Degree Distributions in Preferential Attachment

Introduction to Network Science
Carlos Castillo
Topic 12

Universitat Pompeu Fabra Barcelona

Contents

- Degree distribution under the BA model
- Distance distribution under the BA model
- Clustering coefficient under the BA model

BA model means Barabási-Albert model (preferential attachment)

Sources

- Albert László Barabási (2016) Network Science
  - Preferential attachment follows chapter 05
- Ravi Srinivasan 2013 Complex Networks Ch 12
- Networks, Crowds, and Markets Ch 18
- Data-Driven Social Analytics course by Vicenç Gómez and Andreas Kaltenbrunner

Remember the BA model

- Network starts with m_0 nodes connected arbitrarily as long as their degree is ≥ 1
- At every time step we add 1 node
- This node will have m outlinks $(m \leq m_0)$
- The probability of an existing node of degree k_i to gain one such link is $$\Pi(k_i) = \frac{k_i}{\sum_{j=1}^{N-1} k_j}$$

Degree $k_i(t)$ as a function of time

$$\frac{d}{dt}k_i = m\Pi(k_i) = m\frac{k_i}{\sum_{j=1}^{N-1} k_j}$$

$$\sum_{j=1}^{N-1} k_j = L(0) + 2m(t-1) \approx 2m(t-1)$$

(For large t)

$$\frac{d}{dt}k_i = \frac{mk_i}{2m(t-1)} = \frac{k_i}{2t-2} \approx \frac{k_i}{2t}$$

Degree $k_i(t)$ ... continued

$$\frac{d}{dt}k_i(t) = \frac{k_i(t)}{2t}$$

$$\frac{1}{k_i(t)}\frac{d}{dt}k_i(t) = \frac{1}{2t}$$

$$\int_{t=t_i}^{t}\frac{1}{k_i(t)}\frac{d}{dt}k_i(t)dt = \int_{t=t_i}^{t}\frac{1}{2t}dt$$

$$\log k_i(t) - \log k_i(t_i) = \frac{1}{2}\log t - \frac{1}{2}\log t_i$$

$$\log k_i(t) = \frac{1}{2}\log t - \frac{1}{2}\log t_i + \log m$$

Note: in exams for this course, you will **not** be asked to solve differential equations on your own

(t_i is the creation time of node i)

Degree $k_i(t)$ … continued

$$\log k_i(t) = \frac{1}{2}\log t - \frac{1}{2}\log t_i + \log m$$

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

Is the degree growth linear, super-linear, or sub-linear? Intuitively, why?

