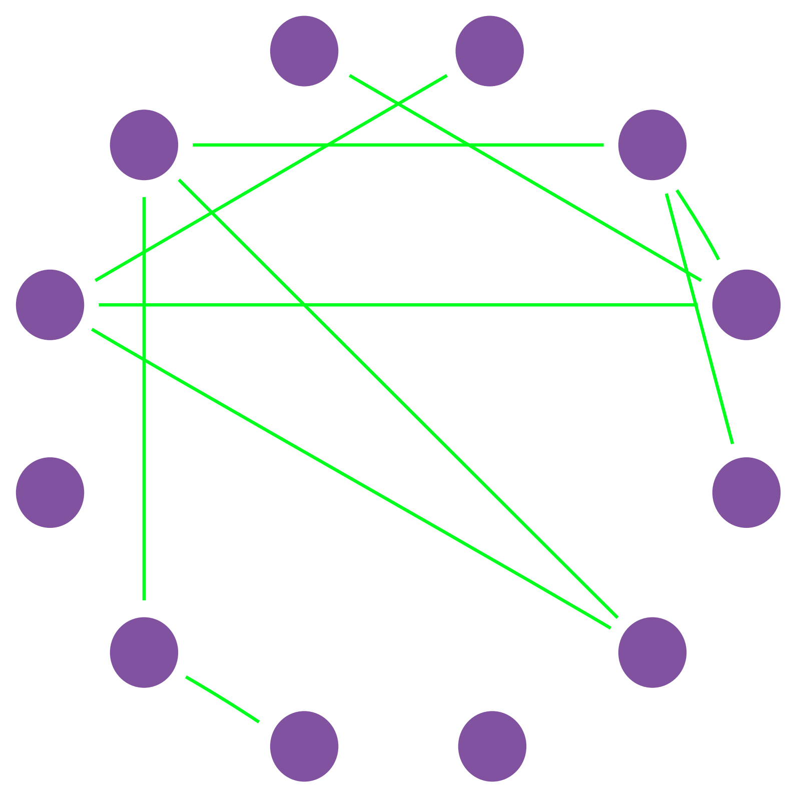
Michele Tizzoni
Dipartimento di Sociologia e Ricerca Sociale
Via Verdi 26, Trento
Ufficio 6, 3 piano

Network theory: models

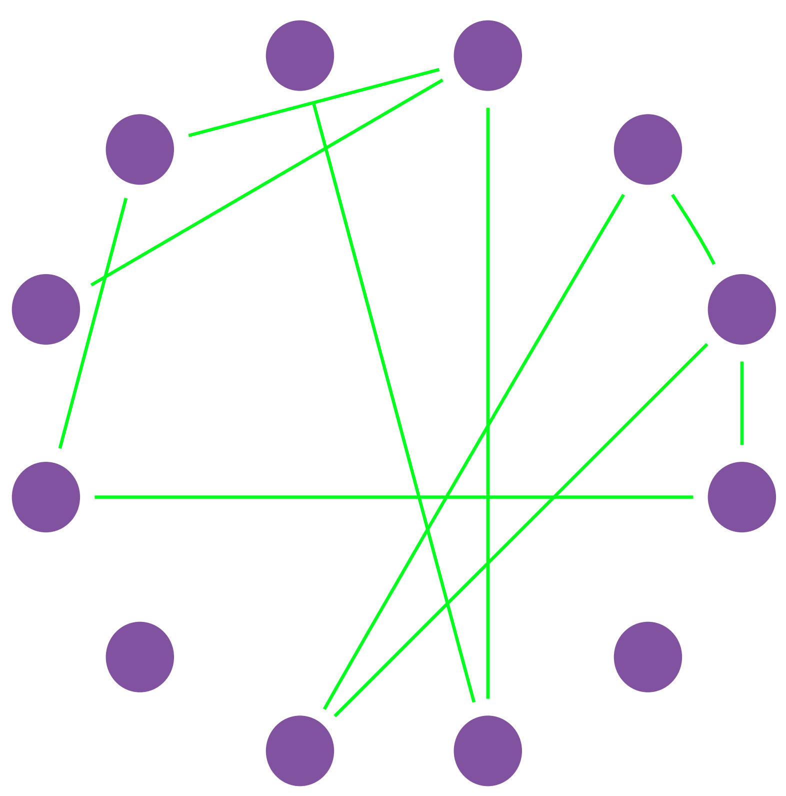
Erdős-Renyi random model

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

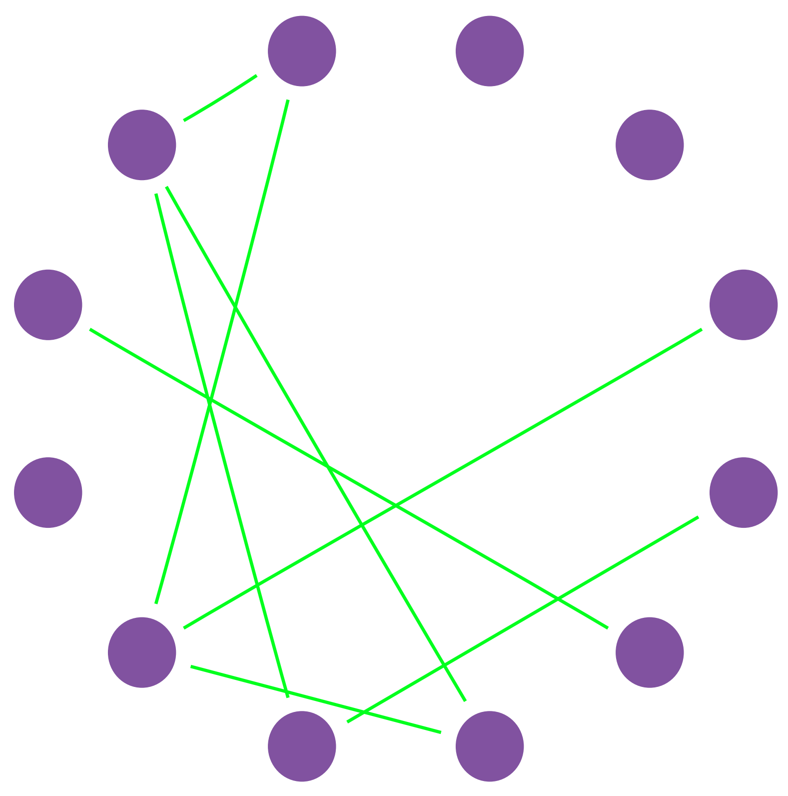
N labeled nodes are connected with *L* randomly placed links. Erdős and Rényi used this definition in their string of papers on random networks [2-9].



L=8



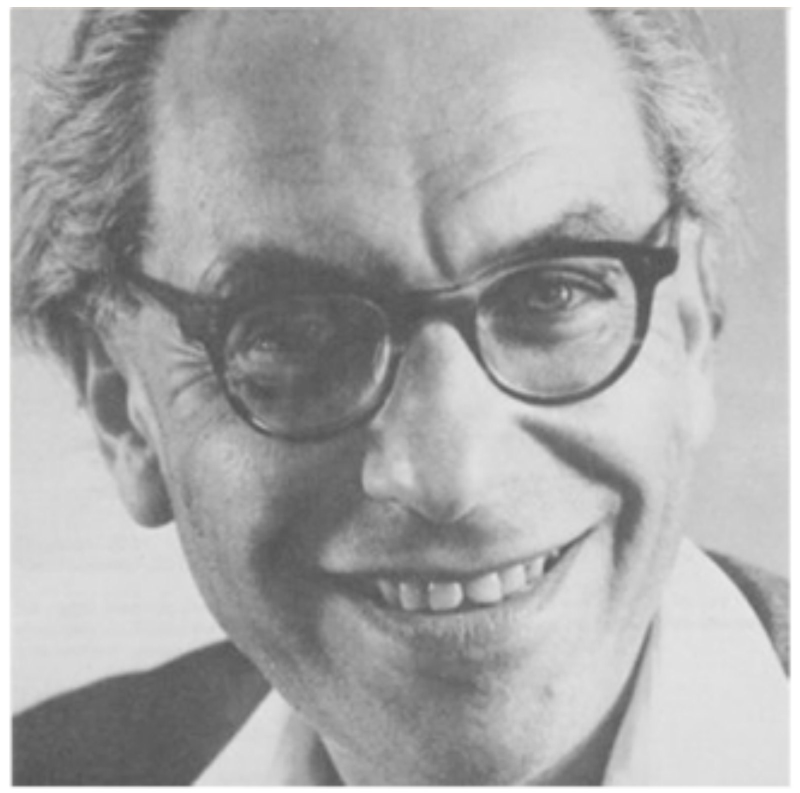
L=10



L=7

$$N = 12$$

Pál Erdös
(1913-1996)



Alfréd Rényi
(1921-1970)

Erdös-Rényi model (1960)

Erdős-Renyi random model

Probability to have exactly *L* links in a network of *N* nodes and probability *p* of placing a link:

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

- Probability that we have L links with L successful attempts: p^L
- Probability that the remaining attempts didn't create a link: $(1-p)^{N(N-1)/2 - L}$
- The binomial coefficient counting the number of different ways we can place L links among N nodes

Degree distribution

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

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

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

We are constraining the average degree!
So if we want SPARSENESS, we need small p

Degree distribution

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

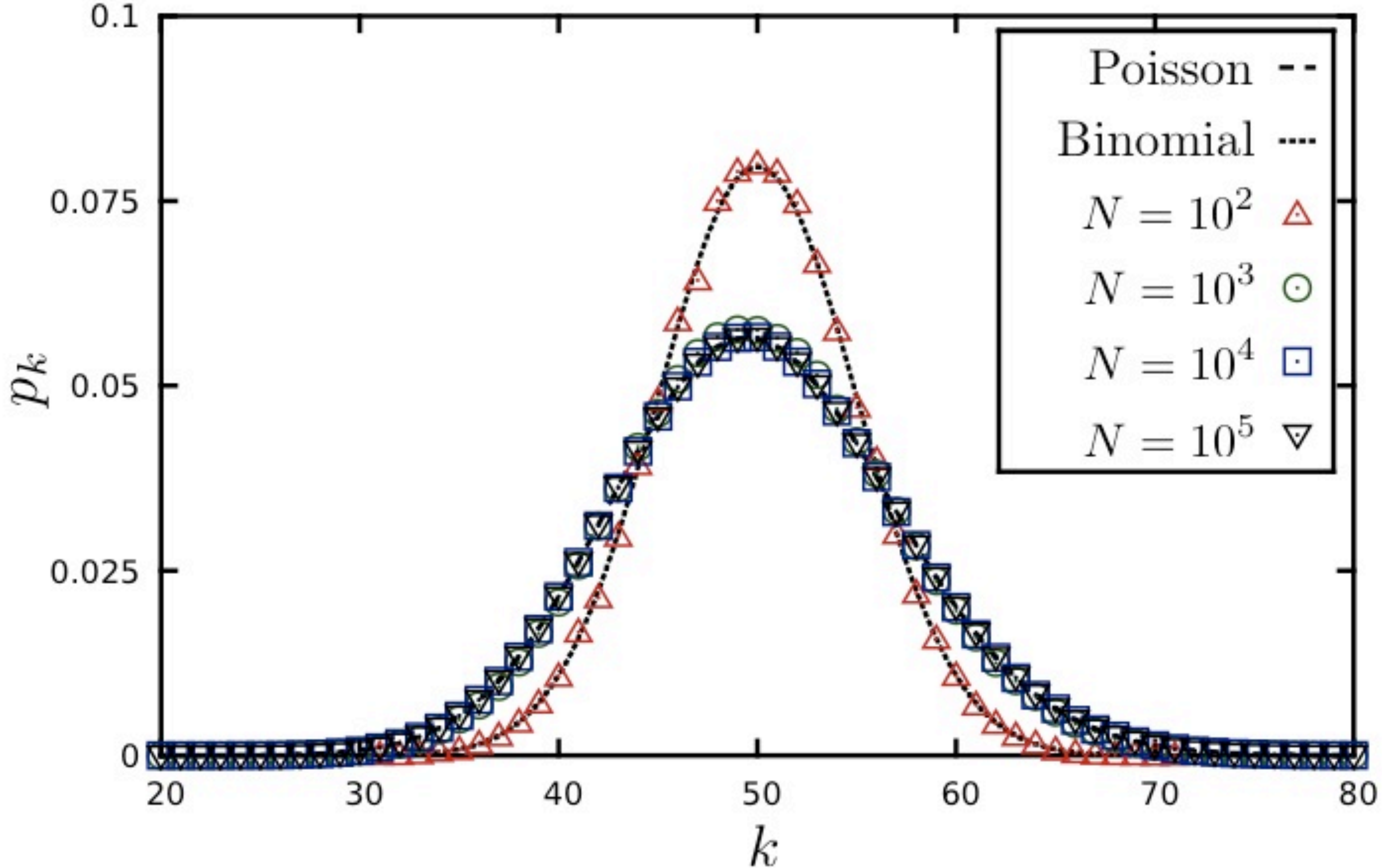
- Probability that we have degree k with k successful attempts: p^k
- Probability that the remaining attempts didn't create a link: $(1\text{-}p)^{N\text{-}1\text{-}k}$
- The binomial coefficient counting the number of different ways we can place k links among the N-1 nodes

