Scale-free networks

Lecture 6

Claudio J. Tessone

Blockchain & Distributed Ledger Technologies
UZH Blockchain Center

Lecture Objectives

1. Learn what is a scale-free distribution
2. Learn what are the mechanisms that lead to the formation of scale-free distributions
3. Learn how to practically determine the parameters of a scale-free distribution

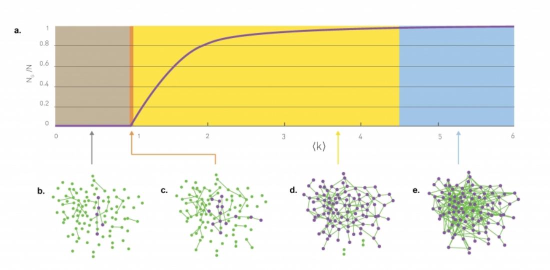
Contents

Large-scale phenomena that have their roots in micro-behaviour

Global order emerges (appears by itself) bottom-up and not enforced (top-down) by a designer or system planner

Distributions

Random Networks: ER Model

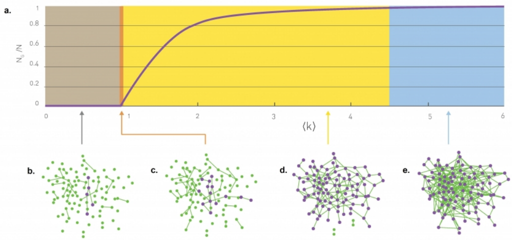


The network is static

Mechanism under study

Random (non-preference) link between nodes

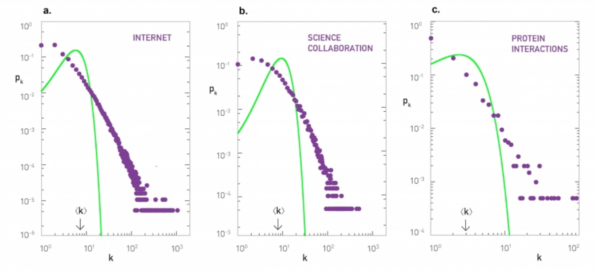
Random Networks: ER Model



It produces Poisson degree distributions (and well-defined average degree), which are unrealistic

Real networks are not Poisson

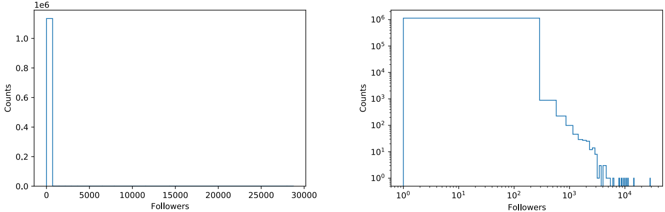
Figure: The green line corresponds to the Poisson prediction, obtained by measuring $\langle k \rangle$ for the real network and then plotting Poisson curve.



Random network model:

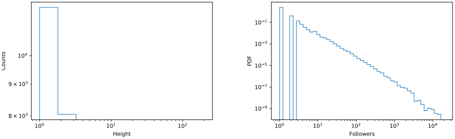
- *underestimates* size and frequency of high degree nodes;
- *underestimates* the number of low degree nodes.
- *overestimates* number of nodes in the vicinity of $\langle k \rangle$.

Heavy-tailed distributions



- In many situations (**and in networks it happens very often**), the distributions involved are *heavy-tailed*
- These distributions when plotted in linear scales do not convey information

Distributions in log-scale



- When plotting distributions in logarithmic scale, you bin logarithmically the data. I.e. $x_2 = a\, x_1, x_2 = a^2\, x_1,$ $...x_N = a^N\, x_1$
- Then, you must plot the distributions, i.e.