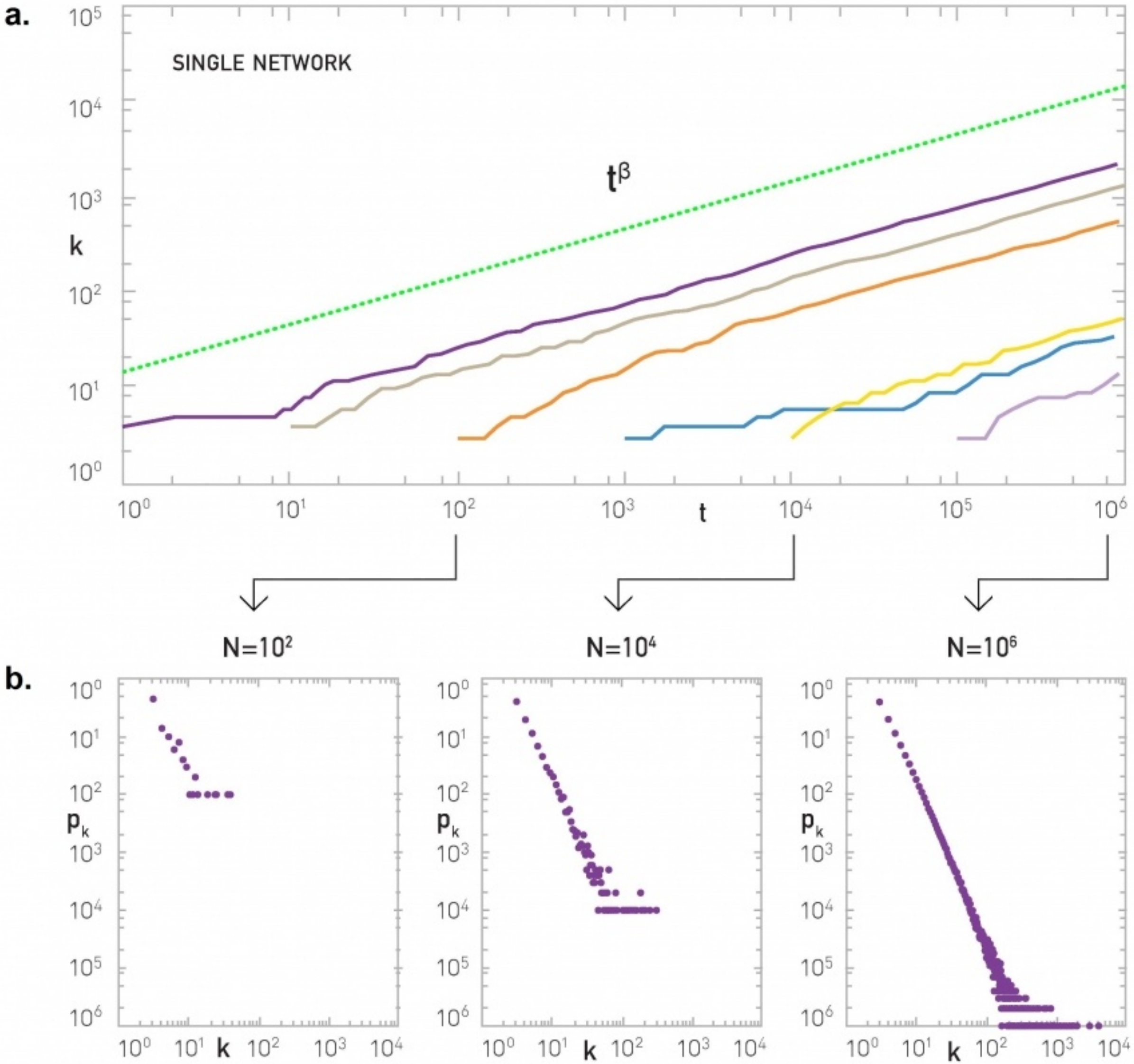
Degree $k_i(t)$ … consequences

$$\log k_i(t) = \frac{1}{2}\log t - \frac{1}{2}\log t_i + \log m$$

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

$$\frac{dk_i(t)}{dt} = \frac{k_i(t)}{2t} = \frac{m\left(\frac{t}{t_i}\right)^{\frac{1}{2}}}{2t} = \frac{m}{2\left(t \cdot t_i\right)^{\frac{1}{2}}}$$

If $t_i < t_j$ (node i is older than node j), what do we expect of k_i and k_j?



Simulation results

Model

Nodes with t_i = 1, 10, 100, 1000, 10000, ...

Degree distribution

- Let's calculate the CDF of the degree distribution

$$
\begin{aligned}
Pr(k_i \leq k) &= 1 - Pr(k_i > k) \\
&= 1 - Pr\left(m \left(\frac{t}{t_i} \right)^{\beta} > k \right) \\
&= 1 - Pr\left(\left(\frac{m}{k} \right)^{1/\beta} > \frac{t_i}{t} \right) \\
&= 1 - \left(\frac{m}{k} \right)^{1/\beta}
\end{aligned}
$$

$$\frac{ti}{t} \sim \mathrm{Uniform}(0,1)$$

Degree distribution

Now let's take the derivative of the CDF to obtain the PDF

$$
\begin{aligned}
p_k &= \frac{d}{dk} Pr(k_i \leq k) = \frac{d}{dk}\left(1 - \left(\frac{m}{k}\right)^{1/\beta}\right) \\
&= -\frac{d}{dk}\left(\left(\frac{m}{k}\right)^{1/\beta}\right) = -m^{1/\beta}\frac{d}{dk}\left(\frac{1}{k^{1/\beta}}\right) \\
&= \frac{1}{\beta}\frac{m^{1/\beta}}{k^{1/\beta+1}} \qquad (\beta = 1/2) \\
&= 2\frac{m^2}{k^3} \longrightarrow p(k) \propto k^{-3}
\end{aligned}
$$

