Degree Under the Preferential Attachment (BA) Model

Social Networks Analysis and Graph Algorithms

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

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

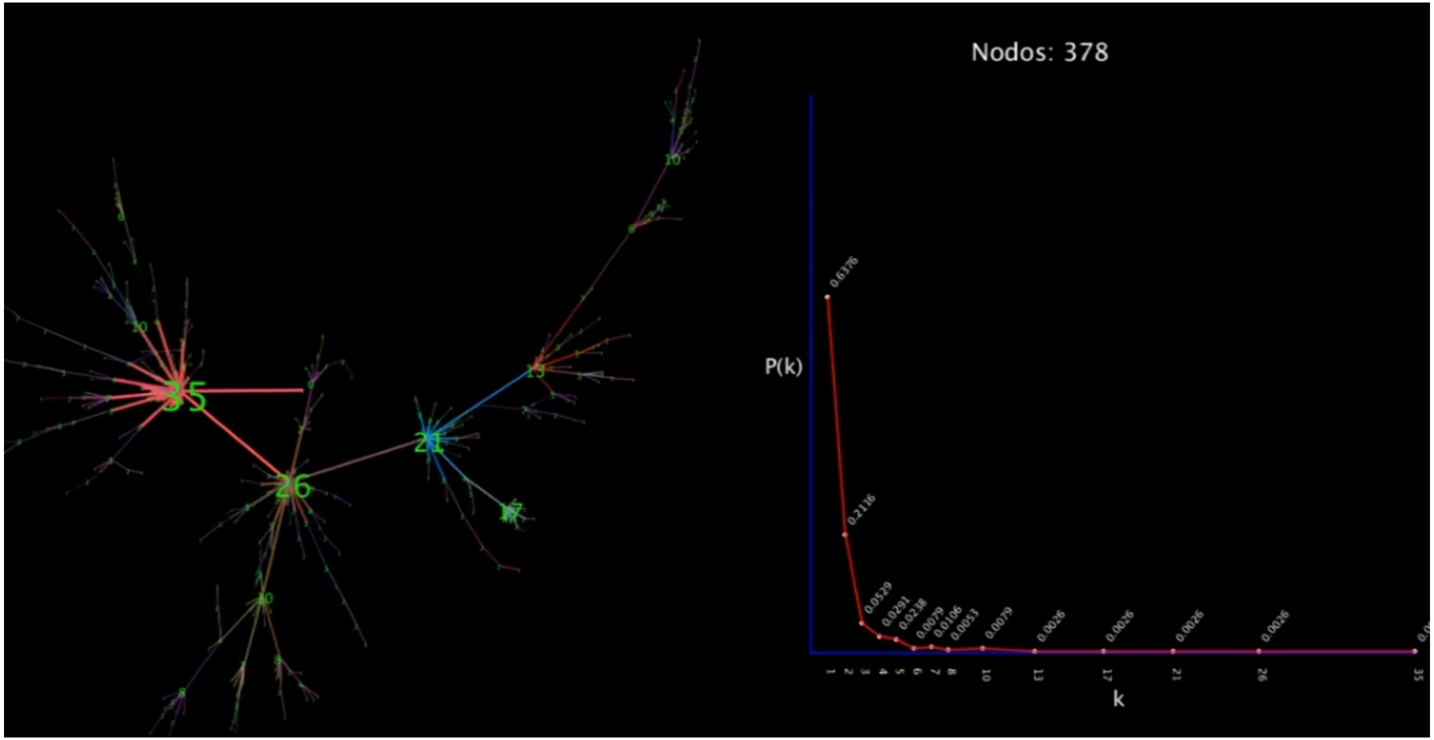
- A. L. Barabási (2016). Network Science – Chapter 05
- R. Srinivasan (2013). Complex Networks – Chapter 12
- D. Easley and J. Kleinberg (2010): Networks, Crowds, and Markets – Chapter 18
- Data-Driven Social Analytics course by Vicenç Gómez and Andreas Kaltenbrunner

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

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

Video of degree distribution



https://www.youtube.com/watch?v=5RIQweqPT6A

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

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

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

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

(t_i is the creation time of node i)

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

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

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

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?

