Other growth models

Introduction to Network Science
Carlos Castillo
Topic 06

Universitat Pompeu Fabra Barcelona

Sources

- Albert László Barabási: Network Science. Cambridge University Press, 2016.
  - Chapters 05 and 06

Actual network growth is complex

A snapshot of the Autodesk organizational hierarchy was taken each day between May 2007 and June 2011, a span of 1498 days.

Each day the entire hierarchy of the company is constructed as a tree with each employee represented by a circle, and a line connecting each employee with his or her manager.

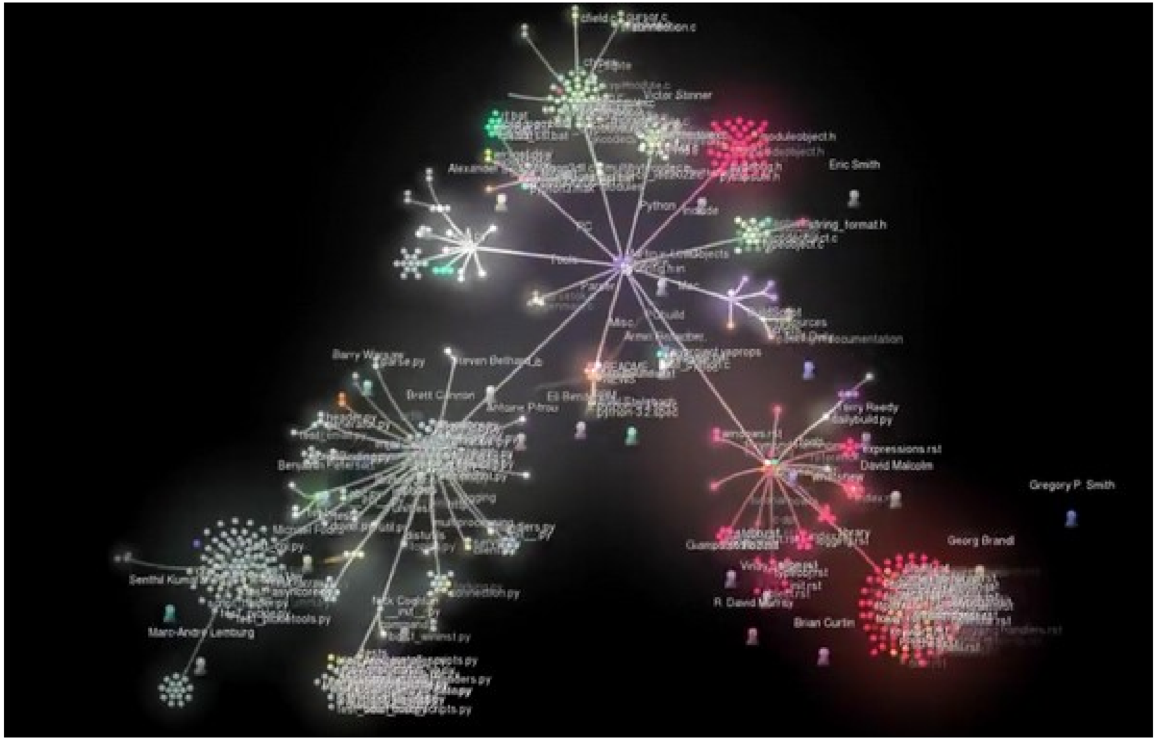
Larger circles represent managers with more employees working under them. The tree is then laid out using a force-directed layout algorithm.

From day to day, there are three types of changes that are possible:

- Employees join the company
- Employees leave the company
- Employees change managers

https://www.youtube.com/watch?v=mkJ-Uy5dt5g

Growth of Python
(Gource visualization)



https://www.youtube.com/watch?v=cNBtDstOTmA

Remember preferential attachment

- Start with m_0 nodes
- At every time step
  - Add one new node u
  - Repeat m times
    - Pick a node v with probability $\Pi(k_v) = \frac{k_v}{\sum_j k_j}$
    - Connect u to v

Two simple variants

- No preference
  - Nodes receiving inlinks are picked uniformly at random
- No growth
  - The network starts with N nodes
  - No new nodes are created

No preference model

- Write the process on paper
- Write $\Pi(k_i)$
- Noting that $\frac{d}{dt}k_i = m\Pi(k_i)$ obtain $k_i(t)$

