Example: COVID-19 Control

• Formally defined optimal control in the risk networks: $x(k + 1) = F[x(k)] + G[x(k), E] + BU$

• Established a function of controllability index and corresponding optimal energy and conditions for nonnegative optimal control

• Provided a universal methodology of applying the LQR control in real world networked systems

• Qualitative analysis of COVID-19 governmental policies
Optimize functions

(a) Optimal energy

- Energy: 0.2331
- Control cost: 0.2331
- Total cost: 60.433

(b) LQR $R=I$, $Q=I$

- Energy: 3.1566
- Control cost: 3.1566
- Total cost: 12.403

Low cost of intermediate states
Contributions:

• Formally defined optimal control in the risk networks:
  \[ x(k + 1) = F[x(k)] + G[x(k), E] + BU \]

• Established a function of controllability index and corresponding optimal energy:
  - Controllability index \( \zeta = N / N_D \)
  - Upper bound of optimal energy \( \hat{J}_\epsilon = e^{10N/N_D} \)

• Established condition for nonnegative optimal control: \( N = N_D \)

• Quantitively analyzed the tradeoffs between control and state costs in Reactive and Proactive phases:
  - Reactive: cost is almost linearly related to the controlled number of active risks
  - Proactive: cost is proportional to the potential risk activities
  - Prevention is better than Governance: the cost in the proactive phase is much smaller than that in the reactive phase
Contributions:

• Provided a universal methodology of applying the LQR control in real world networked systems:
  • Built a flight-delay network with five million flights record in 2015.
  • Built a delay cost matrix $Q$ and aircraft cost matrix $R$ according to official statistic data

• Provided significant results on flights control:
  • LQR control saves around 90% time for the customer and 70% cost for the society on average.
  • In over 5000 unique flights, almost every single one benefits from the LQR control.

• Provided significant results on airports control:
  • The small airport in the inland area benefits more than large international one in the coastal area
  • In over 300 airports, almost every single one benefits from the LQR control.

• Discovered that the airline ranking by simulated steady states in the CARP model are highly (above 0.8) correlated with Airline Quality Ranking.

• Submitted to:
Questions?
Power laws and scale-free networks
Nodes: WWW documents  
Links: URL links  

Over 3 billion documents  

ROBOT: collects all URL’s found in a document and follows them recursively

Discrete vs. Continuum formalism

**Discrete Formalism**
As node degrees are always positive integers, the discrete formalism captures the probability that a node has exactly $k$ links:

$$\sum_{k=1}^{\infty} p_k = 1.$$  
$$C \sum_{k=1}^{\infty} k^{-\gamma} = 1.$$ 
$$C = \frac{1}{\zeta(\gamma)}.$$ 
$$p_k = \frac{k^{-\gamma}}{\zeta(\gamma)}.$$ 

**Continuum Formalism**
In analytical calculations it is often convenient to assume that the degrees can take up any positive real value:

$$p(k) = C k^{-\gamma}.$$  
$$\int_{k_{\min}}^{\infty} p(k) dk = 1.$$ 
$$C = \frac{1}{\int_{k_{\min}}^{\infty} k^{-\gamma} dk} = (\gamma - 1) k_{\min}^{\gamma - 1}.$$ 
$$p(k) = (\gamma - 1) k_{\min}^{\gamma - 1} k^{-\gamma}.$$  

**INTERPRETATION:**

$$P_k$$  
$$\int_{k_1}^{k_2} p(k) dk$$
Vilfredo Federico Damaso Pareto (1848 – 1923), Italian economist, political scientist and philosopher, who had important contributions to our understanding of income distribution and to the analysis of individuals choices. A number of fundamental principles are named after him, like Pareto efficiency, Pareto distribution (another name for a power-law distribution), the Pareto principle (or 80/20 law).
The difference between a power law and an exponential distribution.
Let us use the WWW to illustrate the properties of the high-\(k\) regime. The probability to have a node with \(k \sim 100\) is

- About \(p_{100} \approx 10^{-30}\) in a Poisson distribution

- About \(p_{100} \approx 10^{-4}\) if \(p_k\) follows a power law.

