Frontiers of Network Science
Fall 2023

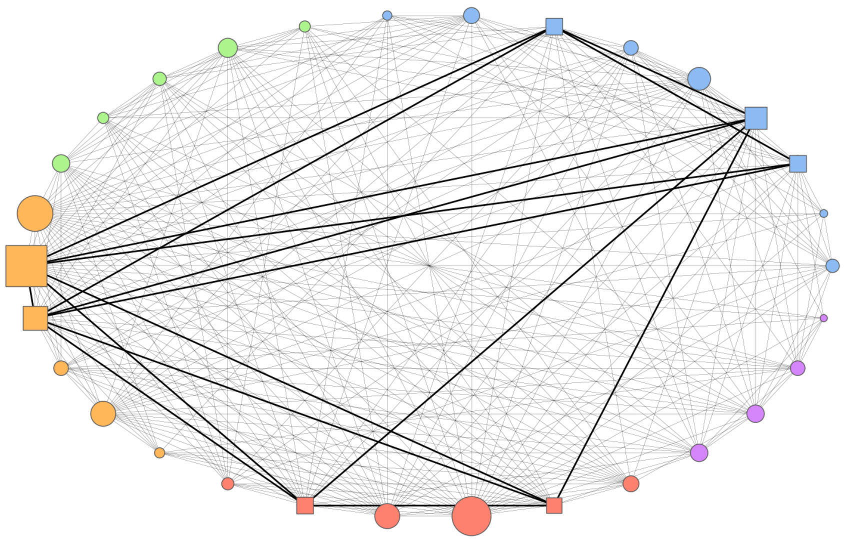
Class 6: Scale Free Networks and Barabasi Model
(Chapters 4-5 in Textbook)

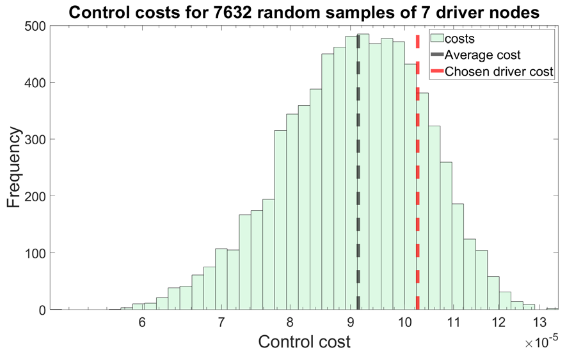
Boleslaw Szymanski

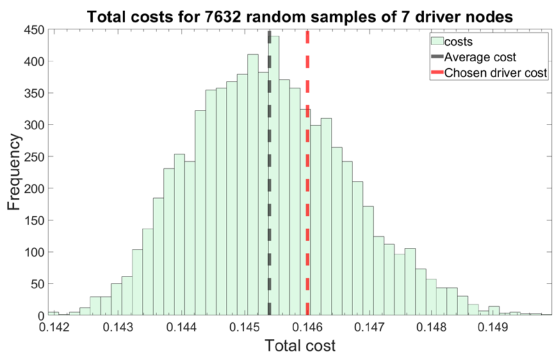
based on slides by Albert-László Barabási & Roberta Sinatra

Example: COVID-19 Control

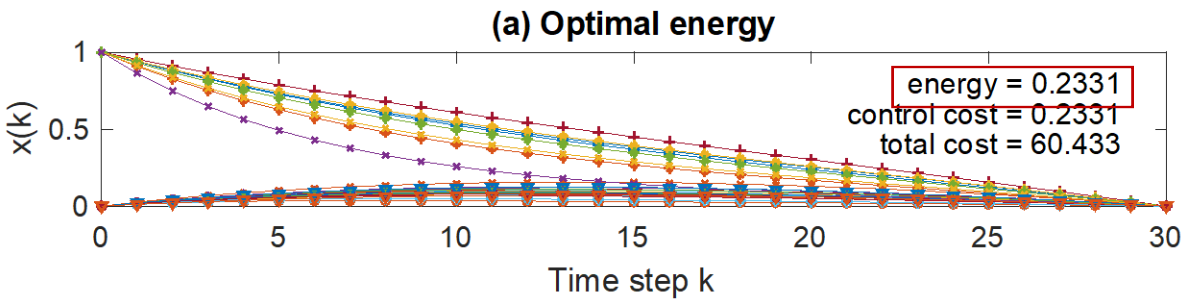
- Formally defined optimal control in the risk networks: $x(k+1) = F[x(k)] + G[x(k), E] + BU$
- Established a function of controllability index and corresponding optimal energy and conditions for nonnegative optimal control
- Provided a universal methodology of applying the LQR control in real world networked systems
- **Qualitative analysis of COVID-19 governmental policies**

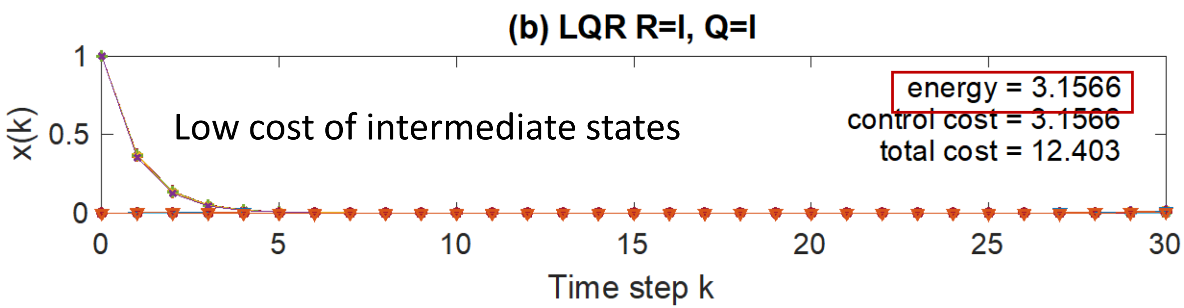






Optimize functions





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:
    - **X. Niu**, C. Jiang, J. Gao, G. Korniss, and B. K. Szymanski. Data-driven control of networked risks with minimal cost. *Nature Communications*, 2019

Questions?

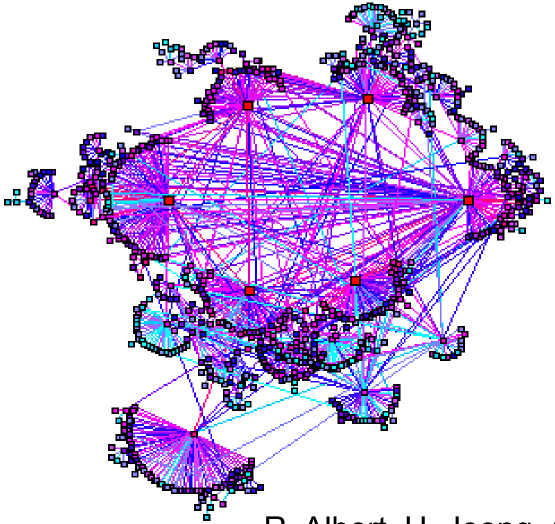
Power laws and scale-free networks

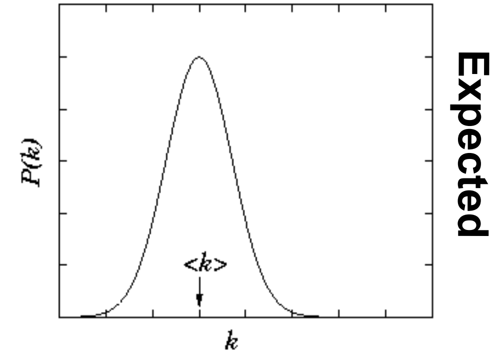
WORLD WIDE WEB

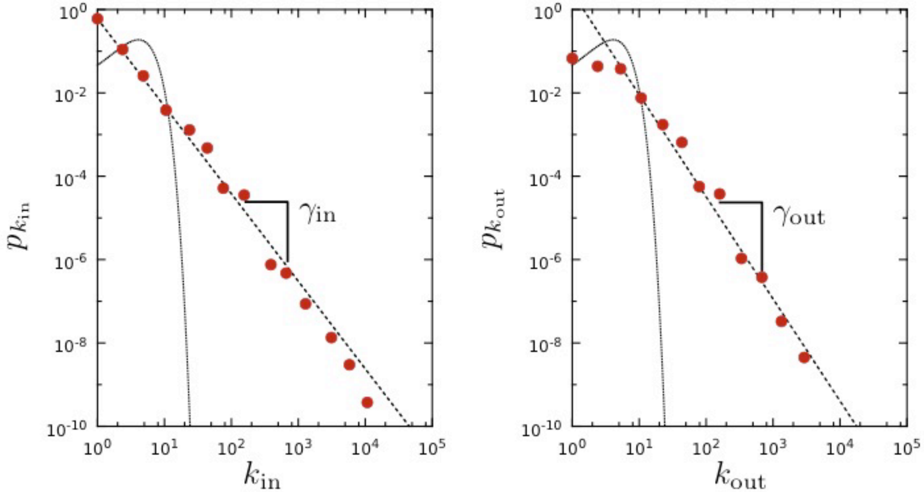
Nodes: **WWW documents**
Links: **URL links**

Over 3 billion documents

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







R. Albert, H. Jeong, A-L Barabasi, *Nature*, 401 130 (1999).

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:

$$p_k = Ck^{-\gamma}.$$

$$\sum_{k=1}^{\infty} p_k = 1.$$

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

$$p_k = \frac{k^{-\gamma}}{\zeta(\gamma)}$$

Continuum Formalism

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

$$p(k) = Ck^{-\gamma}.$$

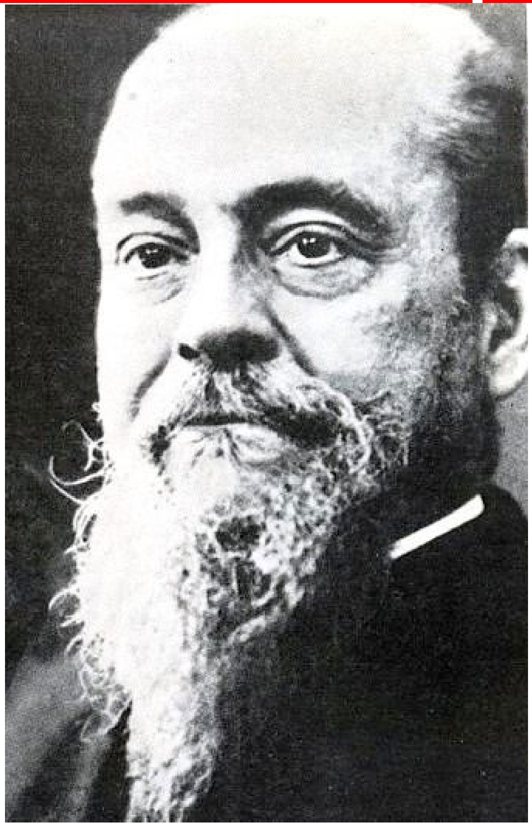
$$\int_{k_{\min}}^{\infty} p(k)dk = 1$$

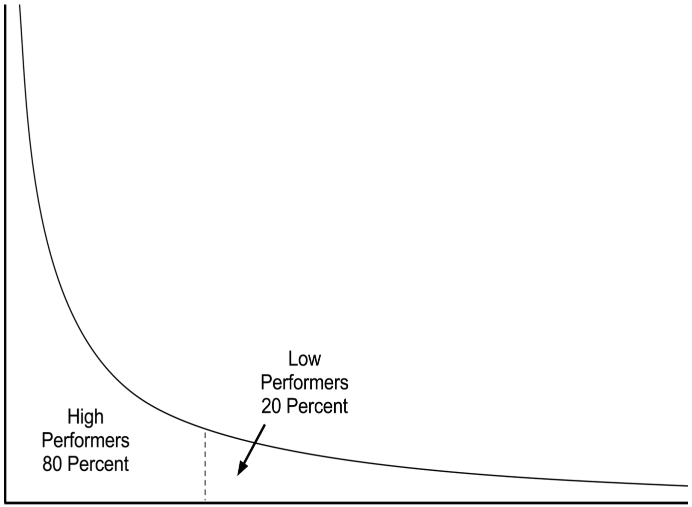
$$C = \frac{1}{\int_{k_{\min}}^{\infty} k^{-\gamma}dk} = (\gamma - 1)k_{\min}^{\gamma-1}$$