Degree distribution

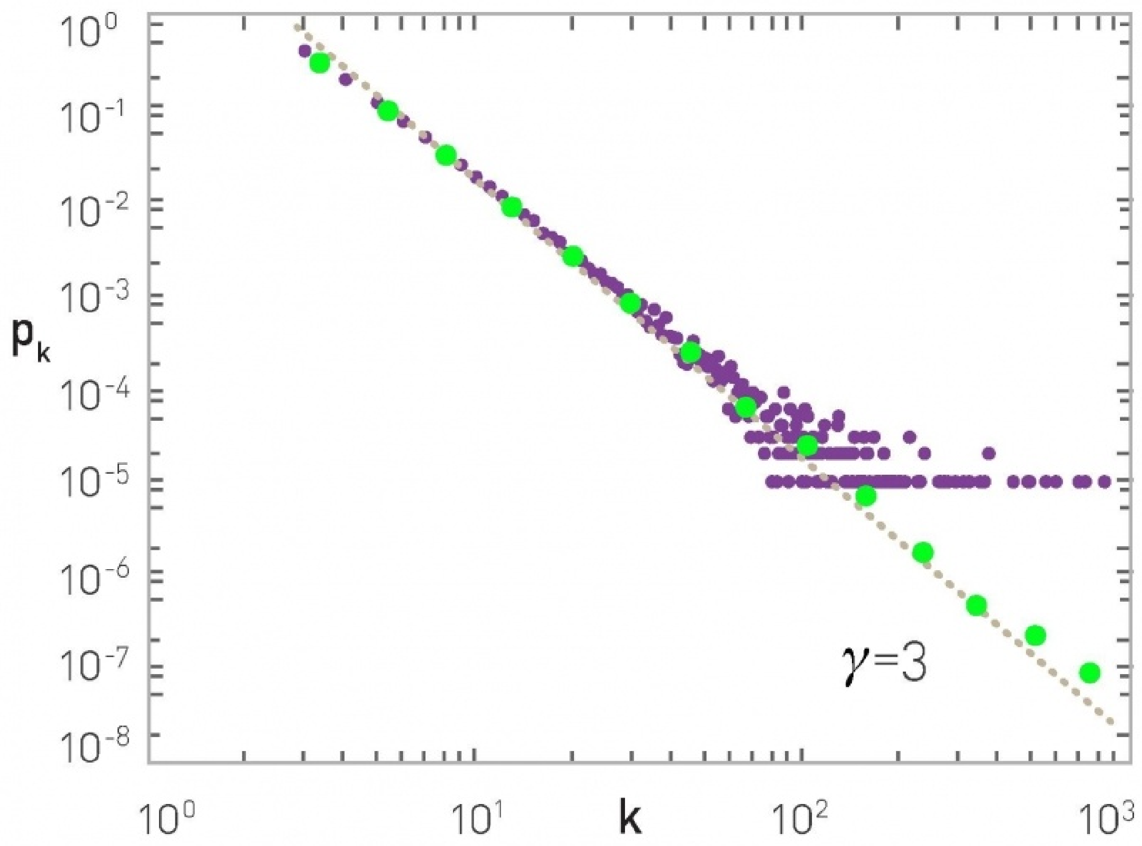
- $\beta = 1/2$ is called the dynamical exponent
- $\gamma = \dfrac{1}{\beta} + 1 = 3$ is the power-law exponent

