Preferential Attachment (BA Model)

Introduction to Network Science
Carlos Castillo
Topic 11

Universitat Pompeu Fabra Barcelona

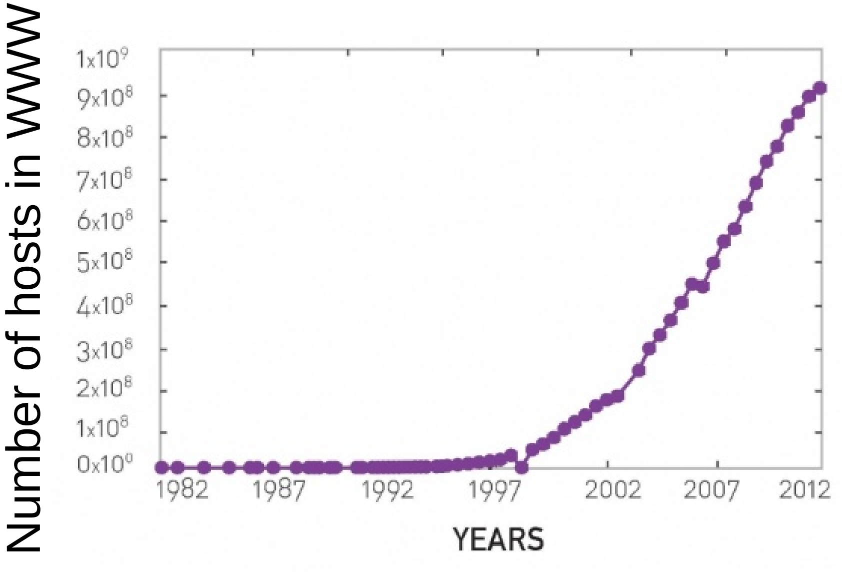
Contents

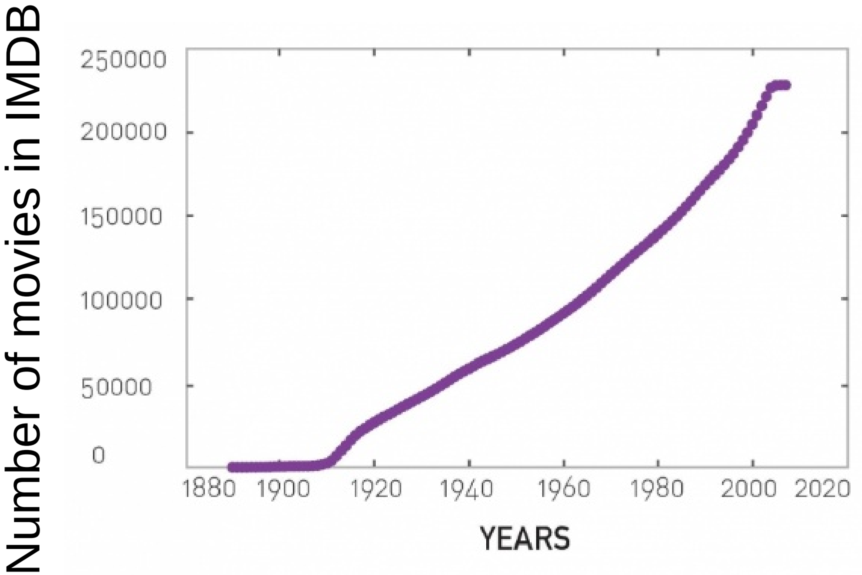
- The uniform random attachment model
- The BA or preferential attachment model
- Degree distribution under the BA model
- Distance distribution under the BA model
- Clustering coefficient under the BA model

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

The number of nodes N increases: we need models of network growth





Preliminary:
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)

