# Network Science

# Class 5: BA model

**Albert-László Barabási**
With
Emma K. Towlson, Sebastian Ruf, Michael Danziger and Louis Shekhtman

# Introduction

Hubs represent the most striking difference between a random and a scale-free network. Their emergence in many real systems raises several fundamental questions:

- Why does the random network model of Erdős and Rényi fail to reproduce the hubs and the power laws observed in many real networks?

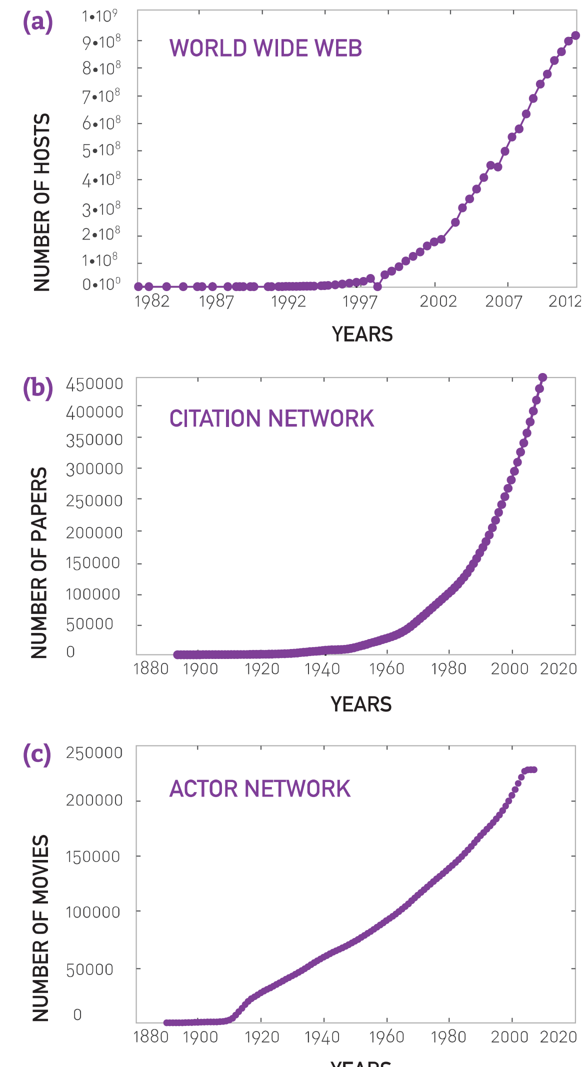
- Why do so different systems as the WWW or the cell converge to a similar scale-free architecture?

# Growth and preferential attachment

# BA MODEL: Growth

**ER model**:
the number of nodes, N, is fixed (static models)

**networks expand through the addition of new nodes**




Barabási & Albert, *Science* **286,** 509 (1999)

# BA MODEL: Preferential attachment

ER model: links are added randomly to the network

**New nodes prefer to connect to the more connected nodes**

Barabási & Albert, *Science* **286,** 509 (1999)

# Section 2: Growth and Preferential Sttachment

The random network model differs from real networks in two important characteristics:

**Growth:** While the random network model assumes that the number of nodes is fixed (time invariant), real networks are the result of a growth process that continuously increases.

**Preferential Attachment:** While nodes in random networks randomly choose their interaction partner, in real networks new nodes prefer to link to the more connected nodes.

Barabási & Albert, *Science* **286,** 509 (1999)

# The Barabási-Albert model

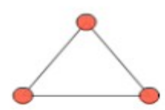
# Origin of SF networks: Growth and preferential attachment

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

WWW : addition of new documents

(2) New nodes prefer to link to highly connected nodes.

