

Frontiers of Network Science

Fall 2021

Class 14: Barabasi-Albert Model II **(Chapter 5 in Textbook)**

Boleslaw Szymanski



MECHANISMS RESPONSIBLE FOR PREFERENTIAL ATTACHMENT

1. Copying mechanism
 - directed network
 - select a node and an edge of this node
 - attach to the endpoint of this edge
2. Walking on a network
 - directed network
 - the new node connects to a node, then to every first, second, ... neighbor of this node
3. Attaching to edges
 - select an edge
 - attach to both endpoints of this edge
4. Node duplication
 - duplicate a node with all its edges
 - randomly prune edges of new node



Section 9

Copying model

(a) Random Connection: with probability p the new node links to u .

(b) Copying: with probability p we randomly choose an outgoing link of node u and connect the new node to the selected link's target. Hence the new node “copies” one of the links of an earlier node

(a) the probability of selecting a node is $1/N$.

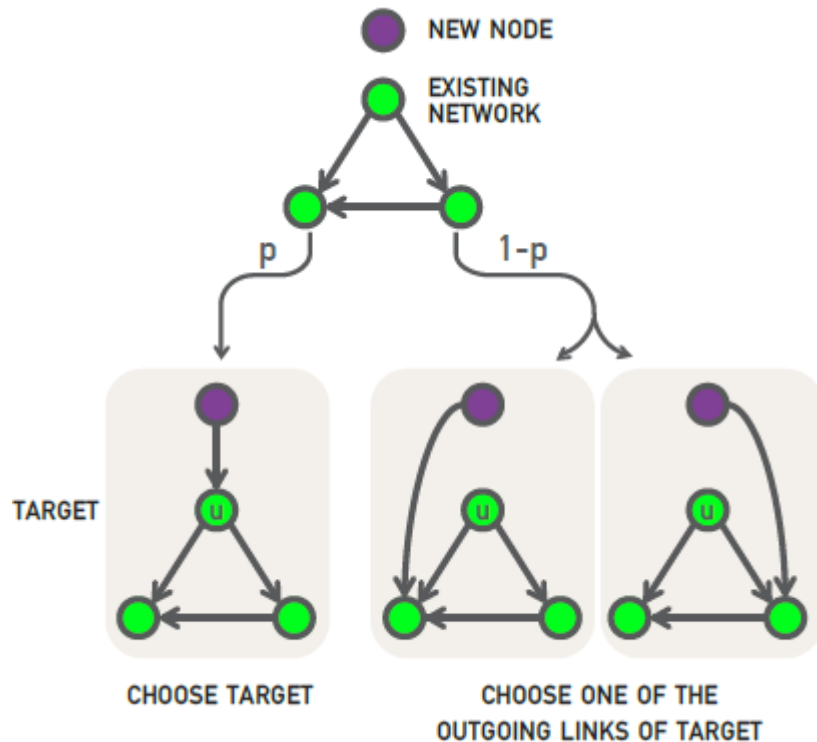
(b) is equivalent with selecting a node linked to a randomly selected link. The probability of selecting a degree- k node through the copying process of step (b) is $k/2L$ for undirected networks.

The likelihood that the new node will connect to a degree- k node follows preferential attachment

Social networks: Copy your friend's friends.

Citation Networks: Copy references from papers we read.

Protein interaction networks: gene duplication,



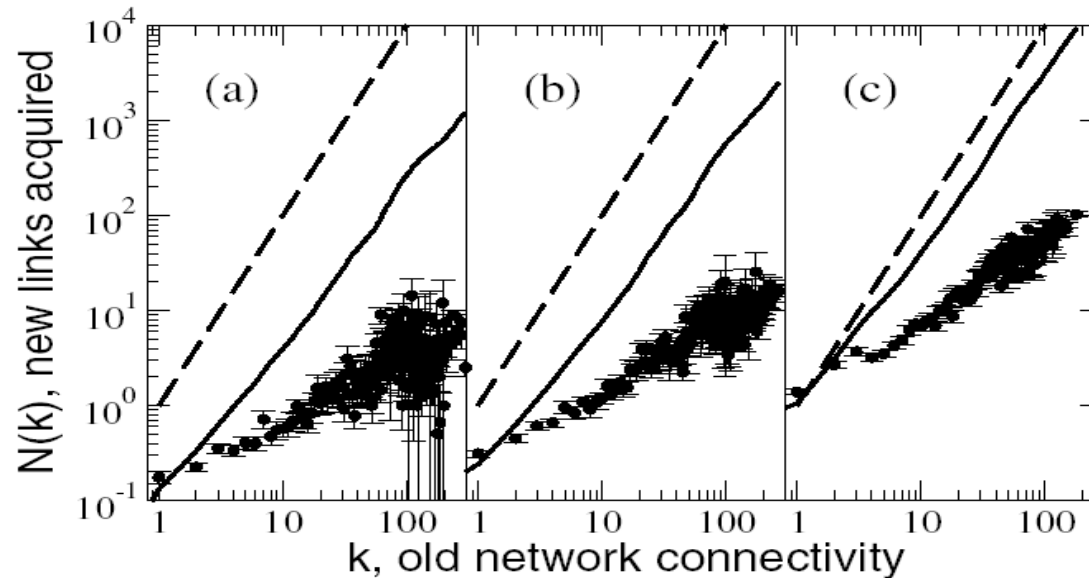
$$\Pi(k) = p / N + (1 - p)k / (2L)$$

Preferential Attachment!

$$\frac{\partial k_i}{\partial t} \propto \Pi(k_i) \sim \frac{\Delta k_i}{\Delta t}$$