$$p(k) = (\gamma - 1)k_{\min}^{\gamma-1}k^{-\gamma}.$$

INTERPRETATION: p_k $\qquad$ $\int_{k_1}^{k_2} p(k)dk$

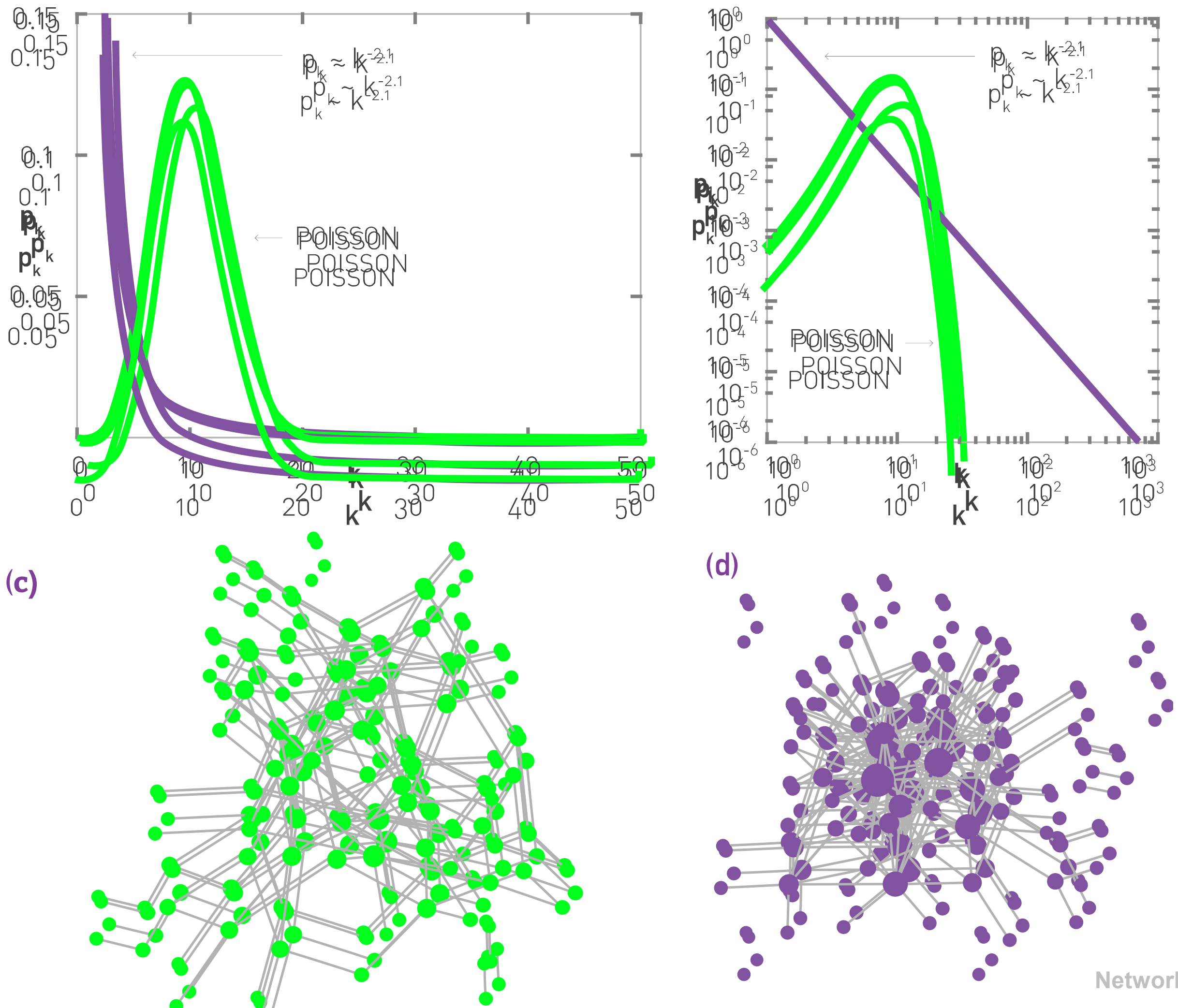
80/20 RULE





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



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~100 is

- *About* $p_{100} \simeq 10^{-30}$ *in a Poisson distribution*
- *About* $p_{100} \simeq 10^{-4}$ *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_{max}

<u>Estimating k_{max}</u>

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

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

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

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

The size of the biggest hub

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

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 $N \simeq 3 \times 10^5$ nodes. As $k_{min} = 1$, if the degree distribution were to follow an exponential, (4.17) predicts that the maximum degree should be $k_{max} \simeq 13$. In a scale-free network of similar size and $\gamma = 2.1$, (4.18) predicts $k_{max} \simeq 85{,}000$, 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_{max} predicted by a scale-free network. This reinforces our conclusion that *in a random network hubs are effectivelly forbidden, while in scale-free networks they are naturally present.*

Finite scale-free networks

Expected maximum degree, k_{max}

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

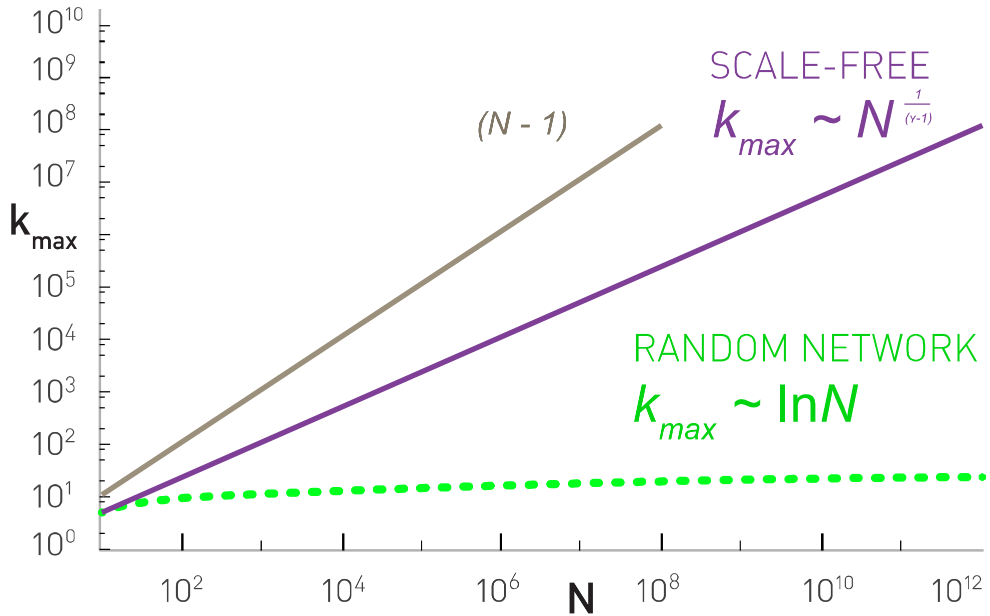
- k_{max}, increases with the size of the network
  → the larger a system is, the larger its biggest hub

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

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

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

The size of the largest hub



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

The meaning of scale-free

Scale-free networks: Definition

Definition:

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

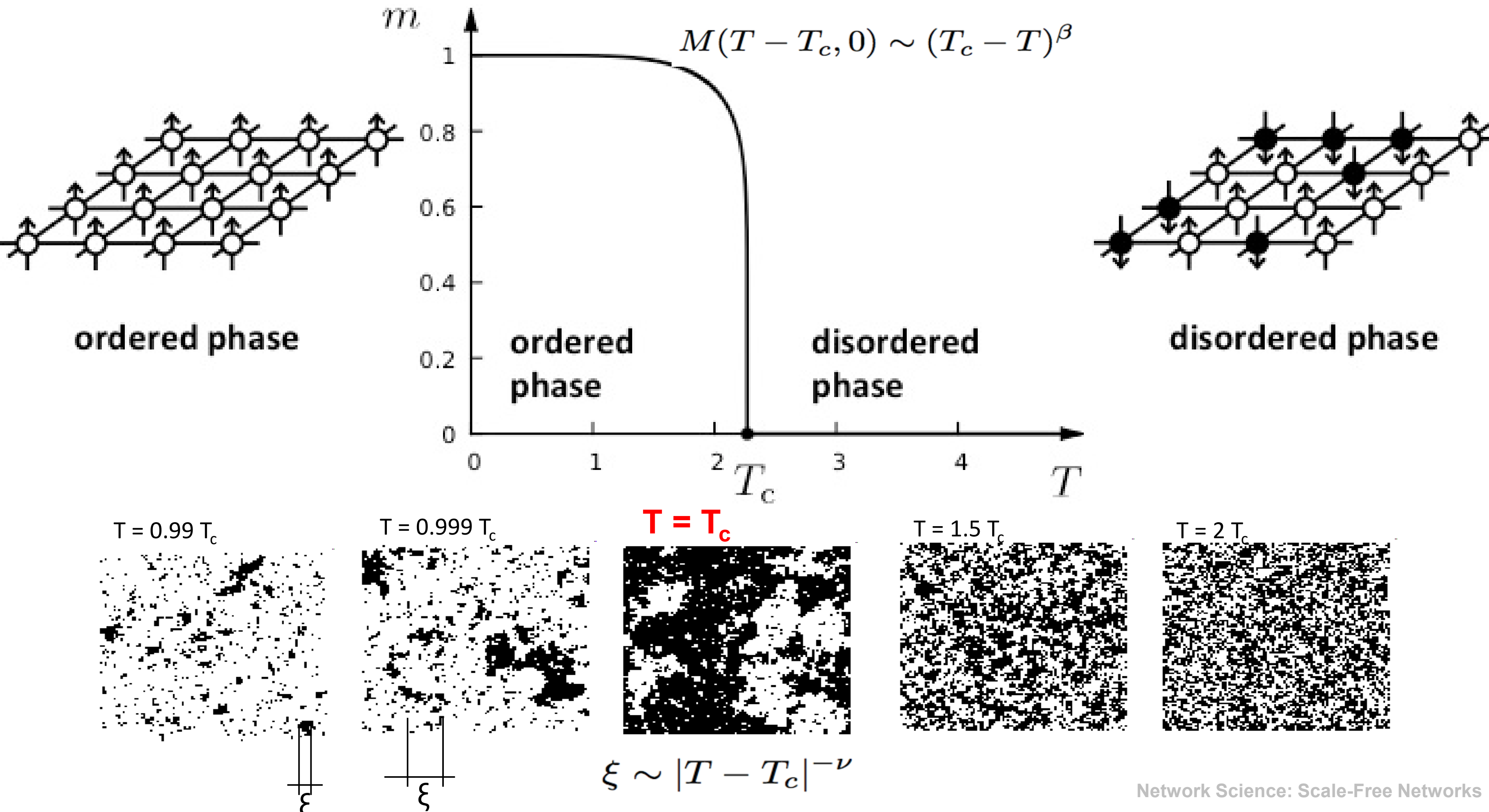
Where does the name come from?

Critical Phenomena and scale-invariance

(a detour)

Slides after **Dante R. Chialvo**

Phase transitions in complex systems I: Magnetism



CRITICAL PHENOMENA

- 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

$$P(k) = Ck^{-\gamma} \qquad k = [k_{\min}, \infty)$$

$$\int_{k_{\min}}^{\infty} P(k)dk = 1$$

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

$$P(k) = (\gamma - 1)k_{\min}^{\gamma-1}k^{-\gamma}$$

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

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

If m-γ+1>0, the integral diverges.

For a fixed γ this means that all moments with m>γ-1 diverge.

DIVERGENCE OF THE HIGHER MOMENTS

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

For a fixed γ this means all moments m>γ-1 diverge.

