Advanced Algorithms and Computational Models (module A)

The Barabasi-Albert Model

Giacomo Fiumara
giacomo.fiumara@unime.it

2020-2021

The Barabasi-Albert Model

Hubs represent the most striking difference between a random and a scale-free network

The existence of hubs such as `google.com` or molecules like ATP (involved in a number of chemical reactions) raises two questions:

- Why do so different system as the WWW or the cell converge to a similar scale-free architecture?
- Why does the random network model fail to reproduce the hubs and the power laws observed in real networks?

Growth and Preferential Attachment

Why are hubs and power laws absent in random networks?

The answer was found in 1999 (Barabasi and Albert, 23944 citations)

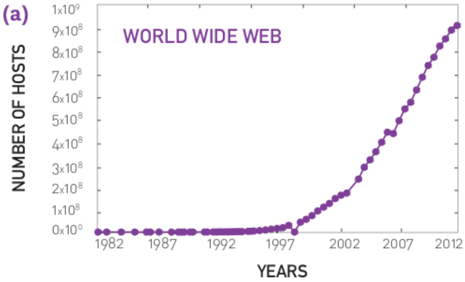
Author highlighted two hidden assumptions of the Erdös-Rényi model

- Networks expand through the addition of new nodes
- Nodes prefer to link to the more connected nodes

Growth and Preferential Attachment

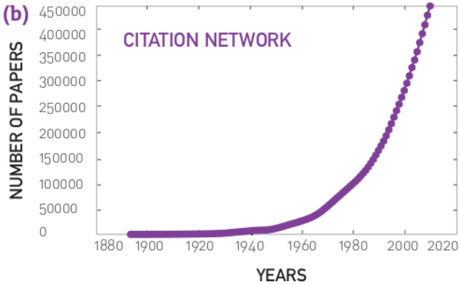
Networks expand through the addition of new nodes

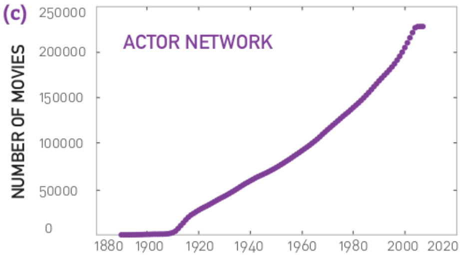
- In 1991 the WWW had a single node, the first webpage built by TBL, the creator of the Web
- Today the Web has over a trillion (10^{12}) documents
- This number was reached through the continuous addition of new documents by millions of individuals and institutions



Growth and Preferential Attachment

Networks expand through the addition of new nodes





Growth and Preferential Attachment

Nodes prefer to link to the more connected nodes

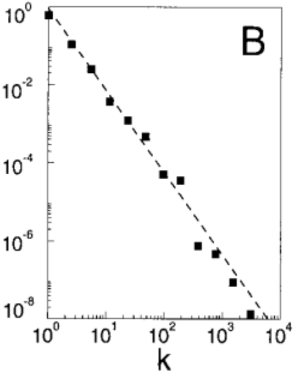
The random network model assumes that we randomly choose the interaction partner of a node

Most real network nodes prefer to link to the more connected nodes (*preferential attachment*)

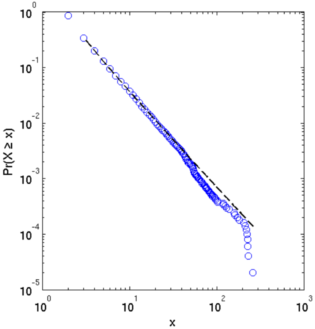
Growth and Preferential Attachment

Nodes prefer to link to the more connected nodes

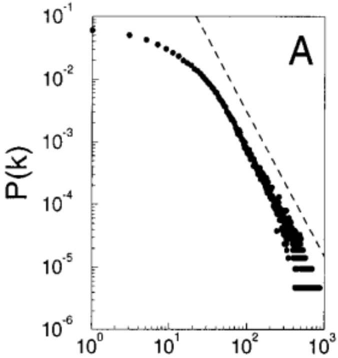
– We are familiar with only a tiny fraction of the trillion web documents. We are more likely yo link to a high-degree node than to a node with only few links