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

$\beta = 1/2$ is called the dynamical exponent

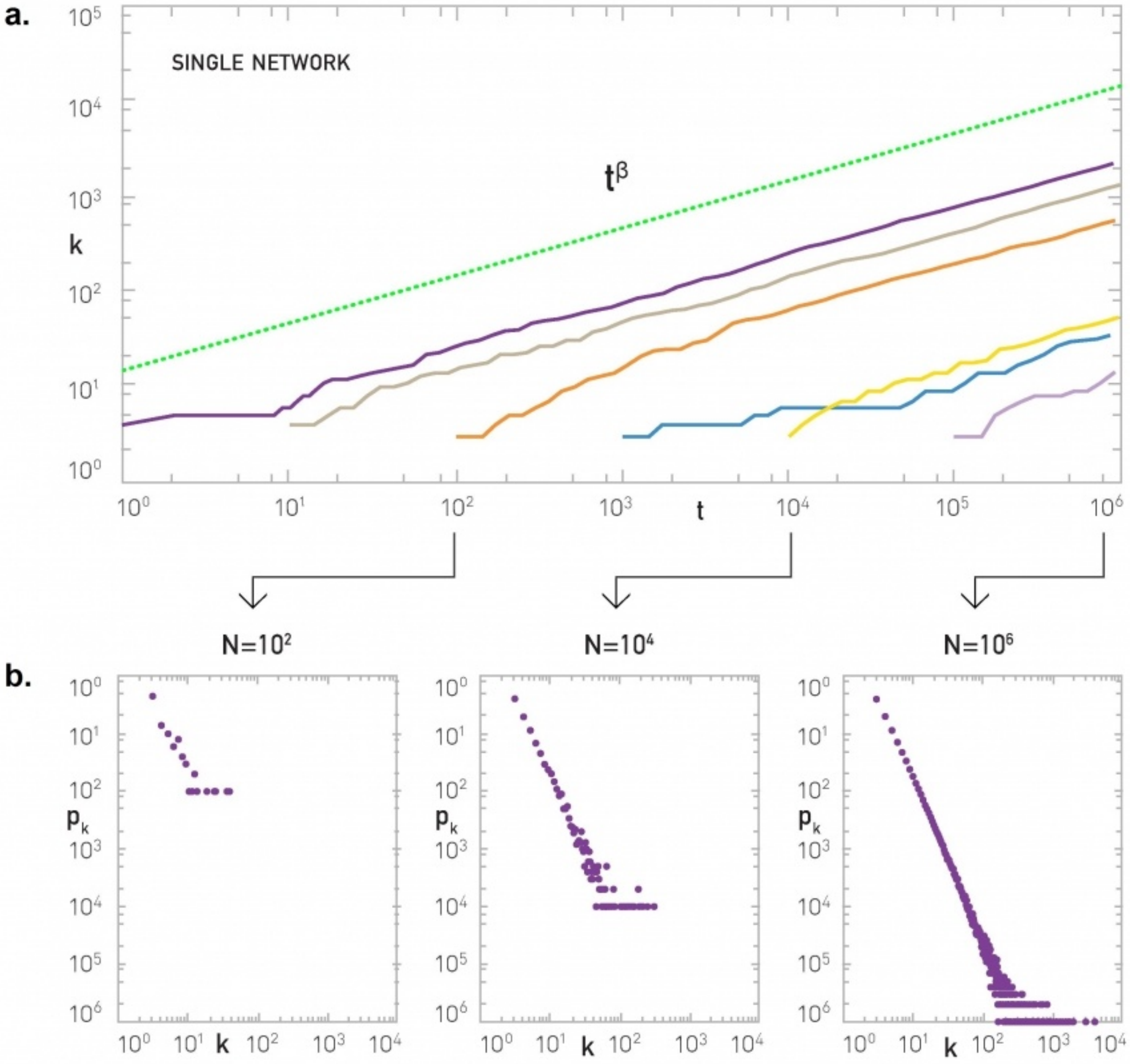
Degree $k_i(t)$ ... consequences

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

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

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

