Class 10: Evolving Networks
(Chapter 6 in Textbook)

Boleslaw Szymanski
Fitness Model
Fitness Model: Can Latecomers Make It?

**SF model:** \( k(t) \sim t^{\frac{1}{2}} \) (first mover advantage)

**Fitness model:**

\[
\text{fitness } (\eta) \quad \Pi(k_i) \approx \frac{\eta_i k_i}{\sum_j \eta_j k_j}
\]

\( k(\eta,t) \sim t^{\beta(\eta)} \)

\( \beta(\eta) = \eta/C \)

Section 5.3

- The degree of each node increases following a power-law with the same dynamical exponent \( \beta = 1/2 \) (Figure 5.6a). Hence all nodes follow the same dynamical law.

- The growth in the degrees is sublinear (i.e. \( \beta < 1 \)). This is a consequence of the growing nature of the Barabási-Albert model: Each new 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.

- 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 the derivative of (5.7)

\[
\frac{dk_i(t)}{dt} = \frac{m}{2} \frac{1}{\sqrt{t_i t}}
\]

indicating that in each time frame older nodes acquire more links (as they have smaller \( t_i \)). Furthermore, the rate at which a node acquires links decreases with time as \( t^{-1/2} \). Hence, fewer and fewer links go to a node.
Absence of growth and preferential attachment
\[ \Pi(k_i) : \text{uniform} \]

\[ \frac{\partial k_i}{\partial t} = A \Pi(k_i) = \frac{m}{m_0 + t - 1} \]

\[ k_i(t) = m \ln\left(\frac{m_0 + t - 1}{m + t_i - 1}\right) + m \]

\[ P(k) = \frac{e}{m} \exp\left(-\frac{k}{m}\right) \sim e^{-k} \]
\[ \frac{\partial k_i}{\partial t} = A \Pi(k_i) + \frac{1}{N} = \frac{N}{N-1} \frac{k_i}{2t} + \frac{1}{N} \]

\[ k_i(t) = \frac{2(N-1)}{N(N-2)} t + C t^{\frac{N}{2(N-1)}} \sim \frac{2}{N} t \]

\( p_k : \) power law (initially) \( \rightarrow \) Gaussian \( \rightarrow \) Fully Connected
Do we need both growth and preferential attachment?

YEP
EMPIRICAL DATA FOR REAL NETWORKS

Pathlengtgh

Clustering

Degree Distr.

Regular network

Erdos-Renyi

Watts-Strogatz

Barabasi-Albert

$P(k) \sim k^{-\gamma}$

$P(k) = \delta(k - k_d)$

$P(k) = e^{-k^<} \frac{<k^>^k}{k!}$

$C = \text{const}$

$C_{\text{rand}} = p = \frac{<k>}{N}$

$C_{\text{rand}} \approx n^{1/2}$

$P(k) = \frac{\log N}{\log<k>}$
The origins of preferential attachment
Link selection model -- perhaps the simplest example of a local or random mechanism capable of generating preferential attachment.

**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 nodes at the two ends of the selected link.

To show that this simple mechanism generates linear preferential attachment, we write the probability that the node at the end of a randomly chosen link has degree $k$ as

$$q_k = C k p_k$$

In (5.26) $C$ can be calculated using the normalization condition $\sum q_k = 1$, obtaining $C=1/\langle k \rangle$. Hence the probability to find a degree-$k$ node at the end of a randomly chosen link is

$$q_k = \frac{k p_k}{\langle k \rangle} ,$$
In *An Informal Theory of the Statistical Structure of Languages* [26], Benoit Mandelbrot proposes optimization as the origin of power laws.

In 1953, Mandelbrot publishes a comment on Simon's paper [27], writing:

Simon’s model is analytically circular...

In 1955, Herbert Simon [6] proposes randomness as the origin of power laws and dismisses Mandelbrot’s claim that power laws are rooted in optimization.

