Other Graph Evolution Models

Social Networks Analysis and Graph Algorithms

Prof. Carlos "ChaTo" Castillo — https://chato.cl/teach

Universitat
Pompeu Fabra
Barcelona

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

When a company is aquired, they appear to "explode" out of their new location within the company.

0:57 / 2:35

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

Other graph evolution models

- Uniform random attachment

- Sub-linear and super-linear preferential attachment

- Good-get-richer

- Aging effects

- Link selection

- Copy model

- No preference and no growth

Sources

- A. L. Barabási (2016). Network Science –
  Chapter 05 and Chapter 06

Uniform Random Attachment

Growth in an ER network

- Two assumptions in ER networks:

  - There are N nodes that **pre-exist**

  - Nodes connect **at random**

- Let's challenge the first assumption

Uniform Attachment

- Network starts with m fully-connected nodes

- Time starts at $t_0 = m$

- At every time step we add 1 node

- **This node will have** m outlinks

Expected degree over time

- Probability of obtaining one link: m/t

  - Decreases over time

- Expected degree of node born at $m < i < t$

$$m + \frac{m}{i} + \frac{m}{i+1} + \frac{m}{i+2} + \cdots + \frac{m}{t} \approx m \left(1 + \log \left(\frac{t}{i} \right) \right)$$

Tail of degree distribution

- How many nodes of degree larger than K are there at time t? (Computation in "Advanced materials" at the end of these slides)

$$e^{-\frac{K-m}{m}}$$

- Decreases exponentially with K: it's vanishingly rare to find high-degree nodes
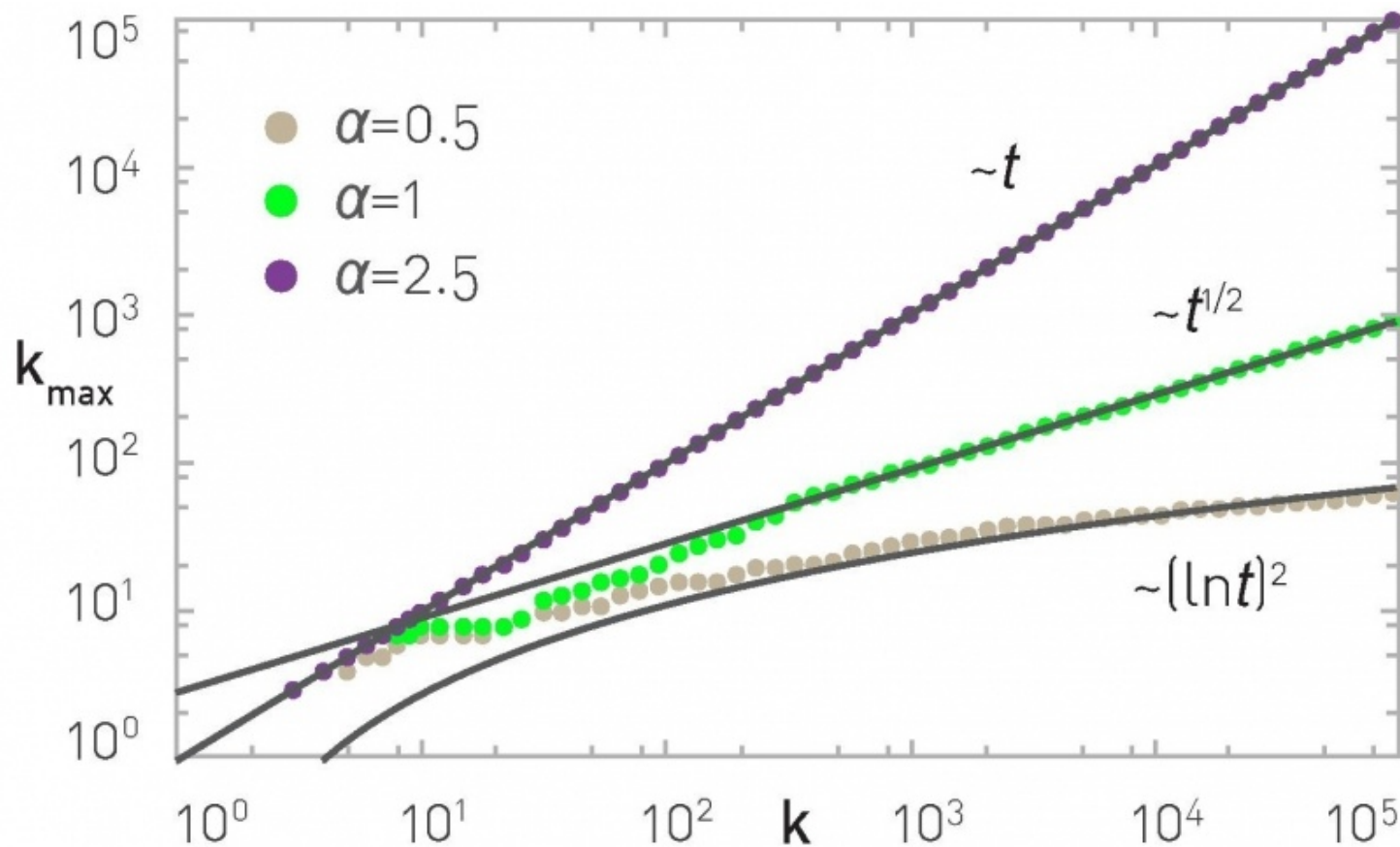
Sub-linear and super-linear preferential attachment