WWW : linking to well known sites



Barabási & Albert, *Science* **286,** 509 (1999)

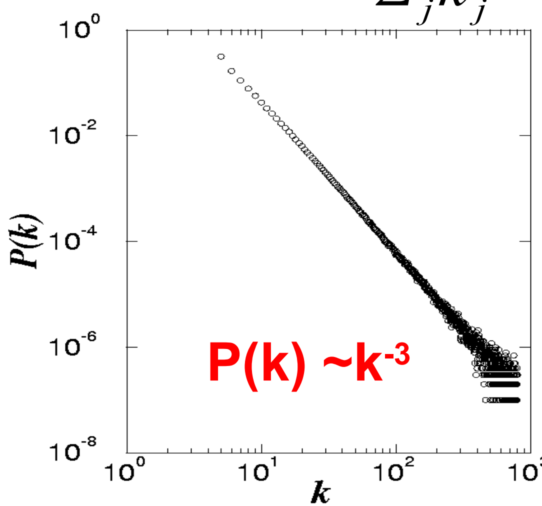
**GROWTH:**

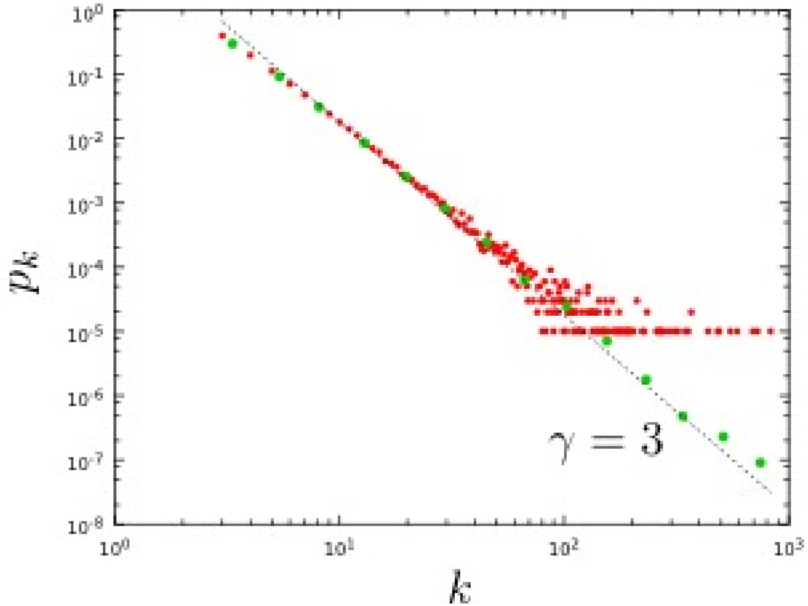
add a new node with m links

**PREFERENTIAL ATTACHMENT:**

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

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



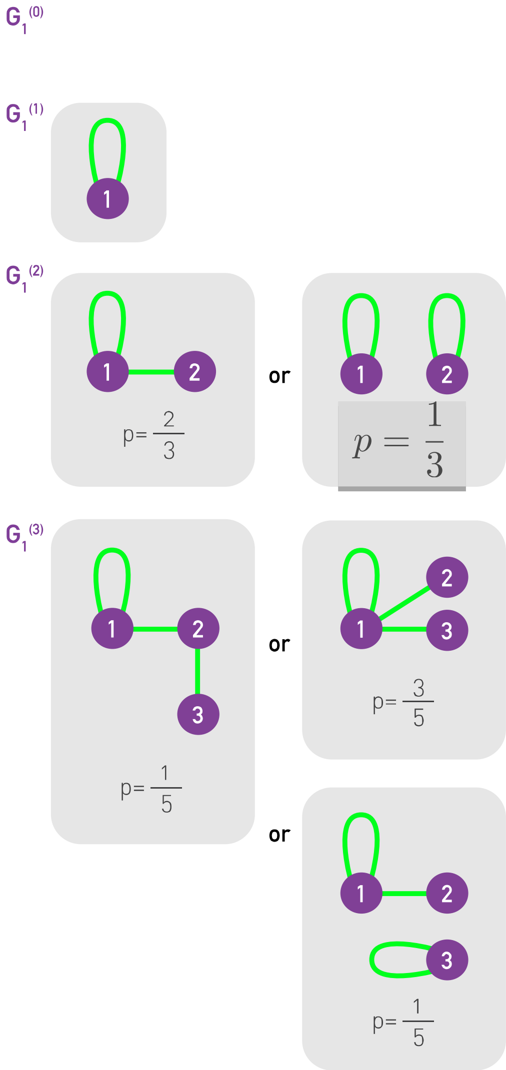


$10^0$
$10^{-1}$