For given Δt : $\Delta k \propto \Pi(k)$

k vs. Δk : linear increase in the # of links



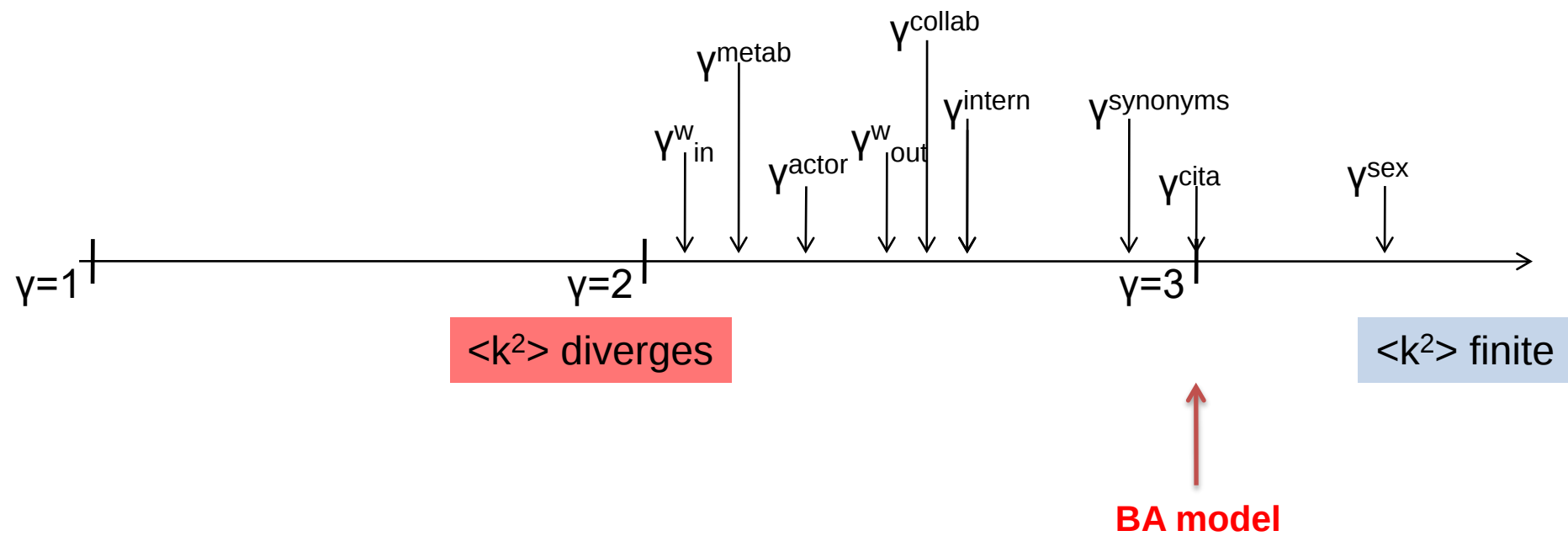
S. Cerevisiae PIN: proteins classified into 4 age groups

SUMMARY: PROPERTIES OF THE BA MODEL

- **Nr. of nodes:** $N = t$
- **Nr. of links:** $L = m t$
- **Average degree:** $\langle k \rangle = \frac{2L}{N} \rightarrow 2m$
- **Degree dynamics** $k_i(t) = m \left(\frac{t}{t_i} \right)^\beta$ $\beta = \frac{1}{2}$ β : dynamical exponent
- **Degree distribution:** $P(k) \sim k^{-\gamma}$ $\gamma = 3$ γ : degree exponent
- **Average Path Length:** $l \approx \frac{\ln N}{\ln \ln N}$
- **Clustering Coefficient:** $C \sim \frac{(\ln N)^2}{N}$

The network grows, but the degree distribution is stationary.

DEGREE EXPONENTS



Can we change the degree exponent?



Section 9

Optimization model for connecting new router

(a)

(b)

h_j denotes the distance of node j to the central node h_0

d_{ij} denotes the bandwidth between nodes i, j

δ denotes the ratio of cost of cable to delay

Question: where to place a new router ?

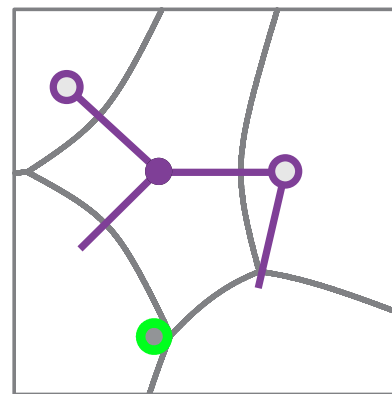
$$C_i = \min_j [\delta d_{ij} + h_j]$$

(c)

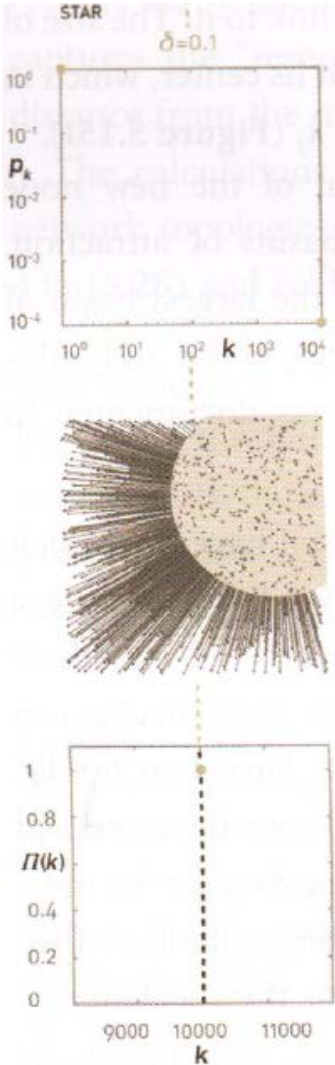
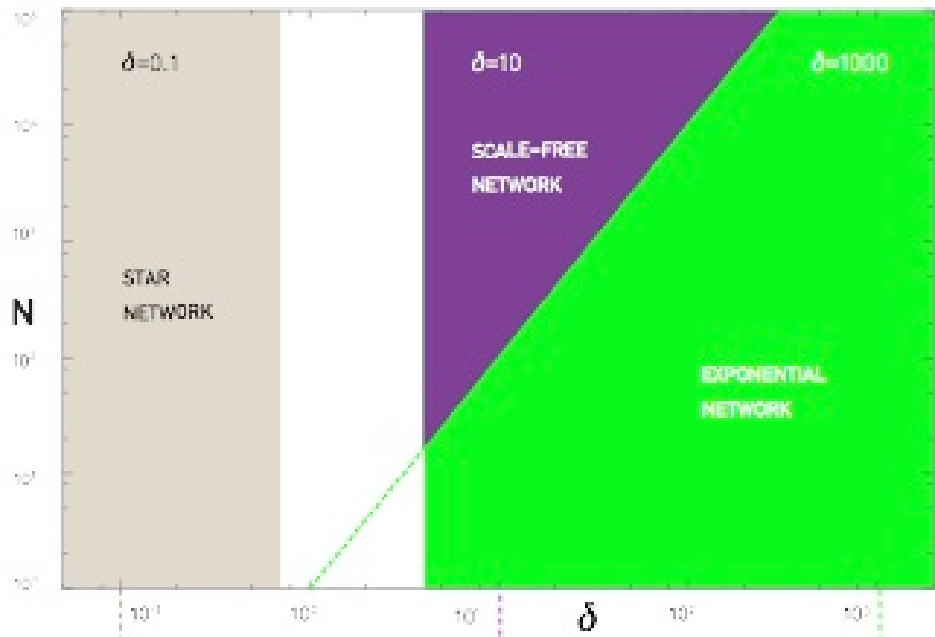
(d)

