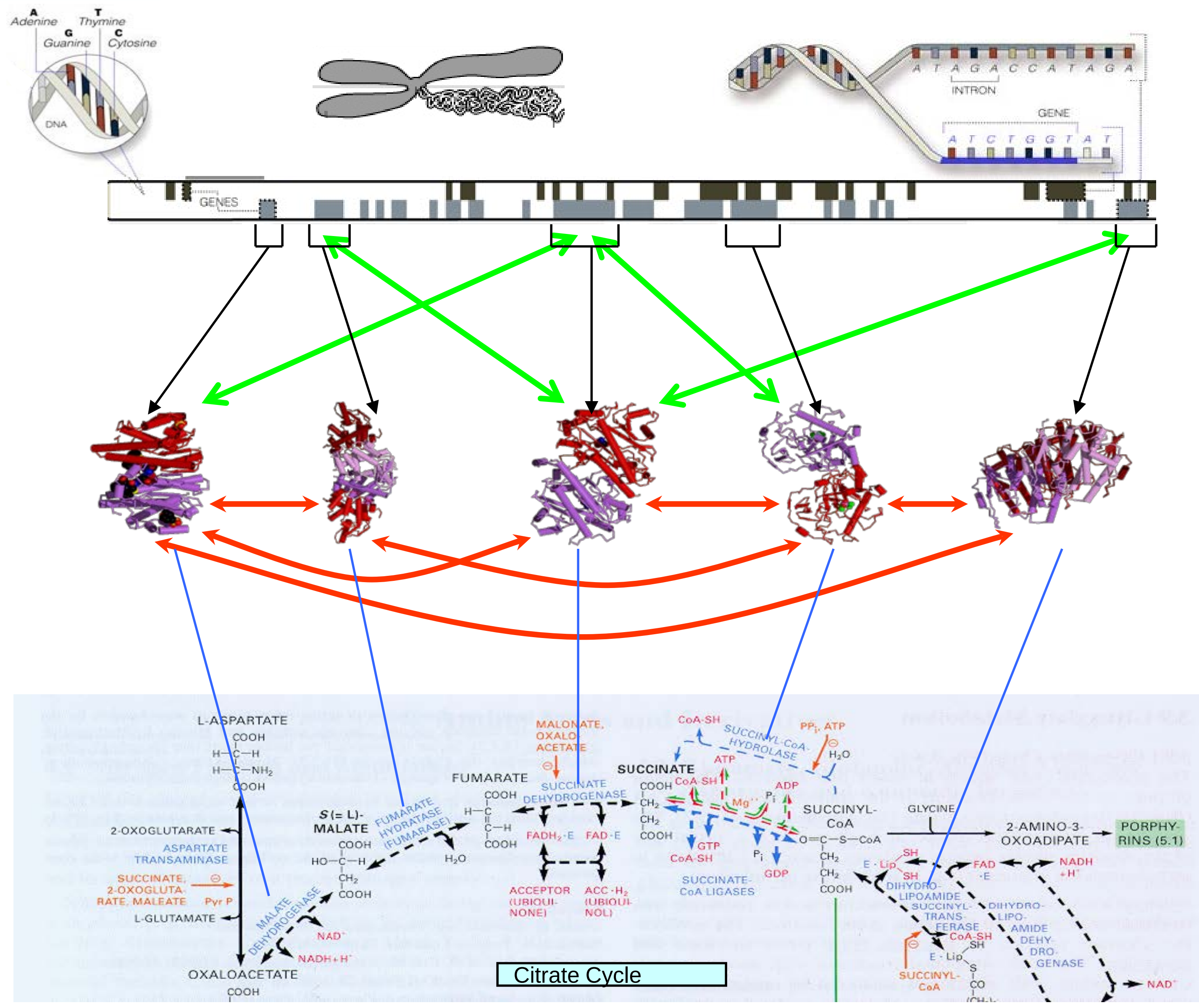
Frontiers of Network Science
Fall 2021

Class 10: Barabasi-Albert Model
(Chapter 5 in Textbook)

