Network theory

Part III

Complexity in Social Systems
AA 2023/2024
Maxime Lucas
Lorenzo Dall'Amico

Recap last lecture

Models

Erdos-Renyi model
Watts-Strogatz model
Configuration model
Chung-Lu model

Concepts

Importance of scale-freeness
Giant component
Phase transitions

Today's topics

Models

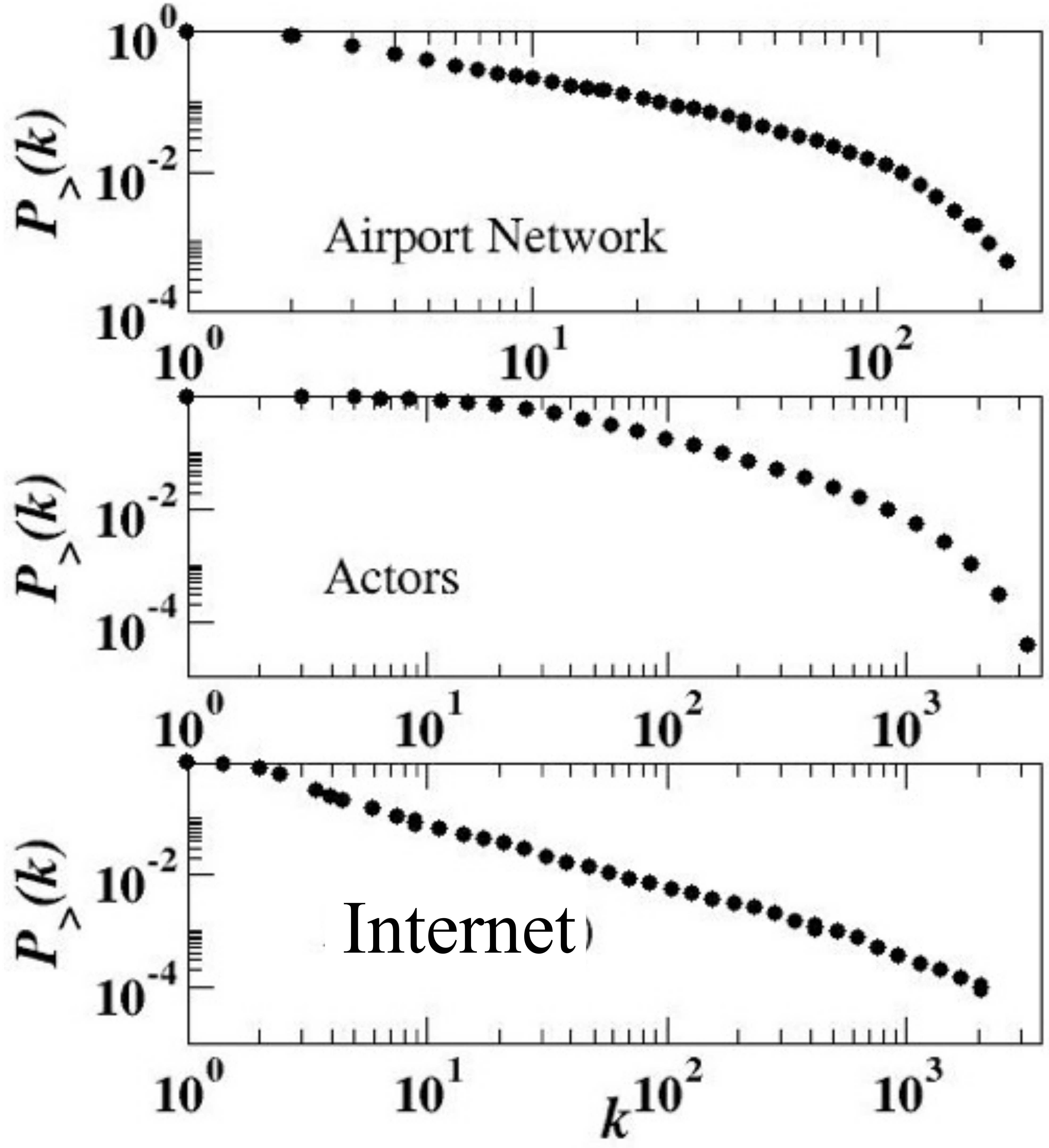
Barabasi-Albert model
Bianconi-Barabasi model
Link/Copying model

Concepts

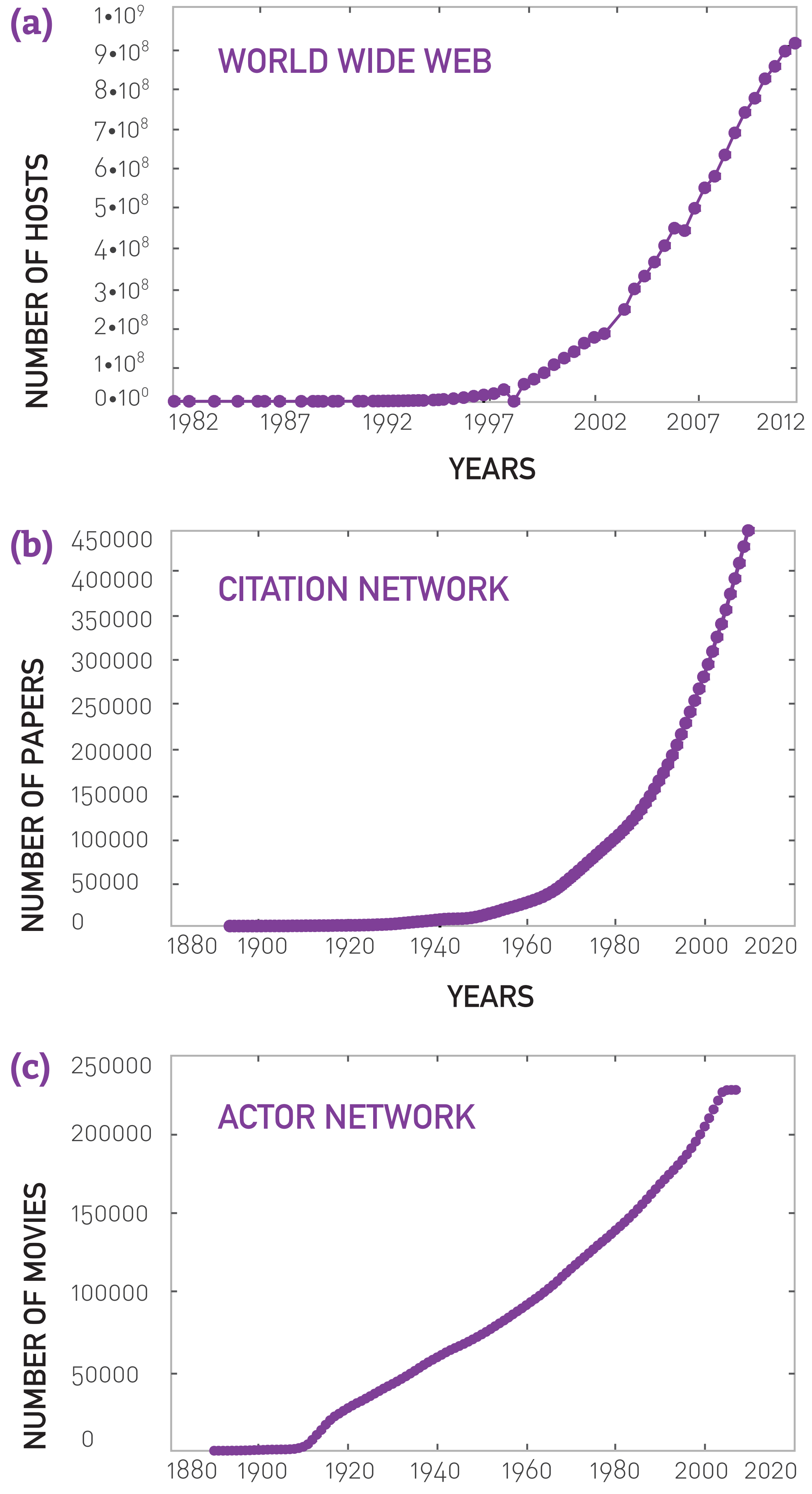
Origins of scale-free distributions
Growing networks
Assortativity and correlations
Bonus: Robustness

Network heterogeneity

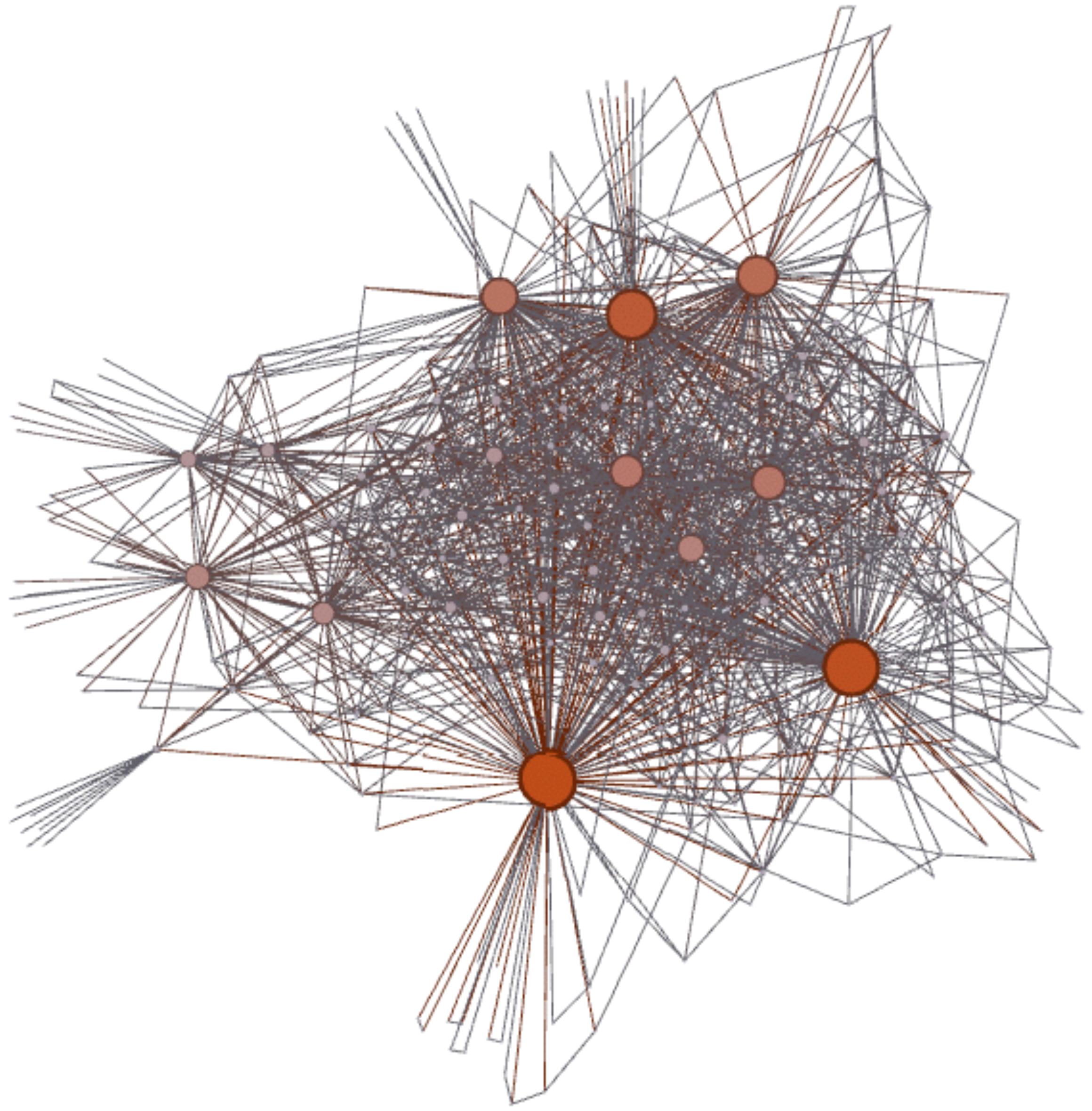
Where does it come from?



First ingredient:
Growth/time



Second ingredient:
Not all links are equally likely!



Network heterogeneity

Microscopic mechanisms for macroscopic observables

In static ensemble models (last lecture) we defined network by constraints
In evolving/growing network, we define growth rules and look for asymptotic stationary behaviour

Barabasi-Albert network model

(1) Networks continuously expand by the addition of new nodes
WWW : addition of new documents

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.

(2) New nodes prefer to link to highly connected nodes.
(WWW : linking to well known sites)

PREFERENTIAL ATTACHMENT:

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

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

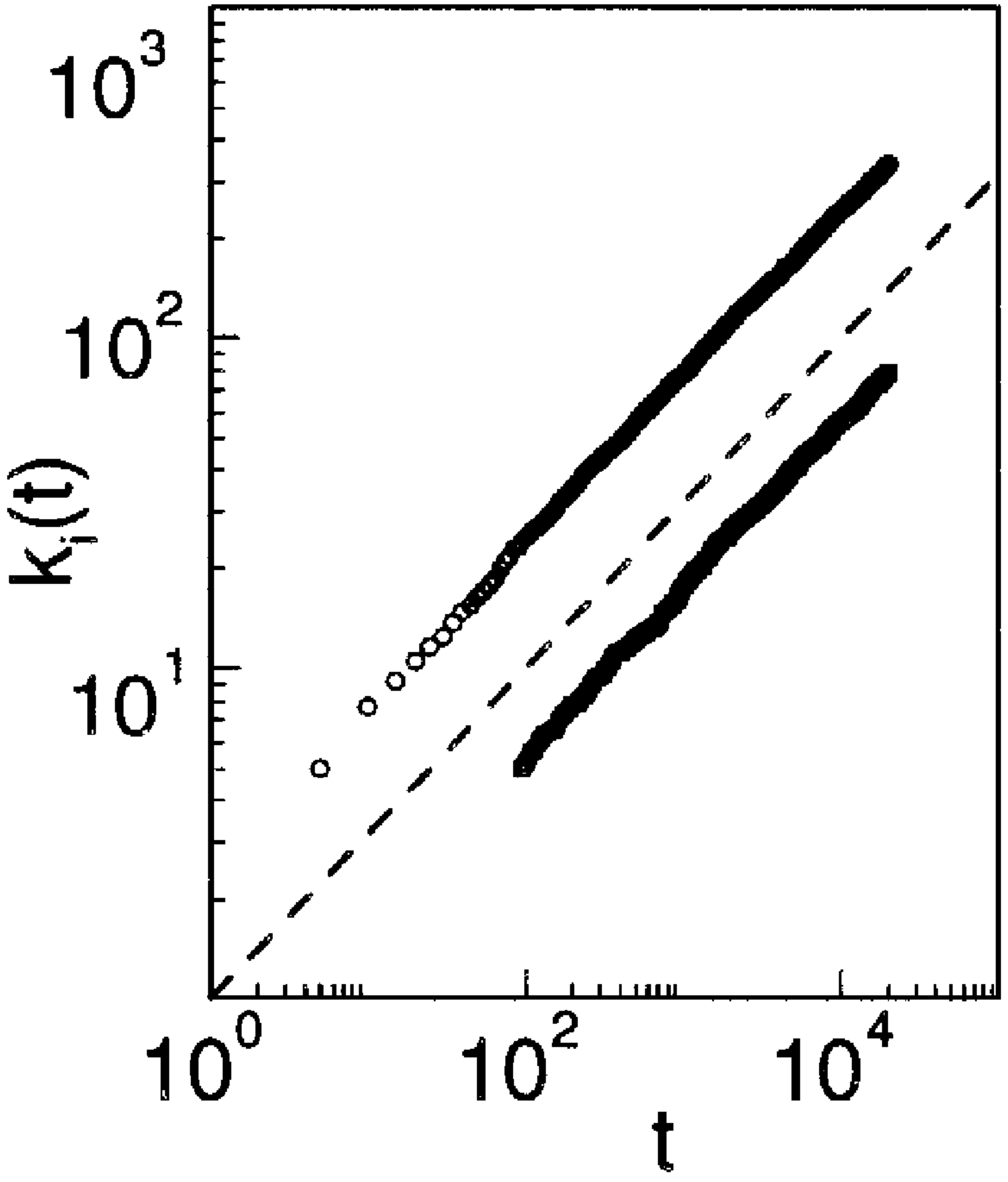
Barabasi-Albert model

Degree dynamics

$$\frac{dk_i}{dt} = m\Pi(k_i) = m\frac{k_i}{\sum_{j=1}^{N-1} k_j}$$

$$\sum_{j=1}^{N-1} k_j = 2mt - m \overset{t\gg 1}{\approx} 2mt$$

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



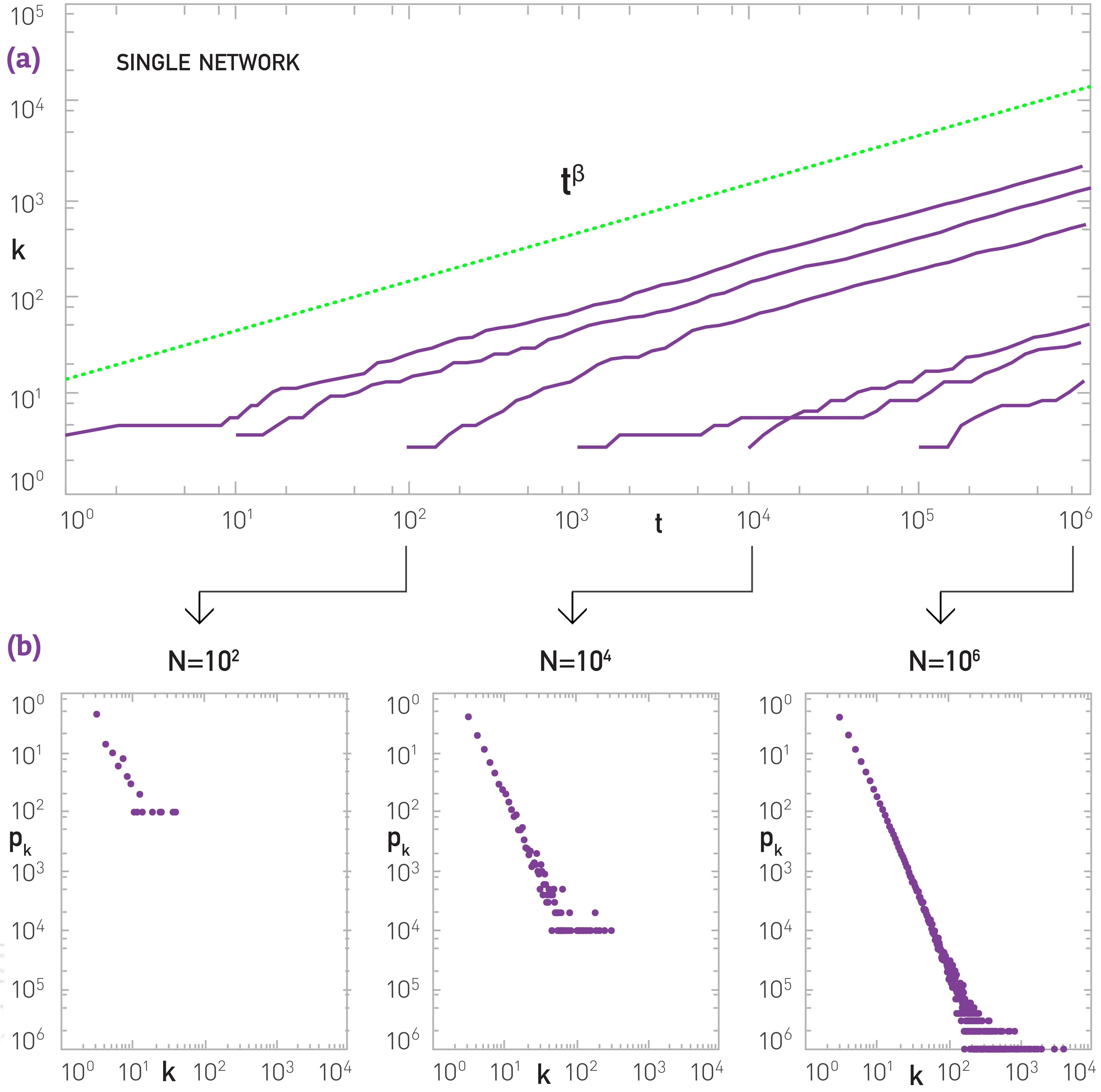
$$k_i(t) = m\left(\frac{t}{t_i}\right)^{\beta} \qquad \beta = \frac{1}{2}$$

dynamical exponent

Barabási, Albert-László, Réka Albert, and Hawoong Jeong. "Mean-field theory for scale-free random networks." *Physica A: Statistical Mechanics and its Applications* 272.1-2 (1999): 173-187.

Barabasi-Albert model

Degree distribution: simple derivation



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

$$P(k_i(t) < k) = 1 - P(k_i(t) > k) = 1 - P(t_i < \frac{m^2 t}{k^2}) = \dots$$

A node *i* can come with equal probability any time between $t_i = m_0$ and t, hence:

$$t_i < t\left(\frac{m}{k}\right)^{1/\beta}$$

== Number of nodes with degree larger than k

There are overall $N = m_0 + t \sim t$

So the probability that a random node has degree k or less is:

$$P(k) \sim 1 - \left(\frac{m}{k}\right)^{1/\beta}$$

$$p_k \sim \frac{\partial P(k)}{\partial k} = \frac{1}{\beta}\frac{m^{1/\beta}}{k^{1/(\beta+1)}}$$

$$p_k \sim 2m^2k^{-3} \qquad \beta = \frac{1}{2}$$

This is great but there are errors in the coefficient

Barabasi-Albert model

Rate-equation derivation

Number of degree k nodes at time t: $< N(k,t) >= tP(k,t)$

Number of links added to degree k nodes after the arrival of a new node: $\frac{k}{2mt} \cdot NP(k,t) \cdot m = \frac{k}{2}P(k,t)$

$$(N+1)P(k,t+1) = NP(k,t) + \frac{k-1}{2}P(k-1,t) - \frac{k}{2}P(k,t)$$

k-nodes at time t+1 | # k-nodes at time t | Gain of k nodes via k-1 to k | Loss of k nodes via k to k+1

We do not have *k=0,1,...,m-1* nodes in the network (each node arrives with degree *m*)

Requires separate equation for degree *m* nodes.

$$(N+1)P(m,t+1) = NP(m,t) + 1 - \frac{m}{2}P(m,t)$$

Barabasi-Albert model

Rate-equation derivation

$$(N+1)P(k,t+1) = NP(k,t) + \frac{k-1}{2}P(k-1,t) - \frac{k}{2}P(k,t)$$

$$(N+1)P(m,t+1) = NP(m,t) + 1 - \frac{m}{2}P(m,t)$$

Impose stationarity! $P(k,\infty) = P(k)$

$$(N+1)P(k,\infty) - NP(k,\infty) = P(k,\infty) = P(k)$$

$$P(k) = \frac{k-1}{2}P(k-1) - \frac{k}{2}P(k) \qquad P(k) = \frac{k-1}{k+2}P(k-1)$$

$$P(m) = 1 - \frac{m}{2}P(m) \qquad \rightarrow \qquad P(m) = \frac{2}{m+2}$$

Barabasi-Albert model

Rate-equation derivation

$$P(m) = \frac{2}{m+2}$$

$$P(m+1) = \frac{m}{m+3} P(m) = \frac{2m}{(m+2)(m+3)}$$

$$P(m+2) = \frac{m+1}{m+4} P(m+1) = \frac{2m(2m+1)}{(m+2)(m+3)(m+3)}$$

$$P(m+3) = \frac{m+2}{m+5} P(m+2) = \frac{2m(2m+1)}{(m+3)(m+4)(m+5)}$$

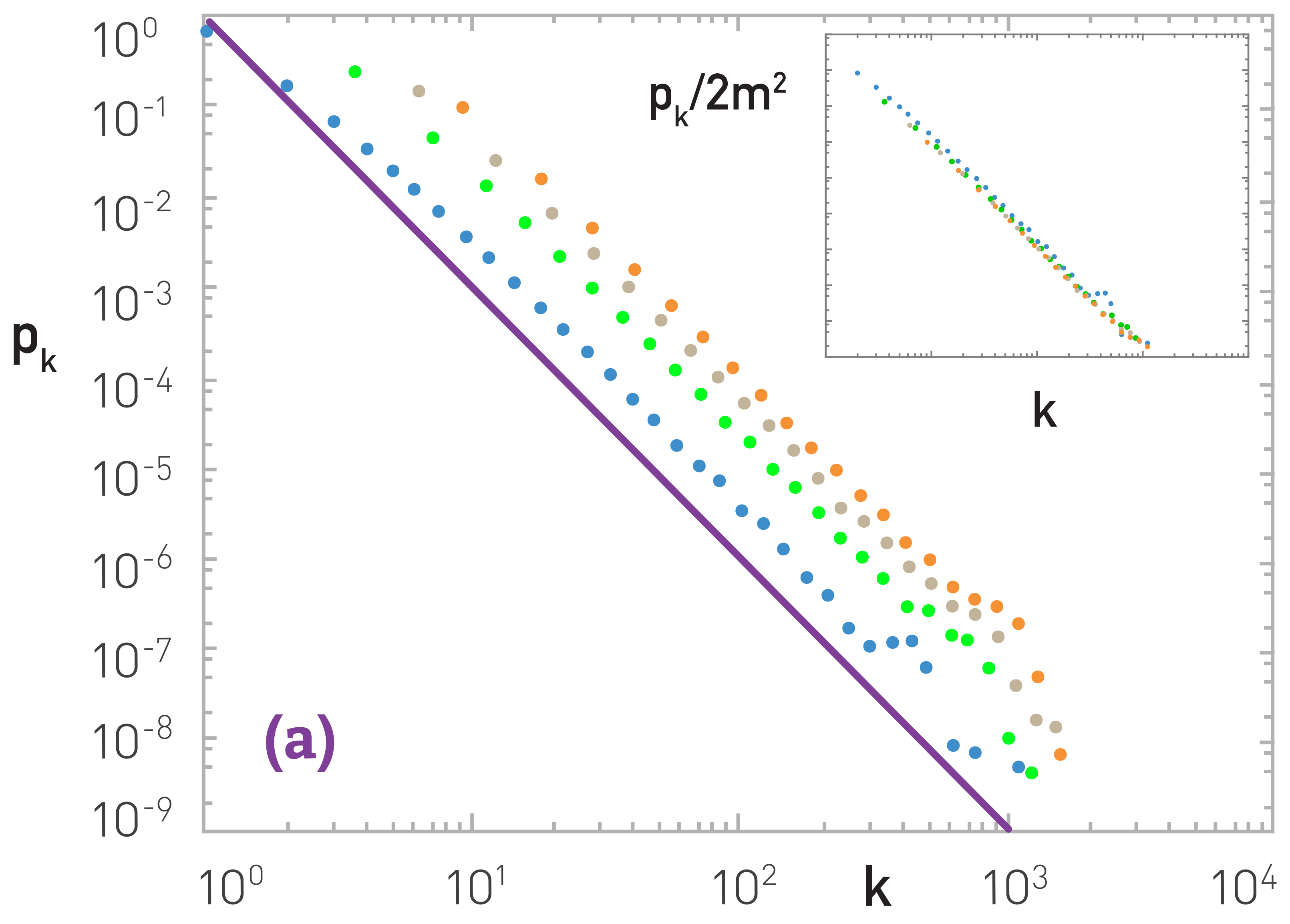
$$\ldots = \ldots$$

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

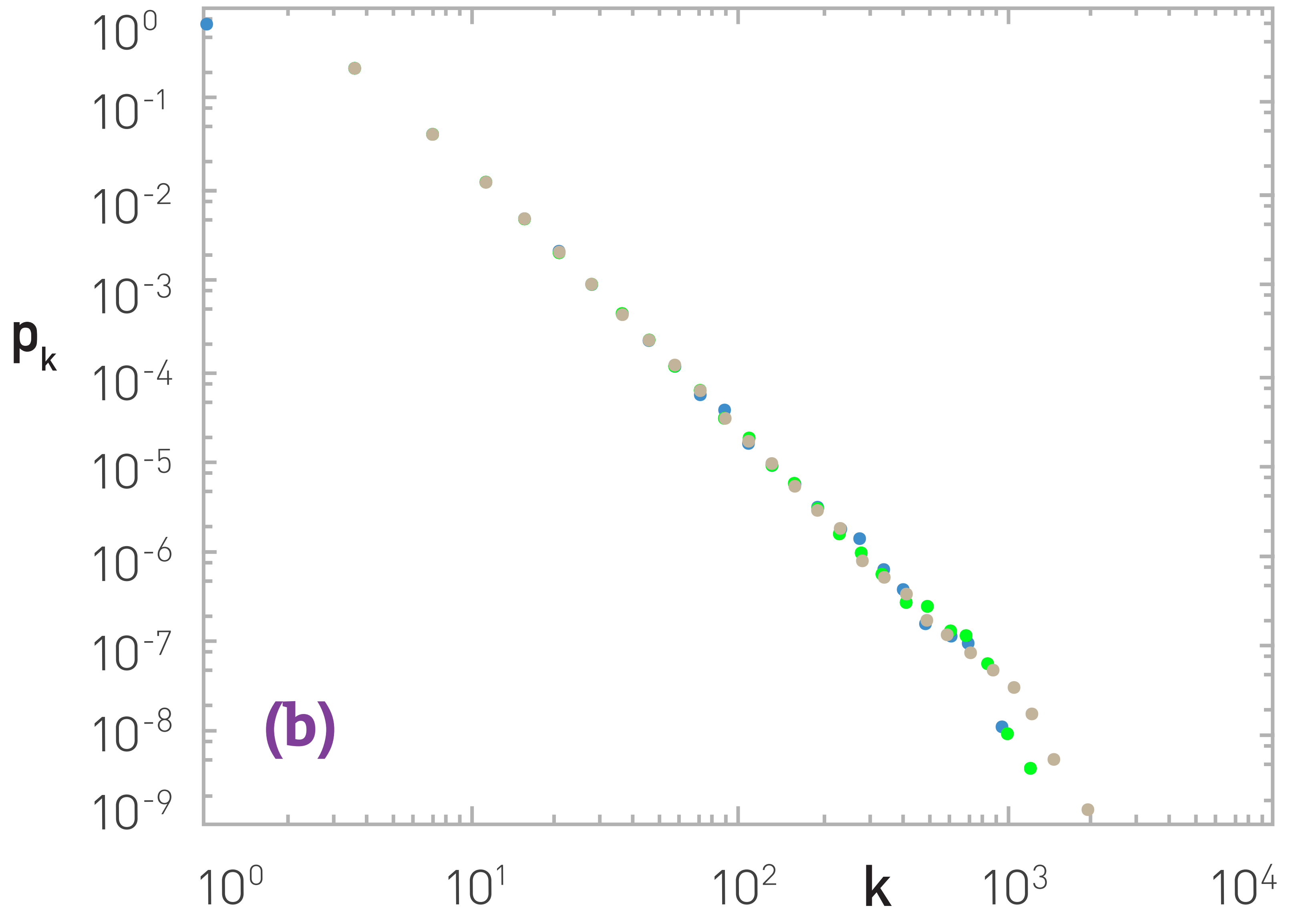
Krapivsky, Redner, Leyvraz, PRL 2000, Dorogovtsev, Mendes, Samukhin, PRL 2000, Bollobas et al, Random Struc. Alg. 2001