$$\int \frac{a}{b+x} = a\log(b+x) + C$$

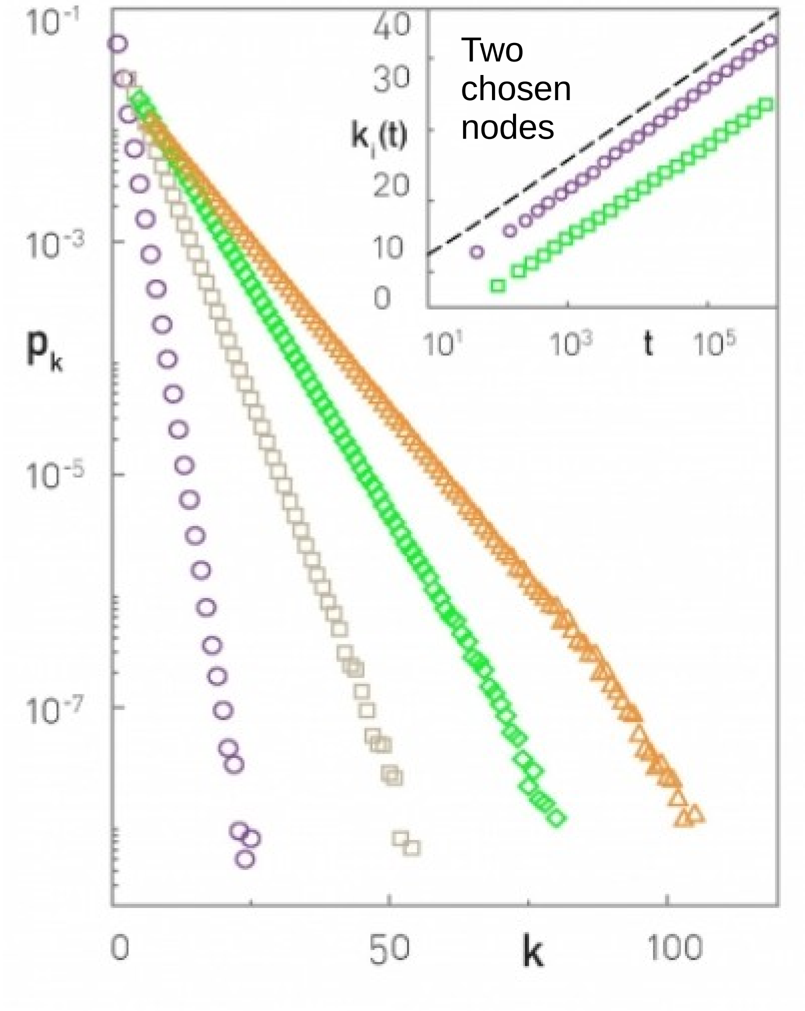
No preference model (cont.)

- Compute $Pr(k_i(t) > k)$ assuming large t, t_i
- Use it to compute

$$Pr(k_i(t) \leq k) = 1 - Pr(k_i(t) > k)$$

- Derive to obtain $p_k = Pr(k_i(t) = k)$

Consequences of the "no preference" model

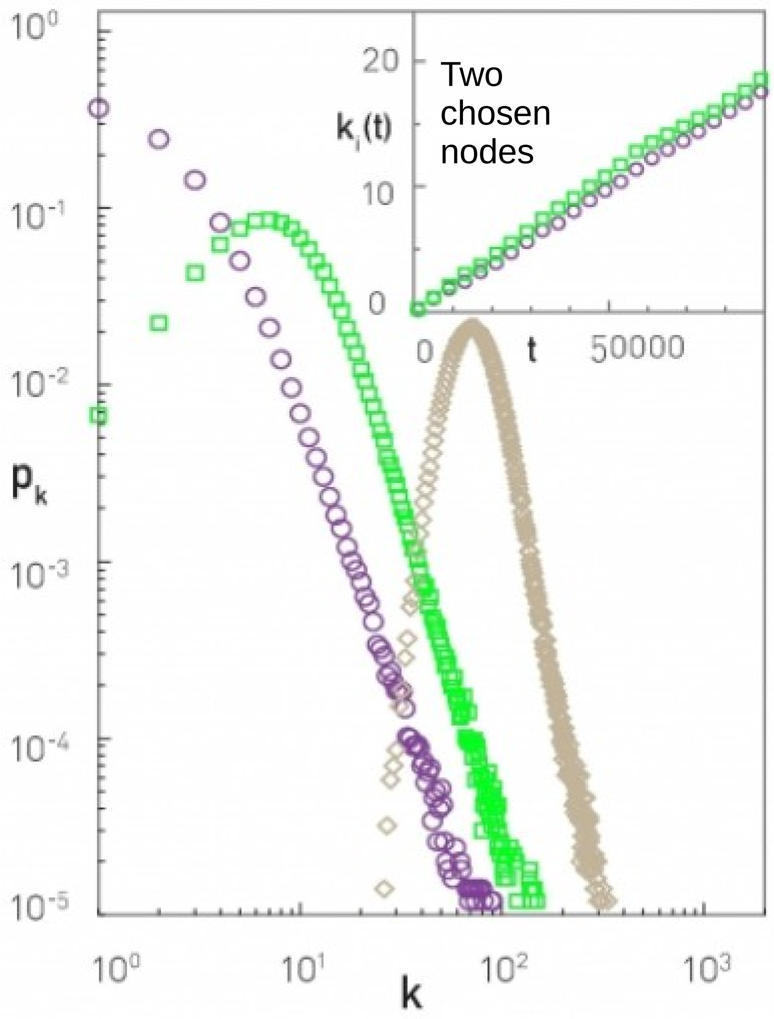


m=1, m=3, m=5, m=7

- Degree decays exponentially $p_k \propto e^{-k/m}$
- No power-law
- No large hubs

No growth model

- Write the process on paper
- You will need to impose $k_i(t_i) \neq 0$ why?
- Write $\Pi(k_i)$
- Noting that $\frac{d}{dt}k_i = \Pi(k_i)$ obtain $k_i(t)$



N=100K t=N, t=5N, t=40N

Consequences of the "no growth" model

- Degree grows linearly $k_i(t) \propto t$
- Degree distribution is not stationary

Sub-linear and super-linear preferential attachment

- The model we have studied so far has **linear preferential attachment** because $\frac{d}{dt}k_i \propto k_i$
- We could imagine cases where $\frac{d}{dt}k_i \propto k_i^{\alpha}$ for $\alpha > 1$ or $\alpha < 1$

