Frontiers of Network Science
Fall 2023

Class 10: Evolving Networks
(Chapter 6 in Textbook)

Boleslaw Szymanski

www.BarabasiLab.com

based on slides by
Albert-László Barabási
and Roberta Sinatra

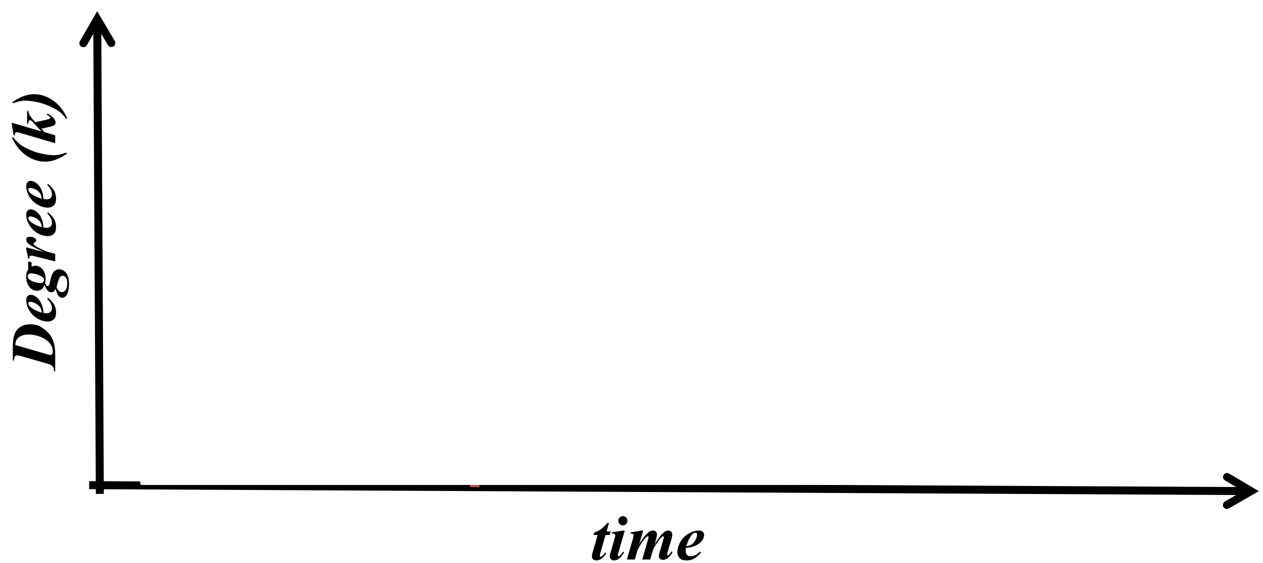
Fitness Model

Fitness Model: Can Latecomers Make It?

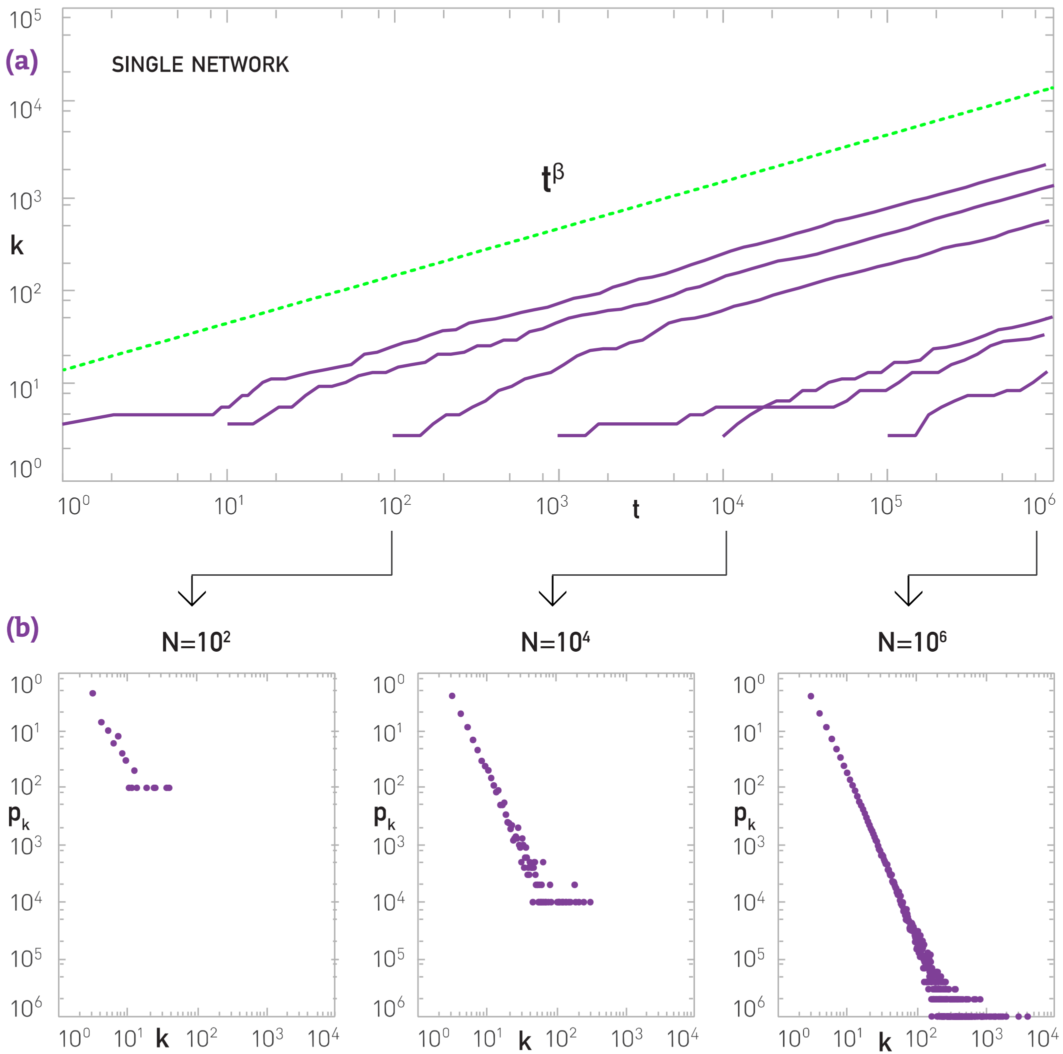
SF model: $k(t) \sim t^{1/2}$ **(first mover advantage)**