- No scientist can attempt to read the more than a million scientific papers published each year. The more cited is a paper, the more likely that we hear about it and eventually read it. We cite what we read, therefore our citations are biased towards the more cited publications (the high-degree nodes of the citation network)

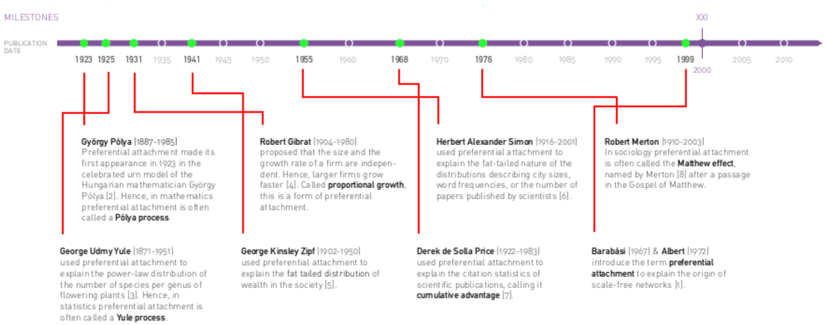


- The more movies an actor has played, the more familiar is a casting director with her skills. Hence, the higher the degree of an actor in the actor network, the higher are the chances that she will be considered for a new role



Growth and Preferential Attachment

Nodes prefer to link to the more connected nodes



Growth and Preferential Attachment

The random network model differs from real networks in two important characteristics:

- Growth

  Real networks are the result of a growth process that continuously increases N. In contrast, the random network model assumes that the number of nodes N is fixed

- Preferential Attachment

  In real networks new nodes tend to link to the more connected nodes. In contrast, nodes in random networks randomly choose their interaction partners

The Barabasi-Albert Model

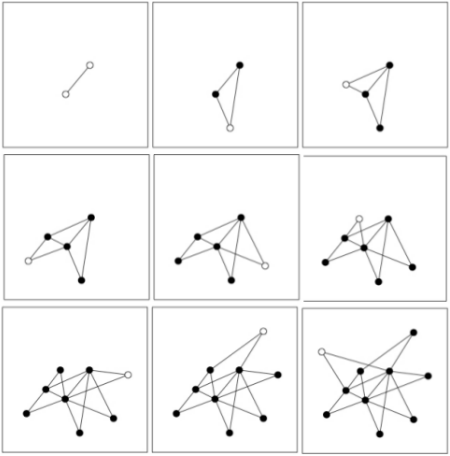
The *Barabasi-Albert* (BA) model is defined as follows:

We start with m_0 nodes, the links between which are chosen arbitrarily, as long as each node has at least one link. The network develops following two steps:

- **Growth** – At each timestep we add a new node with m ($\leq m_0$) links that connect the new node to m nodes already in the network
- **Preferential attachment** – The probability $\Pi(k)$ that a link of the new node connects to node i depends con the degree k_i as

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

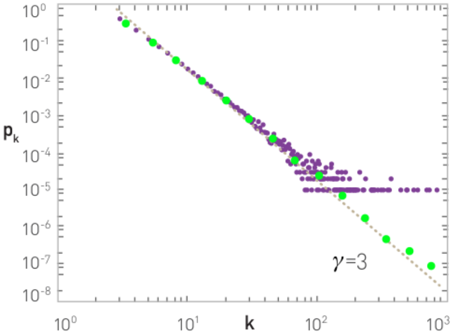
The Barabasi-Albert Model



The Barabasi-Albert Model

- Preferential attachment is a probabilistic mechanism
- A new node is free to connect to *any* node in the network, whether it is a hub or has a single link
- However, it is highly probable that a new node connects to a high degree node
- After t timesteps, the BA model generates a network with $N = t + m_0$ nodes and $m_0 + mt$ links
- The obtained network has a power-law degree distribution with degree exponent $\gamma = 3$

The Barabasi-Albert Model



Degree Dynamics

To understand the emergence of the scale-free property it is necessary to focus on the time evolution of the BA model