Sub-linear and super-linear preferential attachment

- The model we have studied so far has **linear preferential attachment** because
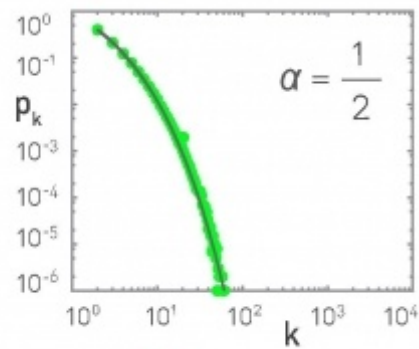
$$\frac{d}{dt}k_i \propto k_i$$

$$\frac{d}{dt}k_i \propto k_i^{\alpha}$$

- We $\alpha > 1$ ima $\alpha < 1$ cases where

  for   or

What do you think should happen in each case?

The degree of the largest hub k_{max}

SUBLINEAR

$$\alpha = \frac{1}{2}$$

LINEAR

$$\alpha = 1$$

SUPERLINEAR

$$\alpha = \frac{3}{2}$$

NO PREFERENTIAL ATTACHMENT   SUBLINEAR $(0<\alpha<1)$ LINEAR $(\alpha=1)$ SUPERLINEAR $(\alpha>1)$ $\pi(k) \sim k^{\alpha}$

EXPONENTIAL   STRETCHED EXPONENTIAL   POWER LAW m WINNERS TAKE ALL p_k

$\sim \ln t$ $\sim (\ln t)^{\frac{1}{1-\alpha}}$ $\sim t^{\frac{1}{\gamma-1}}$ $\sim t$ k_{max}