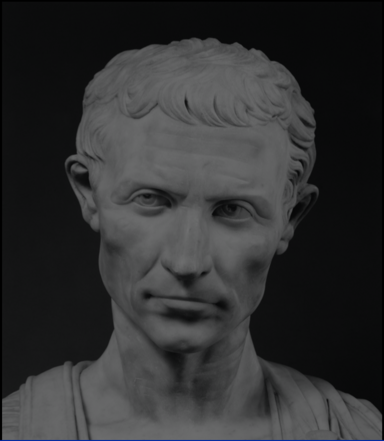
$$p(a \leq x \leq b) \approx \text{count}(a \leq x \leq b)/(b-a)$$

- *Bins should be of similar size in the scale in which they are plotted*

When plotting data distributions, they have their natural scales, and the bins should accompany them

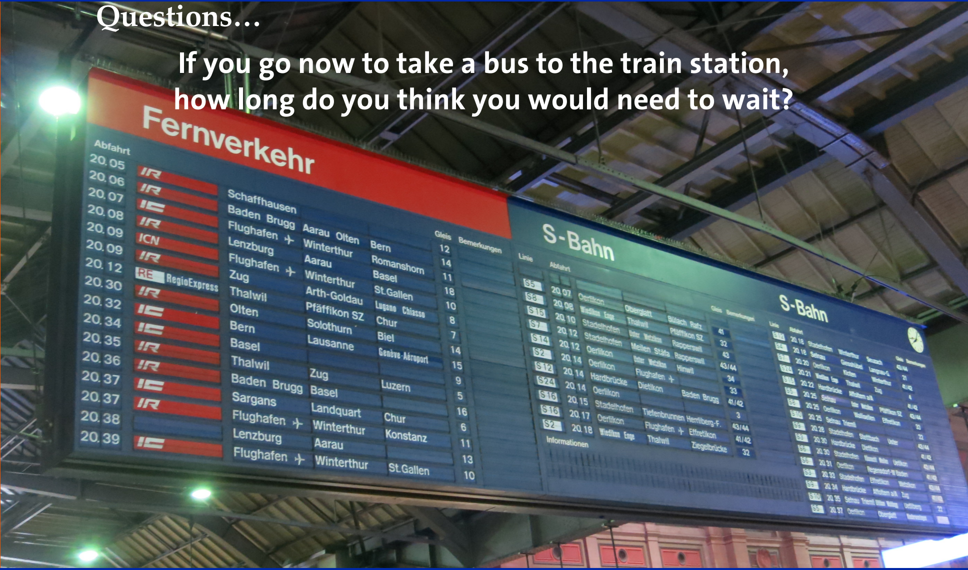
Questions…

How tall was Julius Caesar?

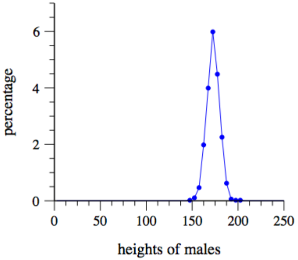


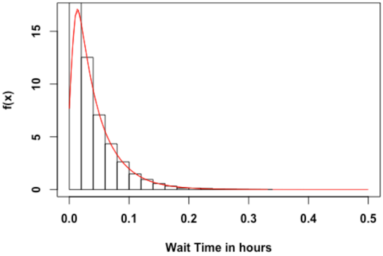
Questions…

If you go now to take a bus to the train station, how long do you think you would need to wait?



Empirical distributions





Many empirical quantities cluster around a typical value

Single-scale distribution: the underlying processes that generate these distributions fall into the general class well-described by the central limit theorem

Questions…

How many connections does this person have in LinkedIn?

Questions…

Which is the wealth of this person?

Degree distribution

$$P(k) = Probability(k_i = k)$$

In many contexts of interest, the degree distribution does not have a characteristic scale