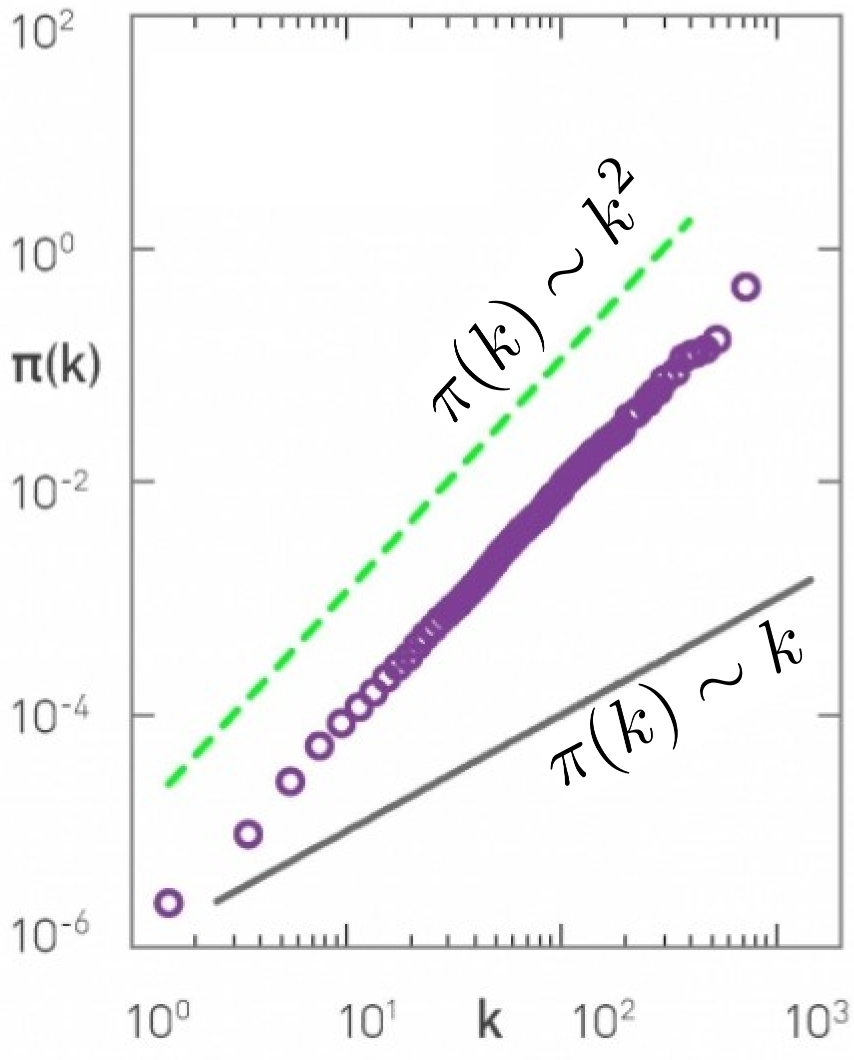
What do you think should happen in each case?

Let's measure preferential attachment

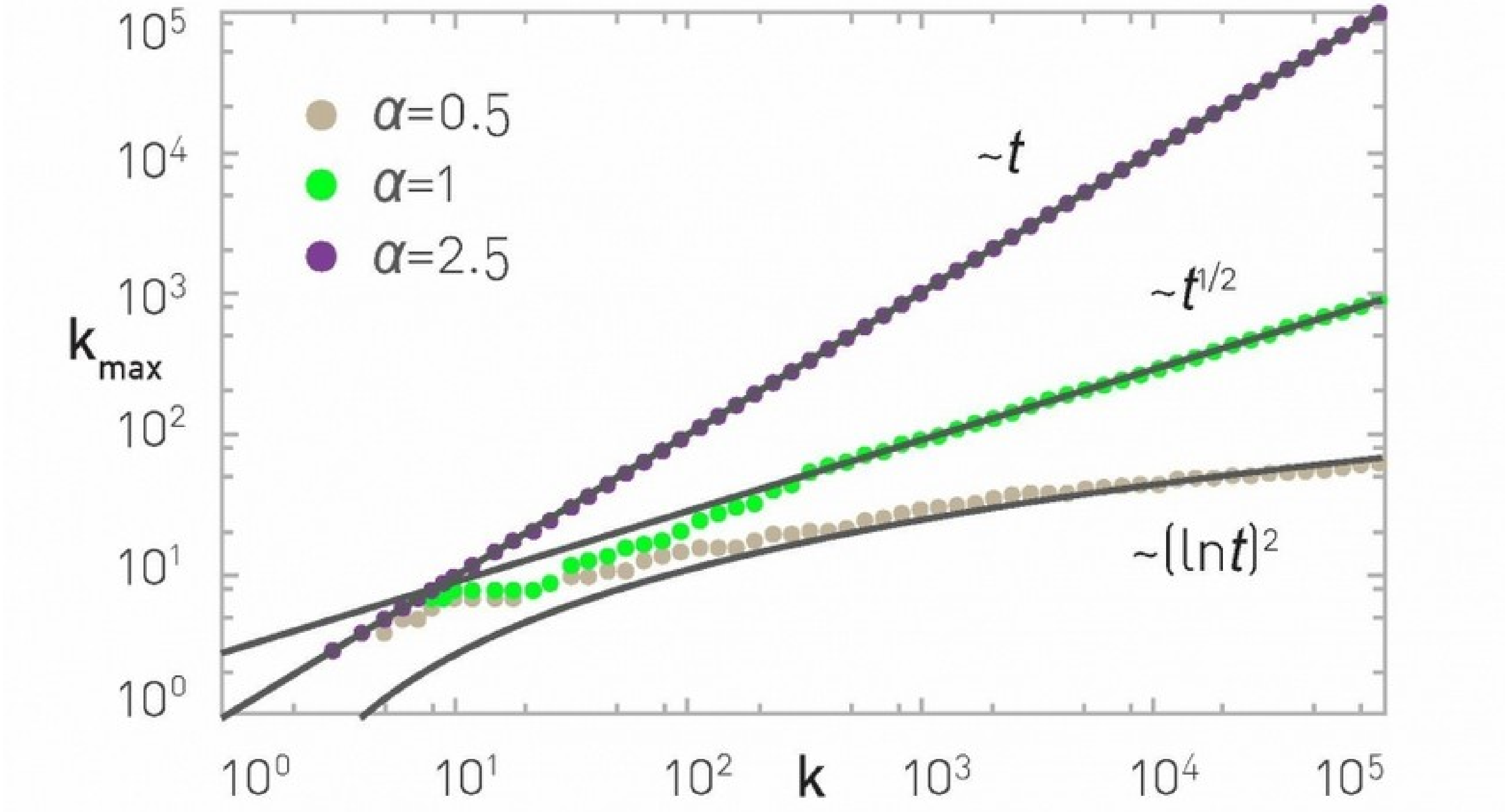
- We should try to measure $\Pi(k_i) \approx \frac{\Delta k_i}{\Delta t}$
- This can be too noisy
  - Why?
- Instead we will measure $\pi(k) = \sum_{k_i=0}^{k} \Pi(k_i)$
- If $\Pi(k_i)$ is constant $\pi(k) \propto k$
- If $\Pi(k_i) \propto k$ then $\pi(k) \propto k^2$