- Note that $p(k) \approx 2m^2/k^3$

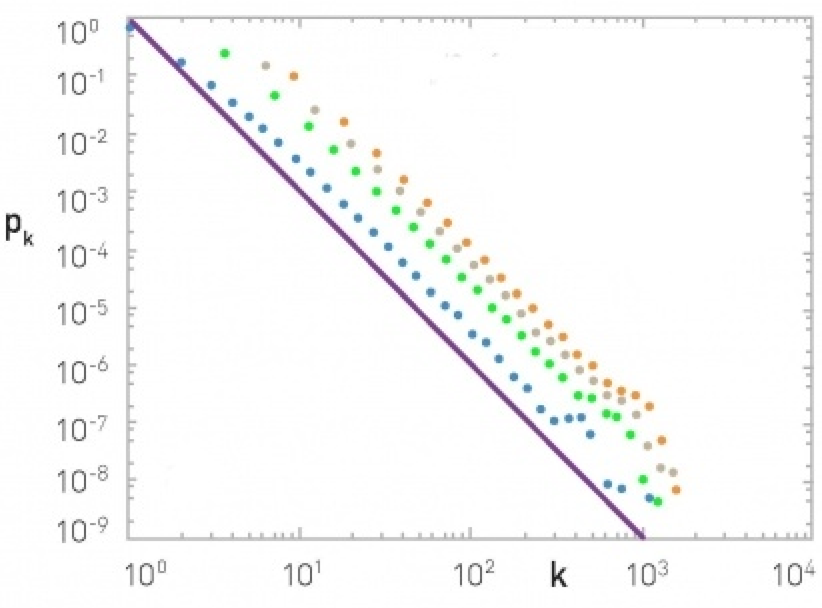
  does not depend on t

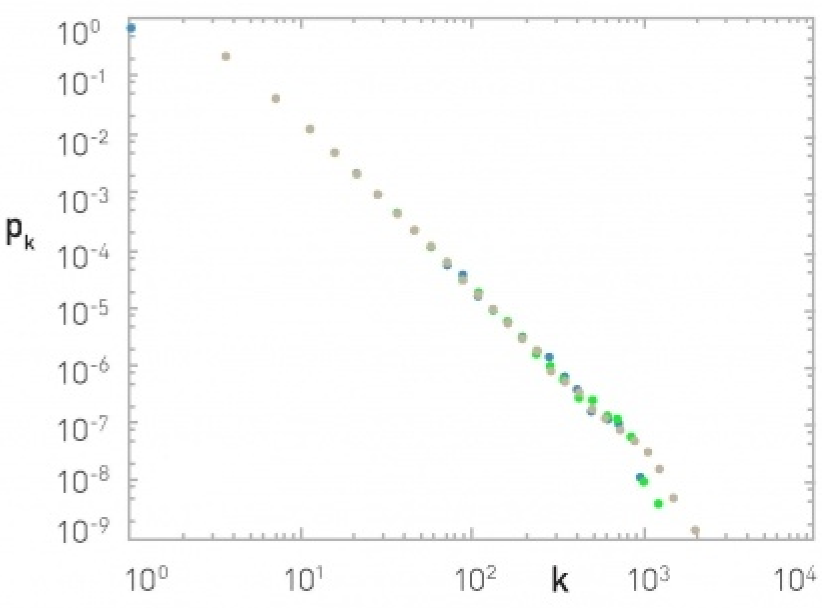
  hence, it describes a stationary network

Degree distribution, simulation results
N=100,000   m=3



More simulations





N = 100,000; m_0 = m =
1 (blue), 3 (green), 5 (gray), 7 (orange)

Observe γ is independent of m (and m_0)

The slope of the purple line is -3

m_0 = m = 3; N =
50K (blue), 100K (green), 200K (gray)

Observe p_k is independent of N

Processes that generate scale-free networks

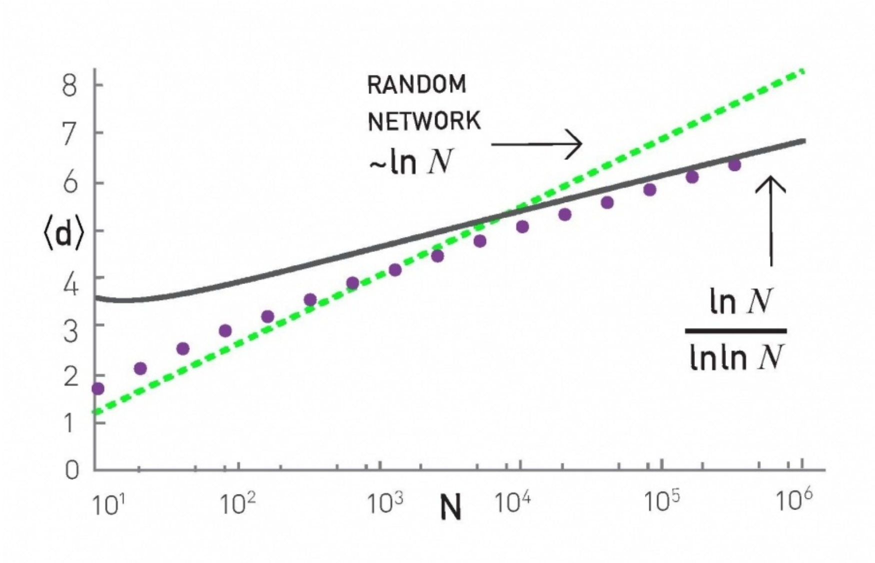
- Link-selection model — step:
  - Add one new node v to the network
  - Select an existing link at random and connect v to one of the edges of that existing link
- Copy model — step:
  - Add one new node v to the network
  - Pick a random existing node u
  - With probability p link to u
  - With probability 1-p link to a neighbor of u