Degree distribution

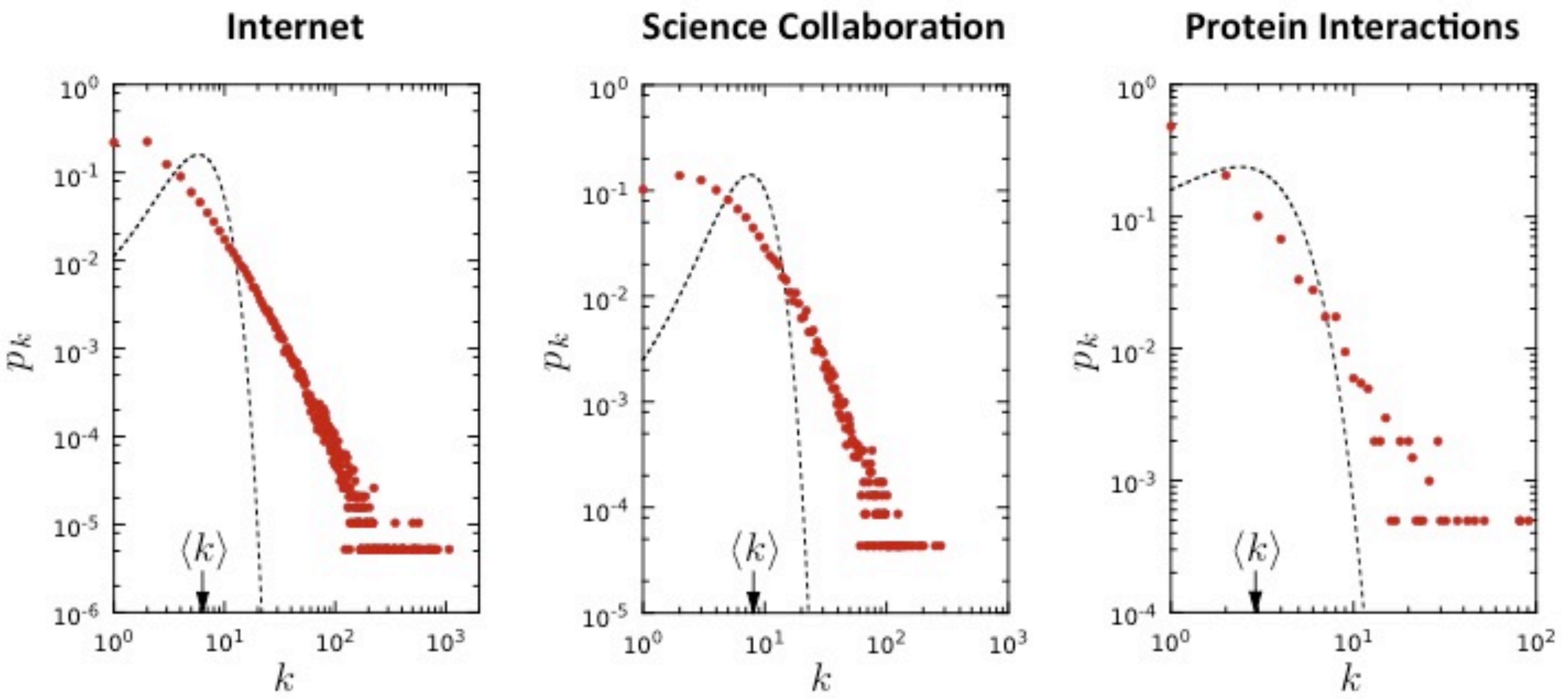
$$p(k) = \binom{N-1}{k} p^k (1-p)^{(N-1)-k}$$

For large N, and $\langle k \rangle \ll N$

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



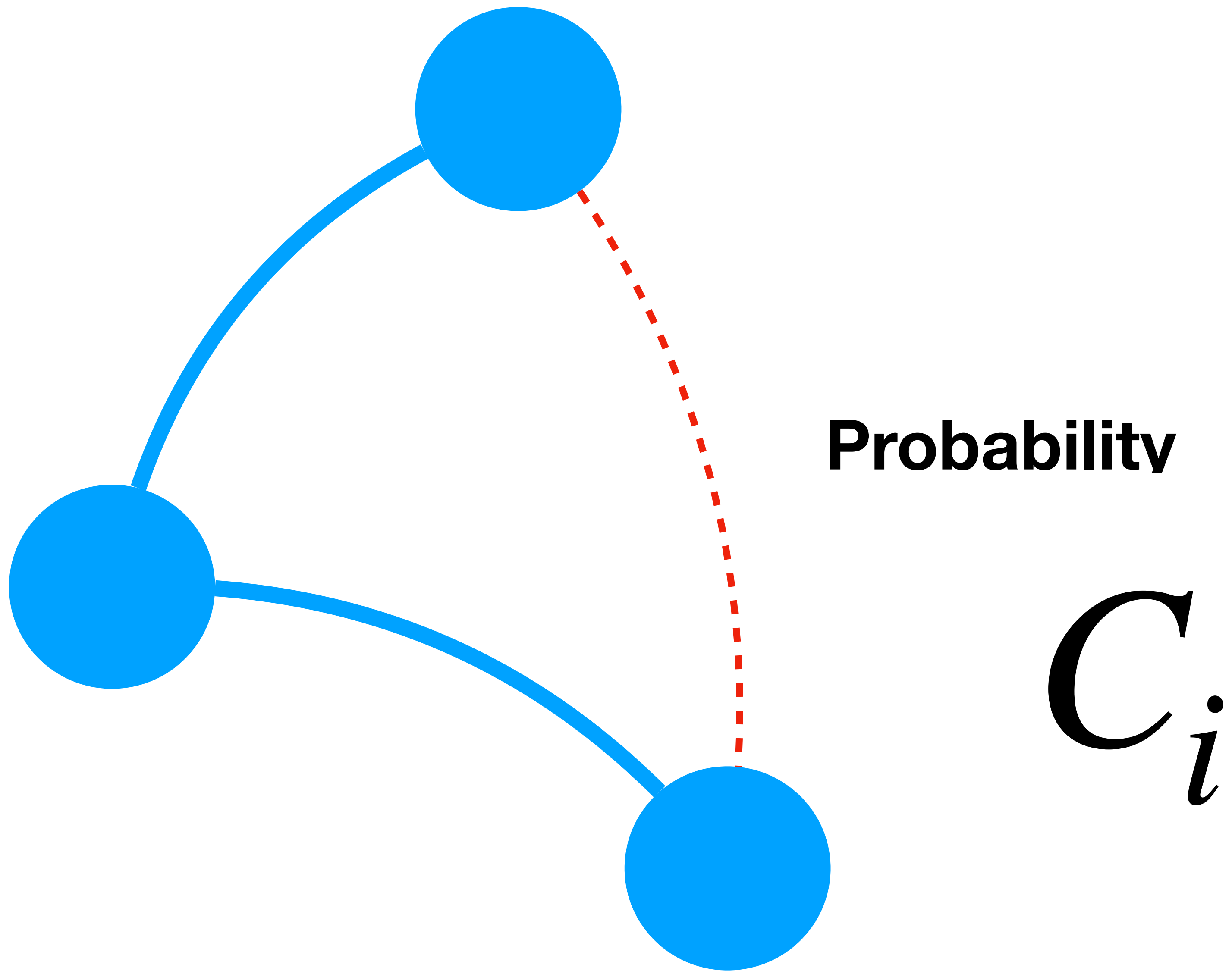
Degree distribution



Clustering

What about clustering?

$$C_i = \frac{2L_i}{k_i(k_i - 1)}$$



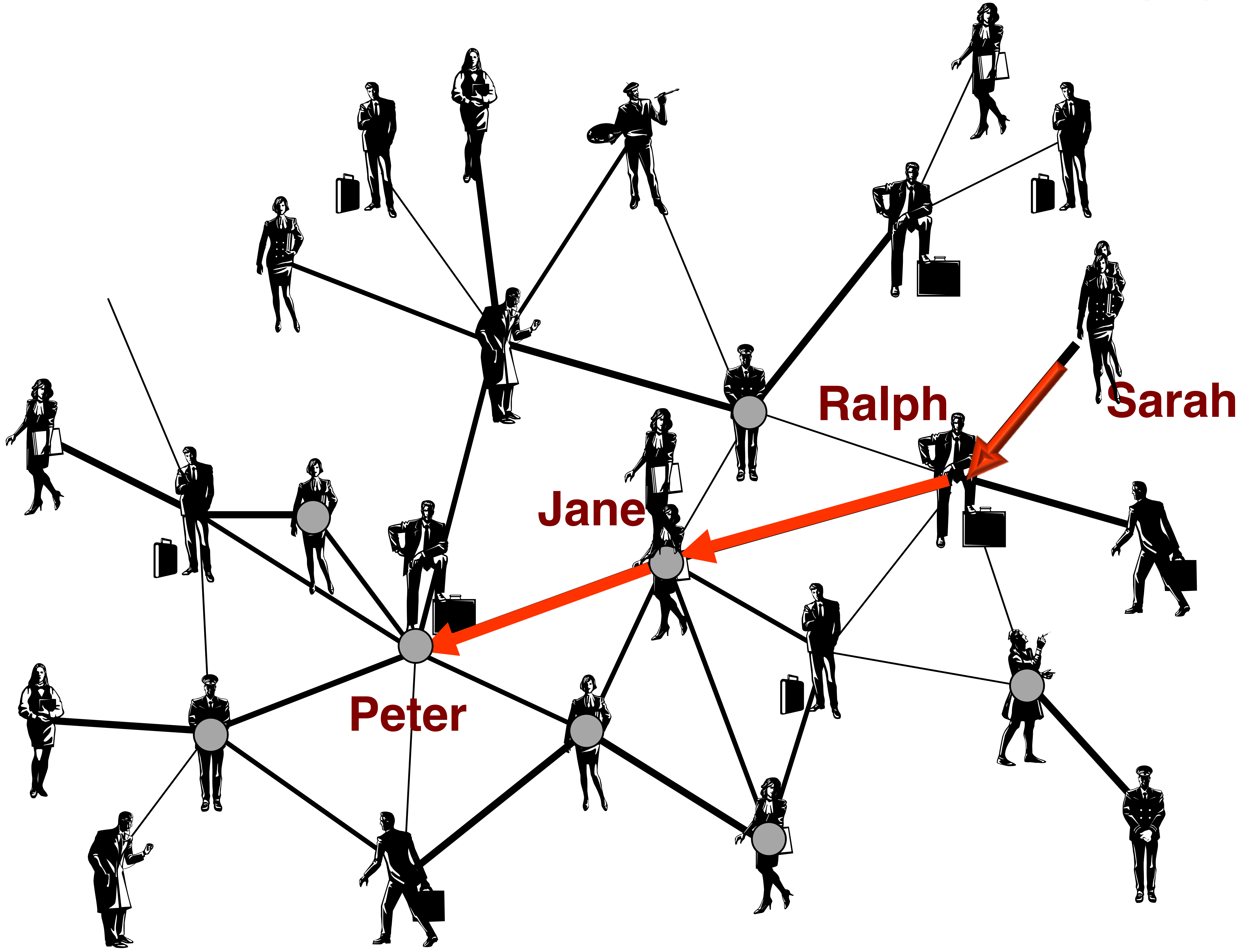
$$C_i = p$$

We CAN constrain the clustering (but uniform)!
So if we want high clustering, we need large p!