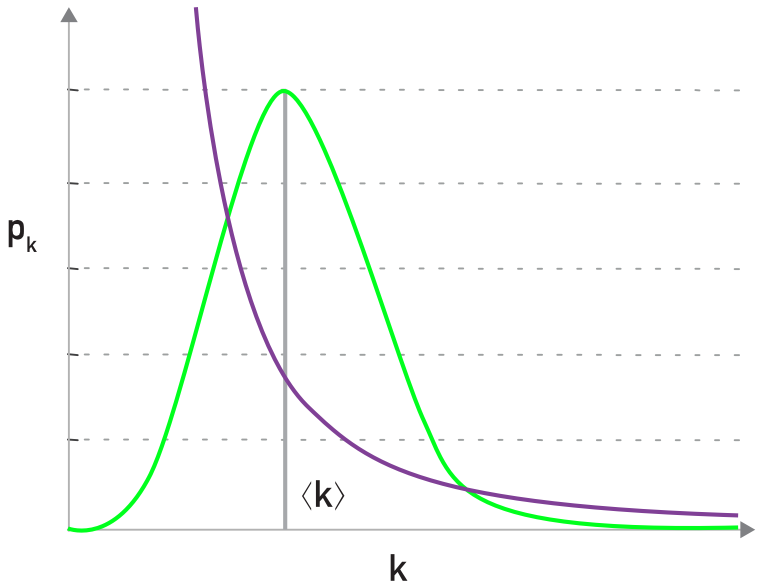
Network	N	L	$\langle k \rangle$	$\langle k_{in}^2 \rangle$	$\langle k_{out}^2 \rangle$	$\langle k^2 \rangle$	γ_{in}	γ_{out}	γ
Internet	192,244	609,066	6.34	-	-	240.1	-	-	3.42*
WWW	325,729	1,497,134	4.60	1546.0	482.4	-	2.00	2.31	-
Power Grid	4,941	6,594	2.67	-	-	10.3	-	-	Exp.
Mobile-Phone Calls	36,595	91,826	2.51	12.0	11.7	-	4.69*	5.01*	-
Email	57,194	103,731	1.81	94.7	1163.9	-	3.43*	2.03*	-
Science Collaboration	23,133	93,437	8.08	-	-	178.2	-	-	3.35*
Actor Network	702,388	29,397,908	83.71	-	-	47,353.7	-	-	2.12*
Citation Network	449,673	4,689,479	10.43	971.5	198.8	-	3.03*	4.00*	-
E. Coli Metabolism	1,039	5,802	5.58	535.7	396.7	-	2.43*	2.90*	-
Protein Interactions	2,018	2,930	2.90	-	-	32.3	-	-	2.89*-

Many degree exponents are smaller than 3

→ $<k^2>$ diverges in the N→∞ limit!!!

→ $<k>$ diverges in the N→∞ limit!!!

The meaning of scale-free



Random Network

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

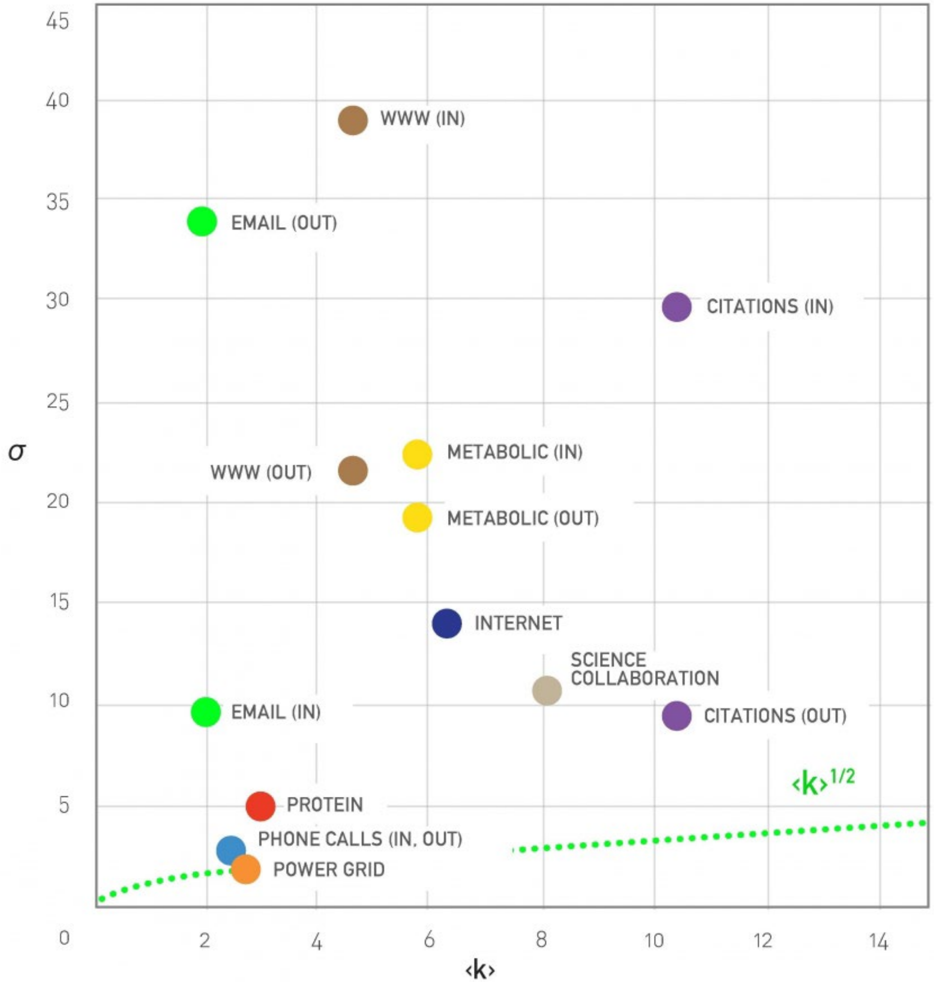
Scale: $\langle k \rangle$

Scale-Free Network

Randomly chosen node: $k = \langle k \rangle \pm \infty$

Scale: none

The meaning of scale-free

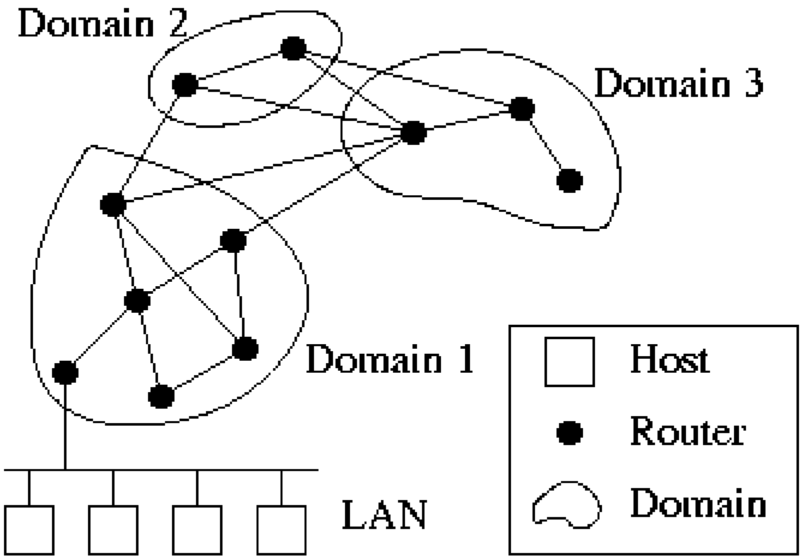


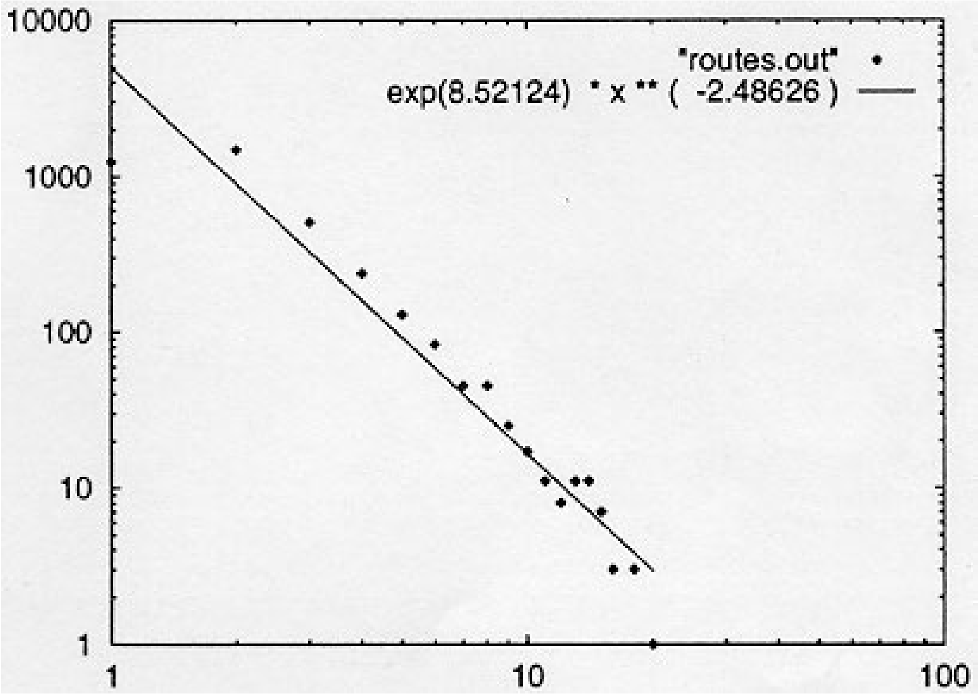
$$k = \langle k \rangle \pm \sigma_k$$

universality

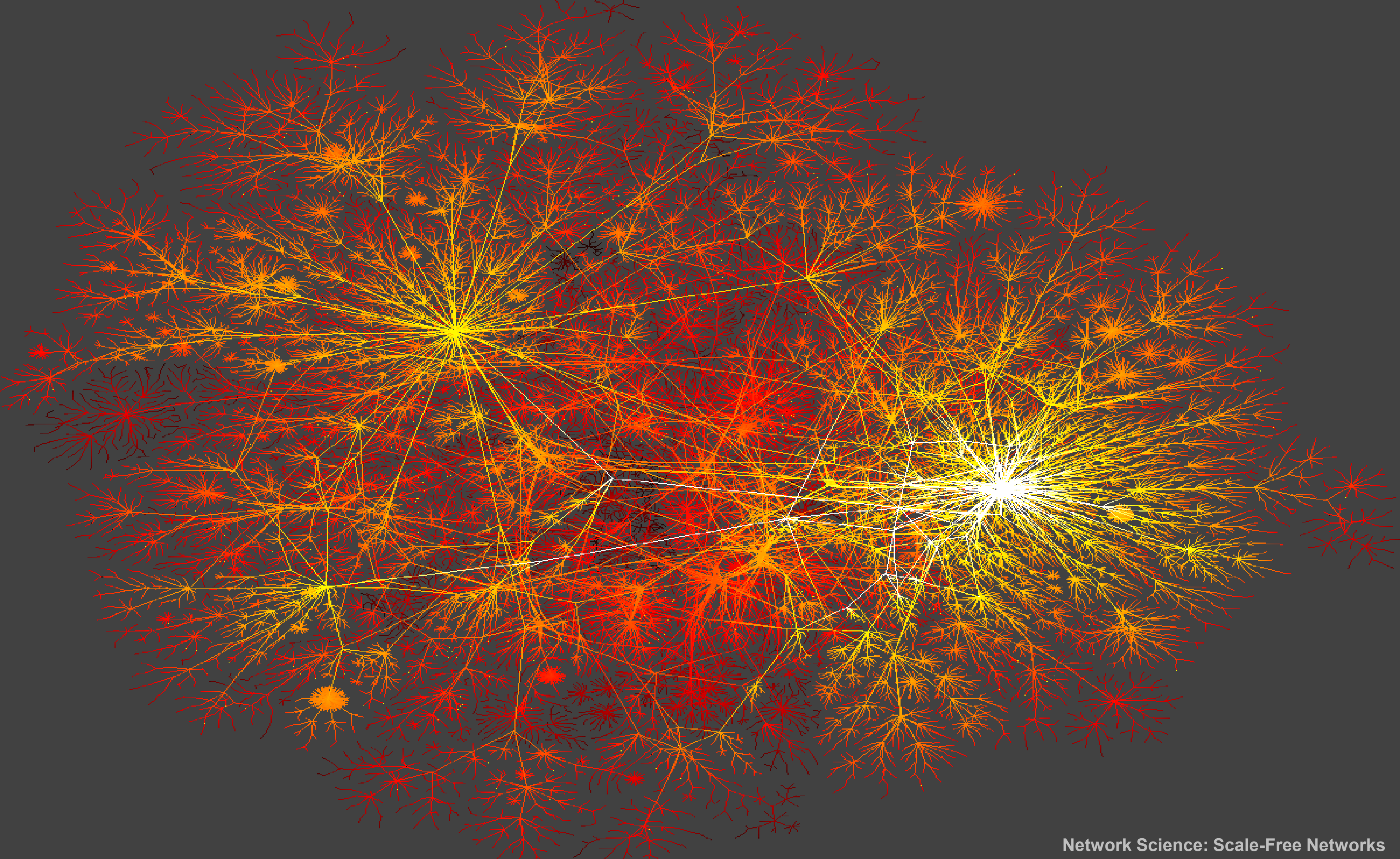
INTERNET BACKBONE

Nodes: computers, routers
Links: physical lines





(Faloutsos, Faloutsos and Faloutsos, 1999)