Consequently, if the WWW were to be a random network, according to the Poisson prediction we would expect \(10^{-18}\) \(k > 100\) degree nodes, or none.

For a power law degree distribution, we expect about \(N_{k > 100} = 10^9\) \(k > 100\) degree nodes
geles
The size of the biggest hub

All real networks are finite → let us explore its consequences.

→ We have an expected maximum degree, $k_{\text{max}}$

Estimating $k_{\text{max}}$

$$\int_{k_{\text{max}}}^{\infty} P(k) dk \approx \frac{1}{N}$$

Why: the probability to have a node larger than $k_{\text{max}}$ should not exceed the prob. to have one node, i.e. $1/N$ fraction of all nodes

$$\int_{k_{\text{max}}}^{\infty} P(k) dk = (\gamma - 1)k_{\text{min}}^{\gamma - 1} \int_{k_{\text{max}}}^{\infty} k^{-\gamma} dk = \frac{(\gamma - 1)}{(-\gamma + 1)} k_{\text{min}}^{\gamma - 1} \left[ k^{-\gamma + 1} \right]_{k_{\text{min}}}^{\infty} = \frac{k_{\text{min}}^{\gamma - 1}}{k_{\text{max}}^{\gamma - 1}} \approx \frac{1}{N}$$

$$k_{\text{max}} = k_{\text{min}} N^{\gamma - 1}$$
The size of the biggest hub

\[ k_{\text{max}} = k_{\text{min}} N^{\gamma - 1} \]

To illustrate the difference in the maximum degree of an exponential and a scale-free network let us return to the WWW sample of Figure 4.1, consisting of \( N = 3 \times 10^5 \) nodes. As \( k_{\text{min}} = 1 \), if the degree distribution were to follow an exponential, (4.17) predicts that the maximum degree should be \( k_{\text{max}} \approx 13 \). In a scale-free network of similar size and \( \gamma = 2.1 \), (4.18) predicts \( k_{\text{max}} \approx 85,000 \), a remarkable difference. Note that the largest in-degree of the WWW map of Figure 4.1 is 10,721, which is comparable to \( k_{\text{max}} \) predicted by a scale-free network. This reinforces our conclusion that in a random network hubs are effectively forbidden, while in scale-free networks they are naturally present.
Finite scale-free networks

Expected maximum degree, $k_{\text{max}}$

$$k_{\text{max}} = k_{\text{min}} N^{\gamma - 1}$$

- $k_{\text{max}}$ increases with the size of the network
  → the larger a system is, the larger its biggest hub

- For $\gamma > 2$ $k_{\text{max}}$ increases slower than $N$
  → the largest hub will contain a decreasing fraction of links as $N$ increases.

- For $\gamma = 2$ $k_{\text{max}} \sim N$.
  → The size of the biggest hub is $O(N)$

- For $\gamma < 2$ $k_{\text{max}}$ increases faster than $N$: condensation phenomena
  → the largest hub will grab an increasing fraction of links. Anomaly!
The size of the largest hub

\[ k_{\text{max}} = k_{\text{min}} N^{\frac{1}{\gamma-1}} \]

\[ k_{\text{max}} \sim N^{\frac{1}{\gamma-1}} \]

\[ (N - 1) \]

Scale-Free

Random Network

Network Science: Scale-Free Networks
The meaning of scale-free
Definition:

Networks with a power law tail in their degree distribution are called ‘scale-free networks’

Where does the name come from?

Critical Phenomena and scale-invariance
(a detour)
Phase transitions in complex systems I: Magnetism

\[ M(T - T_c, 0) \sim (T_c - T)^\beta \]

\[ \xi \sim |T - T_c|^{-\nu} \]
• Correlation length diverges at the critical point: the whole system is correlated!

• **Scale invariance**: there is no characteristic scale for the fluctuation (**scale-free behavior**).