(e)

(f)



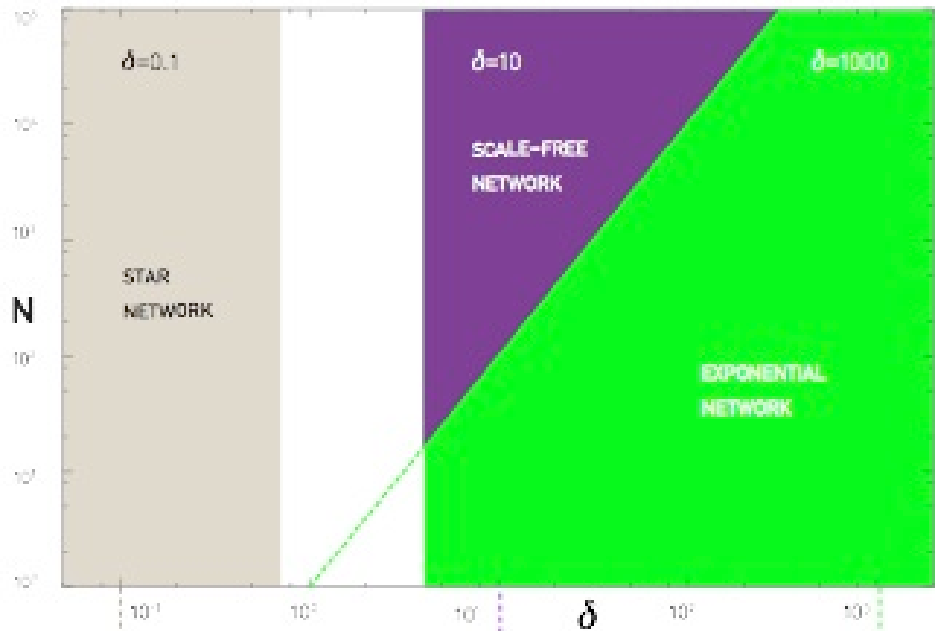
$$C_i = \min_j [\delta d_{ij} + h_j]$$



Star Network

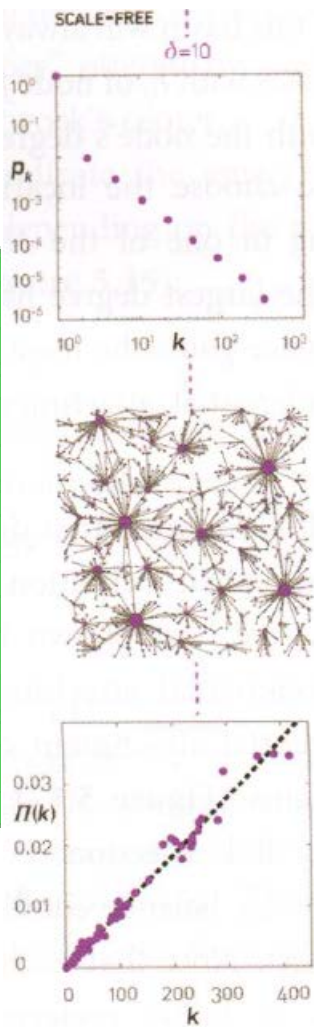
The vertical boundary of the star configuration is at $\delta = (1/2)^{1/2}$. This is the inverse of the maximum distance between two nodes on a square lattice with unit length, over which the model is defined. Therefore, if $\delta < (1/2)^{1/2}$, for any new node $\delta d_{ij} < 1$ and the cost (5.28) of connecting to the central node is $C_i = \delta d_{ij} + 0$, always lower than connecting to any other node at a cost of $f(i,j) = \delta d_{ij} + 1$. Therefore, for $\delta < (1/2)^{1/2}$ all nodes connect to node 0 (star-and-spoke network (c)).

$$C_i = \min_j [\delta d_{ij} + h_j]$$

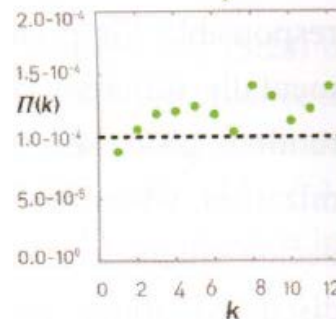
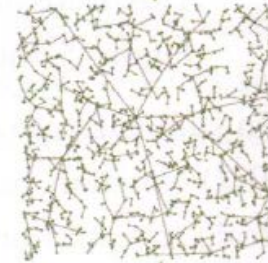
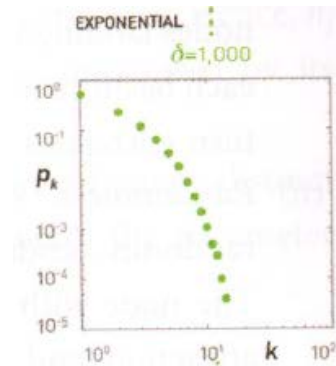
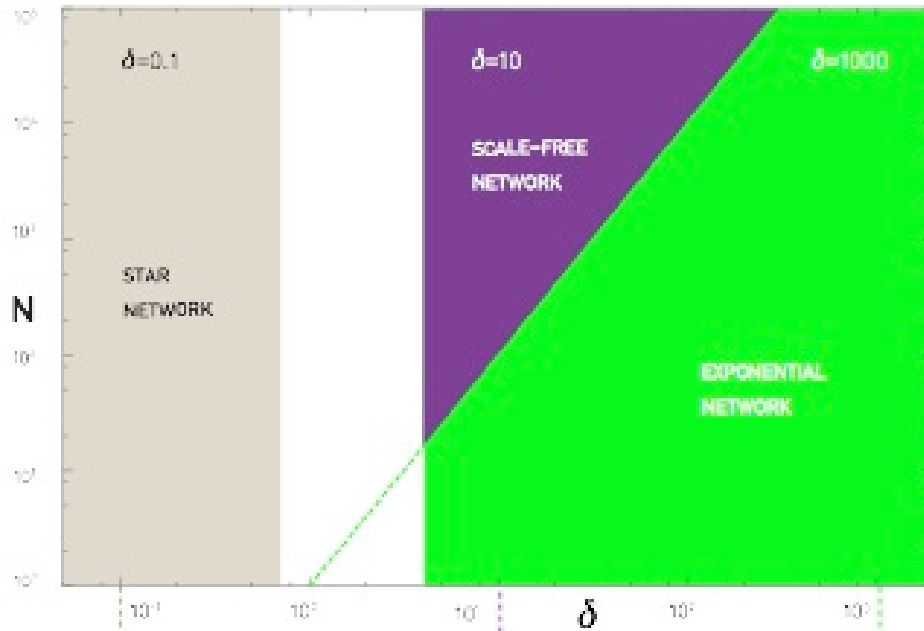