$10^{-2}$
$10^{-3}$
$10^{-4}$
$10^{-5}$
$10^{-6}$
$10^{-7}$
$10^{-8}$
$p_k$
$\gamma = 3$
$10^0$
$10^1$
$10^2$
$10^3$
$k$

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.

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



# Degree dynamics

# All nodes follow the same growth law

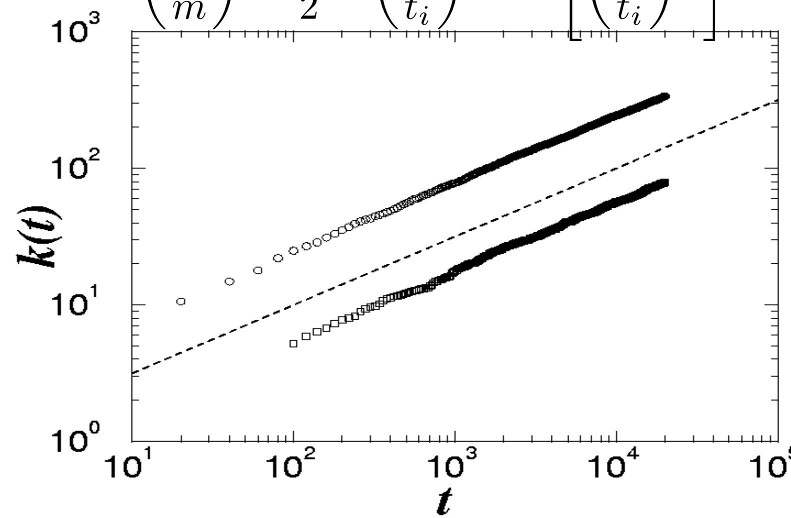
$$\frac{\partial k_i}{\partial t} \propto \Pi(k_i) = A\frac{k_i}{\sum_j k_j}$$