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

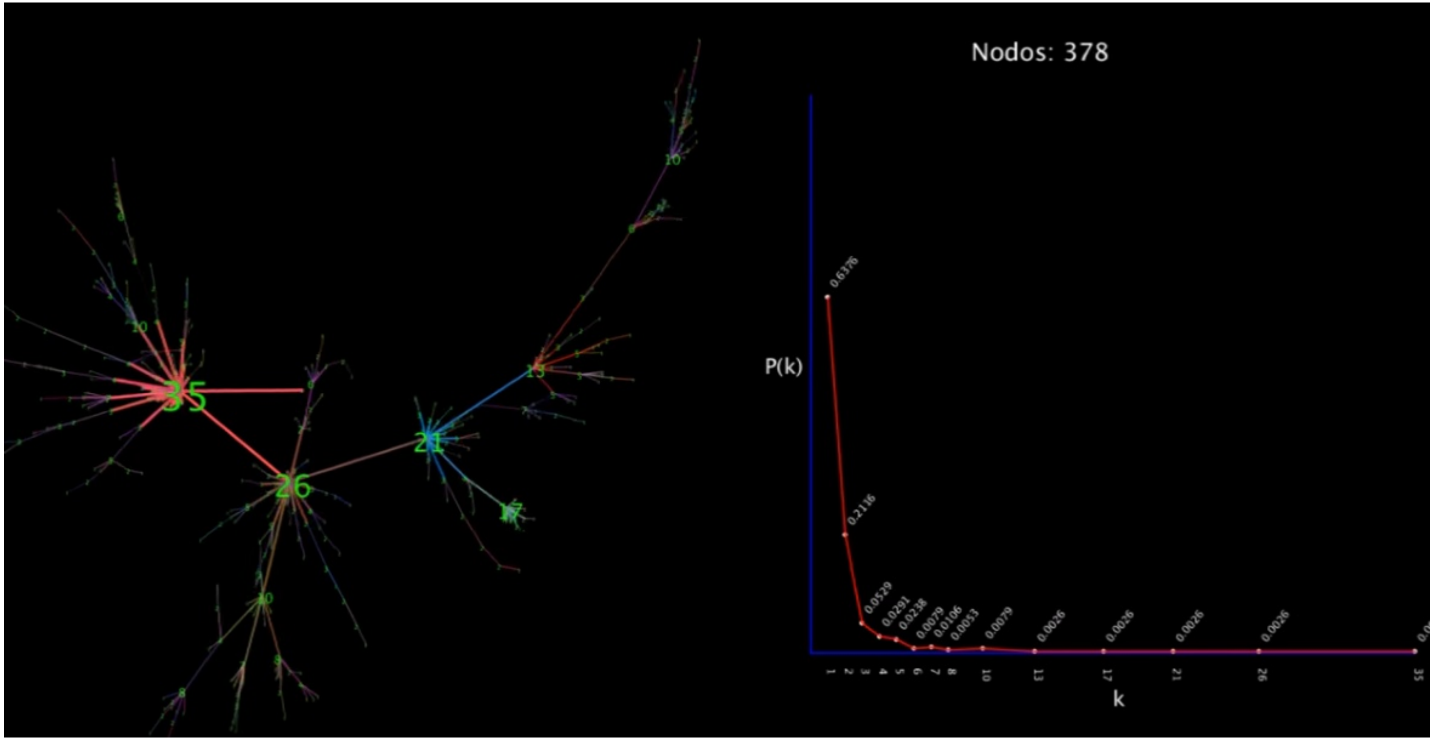
- Decreases exponentially with *K*: it’s vanishingly rare to find high-degree nodes

Preferential Attachment

Preferential attachment simulation

https://www.youtube.com/watch?v=4GDqJVtPEGg

Degree distribution in simulation



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

We have seen what but not why

- Power-law degree distributions are prevalent
  - Why?
- Two assumptions in ER networks:
  - There are N nodes that **pre-exist**
  - Nodes connect **at random**
- Let’s challenge both assumptions

Growth

- Suppose there are two web pages on a topic, one with many inlinks the other with few, which one am I most likely to link to?
- Which scientific papers are read?
- Which book authors sell more?
- Which actors are more sought after?