- In the model, an existing node can increase its degree each time a *new* node enters the network
- This node will link to m of the $N(t)$ nodes already present in the system
- The probability that one of these links connects to node i is

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

Degree Dynamics

- Let us approximate the degree k_i with a continuous real variable
- The rate at which an existing node i acquires links as a result of new nodes connecting to it is

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

- The coefficient m describes that each node arrives with m links
- Hence, node i has m chances to be chosen
- Moreover

$$\sum_{j=1}^{N-1} k_j = 2mt - m = m(2t-1)$$

Degree Dynamics

Therefore

$$\frac{dk_i}{dt} = m\frac{k_i}{m(2t-1)} = \frac{k_i}{2t-1}$$

For large t the -1 term can be neglected in the denominator, obtaining

$$\frac{dk_i}{k_i} = \frac{1}{2}\frac{dt}{t}$$

which can be integrated using the fact that $k_i(t_i) = m$ (node i joins the network at time t_i with m links). We obtain:

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

Where $\beta = \frac{1}{2}$ is called *dynamical exponent*

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

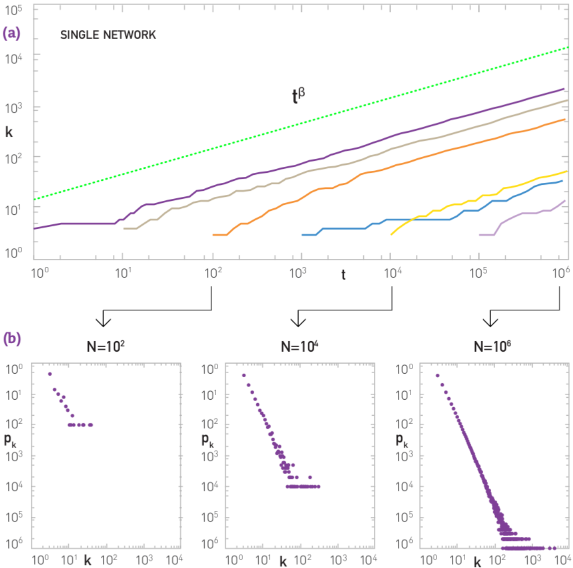
- The degree of each node increases following a power-law with the same dynamical exponent β. Hence all nodes follow the same dynamical law
- The growth in the degrees is sublinear ($\beta < 1$). This is a consequence of the growing nature of the BA model: each node has more nodes to link to than the previous node. Hence, with time the existing nodes compete for links with an increasing pool of other nodes

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

- The earlier node i was added, the higher is its degree $k_i(t)$. Hence, hubs are large because they arrived earlier, a phenomenon called *first-mover advantage* in marketing and business
- The rate at which the node i acquires new links is given by
  $$\frac{dk_i(t)}{dt} = \frac{m}{2}\frac{1}{\sqrt{t_i t}}$$
  indicating that in each time step older nodes acquire more links (as they have smaller t_i). The rate at which a node acquires links decreases with time as $t^{-1/2}$

Degree Dynamics

Some considerations



Degree Distribution

- To calculate the degree distribution of the BA model in the continuum approximation we first calculate the number of nodes with degree smaller than k, i.e. $k_i(t) < k$
- Using

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

  we can obtain

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

- In the BA model, a node is added at equal time step. Therefore, the number of nodes with degree smaller than k is

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

Degree Distribution

- Altogether there are $N = m_0 + t$ nodes, which becomes $N \approx t$ in the large t limit
- Therefore the probability that a randomly chosen node has degree k or smaller, which is the cumulative degree distribution, follows