We are constraining the average degree!
So if we want SPARSENESS, we need small p

Distances

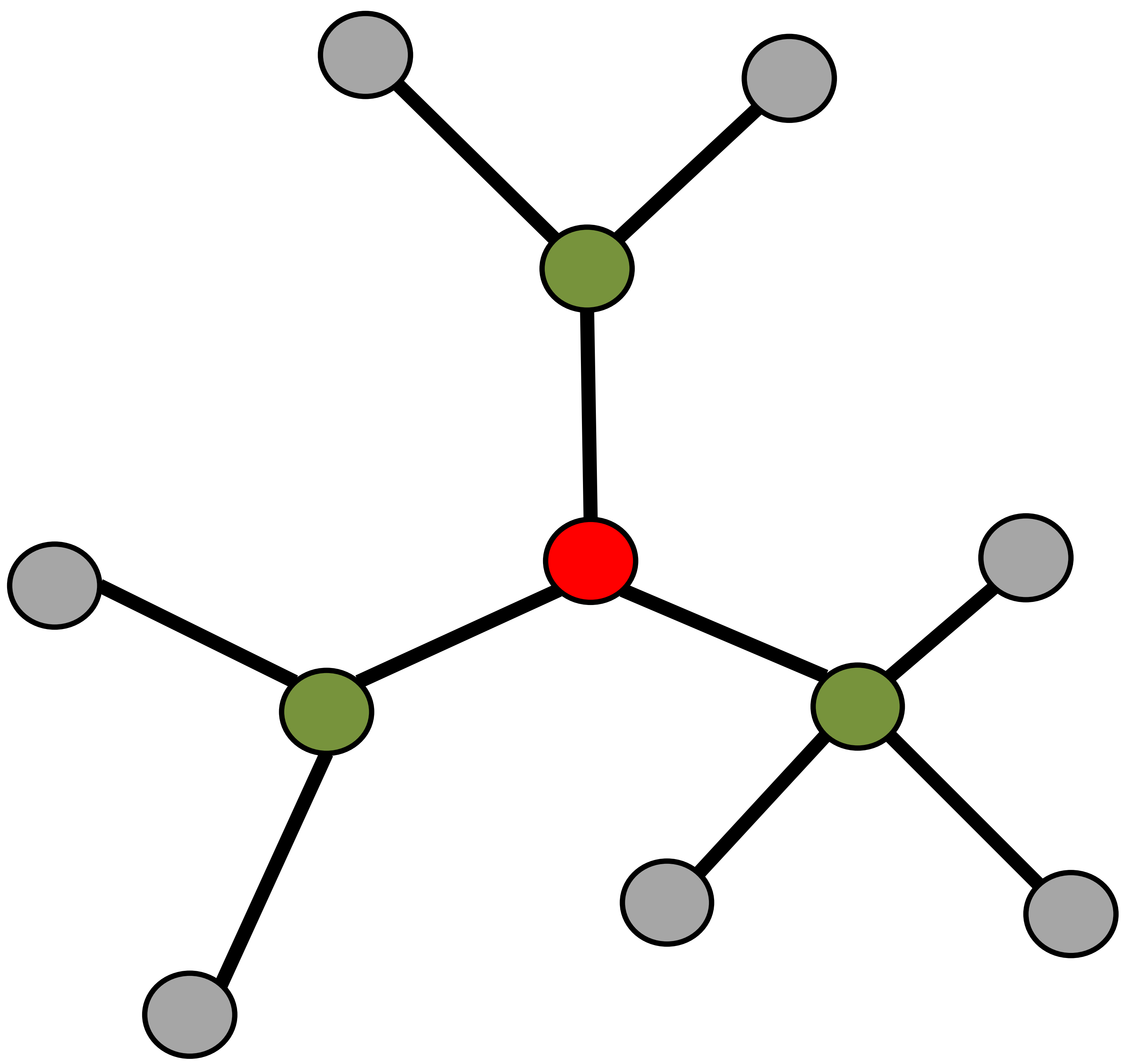
What about distances?



Frigyes Karinthy, 1929
Stanley Milgram, 1967

Distances

Let's try an easy case



<k> nodes at distance d=1

<k>^{2} nodes at distance d=2

<k>^{3} nodes at distance d=3

...

$$1 + \langle k \rangle + \langle k \rangle^2 + \langle k \rangle^3 + \ldots + \langle k \rangle^{d_{max}} = N$$

$$\frac{\langle k \rangle^{d_{max}+1} - 1}{\langle k \rangle - 1} = N \quad \rightarrow \quad d_{max} \simeq \frac{\log N}{\log \langle k \rangle}$$

Wrong! This is actually closer to the average distance!

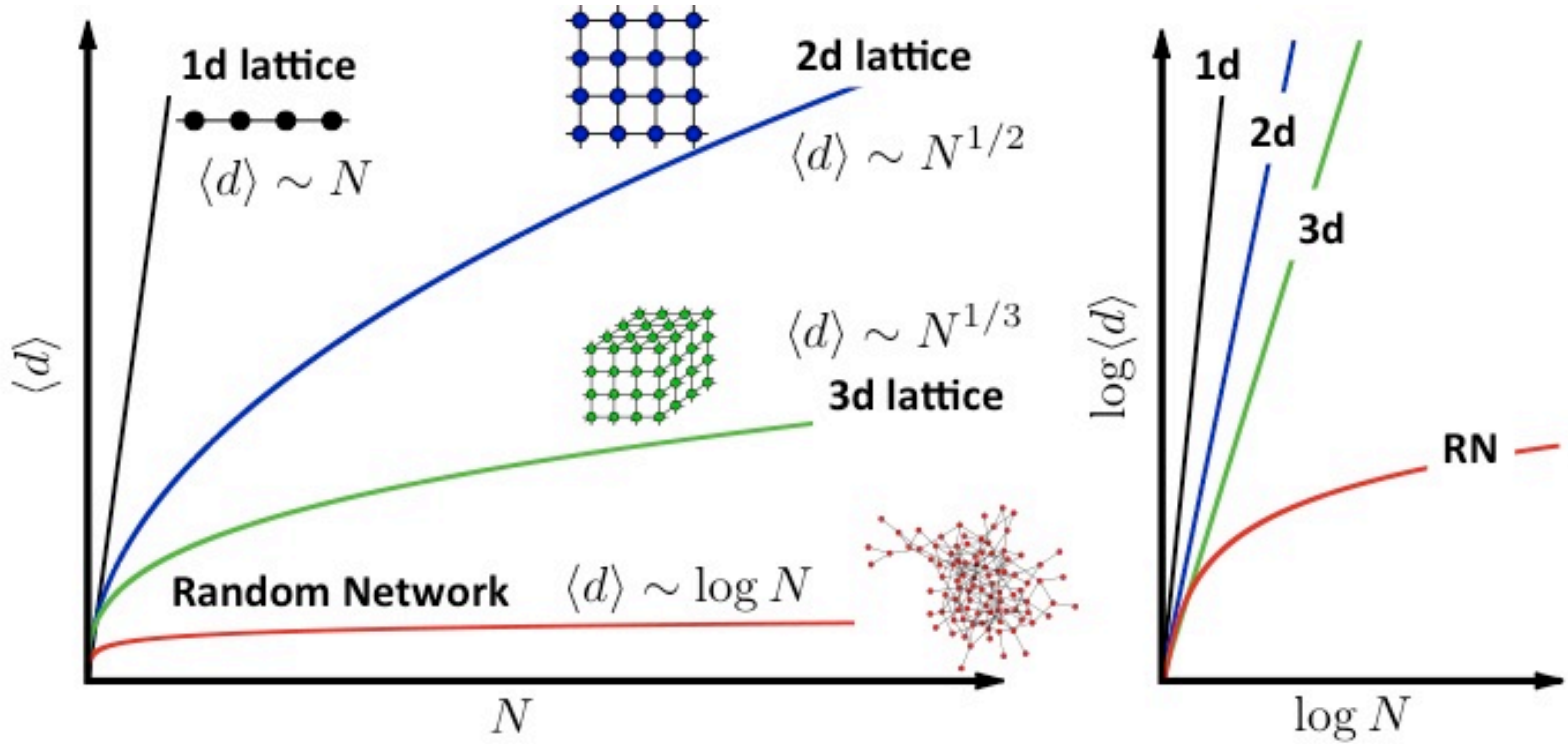
$$\langle d \rangle \simeq \frac{\log N}{\log \langle k \rangle}$$

Erdős-Renyi random model

List of main results:

- We can reproduce sparseness using N and *p*
- Degree distribution is Poissonian and not power-law/broad
- We can reproduce high-clustering but not low density, or viceversa
- Small world property emerges naturally