Use: $\sum_j k_j = 2mt$

During a unit time (time step): $\Delta k = m$ → $A = m$

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

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



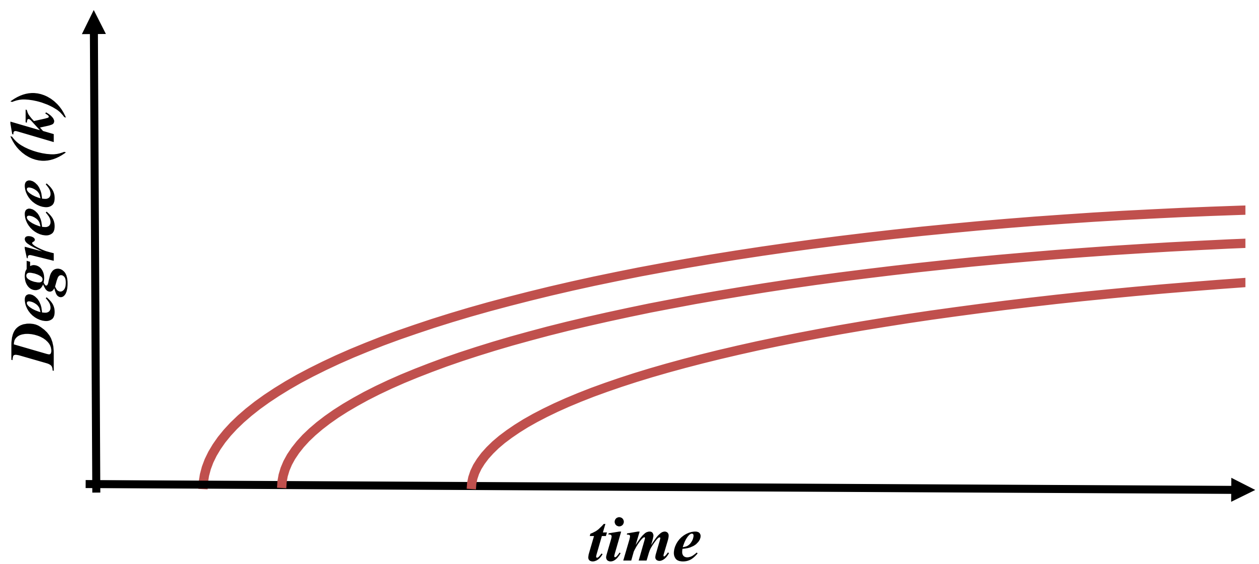

$$k_i(t) = m\left(\frac{t}{t_i}\right)^{\beta} \qquad \beta = \frac{1}{2}$$

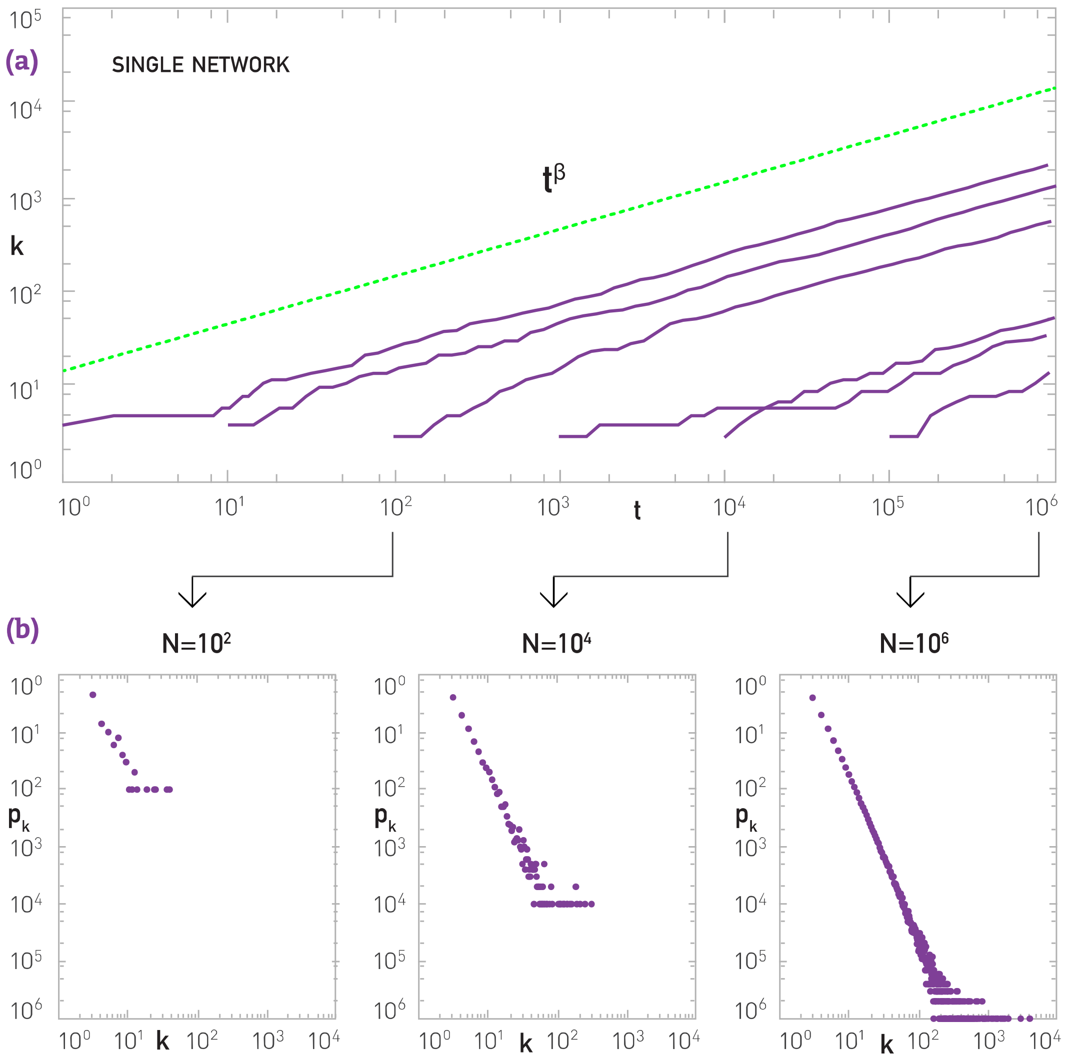
$\beta$: dynamical exponent

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

# All nodes follow the same growth law

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






- The degree of each node increases following a power-law with the same dynamical exponent $\beta$ =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)

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

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.

# Degree distribution

# Degree distribution

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

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

$$P(t_i) = \frac{1}{m_0 + t} \qquad P(t_i < \tau) = \frac{1}{m_0 + t}\int_0^{\tau} dt_i = \frac{\tau}{m_0 + t}$$

$$P(k) = P\left(t_i \le \frac{m^{1/\beta} t}{k^{1/\beta}}\right) = 1 - \frac{m^{1/\beta} t}{k^{1/\beta}(t + m_0)}$$