Barabasi-Albert model

Validation



(a) We generated networks with N=100,000 and m_0=m=1 (blue), 3 (green), 5 (grey), and 7 (orange). The fact that the curves are parallel to each other indicates that γ is independent of m and m_0. The slope of the purple line is -3, corresponding to the predicted degree exponent γ=3. Inset: (5.11) predicts $p_k \sim 2m^2$, hence $p_k/2m^2$ should be independent of m. Indeed, by plotting $p_k/2m^2$ vs. k, the data points shown in the main plot collapse into a single curve.



(b) The Barabási-Albert model predicts that p_k is independent of N. To test this we plot p_k for N = 50,000 (blue), 100,000 (green), and 200,000 (grey), with m_0=m=3. The obtained p_k are practically indistinguishable, indicating that the degree distribution is stationary, i.e. independent of time and system size.

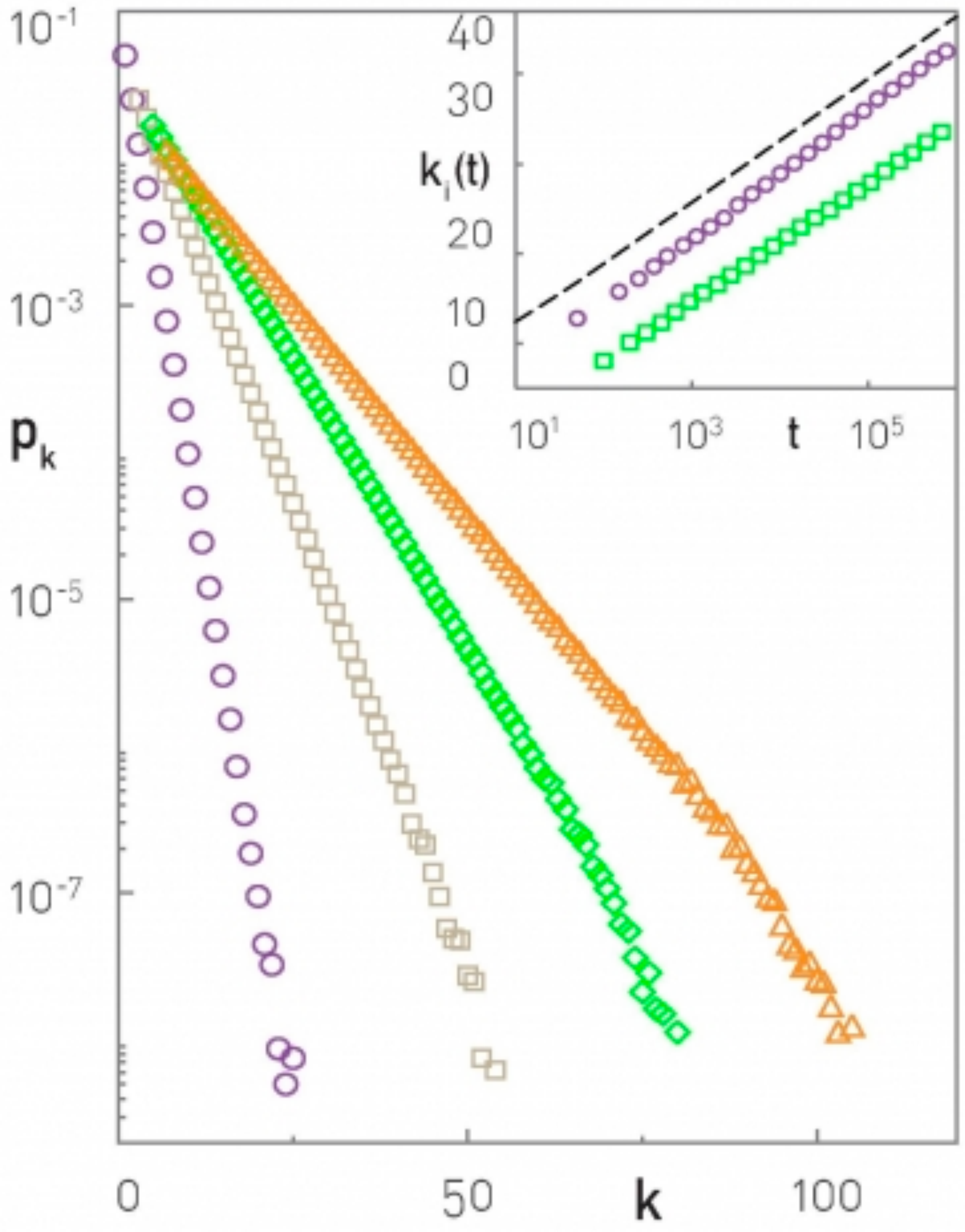
Barabasi-Albert model

Necessity of ingredients: growth

$$\frac{\partial k_i}{\partial t} = \frac{m}{m_0 + t - 1}$$

$$k_i(t) = m \ln\left(\frac{m_0 + t - 1}{m + t_i - 1}\right) + m$$

$$P(k) = \frac{e}{m} e^{-\frac{k}{m}} \sim e^{-k}$$



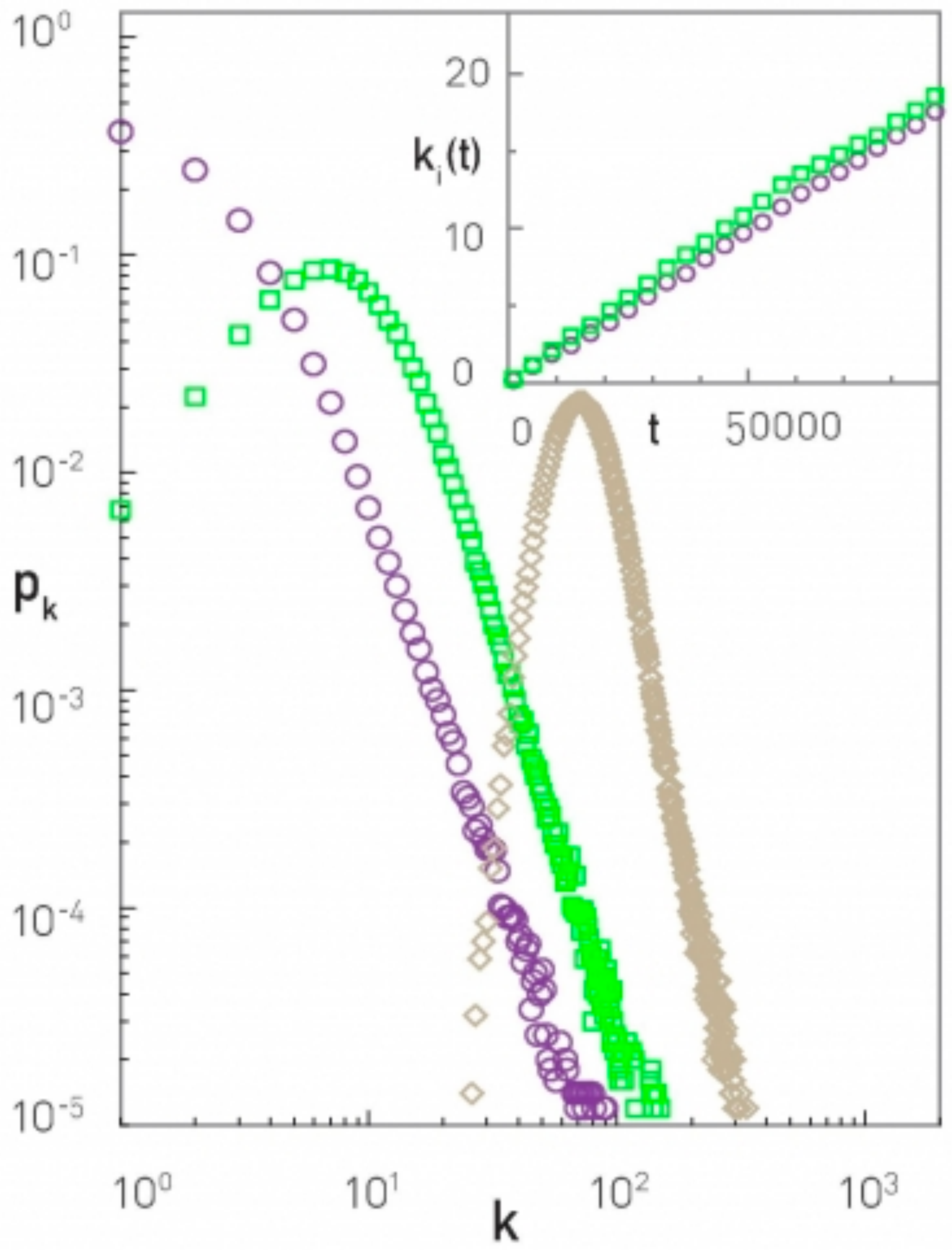
Barabasi-Albert model

Necessity of ingredients: preferential attachment

$$\frac{\partial k_i}{\partial t} = \frac{N}{N-1}\frac{k_i}{2t} + \frac{1}{N}$$

$$k_i(t) = \frac{2(N-1)}{N(N-2)}t + Ct^{\frac{N}{2(N-1)}} \sim \frac{2}{N}t$$

p_k : power law (initially) -> Gaussian -> Fully Connected



Barabasi-Albert model

Necessity of ingredients?

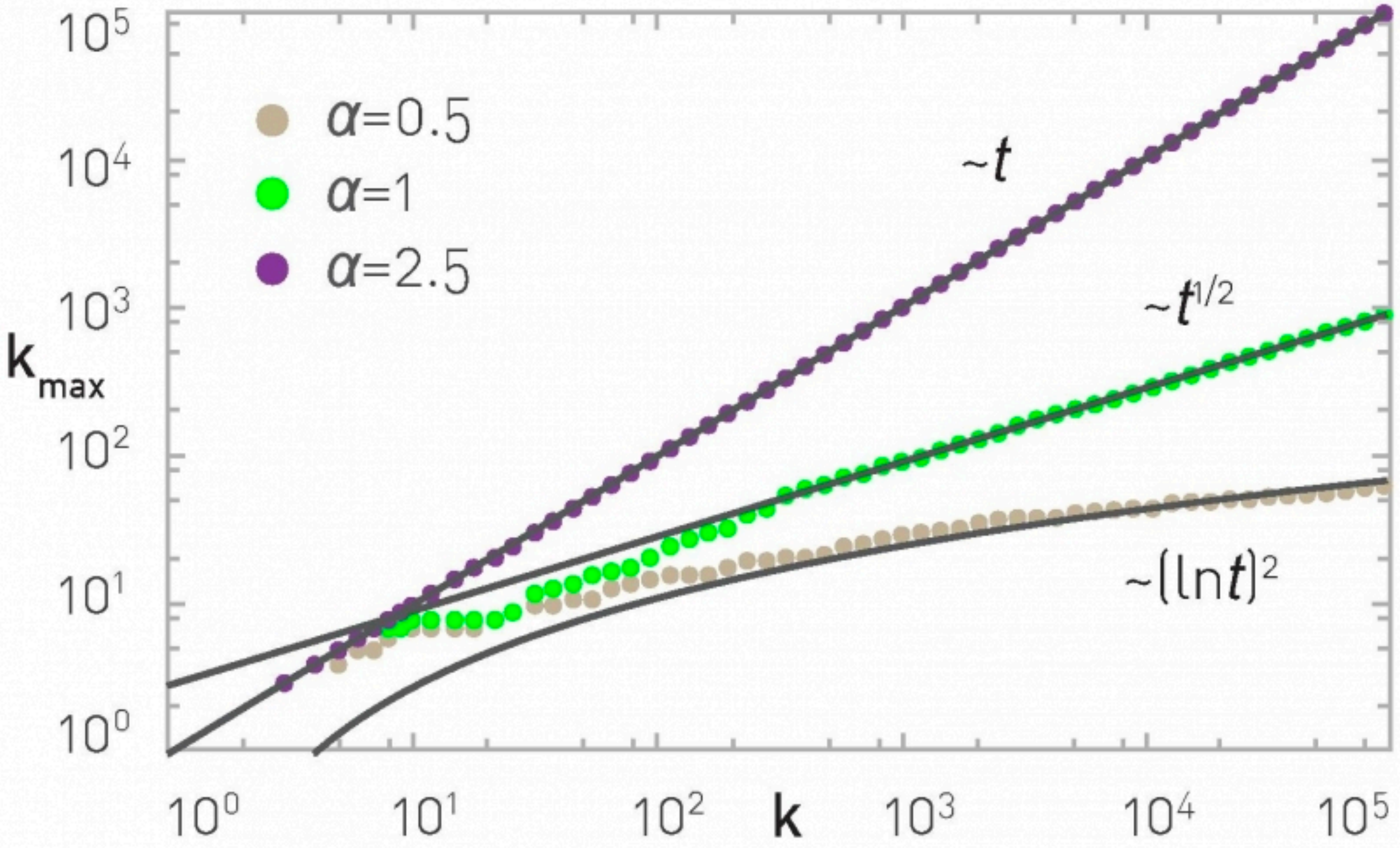
Do we need both growth and preferential attachment?

YEP.