Scale-Free Network

The oblique boundary of the scale-free regime is $\delta = N^{1/2}$. Indeed, if nodes are placed randomly on the unit square, then the typical distance between neighbors decreases as $N^{-1/2}$. Hence, if $d_{ij} \sim N^{-1/2}$ then $\delta d_{ij} \geq h_{ij}$ for most node pairs. Typically the path length to the central node h_j grows slower than N (in small-world networks $h \sim \log N$). Therefore, C_i is dominated by the δd_{ij} term and the smallest C_i is achieved by minimizing the distance-dependent term. Note that, strictly speaking, the transition only occurs in the $N \rightarrow \infty$ limit. In the white regime we lack an analytical form for the degree distribution.



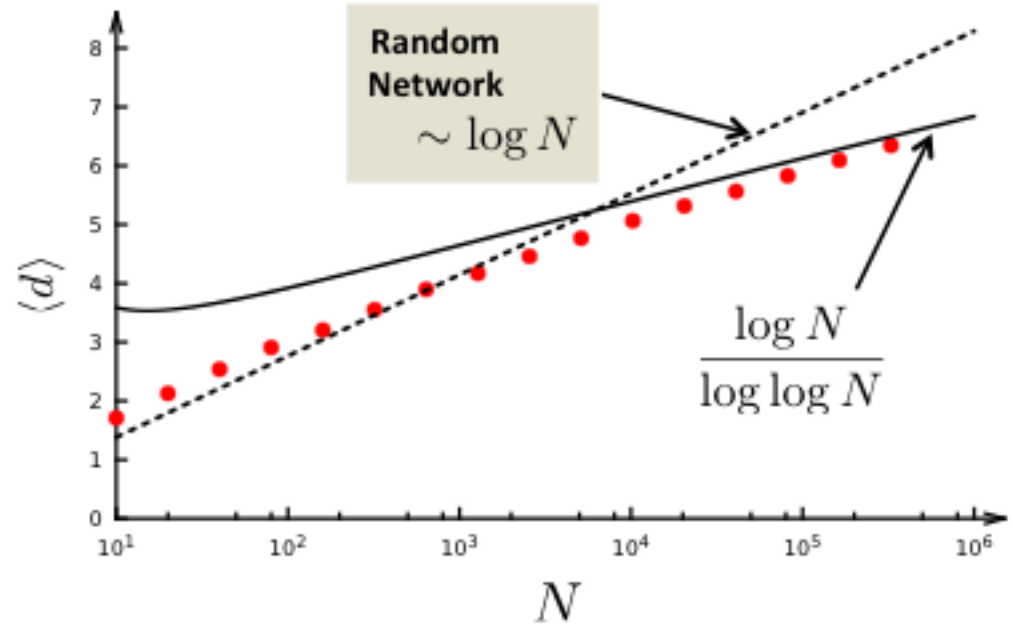
$$C_i = \min_j [\delta d_{ij} + h_j]$$



Exponential Networks

Diameter and clustering coefficient

$$D \sim \frac{\log N}{\log \log N}$$

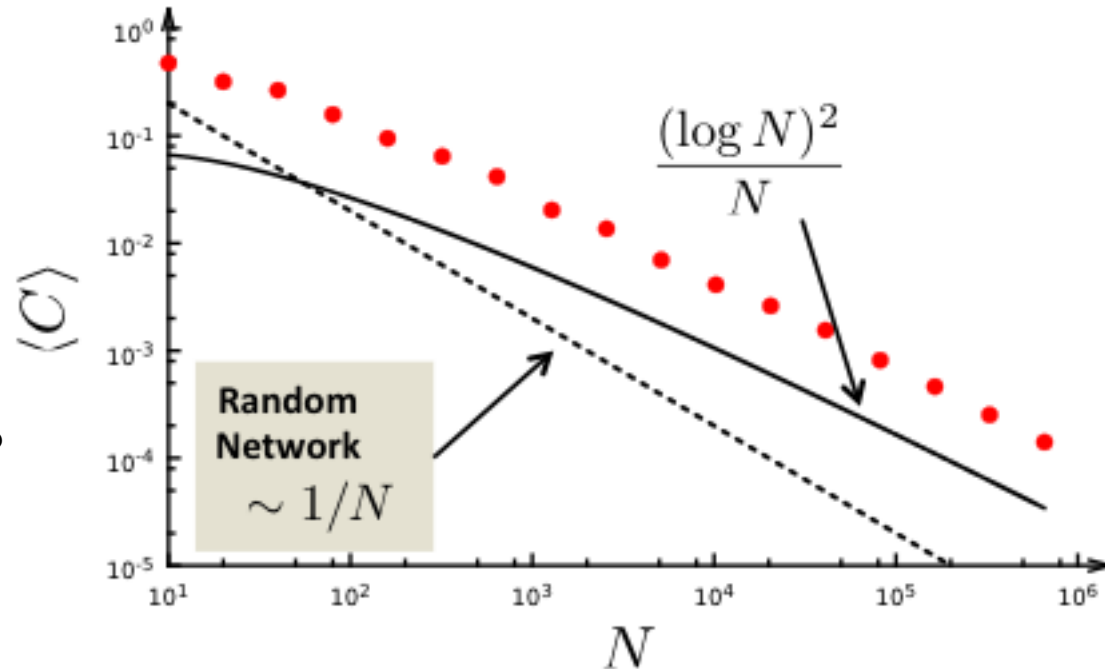


Reminder: for a random graph we have:

$$C_{rand} = \frac{\langle k \rangle}{N} \sim N^{-1}$$

What is the functional form of $C(N)$?