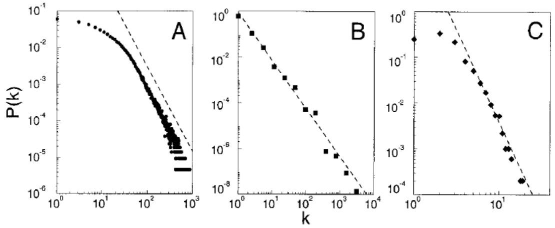
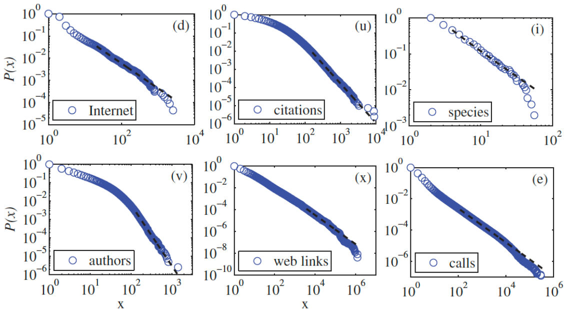


Fig. 1. The distribution function of connectivities for various large networks. (**A**) Actor collaboration graph with $N = 212{,}250$ vertices and average connectivity $\langle k \rangle = 28.78$. (**B**) WWW, $N = 325{,}729$, $\langle k \rangle = 5.46$ (6). (**C**) Power grid data, $N = 4941$, $\langle k \rangle = 2.67$. The dashed lines have slopes (A) $\gamma_{actor} = 2.3$, (B) $\gamma_{www} = 2.1$ and (C) $\gamma_{power} = 4$.

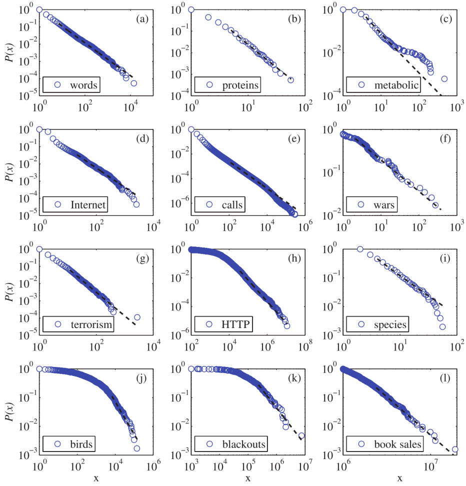
Degree distribution

$$P(k) = Probability(k_i = k)$$

In many contexts of interest, the degree distribution does not have a characteristic scale



Scale-free property



Scale-free distributions are ubiquitous in social and economic systems

Scale-free property

- Given a function $f(x)$
- ...*Scaling* its argument, means computing $c\,x$ for all x

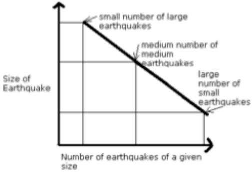
A function is said to be scale-free if $f(c\,x) = b f(x), \forall c, x$

- Let us take the power-law function $f(x) = a\,x^{\gamma}$.

$$f(c\,x) = a\,(c\,x)^{\gamma} = c^{\gamma} f(x) = b f(x)$$

- Regardless of the value of c, the function evaluated at $c\,x$ looks the same as the function at x (because all the derivatives are the same!)

Scale Invariance



Let us consider x to be the size of an earthquake.
$p(x)$ is the PDF of their frequency, which is compatible with a power-law

$$p(x) = N\,x^{-\alpha}$$

The transformation $x \to c\,x$ implies

$$p(c\,x) = c^{-\alpha} N\,x^{-\alpha} \propto p(x)$$

We can also compute

$$\frac{p(100\,x)}{p(10\,x)} = \frac{p(10\,x)}{p(x)} = \frac{p(x)}{p(x/10)}$$

for the same scaling, the ratio between function values is the same, irrespective of x

Preferential attachment

This process is any in which some quantity, typically some form of wealth or size, is distributed among a number of elements ***according to how much they already have****. Therefore, those who are already wealthy receive more than those who are not.*

Other names:

- rich get richer
- cumulative advantage
- Yule process
- Matthew effect

Barabási-Albert model

Barabási-Albert model

- Was proposed in 1999, by Barabási and Albert for studying of the growth of the World Wide Web; (but it is not the first to report the mechanism!0
- *Proposes the **preferential attachment** mechanism to generate random scale-free networks*
- Degree distribution follows a **power-law**, unlike random (ER-model) or small-world (WS-model) graphs.