RANDOM   RANDOM   SCALE-FREE   HUB-AND-SPOKE TOPOLOGY   NETWORK
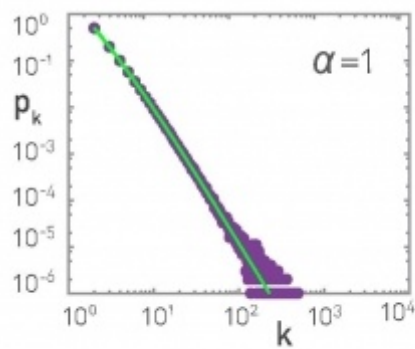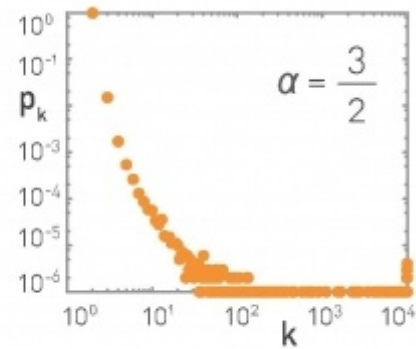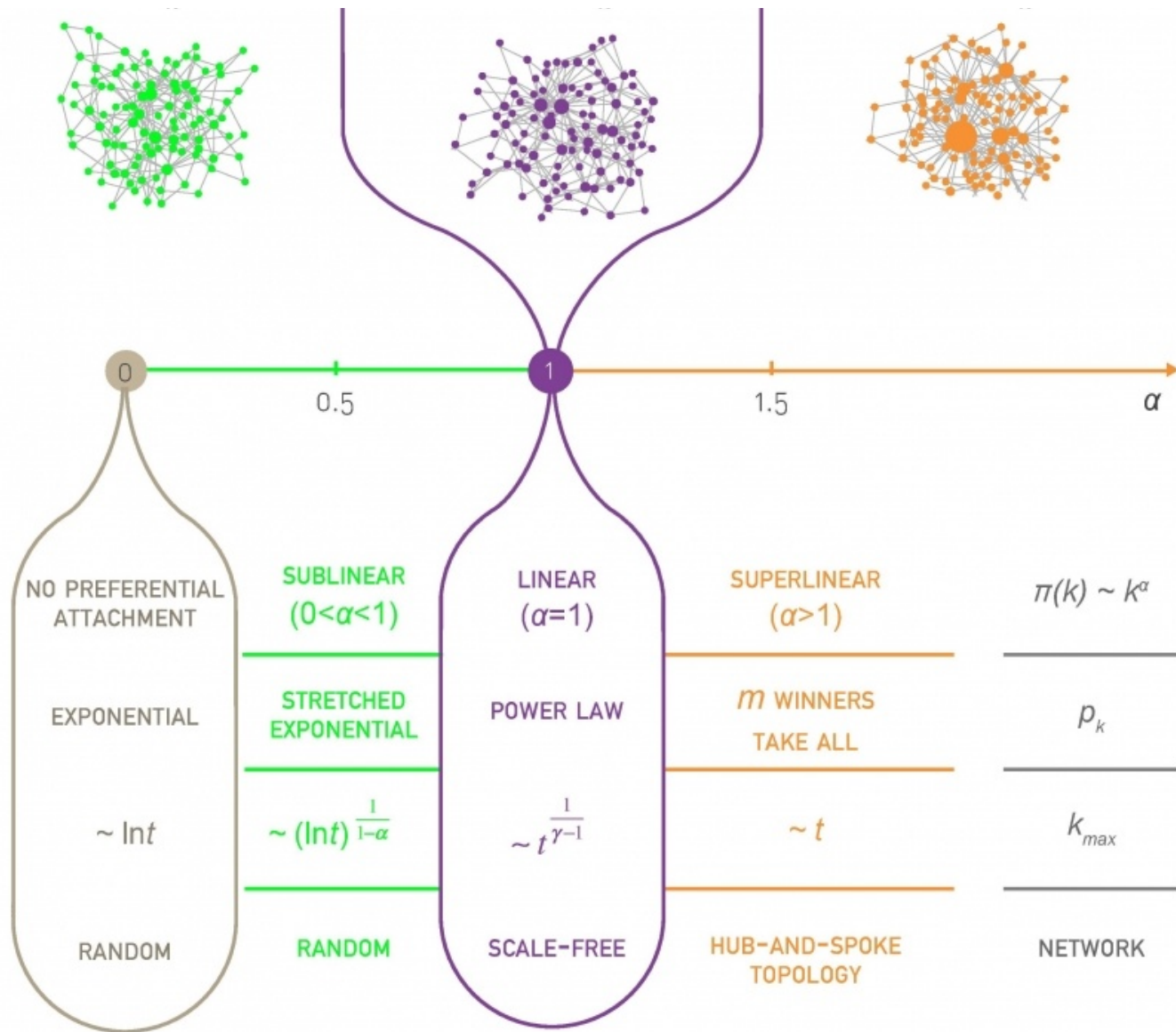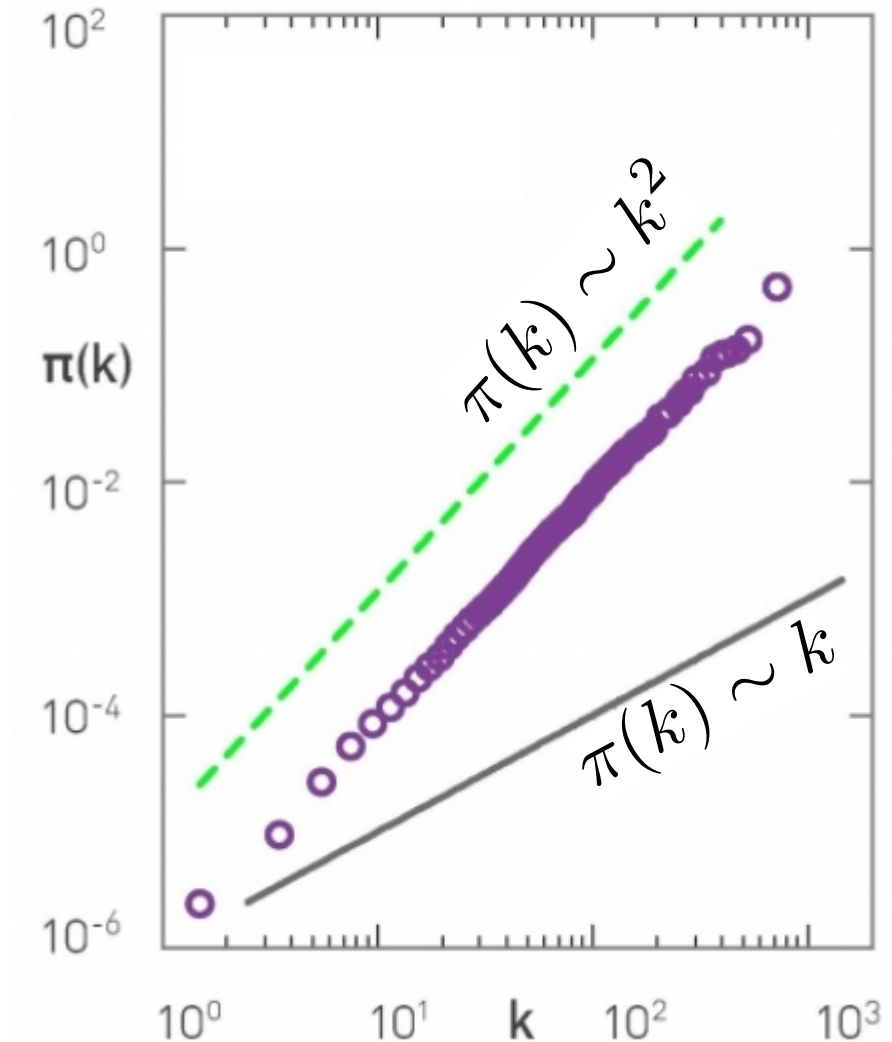
Measuring preferential attachment

Measuring 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 it follows preferential attachment (with $\alpha = 1$) in this case