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

- By taking the derivative we obtain the degree distribution

$$p_k = \frac{\partial P(k)}{\partial k} = \frac{1}{\beta} \frac{m^{1/\beta}}{k^{1/\beta+1}} = 2m^2 k^{-\gamma}$$

with $\gamma = \frac{1}{\beta} + 1 = 3$

Degree Distribution

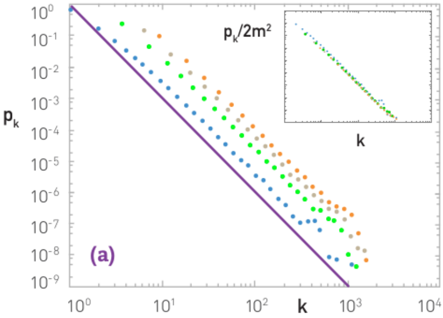
The continuum theory predicts the correct degree exponent, but it fails to accurately predict the pre-factors. The exact degree distribution of the BA model can be obtained using other approaches and is

$$p_k = \frac{2m(m+1)}{k(k+1)(k+2)}$$

Degree Distribution

$$p_k = \frac{2m(m+1)}{k(k+1)(k+2)}$$

The degree exponent γ is independent of m, a prediction that agrees with the numerical results

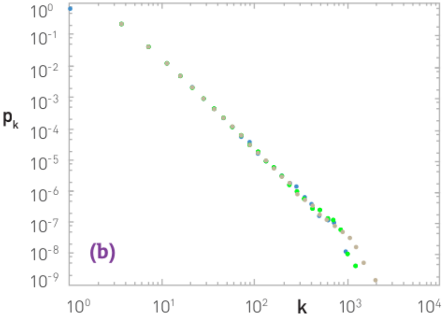


$m_0 = m = 1$ (blue), 3 (green), 5 (grey), 7 (orange)

Degree Distribution

$$p_k = \frac{2m(m+1)}{k(k+1)(k+2)}$$

The degree distribution of the BA model is independent of both t and N. This is in agreement with real networks that differ in age and size



Absence of Growth or Preferential Attachment

The coexistence of growth and preferential attachment in the BA model leads to a question: Are they both necessary for the emergence of the scale-free property?

Is it possible to generate a scale-free network with only one of the two ingredients?

Absence of Growth or Preferential Attachment

Model A: absence of preferential attachment

Model A starts with m_0 nodes and follows these steps:

- **Growth** – At each time step we add a new node with m ($\leq m_0$) links that connect to m nodes added earlier
- **Preferential Attachment** – The probability that a new node links to a node with degree k_i is

$$\Pi(k_i) = \frac{1}{m_0 + t - 1}$$

- This means that $\Pi(k_i)$ is independent of k_i, indicating that new nodes choose randmly the nodes they link to

Absence of Growth or Preferential Attachment

Model A: absence of preferential attachment

- For Model A the continuum theory predicts that $k_i(t)$ increases logarithmically with time

$$k_i(t) = mln\left(e\frac{m_0 + t - 1}{m_0 + t_i - 1}\right)$$

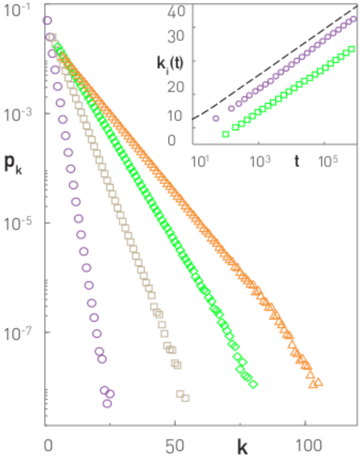
- Consequently, the degree distribution follows an exponential

$$p(k) = \frac{e}{m}exp\left(-\frac{k}{m}\right)$$

- An exponential function decays faster than a power law, hence it does not support hubs
- The lack of preferential attachment eliminates the scale-free character of hubs from network. Indeed, as all nodes acquire links with equal probability, we lack a *rich-get-richer* process

Absence of Growth or Preferential Attachment

Model A: absence of preferential attachment



$m_0 = m = 1$ (circles), 3 (squares), 5 (diamonds), 7 (triangles)