Fitness model: **fitness** (η) $\Pi(k_i) \cong \frac{\eta_i k_i}{\sum_j \eta_j k_j}$ $k(\eta,t) \sim t^{\beta(\eta)}$

$\beta(\eta) = \eta / C$



Bianconi & Barabási, *Physical Review Letters* 2001; *Europhys. Lett.* 2001.



- The degree of each node increases following a power-law with the same dynamical exponent β =1/2 (Figure 5.6a). Hence all nodes follow the same dynamical law.

- The growth in the degrees is sublinear (i.e. $\beta < 1$). This is a consequence of the growing nature of the Barabási-Albert model: Each new node has more nodes to link to than the previous node. Hence, with time the existing nodes compete for links with an increasing pool of other nodes.

- The earlier node i was added, the higher is its degree $k_i(t)$. Hence, hubs are large because they arrived earlier, a phenomenon called *first-mover advantage* in marketing and business.

- The rate at which the node i acquires new links is given by the derivative of (5.7)

$$\frac{dk_i(t)}{dt} = \frac{m}{2}\frac{1}{\sqrt{t_i t}}, \qquad (5.8)$$

indicating that in each time frame older nodes acquire more links (as they have smaller t_i). Furthermore the rate at which a node acquires links decreases with time as $t^{-1/2}$. Hence, fewer and fewer links go to a node.

Absence of growth and preferential attachment

MODEL A

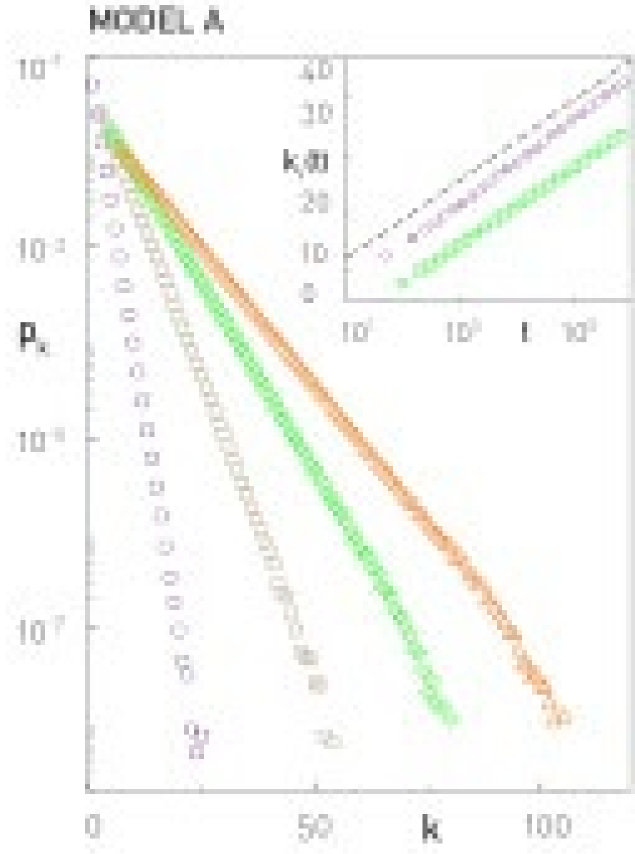
growth ~~preferential attachment~~

$\Pi(k_i)$: uniform

$$\frac{\partial k_i}{\partial t} = A\Pi(k_i) = \frac{m}{m_0 + t - 1}$$

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

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



MODEL B

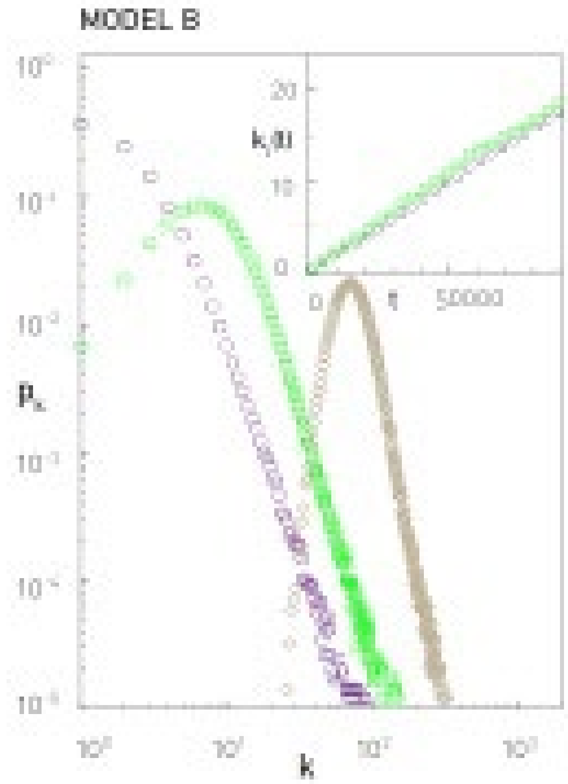
~~growth~~	preferential attachment

$$\frac{\partial k_i}{\partial t} = A\Pi(k_i) + \frac{1}{N} = \frac{N}{N-1}\frac{k_i}{2t} + \frac{1}{N}$$