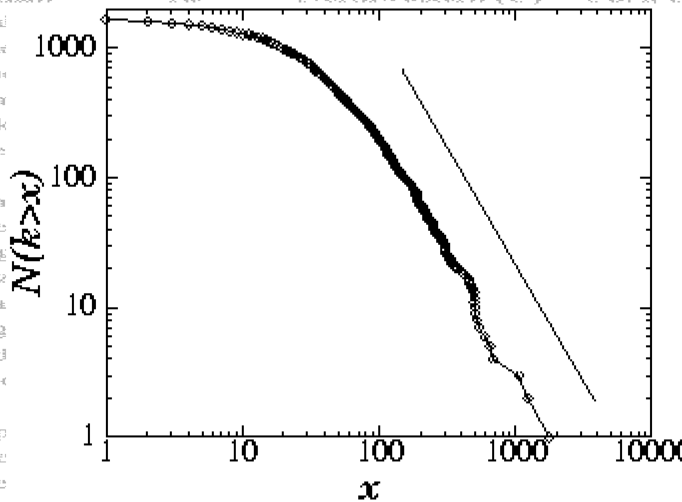


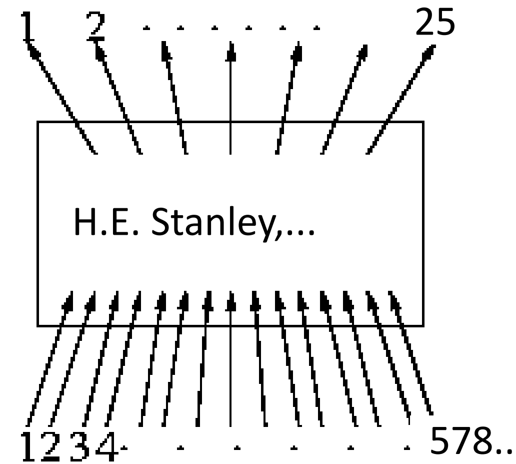
SCIENCE CITATION INDEX

Out of over 500,000 Examined
(see http://www.ssi.nrel.gov)

Nodes: papers
Links: citations

1736 PRL papers (1988)





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

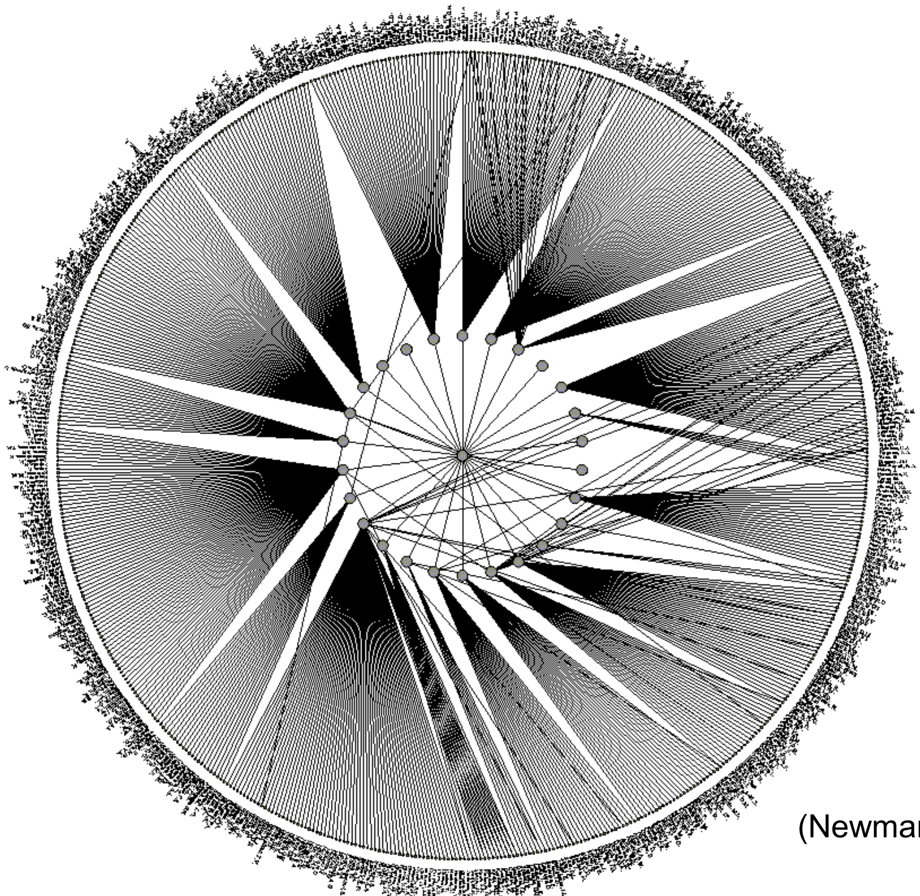
$(\gamma = 3)$

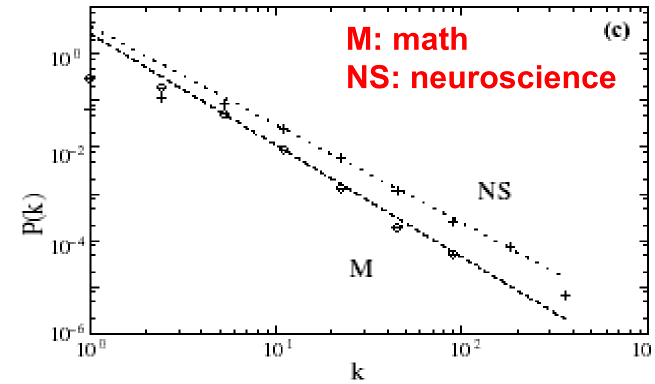
(S. Redner, 1998)

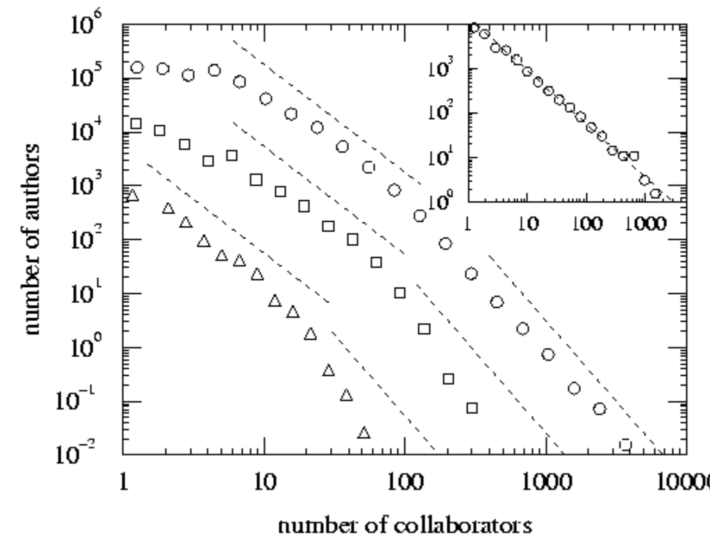
* citation total may be skewed because of multiple authors with the same name

SCIENCE COAUTHORSHIP

Nodes: scientist (authors)
Links: joint publication





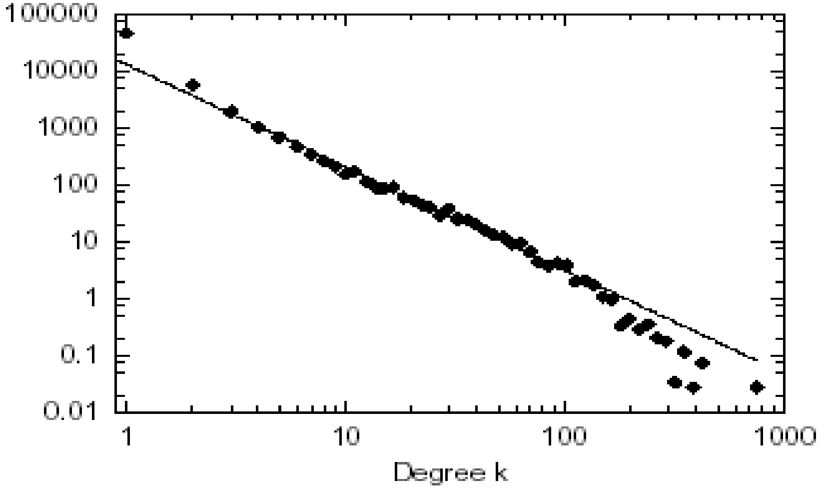


(Newman, 2000, Barabasi et al 2001)

ONLINE COMMUNITIES

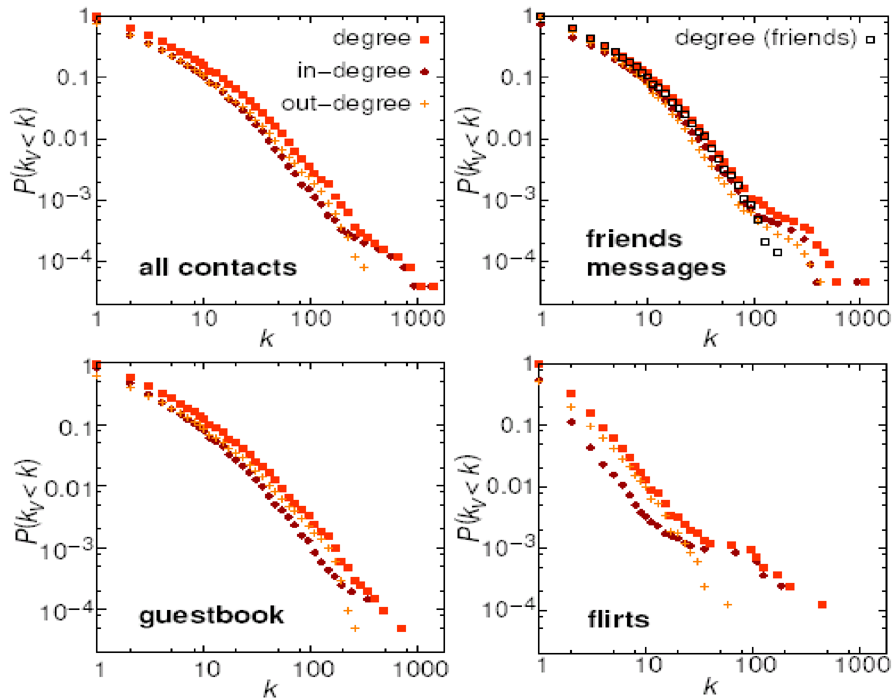
Nodes: online user
Links: email contact

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



Ebel, Mielsch, Bornholdtz, PRE 2002.

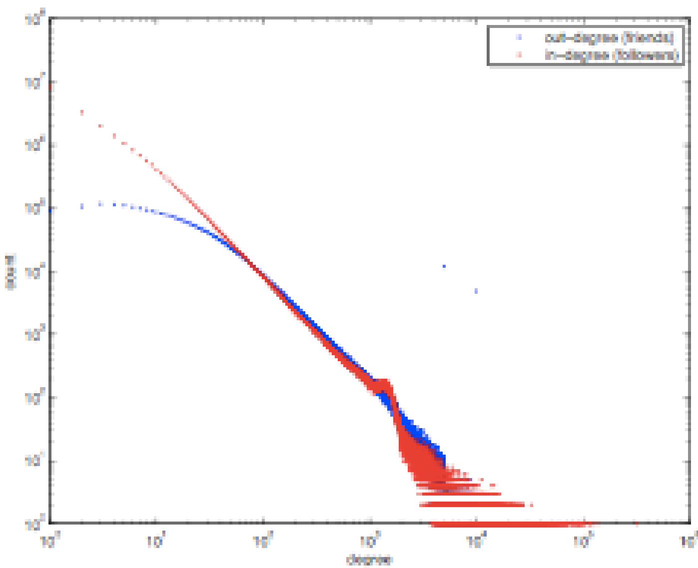
Pussokram.com online community;
512 days, 25,000 users.