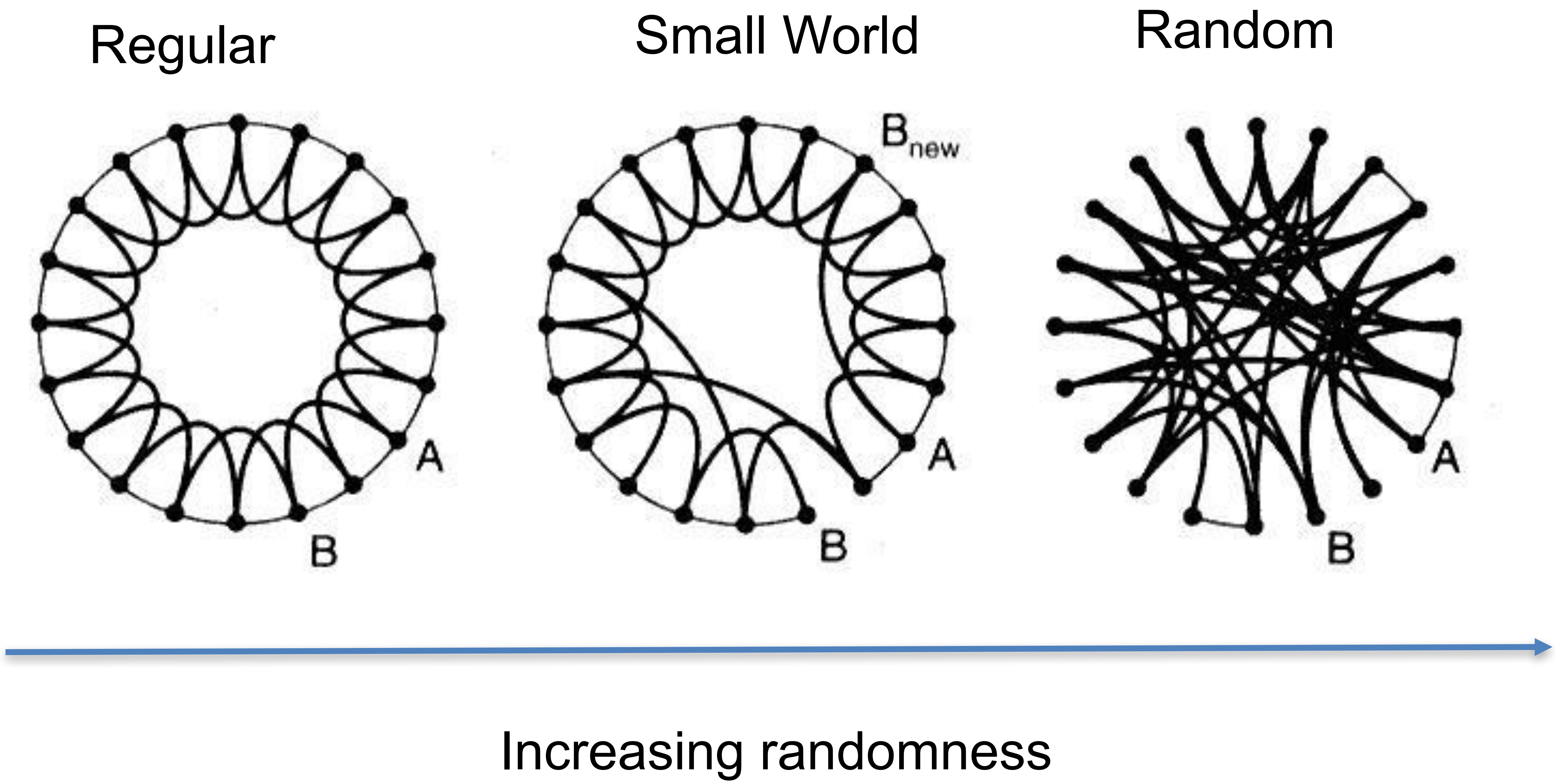
Distances

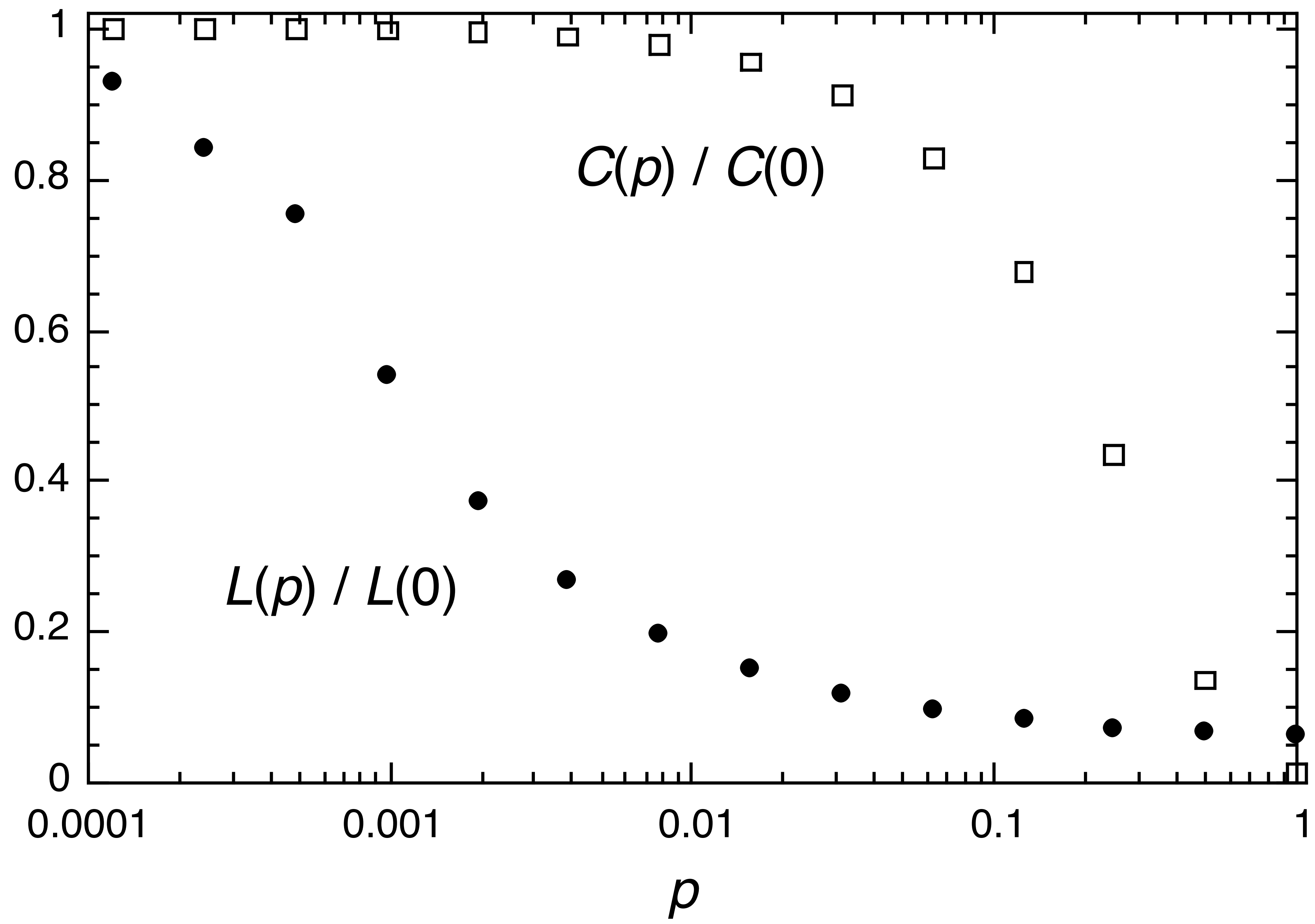


1d lattice
$\langle d \rangle \sim N$
2d lattice
$\langle d \rangle \sim N^{1/2}$
$\langle d \rangle \sim N^{1/3}$
3d lattice
Random Network
$\langle d \rangle \sim \log N$
$\langle d \rangle$
N
1d
2d
3d
RN
$\log \langle d \rangle$
$\log N$

Watts-Strogatz random model

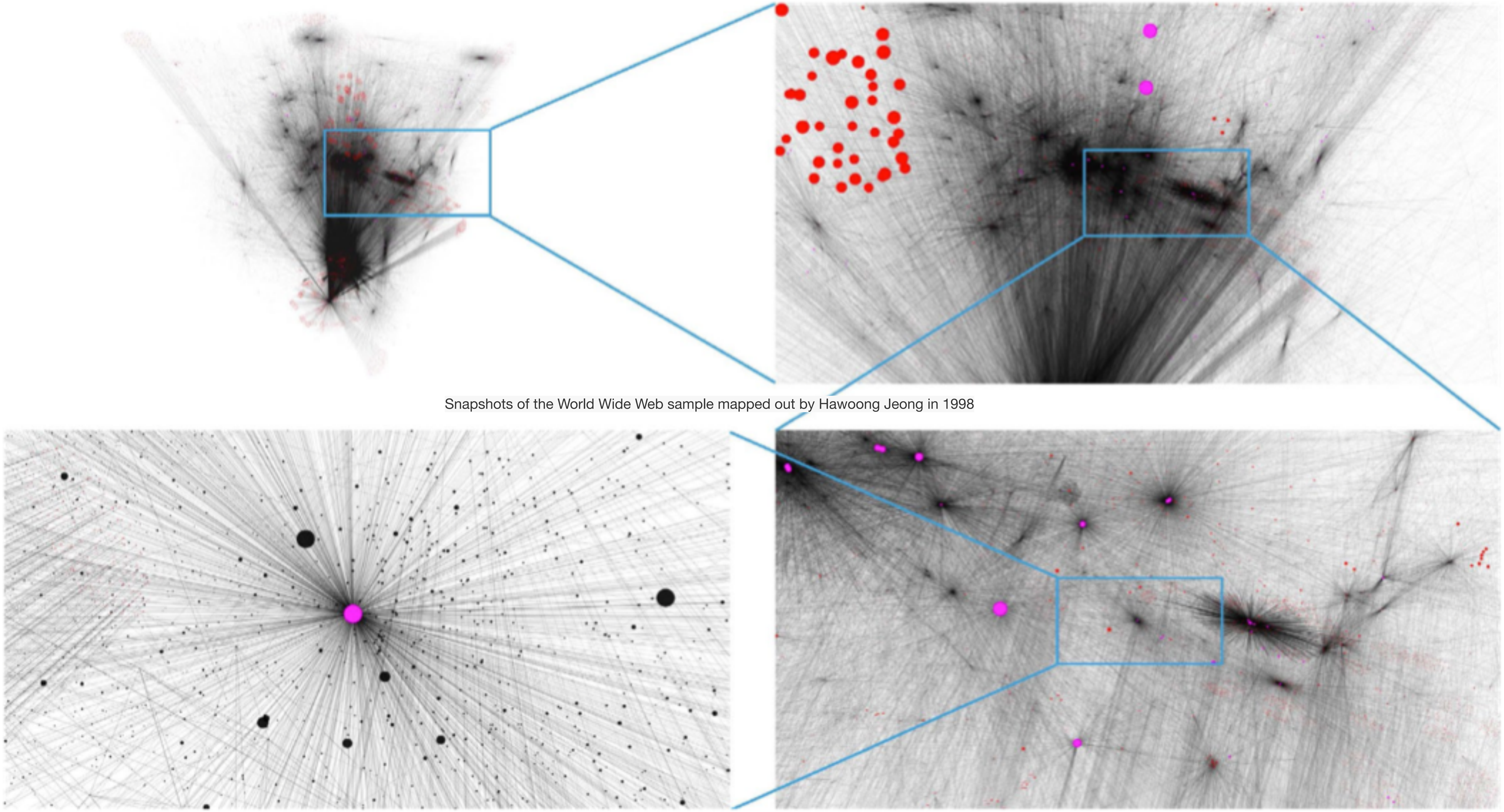
Can we reconcile small world and high clustering in a single model? Yes





Watts, Duncan J., and Steven H. Strogatz. "Collective dynamics of 'small-world' networks." nature 393.6684 (1998): 440-442.

The meaning of scale-free



Snapshots of the World Wide Web sample mapped out by Hawoong Jeong in 1998

The meaning of scale-free

A scale-free network is a network whose degree distribution follows a power law.

Discrete formalism

$$p_k = Ck^{-\gamma}$$

$$\sum_{k=1}^{\infty} p_k = 1$$

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

$$p_k = \frac{k^{-\gamma}}{\zeta(\gamma)}$$

Continuous formalism

$$p(k) = Ck^{-\gamma}$$

$$\int_{k_{min}}^{\infty} p(k)dk = 1$$

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

$$p(k) = (\gamma - 1)k_{min}^{\gamma-1}k^{-\gamma}$$

Scale-free networks

- Real world networks can be very large but they are finite.
- It exists a **maximum degree** in each finite network.

$$\int_{k_{max}}^{\infty} p(k)dk \simeq \frac{1}{N}$$

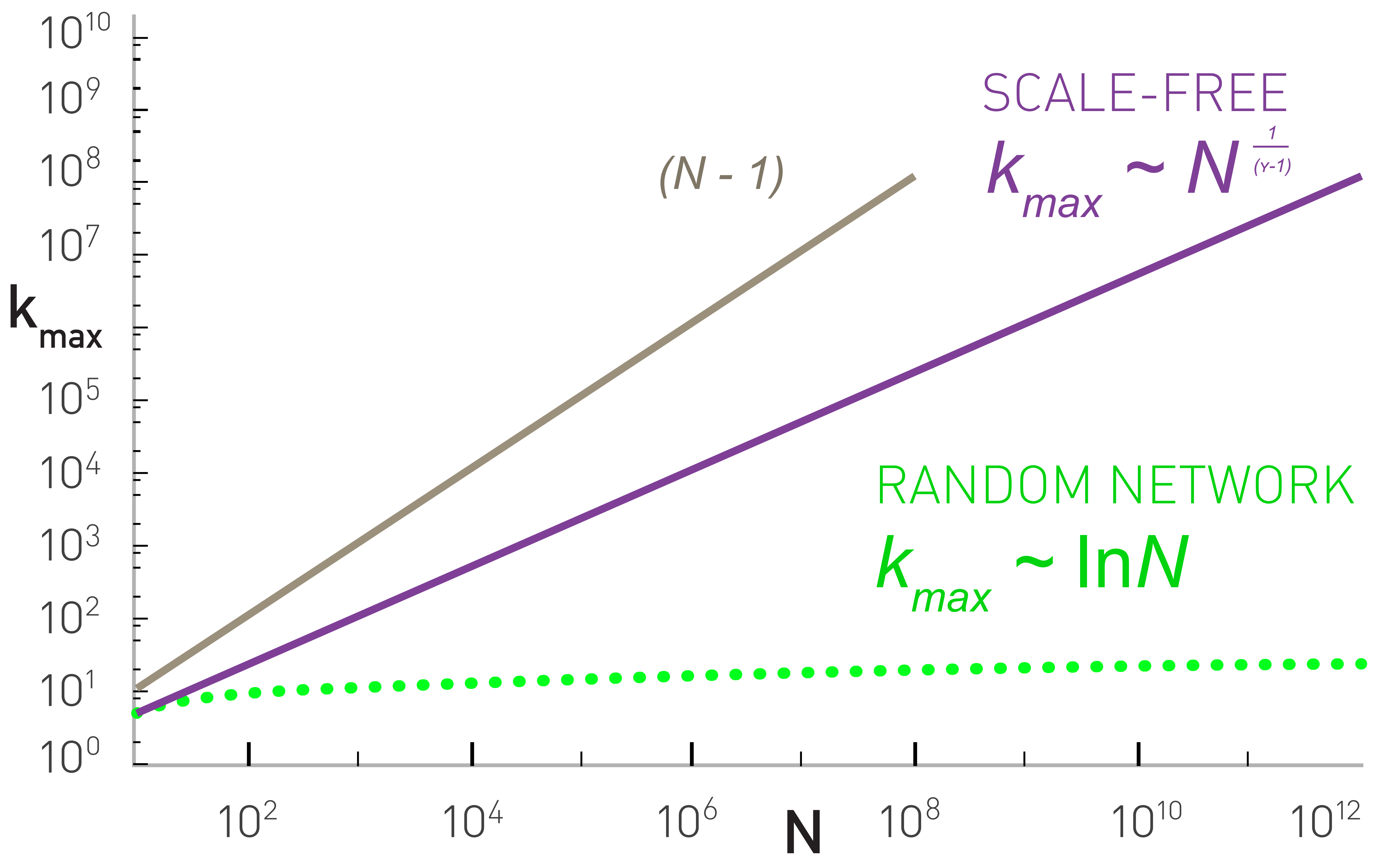
We **assume** there is only one node with degree max or above.

$$\int_{k_{max}}^{\infty} p(k)dk = (\gamma - 1)k_{min}^{\gamma-1}\int_{k_{max}}^{\infty} k^{-\gamma}dk = \frac{\gamma-1}{-\gamma+1}k_{min}^{\gamma-1}[k^{-\gamma+1}]_{k_{max}}^{\infty} = \frac{k_{min}^{\gamma-1}}{k_{max}^{\gamma-1}} \simeq \frac{1}{N}$$

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

Scale-free networks

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



- k_{max} increases with the size of the network ==> bigger system, bigger hub
- For γ>2, k_{max} increases slower than N ==> decreasing fraction of links as N increases.
- For γ=2, k_{max}~N ==> The size of the biggest hub is O(N)
- For γ<2 k_{max} increases faster than N: condensation phenomena ==> the largest hub will grab an increasing fraction of links.