Boleslaw Szymanski

based on slides by
Albert-László Barabási
and Roberta Sinatra

www.BarabasiLab.com



GENOME

protein-gene interactions

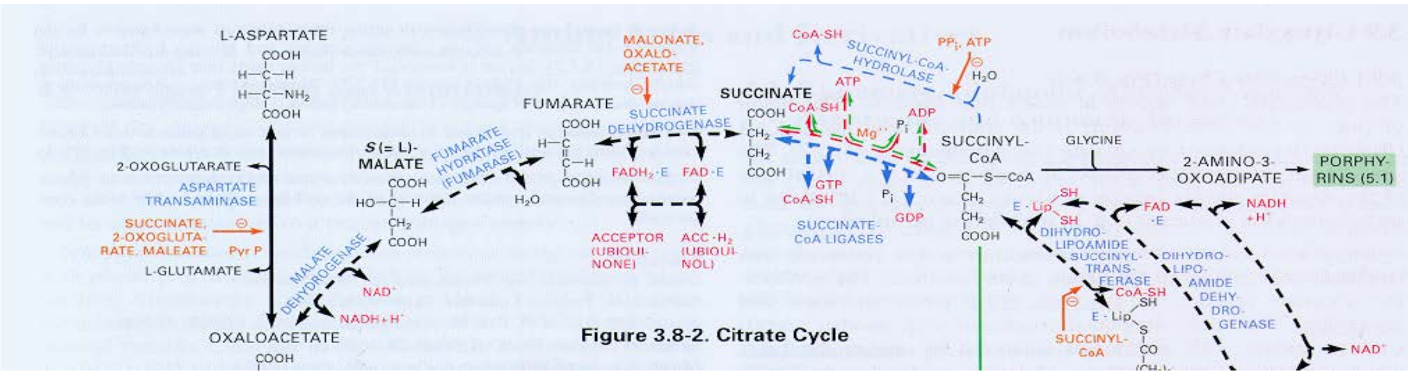
PROTEOME

protein-protein interactions

METABOLISM

Bio-chemical reactions

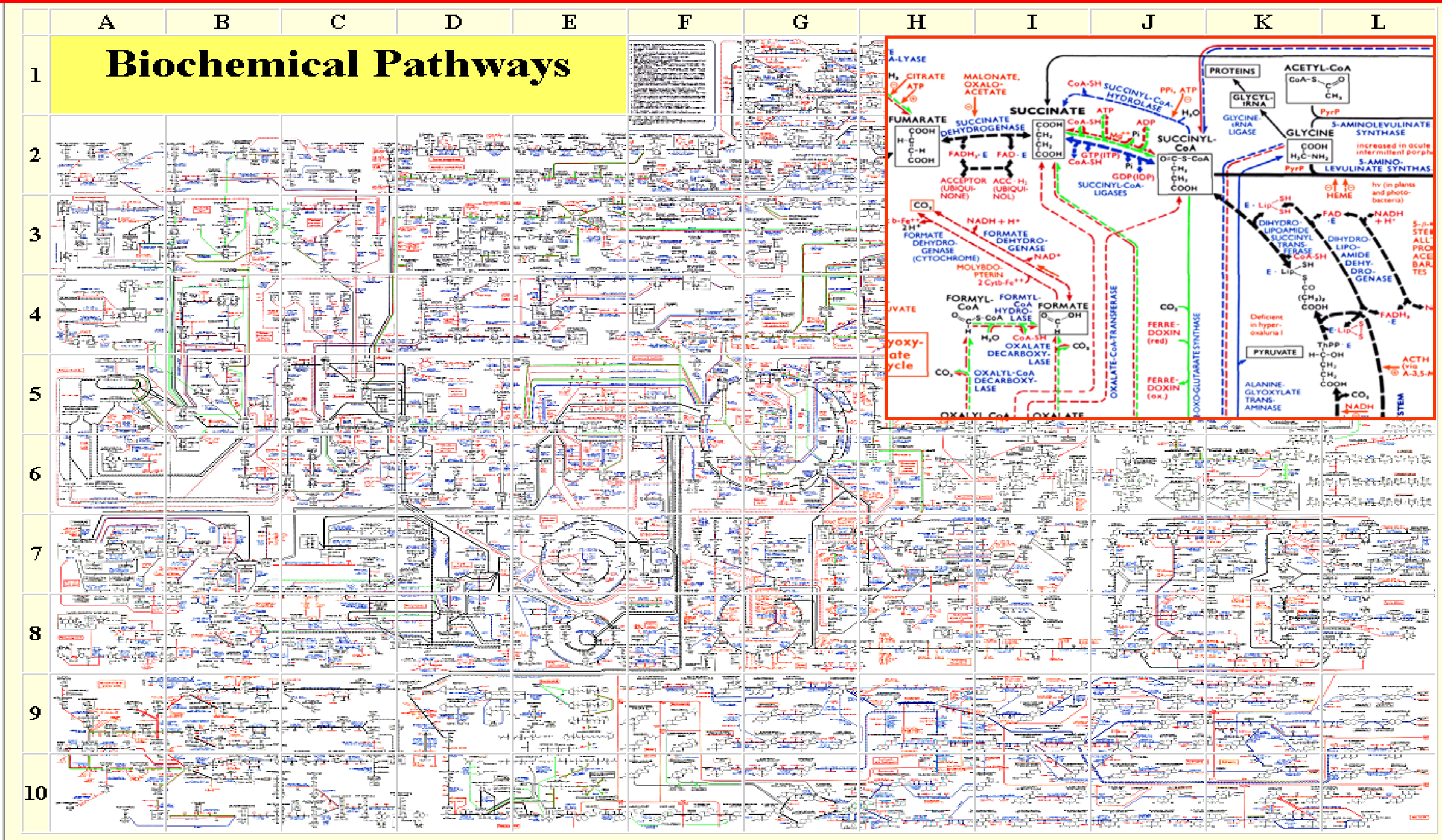
Citrate Cycle



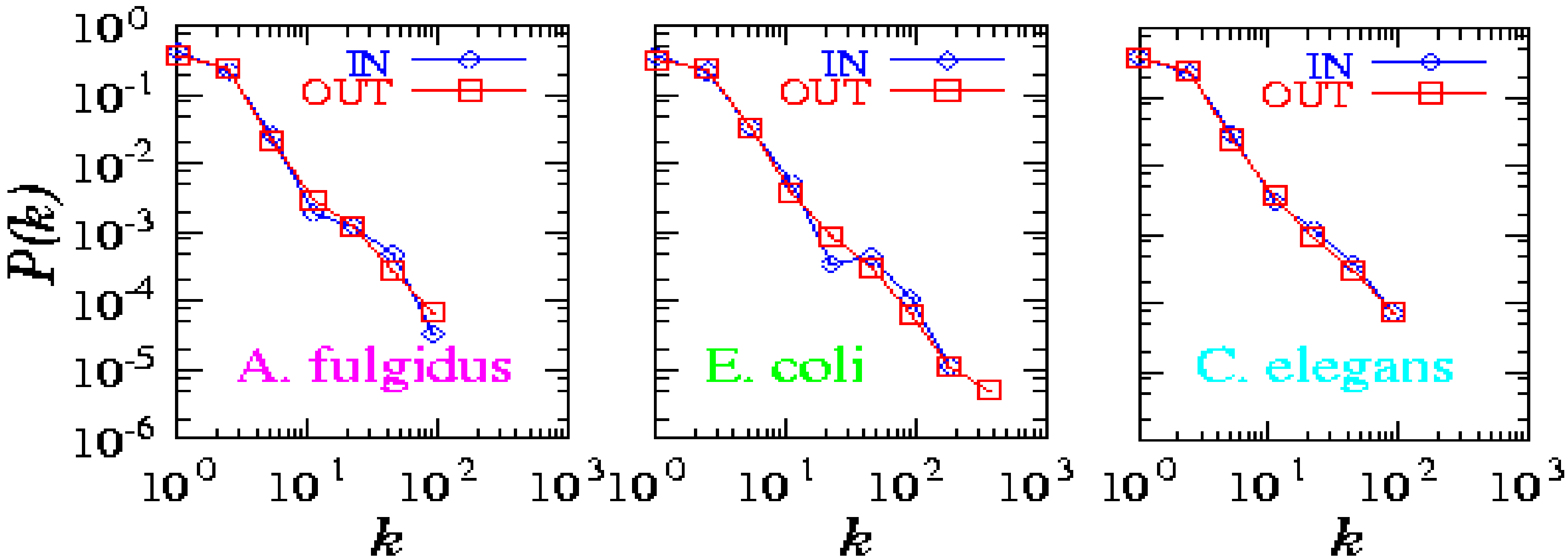
METABOLISM

Bio-chemical reactions

BOEHRING-MENNHEIN



METABOLIC NETWORK



Archaea

Bacteria

Eukaryotes

Organisms from all three domains of life are scale-free!