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

p_k : power law (initially) →

→ Gaussian → Fully Connected

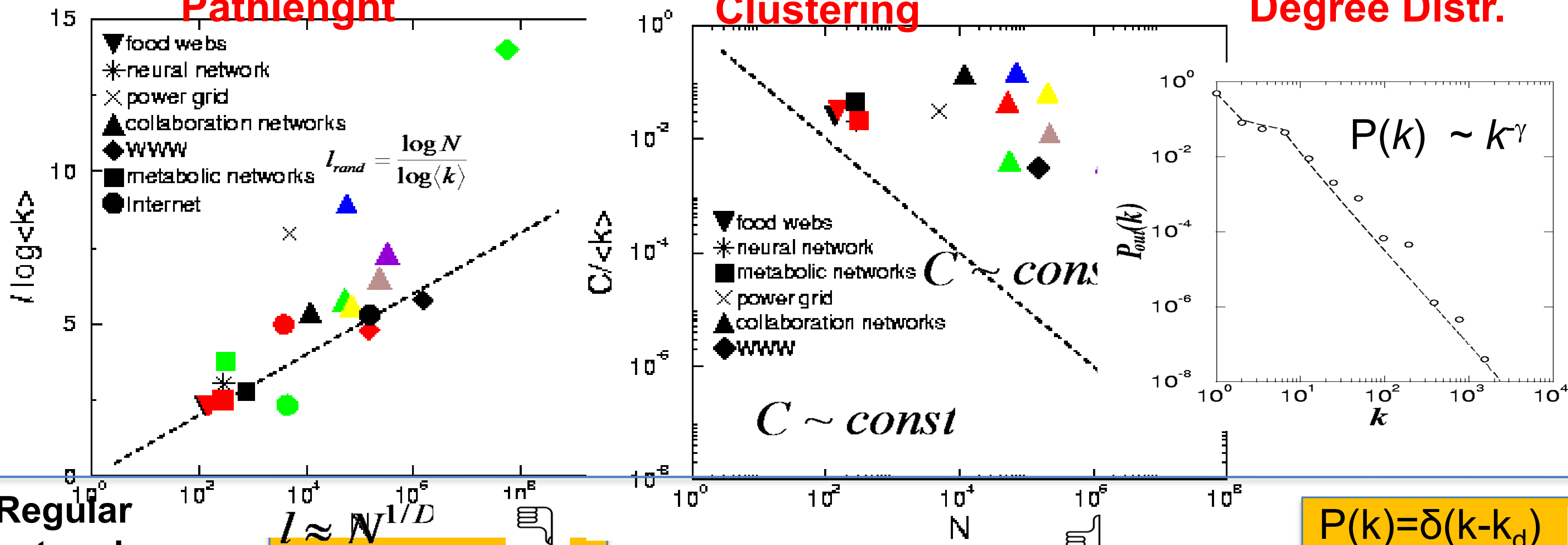


Do we need both growth and preferential attachment?

YEP

EMPIRICAL DATA FOR REAL NETWORKS

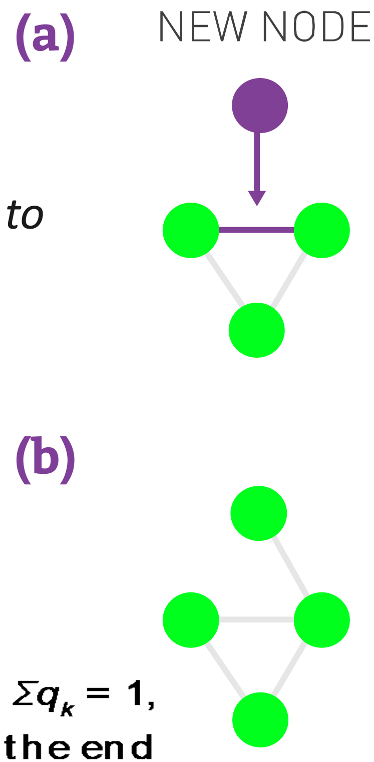
Pathlenght **Clustering** **Degree Distr.**



	Pathlength	Clustering	Degree Distr.
Regular network	$l \approx N^{1/D}$		$P(k)=\delta(k-k_d)$
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			P(k) ~ k^{γ}

The origins of preferential attachment

Link selection model -- perhaps the simplest example of a local or random mechanism capable of generating 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 nodes at the two ends of the selected link.*

To show that this simple mechanism generates linear preferential attachment, we write the probability that the node at the end of a randomly chosen link has degree k as

$$q_k = Ckp_k$$

In (5.26) C can be calculated using the normalization condition $\Sigma q_k = 1$, obtaining $C=1/\langle k\rangle$. Hence the probability to find a degree-k node at the end of a randomly chosen link is

$$q_k = \frac{kp_k}{\langle k\rangle}, \quad (5.27)$$

Section 9 | Originators of preferential attachments

1

1953 In **An Informal Theory of the Statistical Structure of Languages** [26] Benoit Mandelbrot proposes optimization as the origin of power laws.