Barabasi-Albert model

Non-linear preferential attachments

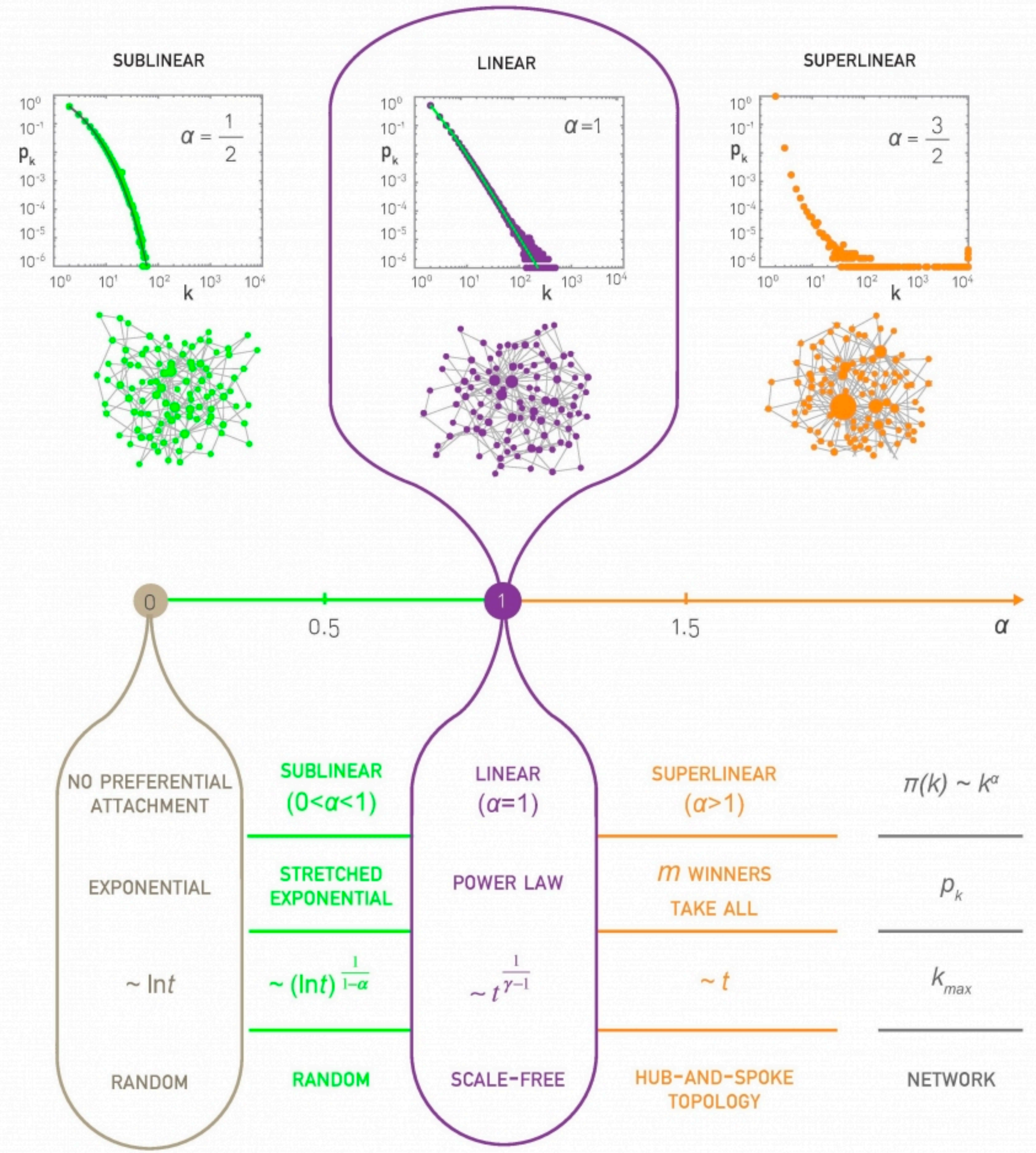
$$\Pi(k) = \frac{k^{\alpha}}{t\sum_k k^{\alpha} p_k(t)}$$



Barabasi-Albert model

Non-linear preferential attachments

$$\Pi(k) = \frac{k^\alpha}{t\sum_k k^\alpha p_k(t)}$$



Explicit derivation in Barabasi's book Sec 5.14

Barabasi-Albert model

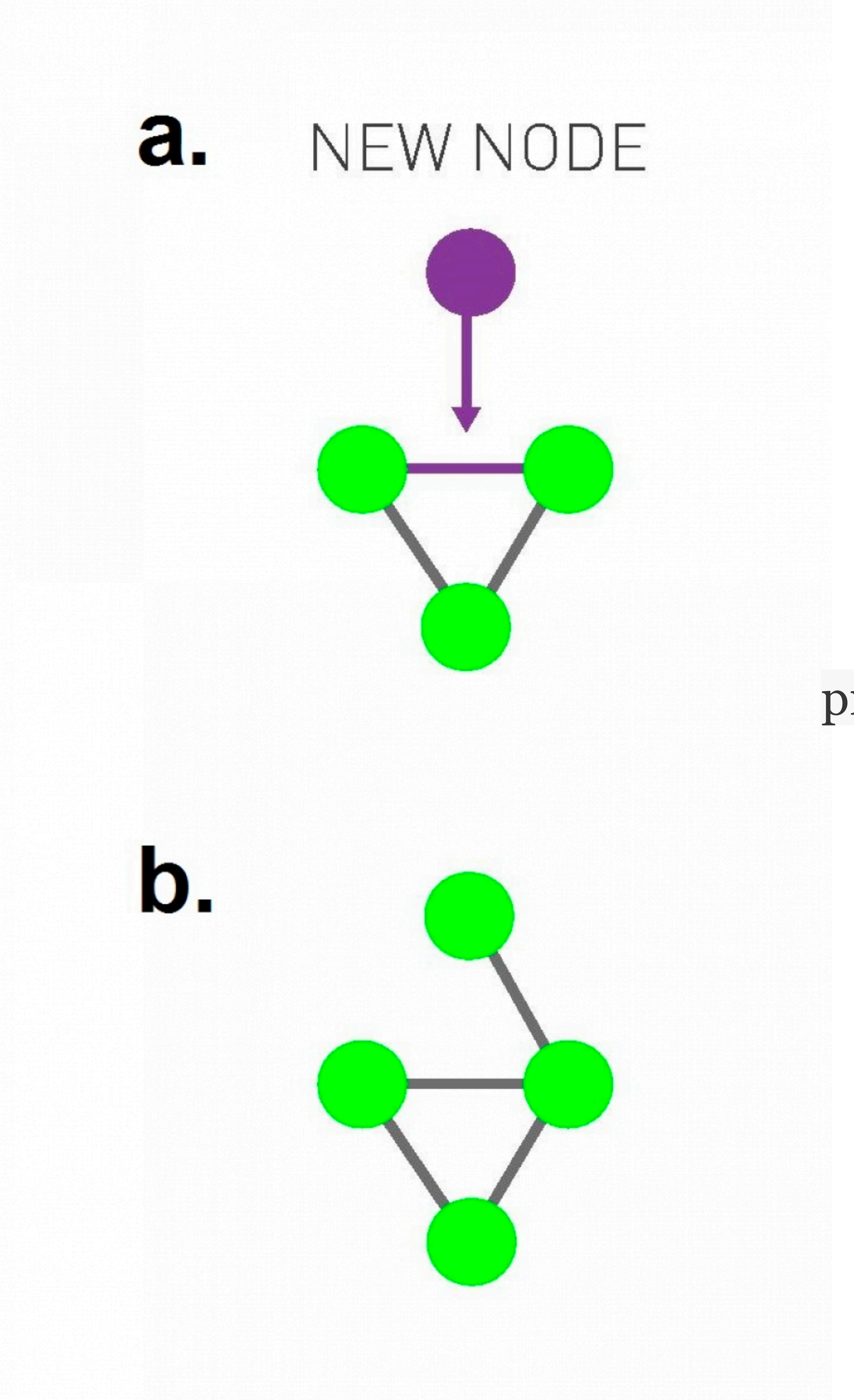
Multiple questions:

- why preferential attachment depends on k?
- Why linearly?
- Global (optimum) versus local (random) mechanisms?

Barabasi-Albert model

Potential local mechanisms: link model

simplest example of a local mechanism that generates a scale-free network without preferential attachment



- *Growth*: At each time step we add a new node to the network.
- *Link Selection*: We select a link at random and connect the new node to one of the two nodes at the two ends of the selected link.

lacks a built-in *Π(k)* function. Yet, it generates preferential attachment.

probability q_k that the node at the end of a randomly chosen link has degree k

$$q_k = C k p_k \qquad \sum_j q_k = 1$$

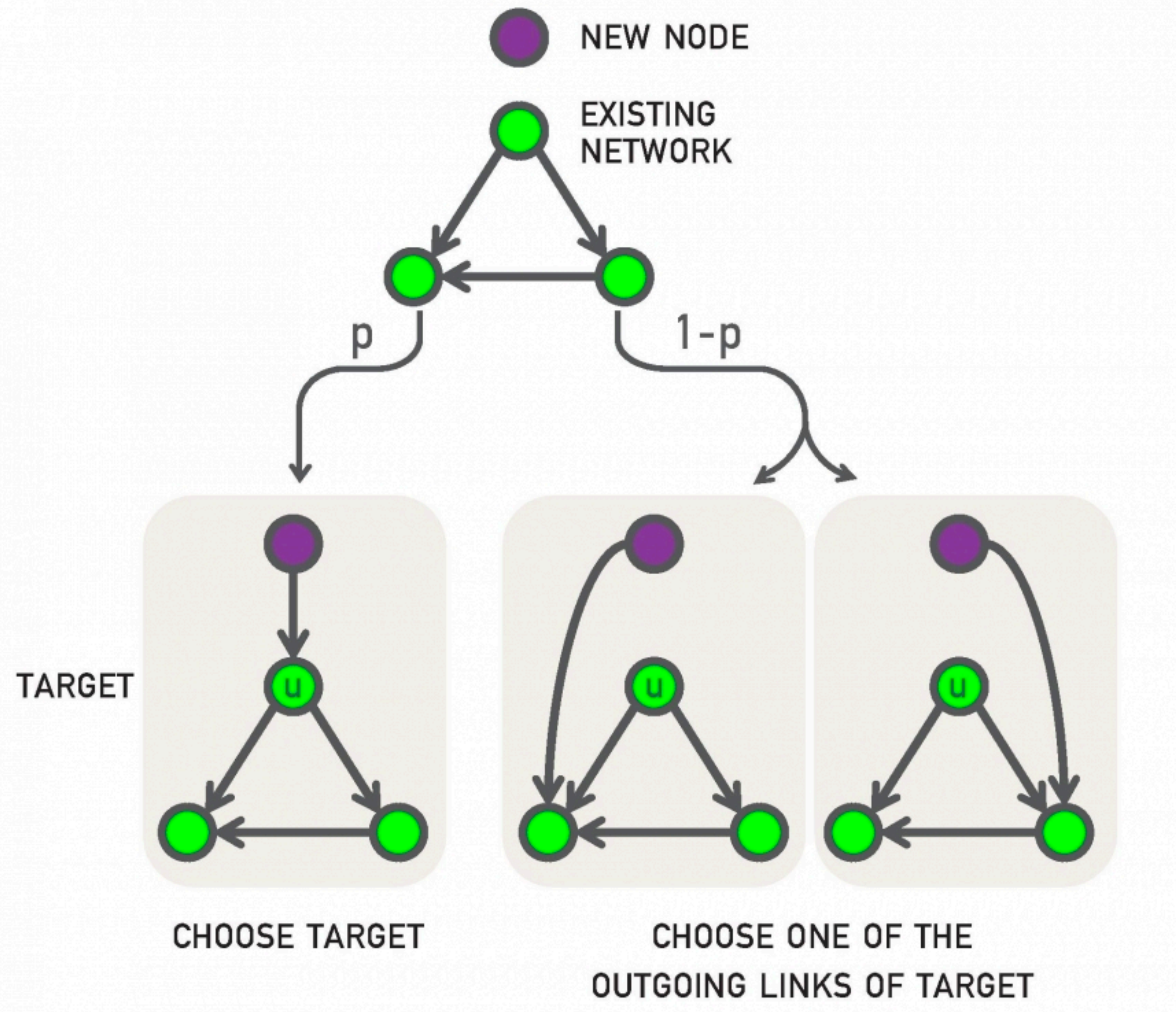
$$C = \frac{1}{<k>} \qquad q_k = \frac{k p_k}{<k>}$$

Excess degree distribution!!!!

Linear in *k*: pref attachment

Barabasi-Albert model

Potential local mechanisms: copy model



- *Random Connection*: With probability *p* the new node links to *u*, which means that we link to the randomly selected web document.
- *Copying*: With probability 1-*p* we randomly choose an *outgoing link* of node *u* and link the new node to the link's target. In other words, the new webpage *copies* a link of node *u* and connects to its target, rather than connecting to node u directly.