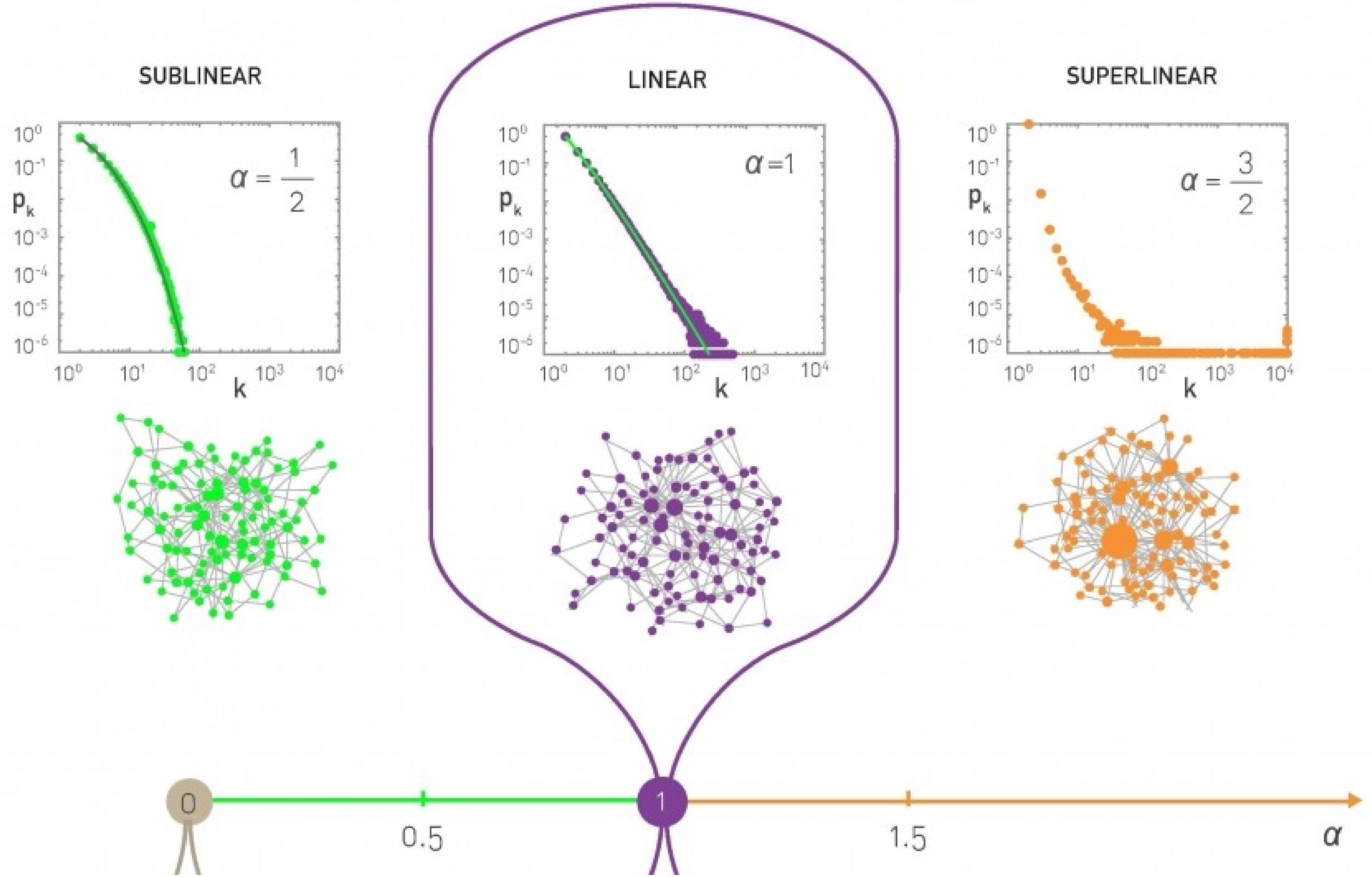
Preferential attachment in a citation network