Applicability of BA model

A number of natural and artificial systems are found to be approximately scale-free:

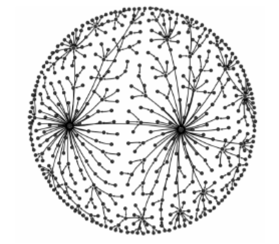
- Internet
- World-Wide-Web
- Citation networks
- Social networks
- etc.

A model is an abstract, simplified description of a real phenomenon aimed at explaining an observation

A model is an abstract, simplified description of a real phenomenon aimed at explaining an observation

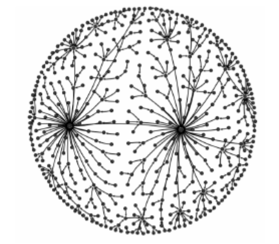
Barabási-Albert model explained

- Model incorporates two mechanisms: **network growth** and **preferential attachment** rule
- **Network Growth**: the network grows in discrete time-steps
- At each time step t, a new node i is added to the network
- The new node is connected to m existing nodes.
- Node degrees ar a function of time $k_i(t)$



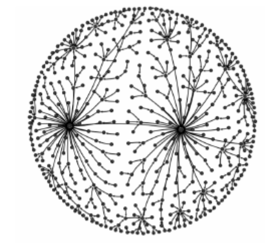
Barabási-Albert model explained

+ Model incorporates two mechanisms: **network growth** and **preferential attachment** rule
+ **Network Growth**: the network grows in discrete time-steps
+ At each time step t, a new node i is added to the network
+ The new node is connected to m existing nodes.
+ Node degrees ar a function of time $k_i(t)$



Barabási-Albert model explained

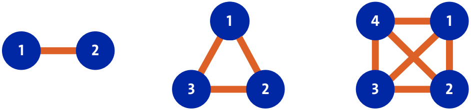
- Model incorporates two mechanisms: **network growth** and **preferential attachment** rule
- **Network Growth**: the network grows in discrete time-steps
- At each time step t, a new node i is added to the network
- The new node is connected to m existing nodes.
- Node degrees ar a function of time $k_i(t)$



Limiting cases

- **Model A**: only growth, no preferential attachment
  - For a new node, probability of connecting to any pre-existing node is equal
  - Resulting degree distribution - geometric
  - **Thus**, only growth is not sufficient for a scale-free property
- **Model B**: only prerefential attachment, no growth
  - Model with a set of existing nodes, adding only links
  - Probability of new connection to node j is $\propto k_j$
  - Degree distribution is scale-free early in simulation, but eventually becomes nearly Gaussian
  - **Thus**, only preferential attachment is not sufficient for a scale-free property

Initial Condition



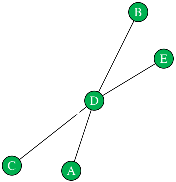
At time $t = 0$ a clique of m_0 nodes is created and serves as initial condition

Preferential attachment

- **Preferential attachment**: The probability that the new node i establishes links to an existing one j is:

$$p(i \leftrightarrow j) \propto k_j(t) = \frac{k_j(t)}{\sum_{j'} k_{j'}(t)} \tag{1}$$

Toy example

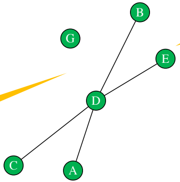


At T=0, there are five nodes in the network.

Toy example

$$p(G \leftrightarrow D) = \frac{4}{4+1+1+1+1} = 50\%$$

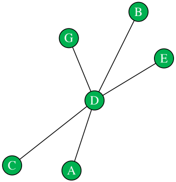
The probabilities that G establishes a link to any other node is 12.5%.



The probability that G establishes a link to D is 50%.

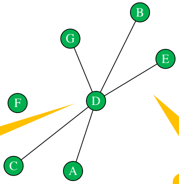
$$p(G \leftrightarrow E) = \frac{1}{4+1+1+1+1} = 12.5\%$$

Toy example



We assume that node G connects to the most likely one, D

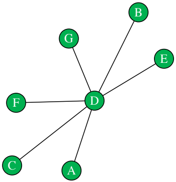
Toy example



The probability that F establishes a link to D is 50%.