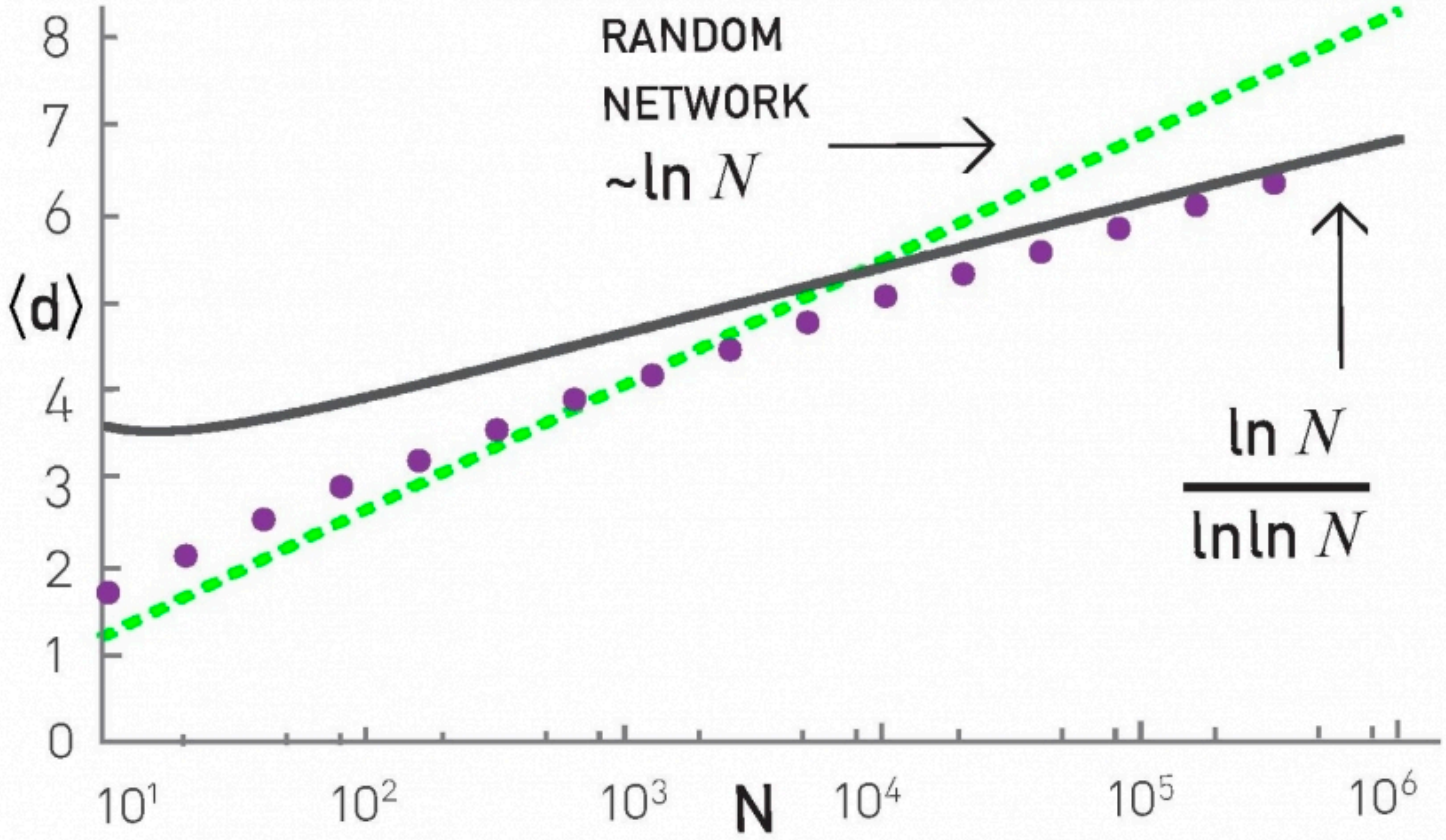
$$\Pi(k) = \frac{p}{N} + \frac{1-p}{2L}k$$

probability of selecting a degree-k node

Barabasi-Albert model

Properties: diameter

$$d \sim \frac{\ln N}{\ln \ln N}$$



Barabasi-Albert model

time when node j arrived with $t_j = j$

Properties: clustering

Node l with degree k_l has a number of triangles:

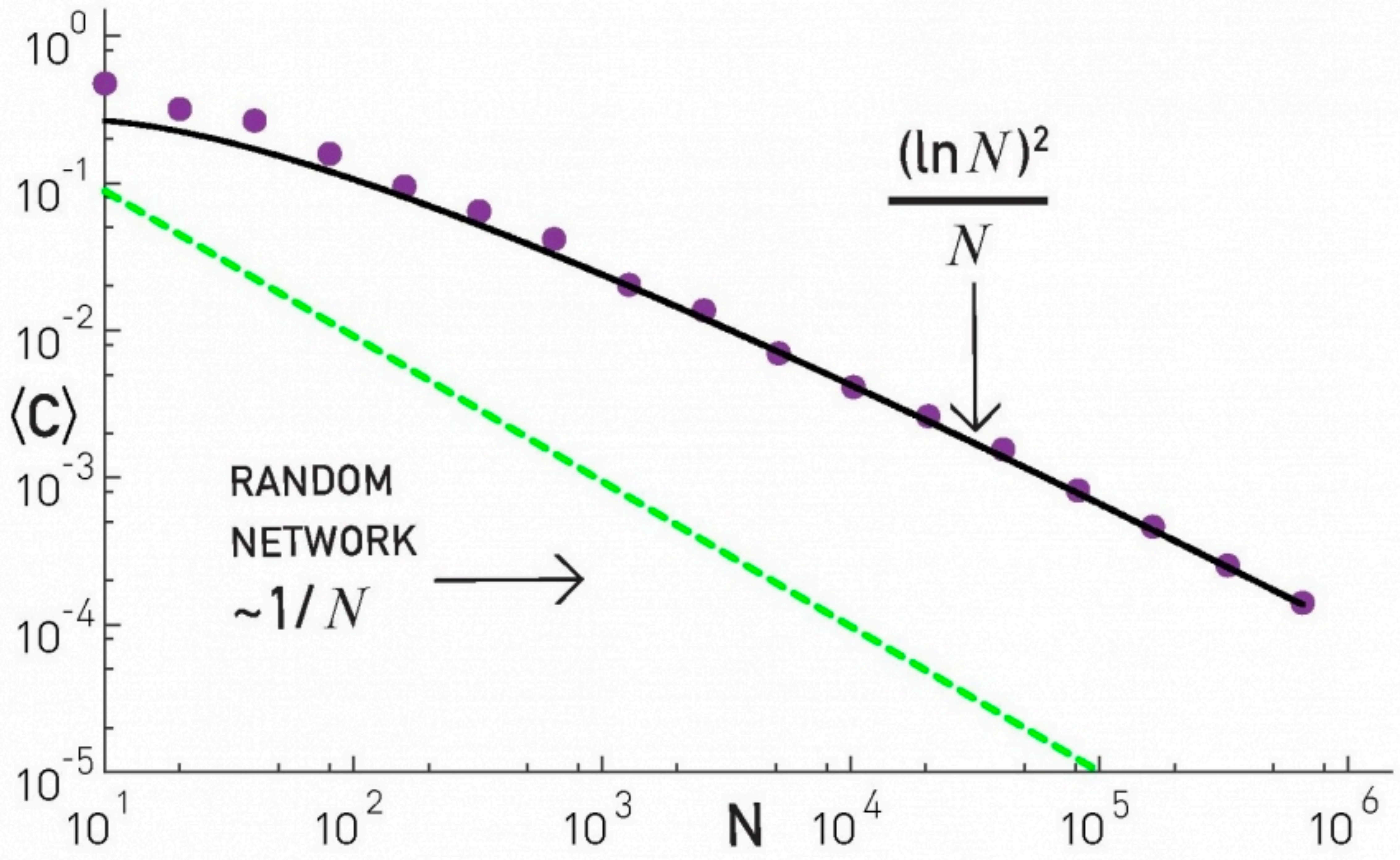
$$N_{r_l} = \int_{j=1}^{N} di \int_{i=1}^{N} dj P(i,j) P(i,l) P(j,l)$$

$$P(i,j) = m\Pi(k_i(j)) = m\frac{k_i(j)}{\sum_{l=1}^{j} k_l(j)} = m\frac{k_i(j)}{2mj}$$

using $\quad k_i(t) = m\left(\frac{t}{t_i}\right)^{\frac{1}{2}} = m\left(\frac{j}{i}\right)^{\frac{1}{2}} \rightarrow P(i,j) = \frac{m}{2}(ij)^{-\frac{1}{2}}$

arrival time of node j is $t_j = j$ and
the arrival time of node i is $t_i = i$

$$N_{r_l} = \frac{m^3}{8l}(\ln N)^2 \quad \rightarrow \quad C = \frac{2N_{r_l}}{k_l(N)(k_l(N)-1)} \simeq \frac{m}{4}\frac{(\ln N)^2}{N}$$



Konstantin Klemm, Victor M. Eguiluz,
Growing scale-free networks with small-world behavior, Phys. Rev. E 65, 057102 (2002),

Barabasi-Albert model

Summary

- Power law with exp -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

- No other realistic methods:
  - Link deletion
  - Internal links
  - Ageing…
  - **Fitness**

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

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

Beyond Barabasi-Albert model

The Bianconi-Barabasi fitness model

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

BA model: first mover advantage!
Can latecomers make it? **Fitness model!**

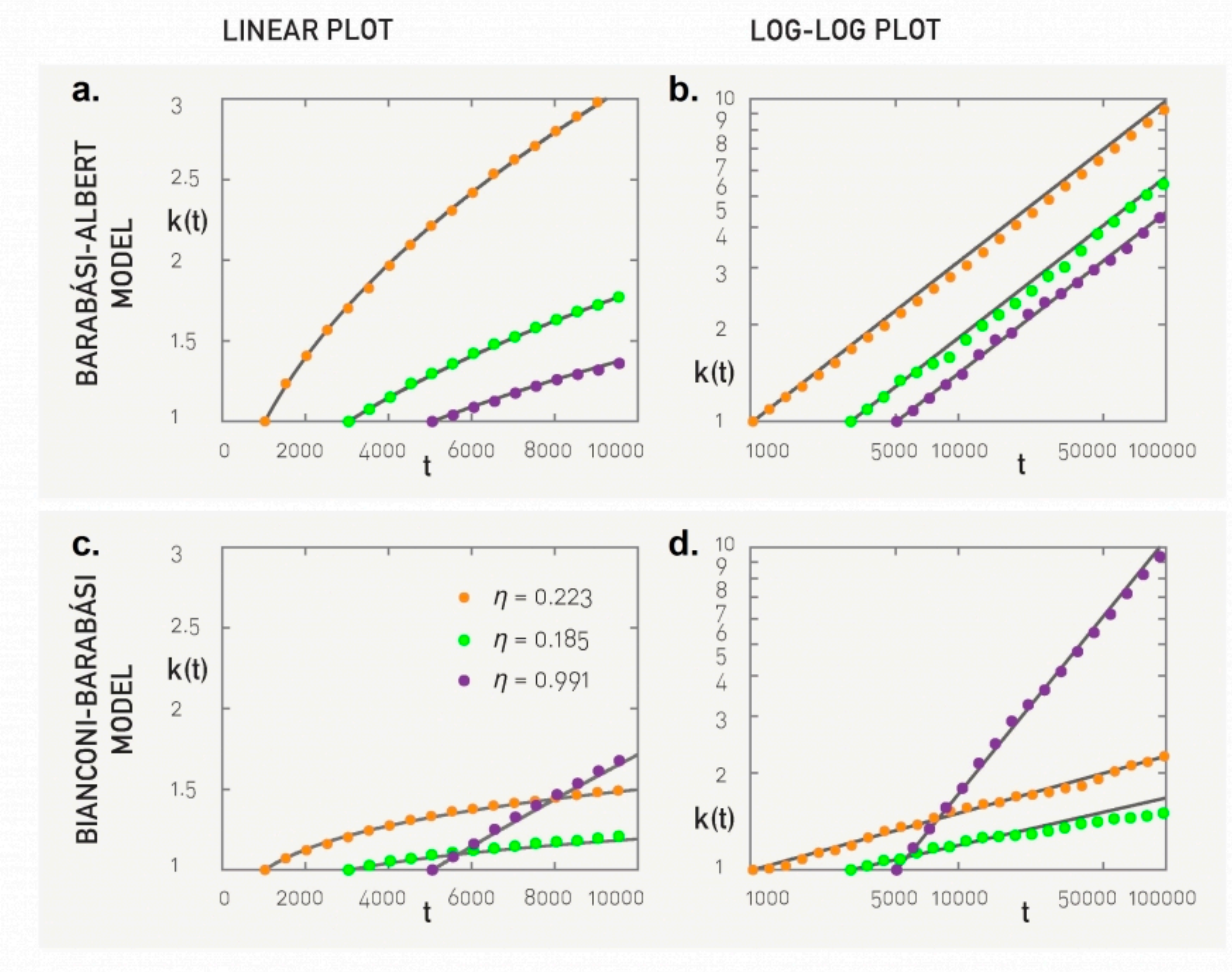
$$\frac{\partial k_i}{\partial t} = m\frac{\eta_i k_i}{\sum_j \eta_j k_j} \qquad \text{ansatz:} \qquad k(t,t_i,\eta_i) = m\left(\frac{t}{t_i}\right)^{\beta(\eta_i)}$$

$$\left\langle \sum_j \eta_j k_j \right\rangle = \int d\eta \rho(\eta)\eta \int_i^t dt_0 k_\eta(t,t_0) = \int d\eta\rho(\eta)\eta m \frac{t - t^{\beta(\eta)}}{1-\beta(\eta)}$$

$$\left\langle \sum_j \eta_j k_j \right\rangle \overset{t\to\infty}{=} Cmt[1 - O(t^{-\epsilon})]_{\epsilon=(1-\max_\eta \beta(\eta))>0}$$

$$\frac{\partial k_\eta}{\partial k} = \frac{\eta k_\eta}{Ct} \qquad \beta(\eta) = \frac{\eta}{C} \qquad 1 = \int_0^{\eta_{max}} d\eta\rho(\eta)\frac{1}{\frac{C}{\eta}-1}$$

$$Cmt = \sum_j k_j \le \eta_{max}\sum_j k_j = 2mt\eta_{max} \qquad \to C \le 2\eta_{max}, \quad C > \eta_{max}$$



Beyond Barabasi-Albert model

Derivation degree distribution

Number of nodes with degree larger than k

$$t_0 < t\left(\frac{m}{k}\right)^{C/\eta}$$

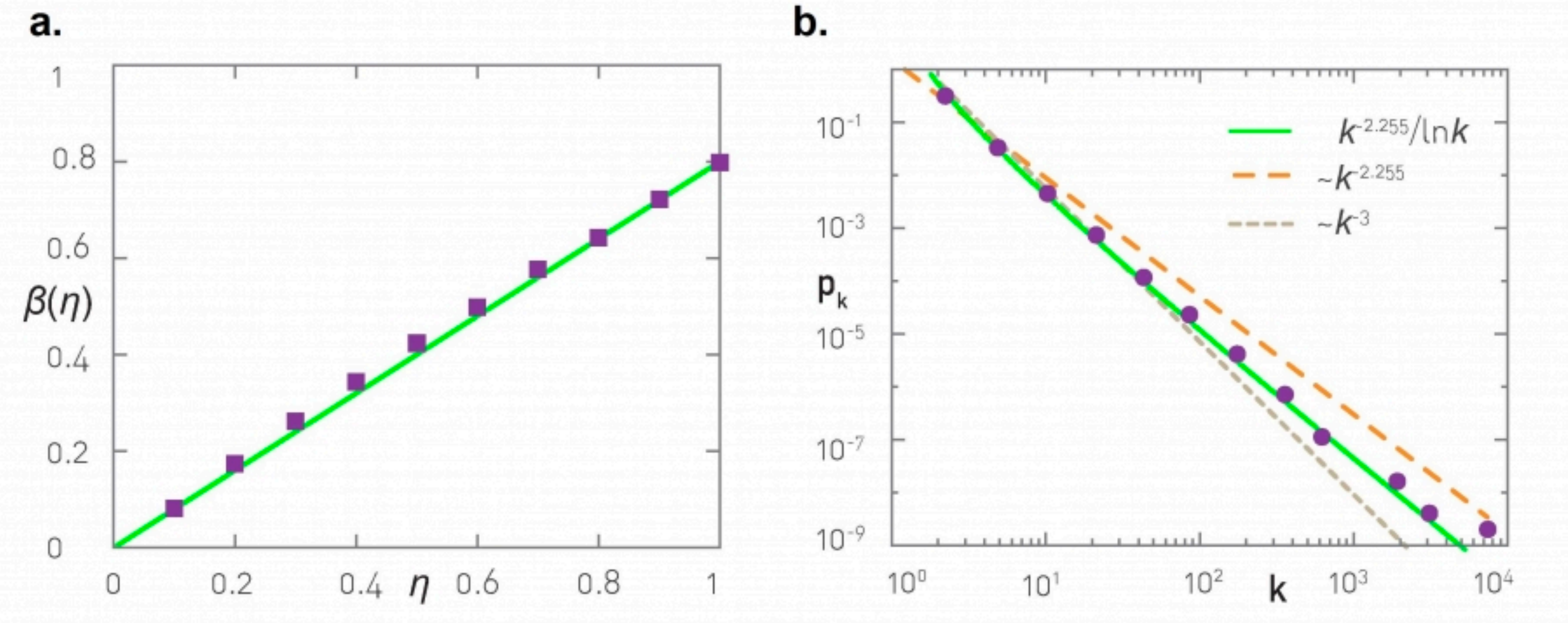
$$P(k_i < k) = 1 - P(k_i > k) = 1 - \frac{1}{m_0 + t}\int_0^{\eta_{max}} t\left(\frac{m}{k}\right)^{C/\eta}\rho(\eta)d\eta$$
$$\sim 1 - \int_0^{\eta_{max}} \left(\frac{m}{k}\right)^{C/\eta}\rho(\eta)d\eta$$

$$p(k) = \frac{\partial P(k)}{\partial k} = \int_0^{\eta_{max}} \frac{C}{\eta} m^{C/\eta} k^{-C/\eta+1)}\rho(\eta)d\eta$$