2

1955 In **On a Class of Skew Distribution Functions** Herbert Simon [6] proposes randomness as the origin of power laws and dismisses Mandelbrot's claim that power law are rooted in optimization.

3 Mandelbrot publishes a comment on Simon's paper [27] writing:

Benoit

> Simon's model is analytically circular... 1959

> Dr. Mandelbrot's principal and mathematical objections to the model are shown to be unfounded 1960

Herbert

The essence of Simon's lengthy reply a year later is well summarized in its abstract [28]. 4

5 In a 19 page response entitled **Final Note**, Mandelbrot states [29]:

Benoit

> ...Most of Simon's (1960) reply was irrelevant. 1961

> This present 'Reply' refutes the almost entirely new arguments introduced by Dr. Mandelbrot in his "Final Note"... 1961

Herbert

Simon's subsequent **Reply to 'Final Note' by Mandelbrot** does not concede [30] 6

7 In the creatively titled **Post Scriptum to Final Note** Mandlebrot [31] writes

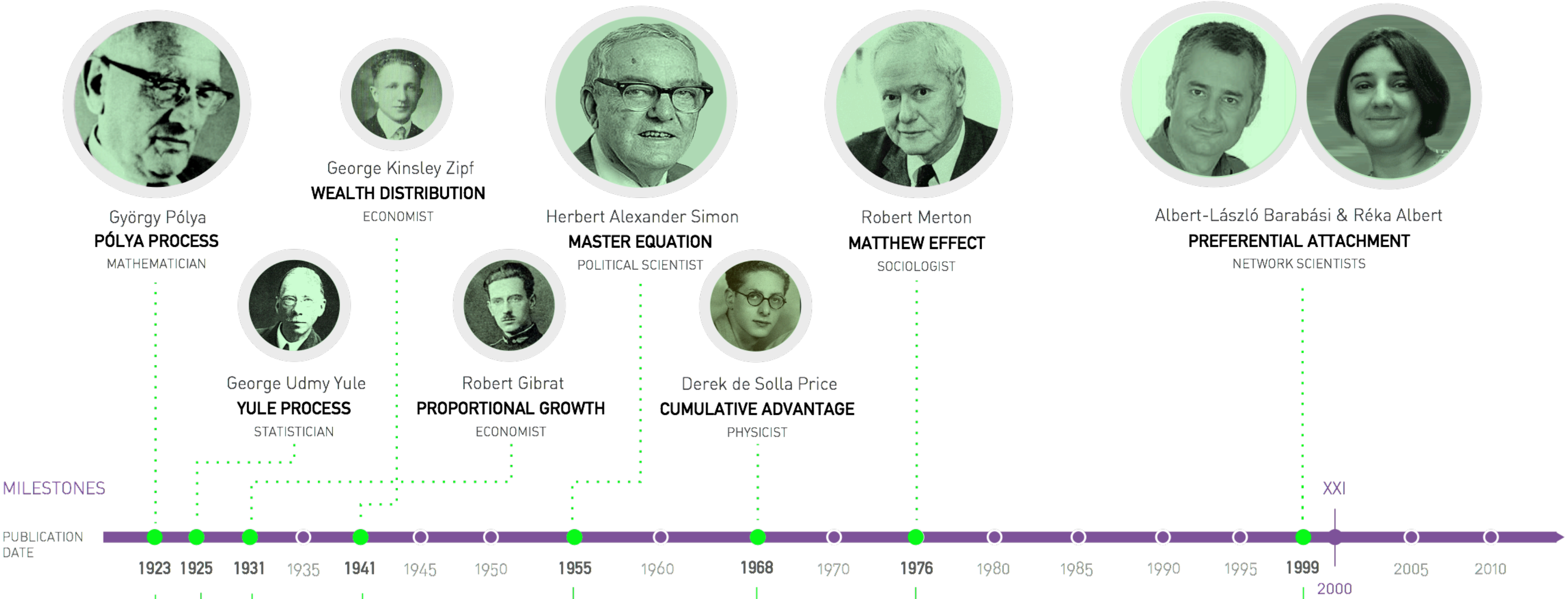
Benoit

> My criticism has not changed since I first had the privilege of commenting upon a draft of Simon. 1961

> Dr. Mandelbrot has proposed a new set of objections to my 1955 models of Yule distributions. Like earlier objections, these are invalid. 1961

Herbert

Simon's final reply ends but does not resolve the debate [32] 8



György Pólya (1887-1985)
Preferential attachment made its first appearance in 1923 in the celebrated urn model of the Hungarian mathematician György Pólya [2]. Hence, in mathematics preferential attachment is often called a **Pólya process**.

Robert Gibrat (1904-1980)
proposed that the size and the growth rate of a firm are independent. Hence, larger firms grow faster [4]. Called **proportional growth**, this is a form of preferential attachment.

Herbert Alexander Simon (1916-2001)
used preferential attachment to explain the fat-tailed nature of the distributions describing city sizes, word frequencies, or the number of papers published by scientists [6].

Robert Merton (1910-2003)
In sociology preferential attachment is often called the **Matthew effect**, named by Merton [8] after a passage in the Gospel of Matthew.

George Udmy Yule (1871-1951)
used preferential attachment to explain the power-law distribution of the number of species per genus of flowering plants [3]. Hence, in statistics preferential attachment is often called a **Yule process**.

George Kinsley Zipf (1902-1950)
used preferential attachment to explain the fat tailed distribution of wealth in the society [5].

Derek de Solla Price (1922-1983)
used preferential attachment to explain the citation statistics of scientific publications, referring to it as **cumulative advantage** [7].