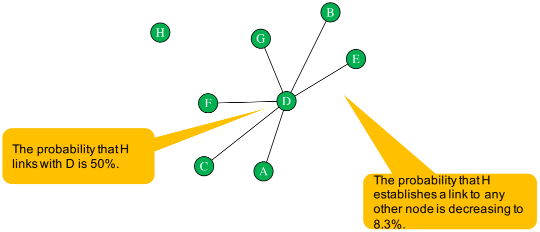
The probabilities that F establishes a link to any other node is decreasing to 10%.

Toy example

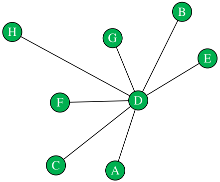


We assume that node F connects to the most likely one, D

Toy example

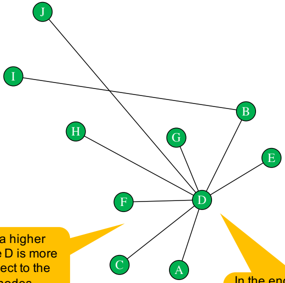

H
G
B
E
F
D
C
A
The probability that H links with D is 50%.
The probability that H establishes a link to any other node is decreasing to 8.3%.

Toy example



We assume that node H connects to the most likely one, D

Toy example



Starting with a higher degree, node D is more likely to connect to the new coming nodes.

In the end, the degree of node D increases a lot while the degrees of the other nodes remain more or less the same.

Node degree distribution: For large t and k, this process produces a scale-free degree distribution $p(k) \propto 1/k^3$

Power-law distribution of node degrees

Node degree

Preferential attachment process generates a “long-tailed” distribution following a Pareto distribution or **power-law** in its tail

$$P(k) \propto k^{-\gamma} \tag{2}$$

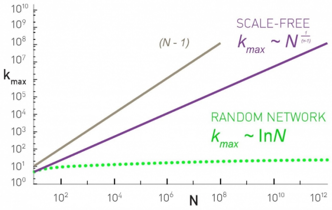
Mathematical derivation in Barabasi’s book

We will now study several models for generation of scale-free networks (networks with power-law distribution of the node degree).

Some characteristics of PA process

- Earlier appearing nodes have more connections
- Heavily connected nodes, **hubs**, quickly accumulate even more connections
- Nodes with few links are unlikely to be chosen for new connections
- New nodes tend to connect themselves to more ‘popular’ nodes

Some characteristics of PA process



The typical maximum degree in the network grows with network size!

Some notes on realism of the PA model

*Preferential attachment **captures a mechanism**: The more acquaintances the users already have, the more likely is that they will be known (or be "attractive" to others).*

- Model suggests that users come one at a time and do not change the connections over time
- This is not correct description of reality
- The origin of the mechanism is unexplained. May be due to:
  1. new users are more likely to know someone who has several acquaintances;
  2. an extroverted person will be signaled by a large social neighborhood and will be more likely to increase it even further.

Extensions of BA model

- Plenty of extensions are built on top of BA model
- They were verified on different types of real data
- Some extensions are:
  - Popularity and similarity (Papadopoulos et al, 2012)
    Nodes connect to popular ones and similar ones
  - Homophily (Centola, 2011)
    People who are connected in a social network tend to like similar things
  - Influentiality and susceptability (Aral and Walker, 2012)
    If influential and susceptible users are connected into scale-free network, information spreads efficiently
  - Positive and negative links
    Friends and foes in a social network
  - Bipartite networks (Li et.al, 2013)
    Users and items connected in a network with scale-free property

A long story

Yule model

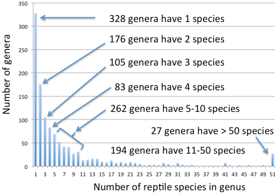
Yule model: Species and Taxonomies

Canis
Temporal range: Miocene to Recent **6 - 0** Ma

Grey wolf (top), coyote and African golden wolf (top middle), Ethiopian wolf and golden jackal (bottom middle), black-backed jackal and side-striped jackal (bottom)