Beyond Barabasi-Albert model

Fitness examples: uniform fitness

$$\eta \in [0,1] \qquad e^{-2/C} = 1 - \frac{1}{C} \quad \rightarrow \quad C^* \sim 1.255 \quad \beta(\eta_i) = \frac{\eta_i}{C^*} p_k \sim \int_0^1 d\eta \frac{C^*}{\eta} \frac{1}{k^{1+C^*/\eta}} \sim \frac{k^{-(1+C^*)}}{\ln k}$$



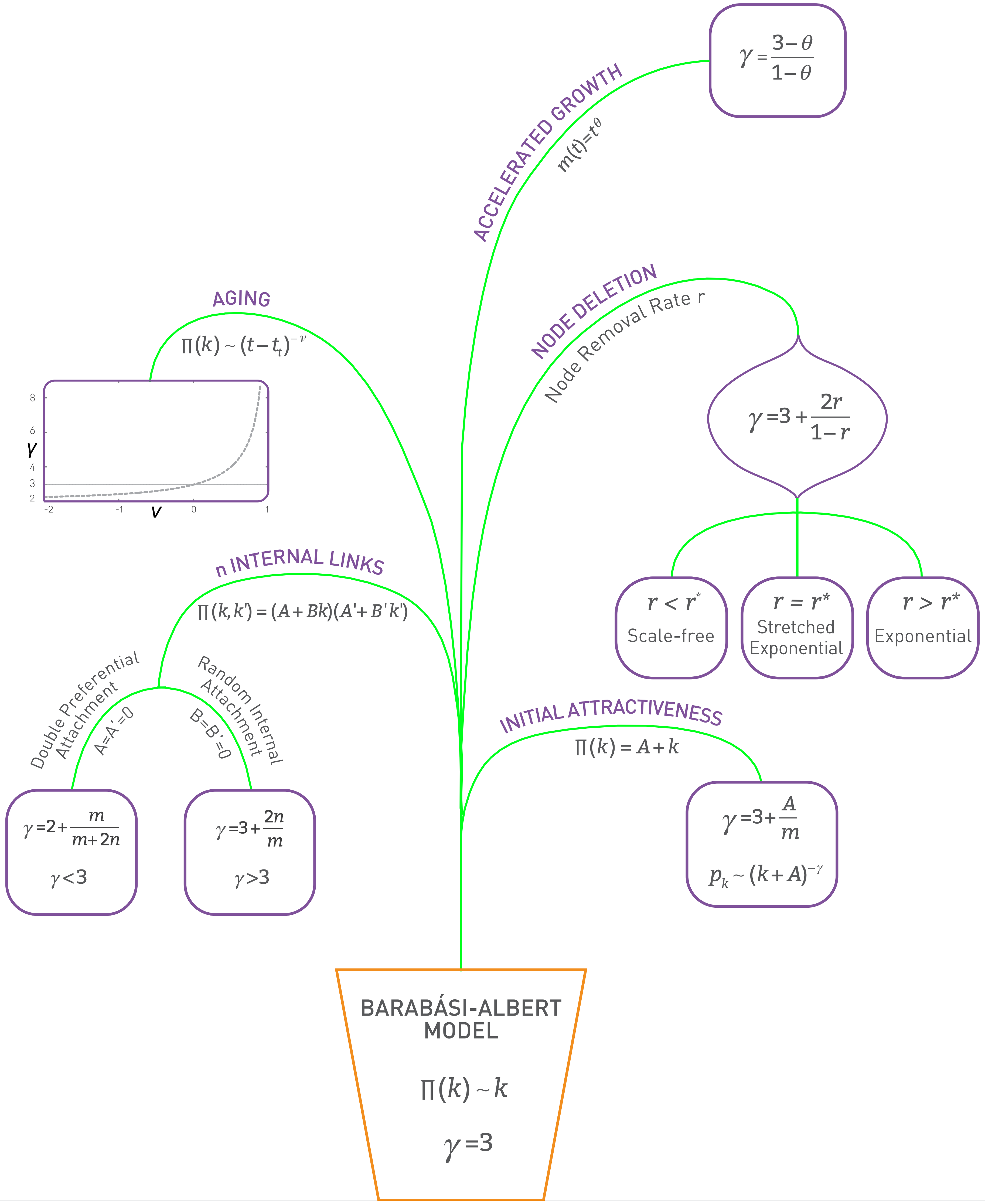
Beyond Barabasi-Albert model

Fitness examples: equal fitness

?

Beyond Barabasi-Albert model

Limitations and extensions



MODEL CLASS	EXAMPLES	CHARACTERISTICS
Static Models	✓ Erdős–Rényi ✓ Watts-Strogatz	• N fixed • p_k exponentially bounded • Static, time independent topologies
Generative Models	✓ Configuration Model Hidden Parameter Model	• Arbitrary pre-defined p_k • Static, time independent topologies
Evolving Network Models	✓ Barabási–Albert Model ✓ Bianconi-Barabási Model Initial Attractiveness Model Internal Links Model Node Deletion Model Accelerated Growth Model Aging Model	• p_k is determined by the processes that contribute to the network's evolution. •Time-varying network topologies

Introduction to network correlations

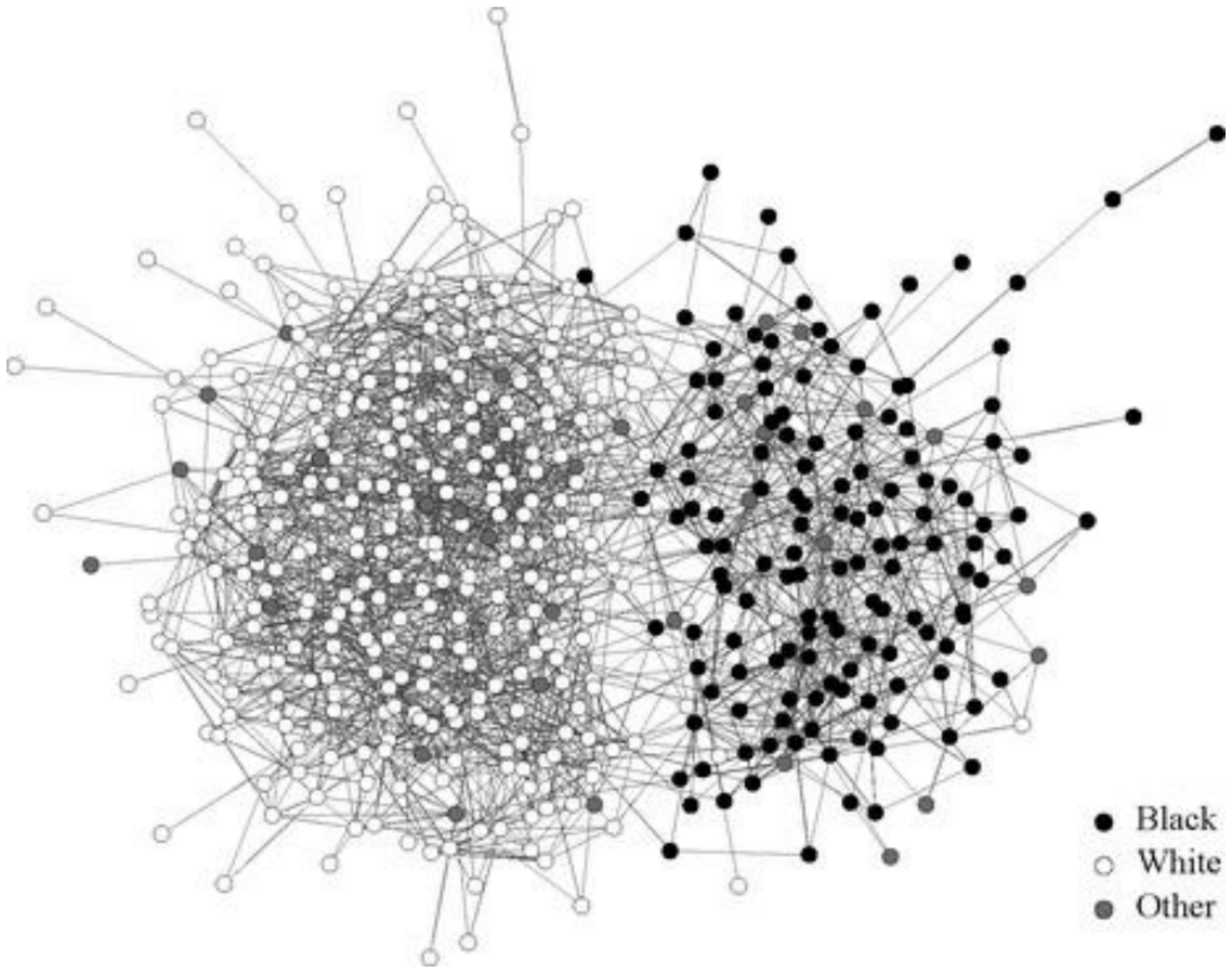
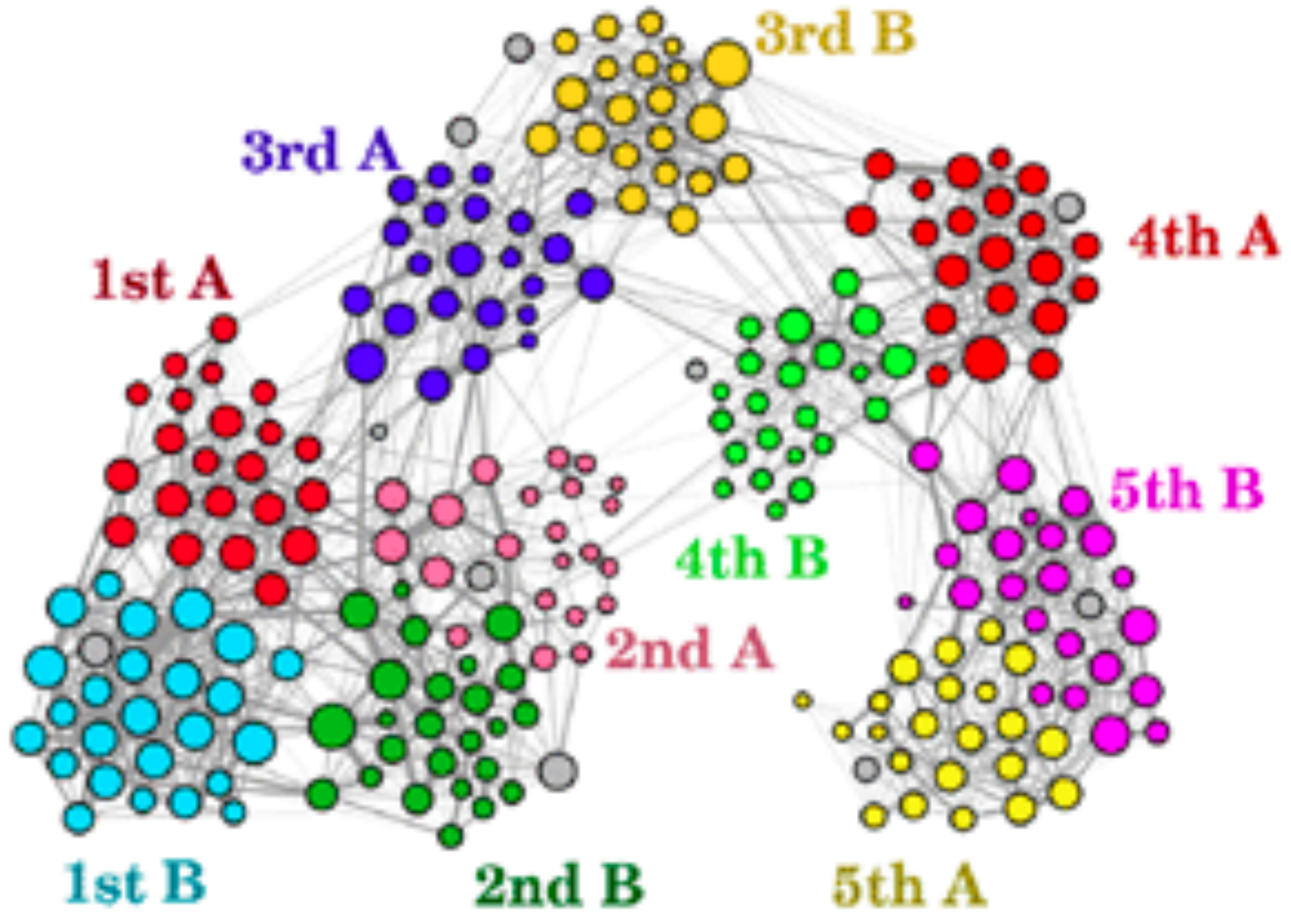


Figure 7.10: Friendship network at a US high school. The vertices in this network represent 470 students at a US high school (ages 14 to 18 years). The vertices are color coded by race as indicated in the key. Data from the National Longitudinal Study of Adolescent Health [34, 314].