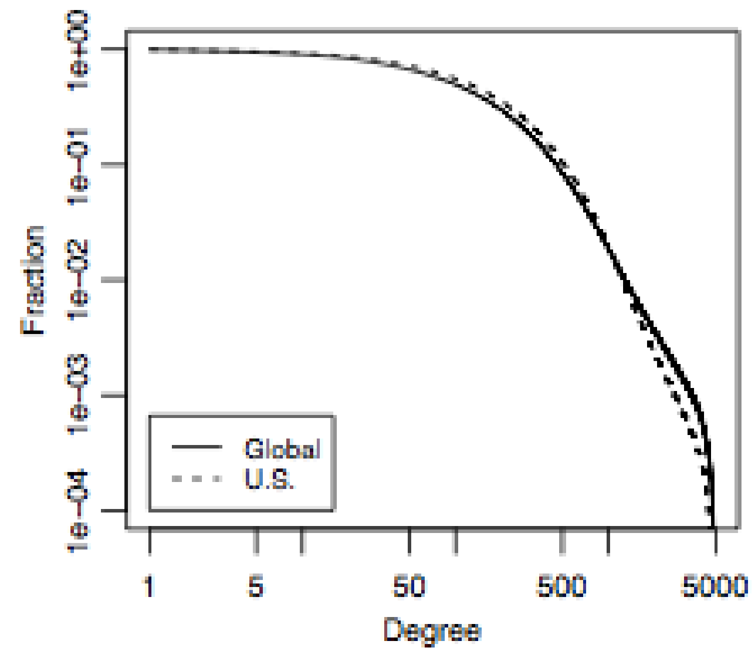
Holme, Edling, Liljeros, 2002.

ONLINE COMMUNITIES

Twitter:



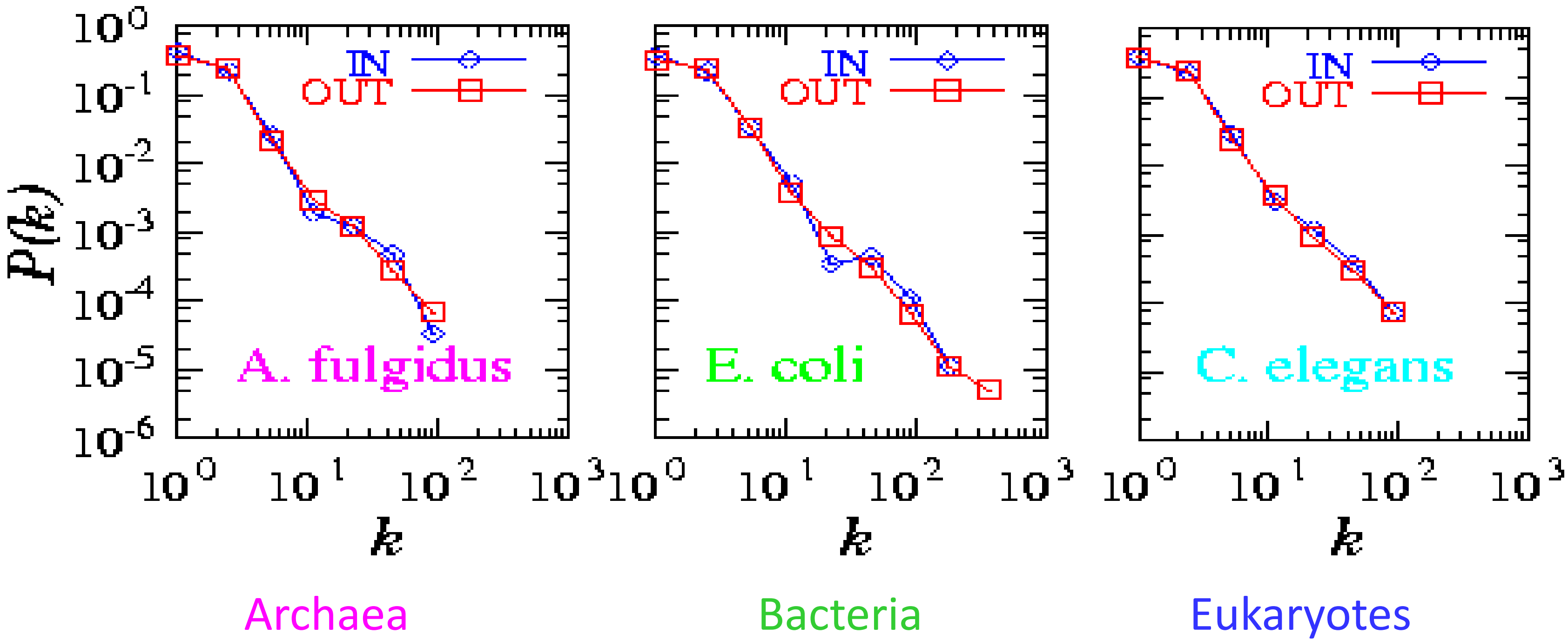
Facebook



Brian Karrer, Lars Backstrom, Cameron Marlowm 2011

Barabasi-Albert Model

METABOLIC NETWORK



Organisms from all three domains of life are scale-free!

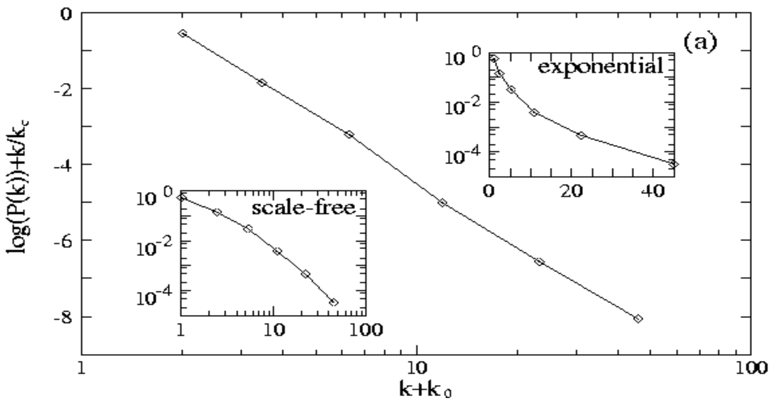
$$P_{in}(k) \approx k^{-2.2}$$

$$P_{out}(k) \approx k^{-2.2}$$

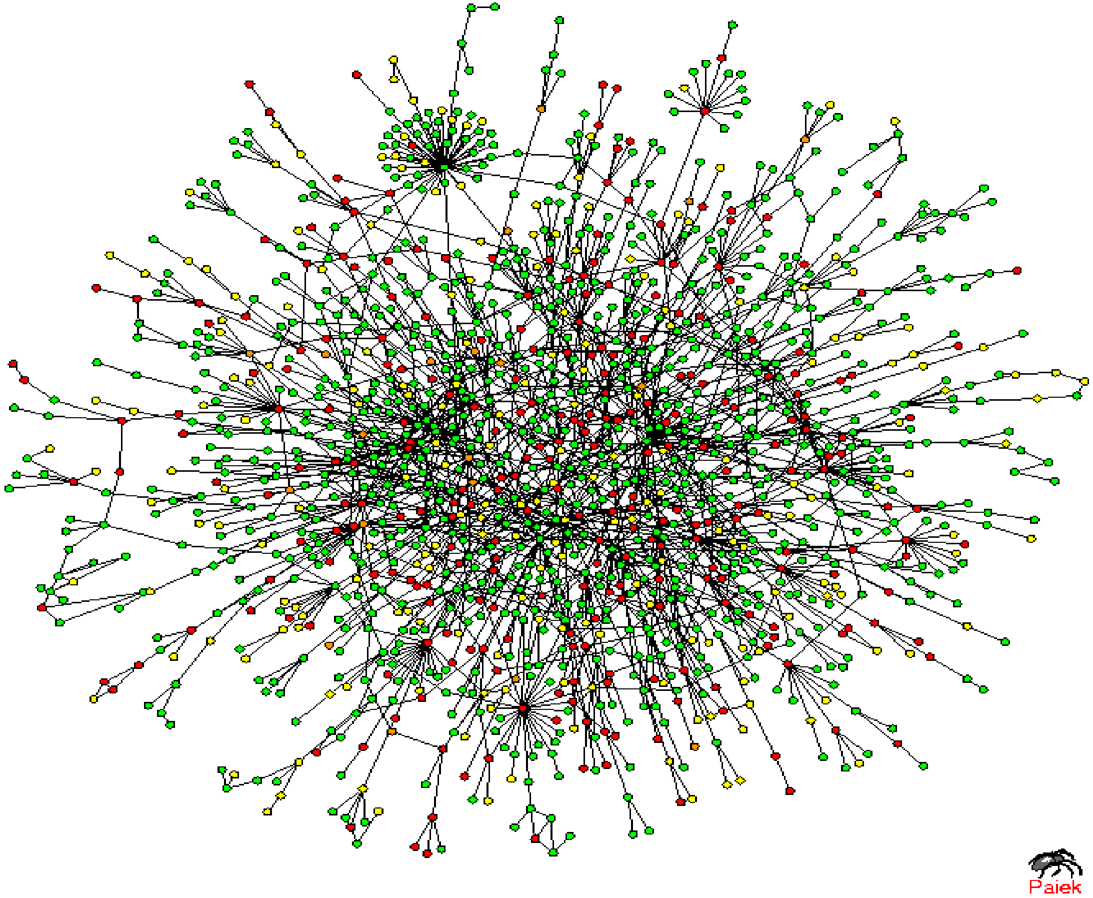
H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, and A.L. Barabasi, *Nature*, 407 651 (2000)

TOPOLOGY OF THE PROTEIN NETWORK

Nodes: proteins
Links: physical interactions-binding

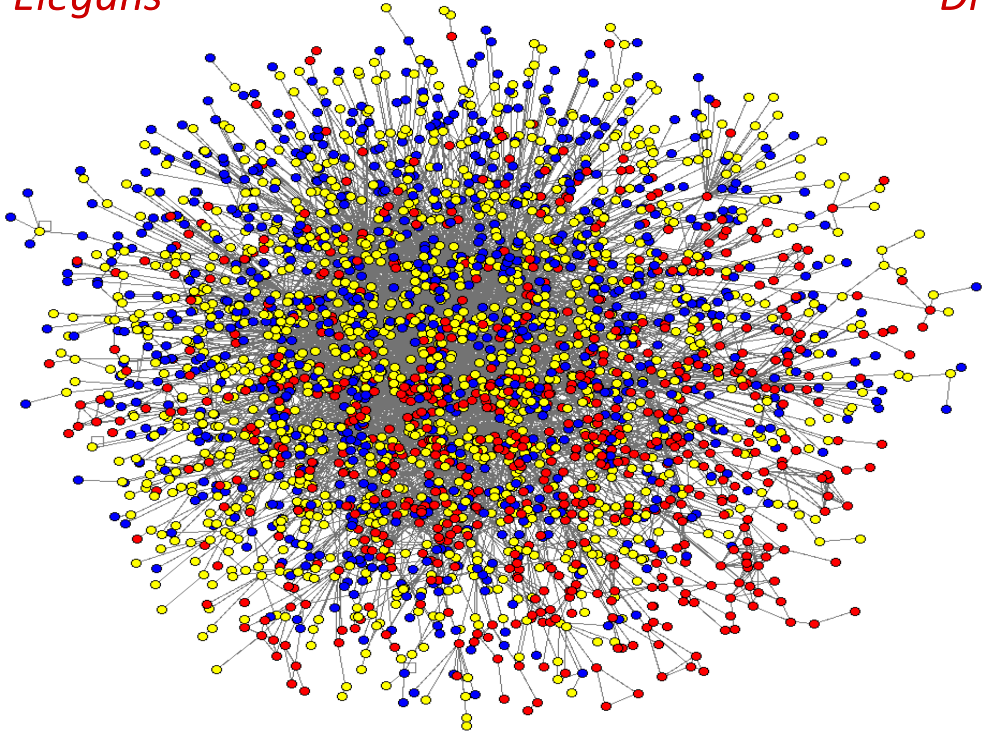


$$P(k) \sim (k + k_0)^{-\gamma} \exp(-\frac{k + k_0}{k_\tau})$$

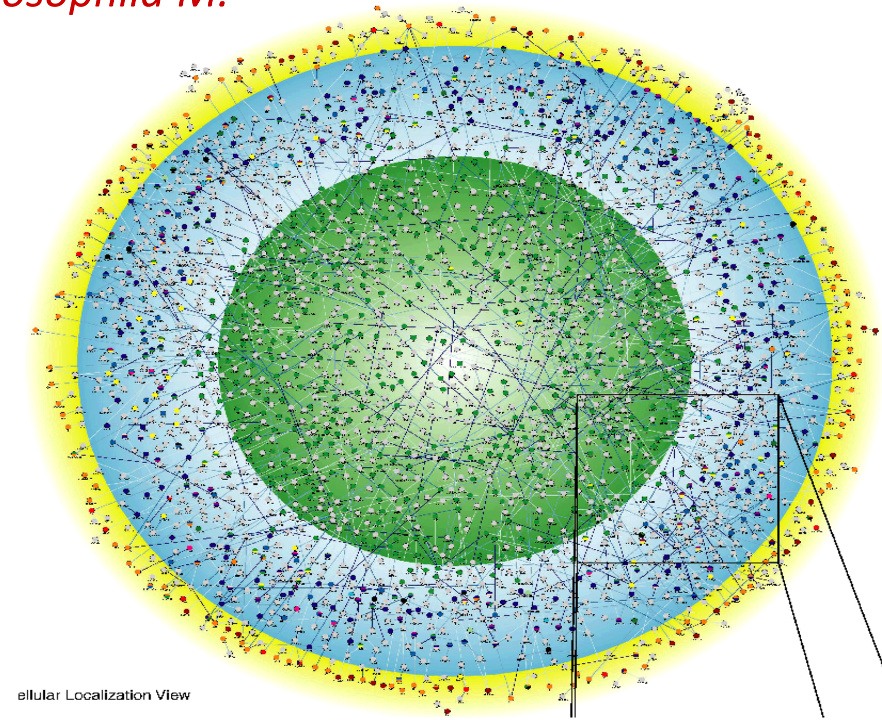


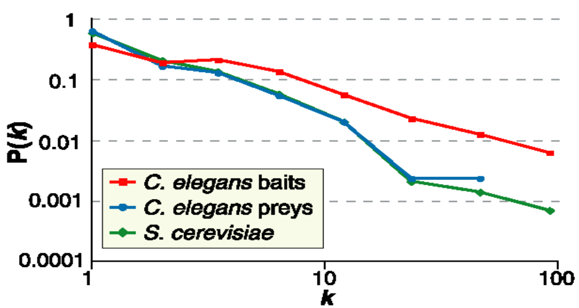
H. Jeong, S.P. Mason, A.-L. Barabasi, Z.N. Oltvai, Nature 411, 41-42 (2001)