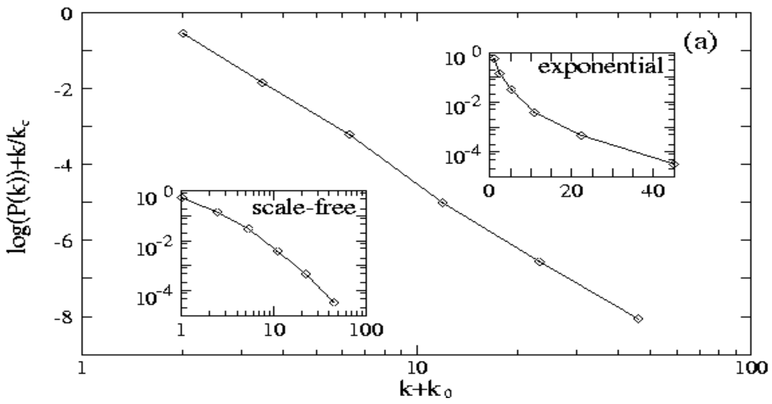
$$P_{in}(k) \approx k^{-2.2}$$

$$P_{out}(k) \approx k^{-2.2}$$

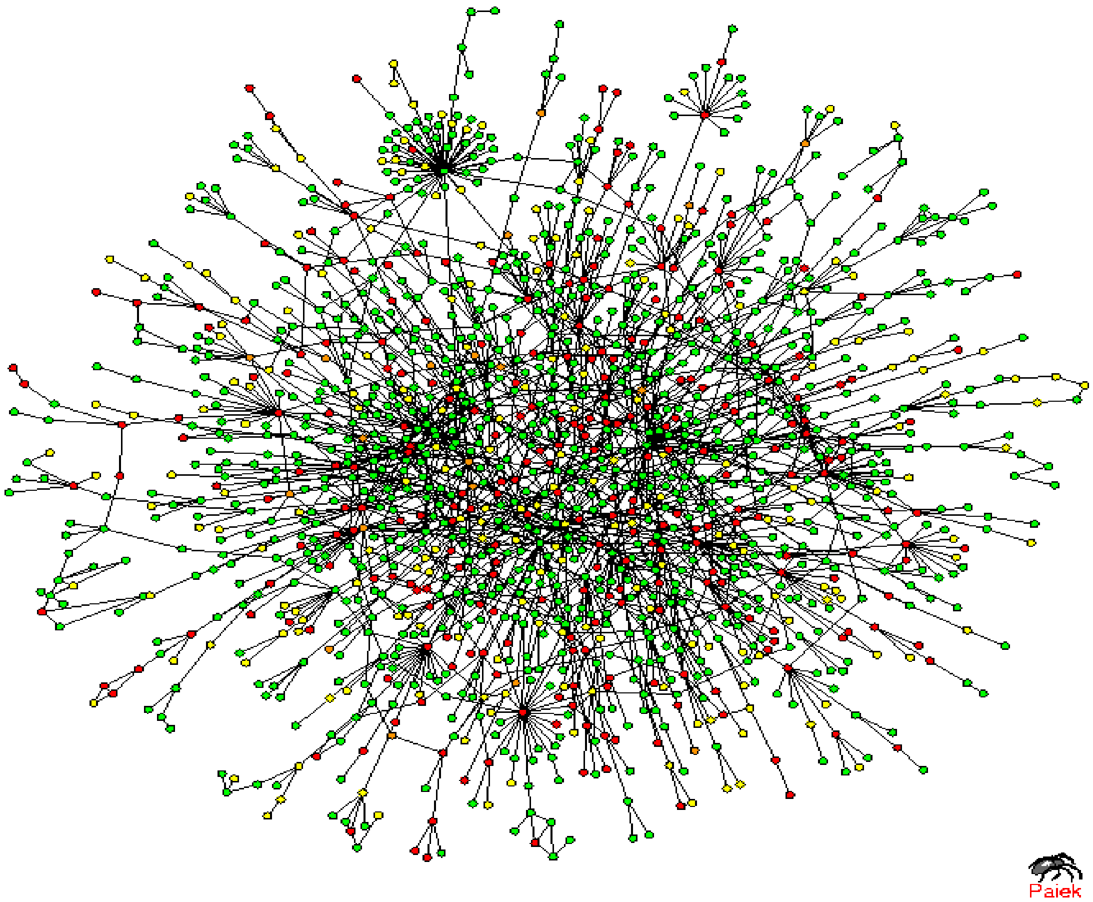
H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, and A.L. Barabasi, *Nature*, 407 651 (2000)

TOPOLOGY OF THE PROTEIN NETWORK

Nodes: proteins
Links: physical interactions-binding

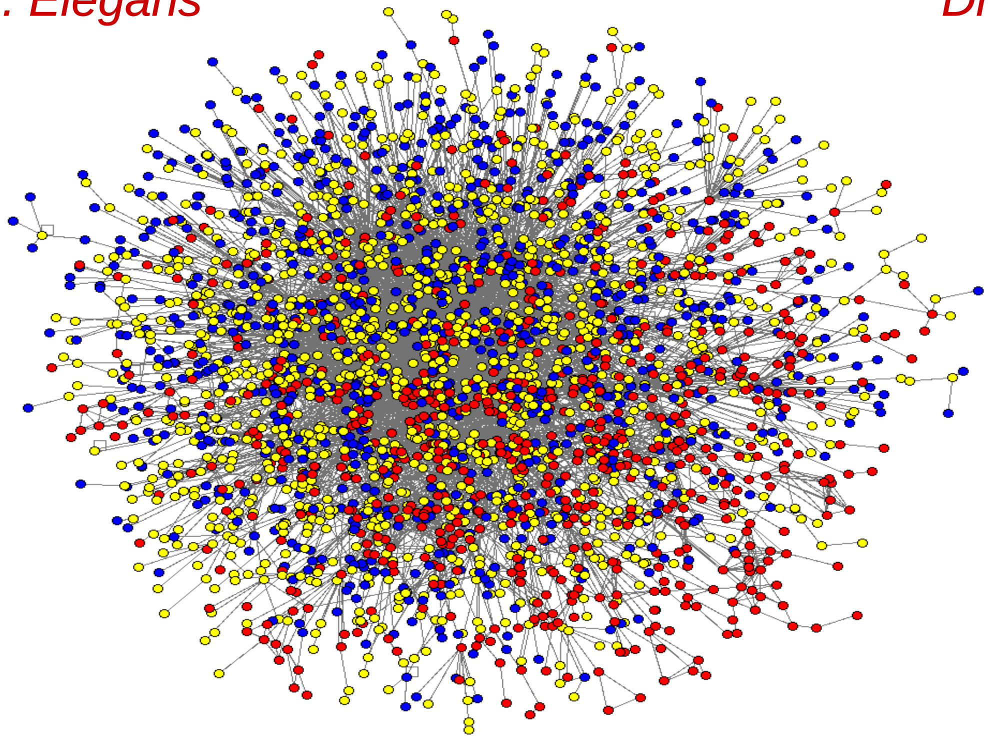


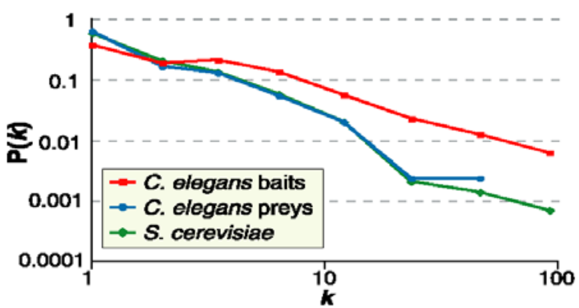
$$P(k) \sim (k+k_0)^{-\gamma} \exp\left(-\frac{k+k_0}{k_\tau}\right)$$



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

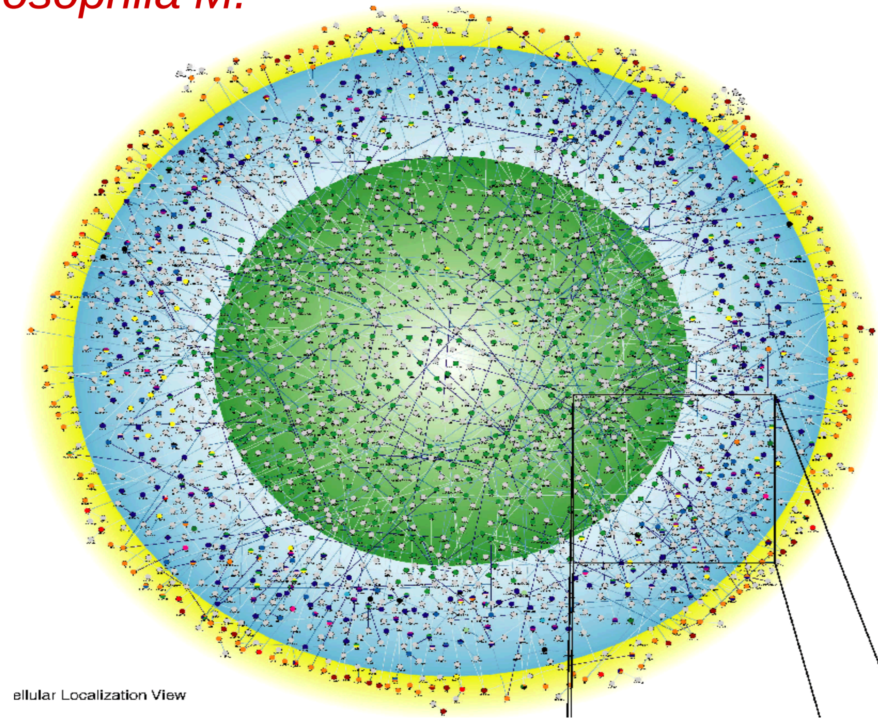
C. Elegans

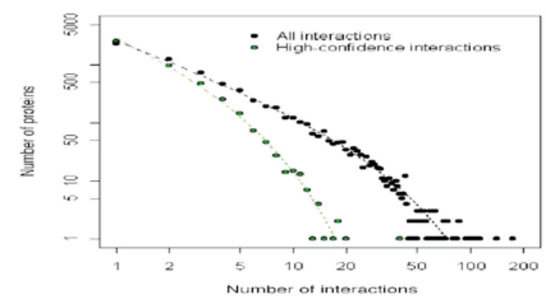




Li *et al.* Science 2004

Drosophila M.