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

- The distribution of the degree follows

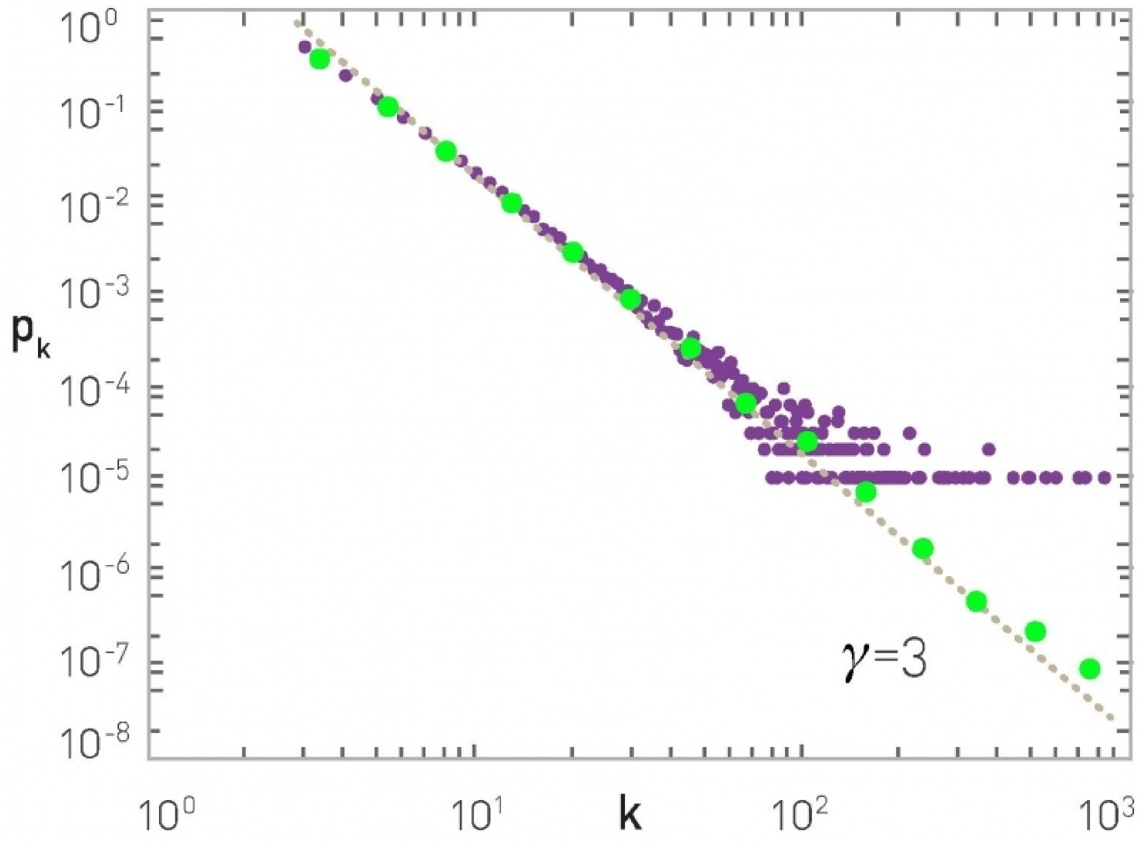
$$p(k) \approx 2m^2/k^3$$

  (Proof at the end of this deck)

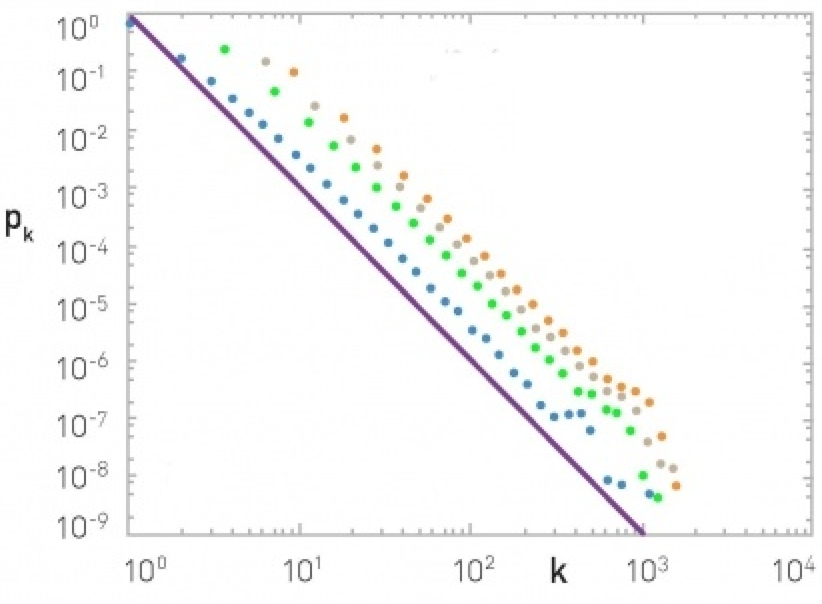
- Note that it does not depend on the time, hence, it describes a stationary network

