Advanced Algorithms and Computational Models (module A)

Evolving Networks

Giacomo Fiumara
giacomo.fiumara@unime.it

2020-2021

- Google was founded six yers after birth of WWW. By the end of 1990 Altavista, Yahoo and Inktomi, three search engines with an early start, dominated the search market

- Yet Google not only became the leading search engine, but acquired links at such an incredible rate that by 2000 became the biggest hub of the Web as well

- Even more impressive, in 2011 Facebook took over as the biggest node of the Web

- These examples highlight that none of the network models studied so fare are able to account for it
- In the Erdös-Rényi model, the identity of the biggest node is driven entirely by chance
- The Barabási-Albert model offers a more realist picture, predicting that each node increases its degree following $k(t) \sim t^{1/2}$

  This means that the oldest node always has the most links, a phenomenon called the *first mover's advantage* in the business literature
- In reality the growth rate of a node does not depend on its age only.

- Some nodes have some intrinsic property that propels them ahead of the pack: *fitness*

  Some companies turn each consumer into a loyal partner

  Some webpages turn visitors into addicts

  Some people turn each random encounter into a lasting social link

- In the Barabási-Albert model we assumed that the rate of growth of the node is determined solely by its degree
- To incorporate the role of fitness we assume that preferential attachment is driven by the product of two factors: fitness (η) and degree (k)

The Bianconi-Barabási (BB) model or the *fitness model*, consists of the following steps:

- **Growth** – In each timestep a new node j with m links and fitness η_j is added to the network, where η_j is a random number chosen from a *fitness distribution* $\rho(\eta)$. Once assigned, the fitness of a node does not change
- **Preferential Attachment** – The probability that a link of a new node connects to node i is proportional to the product of its degree and its fitness

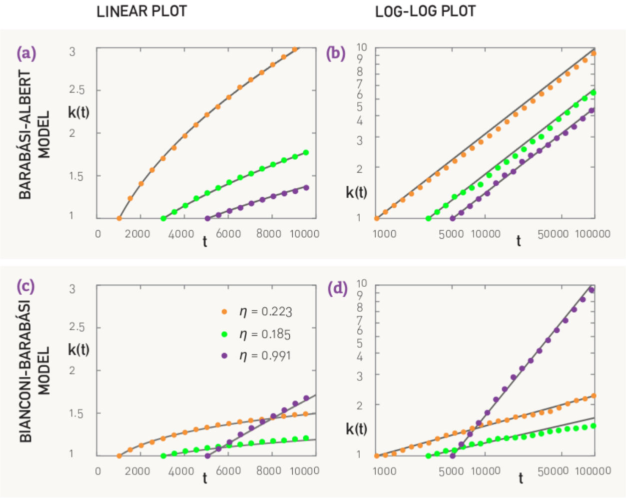
$$\Pi_i = \frac{\eta_i k_i}{\sum_j \eta_j k_j}$$

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

- The dependence of Π_i on k_i captures the fact that higher-degree nodes have more visibility, hence we are more likely to link to them
- The dependence of Π_i on η_i implies that between two nodes with the same degree, the one with higher fitness is selected with a higher probability
- Therefore a relatively young node, with initially only a few links, can acquire links rapidly if it has larger fitness than the rest of the nodes

Evolving Networks

The Bianconi-Barabasi Model



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

We can ue the continuum theory to predict each node's temporal evolution. The degree of node i changes at the rate

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

Let us assume that the time evolution of k_i follows a power law with a fitness-dependent exponent $\beta(\eta_i)$

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

It can be demonstrated that the *dynamic exponent* β is

$$\beta(\eta) = \frac{\eta}{C}$$

with

$$C = \int \rho(\eta) \frac{\eta}{1 - \beta(\eta)} d\eta$$

- In the BA model we have $\beta = 1/2$, hence the degree of each node increases as a square root of time
- In the BB model the dynamic exponent is proportional to the fitness of the node η, hence each node has its own dynamic exponent
- Therefore, a node with higher fitness will increase its degree faster. Given sufficient time, the fitter node will leave behind nodes with a smaller fitness
- This is the case of Facebook, a latecomer with an addictive product, which acquired links faster than its competitors

- The degree distribution of the network generated by the BB model can be calculated using the continuum theory, obtaining

$$p_k = C \int d\eta \frac{\rho(\eta)}{\eta} \left(\frac{m}{k} \right)^{\frac{C}{\eta}+1}$$

  which is a weighted sum of multiple power-laws.