Giot *et al.* Science 2003

Growth and preferential attachment

Barabasi-Albert model Definition

The recognition that growth and preferential attachment coexist in real networks has inspired a minimal model called the **Barabási-Albert model (BA model),** which generates scale-free networks [1], defined as follows:

We start with m_0 nodes, the links between which are chosen arbitrarily, as long as each node has at least one link. The network develops following two steps:

1. **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. **Preferential attachment:** **t**he probability $\Pi(k)$ that a link of the new node connects to node i depends on the degree k_i as $\Pi(k_i)=k_i \sum_j k_j$

Preferential attachment is a probabilistic mechanism: a new node is free to connect to any node in the network, whether it is a hub or has a single link. However, that if a new node has a choice between a degree-two and a degree-four node, it is twice as likely that it connects to the degree-four node.

[1] A.-L.Barabási, R. Albert and H. Jeong, *Physica* A **272,** 173 (1999)

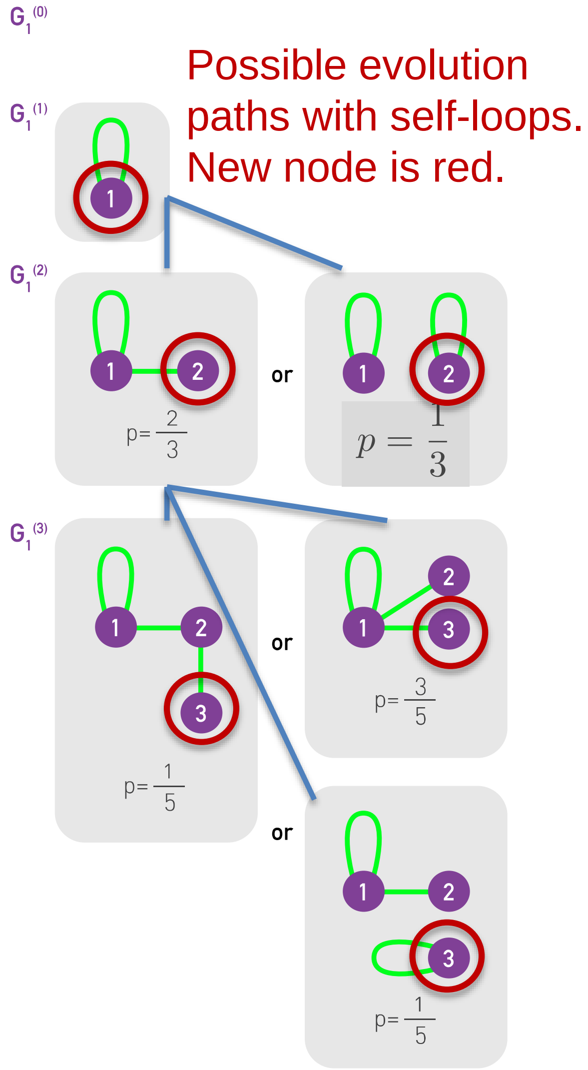
Linearized Chord Diagram

The definition of the Barabási-Albert model leaves many mathematical details open:

- It does not specify the precise initial configuration of the first m_0 nodes.
- It does not specify whether the m links assigned to a new node are added one by one, or simultaneously. This leads to potential mathematical conflicts: If the links are truly independent, they could connect to the same node i, leading to multi-links.

One possible definition with self-loops

$$p(i=s)=\begin{cases}\dfrac{k_i}{2t-1} & if\ 1\leq s\leq t-1\\ \dfrac{1}{2t-1}, & if\ s=t\end{cases}$$



Degree dynamics

Degree distribution for Barabasi-Albert model

$$k_i(t) = m\left(\frac{t}{t_i}\right)^{\beta} \qquad \beta = \frac{1}{2}$$ for $t \geq m_0 + i$ and 0 otherwise as system size at t is $N = m_0 + t - 1$

We assume the initial m_0 nodes create a fully connected graph.
A random node j arriving at time t is with equal probability $1/N = 1/(m_0+t-1)$ one of the nodes $1, 2, \ldots. N$, its degree will grow with the above equation, so

$$P\big(k_j(t)) < k\big) = P\left(t_j > \frac{m^{\frac{1}{\beta}} t}{k^{\frac{1}{\beta}}}\right) = 1 - P\left(t_j \leq \frac{m^{\frac{1}{\beta}} t}{k^{\frac{1}{\beta}}}\right) = 1 - \frac{m^{\frac{1}{\beta}} t}{k^{\frac{1}{\beta}}(t + m_0 - 1)}$$

For the large times t (and so large network sizes) we can replace t-1 with t above, so

$$\therefore P(k) = \frac{\partial P(k_i(t) < k)}{\partial k} = \frac{2m^2 t}{m_o + t}\frac{1}{k^3} \sim k^{-\gamma} \qquad \boxed{\gamma = 3}$$

A.-L.Barabási, R. Albert and H. Jeong, *Physica* A **272,** 173 (1999)

Degree distribution

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

$$\mathrm{P(k)} = \frac{2m^2 t}{t - t_0}\frac{1}{k^3} \sim k^{-\gamma}$$

$$\boxed{\gamma = 3}$$

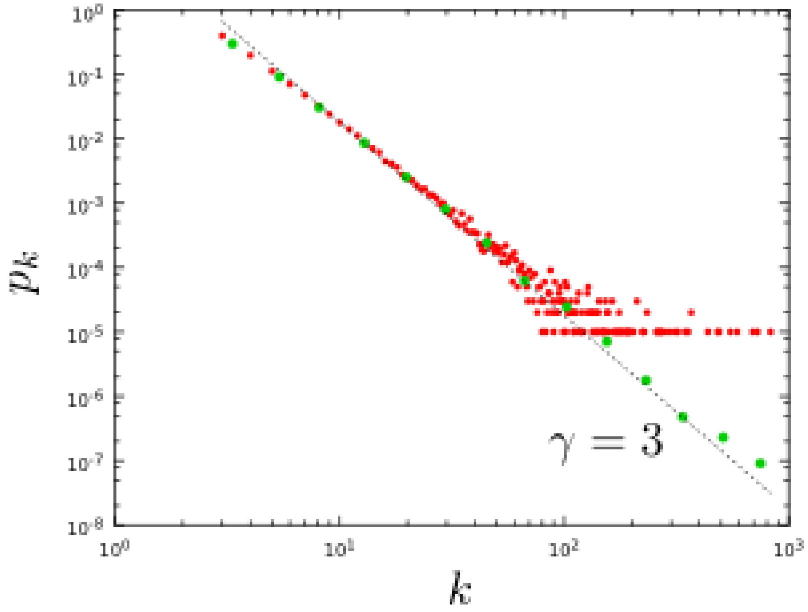
(i) The degree exponent is independent of m.

(ii) As the power-law describes systems of rather different ages and sizes, it is expected that a correct model should provide a time-independent degree distribution. Indeed, asymptotically the degree distribution of the BA model is independent of time (and of the system size N)
→ the network reaches a stationary scale-free state.

(iii) The coefficient of the power-law distribution is proportional to m^2.