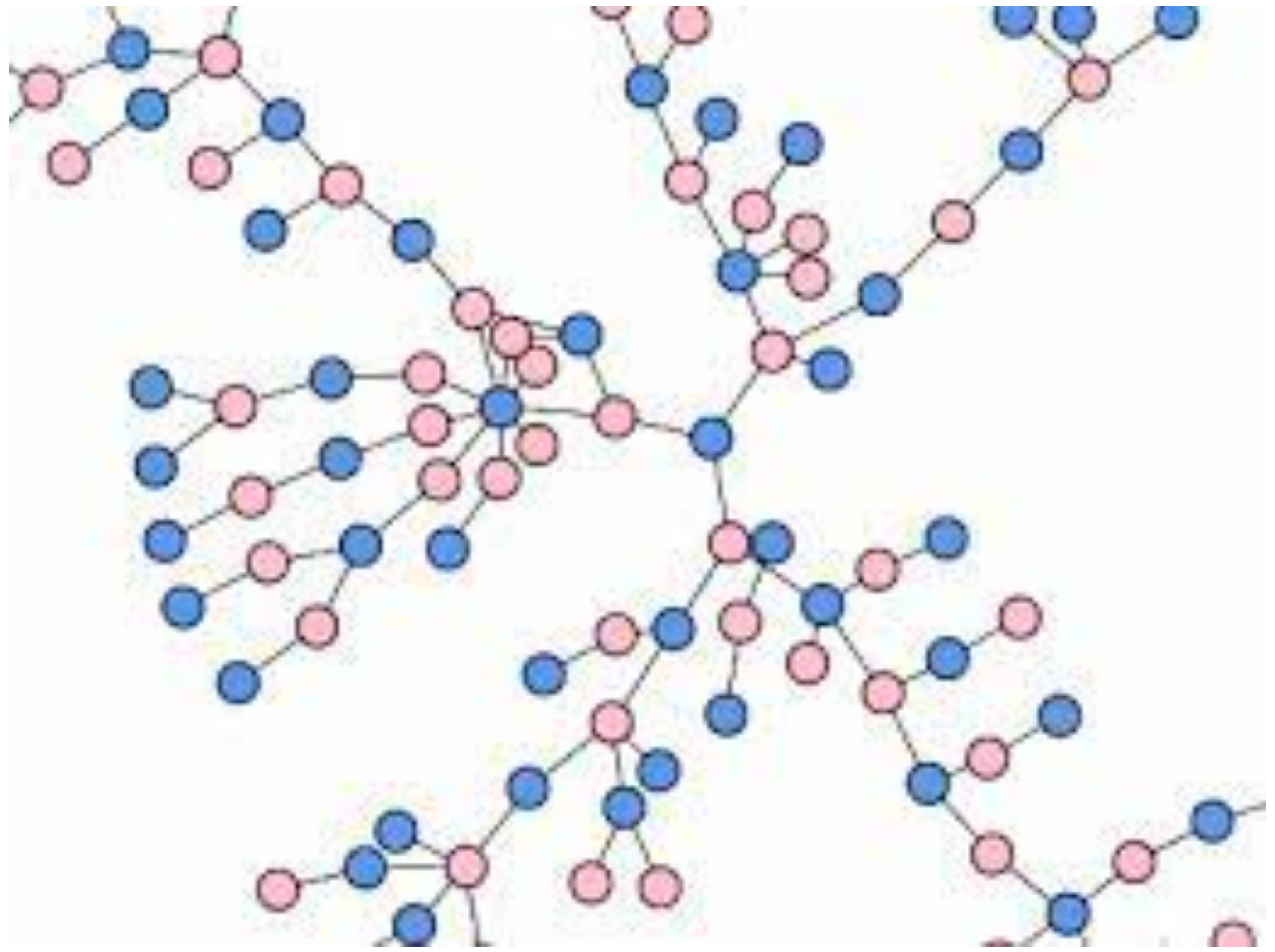
Introduction to network correlations

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



Enumerative assortativity

Given c_i class or type of vertex i (1,...,n_C = total number of classes), then the total number of edges that run between vertices of the same type is:

$$\sum_{edges(i,j)} \delta(c_i, c_j) = \frac{1}{2} \sum_{ij} A_{ij} \delta(c_i, c_j)$$

However, we want to control for the random expectation of the mixing: $\frac{1}{2} \sum ij \frac{k_i k_j}{2m} \delta(c_i, c_j)$

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j)$$

Modularity: It is strictly less than 1, takes positive values if there are more edges between vertices of the same type than we would expect by chance, and negative ones if there are less.

Enumerative assortativity

How can we normalise it?

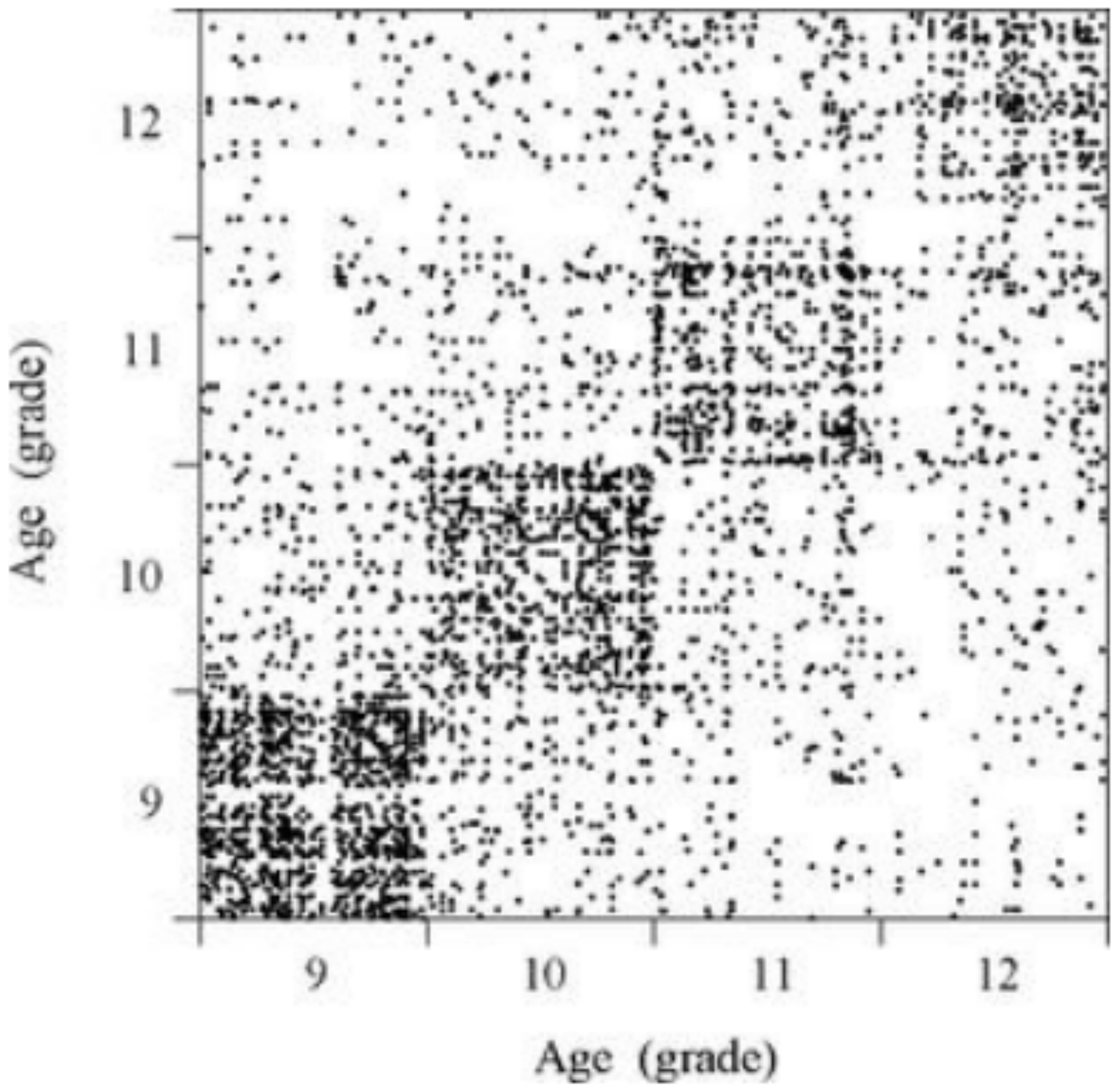
Obtained for full mixing:

$$Q_{max} = \frac{1}{2m}\left(2m - \sum_{ij} \frac{k_i k_j}{2m}\delta(c_i, c_j)\right)$$

Assortativity coefficient

$$r = \frac{Q}{Q_{max}} = \frac{\sum_{ij}\left(A_{ij} - \frac{k_i k_j}{2m}\right)\delta(c_i, c_j)}{\left(2m - \sum_{ij}\frac{k_i k_j}{2m}\delta(c_i, c_j)\right)}$$

Scalar assortativity



$$\mu = \frac{\sum_{ij} A_{ij} x_i}{\sum_{ij} A_{ij}} = \frac{\sum_i k_i x_i}{\sum_i k_i} = \frac{1}{2m} \sum_i k_i x_i$$

$$cov(x_i, x_j) = \frac{\sum_{ij} A_{ij}(x_i - \mu)(x_j - \mu)}{\sum_{ij} A_{ij}}$$

$$= \frac{1}{2m} \sum_{ij} (A_{ij} - \frac{k_i k_j}{2m}) x_i x_j$$

$$r = \frac{\sum_{ij}(A_{ij} - k_i k_j / 2m) x_i x_j}{\sum_{ij}(k_i \delta_{ij} - k_i k_j / 2m) x_i x_j}$$

Degree correlations

Degree assortativity coefficient

$$r = \frac{\sum_{ij}(A_{ij} - k_i k_j/2m)k_i k_j}{\sum_{ij}(k_i \delta_{ij} - k_i k_j/2m)k_i k_j}$$

It can be misleading when:

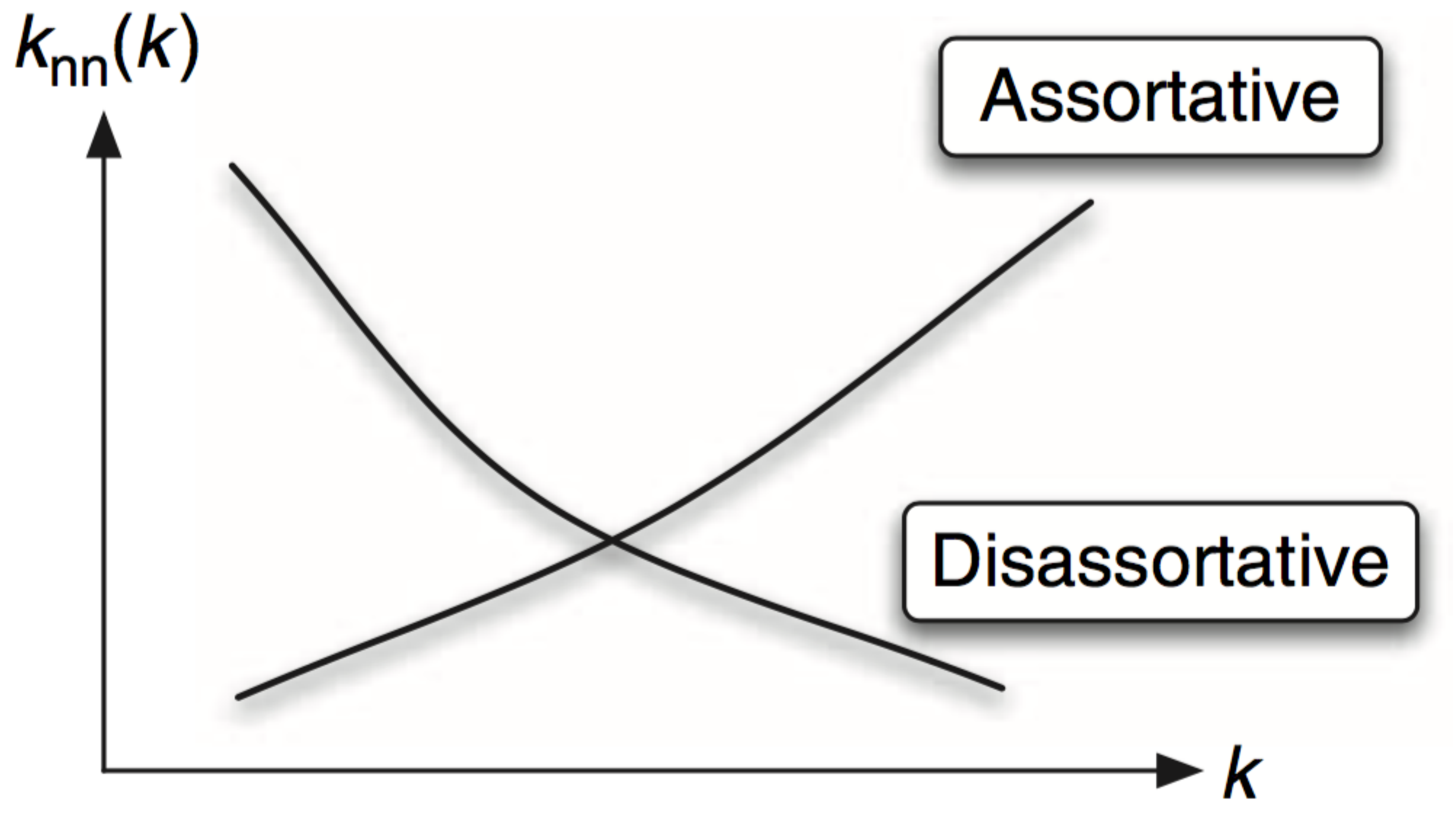
- complicated behavior of the correlation functions (non-monotonous behavior)
- Pearson coefficient gives a larger weight to the more abundant degree classes

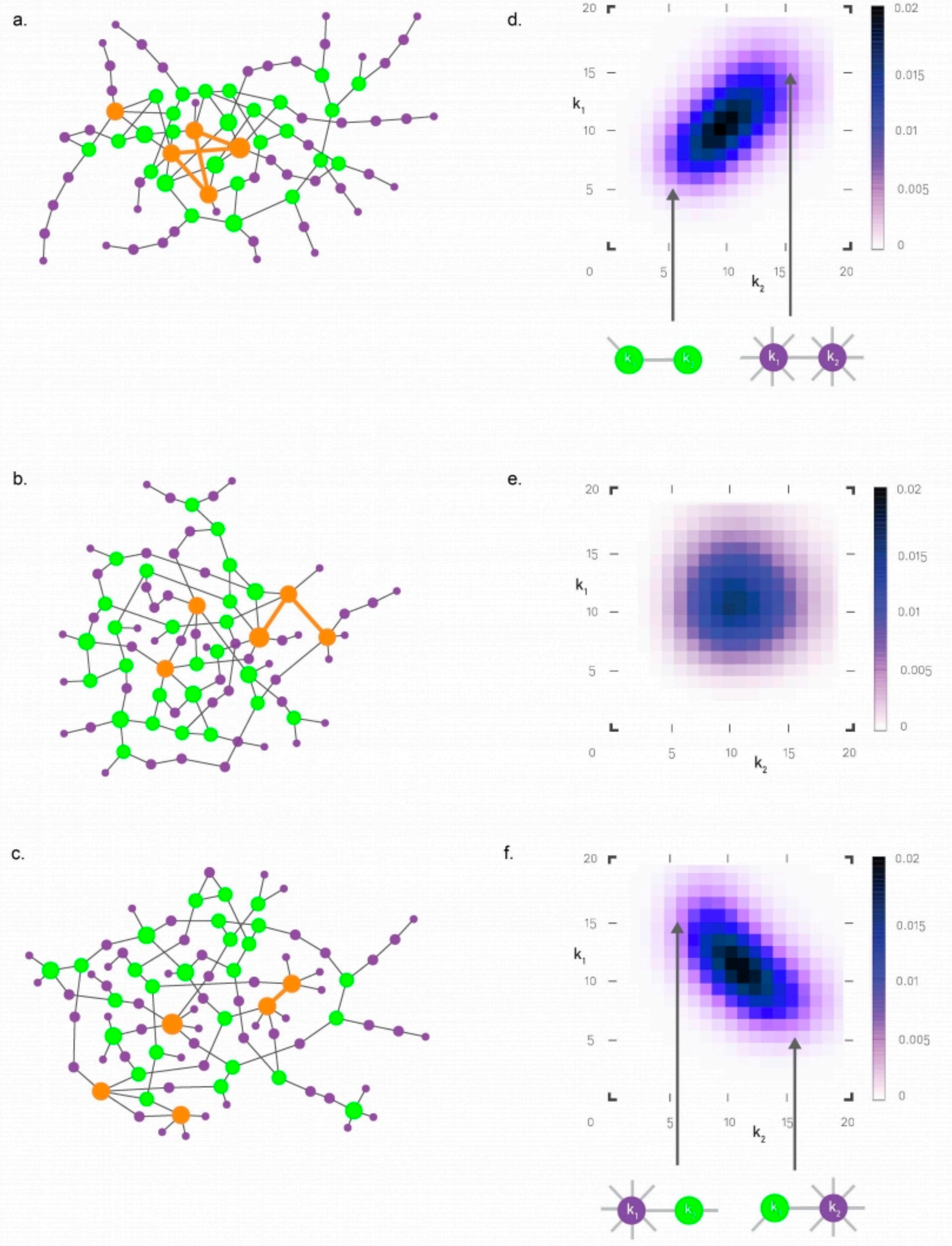
Average nearest neighbour degree

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

Degree correlations

What are the possible scenarios?