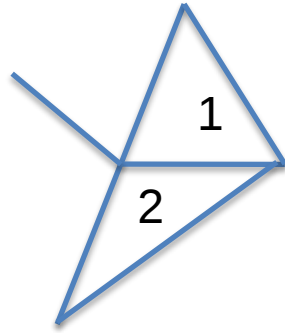
$$C = \frac{m}{8} \frac{(\ln N)^2}{N}$$



Konstantin Klemm, Victor M. Eguiluz,
 Growing scale-free networks with small-world behavior,
 Phys. Rev. E 65, 057102 (2002), cond-mat/0107607

CLUSTERING COEFFICIENT OF THE BA MODEL

$$C = \frac{Nr(\Delta)}{\frac{k(k-1)}{2}}$$



$$C = \frac{2}{6}$$

Denote the probability to have a link between node i and j with $P(i,j)$

The probability that three nodes i,j,l form a triangle is $P(i,j)P(i,l)P(j,l)$

The expected number of triangles in which a node l with degree k_l participates is thus:

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

We need to calculate $P(i,j)$.

CLUSTERING COEFFICIENT OF THE BA MODEL

Calculate P(i,j).

Node j arrives at time $t_j=j$ and the probability that it will link to node i with degree k_i already in the network is determined by preferential attachment:

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

$$k_i(t) = m \left(\frac{t}{t_i} \right)^{1/2} = m \left(\frac{j}{i} \right)^{1/2}$$

Where we used that the arrival time of node j is $t_j=j$ and the arrival time of node i is $t_i=i$

$$P(i,j) = \frac{m}{2} (ij)^{-\frac{1}{2}}$$

$$Nr_l(\Delta) = \int_{i=1}^N di \int_{j=1}^N dj P(i,j) P(i,l) P(j,l) = \frac{m^3}{8} \int_{i=1}^N di \int_{j=1}^N dj (ij)^{-\frac{1}{2}} (il)^{-\frac{1}{2}} (jl)^{-\frac{1}{2}} = \frac{m^3}{8l} \int_{i=1}^N \frac{di}{i} \int_{j=1}^N \frac{dj}{j} = \frac{m^3}{8l} (\ln N)^2$$

$$C = \frac{\frac{m^3}{8l} (\ln N)^2}{k_l(k_l - 1)/2}$$

$$k_l(t) = m \left(\frac{N}{l} \right)^{1/2}$$

Which is the degree of node l at current time, at time $t=N$

Let us approximate:

$$k_l(k_l - 1) \approx k_l^2 = m^2 \frac{N}{l}$$

$$C = \frac{m}{8} \frac{(\ln N)^2}{N}$$

There is a factor of two difference... Where does it come from?

Evolving network models

EVOLVING NETWORK MODELS

The BA model is only a minimal model.

Makes the simplest assumptions:

- linear growth
- linear preferential attachment

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

$$\Pi(k_i) \propto k_i$$

Does not capture

variations in the shape of the degree distribution

variations in the degree exponent

the size-independent clustering coefficient

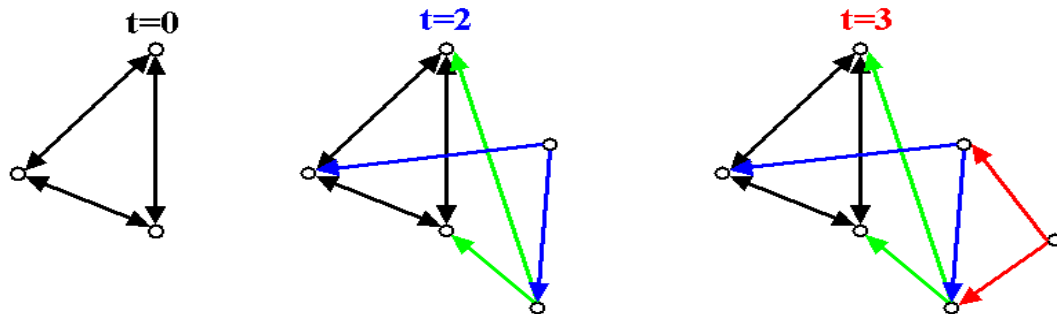
Hypothesis:

The BA model can be adapted to describe most features of real networks.

We need to incorporate mechanisms that are known to take place in real networks: addition of links without new nodes, link rewiring, link removal; node removal, constraints or optimization

BA ALGORITHM WITH DIRECTED EDGES

(the simplest way to change the degree exponent)



$$\frac{\partial k_i}{\partial t} \propto \Pi(k_i) = A \frac{k_i}{\sum_j k_j} = \frac{k_i}{t}$$

Undirected BA network: $\sum_j k_j = 2t$

Directed BA network: $\sum_j k_j = t$

$$k_i(t) = m \frac{t}{t_i}$$

$$P_{in}(k) \sim k^{-2}$$

$\beta=1$: dynamical exponent

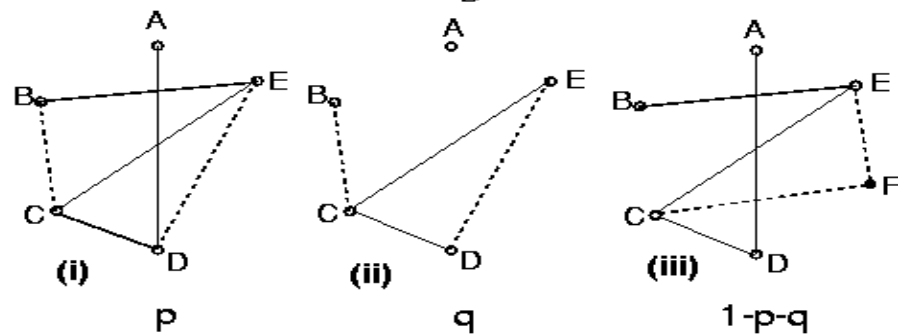
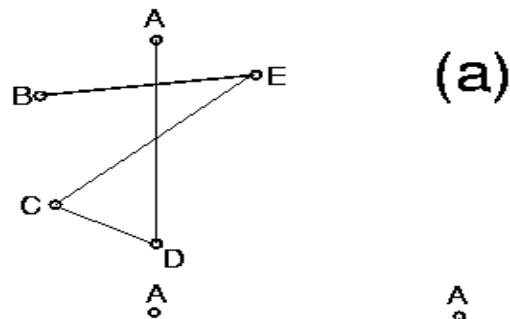
Undirected BA: $\beta=1/2$;