• **Universality**: exponents are independent of the system’s details.
Divergences in scale-free distributions

\[ P(k) = Ck^{-\gamma} \quad k \in [k_{\text{min}}, \infty) \]
\[ \int_{k_{\text{min}}}^{\infty} P(k) \, dk = 1 \]
\[ C = \frac{1}{\int_{k_{\text{min}}}^{\infty} k^{-\gamma} \, dk} = (\gamma - 1)k_{\text{min}}^{\gamma - 1} \]

\[ P(k) = (\gamma - 1)k_{\text{min}}^{\gamma - 1}k^{-\gamma} \]

\[ < k^m > = \int_{k_{\text{min}}}^{\infty} k^m P(k) \, dk \]
\[ < k^m > = (\gamma - 1)k_{\text{min}}^{\gamma - 1} \int_{k_{\text{min}}}^{\infty} k^{m-\gamma} \, dk = \frac{(\gamma - 1)}{(m - \gamma + 1)} k_{\text{min}}^{\gamma - 1} \left[ k_{\text{min}}^{m-\gamma+1} \right]_{k_{\text{min}}}^{\infty} \]

If \( m-\gamma+1 < 0 \):
\[ < k^m > = -\frac{(\gamma - 1)}{(m - \gamma + 1)} k_{\text{min}}^m \]

If \( m-\gamma+1 > 0 \), the integral diverges.

For a fixed \( \gamma \) this means that all moments with \( m > \gamma - 1 \) diverge.
For a fixed \( \gamma \) this means all moments \( m > \gamma - 1 \) diverge.

Many degree exponents are smaller than 3

→ \(<k^2>\) diverges in the \( N \to \infty \) limit!!!
The meaning of scale-free

Random Network
Randomly chosen node: $k = \langle k \rangle \pm \langle k \rangle^{1/2}$
Scale: $\langle k \rangle$

Scale-Free Network
Randomly chosen node: $k = \langle k \rangle \pm \infty$
Scale: none
The meaning of scale-free

\[ k = \langle k \rangle \pm \sigma_k \]
Section 5

universality
Nodes: computers, routers
Links: physical lines

(Faloutsos, Faloutsos and Faloutsos, 1999)
SCIENCE CITATION INDEX

Nodes: papers
Links: citations

1736 PRL papers (1988)

\[ P(k) \sim k^{-\gamma} \]
\[ (\gamma = 3) \]

(S. Redner, 1998)
Nodes: scientist (authors)
Links: joint publication

(Newman, 2000, Barabasi et al 2001)
ONLINE COMMUNITIES

Nodes: online user
Links: email contact

Pussokram.com online community;
512 days, 25,000 users.

Kiel University log files
112 days, N=59,912 nodes

Ebel, Mielsch, Bornholdt, PRE 2002.
ONLINE COMMUNITIES

Twitter:

Facebook

Brian Karrer, Lars Backstrom, Cameron Marlow 2011
Barabasi-Albert Model
Organisms from all three domains of life are scale-free!

\[ P_{in}(k) \approx k^{-2.2} \]
\[ P_{out}(k) \approx k^{-2.2} \]

TOPOLOGY OF THE PROTEIN NETWORK

Nodes: proteins
Links: physical interactions-binding

\[ P(k) \sim (k + k_0)^{-\gamma} \exp\left(-\frac{k + k_0}{k_r}\right) \]

Growth and preferential attachment
Barabasi-Albert model Definition

The recognition that growth and preferential attachment coexist in real networks has inspired a minimal model called the Barabási-Albert model (BA model), which generates scale-free networks [1], 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:

1. **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.

2. **Preferential attachment:** the probability $\Pi(k)$ that a link of the new node connects to node $i$ depends on the degree $k_i$ as $\Pi(k_i) = k_i \Sigma_j k_j$

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, that if a new node has a choice between a degree-two and a degree-four node, it is twice as likely that it connects to the degree-four node.

The definition of the Barabási-Albert model leaves many mathematical details open:

- It does not specify the precise initial configuration of the first $m_0$ nodes.
- It does not specify whether the $m$ links assigned to a new node are added one by one, or simultaneously. This leads to potential mathematical conflicts: If the links are truly independent, they could connect to the same node $i$, leading to multi-links.