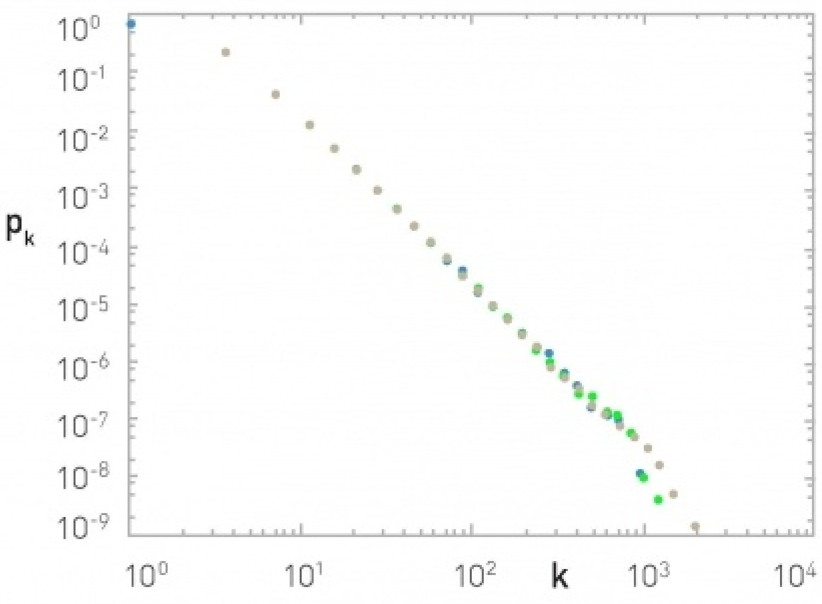
Degree distribution, simulation results

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

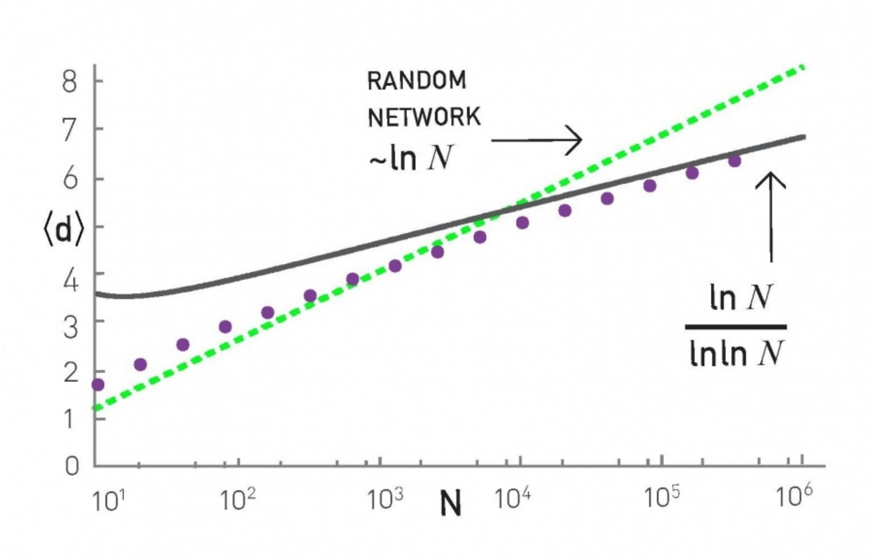
m_0 = m = 3; N =
50K (blue), 100K (green), 200K (gray)