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

$$\gamma = 3$$

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

# Degree distribution

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

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

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

(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)

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.

# Degree distribution

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

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

$$\gamma = 3$$

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

(i) The degree exponent is independent of m.

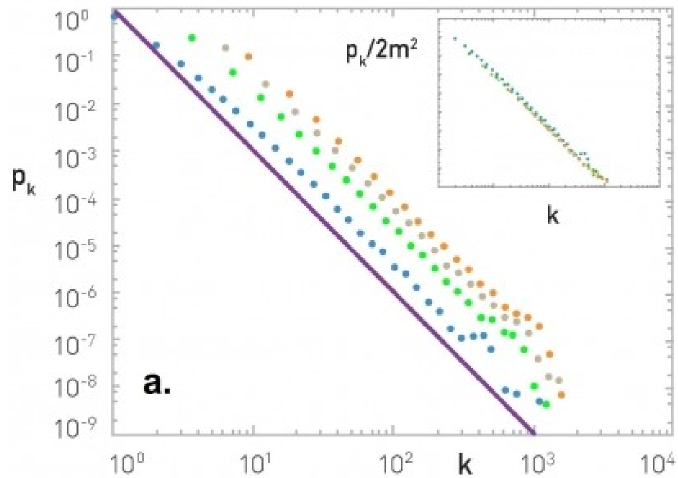
(ii) As the power-law describes systems of rather different ages and sizes, it is expected that a correct model should provide a time-independent degree distribution. Indeed, asymptotically the degree distribution of the BA model is independent of time (and of the system size N)

→ the network reaches a stationary scale-free state.

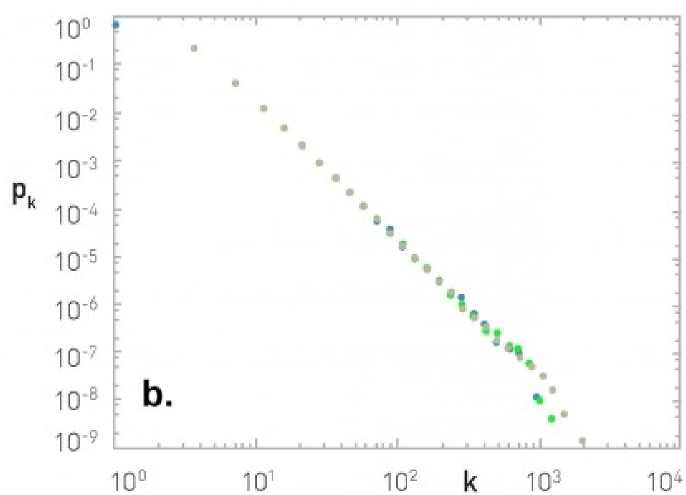
(iii) The coefficient of the power-law distribution is proportional to $m^2$.