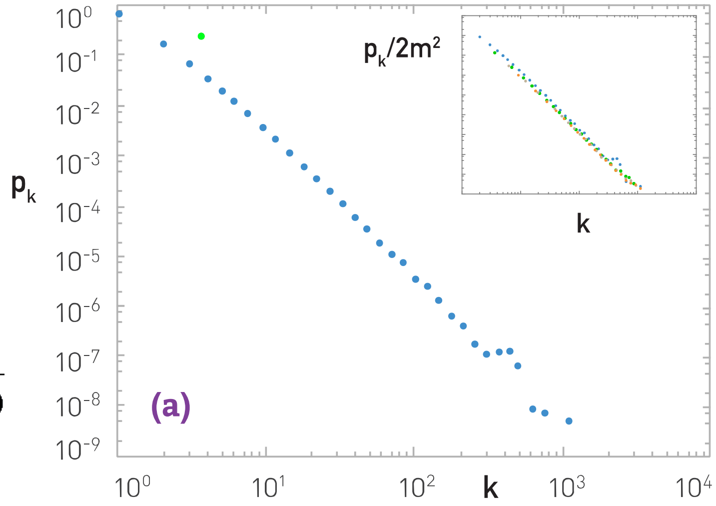
A.-L.Barabási, R. Albert and H. Jeong, *Physica* A **272,** 173 (1999)

Section 4

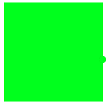


NUMERICAL SIMULATION OF THE BA MODEL

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



(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~$2m^2$, 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)

(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.

NUMERICAL SIMULATION OF THE BA MODEL

m=1,3,5,7

N=100,000;150,000;200,000

m-dependence

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

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

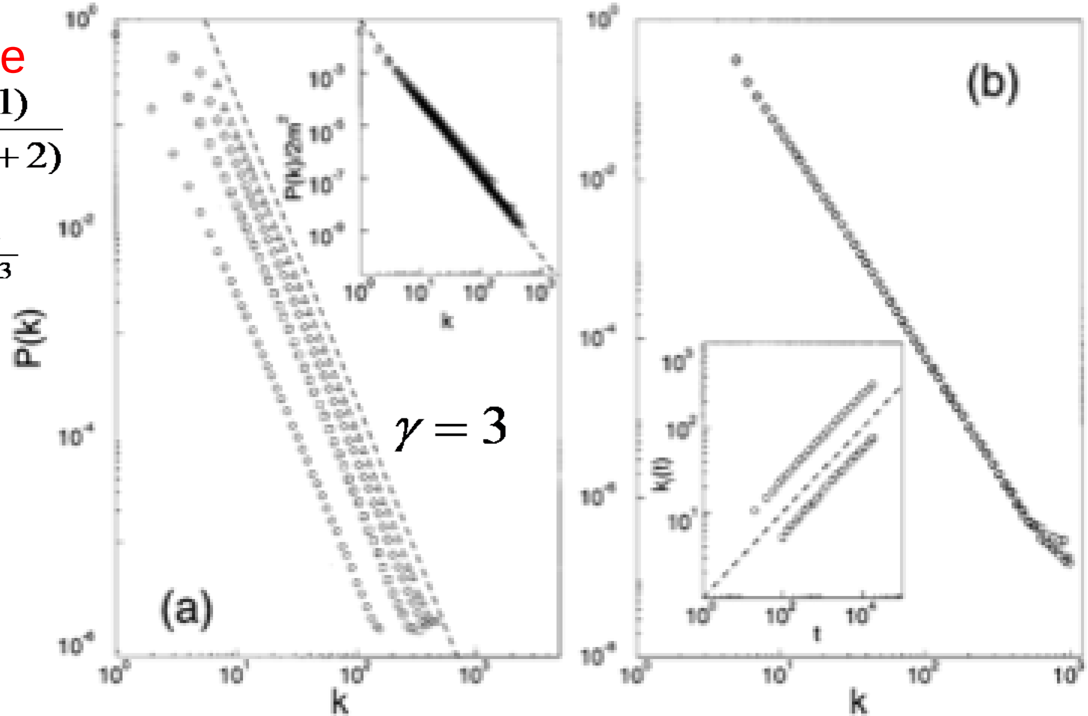


FIG. 21. Numerical simulations of network evolution: (a) Degree distribution of the Barabási-Albert model, with $N = m_0 + t = 300\,000$ and ○, $m_0 = m = 1$; □, $m_0 = m = 3$; ◇, $m_0 = m = 5$; and △, $m_0 = m = 7$. The slope of the dashed line is $\gamma = 2.9$, providing the best fit to the data. The inset shows the rescaled distribution (see text) $P(k)/2m^2$ for the same values of m, the slope of the dashed line being $\gamma = 3$; (b) $P(k)$ for $m_0 = m = 5$ and various system sizes, ○, $N = 100\,000$; □, $N = 150\,000$; ◇, $N = 200\,000$. The inset shows the time evolution for the degree of two vertices, added to the system at $t_1 = 5$ and $t_2 = 95$. Here $m_0 = m = 5$, and the dashed line has slope 0.5, as predicted by Eq. (81). After Barabási, Albert, and Jeong (1999).

Stationarity:
P(k) independent of *N*

Insert:
degree dynamics

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

The mean field theory offers the correct scaling, BUT it provides the wrong coefficient of the degree distribution.

So assymptotically it is correct ($k \rightarrow \infty$), but not correct in details (particularly for small k).

To fix it, we need to calculate P(k) exactly, which we will do next using a rate equation based approach.

MFT - Degree Distribution: Rate Equation

$< N(k,t) >= tP(K,t)$ Number of nodes with degree *k* at time *t*.

Since at each timestep we add one node, we have N=t (total number of nodes = number of timesteps)

$\Pi(k) = \frac{k}{\sum_j k_j} = \frac{k}{2mt}$ *2m*: each node adds *m* links, but each link contributed to the degree of 2 nodes

Number of links added to degree *k* nodes after the arrival of a new node: $$\underbrace{\frac{k}{2mt}}_{\text{Preferential attachment}} \times \overbrace{NP(k,t)}^{\text{Total number of k-nodes}} \times \underbrace{m}_{\text{New node adds m new links to other nodes}} = \frac{k}{2}P(k,t)$$

Nr. of degree *k-1* nodes that acquire a new link, becoming degree *k* $\frac{k-1}{2}P(k-1,t)$

Nr. of degree *k* nodes that acquire a new link, becoming degree *k+1* $\frac{k}{2}P(k,t)$

$$\underbrace{(N+1)P(k,t+1)}_{\text{\# k-nodes at time t+1}} = \underbrace{NP(k,t)}_{\text{\# k-nodes at time t}} + \underbrace{\frac{k-1}{2}P(k-1,t)}_{\text{Gain of k-nodes via k-1}\rightarrow\text{k}} - \underbrace{\frac{k}{2}P(k,t)}_{\text{Loss of k-nodes via k}\rightarrow\text{k+1}}$$

A.-L.Barabási, R. Albert and H. Jeong, *Physica* A **272,** 173 (1999)

MFT - Degree Distribution: Rate Equation

$$\underbrace{(N+1)P(k,t+1)}_{\text{\# k-nodes at time t+1}} = \underbrace{NP(k,t)}_{\text{\# k-nodes at time t}} + \underbrace{\frac{k-1}{2}P(k-1,t)}_{\text{Gain of k-nodes via k-1} \rightarrow \text{k}} - \underbrace{\frac{k}{2}P(k,t)}_{\text{Loss of k-nodes via k} \rightarrow \text{k+1}}$$

We do not have *k=0,1,...,m-1* nodes in the network (each node arrives with degree *m*)
→ We need a separate equation for degree *m* modes

$$\underbrace{(N+1)P(m,t+1)}_{\text{\# m-nodes at time t+1}} = \underbrace{NP(m,t)}_{\text{\# m-nodes at time t}} + \underbrace{1}_{\text{Add one m-degeree node}} - \underbrace{\frac{m}{2}P(m,t)}_{\text{Loss of an m-node via m} \rightarrow \text{m+1}}$$

A.-L.Barabási, R. Albert and H. Jeong, *Physica* A **272,** 173 (1999)

MFT - Degree Distribution: Rate Equation

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

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

We assume that there is a stationary state in the N=t→∞ limit, when P(k,∞)=P(k)

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

$$(N+1)P(m,t+1) - NP(m,t) \to P(m)$$

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

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

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

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

MFT - Degree Distribution: Rate Equation

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

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

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

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

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

$m+3 \rightarrow k$

...