Aging effects

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

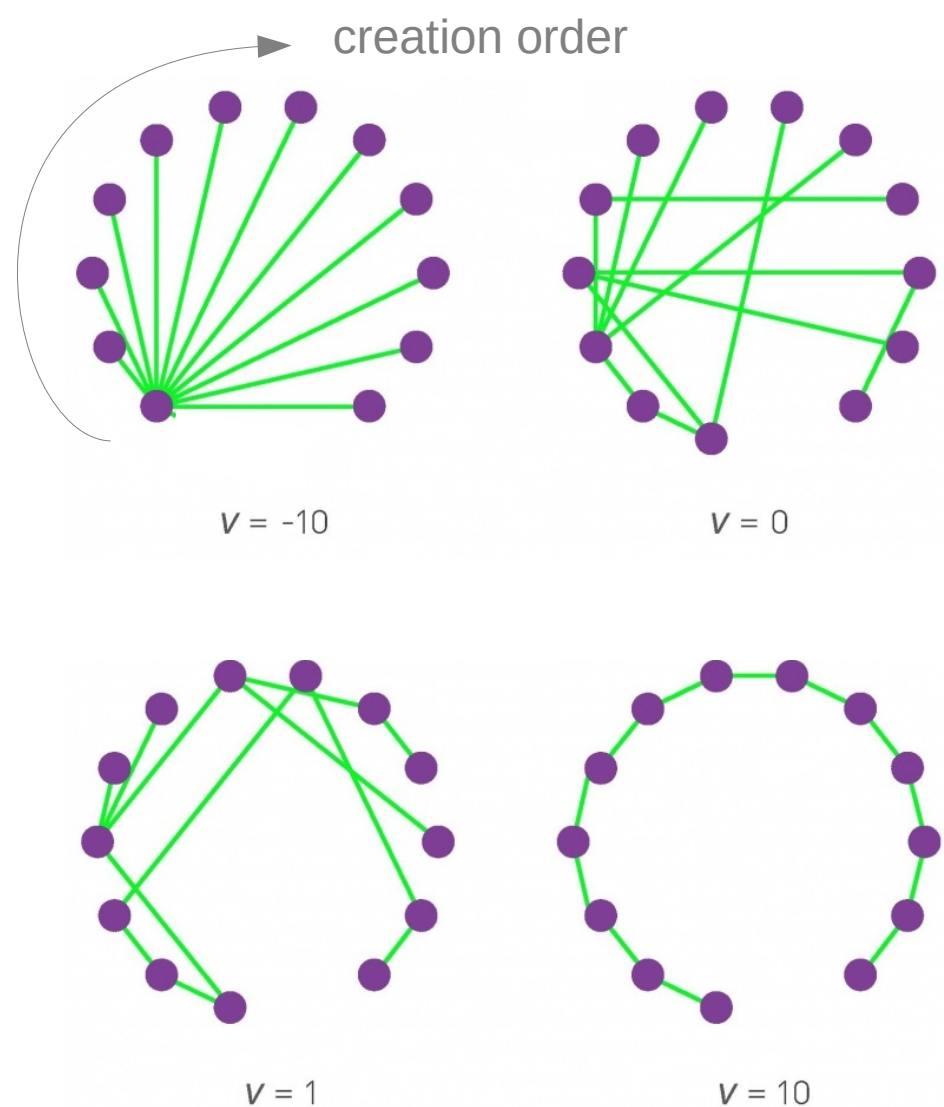- Models without fitness but with a negative effect of age

$$\Pi(k_i, t - t_i) \approx k_i(t - t_i)^{-v}$$

- Older nodes accumulate links more slowly

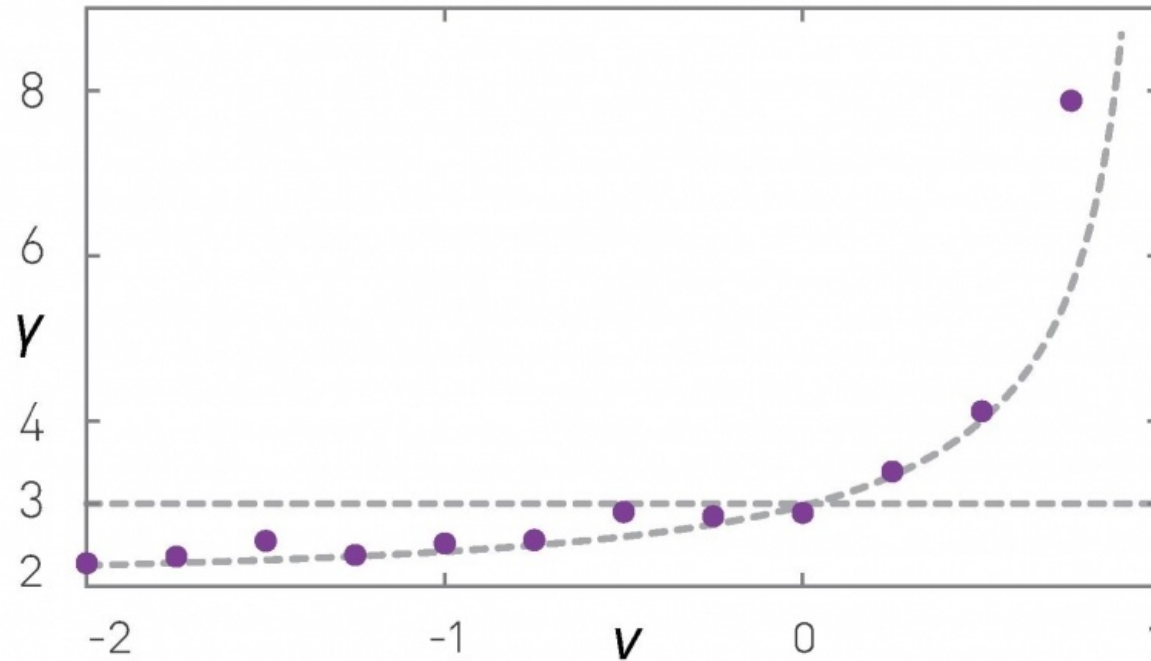- Parameter v is the decay factor

Qualitatively, what would you expect if:
$$v < 0 \qquad v = 0 \qquad v \approx 1 \qquad v \gg 1$$