# 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)** 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.

# absence of growth and preferential attachment

# MODEL A

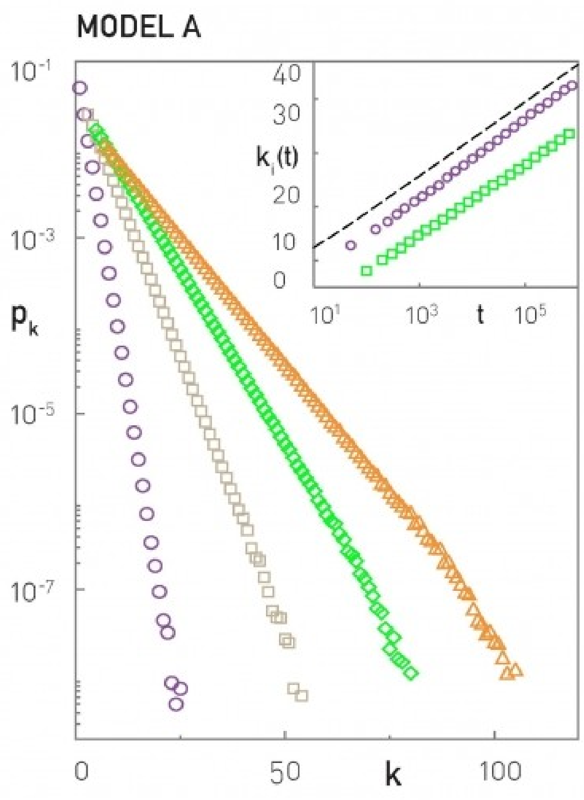
growth ~~preferential attachment~~

$\Pi(k_i)$ : 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

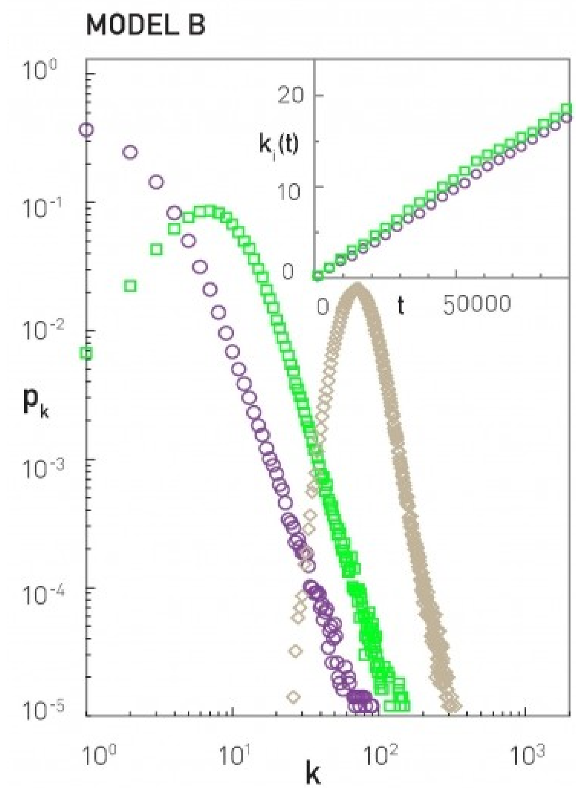
~~growth~~ preferential attachment

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

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

$p_k$ : power law (initially) →

→ Gaussian → Fully Connected



Do we need both growth and preferential attachment?

YEP.

# Preliminary project presentation (Apr. 28th)

5 slides

**Discuss:**

What are your nodes and links

How will you collect the data, or which dataset you will study

Expected size of the network (# nodes, # links)

What questions you plan to ask (they may change as we move along with the class).

Why do we care about the network you plan to study.