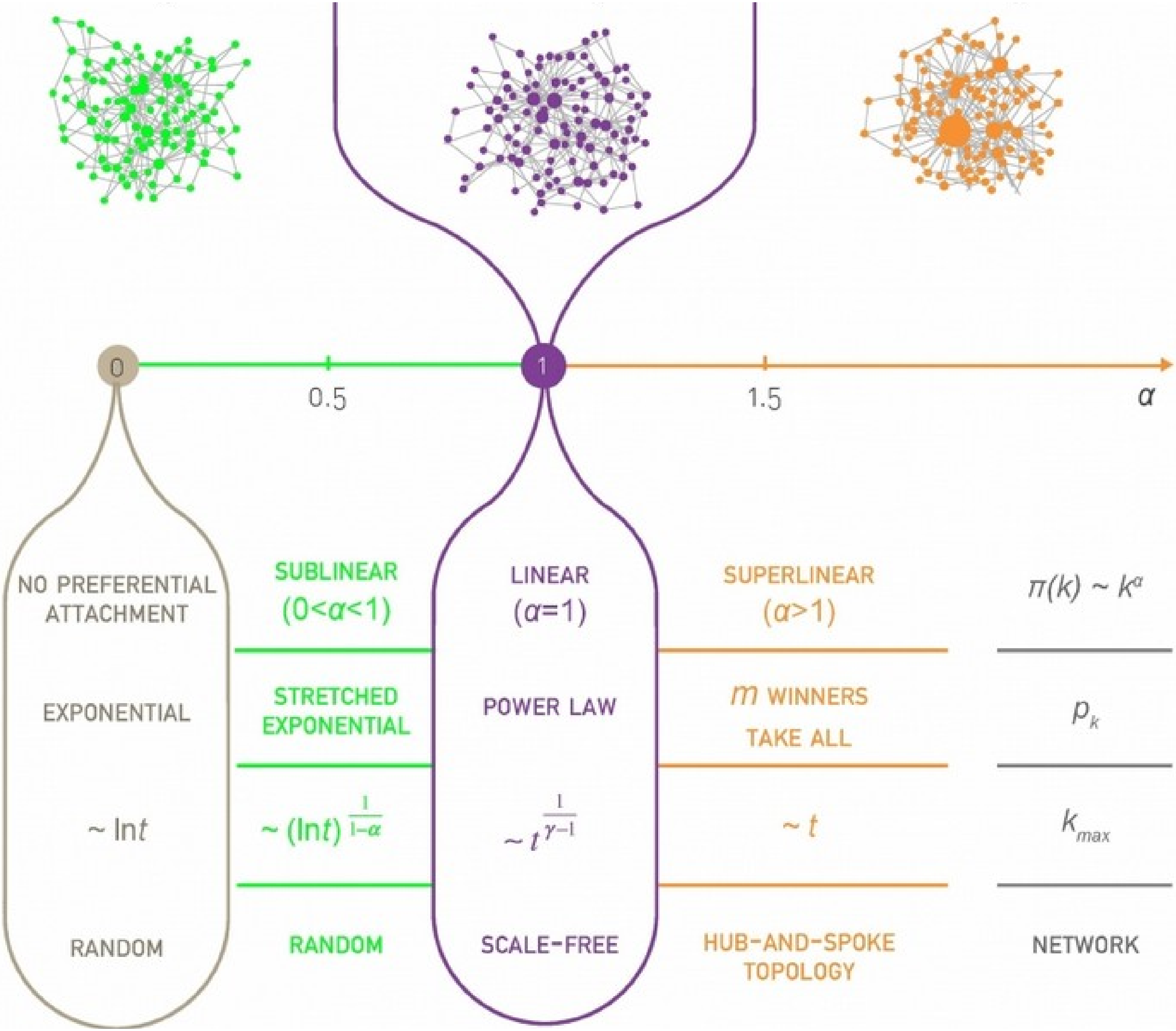
- We observe this follows preferential attachment (with $\alpha = 1$)
- But there may be cases where this does not hold