Barabási (1967) & **Albert** (1972)
introduce the term **preferential attachment** in the context of networks [1] to explain the origin of their power-law degree distribution.

MECHANISMS RESPONSIBLE FOR PREFERENTIAL ATTACHMENT

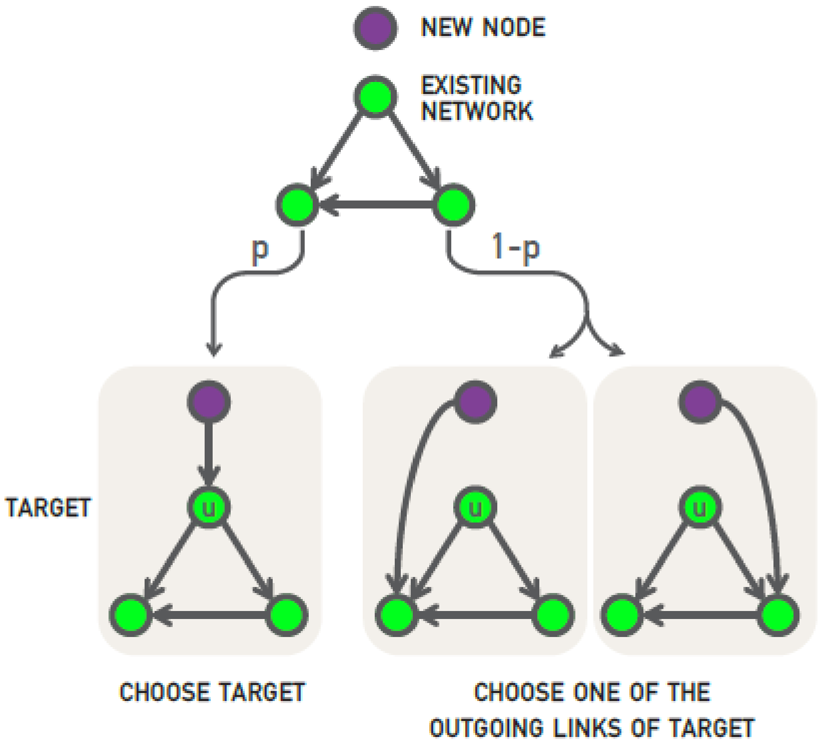
1. Copying mechanism
   directed network
   select a node and an edge of this node
   attach to the endpoint of this edge

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

Copying model



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

(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

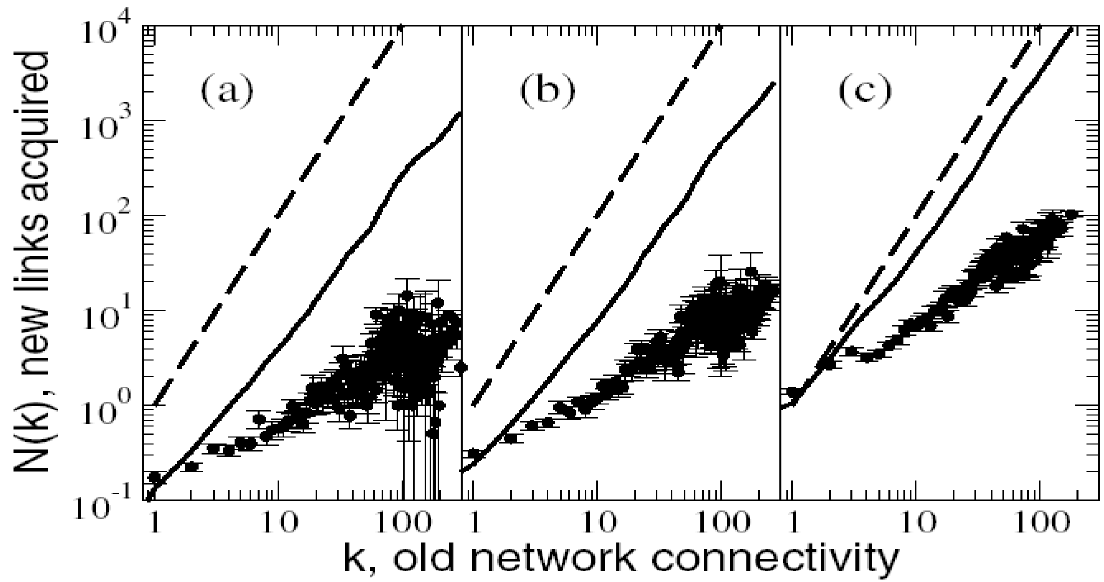
Social networks: *Copy your friend’s friends.*
Citation Networks*: Copy references from papers we read.*
Protein interaction networks*: gene duplication,*