Observe p_k is independent of N

Average distance

- Distances grow slower than log N

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

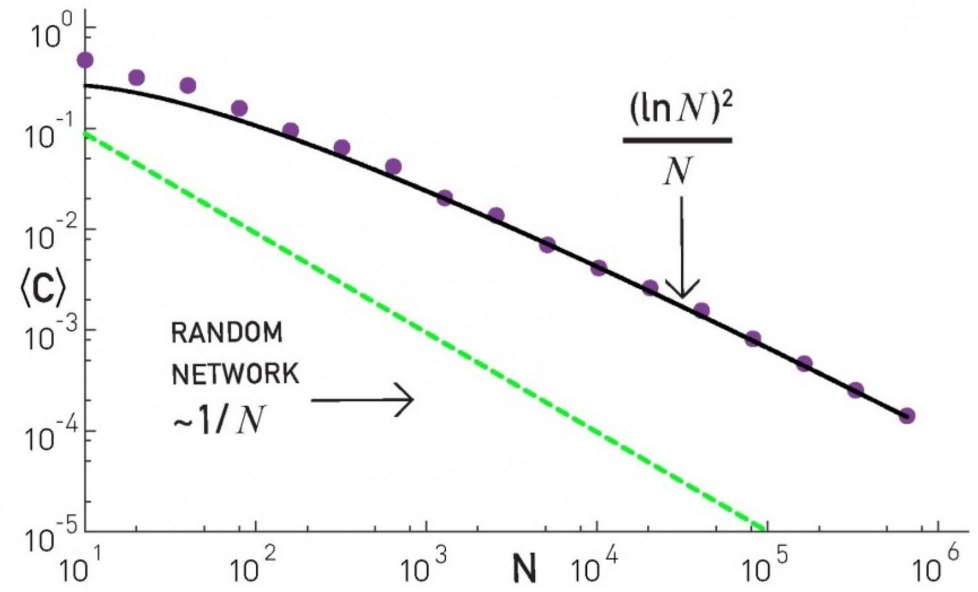


(Why: scale free network with $\gamma = 3$)

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

Summary

Things to remember

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

Practice on your own

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

Additional contents
(not included in exams)

Cumulative Distribution Function

Let's calculate the CDF of the degree distribution

By definition of CDF, this is equal to:

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

CDF (cont.)

Let's calculate $Pr(k_i(t) > k)$

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

$$k_i(t) > k \Rightarrow m\left(\frac{t}{t_i}\right)^{\beta} > k$$

$$m^{\frac{1}{\beta}}\left(\frac{t}{t_i}\right) > k^{\frac{1}{\beta}}$$

$$\left(\frac{m}{k}\right)^{\frac{1}{\beta}}\left(\frac{t}{t_i}\right) > 1$$

$$\left(\frac{m}{k}\right)^{\frac{1}{\beta}} > \left(\frac{t_i}{t}\right)$$

$$t_i < t\left(\frac{m}{k}\right)^{\frac{1}{\beta}}$$

This means that nodes i with degree larger than k were created at time t_i **before** a certain timestep, which is expected because older nodes have larger degree.

CDF (cont.)

From the previous slide, we have: $Pr(k_i(t) > k) = Pr\left(\left(\frac{m}{k}\right)^{\frac{1}{\beta}} > \frac{t_i}{t}\right)$

Remember there is one node created at each timestep, so by time t there are $N(t) = m_o + t$ nodes, and for large t, we have $N(t) \approx t$

Now, what is $Pr(x > t_i/t)$ if you pick a node i at random?

It is x, because $\boldsymbol{t_i/t}$ **is distributed uniformly in** $\boldsymbol{[0,1]}$

Hence:

$$Pr(k_i(t) > k) = \left(\frac{m}{k}\right)^{\frac{1}{\beta}}$$

Imagine the following "game", in which the larger number wins

- You pick a number x in $[0,1]$
- Your opponent picks a number y uniformly at random in $[0,1]$

The probability that $x > y$ and hence you win is exactly x

CDF (cont.)

Hence:

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

Probability Density Function (PDF)

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

$$p_k = \frac{d}{dk} Pr(k_i \leq k) = \frac{d}{dk}\left(1 - \left(\frac{m}{k}\right)^{1/\beta}\right)$$

$$= -\frac{d}{dk}\left(\left(\frac{m}{k}\right)^{1/\beta}\right) = -m^{1/\beta}\frac{d}{dk}\left(\frac{1}{k^{1/\beta}}\right)$$

$$= \frac{1}{\beta}\frac{m^{1/\beta}}{k^{1/\beta+1}} \qquad (\beta = 1/2)$$

$$= 2\frac{m^2}{k^3} \longrightarrow p(k) \propto k^{-3}$$

Degree distribution

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