- p_k depends on the precise form of the fitness distribution $\rho(\eta)$

To illustrate the properties of the BB model equations

$$C = \int \rho(\eta) \frac{\eta}{1-\beta(\eta)} d\eta$$

and

$$p_k = C \int d\eta \frac{\rho(\eta)}{\eta} \left(\frac{m}{k}\right)^{\frac{C}{\eta}+1}$$

are used to calculate $\beta(\eta)$ and p_k for the following fitness distributions:

- **Equal fitness**
- **Uniform fitness distribution**

Equal fitness

- When all fitnesses are equal, the BB model reduces to the BA model
- Let us use $\rho(\eta) = \delta(\eta - 1)$, which means that each node has the same fitness $\eta = 1$
- In this case from the equation

$$C = \int \rho(\eta) \frac{\eta}{1 - \beta(\eta)} d\eta$$

one obtains

$$C = 2$$

From the equation

$$\beta(\eta) = \frac{\eta}{C}$$

follows $\beta = 1/2$

Equal fitness (cont.d)

– Using equation

$$p_k = C \int d\eta \frac{\rho(\eta)}{\eta} \left(\frac{m}{k}\right)^{\frac{C}{\eta}+1}$$

one obtains

$$p_k = 2 \int d\eta \frac{\delta(\eta - 1)}{\eta} \left(\frac{m}{k}\right)^{\frac{C}{\eta}+1} = 2\left(\frac{m}{k}\right)^{2+1} = 2\left(\frac{m}{k}\right)^{3}$$

namely

$$p_k \approx k^{-3}$$

the knwon scaling of the degree distribution in the BA model

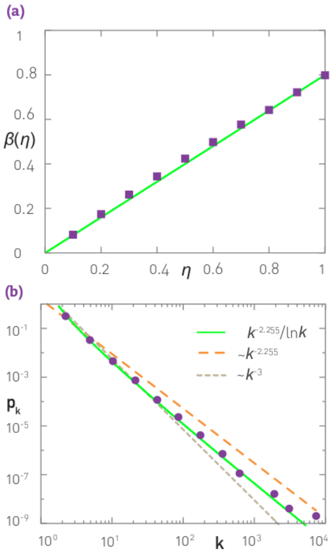
Uniform fitness distribution

- The behavior of the model is more interesting when nodes have different fitnesses
- Let us choose η to be uniformly distributed in the $[0, 1]$ interval
- In this case C can be obtained using a numerical approach which yelds $C^{\star} = 1.255$
- The degree distribution is therefore

$$p_k \approx \int_0^1 d\eta \frac{C^{\star}}{\eta} \frac{1}{k^{1+C^{\star}/\eta}} \approx \frac{k^{-(1+C^{\star})}}{lnk}$$

predicting that the degree distribution follows a power law with degree exponent $\gamma = 2.255$

The Bianconi-Barabasi Model: Degree Distribution



- In summary, the BB model can account for the fact that nodes with different internal characteristics acquire links at different rates
- A node's growth rate is determined by its fitness η and allows us to calculate the dependence of the degree distribution on the fitness distribution $\rho(\eta)$

- Measuring the fitness of a node could help us identify web sites that are poised to grow in visibility, research papers that will become influential, or actor on their way to stardom
- Fitness is the collective perception of a node's importance relative to the other nodes
- Therefore, it is possible to determine a node's fitness by comparing its time evolution to the time evolution of other nodes in the network
- If we have dynamical information about the evolution of the individual nodes, the BB model allows us to determine the fitness of each node