Absence of Growth or Preferential Attachment

Model B: absence of growth

Model B starts with N nodes and evolves following this step

- **Preferential Attachment** – At each time step a node is selected randomly and connected to node i with degree k_i already present in the network, where i is chosen with probability $\Pi(k)$. $\Pi(0) = 0$, therefore nodes with $k = 0$ are assumed to have $k = 1$, otherwise they can not acquire links
- In Model B the number of nodes remains constant during the evolution of the network, while the number of links increases linearly with time. As a result:

$$k_i(t) = \frac{2}{N} t$$

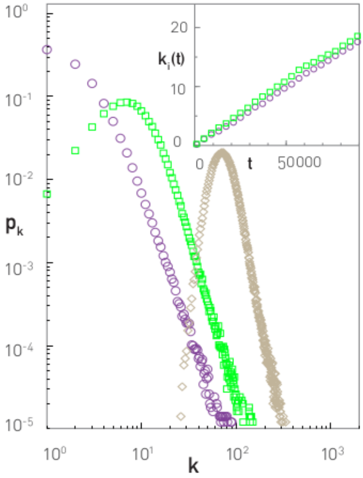
Absence of Growth or Preferential Attachment

Model B: absence of growth

- At early time, when there are only a few links in the network ($L \ll N$) each new link connects previously unconnected nodes. In this stage, the evolution is indistinguishable from the BA model with $m = 1$
- After a transient period, the node degree converge to the average degree and the degree develops a peak. For $t \to N(N-1)/2$ the network becomes a complete graph in which all nodes have degree $k_{max} = N - 1$, hence $p_k = \delta(N-1)$

Absence of Growth or Preferential Attachment

Model B: absence of growth



$t = N$ (circles), $t = 5N$ (squares), $t = 40N$ (diamonds)

Absence of Growth or Preferential Attachment

- The absence of preferential attachment leads to a growing network with a stationary but exponential degree distribution
- The absence of growth leads to the loss of stationarity, forcing the network to converge to a complete graph
- This failure of models A and B to reproduce the empirically observed scale-free distribution indicates that growth and preferential attachment are simultaneously needed for the emergence of the scale-free property

Measuring Preferential Attachment

Growth and preferential attachment are jointly responsible for the scale-free property

- Growth is easily detectable: all large networks have reached their size by adding new nodes
- Preferential attachment is also present in real networks, and can be detected experimentally

Measuring Preferential Attachment

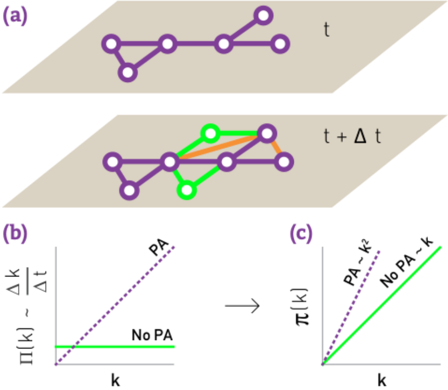
Preferential attachment relies on two distinct hypotheses:

- The likelihood to connect to a node depends on that node's degree k. This is in contrast with the random network model, for which $\Pi(k)$ is independent of k
- The functional form of $\Pi(k)$ is linear in k

Both hypotheses can be tested by measuring $\Pi(k)$, which can be determined for systems for which we known the time at which each node joined the network

Measuring Preferential Attachment

Consider a network for which we have two different maps, taken at time t and $t + \Delta t$



Measuring Preferential Attachment

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

For nodes that changed their degree during the Δt time frame we measure

$$\Delta k_i = k_i(t + \Delta t) - k_i(t)$$

The relative change is

$$\frac{\Delta k_i}{\Delta t} \sim \Pi(k_i)$$

This approximation holds if Δt is small, so that the changes in Δk are modest. But Δt must not be too small so that there are still detectable differences between the two networks

Measuring Preferential Attachment

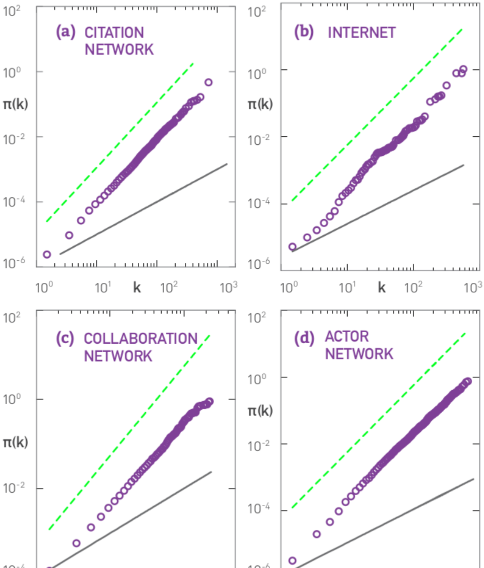
The curve $\Delta k_i/\Delta t$ can be very noisy