One possible definition with self-loops

$$p(i = s) = \begin{cases} \frac{k_i}{2t-1}, & \text{if } 1 \leq s \leq t-1 \\ \frac{1}{2t-1}, & \text{if } s = t \end{cases}$$
Degree dynamics
Degree distribution for Barabasi-Albert model

\[ k_i(t) = m \left( \frac{t}{t_i} \right)^\beta \]
\[ \beta = \frac{1}{2} \]
for \( t \geq m_o+i \) and 0 otherwise as system size at \( t \) is \( N=m_o+t-1 \)

We assume the initial \( m_o \) nodes create a fully connected graph.
A random node \( j \) arriving at time \( t \) is with equal probability \( 1/N=1/(m_o+t-1) \) one of the nodes \( 1, \ldots, N \), its degree will grow with the above equation, so

\[ P(k_j(t)) < k = P \left( t_j > \frac{1}{m^\beta t} \right) = 1 - P \left( t_j \leq \frac{1}{m^\beta k} \right) = 1 - \frac{1}{k^\beta (t + m_0 - 1)} \]

For the large times \( t \) (and so large network sizes) we can replace \( t-1 \) with \( t \) above, so

\[ : P(k) = \frac{\partial P(k_i(t)) < k}{\partial k} = \frac{2m^2 t}{m_0 + t} \frac{1}{k^3} \sim k^{-3} \]
\[ \gamma = 3 \]

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

\[ P(k) = \frac{2m^2 t}{t - t_0} \frac{1}{k^3} \sim k^{-\gamma} \]

(i) The degree exponent is independent of \( m \).

(ii) As the power-law describes systems of rather different ages and sizes, it is expected that a correct model should provide a time-independent degree distribution. Indeed, asymptotically the degree distribution of the BA model is independent of time (and of the system size \( N \))
\[ \implies \text{the network reaches a stationary scale-free state.} \]

(iii) The coefficient of the power-law distribution is proportional to \( m^2 \).

Section 4

\[ p_k \]

\[ \gamma = 3 \]
(a) We generated networks with $N=100,000$ and $m_0=m=1$ (blue), 3 (green), 5 (grey), and 7 (orange). The fact that the curves are parallel to each other indicates that $\gamma$ is independent of $m$ and $m_0$. The slope of the purple line is -3, corresponding to the predicted degree exponent $\gamma=3$. Inset: (5.11) predicts $p_k \sim 2m^2$, hence $p_k/2m^2$ should be independent of $m$. Indeed, by plotting $p_k/2m^2$ vs. $k$, the data points shown in the main plot collapse into a single curve.

(b) The Barabási-Albert model predicts that $p_k$ is independent of $N$. To test this we plot $p_k$ for $N = 50,000$ (blue), 100,000 (green), and 200,000 (grey), with $m_0=m=3$. The obtained $p_k$ are practically indistinguishable, indicating that the degree distribution is stationary, i.e. independent of time and system size.
NUMERICAL SIMULATION OF THE BA MODEL

m-dependence

\[ P(k) = \frac{2m(m+1)}{k(k+1)(k+2)} \]
\[ P(k) = \frac{2m^2 t}{m_o + t} \frac{1}{k^3} \]

\( \gamma = 3 \)

Stationarity:

\( P(k) \) independent of \( N \)

Insert:

degree dynamics

\[ k_i(t) = m \left( \frac{t}{t_i} \right)^{\beta} \quad \beta = \frac{1}{2} \]
The mean field theory offers the correct scaling, BUT it provides the wrong coefficient of the degree distribution. So asymptotically it is correct ($k \to \infty$), but not correct in details (particularly for small $k$).

To fix it, we need to calculate $P(k)$ exactly, which we will do next using a rate equation based approach.
MFT - Degree Distribution: Rate Equation

\[ < N(k, t) > = tP(K, t) \] Number of nodes with degree \( k \) at time \( t \).