More divergences

$$< k^m >= \int_{k_{min}}^{\infty} k^m p(k) dk \qquad p(k) = (\gamma - 1) k_{min}^{\gamma-1} k^{-\gamma}$$

$$< k^m >= (\gamma - 1) k_{min}^{\gamma-1} \int_{k_{min}}^{\infty} k^{m-\gamma} dk = \frac{\gamma - 1}{m - \gamma + 1} k_{min}^{\gamma-1} \left[k^{m-\gamma+1} \right]_{k_{min}}^{\infty}$$

if $m - \gamma + 1 < 0$: $\qquad < k^m >= \frac{\gamma - 1}{m - \gamma + 1} k_{min}^m$

if $m - \gamma + 1 > 0$: $\qquad < k^m > \to \infty$

This implies:

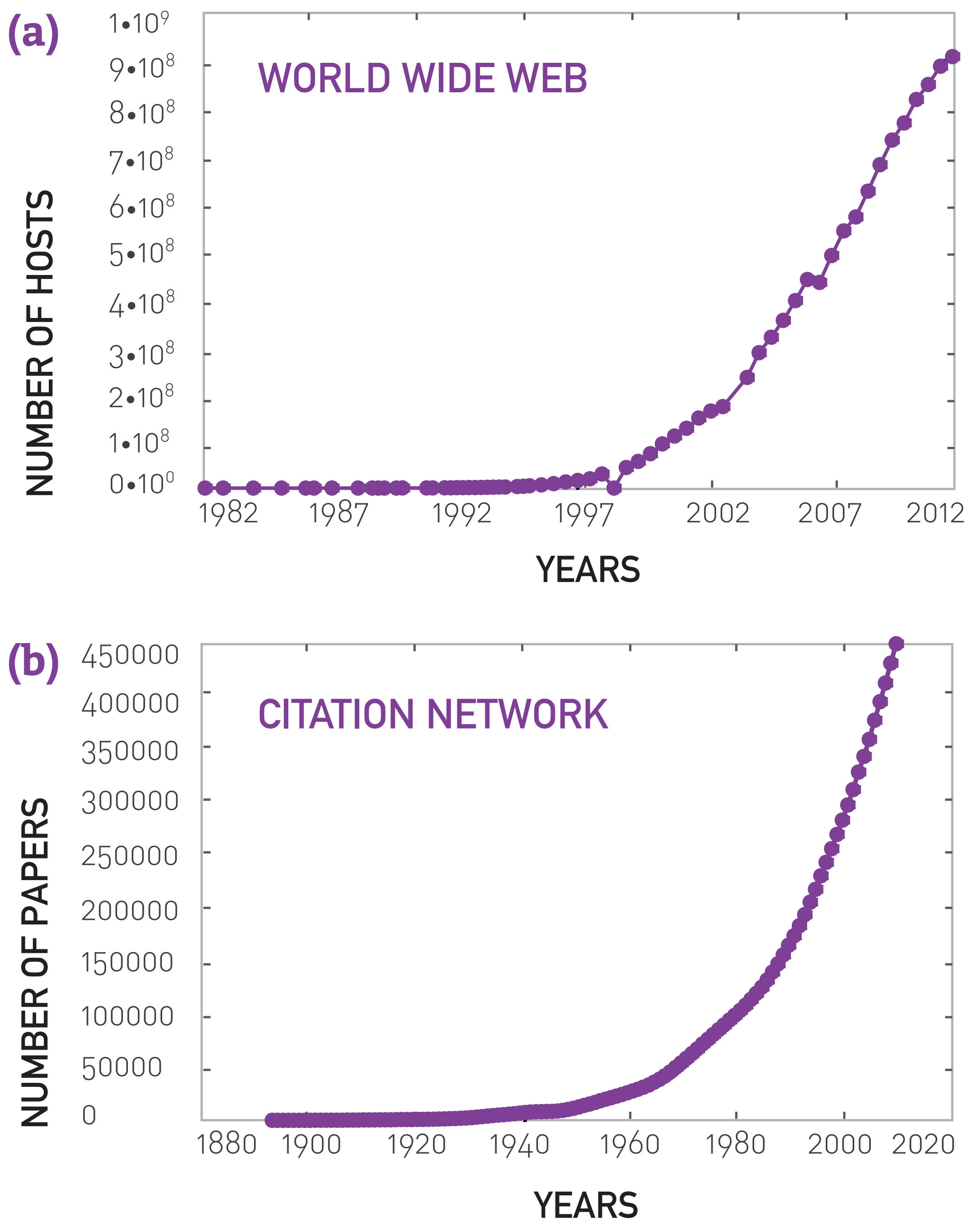
For $\gamma < 3, \quad < k^2 > \to \infty$

Barabási-Albert model

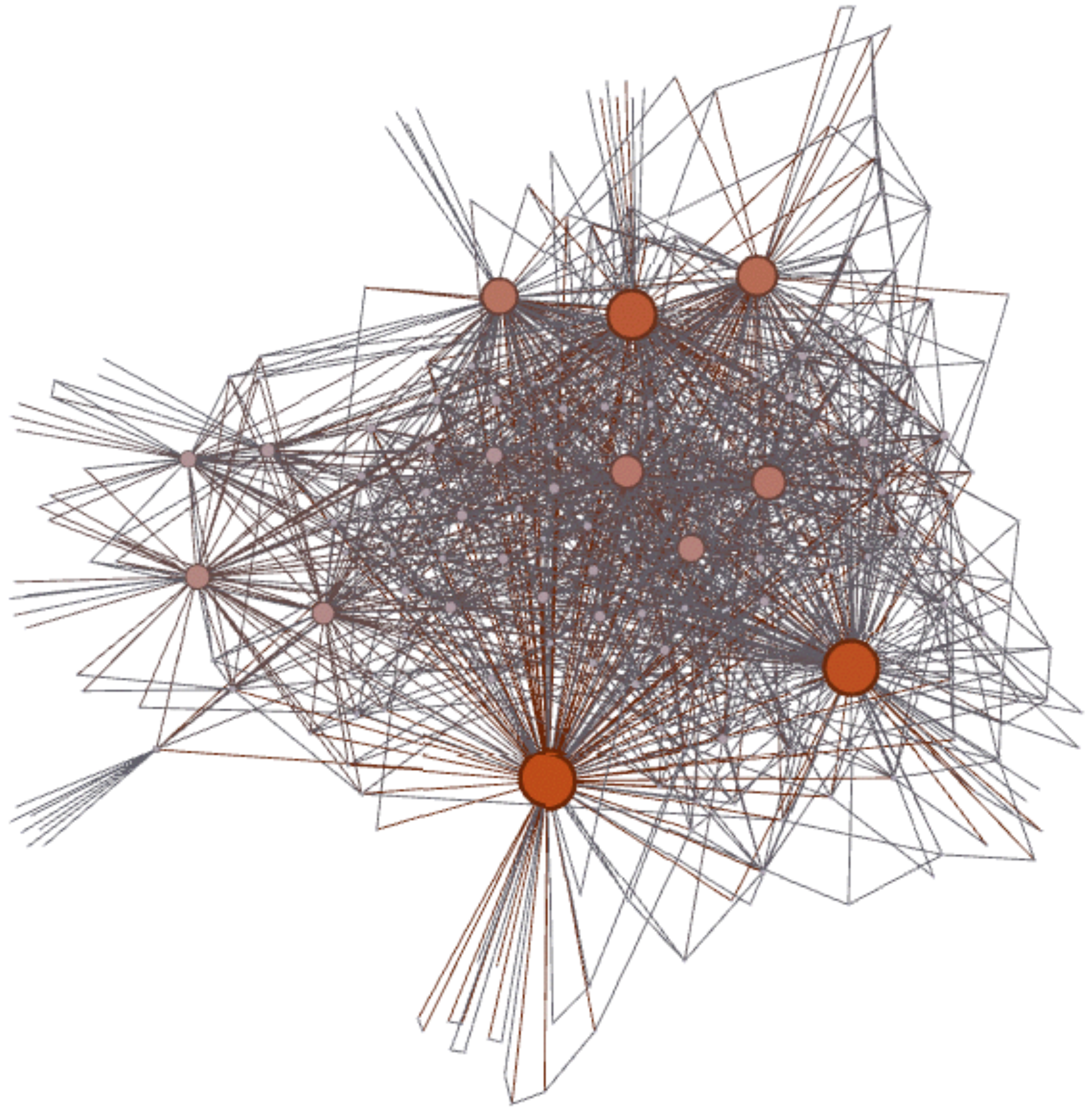
- In static ensemble models we defined network by constraints.
- In evolving/growing network, we **define growth rules** and look for asymptotic stationary behaviour

Barabási-Albert model

First ingredient:
Growth/time



Second ingredient:
Not all links are equally likely!



GROWTH:

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

PREFERENTIAL ATTACHMENT:

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

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}$$

BA degree dynamics

$$\frac{dk_i}{dt} = m\Pi(k_i) = m\frac{k_i}{\sum_{j=1}^{N-1} k_j} \qquad \sum_{j=1}^{N-1} k_j = 2mt - m \overset{t\gg1}{\approx} 2mt$$

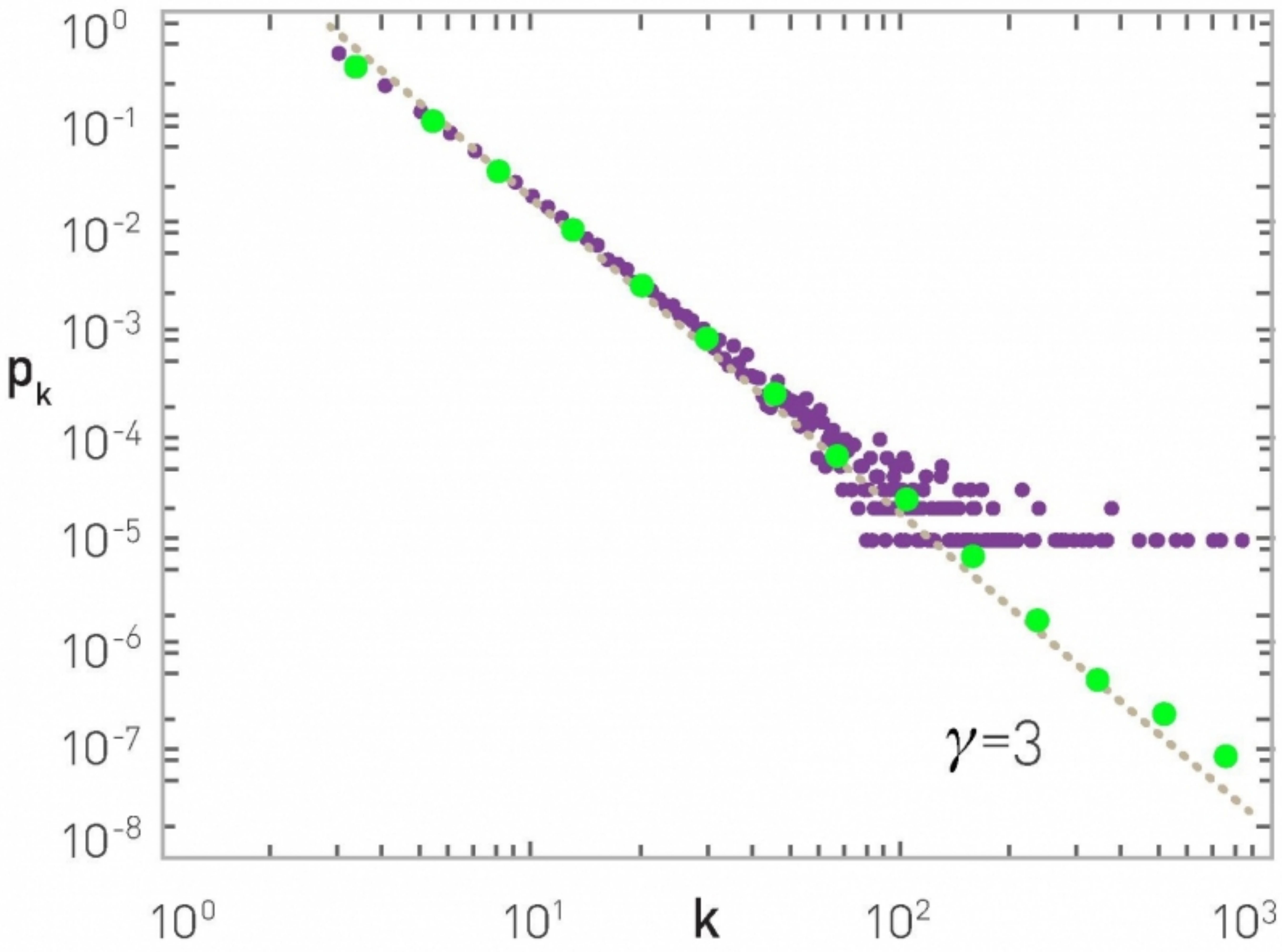
$$\frac{\partial k_i}{k_i} = \frac{\partial t}{2t} \qquad \int_m^k \frac{\partial k_1}{k_i} = \int_{t_i}^t \frac{\partial t}{2t} \qquad \ln\left(\frac{k}{m}\right) = \frac{1}{2}\ln\left(\frac{t}{t_i}\right)$$

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

First mover advantage!