To reduce this noise the *cumulative preferential attachment function* is used instead

$$\pi(k) = \sum_{k_i=0}^{k} \Pi(k_i)$$

- In the absence of preferential attachment we have $\Pi(k_i) =$ *constant*, hence $\pi(k) \sim k$
- If preferential attachment is present, i.e. if $\Pi(k_i) = k_i$, it follows $\pi(k) \sim k^2$

Measuring Preferential Attachment



Measuring Preferential Attachment

- In the previous Figure are shown the measured $\pi(k)$ for four real networks
- For each system we observe a faster than linear increase in $\pi(k)$, indicating the presence of preferential attachment
- $\Pi(k)$ can be approximated with

$$\Pi(k) \sim k^{\alpha}$$

- For the Internet and the citation networks we have $\alpha \approx 1$, indicating that $\Pi(k)$ depends linearly on k
- For the co-authorship and the actor networks the best fit provides $\alpha \approx 0.9 \pm 0.1$, indicating the presence of a *sublinear preferential attachment*

Measuring Preferential Attachment

The theoretical results predict the existence of four scaling regimes:

- **No preferential attachment** ($\alpha = 0$) – The network has a simple exponential degree distribution. Hubs are absent and the resulting network is similar to a random network
- **Sublinear regime** ($0 < \alpha < 1$) – In this region fewer and smaller hubs are present than in a scale-free network. As $\alpha \to 1$ p_k follows a power law over an increasing range of degrees
- **Linear regime** ($\alpha = 1$) –This corresponds to the BA model, hence the degree distribution follows a power law
- **Superlinear regime** ($\alpha > 1$) – The high-degree nodes are very attractive. In this configuration the earliest nodes become super hubs and all subsequent nodes link to them.

The Origins of Preferential Attachment

The key role of preferential attachment poses another question: where does it come from? From this question two more detailed questions derive:

- Why does $\Pi(k)$ depend on k?
- Why is the dependence of $\Pi(k)$ linear in k?

The Origins of Preferential Attachment

Two philosophically different answers emerged to these questions:

- The first view preferential attachment as the interplay between random events and some structural property of a network. These mechanisms rely on random events and are therefore called *local* or *random* mechanisms
- The second assumes that each new node or link balances conflicting needs, hence they are preceeded by a cost-benefit analysis. These models assume familiarity with the whole network and rely on optimization principles and are therefore called *global* or *optimized* mechanisms

The BA model postulates the presence of preferential attachment

Yet, it is possible to build models that generate scale-free networks apparently without preferential attachment. They work by *generating* preferential attachment

Two models have been developed:

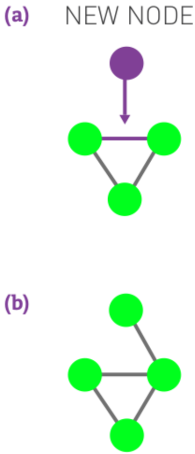
- Link Selection Model
- Copying Model

The *link selection model* is a simple example of a local mechanism that generates a scale-free network without preferential attachment. It is defined as follows:

- **Growth** – At each time step we add a new node to the network
- **Link Selection** – We select a link at random and connect the new node to one of the two nodes at the two ends of the selected link

This model requires no knowledge about the overall network topology, hence is inherently local and random

Yet, it generates preferential attachment



(a)
NEW NODE
(b)

The *link selection model* is a simple example of a local mechanism that generates a scale-free network without preferential attachment. It is defined as follows:

- **Growth** – At each time step we add a new node to the network
- **Link Selection** – We select a link at random and connect the new node to one of the two nodes at the two ends of the selected link