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

$$P(k)\sim k^{-3}$$

for large k

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

MFT - Degree Distribution: A Pretty Caveat

Start from eq. $$P(k) = \frac{k-1}{2} P(k-1) - \frac{k}{2} P(k)$$

$$2P(k) = (k-1)P(k-1) - kP(k) = -P(k-1) - k[P(k) - P(k-1)]$$

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

$$P(k) = -\frac{1}{2}\frac{\partial [kP(k)]}{\partial k}$$

Its solution is: $P(k) \sim k^{-3}$

Dorogovtsev and Mendes, 2003

All nodes follow the same growth law

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

In limit: $A\frac{k_i}{\sum_j k_j} = A\frac{k_i}{2mt}$ So:

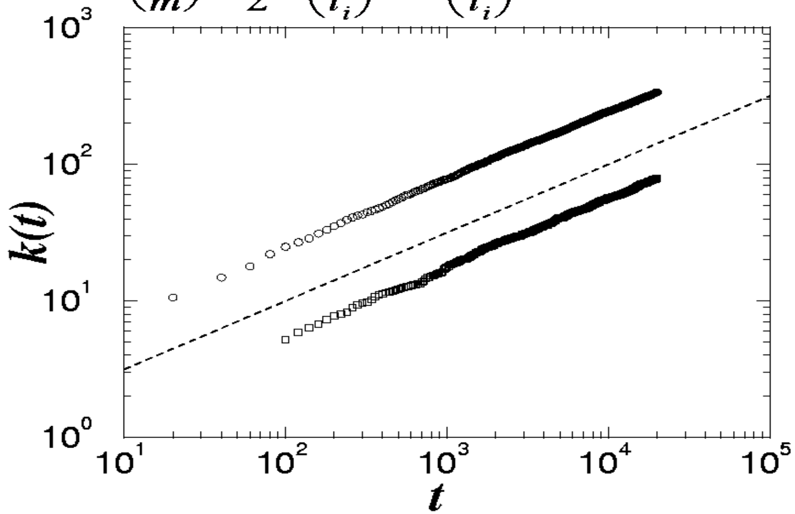
$$m = \sum_i \frac{\Delta k_i}{dt} = \sum_i A\frac{k_i}{2mt} = A$$

Use: $\sum_j k_j = 2m(t-1) + \frac{m_0(m_0-1)}{2}$

During a unit time (time step): *Δk=m* → *A=m*

$$\frac{\partial k_i}{\partial t} = \frac{k_i}{2t} \qquad \frac{\partial k_i}{k_i} = \frac{\partial t}{2t} \qquad \int_m^k \frac{\partial k_i}{k_i} = \int_{t_i}^t \frac{\partial t}{2t}$$

$$\ln\left(\frac{k}{m}\right) = \frac{1}{2}\ln\left(\frac{t}{t_i}\right) = \ln\left(\frac{t}{t_i}\right)^{1/2}$$



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

β: dynamical exponent

A.-L.Barabási, R. Albert and H. Jeong, *Physica* A **272,** 173 (1999)

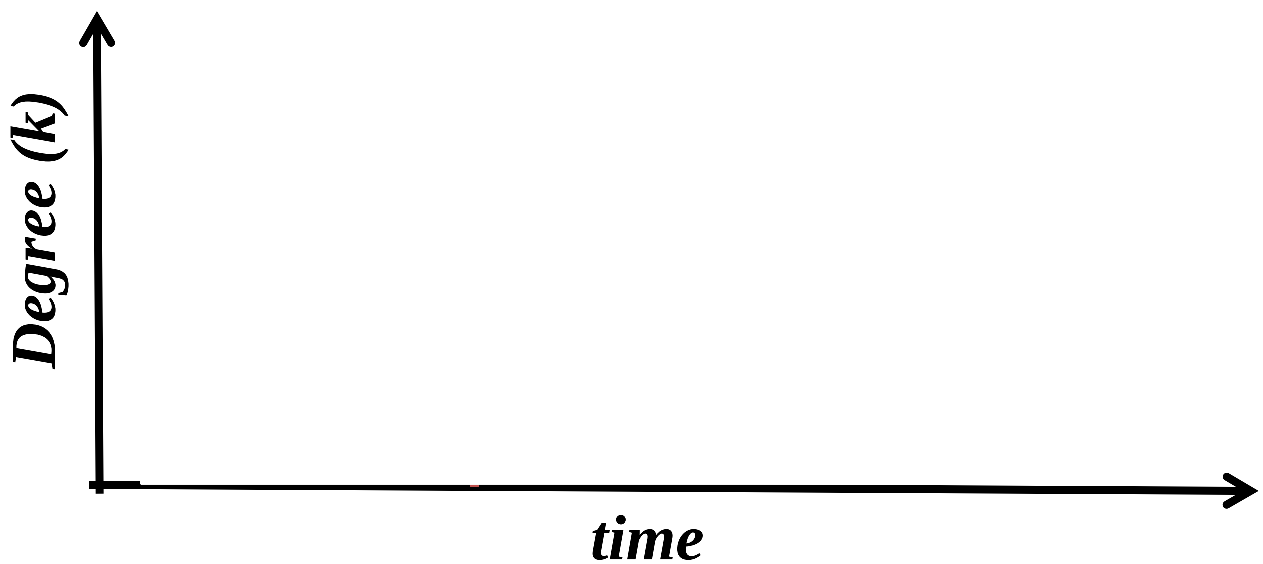
Fitness Model

Fitness Model: Can Latecomers Make It?