$\gamma_{in}=2$: degree exponent; $P(k_{out})=\delta(k_{out}-m)$

$\gamma=3$



EXTENDED MODEL: Other ways to change the exponent

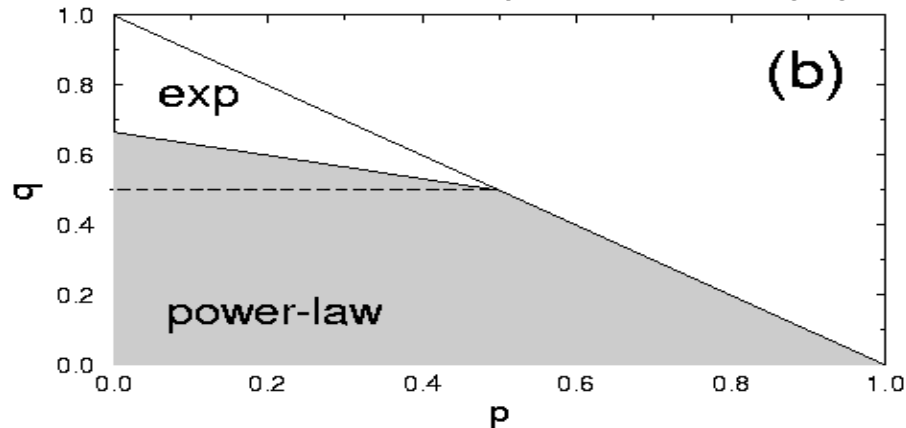


Extended Model

- prob. p : internal links
- prob. q : link deletion
- prob. $1-p-q$: add node

$$P(k) \sim (k + \kappa(p, q, m))^{-\gamma(p, q, m)}$$

$$\gamma \in [1, \infty)$$





EXTENDED MODEL: Small-k cutoff

$$P(k) \sim (k + \kappa(p, q, m))^{-\gamma(p, q, m)} \quad \gamma \in [1, \infty)$$

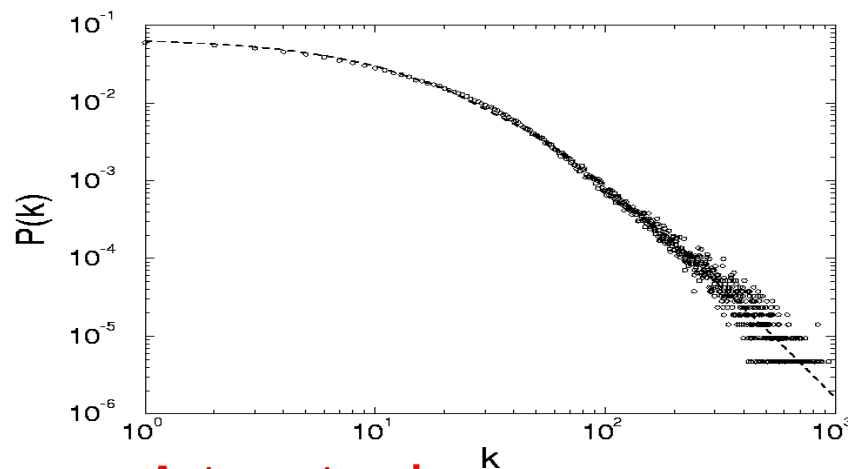
Extended Model

→ Predicts a small-k cutoff

→ a correct model should predict all aspects of the degree distribution, not only the degree exponent.

→ Degree exponent is a continuous function of p, q, m

- prob. p : internal links
- prob. q : link deletion
- prob. $1-p-q$: add node



$$p=0.937$$

$$m=1$$

$$\kappa =$$

$$31.68$$

$$\gamma = 3.07$$

- Non-linear preferential attachment:

$$\Pi(k) = \frac{k^\alpha}{\sum_i k_i^\alpha}$$

→ $P(k)$ does not follow a power law for $\alpha \neq 1$

⇒ $\alpha < 1$: stretch-exponential $P(k) \approx \exp(-(k/k_0)^\beta)$

⇒ $\alpha > 1$: no-scaling ($\alpha > 2$: “gelation”)

INITIAL ATTRACTIVENESS

BA model: $k=0$ nodes cannot acquire links, as $\Pi(k=0)=0$
(the probability that a new node will attach to it is zero)

$$\Pi(k) \approx A + k^\alpha, \quad \alpha \leq 1$$

A - initial attractiveness

Initial attractiveness shifts the degree exponent:

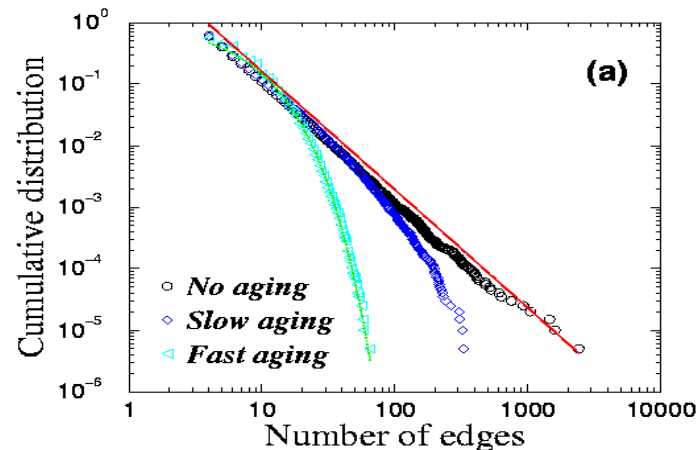
$$\gamma_{in} = 2 + \frac{A}{m}$$

Note: the parameter A can be measured from real data, being the rate at which $k=0$ nodes acquire links, i.e. $\Pi(k=0)=A$

GROWTH CONSTRAINTS AND AGING CAUSE CUTOFFS

- Finite lifetime to acquire new edges

L. A. N. Amaral et al., PNAS 97, 11149 (2000)



- Gradual aging:

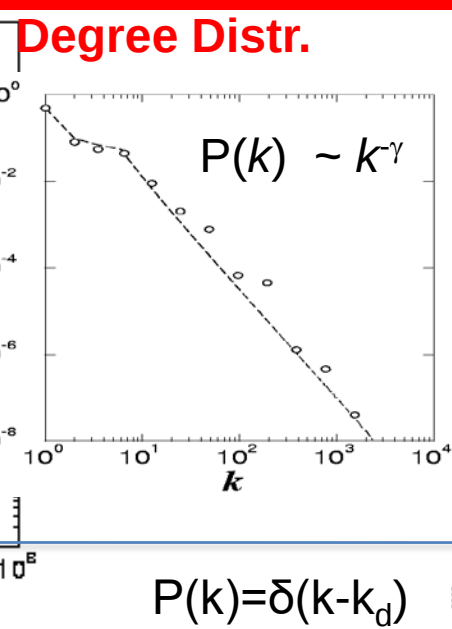
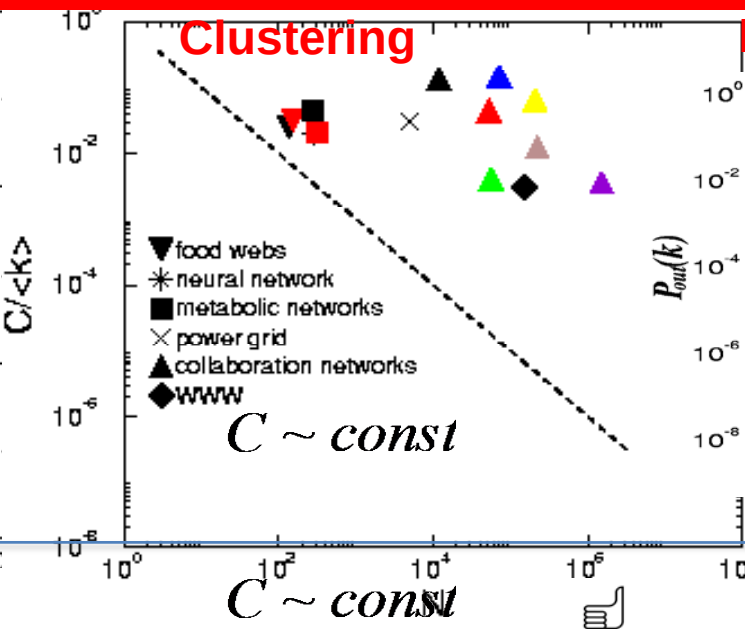
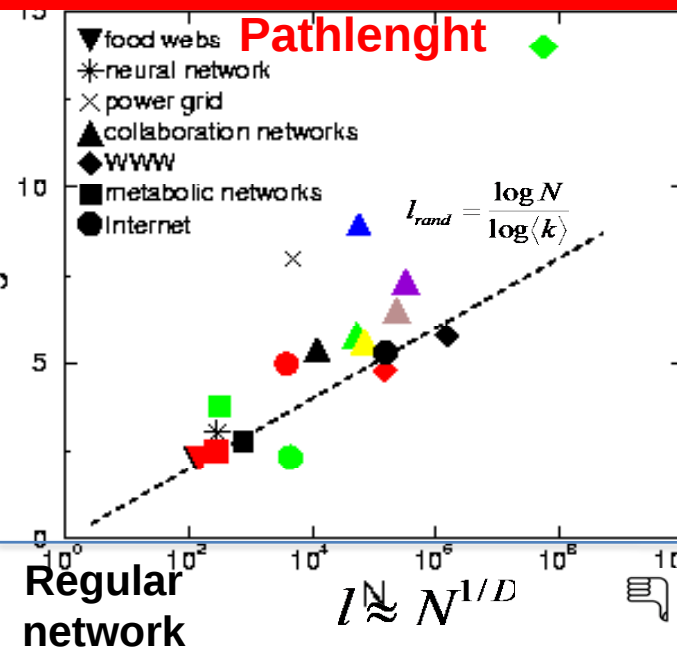
$$\Pi(k_i) \propto k_i(t - t_i)^{-\nu}$$

γ increases with ν

S. N. Dorogovtsev and J. F. F. Mendes, Phys. Rev. E 62, 1842 (2000)



THE LAST PROBLEM: HIGH, SYSTEM-SIZE INDEPENDENT C(N)



Erdos-Renyi

$l_{rand} \approx \frac{\log N}{\log \langle k \rangle}$

$C_{rand} = p = \frac{\langle k \rangle}{N}$

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

Watts-Strogatz

$l_{rand} \approx \frac{\log N}{\log \langle k \rangle}$

$C \sim const$

Exponential

Barabasi-Albert

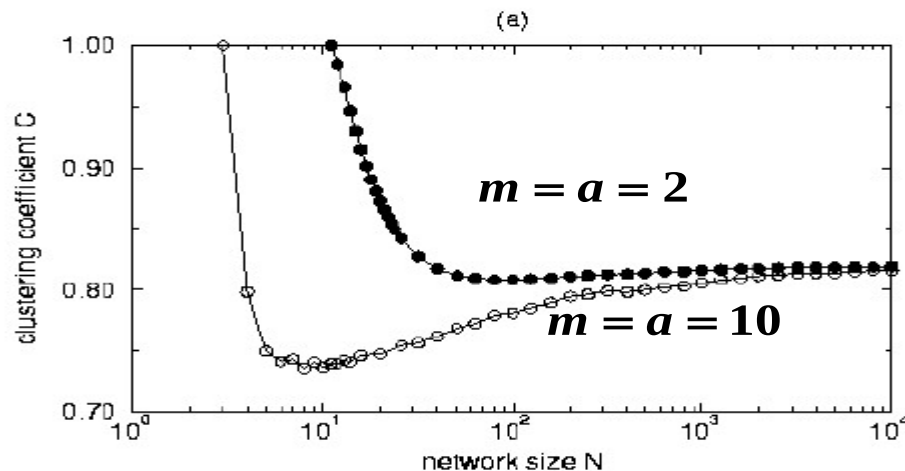
$l \approx \frac{\ln N}{\ln \ln N}$

$C \sim \frac{(\ln N)^2}{N}$

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

A MODEL WITH HIGH CLUSTERING COEFFICIENT

- Each node of the network can be either **active** or **inactive**.
- There are m active nodes in the network in any moment.
- 1. Start with m active, completely connected nodes.
- 2. Each timestep add a new node (active) that connects to m active nodes.
- 3. Deactivate one active node with probability: $P_d(k_i) \propto (a + k_j)^{-1}$