Since at each timestep we add one node, we have \( N = t \) (total number of nodes = number of timesteps)

\[ \Pi(k) = \frac{k}{\sum_j k_j} = \frac{k}{2mt} \]

\( 2m \): each node adds \( m \) links, but each link contributed to the degree of 2 nodes

Number of links added to degree \( k \) nodes after the arrival of a new node:

\[ \frac{k}{2m} \times NP(k, t) \times m = \frac{k}{2} P(k, t) \]

Nr. of degree \( k-1 \) nodes that acquire a new link, becoming degree \( k \)

\[ \frac{k-1}{2} P(k-1, t) \]

Nr. of degree \( k \) nodes that acquire a new link, becoming degree \( k+1 \)

\[ \frac{k}{2} P(k, t) \]

\[ (N + 1)P(k, t + 1) = NP(k, t) + \frac{k-1}{2} P(k-1, t) - \frac{k}{2} P(k, t) \]

# k-nodes at time \( t+1 \)

# k-nodes at time \( t \)

Gain of k-nodes via \( k-1 \rightarrow k \)

Loss of k-nodes via \( k \rightarrow k+1 \)

MFT - Degree Distribution: Rate Equation

\[(N+1)P(k, t+1) = NP(k, t) + \frac{k-1}{2} P(k-1, t) - \frac{k}{2} P(k, t)\]

- \# k-nodes at time t+1
- \# k-nodes at time t
- Gain of k-nodes via k-1 \(\rightarrow\) k
- Loss of k-nodes via k \(\rightarrow\) k+1

We do not have \(k=0,1,...,m-1\) nodes in the network (each node arrives with degree \(m\))
\(\rightarrow\) We need a separate equation for degree \(m\) modes

\[(N+1)P(m, t+1) = NP(m, t) + 1 \frac{m}{2} P(m, t)\]

- \# m-nodes at time t+1
- \# m-nodes at time t
- Add one m-degree node
- Loss of an m-node via m \(\rightarrow\) m+1

We assume that there is a stationary state in the $N=t \to \infty$ limit, when $P(k,\infty)=P(k)$.

\[(N+1)P(k,t+1) = NP(k,t) + \frac{k-1}{2} P(k-1,t) - \frac{k}{2} P(k,t) \quad k>m\]

\[(N+1)P(m,t+1) = NP(m,t) + 1 - \frac{m}{2} P(m,t)\]

\[(N+1)P(k,t+1) - NP(k,t) \to NP(k,\infty) + P(k,\infty) - NP(k,\infty) = P(k,\infty) = P(k)\]

\[(N+1)P(m,t+1) - NP(m,t) \to P(m)\]

\[P(k) = \frac{k-1}{2} P(k-1) - \frac{k}{2} P(k)\]

\[P(m) = 1 - \frac{m}{2} P(m)\]

\[P(k) = \frac{k-1}{k+2} P(k-1) \quad k>m\]

\[P(m) = \frac{2}{2+m}\]
MFT - Degree Distribution: Rate Equation

\[
P(k) = \frac{k - 1}{k + 2} P(k - 1) \implies P(k + 1) = \frac{k}{k + 2} P(k)
\]

\[
P(m) = \frac{2}{m + 2}
\]

\[
P(m + 1) = \frac{m}{m + 3} P(m) = \frac{2m}{(m + 2)(m + 3)}
\]

\[
P(m + 2) = \frac{m + 1}{m + 4} P(m + 1) = \frac{2m(m + 1)}{(m + 2)(m + 3)(m + 4)}
\]

\[
P(m + 3) = \frac{m + 2}{m + 5} P(m + 2) = \frac{2m(m + 1)}{(m + 3)(m + 4)(m + 5)}
\]

\[
\ldots
\]

\[
P(k) = \frac{2m(m + 1)}{k(k + 1)(k + 2)} \quad P(k) \sim k^{-3} \quad \text{for large } k
\]

Krapivsky, Redner, Leyvraz, PRL 2000
Dorogovtsev, Mendes, Samukhin, PRL 2000
Bollobas et al, Random Struc. Alg. 2001