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

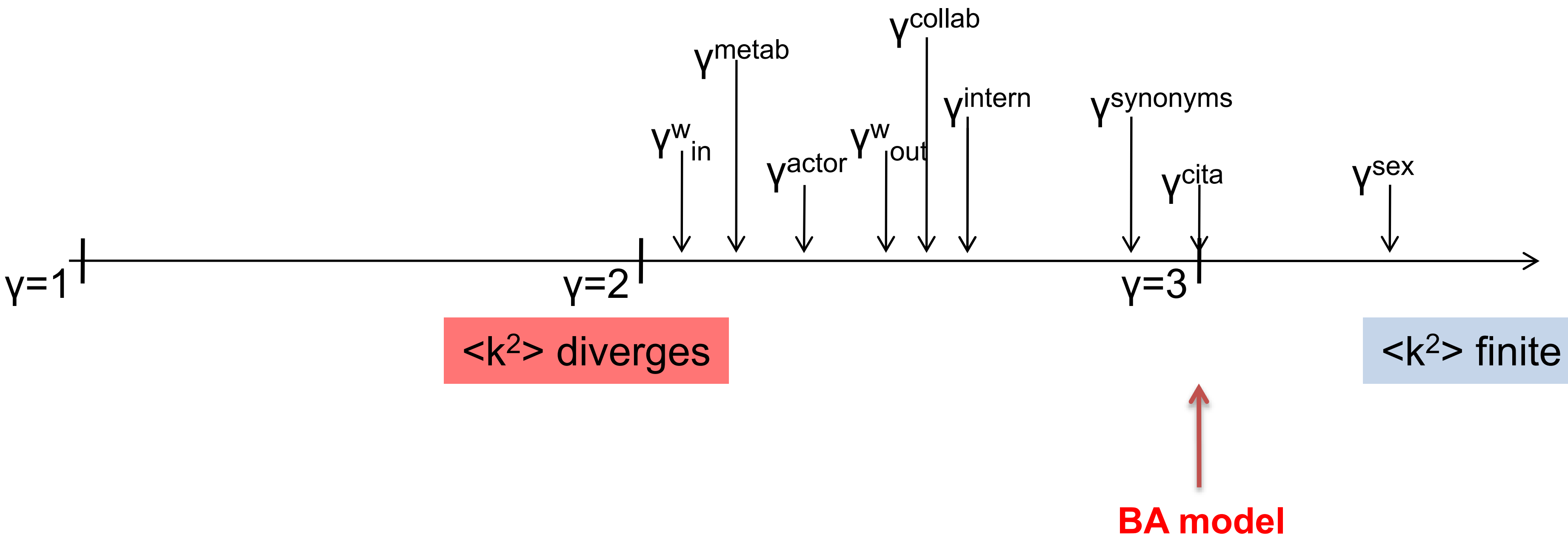
Eisenberg E, Levanon EY, Phys. Rev. Lett. 2003.

SUMMARY: PROPERTIES OF THE BA MODEL

- **Nr. of nodes:** $N = t$
- **Nr. of links:** $L = m\,t$
- **Average degree:** $\langle k \rangle = \frac{2L}{N} \to 2m$
- **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_{jj} denotes the bandwidth between nodes i, j

δ denotes the ratio of cost of cable to delay

Question: where to place a new router ?

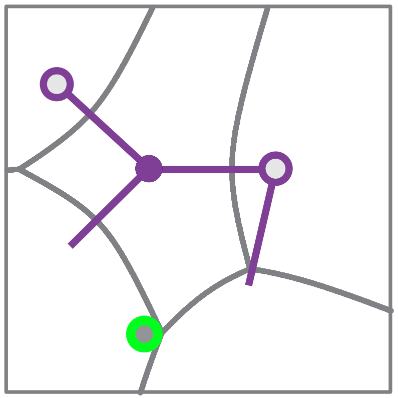
$$C_i = \min_j[\delta d_{ij} + h_j]$$

(c)

(d)

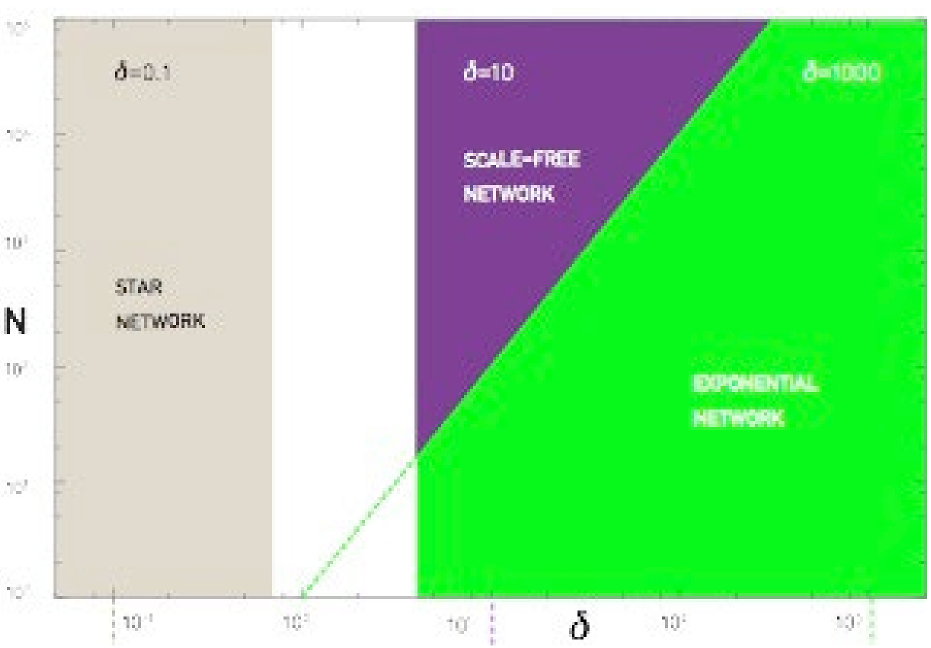
(e)

(f)