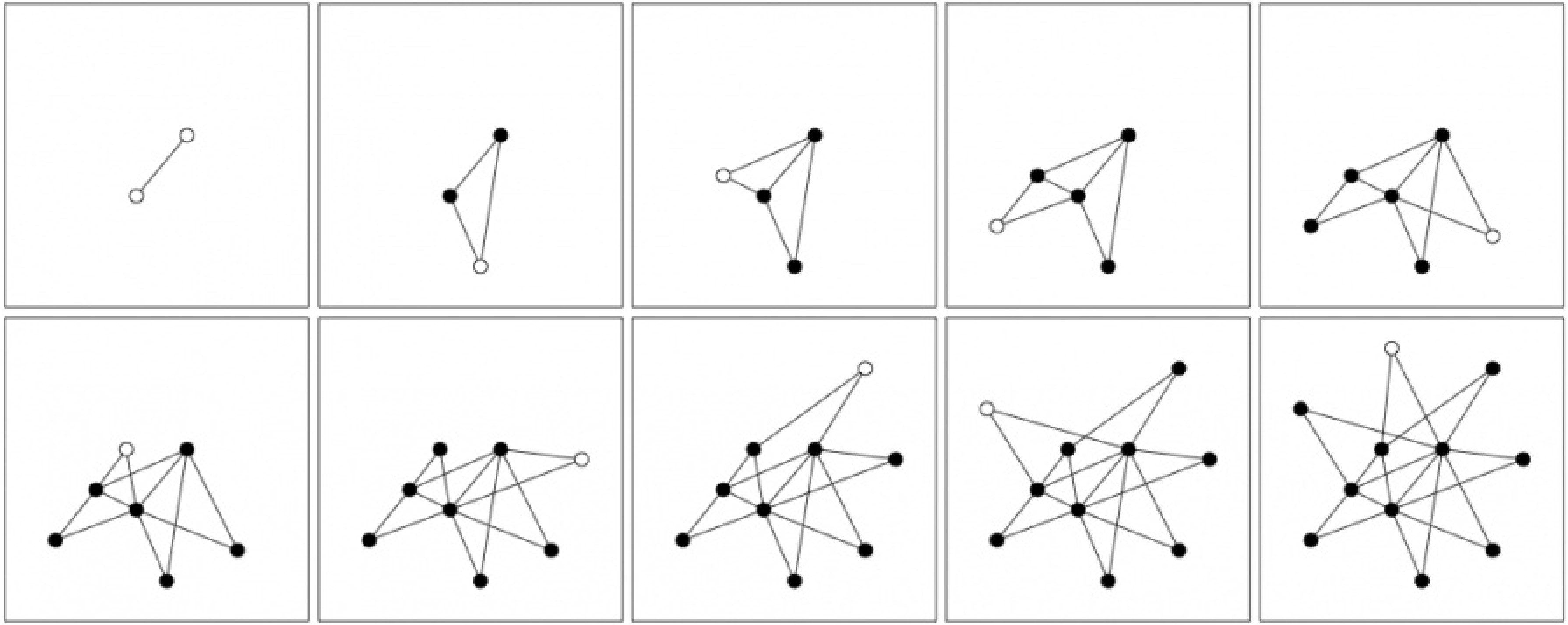

BURNS
UNIVERSITY
Ut
Uberiores
Divites
Our motto: *Ut uberiores divites.*
FOX 32
CHICAGO

The Barabási-Albert (BA) model

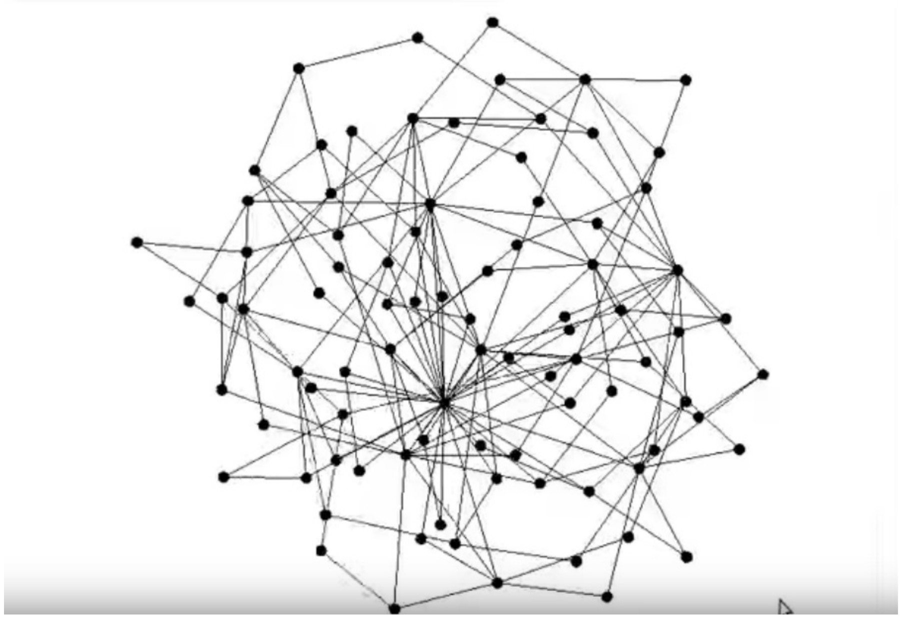
- Network starts with m_0 nodes connected arbitrarily as long as their degree is ≥ 1
- At every time step we add 1 node
- This node will have $m \leq m_0$ outlinks
- The probability of an existing node of degree k_i to gain one such link is $$\Pi(k_i) = \frac{k_i}{\sum_{j=1}^{N-1} k_j}$$