Scientific classification	
Kingdom:	Animalia
Phylum:	Chordata
Class:	Mammalia
Order:	Carnivora
Suborder:	Caniformia
Family:	Canidae
Tribe:	Canini
Genus:	***Canis*** Linnaeus, 1758[1]

- *Species* are grouped in *genera*
- The distribution of number of species that belong to different genera is very skewed



Yule model: Species and Taxonomies

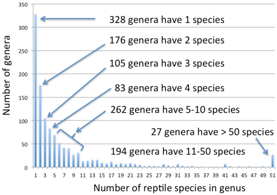
Canis

Temporal range: Miocene to Recent 6 - 0 Ma

Grey wolf (top), coyote and African golden wolf (top middle), Ethiopian wolf and golden jackal (bottom middle), black-backed jackal and side-striped jackal (bottom)

Scientific classification	
Kingdom:	Animalia
Phylum:	Chordata
Class:	Mammalia
Order:	Carnivora
Suborder:	Caniformia
Family:	Canidae
Tribe:	Canini
Genus:	***Canis*** Linnaeus, 1758[1]

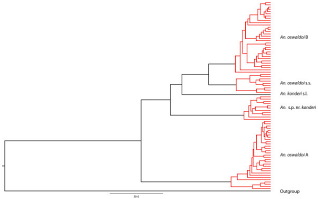
- *Species* are grouped in *genera*
- The distribution of number of species that belong to different genera is very skewed

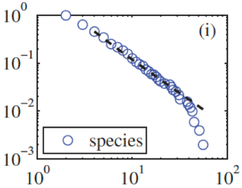


Yule model

[Yule, *Phyl. Trans. Royal Soc. B* (1925)]

- First example of a proportionate growth process
- Intended to explain biological evolution
- Based on the observation of the highly skewed distribution of number of species (of flowers) that compose different genera

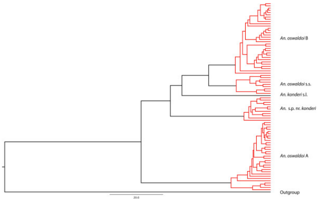


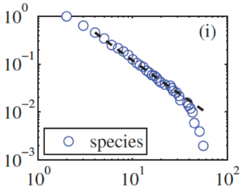


Yule model

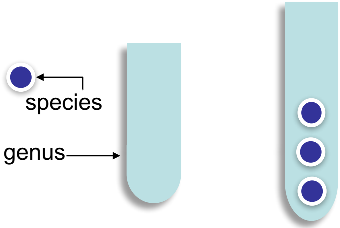
[Yule, *Phyl. Trans. Royal Soc. B* (1925)]

- First example of a proportionate growth process
- Intended to explain biological evolution
- Based on the observation of the highly skewed distribution of number of species (of flowers) that compose different genera



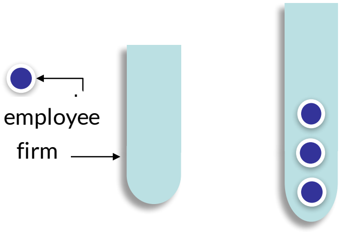


Yule model: species and genera



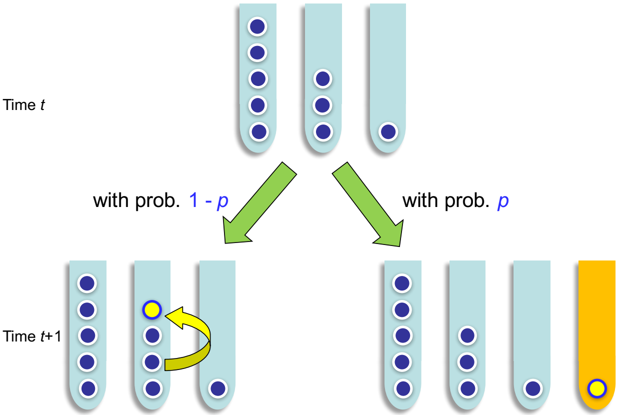
- Species are grouped in genera
- Over time, new species appear by the mutation of existing ones
- At rate $1 - p$, the new species belongs to the same genus
- Otherwise, at rate p, new species generates a new genus