Section 9 Optimization model

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

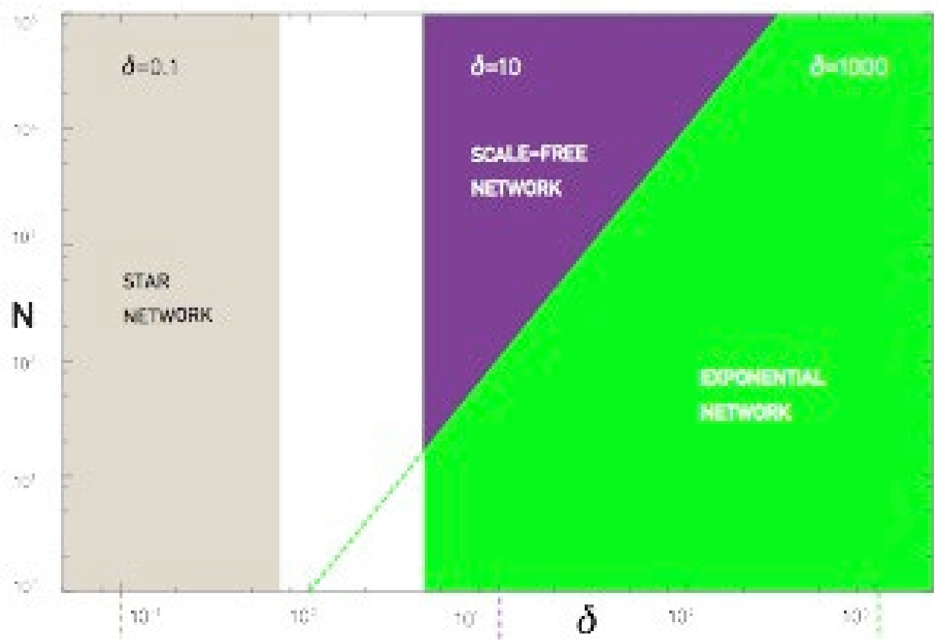


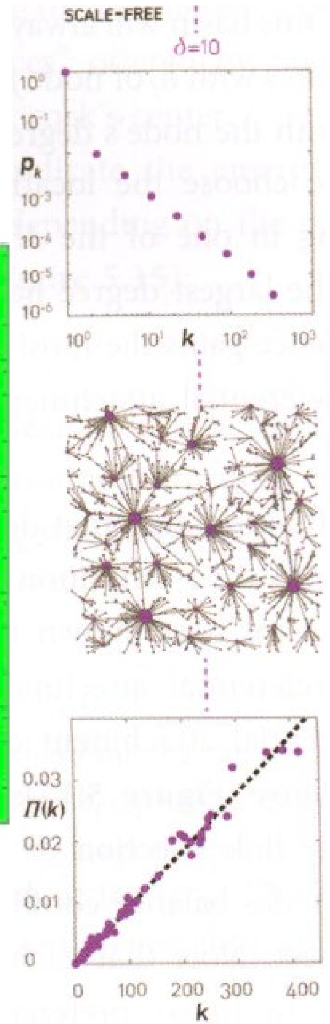
Star Network

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)).

Optimization model

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



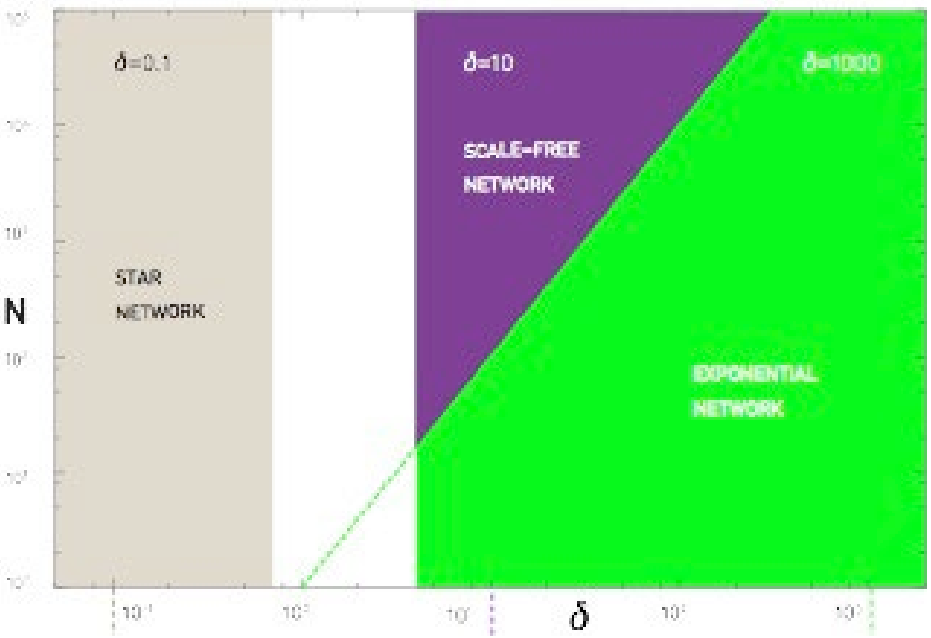


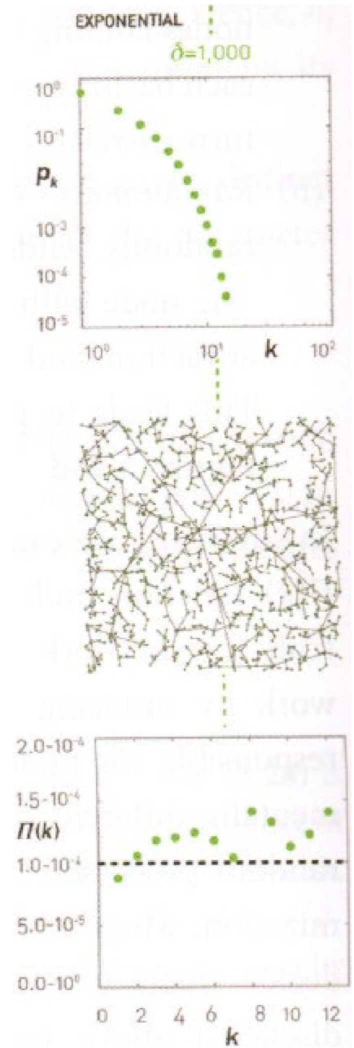
Scale-Free Network

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.