Aging effects

- $v < 0$ favors older nodes

- $v = 0$ is simply preferential attachment

- $v \gg 1$ means only youngest are linked

creation order

$v = -10$

$v = 0$

$v = 1$

$v = 10$

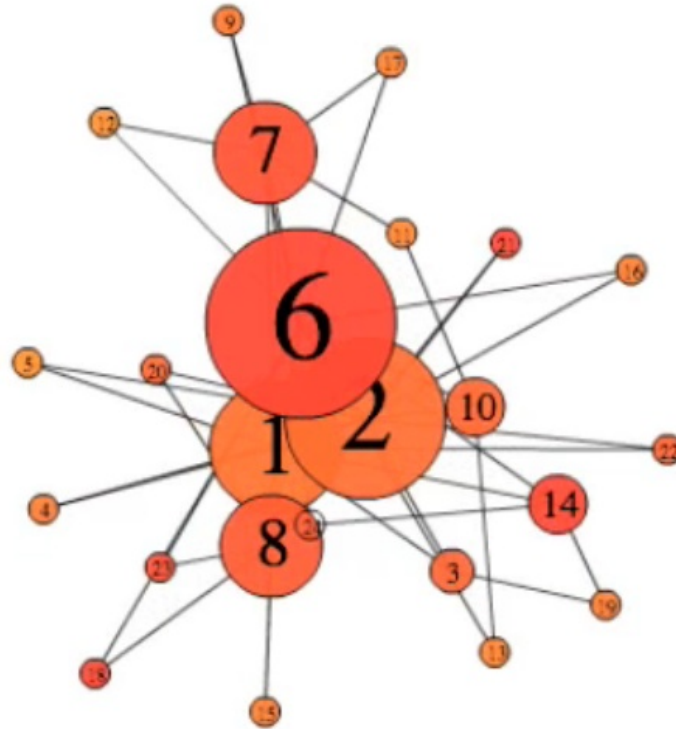# Power-law exponent in models with aging (N=10K, m=1)

More heterogeneous ⟵⟶ More homogeneous

"Good get richer"
(incl. Bianconi-Barabási model)

"Good get richer" simulation (number 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

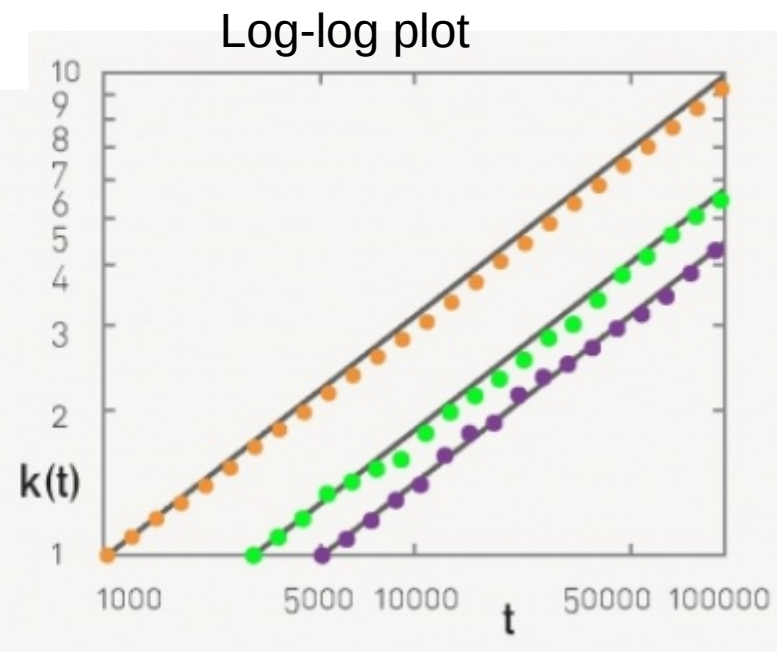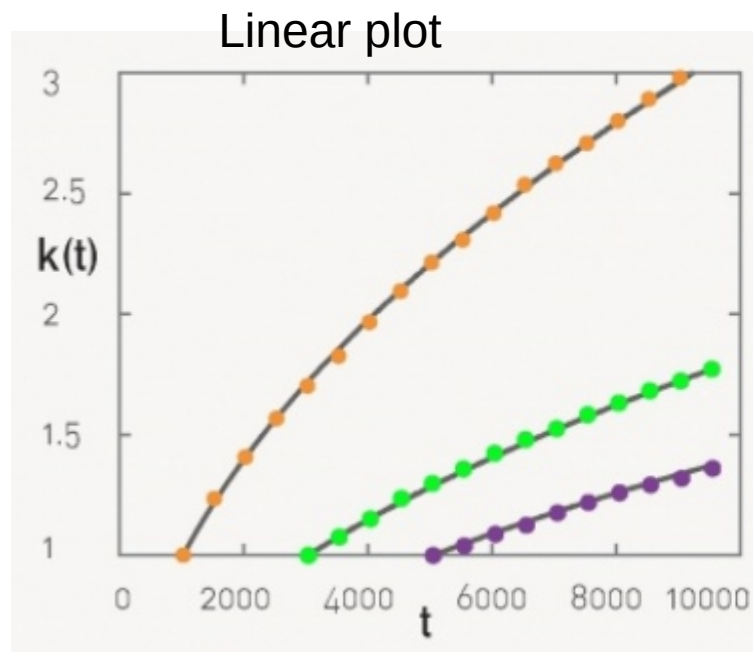- The probability of $\Pi_i = \dfrac{\eta_i k_i}{\sum_j \eta_j k_j}$ o node i is:

Degree dynamics

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

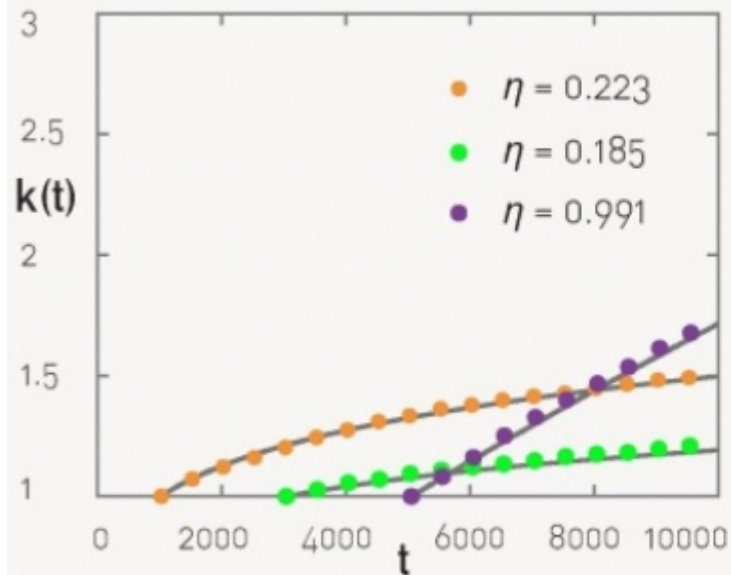$$k_i(t; t_i, \eta_i) = m\left(\frac{t}{t_i}\right)^{\beta(\eta_i)}$$

- With the dynamic exponent $\quad \beta(\eta_i) \propto \eta_i$

- Remember that in linear preferential attachment $\beta = 1/2$ (for all nodes)

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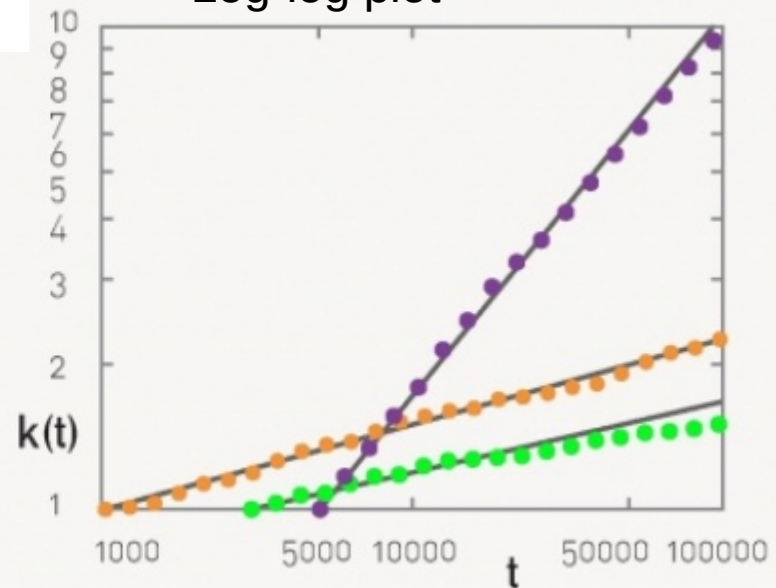

Linear plot   Log-log plot

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

Linear plot

Log-log plot

$\eta = 0.223$
$\eta = 0.185$
$\eta = 0.991$

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?

Link selection model / copy model

Other processes that generate scale-free networks

- Link-selection model — step:

    - Add one new node v to the network

    - Select an existing link (u, w) at random and connect v to either u or w

- Copy model — step:

    - Add one new node v to the network

    - Pick a random existing node u

    - With probability p link to u

    - With probability $1\text{-}p$ link to a neighbor of u

Exercise: the copy model

In the copy model, start at $t=1$ with one node, and at every step t:

- Add one new node v to the network

- Pick a random existing node u

- If u has no out-links, link to u

- If u has out-links choose one of the following:

  - With probability p link to u

  - With probability $1-p$ link to one of the out-neighbors of u chosen at random

- Simulate it on paper (directed graph) for 7 nodes with $p=0.5$

  - Make sure you understand the model fully!

- What is $N(t)$ and $L(t)$? What is k_i^{out}?

Degree distribution in the copy model

Proven in the paper by Kumar et al. (FOCS 2000) "Stochastic models for the web graph" and developed in the advanced materials.