BA degree distribution

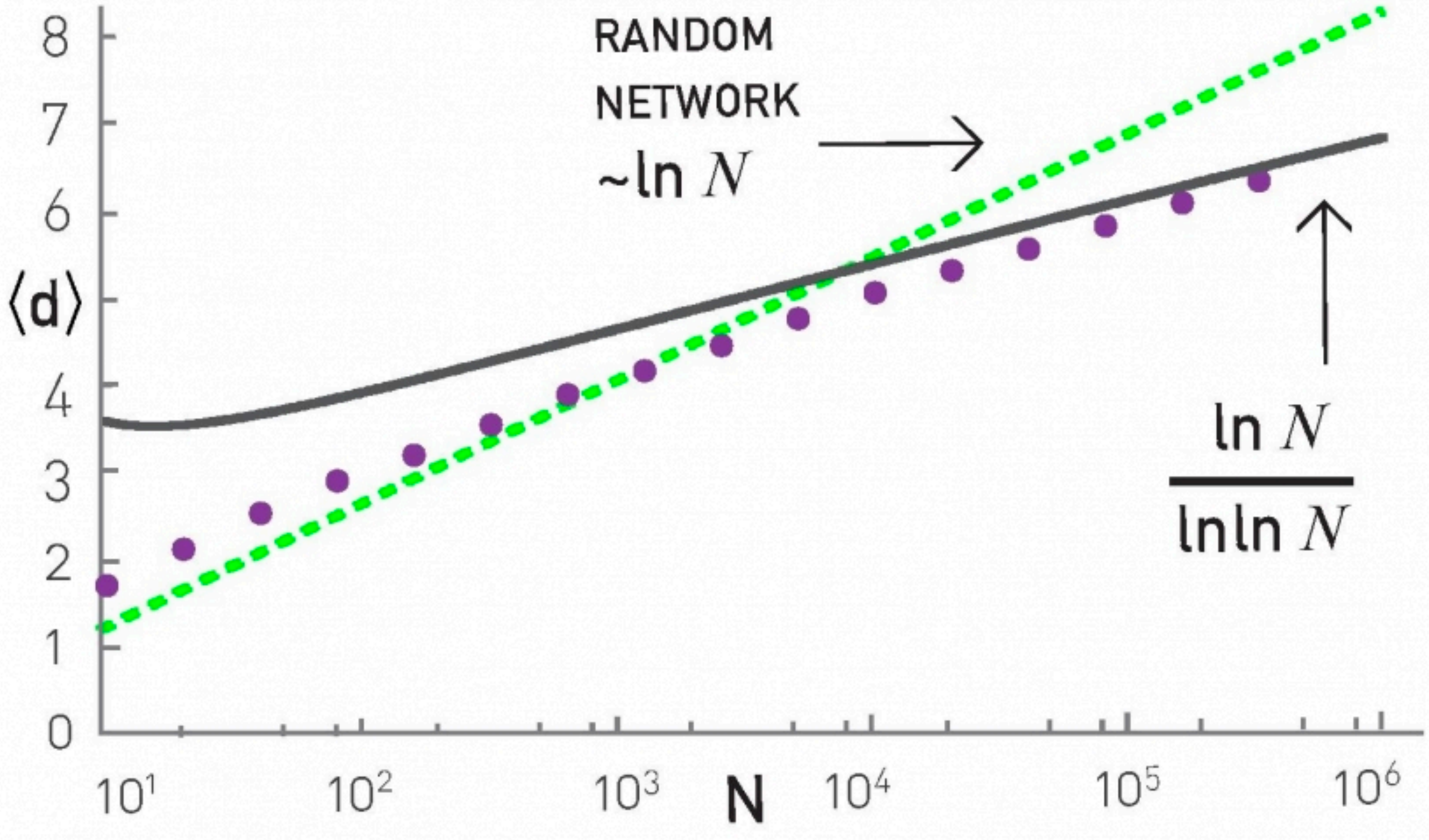
$$P(k) = \frac{2m(2m+1)}{k(k+1)(k+2)} \sim \boxed{k^{-3}}$$



Power-law degree distribution with **fixed** decay exponent: -3

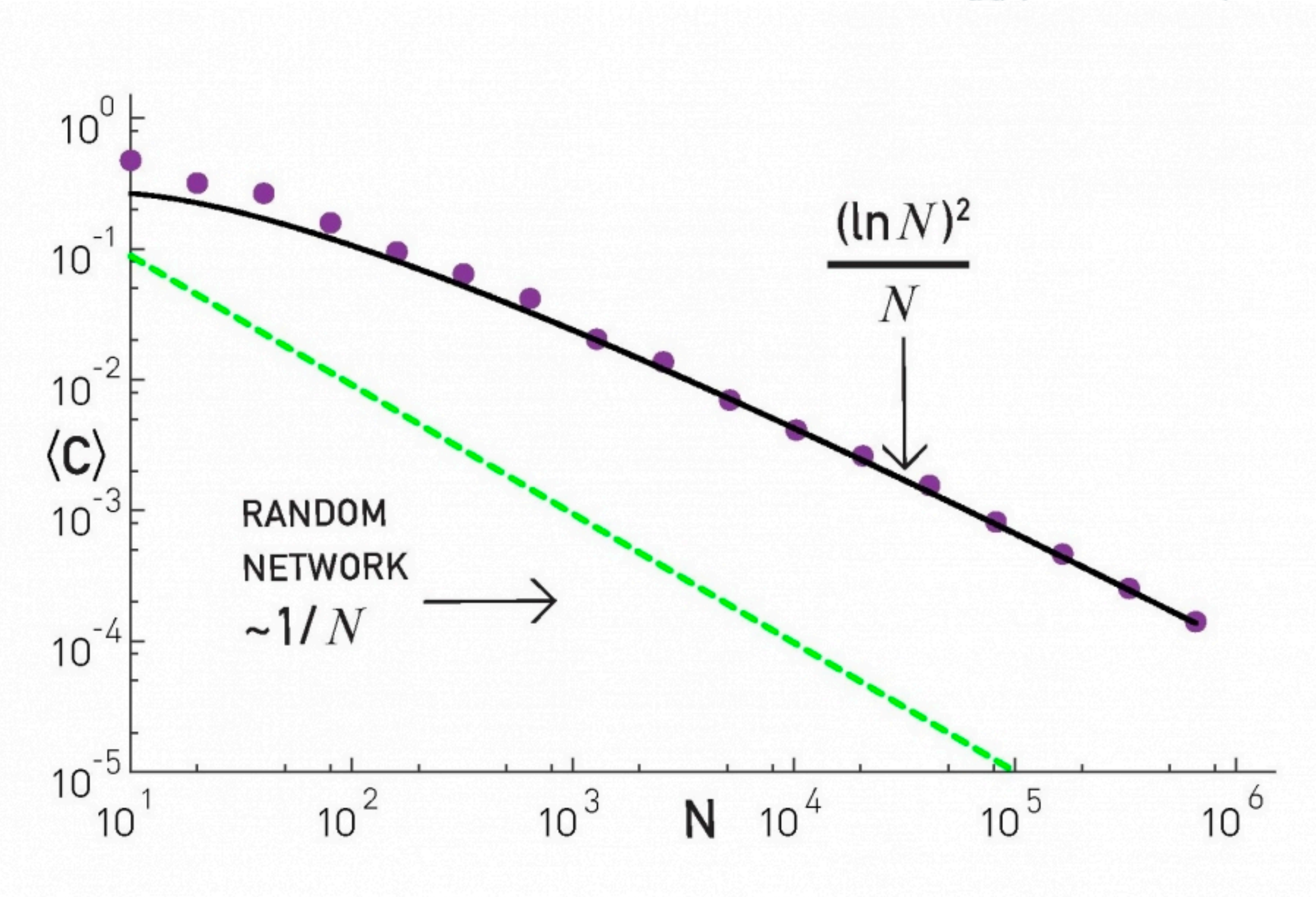
BA diameter

$$\langle d \rangle \sim \frac{\ln N}{\ln(\ln N)}$$



BA clustering

$$\langle C \rangle \simeq \frac{m}{4}\frac{(\ln N)^2}{N}$$



Barabási-Albert summary

- Power law with fixed exponent, equal to -3
- Ultrasmall world
- Undirected
- Vanishing clustering

- Does not capture:
  - variations in the shape of the degree distribution
  - variations in the degree exponent
  - size-independent clustering coefficient

Number of Nodes
$N = t$

Number of Links
$N = mt$

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

Degree Dynamics
$k_i(t) = m\,(t/t_i)^{\beta}$

Dynamical Exponent
$\beta = 1/2$

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

Degree Exponent
$\gamma = 3$

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

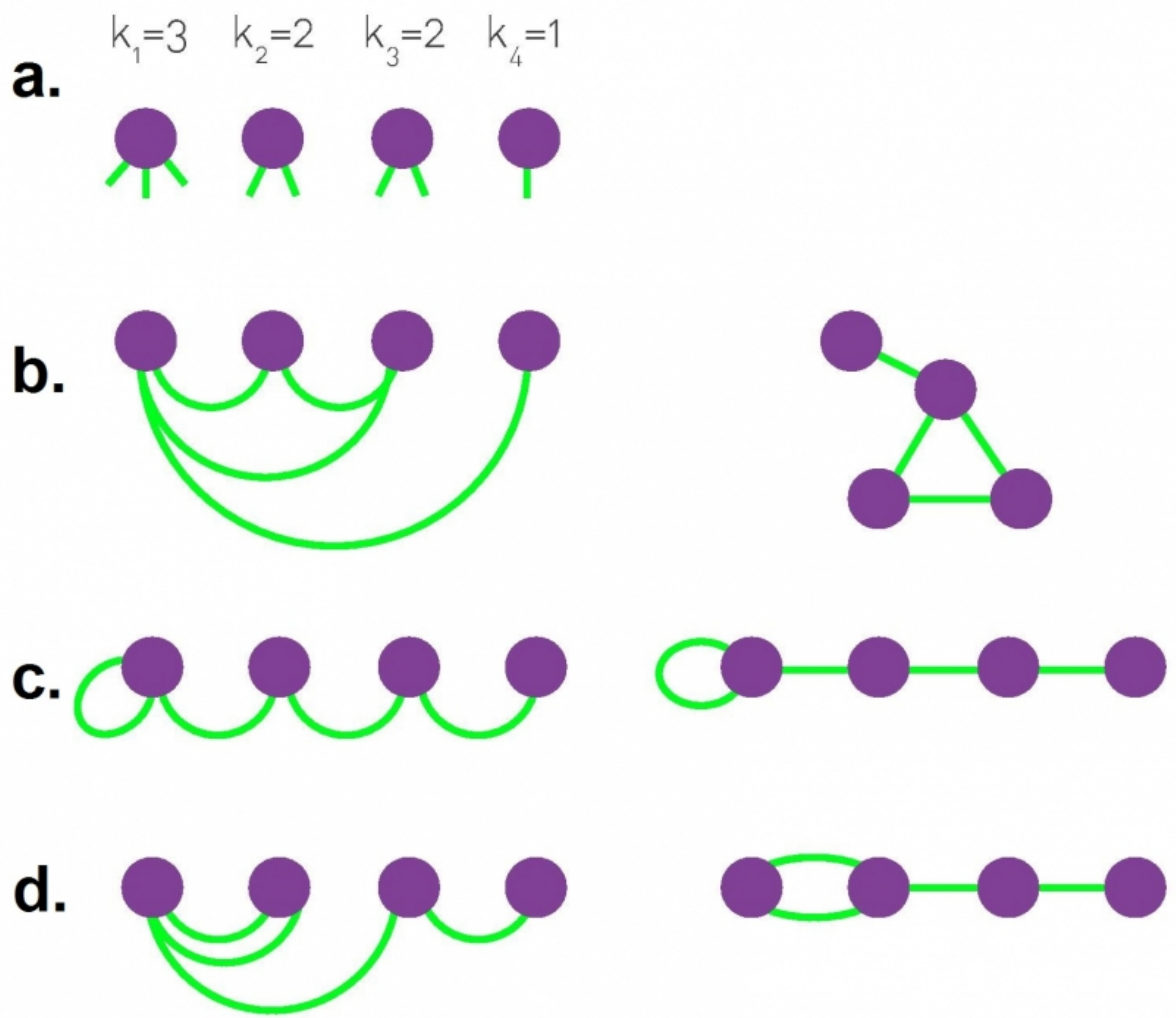
Clustering Coefficient
$\langle C \rangle \sim (\ln N)^2 / N$