C. Elegans

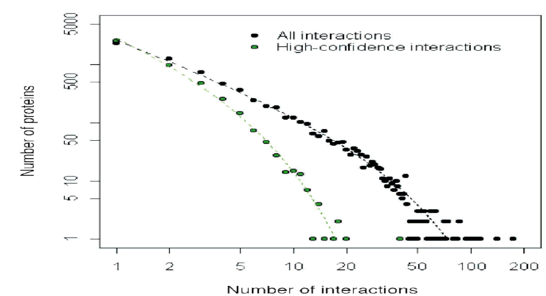


Drosophila M.





Li *et al.* Science 2004



Giot *et al.* Science 2003

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:** **t**he probability $\Pi(k)$ that a link of the new node connects to node i depends on the degree k_i as $\Pi(k_i)=k_i\sum_j k_j$

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.

[1] A.-L.Barabási, R. Albert and H. Jeong, *Physica* A **272,** 173 (1999)

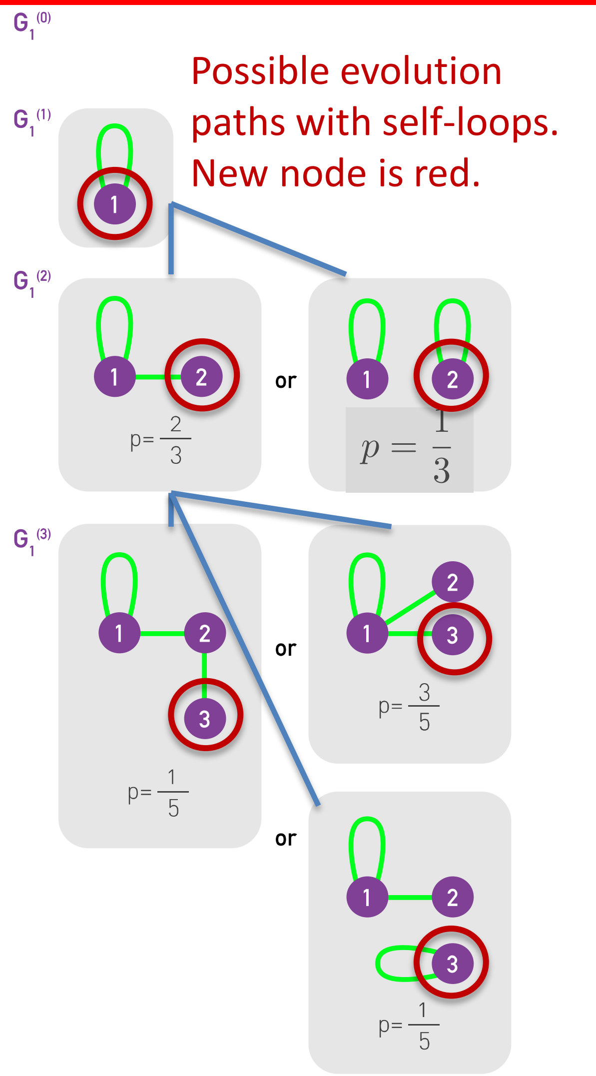
Linearized Chord Diagram

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

$$p(i=s)=\begin{cases} \dfrac{k_i}{2t-1} & if\ 1\le s\le t-1 \\ \dfrac{1}{2t-1}, & if\ s=t \end{cases}$$



Degree dynamics

Degree distribution for Barabasi-Albert model

$$k_i(t) = m\left(\frac{t}{t_i}\right)^{\beta} \qquad \beta = \frac{1}{2}$$ for $t \geq m_0+i$ and 0 otherwise as system size at t is $N=m_0+t-1$

We assume the initial m_0 nodes create a fully connected graph.

A random node j arriving at time t is with equal probability $1/N=1/(m_0+t-1)$ one of the nodes $1, 2, \ldots N$, its degree will grow with the above equation, so

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

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

$$\therefore P(k) = \frac{\partial P(k_i(t) < k)}{\partial k} = \frac{2m^2 t}{m_o + t} \frac{1}{k^3} \sim k^{-\gamma}$$

$$\gamma = 3$$

A.-L.Barabási, R. Albert and H. Jeong, *Physica* A **272,** 173 (1999)

Degree distribution

$$k_i(t) = m\left(\frac{t}{t_i}\right)^{\beta} \qquad \beta = \frac{1}{2}$$

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

$$\gamma = 3$$

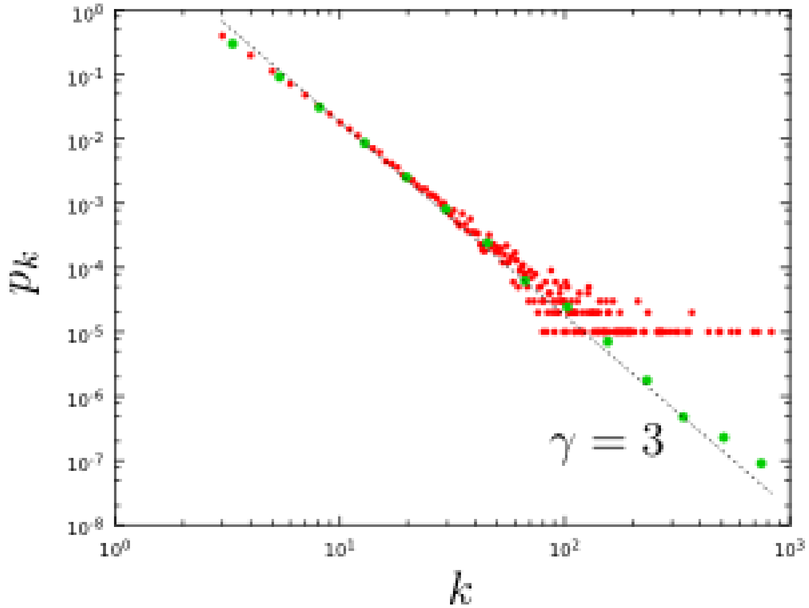
(i) The degree exponent is independent of m.

(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)
→ the network reaches a stationary scale-free state.

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

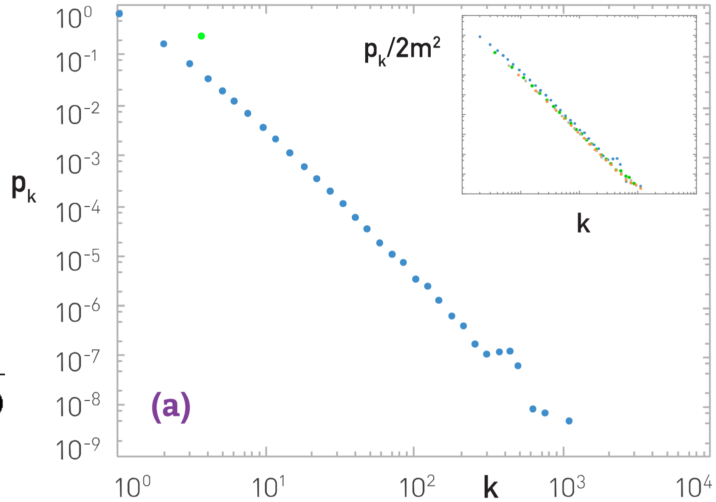
A.-L.Barabási, R. Albert and H. Jeong, *Physica* A **272,** 173 (1999)

Section 4

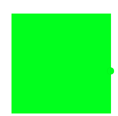


NUMERICAL SIMULATION OF THE BA MODEL

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



(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 γ is independent of m and m_0. The slope of the purple line is -3, corresponding to the predicted degree exponent γ=3. Inset: (5.11) predicts p_k~$2m^2$, 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)

(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

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

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

m=1,3,5,7

N=100,000;150,000;200,000

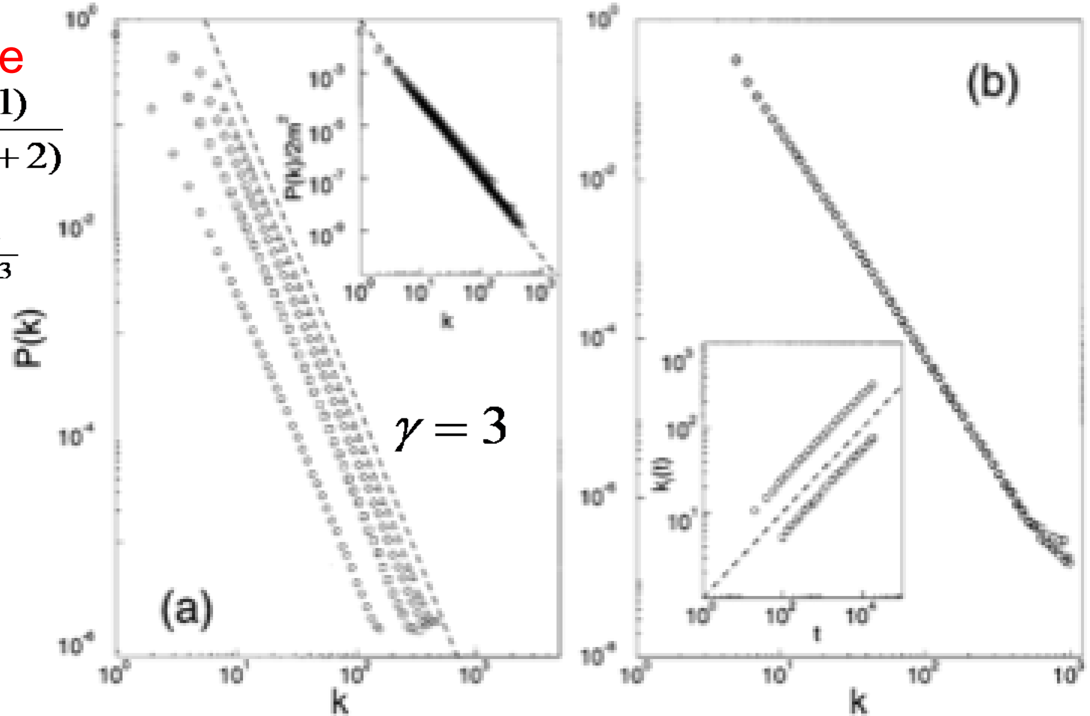


FIG. 21. Numerical simulations of network evolution: (a) Degree distribution of the Barabási-Albert model, with $N=m_0+t=300\,000$ and ○, $m_0=m=1$; □, $m_0=m=3$; ◇, $m_0=m=5$; and △, $m_0=m=7$. The slope of the dashed line is $\gamma=2.9$, providing the best fit to the data. The inset shows the rescaled distribution (see text) $P(k)/2m^2$ for the same values of m, the slope of the dashed line being $\gamma=3$; (b) $P(k)$ for $m_0=m=5$ and various system sizes, ○, $N=100\,000$; □, $N=150\,000$; ◇, $N=200\,000$. The inset shows the time evolution for the degree of two vertices, added to the system at $t_1=5$ and $t_2=95$. Here $m_0=m=5$, and the dashed line has slope 0.5, as predicted by Eq. (81). After Barabási, Albert, and Jeong (1999).