The copy model can generate any exponent between 2 and 3!

$$\gamma = \frac{2-p}{1-p} \in [2,3] \ \ \text{if } p \in [0, 1/2]$$

In the copy model, at every step t:
1) Add one new node v to the network
2) Pick a random existing node u
3) With probability p link to u
4) With probability 1-p link to a neighbor of u

- We will compute k_i^{in} but first …

- How many links on average gets node i at time t? In other words, what is:

$$\frac{d}{dt} k_i^{\text{in}}(t)$$

- Hint: it has a term with p and a term with $1\text{-}p$

- Integrate between t_i and t to obtain an expression for $k_i(t_i)$

  (*we drop the "in" superscript just for simplicity during this exercise*)

- Note that now $k_i(t_i) = 0$

- Once you have a expression for $k_i(t_i)$

- Compute $Pr(k_i(t_i) > k)$

- Now write the cumulative distribution function of $k_i(t_i)$

- And compute its derivative to obtain
$$p_k = Pr(k_i(t) = k) = \frac{d}{dk}Pr(k_i(t) \leq k)$$

- It should show exponent $\gamma = \dfrac{2-p}{1-p}$

Summary

Things to remember

- Uniform attachment

- Sub-linear and super-linear preferential attachment

- Measuring preferential attachment

- "Good-get-richer" and aging effects

- The copy model

Practice on your own

- Practice creating graphs using the different models

  - By hand

  - Or write your own code (it's not a lot of code)

Advanced materials: expected degree under uniform random attachment

EXTRA

Expected degree in uniform random attachment using a differential equation

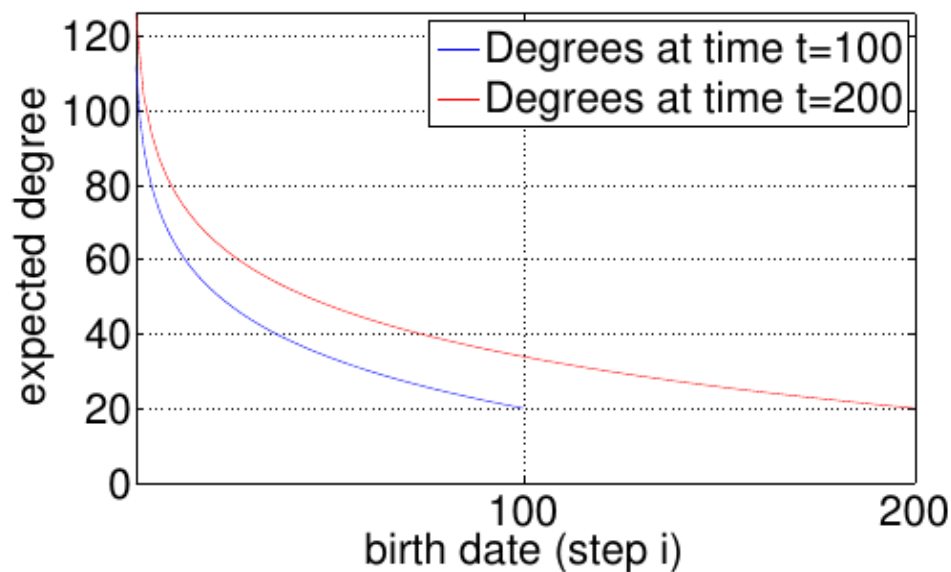$$\frac{d}{dt}k_i(t) = \frac{m}{t}$$

Obtain k_i

(1) Integrate between time i and time t

(2) Use initial condition $k_i(i) = m$

$$\int \frac{1}{t} = \log t + C$$

Degree distribution over time is not static

Degree of node born at time $m < i < t = m \left(1 + \log \left(\dfrac{t}{i} \right) \right)$

Tail of degree distribution

How many nodes of degree larger than K are there at time t?

The fraction is $\dfrac{te^{-\frac{K-m}{m}}}{t} = e^{-\frac{K-m}{m}}$

Decreases exponentially with K: it's vanishingly rare to find high-degree nodes

$$m\left(1 + \log\left(\frac{t}{i}\right)\right) > K$$

$$1 + \log\left(\frac{t}{i}\right) > \frac{K}{m}$$

$$\log\left(\frac{t}{i}\right) > \frac{K-m}{m}$$

$$\frac{t}{i} > e^{\frac{K-m}{m}}$$

$$i < te^{-\frac{K-m}{m}}$$

Advanced materials:

(1) No preference (2) No growth

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
    - Connect u to v

$$\Pi(k_v) = \frac{k_v}{\sum_j k_j}$$

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 $\dfrac{d}{dt}k_i = m\Pi(k_i)$ obtain $k_i(t)$