Configuration model

- Can we constrain a random model of a static network to be scale-free?
- We can do it by **fixing the degree sequence** and then wire nodes together preserving the degree of each node

Configuration model

1. Given a degree sequence $\boldsymbol{k = \{k_1, k_2, k_3, \dots, k_n\}}$
2. Assign to each node k_i stubs
3. Select random pairs of unmatched stubs and connect them
4. Repeat 3 while there are unmatched stubs



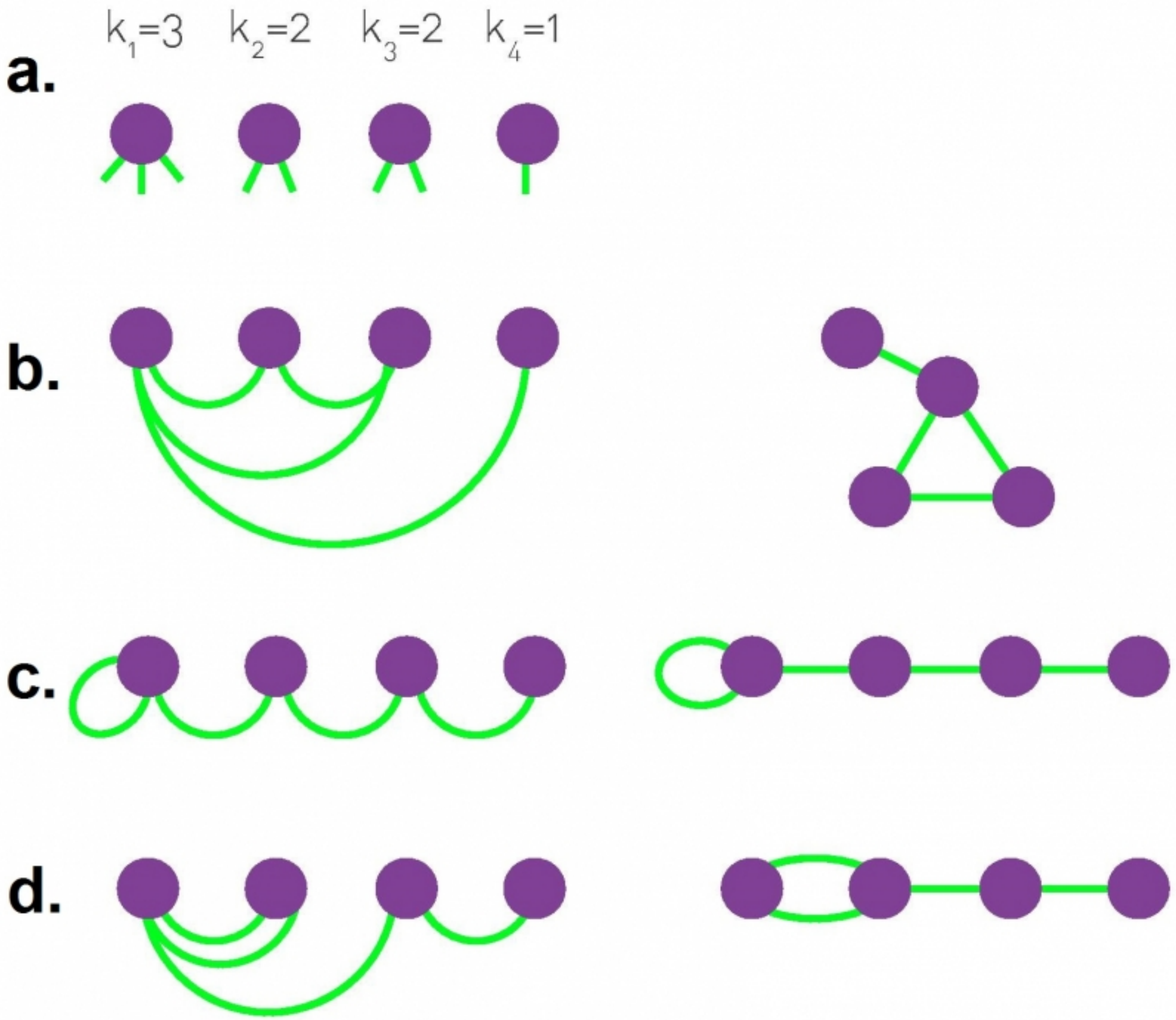
Configuration model

- The probability of having a link between two nodes of degree k_i and k_j is

$$p_{ij} = \frac{k_i k_j}{2L - 1}$$

- Often, this quantity is approximated as

$$p_{ij} = \frac{k_i k_j}{2L}$$



Correlations

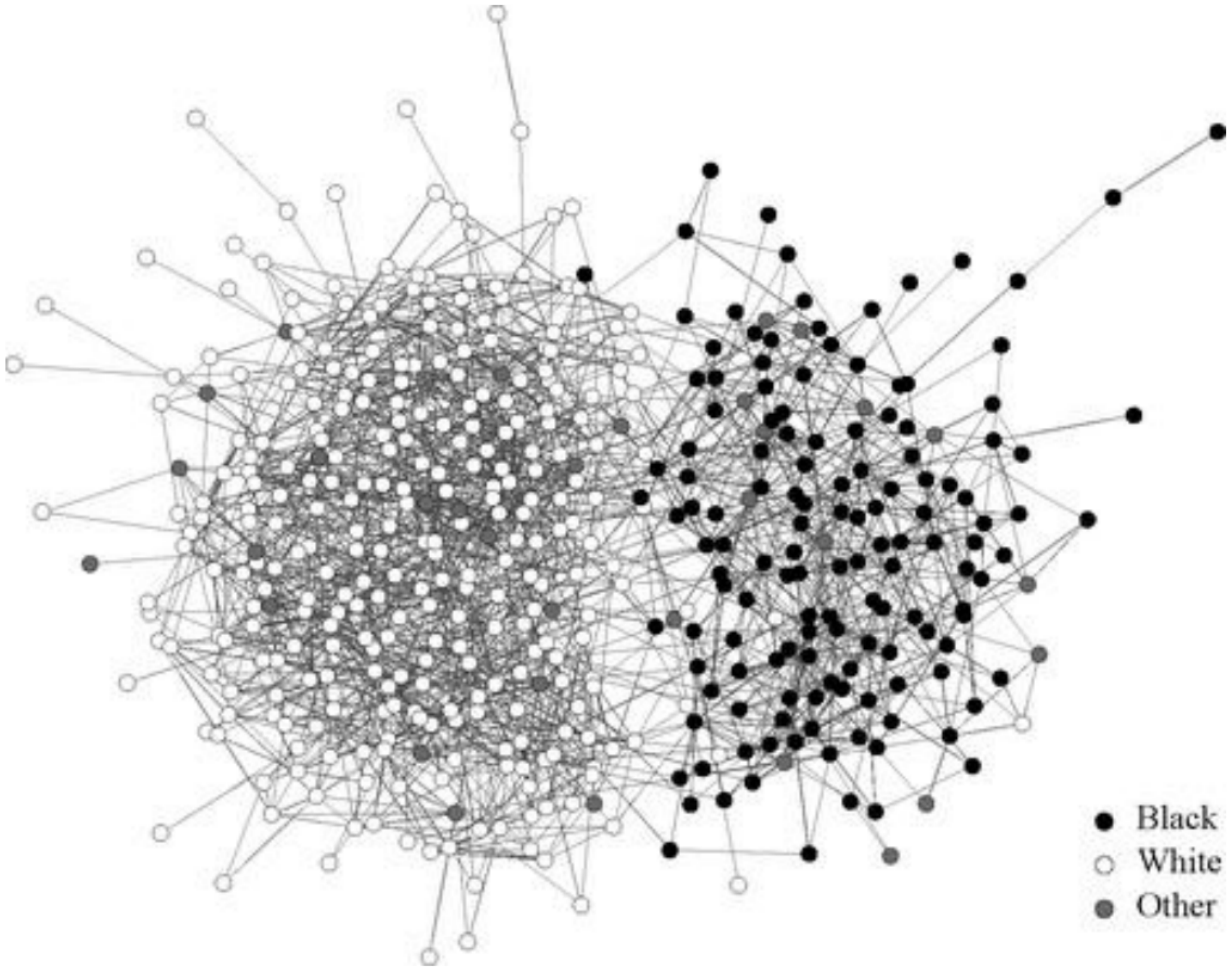
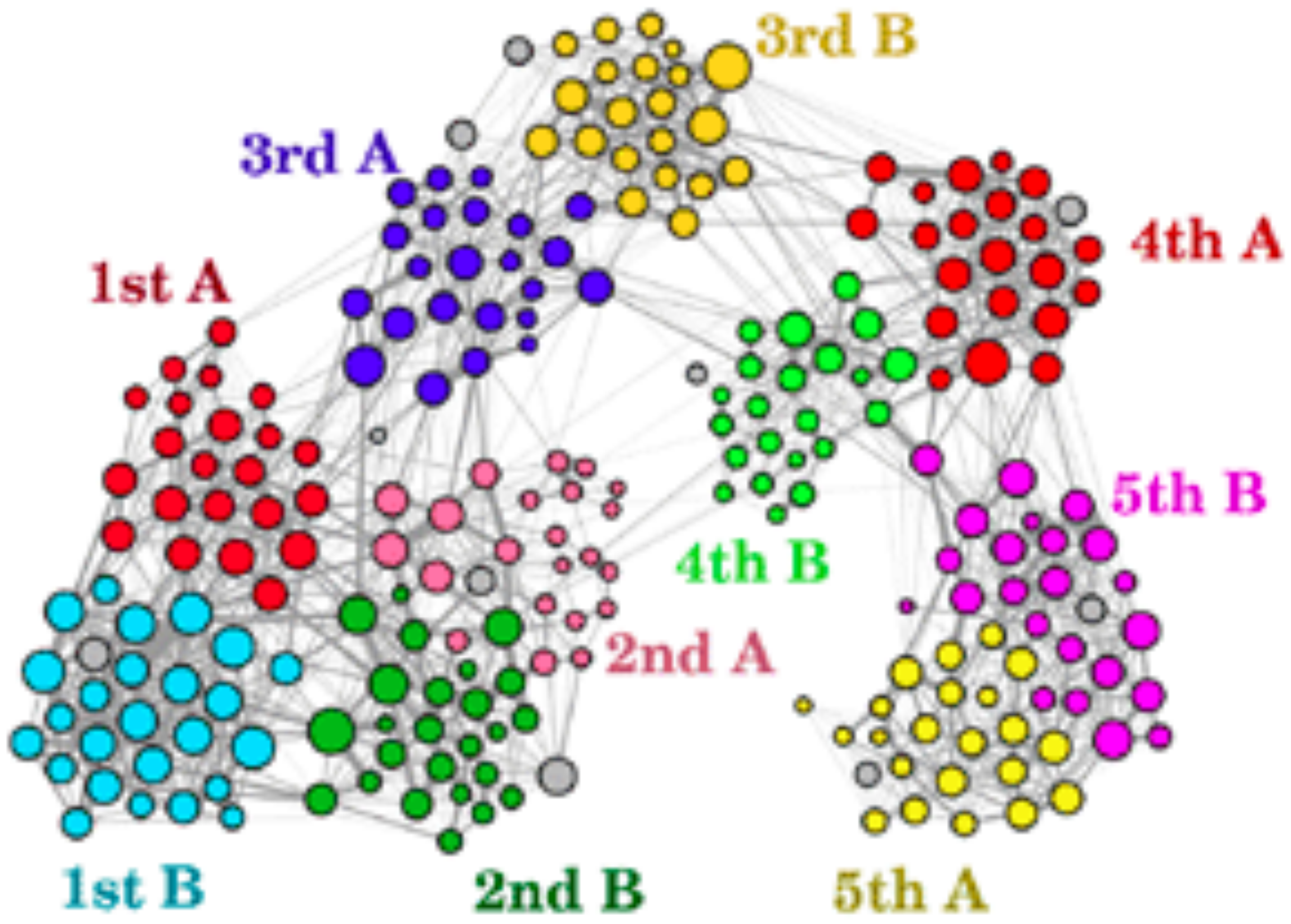


Figure 7.10: Friendship network at a US high school. The vertices in this network represent 470 students at a US high school (ages 14 to 18 years). The vertices are color coded by race as indicated in the key. Data from the National Longitudinal Study of Adolescent Health [34, 314].