Optimization model

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

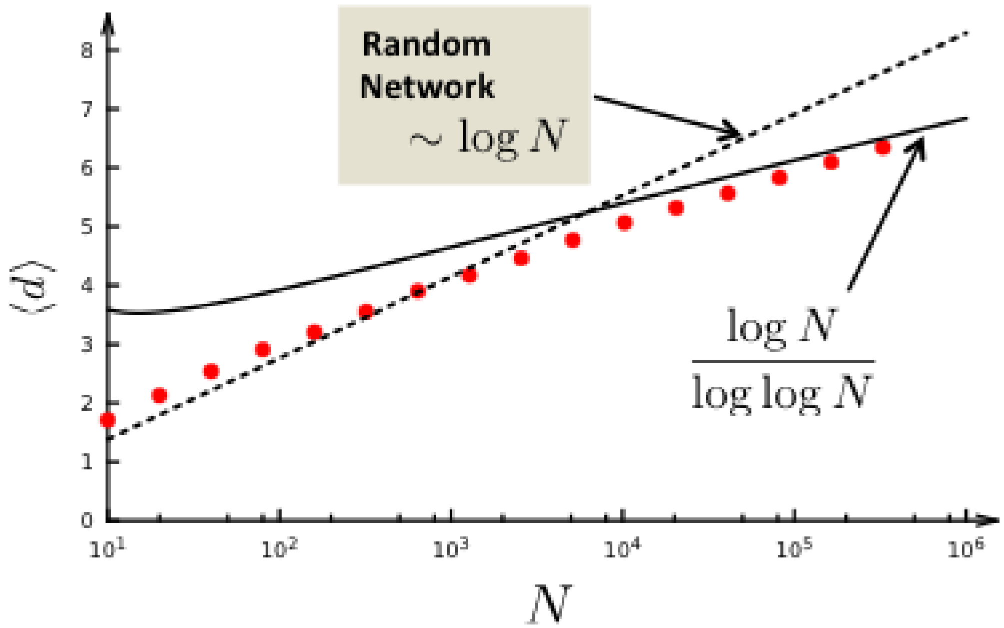




Exponential Networks

Diameter and clustering coefficient

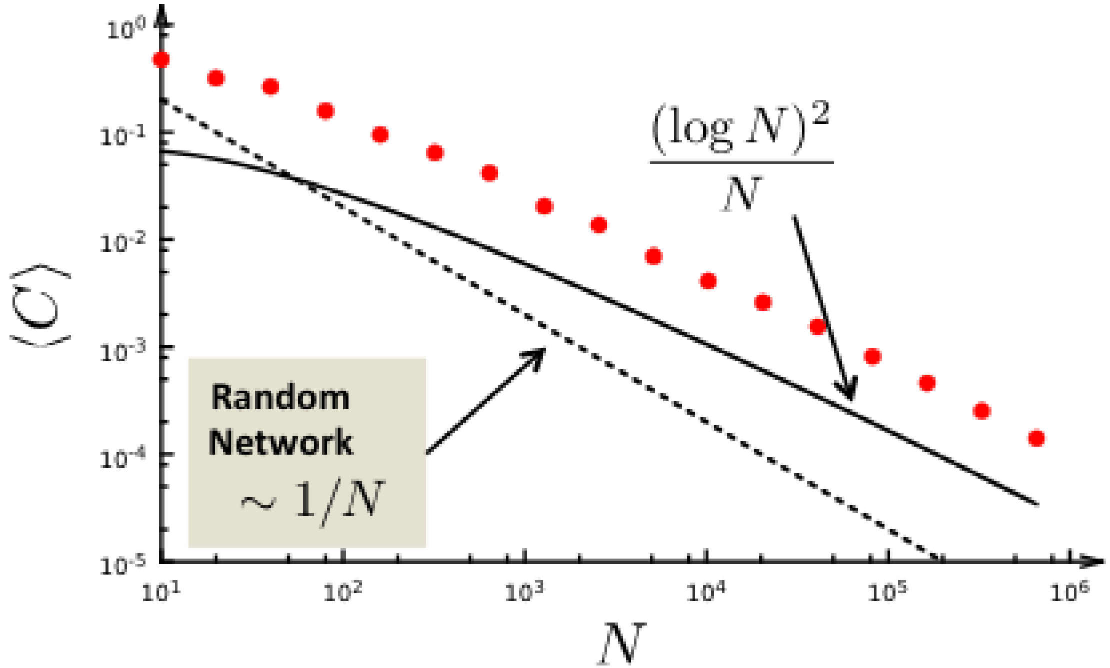
$$D \sim \frac{\log N}{\log\log N}$$



Bollobas, Riordan, 2002

Reminder: for a random graph we have:

$$C_{rand} = \frac{<k>}{N} \sim N^{-1}$$



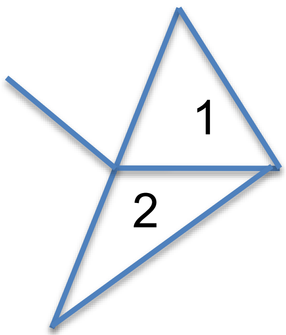
What is the functional form of C(N)?

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

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 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?