Stationarity:
P(k) independent of *N*

Insert:
degree dynamics

$$k_i(t) = m\left(\frac{t}{t_i}\right)^{\beta} \qquad \beta = \frac{1}{2}$$

The mean field theory offers the correct scaling, BUT it provides the wrong coefficient of the degree distribution.

So assymptotically it is correct ($k \rightarrow \infty$), 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:

$$\frac{k}{2mt} \times NP(k,t) \times m = \frac{k}{2}P(k,t)$$

Total number of k-nodes

Preferential attachment

New node adds m new links to other nodes

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→ k

Loss of k-nodes via k→ k+1

A.-L.Barabási, R. Albert and H. Jeong, *Physica* A **272**, 173 (1999)

MFT - Degree Distribution: Rate Equation

$$\underbrace{(N+1)P(k,t+1)}_{\text{\# k-nodes at time t+1}} = \underbrace{NP(k,t)}_{\text{\# k-nodes at time t}} + \underbrace{\frac{k-1}{2}P(k-1,t)}_{\text{Gain of k-nodes via k-1} \rightarrow \text{k}} - \underbrace{\frac{k}{2}P(k,t)}_{\text{Loss of k-nodes via k} \rightarrow \text{k+1}}$$

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

$$\underbrace{(N+1)P(m,t+1)}_{\text{\# m-nodes at time t+1}} = \underbrace{NP(m,t)}_{\text{\# m-nodes at time t}} + \underbrace{1}_{\text{Add one m-degeree node}} - \underbrace{\frac{m}{2}P(m,t)}_{\text{Loss of an m-node via m} \rightarrow \text{m+1}}$$

A.-L.Barabási, R. Albert and H. Jeong, *Physica* A **272,** 173 (1999)

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>m

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

We assume that there is a stationary state in the N=t→∞ limit, when P(k,∞)=P(k)

$$(N+1)P(k,t+1) - NP(k,t) \rightarrow NP(k,\infty) + P(k,\infty) - NP(k,\infty) = P(k,\infty) = P(k)$$

$$(N+1)P(m,t+1) - NP(m,t) \rightarrow P(m)$$

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

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

$$P(k) = \frac{k-1}{k+2}P(k-1)$$ k>m

$$P(m) = \frac{2}{2+m}$$

MFT - Degree Distribution: Rate Equation

$$P(k) = \frac{k-1}{k+2} P(k-1) \quad \rightarrow \quad P(k+1) = \frac{k}{k+2} P(k)$$

$$P(m) = \frac{2}{m+2}$$

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

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

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

$m+3 \rightarrow k$

...

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

$$P(k) \sim k^{-3}$$

for large k

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

MFT - Degree Distribution: A Pretty Caveat

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

$$\frac{\partial k_i}{\partial t} \propto \Pi(k_i) = A\frac{k_i}{\sum_j k_j}$$

In limit: $A\frac{k_i}{\sum_j k_j} = A\frac{k_i}{2mt}$ So:

$$m = \sum_i \frac{\Delta k_i}{dt} = \sum_i A\frac{k_i}{2mt} = A$$

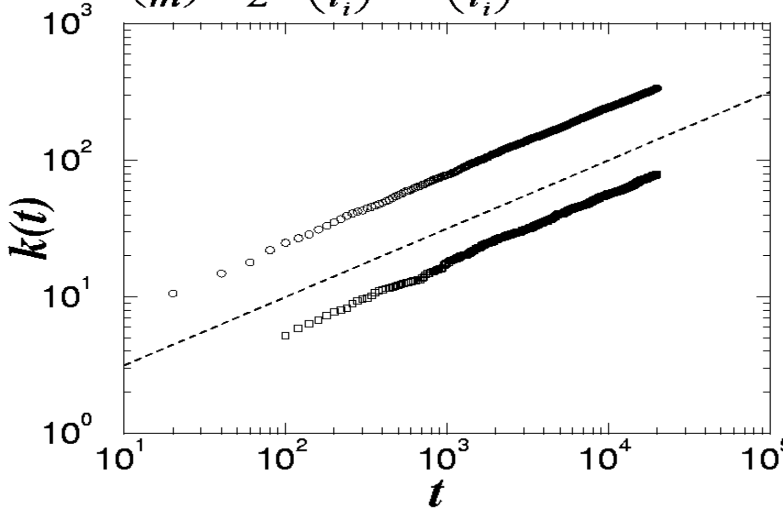
Use: $\sum_j k_j = 2m(t-1) + \frac{m_0(m_0-1)}{2}$

During a unit time (time step): *Δk=m* → *A=m*

$$\frac{\partial k_i}{\partial t} = \frac{k_i}{2t} \qquad \frac{\partial k_i}{k_i} = \frac{\partial t}{2t} \qquad \int_m^k \frac{\partial k_i}{k_i} = \int_{t_i}^t \frac{\partial t}{2t}$$

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

$$k_i(t) = m\left(\frac{t}{t_i}\right)^{\beta} \qquad \beta = \frac{1}{2}$$



β: dynamical exponent

A.-L.Barabási, R. Albert and H. Jeong, *Physica* A **272,** 173 (1999)

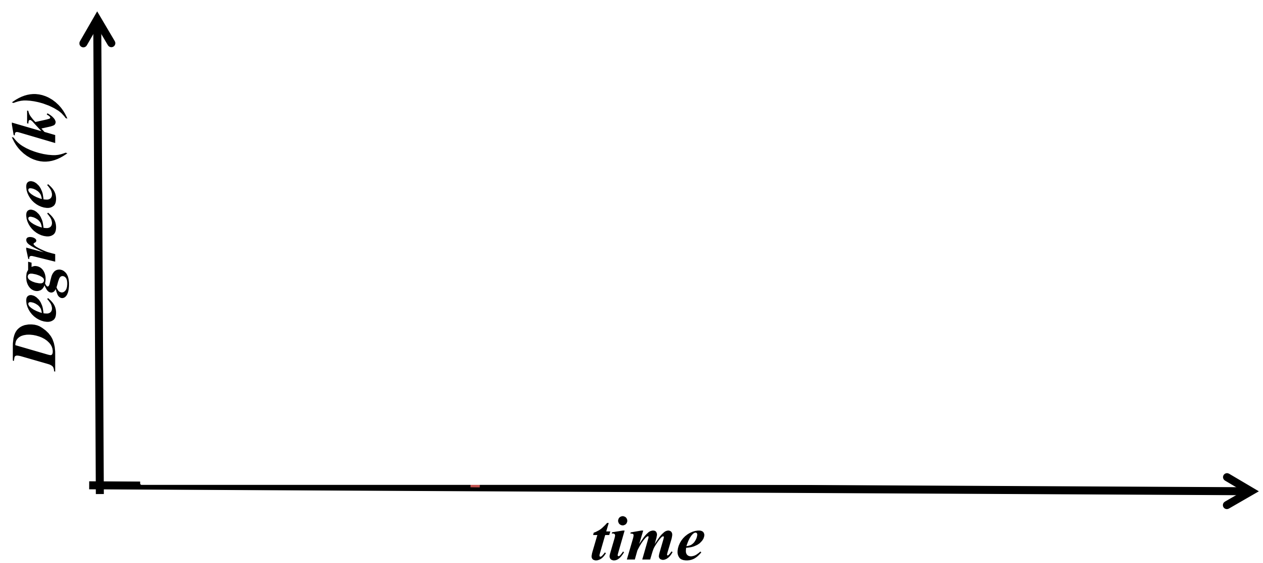
Fitness Model

Fitness Model: Can Latecomers Make It?

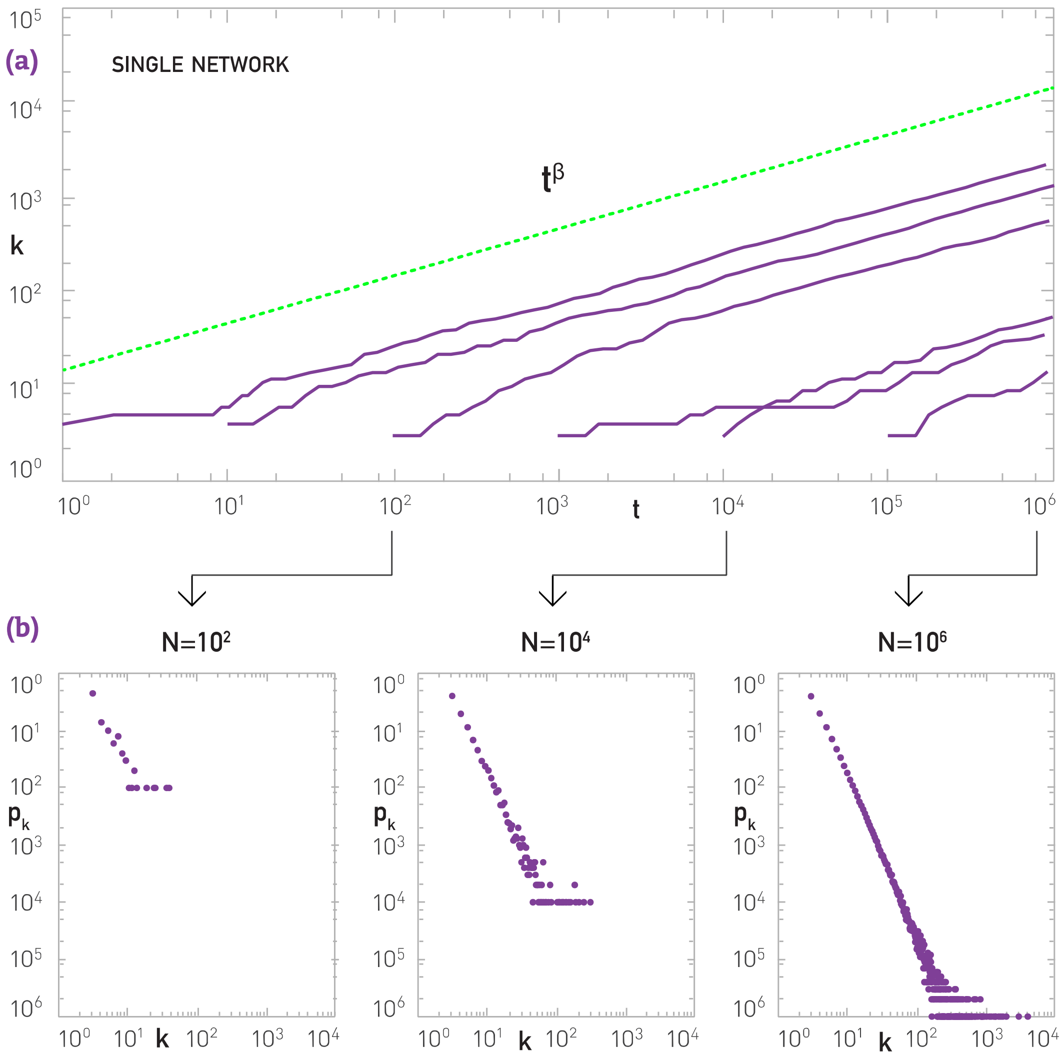
SF model: $k(t) \sim t^{1/2}$ (first mover advantage)

Fitness model: fitness (η) $\Pi(k_i) \cong \frac{\eta_i k_i}{\sum_j \eta_j k_j}$ $k(\eta,t) \sim t^{\beta(\eta)}$

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



Bianconi & Barabási, *Physical Review Letters* 2001; *Europhys. Lett.* 2001.