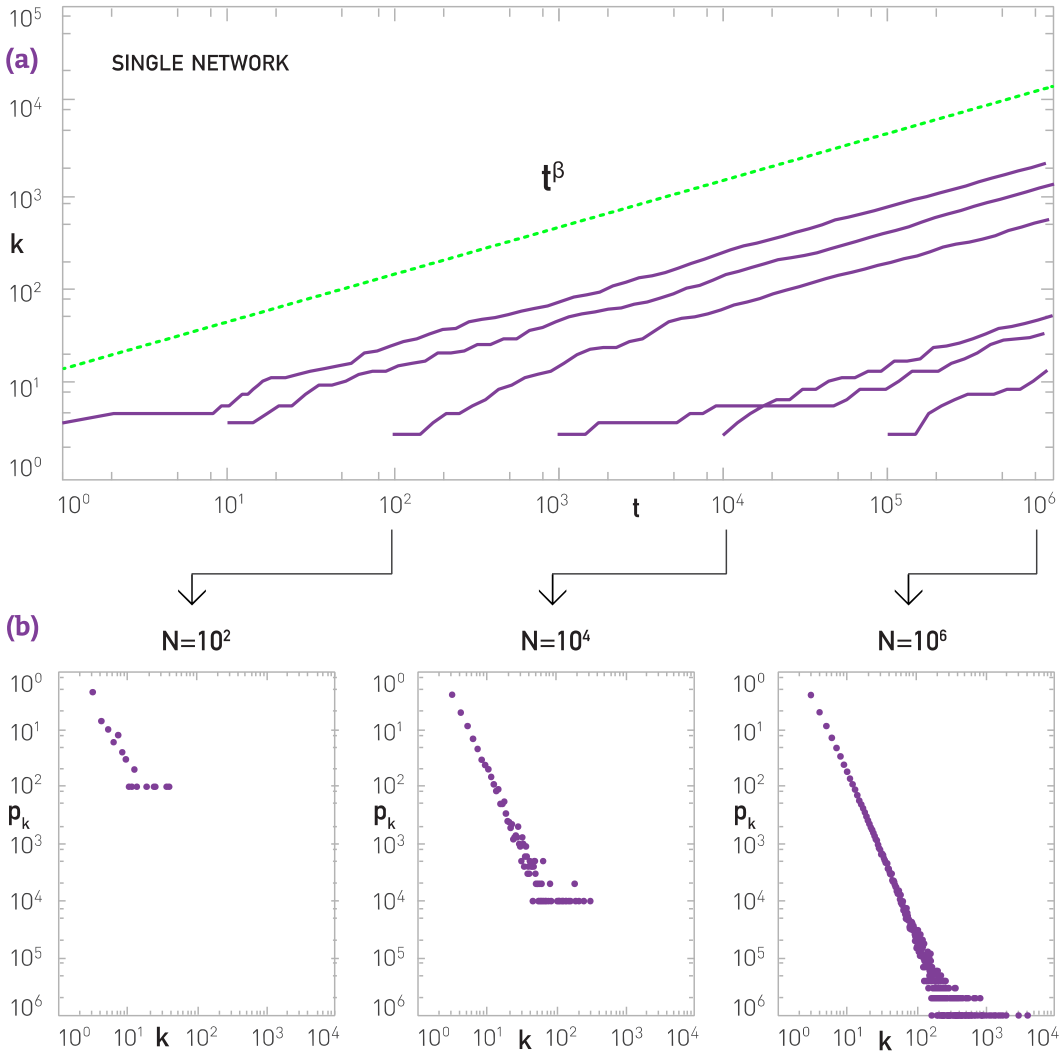
SF model: $k(t) \sim t^{1/2}$ **(first mover advantage)**

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

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



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

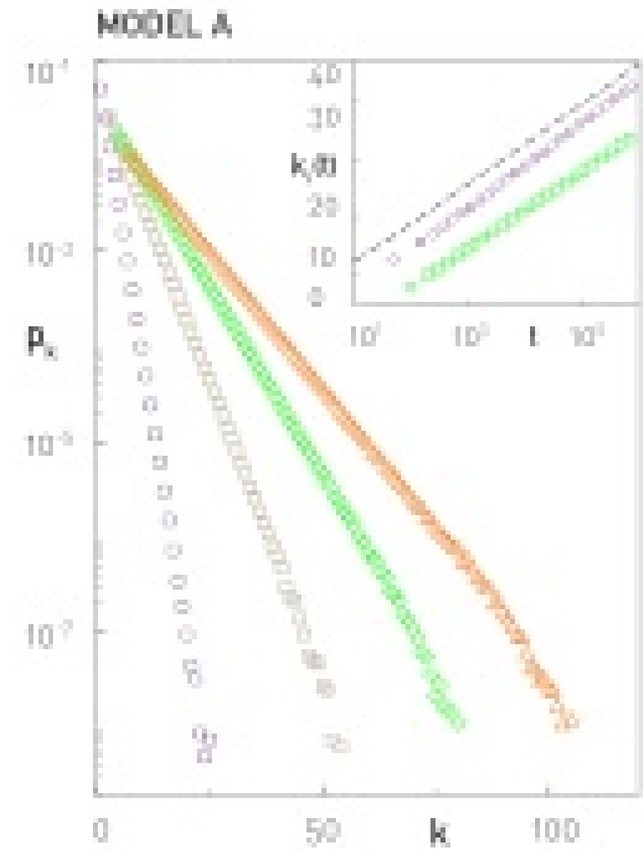
growth	~~preferential attachment~~

$\Pi(k_i)$: uniform

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

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

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

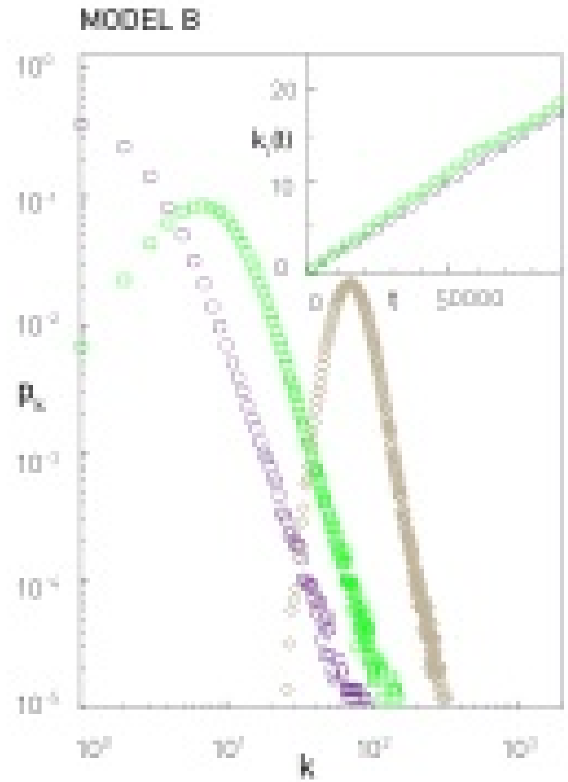


MODEL B

~~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) →

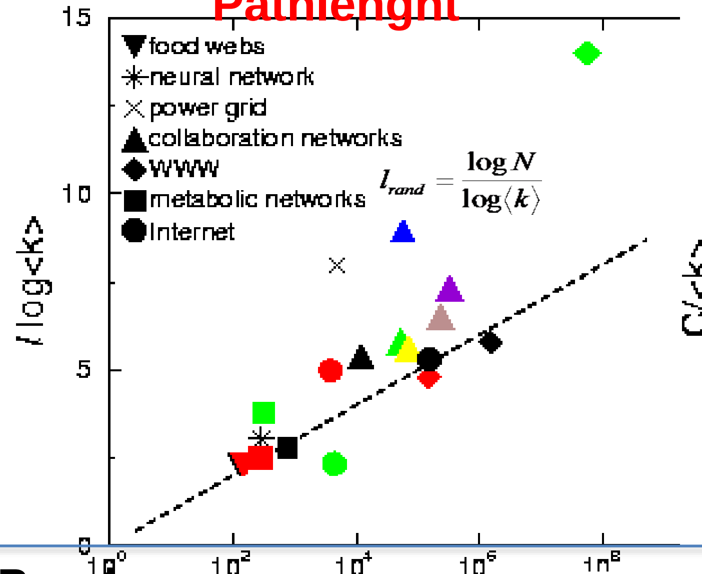
→ Gaussian → Fully Connected

Do we need both growth and preferential attachment?