The degree of the largest hub k_{max}



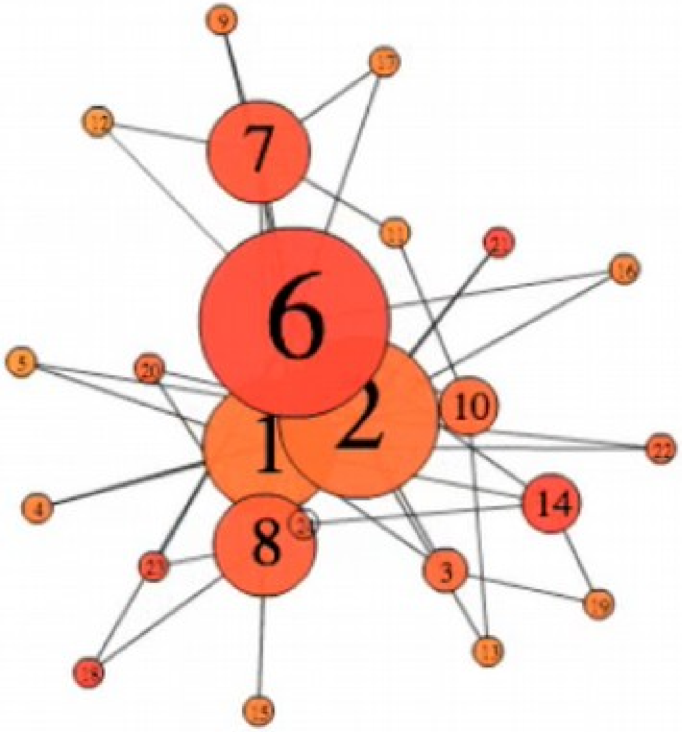

SUBLINEAR
LINEAR
SUPERLINEAR
$\alpha = \frac{1}{2}$
$\alpha = 1$
$\alpha = \frac{3}{2}$
p_k
k
0
0.5
1
1.5
α



0
0.5
1
1.5
α
NO PREFERENTIAL ATTACHMENT
EXPONENTIAL
~ ln$t$
RANDOM
SUBLINEAR
($0<\alpha<1$)
STRETCHED EXPONENTIAL
$\sim (\ln t)^{\frac{1}{1-\alpha}}$
RANDOM
LINEAR
($\alpha=1$)
POWER LAW
$\sim t^{\frac{1}{\gamma-1}}$
SCALE-FREE
SUPERLINEAR
($\alpha>1$)
m WINNERS TAKE ALL
$\sim t$
HUB-AND-SPOKE TOPOLOGY
$\pi(k) \sim k^{\alpha}$
p_k
k_{max}
NETWORK

“Good get richer” (incl. Bianconi-Barabási model)

“Good get richer” simulation (color saturation is attractiveness)



http://networksciencebook.com/images/ch-06/video-6-1.mov

“Good get richer”

- A “good get richer” model is one where
  - Each node has an “attractiveness” (called “fitness”)
  - Preferential attachment is guided by this fitness η_i
- The probability of connecting to node i is:

$$\Pi_i = \frac{\eta_i k_i}{\sum_j \eta_j k_j}$$

Degree dynamics

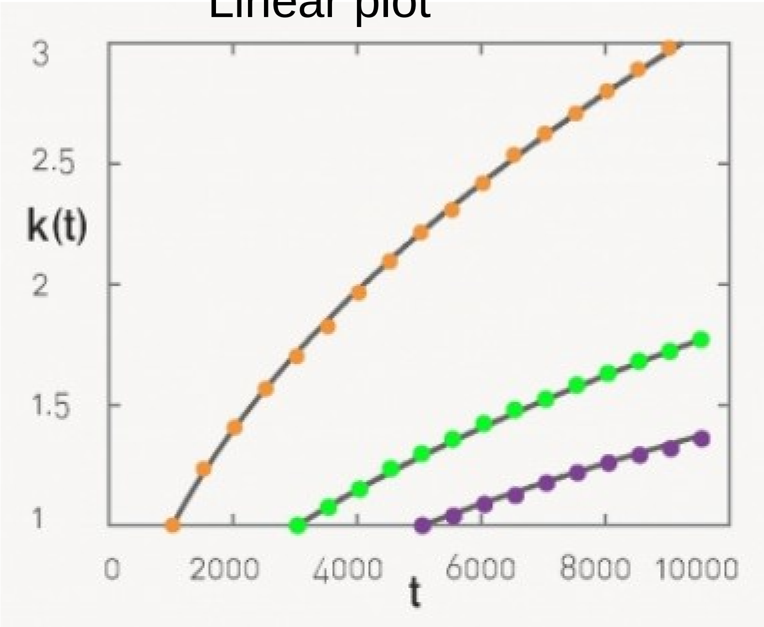
$$\frac{d}{dt}k_i = m\frac{\eta_i k_i}{\sum_j \eta_j k_j}$$

$$k_i(t, t_i, \eta_i) = m\left(\frac{t}{t_i}\right)^{\beta(\eta_i)}$$

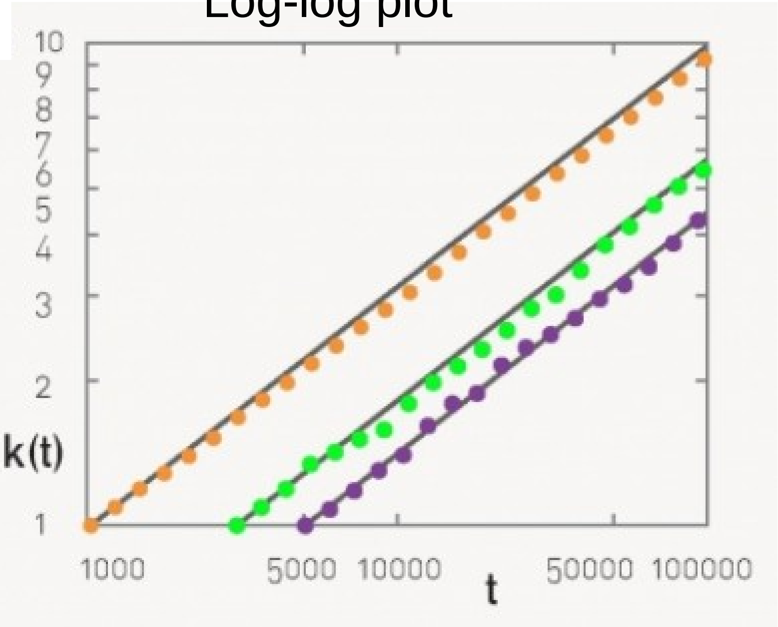
- With the dynamic exponent $\beta(\eta_i) \propto \eta_i$
- Remember that in linear preferential attachment $\beta = 1/2$ (for all nodes)