- The degree of each node increases following a power-law with the same dynamical exponent β =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)

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

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

MODEL A

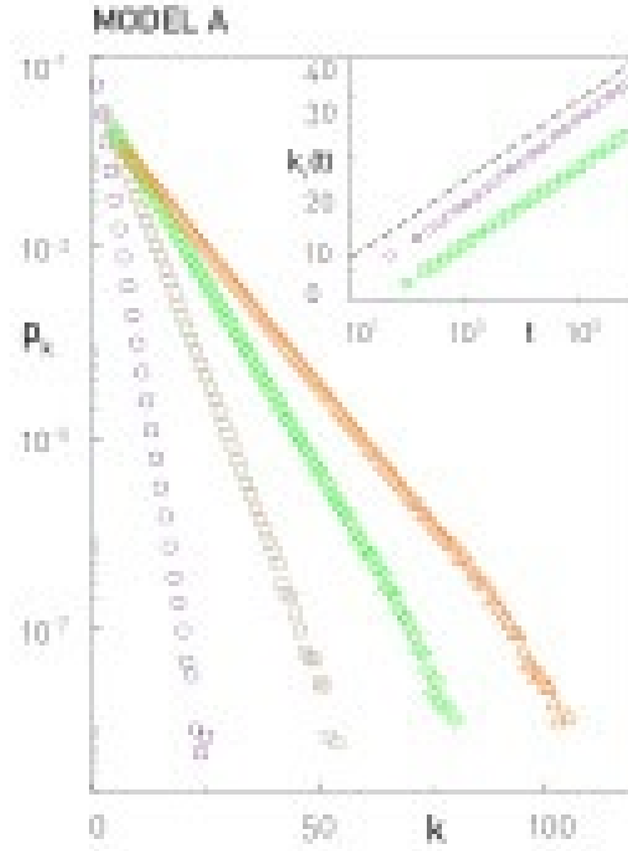
growth ~~preferential attachment~~

$\Pi(k_i)$: uniform

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

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

$$P(k) = \frac{e}{m}\exp\left(-\frac{k}{m}\right) \sim e^{-k}$$



MODEL B

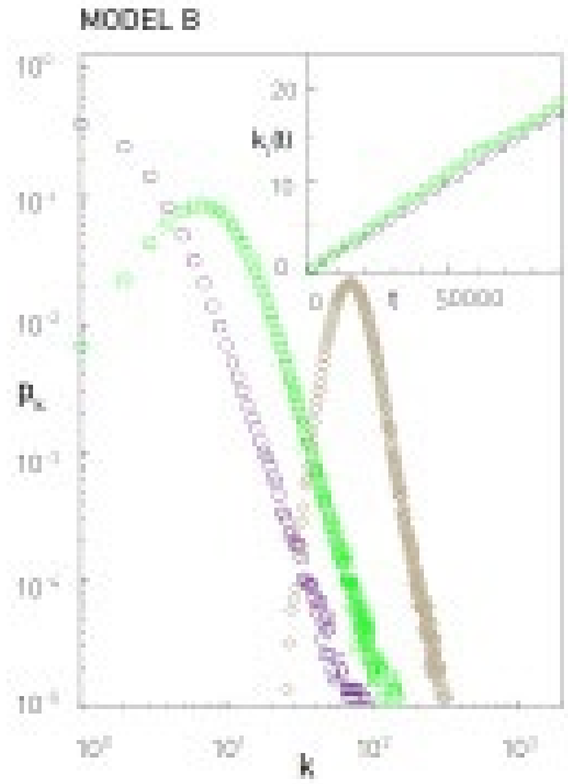
~~growth~~ preferential attachment

$$\frac{\partial k_i}{\partial t} = A\Pi(k_i) + \frac{1}{N} = \frac{N}{N-1}\frac{k_i}{2t} + \frac{1}{N}$$

$$k_i(t) = \frac{2(N-1)}{N(N-2)}t + Ct^{\frac{N}{2(N-1)}} \sim \frac{2}{N}t$$

p_k : power law (initially) →

→ Gaussian → Fully Connected

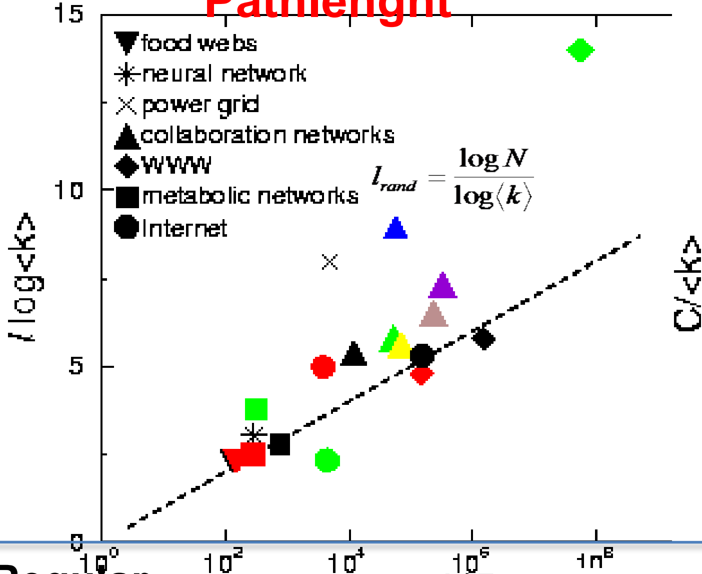


Do we need both growth and preferential attachment?

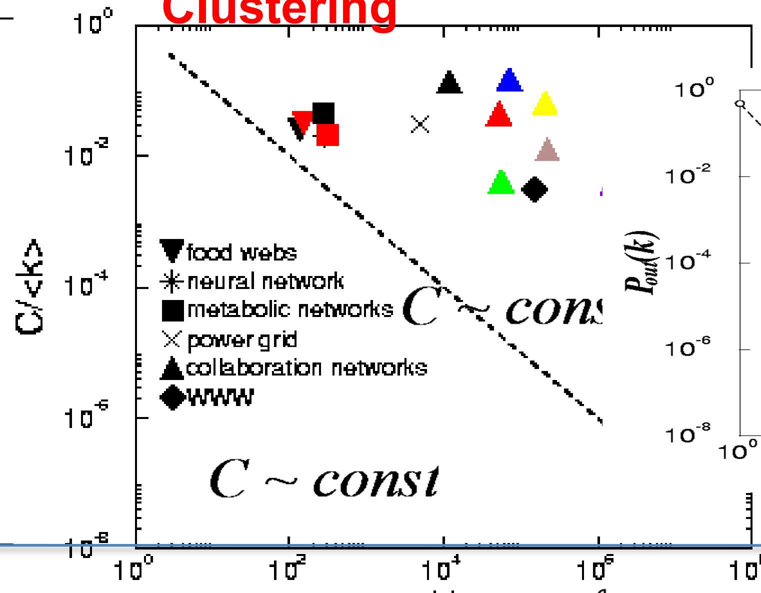
YEP

EMPIRICAL DATA FOR REAL NETWORKS

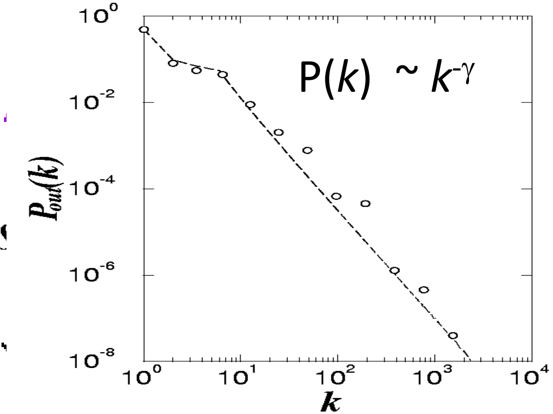
Pathlenght



Clustering



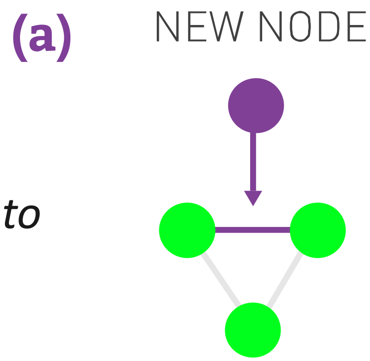
Degree Distr.



	Pathlength		Clustering		Degree Distr.	
Regular network	$l \approx N^{1/D}$	👎		👍	$P(k)=\delta(k-k_d)$	👎
Erdos-Renyi	$l_{rand} \approx \frac{\log N}{\log\langle k\rangle}$	👍	$C_{rand} = p = \frac{\langle k\rangle}{N}$	👎	$P(k) = e^{-\langle k\rangle}\frac{\langle k\rangle^k}{k!}$	👎
Watts-Strogatz	$l_{rand} \approx \frac{\log N}{\log\langle k\rangle}$	👍	$C \sim const$	👍	*Exponential*	👎
Barabasi-Albert					$P(k) \sim k^{-\gamma}$	👍

The origins of preferential attachment

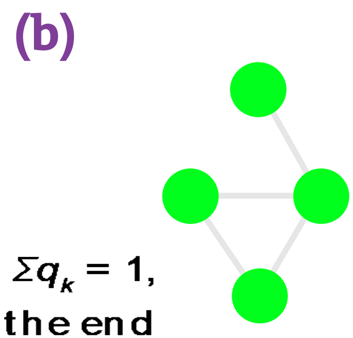
Link selection model -- perhaps the simplest example of a local or random mechanism capable of generating preferential attachment.



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



$$q_k = Ckp_k$$

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

$$q_k = \frac{kp_k}{\langle k\rangle}, \qquad (5.27)$$

Originators of preferential attachments

1

1953

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

2

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.

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

Benoit

Simon's model is analytically circular...

1959

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

Herbert

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

1960

5 In a 19 page response entitled **Final Note**, Mandelbrot states [29]:

Benoit

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

1961

6 Simon's subsequent **Reply to 'Final Note' by Mandelbrot** does not concede [30]

Herbert

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

1961

7 In the creatively titled **Post Scriptum to Final Note** Mandlebrot [31] writes

Benoit

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

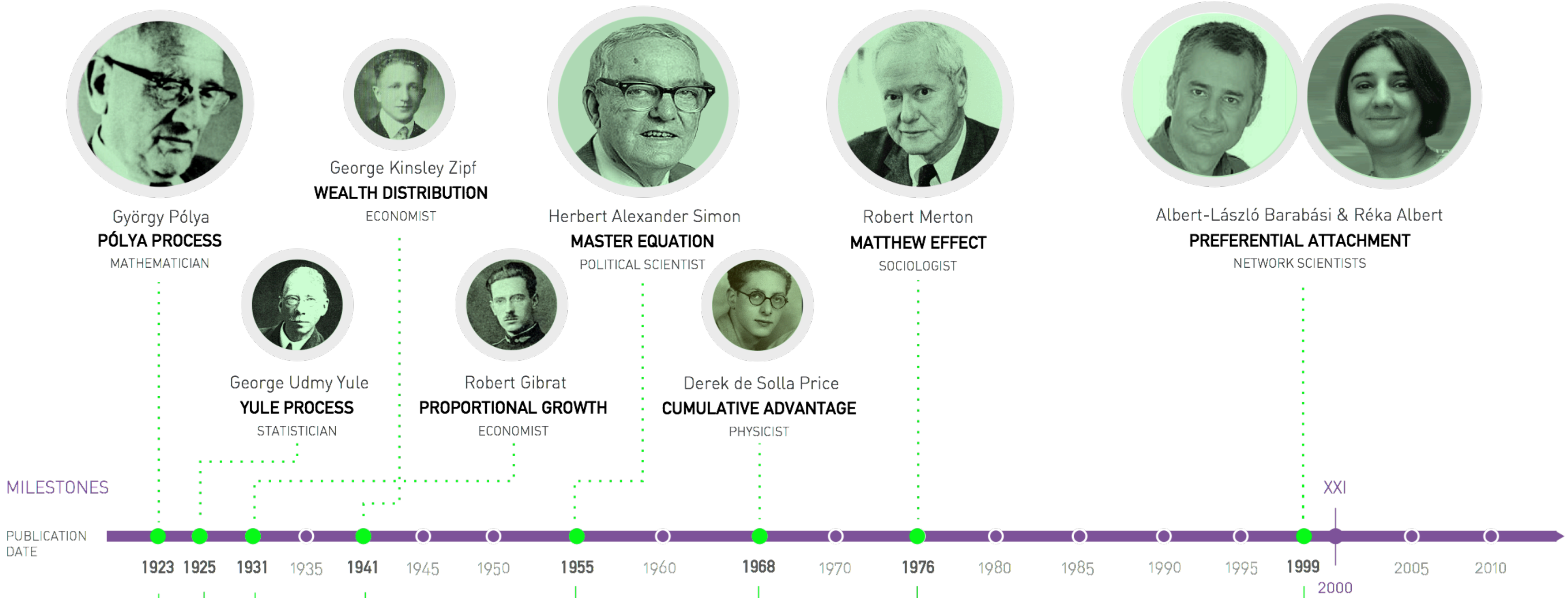
1961

8 Simon's final reply ends but does not resolve the debate [32]

Herbert

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

1961



György Pólya (1887-1985)
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 Udmy 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 Gibrat (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].

Robert Merton (1910-2003)
In sociology preferential attachment is often called the **Matthew effect**, named by Merton [8] after a passage in the Gospel of Matthew.

Barabási (1967) & **Albert** (1972)
introduce the term **preferential attachment** in the context of networks [1] to explain the origin of their power-law degree distribution.