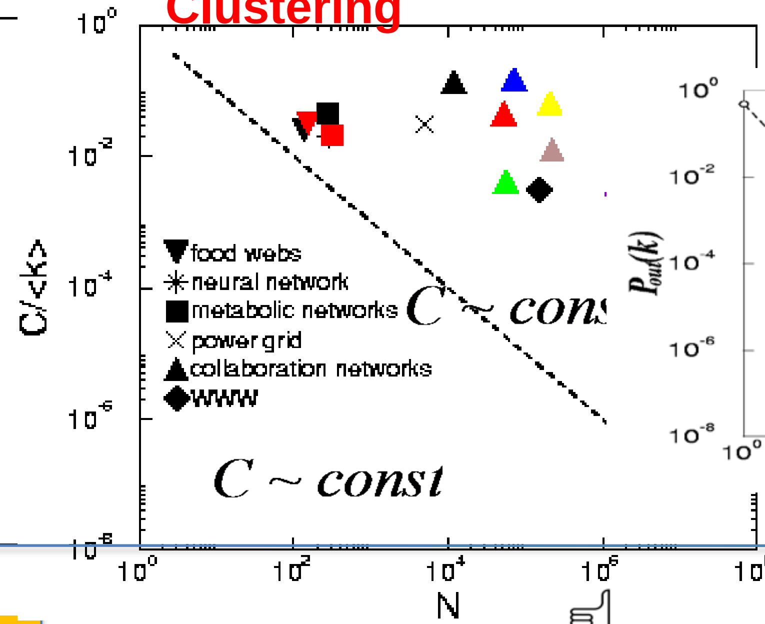
YEP

EMPIRICAL DATA FOR REAL NETWORKS

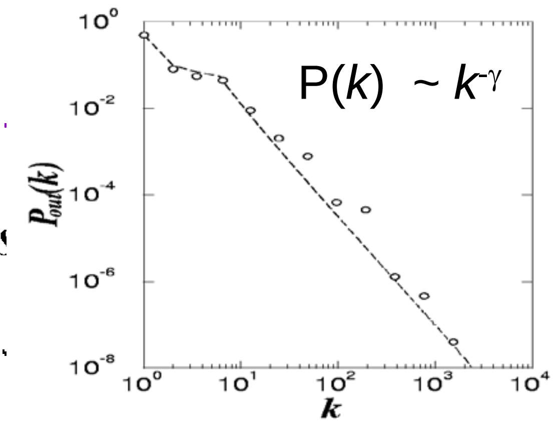
Pathlenght



Clustering



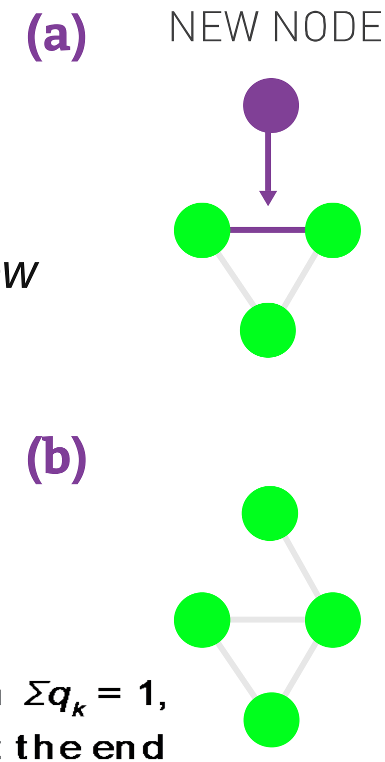
Degree Distr.



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

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

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

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

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

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

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

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.

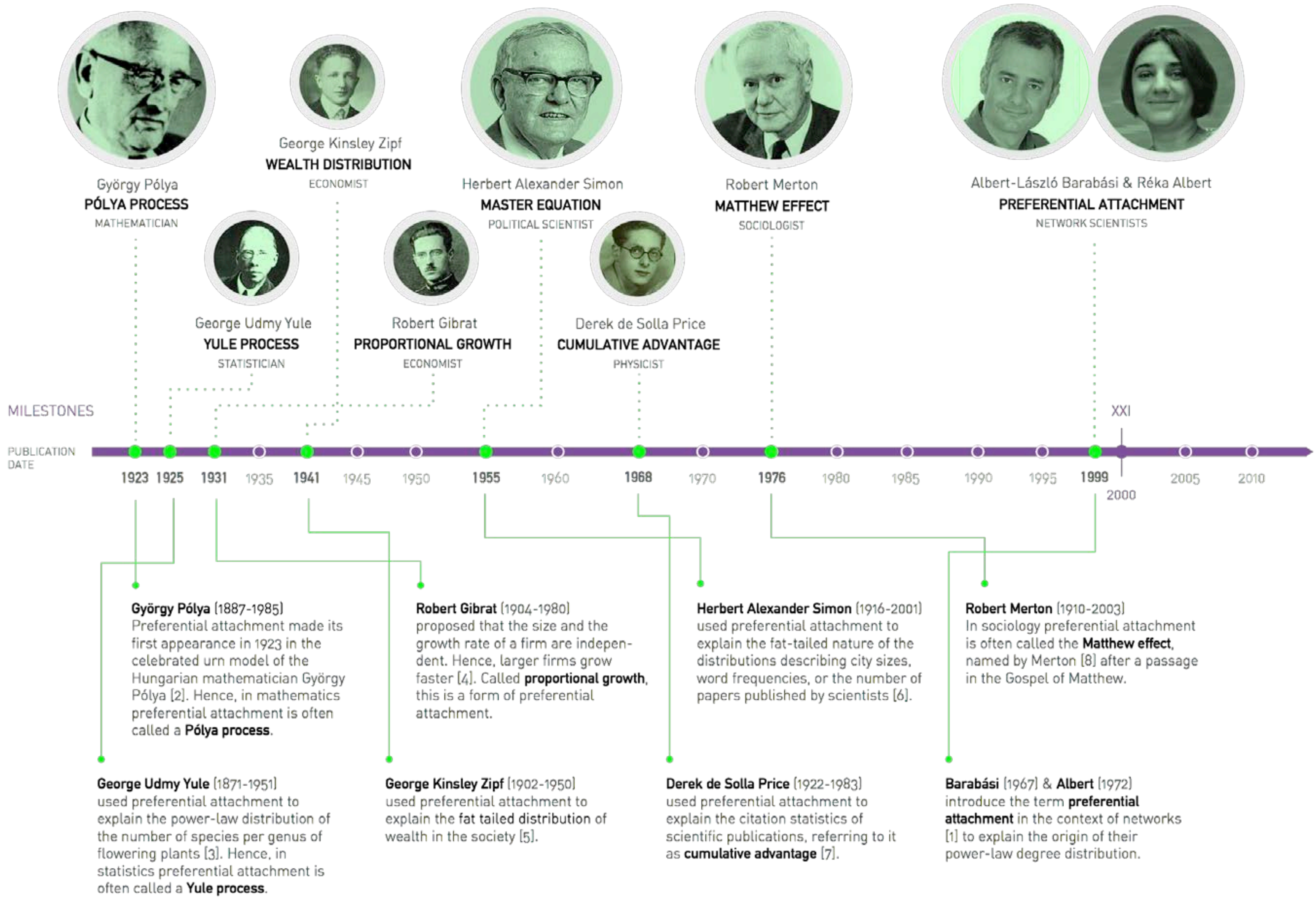
1961

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

1961

Herbert

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


György Pólya
PÓLYA PROCESS
MATHEMATICIAN
George Kinsley Zipf
WEALTH DISTRIBUTION
ECONOMIST
George Udmy Yule
YULE PROCESS
STATISTICIAN
Robert Gibrat
PROPORTIONAL GROWTH
ECONOMIST
Herbert Alexander Simon
MASTER EQUATION
POLITICAL SCIENTIST
Derek de Solla Price
CUMULATIVE ADVANTAGE
PHYSICIST
Robert Merton
MATTHEW EFFECT
SOCIOLOGIST
Albert-László Barabási & Réka Albert
PREFERENTIAL ATTACHMENT
NETWORK SCIENTISTS
MILESTONES
XXI
PUBLICATION DATE
1923
1925
1931
1935
1941
1945
1950
1955
1960
1968
1970
1976
1980
1985
1990
1995
1999
2000
2005
2010
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 indepen-dent. 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.