- To relate the growth rate of a node to its fitness we start from equation

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

and take the logarithm. We obtain

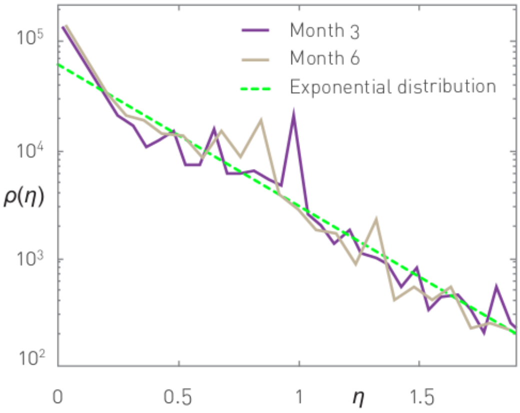
$$lnk(t, t_i, \eta_i) = \beta(\eta_i) lnt + ln\frac{m}{t_i^{\beta(\eta_i)}}$$

- Hence the slope of $lnk(t, t_i, \eta_i)$ is a linear function of the dynamical exponent $\beta(\eta_i)$. In turn we have $beta(\eta) = \eta/C$.
- Therefore, the distribution of $\beta(\eta_i)$ is identical with the fitness distribution $\rho(\eta)$

The Fitness of a Web document

- Node fitnesses were measured in the context of the WWW (dataset: 22 million documents for 13 months)
- While most nodes did not change their degree during this time frame, 6.5% showed sufficient change to determine their dynamical exponent
- The obtained fitness distribution $\rho(\eta)$ has an exponential form, indicating that high fitness nodes are rare

The Fitness of a Web document



The Fitness of a Web document

- The shape of the obtained fitness distribution is somewhat unexpected, as one would be tempted to assume that on the web fitness varies widely

- The exponential form of $\rho(\eta)$ indicates that the fitness of Wen documents is bounded, varying in a relatively narrow range
- As a consequence, the observed large differences in the degree of two web documents is generated by the system's dynamics: growth and preferential attachment amplify the small fitness differences, turning nodes with slightly higher fitness into much bigger nodes

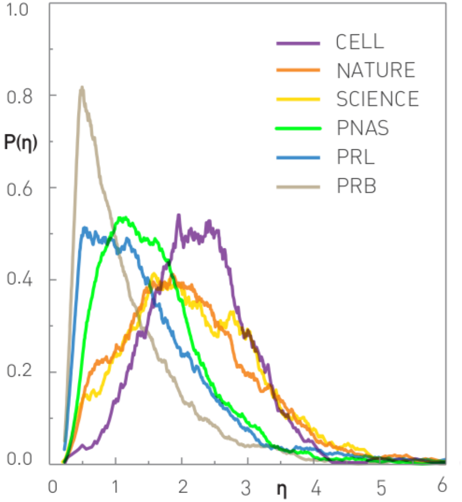
The Fitness of a Scientific Publication

- While most research papers acquire only a few citations, a small number of publications collect thousands citations
- The probability that a research paper i is cited at time t after publication is

$$\Pi_i \sim \eta_i c_i^t P_i(t)$$

- the paper's fitness η_i accounts for the perceived novelty and importance of the reported discovery,
- c_i^t is the cumulative number of citations acquired by paper i at time t (well-cited papers are more likely to be cited than less-cited papers)
- $P_i(t)$ captures the fact that new ideas are integrated into subsequent work, hence the novelty of each paper fades with time.

The Fitness of a Scientific Publication



The Fitness of a Scientific Publication

- The measurements indicate that the fitness distribution of *Cell* is shifted to the right, which means that *Cell* papers tend to have high fitness
- By comparison, the fitness of papers published in *Physical Review* are shifted to the left, indicating that the journal publishes fewer high fitness papers

The Fitness of a Scientific Publication

- The BB model allows us to determine the fitness of individual nodes and the shape of the fitness distribution $\rho(\eta)$
- The fitness distribution is exponentially bounded, meaning that fitness differences between different nodes are small.
- With time these differences are magnified, resulting in a power law degree distribution of incoming links (WWW) or a broad citation distribution (citation networks)