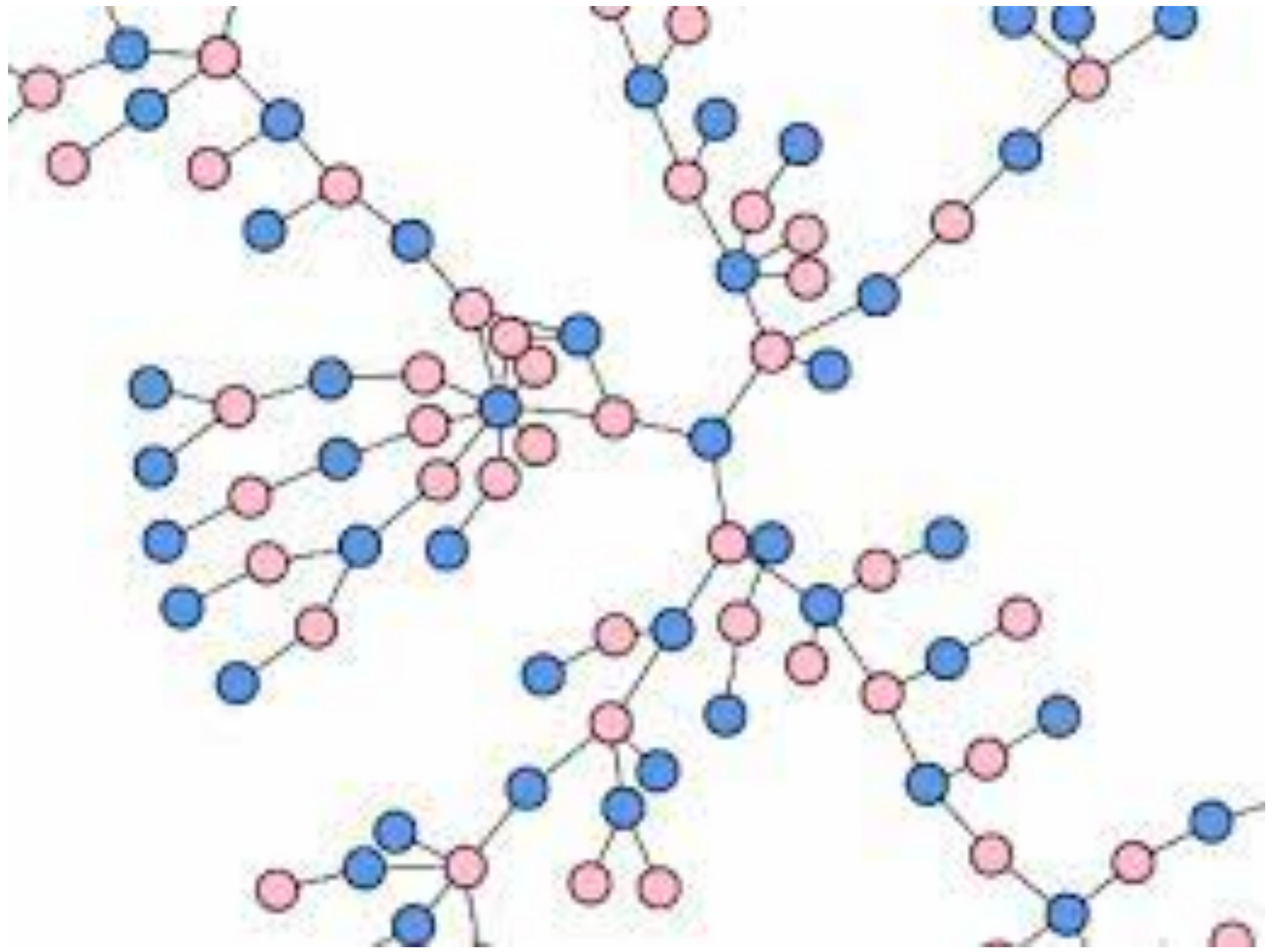
Correlations

Homophily: This is not news to sociologists, who have long observed and discussed such divisions.

Assortative: like is associated with like



Disassortative: like is associated with not-like

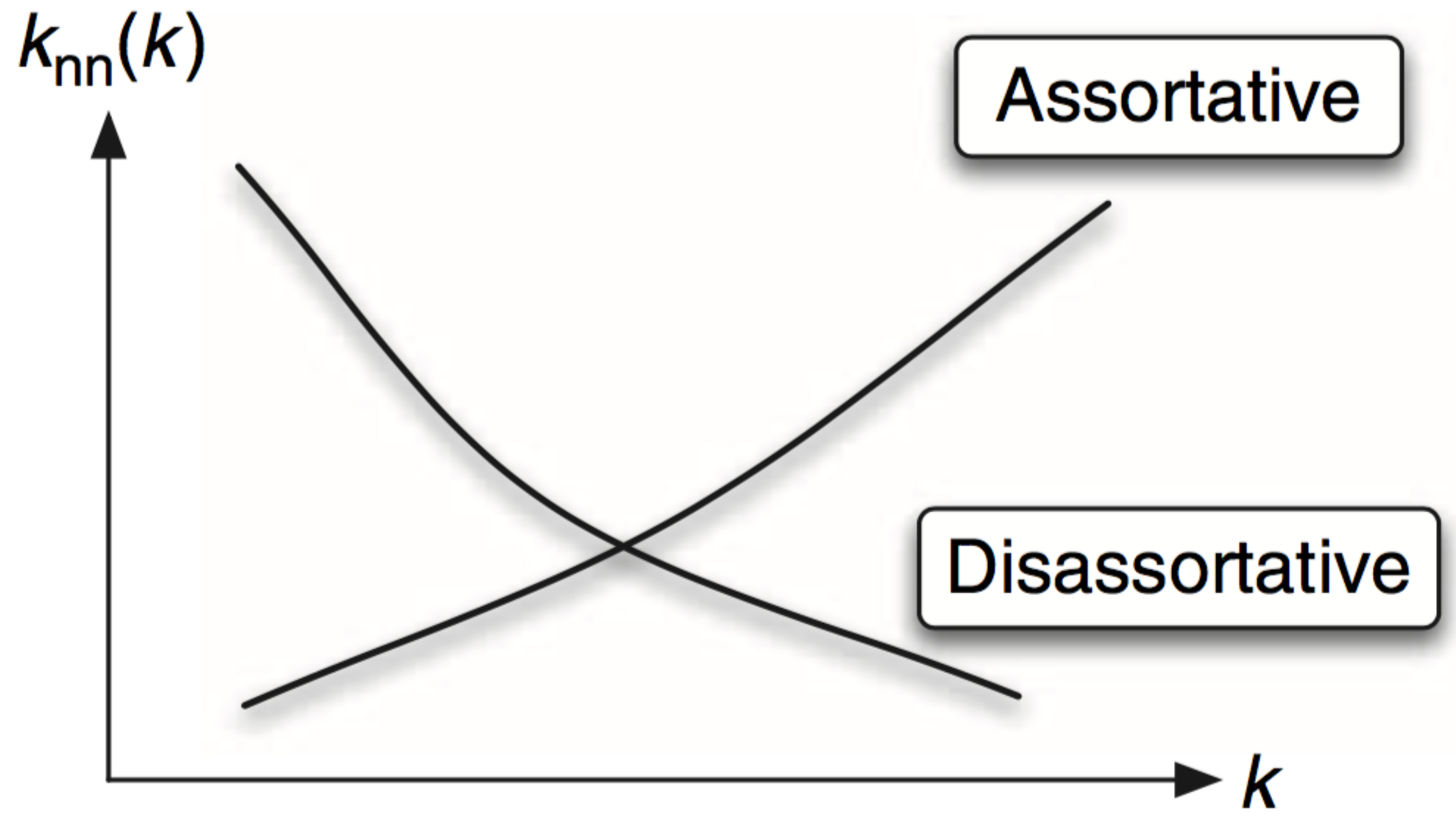


Degree correlations

Average nearest neighbours degree

$$k_{nn,i} = \frac{1}{k_i} \sum_{j \in \nu(i)} k_j$$

$$k_{nn} = \frac{1}{N_k} \sum_{i, k_j = k} k_{nn,i} = \sum_{k'} k' P(k'|k)$$



Random baseline

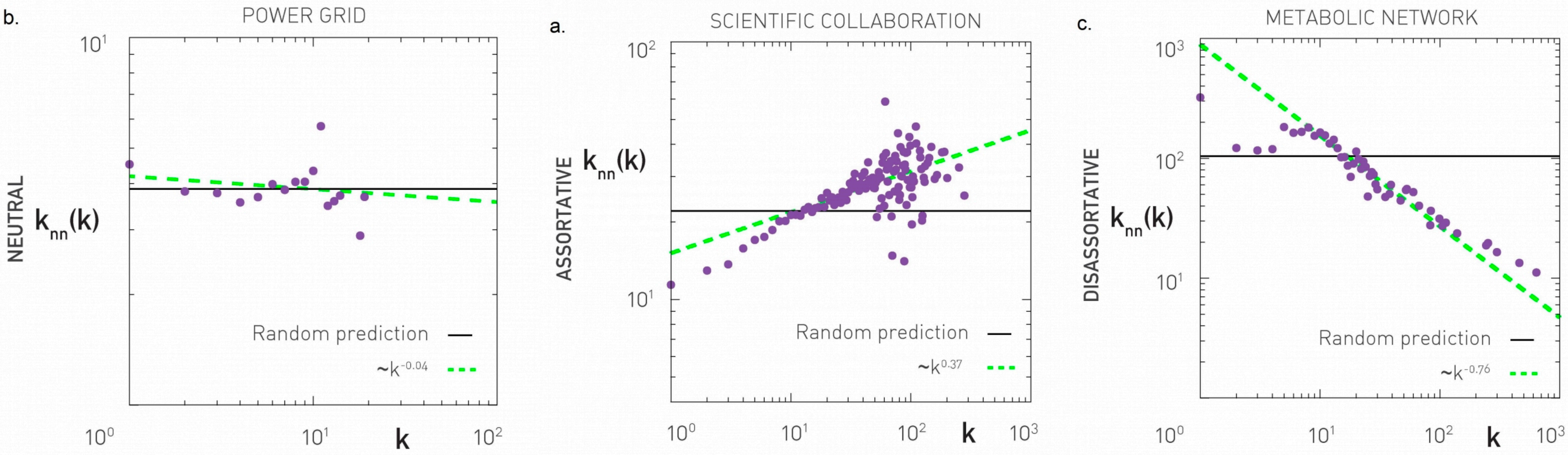
$$k_{nn}^{\mathbf{unc}} = \sum_{k'} k' P^{\mathbf{unc}}(k' \mid k) = \sum_{k'} k' \frac{k' P(k')}{\langle k \rangle} = \frac{\langle k^2 \rangle}{\langle k \rangle}$$

$$\boxed{P^{\mathbf{unc}}(k' \mid k) = \frac{k'}{2L} P(k') N = \frac{k' P(k')}{\langle k \rangle}}$$

This observation will be extensively used to simplify analytical calculations when dealing with dynamical processes (SIR, SIS) on networks.

Random baseline

$$k_{nn}^{\mathbf{unc}} = \sum_{k'} k' P^{\mathbf{unc}}(k' \mid k) = \sum_{k'} k' \frac{k' P(k')}{\langle k \rangle} = \frac{\langle k^2 \rangle}{\langle k \rangle}$$



Next.. Epidemics on networks