Network Science: Evolving Network Models
Start from eq. \[ P(k) = \frac{k-1}{2} P(k-1) - \frac{k}{2} P(k) \]

\[ 2P(k) = (k-1)P(k-1) - kP(k) = -P(k-1) - k[P(k) - P(k-1)] \]

\[ 2P(k) = -P(k-1) - k \frac{P(k) - P(k-1)}{k-(k-1)} = -P(k-1) - k \frac{\partial P(k)}{\partial k} \]

\[ P(k) = -\frac{1}{2} \frac{\partial [kP(k)]}{\partial k} \]

Its solution is: \[ P(k) \sim k^{-3} \]

Dorogovtsev and Mendes, 2003
All nodes follow the same growth law

\[ \frac{\partial k_i}{\partial t} \propto I_1(k_i) = A \frac{k_i}{\sum_j k_j} \]

Use: \( \sum_j k_j = 2m(t-1) + \frac{m_0(m_0-1)}{2} \)

During a unit time (time step): \( \Delta k = m \) \( \Rightarrow A = m \)

\[ \frac{\partial k_i}{\partial t} = \frac{k_i}{2t} \]

\[ \int \frac{\partial k_i}{k_i} = \int \frac{\partial t}{2t} \]

\[ \ln \left( \frac{k}{m} \right) = \frac{1}{2} \ln \left( \frac{t}{t_i} \right) = \ln \left( \frac{t}{t_i} \right)^{1/2} \]

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

\( \beta = \frac{1}{2} \)

\( \beta \): dynamical exponent

Fitness Model
**Fitness Model: Can Latecomers Make It?**

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

**Fitness model:**

Fitness \( (\eta) \)

\[ \Pi(k_i) \approx \frac{\eta_i k_i}{\sum_j \eta_j k_j} \]

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

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

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)

$$\frac{dk_i(t)}{dt} = \frac{m}{2} \frac{1}{\sqrt{t_i}}$$ (5.8)

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

Regular network
\[ l \approx \sqrt{N} \]

Erdos-Renyi
\[ l_{\text{rand}} \approx \frac{\log N}{\log \langle k \rangle} \]

Watts-Strogatz
\[ l_{\text{rand}} \approx \frac{\log N}{\log \langle k \rangle} \]

Barabasi-Albert

Pathlength

Clustering
\[ C \sim \text{const} \]

Degree Distrib.
\[ P(k) \sim k^{-\gamma} \]

\[ P(k) = \delta(k-k_d) \]

\[ P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!} \]

Exponential

\[ C_{\text{rand}} = p = \frac{\langle k \rangle}{N} \]
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

$$q_k = Ck p_k$$

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

$$q_k = \frac{k p_k}{\langle k \rangle}, \quad (5.27)$$
Section 9

Originators of preferential attachments

1953
In *An Informal Theory of the Statistical Structure of Languages* [26] Benoît Mandelbrot proposes optimization as the origin of power laws.

Mandelbrot publishes a comment on Simon’s paper [27] writing:

Simon’s model is analytically circular...

1955
In *On a Class of Skew Distribution Functions* Herbert Simon [6] proposes randomness as the origin of power laws and dismisses Mandelbrot’s claim that power law are rooted in optimization.

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

1959
This present ‘Reply’ refutes the almost entirely new arguments introduced by Dr. Mandelbrot in his “Final Note”...

Herbert

1960
My criticism has not changed since I first had the privilege of commenting upon a draft of Simon.

Herbert

1961
Dr. Mandelbrot has proposed a new set of objections to my 1955 models of Yule distributions. Like earlier objections, these are invalid.

Herbert

1961
Most of Simon’s (1960) reply was irrelevant.

1961
Simon’s subsequent *Reply to ‘Final Note’* by Mandelbrot does not concede [30].

1961
Simon’s final reply ends but does not resolve the debate [32].
György Pólya [1887-1988]
Preferential attachment made its first appearance in 1923 in the celebrated urn model of the Hungarian mathematician György Pólya [2]. Hence, in mathematics preferential attachment is often called a Pólya process.

George Udny Yule [1871-1951]
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 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.

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].

introduce the term preferential attachment in the context of networks [1] to explain the origin of their power-law degree distribution.