This model requires no knowledge about the overall network topology, hence is inherently local and random

Yet, it generates preferential attachment

The probability q_k that the node at the end of a randomly chosen link has degree k is

$$q_k = Ckp_k$$

- The higher the degree of the node, the higher the chance that it is located at the end of the chosen link
- The more degree-k nodes are in the network, the more likely that a degree k node is at the end of the link

- C can be calculated using the normalization condition

$$\sum q_k = 1$$

from which one get

$$\sum q_k = 1 \implies C \sum k p_k = 1$$

$$C\langle k \rangle = 1 \implies C = \frac{1}{\langle k \rangle}$$

and therefore

$$q_k = \frac{k p_k}{\langle k \rangle}$$

which represents the probability that a new node connects to a node with degree k

The *copying model* mimics a simple phenomenon: the authors of a new webpage tend to borrow links from other pages on related topics

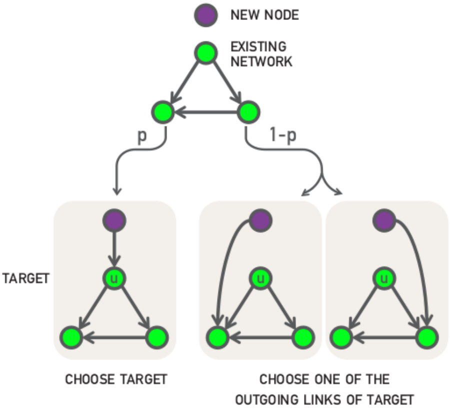
In each time step a new node is added to the network. To decide where it connects we randomly select a node u

Then a two-step procedure is followed:

- *Random connection* – With probability p the new node links to u, which means that we link to the randomly selected node
- *Copying* – With probability $1 - p$ we randomly choose an *outgoing* link of node u and link the new node to the target of the link. In other words, the new node copies a link of node u and connects it to its target rather than connecting to node u directly

The Origins of Preferential Attachment

Local mechanisms: Copying Model



The probability of selecting a particular node in the first step is $1/N$

The probability of selecting a node linked to a degree-k node through this copying step is

$$\frac{k}{2L}$$

for undirected networks. Therefore:

$$\Pi(k) = \frac{p}{N} + \frac{1-p}{2L}k$$

which is linear in k and therefore predicts a linear preferential attachment

The copying model is popular due to its relevance to real systems:

- *Social Networks* – The more acquaintances an individual has, the higher is the chance that she will be introduced to new individuals by her existing acquaintances. We copy the friends of our friends
- *Citation Networks* – Authors decide what to read and cite by copying references from the papers they have read. Papers with more citations are more likely to be studied and studied again
- *Protein Interactions* – Gene duplication, responsible for the emergence of new genes in a cell, can be mapped into the copying model, explaining the scale-free nature of protein interaction networks

The Origins of Preferential Attachment

Global mechanisms: Optimization

- According to an assumption of economics, humans make rational decisions, balancing cost against benefits
- In other words, each individual aims to maximize its personal advantage. Such rational decisions can lead to preferential attachment
- Consider Internet, whose nodes are routers connected via cables. Each new router will choose its link to balance access to good network performance with the cost of laying down a new cable

Let us consider a network. At each time step we add a new node i and calculate the cost function

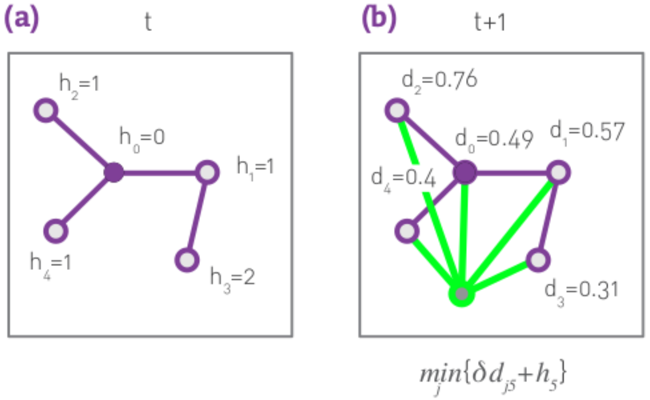
$$C_i = min_j[\delta d_{ij} + h_j]$$

which compares the cost of connecting to each node j already in the network

- d_{ij} is the Euclidean distance between the new node i and the potential target j
- h_j is the network-based distance of node j to the “center” of the network