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

No preference model (cont.)
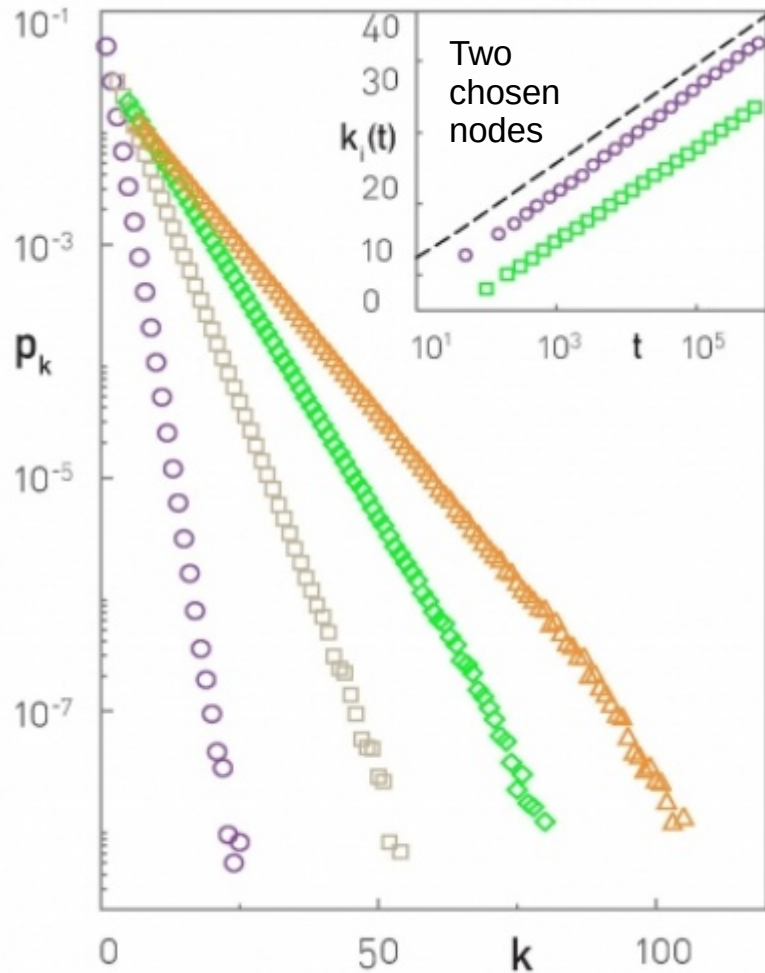
- Compute $Pr(k_i(t) > k)$ assuming large t, t_i

- Use it to compute
$$Pr(k_i(t) \leq k) = 1 - Pr(k_i(t) > k)$$

$$p_k = Pr(k_i(t) = k)$$

- Derive to obtain

Two chosen nodes

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

Consequences of the "no preference" model
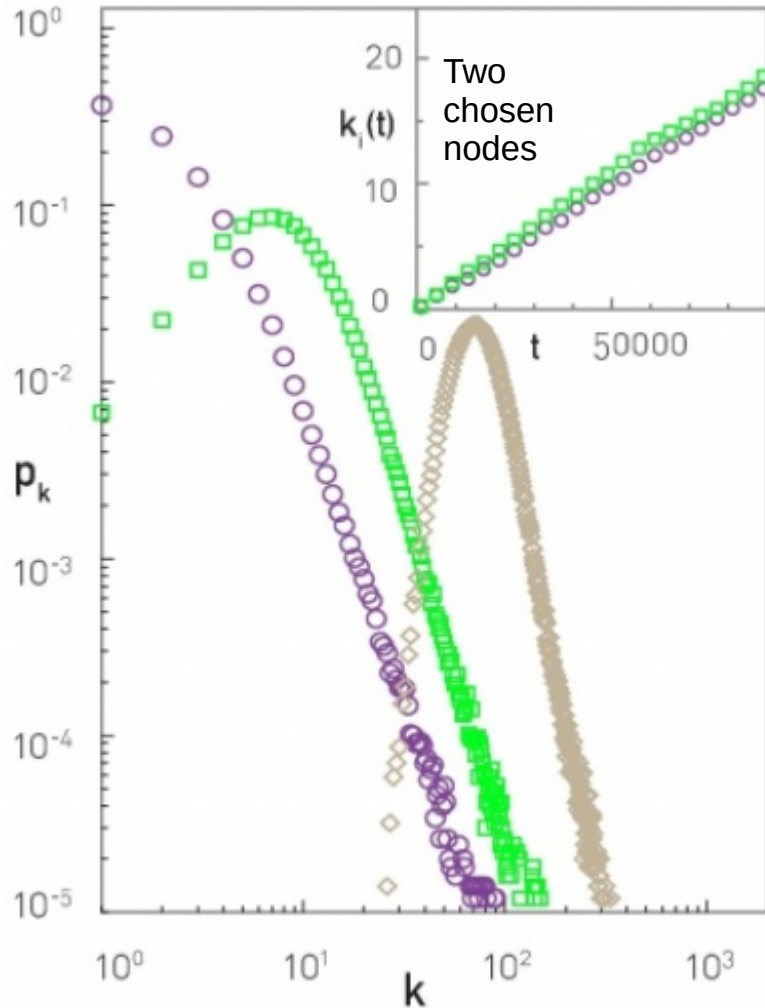
- Degree decays exponentially
  $$p_k \propto e^{-k/m}$$
- No power-law
- No large hubs

No growth model

- Write the process on paper

- You will need to impose $\quad k_i(t_i) \neq 0 \quad$ why?

- Write $\Pi(k_i)$

- Noting that $\quad \dfrac{d}{dt}k_i = \Pi(k_i) \qquad$ obtain $k_i(t)$

Two chosen nodes

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