In preferential attachment (BA) a “younger” node cannot overtake an “older” node

Linear plot

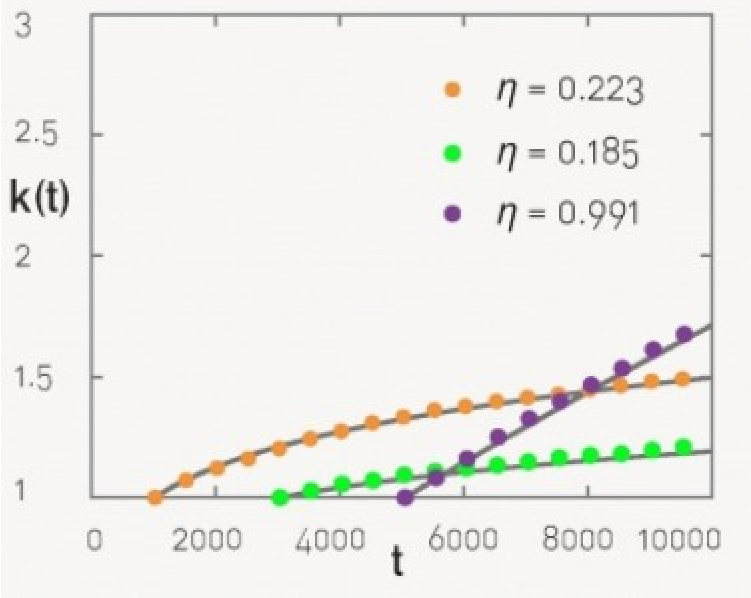


Log-log plot

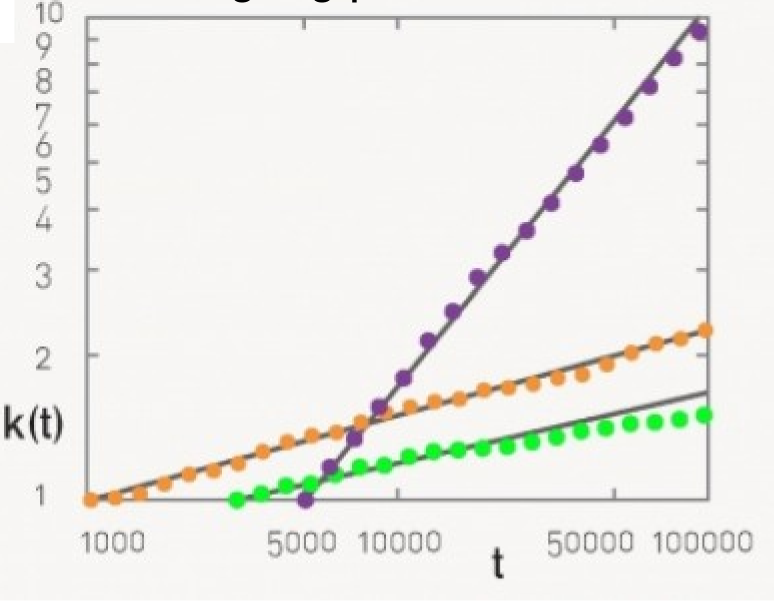


In good-get-richer (Bianconi-Barabási) this depends on node fitness

Linear plot



Log-log plot



Degree distribution

$$p_k \propto \int \frac{\rho(\eta)}{\eta} \left(\frac{m}{k}\right)^{\frac{c}{\eta}+1} d\eta \qquad\qquad \eta \sim \rho(\eta)$$

- When η is constant this reduces to BA
- When η is uniformly distributed in [0, 1] this also yields a power law but instead of $\gamma = 3$ we get $\gamma \approx 2.3$

Which distribution is more heterogeneous?

Sick Boy's unified theory of life from *Trainspotting (1996)*

In English: https://www.youtube.com/watch?v=pQD-dXfHrvk
In Spanish: https://www.youtube.com/watch?v=cN_WbiuqyQU
English (bad audio) subs in Spanish: https://www.youtube.com/watch?v=4xTWD9GNRFA