In an ER network, $\Pi(k_i) = \frac{1}{N-1}$

Example ($m_0 = 2; m=2$)



Network growth with $m=2$



https://www.youtube.com/watch?v=wocaGeNKn7Y

The Barabási-Albert (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 $\Pi(k_i) = \dfrac{k_i}{\sum_{j=1}^{N-1} k_j}$

Write the formula for $N(t)$ and $L(t)$: at t=0 the network has m_0 nodes and $L(0)$ links

Answer in Nearpod Collaborate: https://nearpod.com/student/ – Access to be provided during class

Summary

Things to remember

- Preferential attachment
- How to create a BA network step by step

Practice on your own

- Describe step by step in pseudocode how to create a Barabási-Albert graph with N nodes having m_0 starting nodes and m outlinks per node.
- For your pseudocode to be valid, if at any point there is a randomized step, you must indicate what is the probability of each possible outcome.

Advanced materials: Expected degree under uniform random attachment (not included in the exam)

Expected degree in uniform random attachment using a differential equation

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

Obtain $\boldsymbol{k_i}$

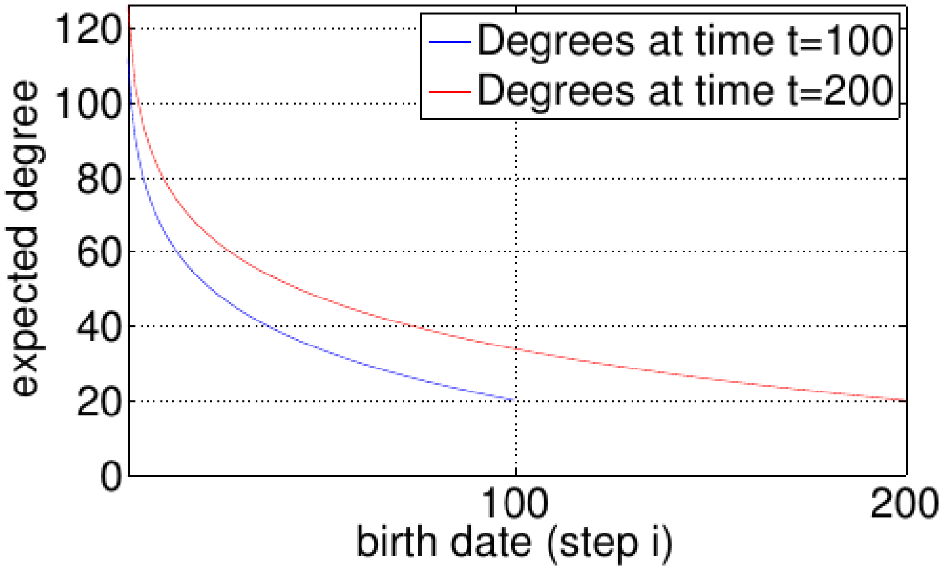
(1) Integrate between time i and time t

(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 $m < i < t = m\left(1 + \log\left(\frac{t}{i}\right)\right)$



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

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

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

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

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

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