The Origins of Preferential Attachment

Global mechanisms: Optimization



$$C_i = min_j[\delta d_{ij} + h_j]$$

Three distinct network topologies emerge, depending on the value of the parameter δ and N

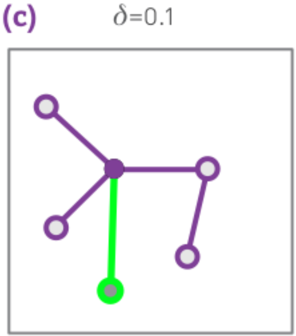
- Star Network ($\delta < (1/2)^{1/2}$)
- Random Network ($\delta \geq N^{1/2}$)
- Scale-free Network ($4 \geq \delta \geq N^{1/2}$)

$$C_i = min_j[\delta d_{ij} + h_j]$$

Star Network ($\delta < (1/2)^{1/2}$)

For $\delta = 0$ the Euclidean distances are irrelevant, hence each node links to the central node, turning the network into a star

We have a star configuration whenever the term h_j dominates over δd_{ij}


(c)
δ=0.1

The Origins of Preferential Attachment

Global mechanisms: Optimization

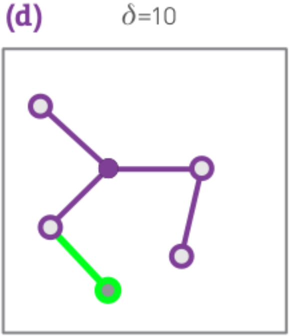
$$C_i = min_j[\delta d_{ij} + h_j]$$

Random Network ($\delta \geq N^{1/2}$)

For very large δ the contribution provided by the distance term δd_{ij} overwhelms h_j

In this case each new node connects to the node closest to it

The resulting network will have a bounded degree distribution, like a random network



(d)
δ=10

The Origins of Preferential Attachment

Global mechanisms: Optimization

$$C_i = min_j[\delta d_{ij} + h_j]$$

Scale-free Network ($4 \geq \delta \geq N^{1/2}$)

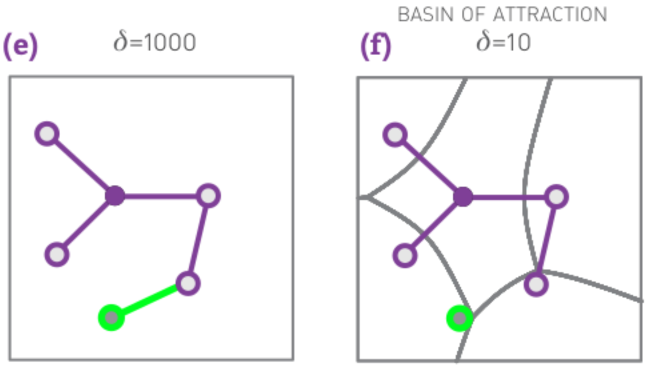
For intermediate values of δ the network develops a scale-free topology

The origin of the power law distribution in this regime is due to:

- **Optimization** – Each node has a basin of attraction, so that nodes landing in this basin will always link to it. The size of each basin correlates with h_j of node j at its center, which in turn correlates with the degree k_j of the node
- **Randomness** – We choose randomly the location of the new node, ending in one of the N basins of attraction. The node with the largest degree has largest basin of attraction, hence gains the most new nodes and links. This leads to preferential attachment

The Origins of Preferential Attachment

Global mechanisms: Optimization



The mechanism responsible for preferential attachment can have two fundamentally different origins

- Random processes, like link selection or copying
- Optimization, when new nodes balance conflicting criteria as they decide where to connect

Each of these mechanisms lead to linear preferential attachment, as assumed in the BA model

Linear preferential attachment is present in so many and so different systems because it can come from both rational choice and random actions