Aging effects

- Models without fitness but with a negative effect of age

$$\Pi(ki, t - t_i) \sim k(t - t_i)^{-v}$$

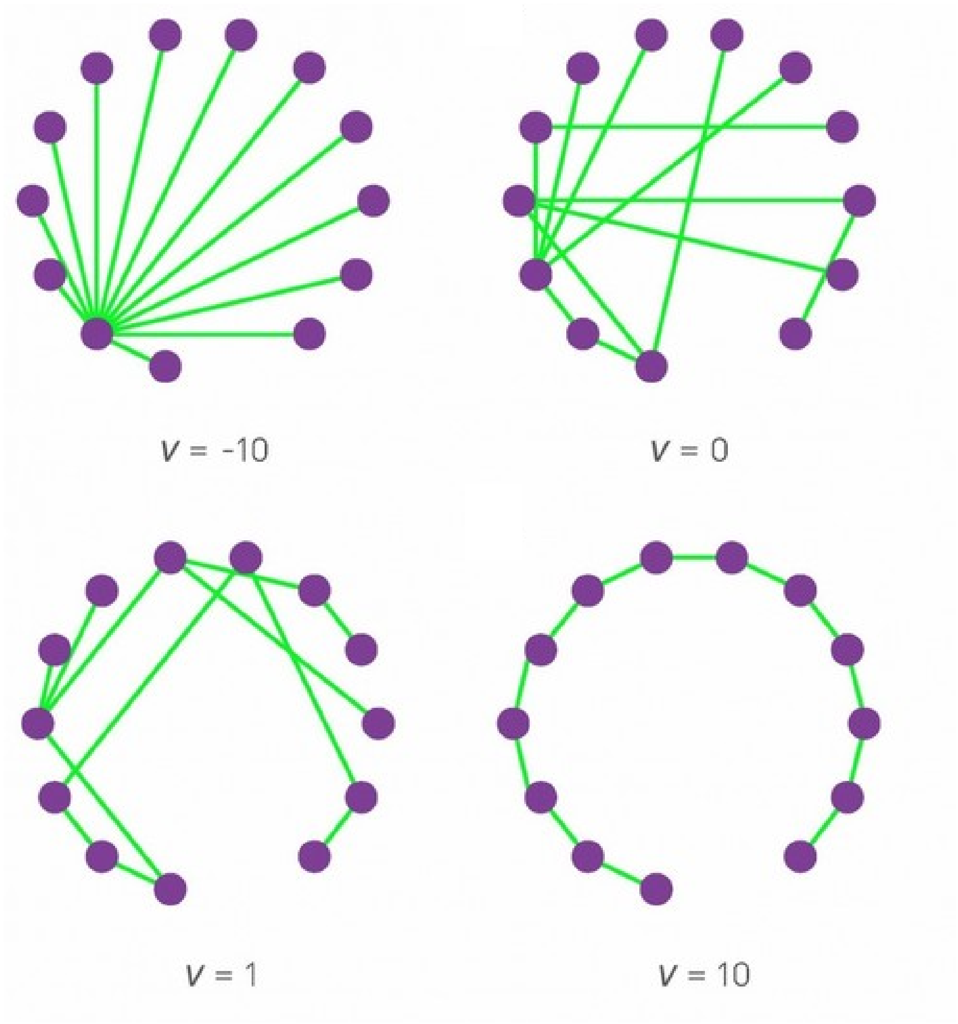
- Older nodes accumulate links more slowly
- Parameter v is the decay factor

Qualitatively, what would you expect if:

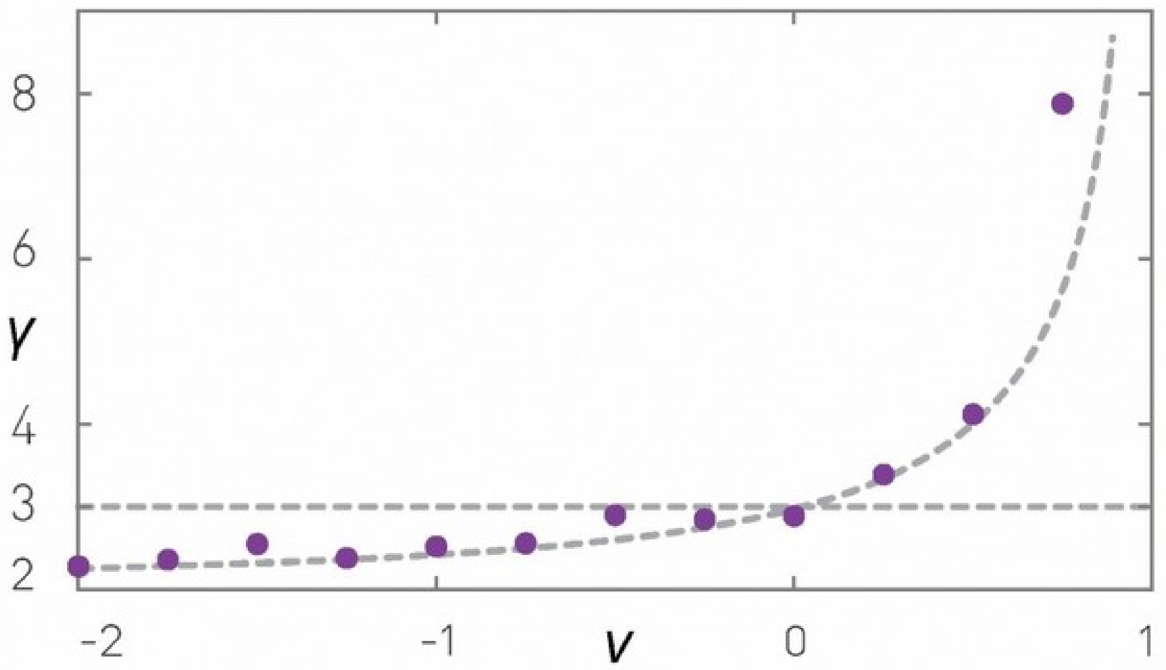
$v < 0 \qquad v = 0 \qquad v \gg 1$

Aging effects

- $v < 0$ favors older nodes
- $v = 0$ is simply preferential attachment
- $v >> 1$ means only youngest are linked



Power-law exponent in models with aging (N=10K, m=1)



More heterogeneous ⟷ More homogeneous