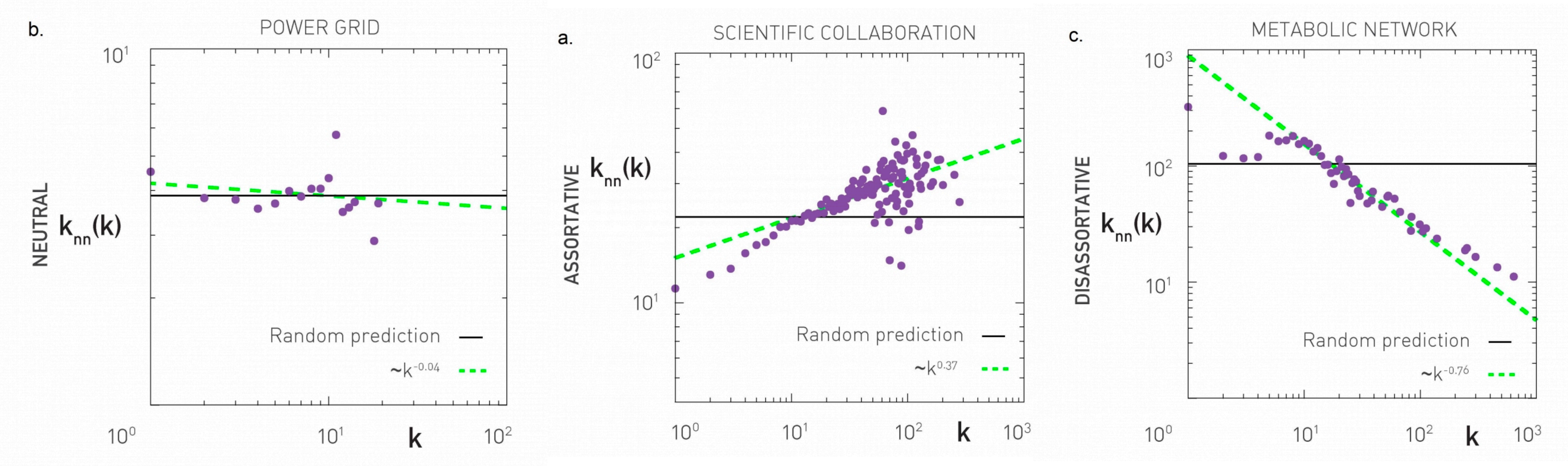


Degree correlations

How do we decide what's expected?

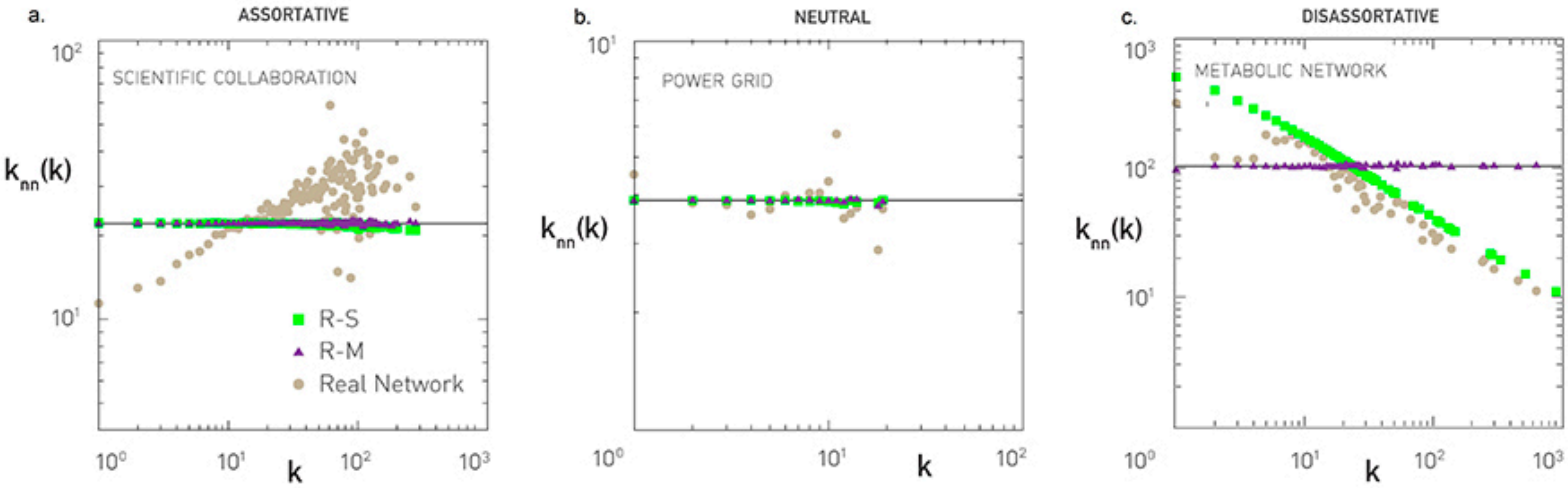
Assume uncorrelated network:

$$k_{nn}^{unc} = \sum_{k'} k' P_{unc}(k'|k) = \sum_{k'} k' \frac{k' p(k')}{< k >} = \frac{< k^2 >}{< k >}$$



Degree correlations

Origins: real or structural ?

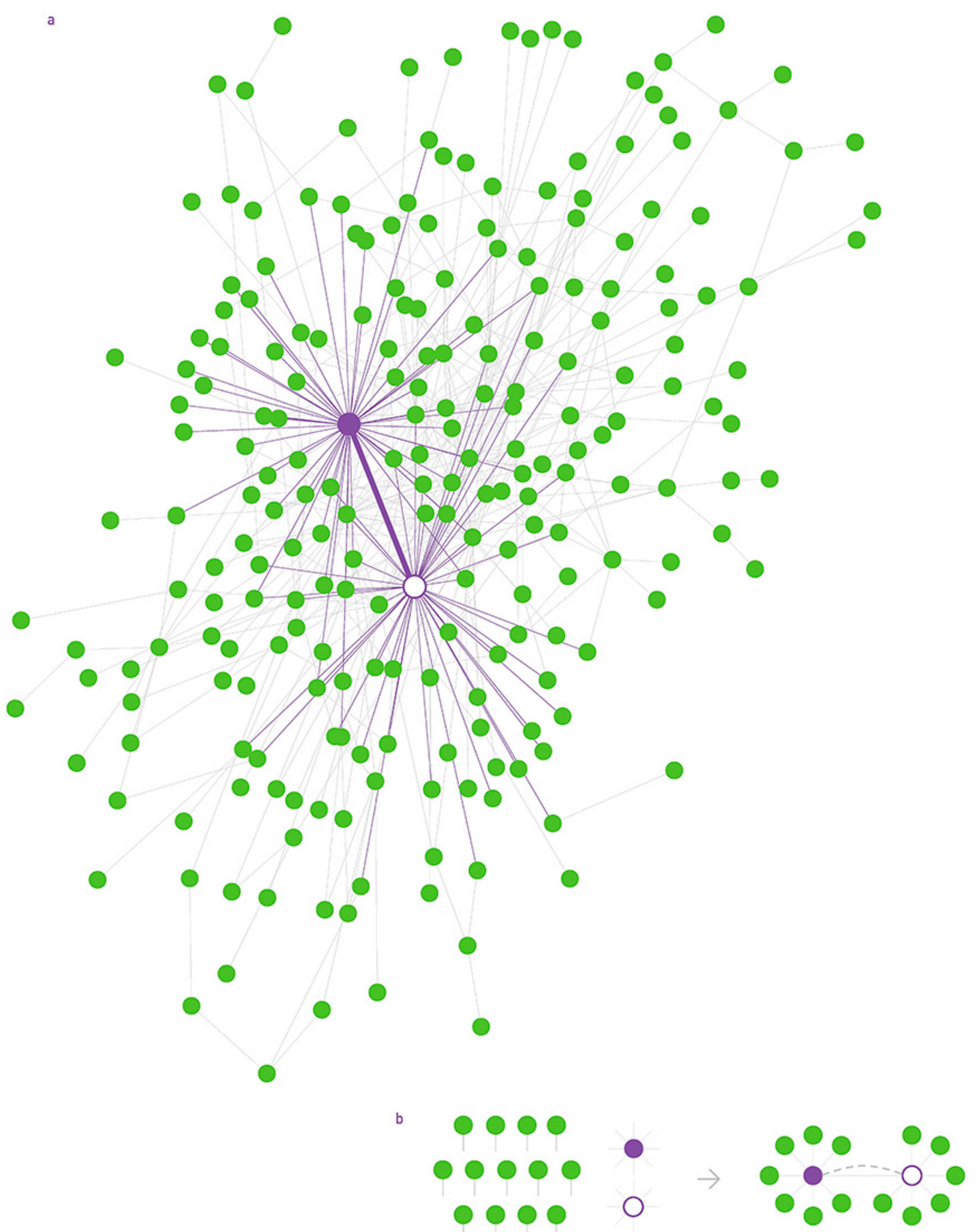


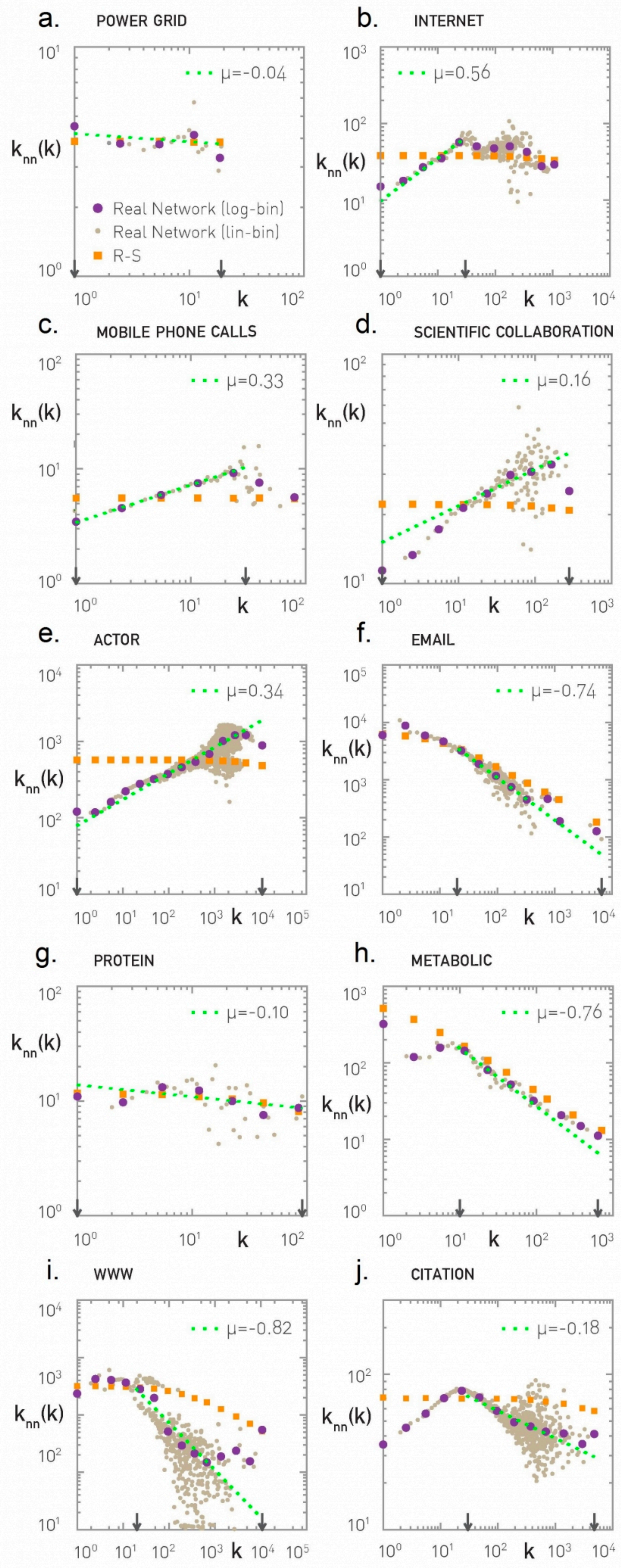
Degree Preserving Randomization with Simple Links (R-S)
We apply degree-preserving randomization to the original network and at each step we make sure that we do not permit more than one link between a pair of nodes. On the algorithmic side this means that each rewiring that generates multi-links is discarded.

Degree Preserving Randomization with Multiple Links (R-M)
For a self-consistency check it is sometimes useful to perform degree-preserving randomization that allows for multiple links between the nodes. On the algorithmic side this means that we allow each random rewiring, even if it leads to multi-links

Degree correlations

Origins: real or structural ?





Recap today's topics

Models

Barabasi-Albert model
Bianconi-Barabasi model
Link/Copying model

Concepts

Origins of scale-free distributions
Growing networks
Assortativity and correlations
"Robustness"

Next time

Percolation!
Spectral properties
Random Walks