Average distance

- Distances grow slower than log N

$$\langle d \rangle \approx \frac{\log N}{\log \log N}$$

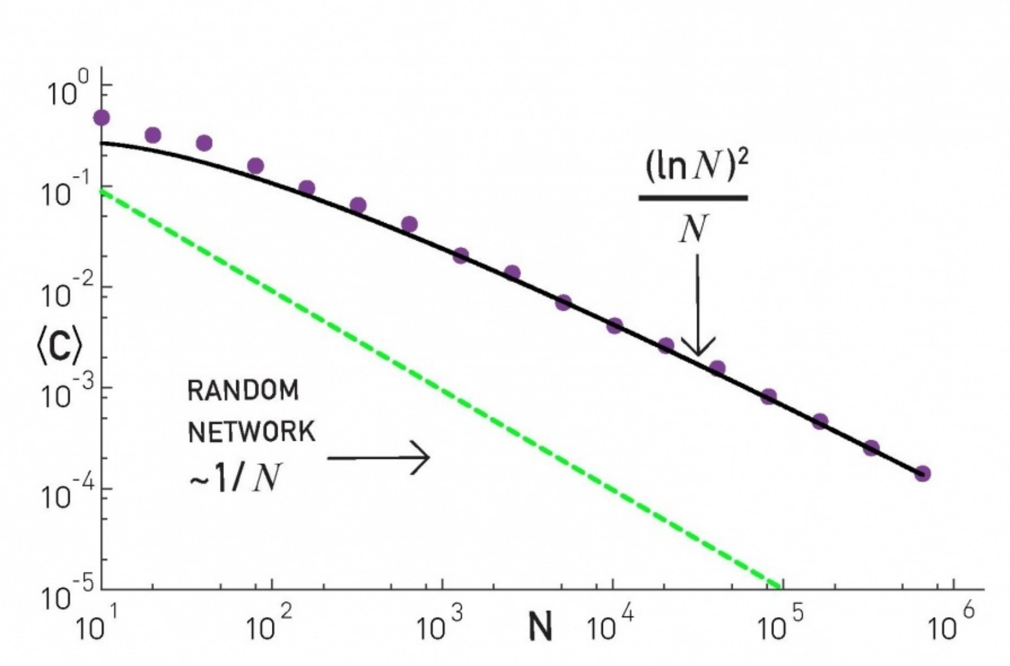
(Scale free network with $\gamma = 3$)



Clustering coefficient

- BA networks are locally more clustered than ER networks

$$\langle C \rangle \approx \frac{(\log N)^2}{N}$$



Limitations of the BA model

- Predicts a fixed exponent of -3
- Assumes an undirected network, while many real complex networks are directed
- Does not consider node deletions or edge deletions which are common in practice
- Considers that all nodes are equal except for their arrival times

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 $1\text{-}p$ 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} ?

Answer in Nearpod Draw-it
https://nearpod.com/student/
Access to be provided during class

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

Answer in Nearpod Draw-it
https://nearpod.com/student/
Access to be provided during class

- We will compute k_i^{in} but before ...
- 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-p*

Summary

Things to remember

- Degree distribution in the BA model
- Distances and clustering coefficient in BA
- The copy model

Practice on your own

- Try to reconstruct the derivations we have done in class, including the exercise
  - Try to understand every step
- Insert a small change in the model and try to recalculate what we have done

Advanced materials:
Copy model cont.
(not included in the exam)

- 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 $\gamma = \frac{2-p}{1-p}$