$$\Pi(k) \approx a + k$$

$$P(k) \approx k^{-2-a/m}$$

$$C \rightarrow C^* \text{ when } N \rightarrow \infty$$

Linear growth, linear pref. attachment	$\gamma=3$	Barabási and Albert, 1999
Nonlinear preferential attachment $\Pi(k_i) \sim k_i^\alpha$	no scaling for $\alpha \neq 1$	Krapivsky, Redner, and Leyvraz, 2000
Asymptotically linear pref. attachment $\Pi(k_i) \sim a_\infty k_i$ as $k_i \rightarrow \infty$	$\gamma \rightarrow 2$ if $a_\infty \rightarrow \infty$ $\gamma \rightarrow \infty$ if $a_\infty \rightarrow 0$	Krapivsky, Redner, and Leyvraz, 2000
Initial attractiveness $\Pi(k_i) \sim A + k_i$	$\gamma = 2$ if $A = 0$ $\gamma \rightarrow \infty$ if $A \rightarrow \infty$	Dorogovtsev, Mendes, and Samukhin, 2000a, 2000b
Accelerating growth $\langle k \rangle \sim t^\theta$ constant initial attractiveness	$\gamma = 1.5$ if $\theta \rightarrow 1$ $\gamma \rightarrow 2$ if $\theta \rightarrow 0$	Dorogovtsev and Mendes, 2001a
Internal edges with probab. p	$\gamma = 2$ if $q = \frac{1-p+m}{1+2m}$	
Rewiring of edges with probab. q	$\gamma \rightarrow \infty$ if $p, q, m \rightarrow 0$	Albert and Barabási, 2000
c internal edges or removal of c edges	$\gamma \rightarrow 2$ if $c \rightarrow \infty$ $\gamma \rightarrow \infty$ if $c \rightarrow -1$	Dorogovtsev and Mendes, 2000c
Gradual aging $\Pi(k_i) \sim k_i(t-t_i)^{-\nu}$	$\gamma \rightarrow 2$ if $\nu \rightarrow -\infty$ $\gamma \rightarrow \infty$ if $\nu \rightarrow 1$	Dorogovtsev and Mendes, 2000b
Multiplicative node fitness $\Pi_i \sim \eta_i k_i$	$P(k) \sim \frac{k^{-1-c}}{\ln(k)}$	Bianconi and Barabási, 2001a Dorogovtsev, Mendes, and Samukhin, 2000c
Edge inheritance $P(k_{in}) = \frac{d}{k_{in}^{\sqrt{2}}} \ln(ak_{in})$		
Copying with probab. p	$\gamma = (2-p)/(1-p)$	Kumar <i>et al.</i> , 2000a, 2000b
Redirection with probab. r	$\gamma = 1 + 1/r$	Krapivsky and Redner, 2001
Walking with probab. p	$\gamma \approx 2$ for $p > p_c$	Vázquez, 2000
Attaching to edges p directed internal edges $\Pi(k_i, k_j) \propto (k_i^{in} + \lambda)(k_j^{out} + \mu)$	$\gamma = 3$ $\gamma_{in} = 2 + p\lambda$ $\gamma_{out} = 1 + (1-p)^{-1} + \mu p/(1-p)$	Dorogovtsev, Mendes, and Samukhin, 2001a Krapivsky, Rodgers, and Redner, 2001

Section 11: Summary

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

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

The network grows, but the degree distribution is stationary.

Section 11: Summary

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

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

Consequently, the modeling philosophy behind the model is simple: *to understand the topology of a complex system, we need to describe how it came into being.*

The network grows, but the degree distribution is stationary.

Section 11: Summary

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

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

- The model predicts $\gamma=3$ while the degree exponent of real networks varies between 2 and 5 (Table 4.2).
- Many networks, like the WWW or citation networks, are directed, while the model generates undirected networks.
- Many processes observed in networks, from linking to already existing nodes to the disappearance of links and nodes, are absent from the model.
- The model does not allow us to distinguish between nodes based on some intrinsic characteristics, like the novelty of a research paper or the utility of a webpage.
- While the Barabási-Albert model is occasionally used as a model of the Internet or the cell, in reality it is not designed to capture the details of any particular real network. It is a minimal, proof of principle model whose main purpose is to capture the basic mechanisms responsible for the emergence of the scale-free property. Therefore, if we want to understand the evolution of systems like the Internet, the cell or the WWW, we need to incorporate the important details that contribute to the time evolution of these systems, like the directed nature of the WWW, the possibility of internal links and node and link removal.

LESSONS LEARNED: evolving network models

1. There is no universal exponent characterizing all networks.
2. Growth and preferential attachment are responsible for the emergence of the scale-free property.
3. The origins of the preferential attachment is system-dependent.
4. Modeling real networks:
 - identify the microscopic processes that take place in the system
 - measure their frequency from real data
 - develop dynamical models that capture these processes.
5. If the model is correct, it should correctly predict not only the degree exponent, but both small and large k -cutoffs.

Philosophical change in network modeling:

ER, **WS** models are static models – the role of the network modeler is to cleverly place the links between a fixed number of nodes so that the network topology mimics the networks seen in real systems.

BA and evolving network models are dynamical models: they aim to reproduce how the network was built and evolved.
Thus their goal is to capture the network dynamics, not the structure.
→ as a byproduct, you get the topology correctly

Philosophical change in network modeling:

ER, **WS** models are static models – the role of the network modeler is to cleverly place the links between a fixed number of nodes so that the network topology mimics the networks seen in real systems.

BA and evolving network models are dynamical models: they aim to reproduce how the network was built and evolved.
Thus their goal is to capture the network dynamics, not the structure.
→ as a byproduct, you get the topology correctly

TOPOLOGY OF THE PROTEIN NETWORK

Nodes: proteins
Links: physical interactions (binding)

Puzzling pattern:

Hubs tend to link to small degree nodes.
Why is this puzzling?

In a random network, the probability that a node with degree k links to a node with degree k' is:

$$P_{kk'} = \frac{kk'}{2L}$$

$k \approx 50$, $k' = 13$, $N = 1,458$, $L = 1746$

$$p_{50,13} = 0.15$$

$$p_{2,1} = 0.0004$$

Yet, we see many links between degree 2 and 1 links, and no links between the hubs.

H. Jeong, S.P. Mason, A.-L. Barabasi, Z.N. Oltvai, Nature 411, 41-42 (2001)