Mandelbrot’s principal and mathematical objections to the model are shown to be unfounded.

The essence of Simon’s lengthy reply a year later is well summarized in its abstract [28].

In a 10-page response entitled *Final Note*, Mendelbrot states [29]:

...Most of Simon’s (1960) reply was irrelevant.

This present ‘Reply’ refutes the almost entirely new arguments introduced by Dr. Mandelbrot in his ‘Final Note’...

Simon’s subsequent *Reply to ‘Final Note’ by Mandelbrot* does not concede [30].

In the creatively titled *Post Scriptum to Final Note*, Mandelbrot [31] writes:

My criticism has not changed since I first had the privilege of commenting upon a draft of Simon. ☹️

Dr. Mandelbrot has proposed a new set of objections to my 1955 models of Yule distributions. Like earlier objections, these are invalid.

Simon’s final reply ends but does not resolve the debate [32].
György Pólya [1887-1985] used preferential attachment to explain the power-law distribution of the number of species per genus of flowering plants [3]. Hence, in statistics preferential attachment is often called a Pólya process.

George Udny Yule [1871-1951] used preferential attachment to explain the power-law degree distribution of the number of species per genus of flowering plants [3]. Hence, in evolution and population genetics preferential attachment is often called a Yule process.

Robert Gibrad [1904-1980] proposed that the size and the growth rate of a firm are independent. Hence, larger firms grow faster [4]. Called proportional growth, this is a form of preferential attachment.

George Kinsley Zipf [1902-1950] used preferential attachment to explain the fat-tailed distribution of wealth in the society [5].

Herbert Alexander Simon [1916-2001] used preferential attachment to explain the fat-tailed nature of the distributions describing city sizes, word frequencies, or the number of papers published by scientists [6].

Derek de Solla Price [1922-1983] used preferential attachment to explain the citation statistics of scientific publications, referring to it as cumulative advantage [7].


1. Copying mechanism
   directed network
   select a node and an edge of this node
   attach to the endpoint of this edge

2. Walking on a network
   directed network
   the new node connects to a node, then to every
   first, second, … neighbor of this node

3. Attaching to edges
   select an edge
   attach to both endpoints of this edge

4. Node duplication
   duplicate a node with all its edges
   randomly prune edges of new node
Section 9

Copying model

(a) **Random Connection**: with probability $p$ the new node links to $u$.

(b) **Copying**: with probability $p$ we randomly choose an outgoing link of node $u$ and connect the new node to the selected link's target. Hence the new node “copies” one of the links of an earlier node.

(a) the probability of selecting a node is $1/N$.
(b) is equivalent with selecting a node linked to a randomly selected link. The probability of selecting a degree-$k$ node through the copying process of step (b) is $k/2L$ for undirected networks.

The likelihood that the new node will connect to a degree-$k$ node follows preferential attachment.

Social networks: Copy your friend’s friends.

Citation Networks: Copy references from papers we read.

Protein interaction networks: gene duplication,
Preferential Attachment!

\[ \frac{\partial k_i}{\partial t} \propto \Pi(k_i) \sim \frac{\Delta k_i}{\Delta t} \]

For given \( \Delta t \): \( \Delta k \propto \Pi(k) \)

\( k \) vs. \( \Delta k \): linear increase in the # of links

S. Cerevisiae PIN: proteins classified into 4 age groups

SUMMARY: PROPERTIES OF THE BA MODEL

- Nr. of nodes: \( N = t \)
- Nr. of links: \( L = m \, t \)
- Average degree: \( \langle k \rangle = \frac{2L}{N} \to 2m \)
- Degree dynamics: \( k_i(t) = m \left( \frac{t}{t_i} \right)^\beta \), \( \beta = \frac{1}{2} \) \( \beta \): dynamical exponent
- Degree distribution: \( P(k) \sim k^{-\gamma} \), \( \gamma = 3 \) \( \gamma \): degree exponent
- Average Path Length: \( l \approx \frac{\ln N}{\ln \ln N} \)
- Clustering Coefficient: \( C \sim \frac{(\ln N)^2}{N} \)

The network grows, but the degree distribution is stationary.
Can we change the degree exponent?
Section 9

Optimization model for connecting new router

(a) (b)

$h_j$ denotes the distance of node $j$ to the central node $h_0$.

$d_{ij}$ denotes the bandwidth between nodes $i$, $j$.

$\delta$ denotes the ratio of cost of cable to delay.

Question: where to place a new router?

$$C_i = \min_j [\delta d_{ij} + h_j]$$
Section 9  Optimization model

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

The vertical boundary of the star configuration is at \( \delta = (1/2)^{1/2} \). This is the inverse of the maximum distance between two nodes on a square lattice with unit length, over which the model is defined. Therefore, if \( \delta < (1/2)^{1/2} \), for any new node \( \delta d_{ij} < 1 \) and the cost (5.28) of connecting to the central node is \( C_i = \delta d_{ij} + 0 \), always lower than connecting to any other node at a cost of \( f(i,j) = \delta d_{ij} + 1 \). Therefore, for \( \delta < (1/2)^{1/2} \) all nodes connect to node 0 (star-and-spoke network (c)).
$C_i = \min_j [\delta d_{ij} + h_j]$
\[ C_i = \min_j [\delta d_{ij} + h_j] \]
Diameter and clustering coefficient
Section 10

Diameter

\[ D \sim \frac{\log N}{\log \log N} \]

Bollobas, Riordan, 2002
Reminder: for a random graph we have:

\[ C_{\text{rand}} = \frac{\langle k \rangle}{N} \sim N^{-1} \]

What is the functional form of \( C(N) \)?

\[ C = \frac{m}{8} \frac{(\ln N)^2}{N} \]

Konstantin Klemm, Victor M. Eguiluz,
Growing scale-free networks with small-world behavior,
Denote the probability to have a link between node \(i\) and \(j\) with \(P(i,j)\).

The probability that three nodes \(i,j,l\) form a triangle is \(P(i,j)P(i,l)P(j,l)\).

The expected number of triangles in which a node \(l\) with degree \(k_l\) participates is thus:

\[
Nr_l(\Delta) = \sum_{i=1}^{N} \int_{i=1}^{N} dj P(i,j)P(i,l)P(j,l)
\]

We need to calculate \(P(i,j)\).
Calculate $P(i,j)$.

Node $j$ arrives at time $t_j=j$ and the probability that it will link to node $i$ with degree $k_i$ already in the network is determined by preferential attachment:

$$P(i,j) = m \prod_{i} (k_i(j)) = m \frac{k_i(j)}{\sum_{i=1}^{k_i} k_i} = m \frac{k_i(j)}{2m t_j}$$

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

Where we used that the arrival time of node $j$ is $t_j=j$ and the arrival time of node is $t_i=i$.

$$P(i,j) = \frac{m}{2} \left( \frac{ij}{t_j} \right)^{-1/2}$$

$$N_{r_i}(\Delta) = \int_{i=1}^{N} \int_{j=1}^{N} di dj P(i,j) P(i,l) P(j,l) = \frac{m^3}{8} \int_{i=1}^{N} di \int_{j=1}^{N} dj (ij)^{-1/2} (il)^{-1/2} (jl)^{-1/2} = \frac{m^3}{8l} \int_{i=1}^{N} di \int_{j=1}^{N} dj = \frac{m^3}{8l} (\ln N)^2$$

$$C = \frac{m^3}{8l} (\ln N)^2 / k_i(k_i - 1)/2$$

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

Which is the degree of node $l$ at current time, at time $t=N$.

Let us approximate:

$$k_i(k_i - 1) \approx k_i^2 = m^2 \frac{N}{l}$$

There is a factor of two difference... Where does it come from?