Yule-Simon model: Firm-size distribution



- Employees work in firms
- Over time, new employees entersinto the market
- With probability $1-p$, the new employee works in an existing firm
- Otherwise, with probability p, the new employee opens a new firm

Yule-Simon process sketch



Growth of a firms is proportional to its size

Firms size distribution

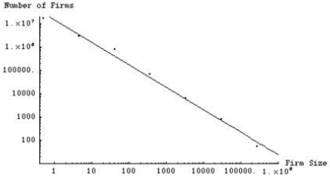


Figure 2: Distribution of U.S. firm sizes (by employees) for 1997, data combined from Census/SBA and Compustat together with self-employment data

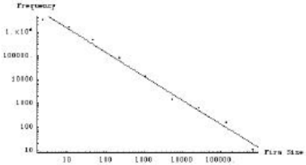
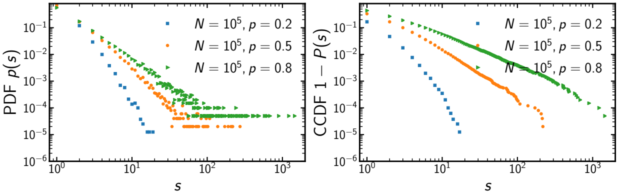


Figure 3: Distribution of U.S. firm sizes (by revenue, in $ million) for 1997, data from Census

The empirical observation is that firms have broad size distributions, by size or revenue [*Axtel, Science* (2001)]

Results of the model



$$p(s) = c\frac{1}{s^{\gamma}} \Rightarrow 1 - P(s) = 1 - \int_0^s p(\sigma)\, d\sigma = c_2 \frac{1}{s^{\gamma-1}}$$

Firm size distribution

$$p(s) \propto \frac{1}{s^{1+\frac{p}{1-p}}}$$

The Matthew Effect

Citations in scientific publications

The Matthew Effect in Science

The reward and communication systems of science are considered.

Robert K. Merton

This paper develops a conception of ways in which certain psychosocial processes affect the allocation of rewards to scientists for their contributions—an allocation which in turn affects the flow of ideas and findings through the communication networks of science. The conception is based

image and the public image of scientists are largely shaped by the communally validating testimony of significant others that they have variously lived up to the exacting institutional requirements of their roles.

A number of workers, in empirical studies, have investigated various as-

reers are more productive later on than those who do not. And the Coles have also found that, at least in the case of contemporary American physics, the reward system operates largely in accord with institutional values of the science, inasmuch as quality of research is more often and more substantially rewarded than mere quantity.

In science as in other institutional realms, a special problem in the workings of the reward system turns up when individuals or organizations take on the job of gauging and suitably rewarding lofty performance on behalf of a large community. Thus, that ultimate accolade in 20th-century science, the Nobel prize, is often assumed to mark off its recipients from all the other scientists of the time. Yet this assumption is at odds with the well-known fact that a good number of scientists who have not received the prize and will not receive it have contributed as much to the advancement

Reprinted from SCIENCE

January 5, 1968, Vol. 159, No. 3810, pages 56-63

Matthew effect

Robert Merton:

"The Matthew effect consists in the accruing of greater increments of recognition for particular scientific contributions to scientists of considerable repute and the withholding of such recognition from scientists who have not yet made their mark."

Matthew 25:29, King James Version:

"For unto every one that hath shall be given, and he shall have abundance: but from him that hath not shall be taken even that which he hath."

Matthew effect

Robert Merton:

“The Matthew effect consists in the accruing of greater increments of recognition for particular scientific contributions to scientists of considerable repute and the withholding of such recognition from scientists who have not yet made their mark.”

Matthew 25:29, King James Version:

“For unto every one that hath shall be given, and he shall have abundance: but from him that hath not shall be taken even that which he hath.”

Examples of Matthew effect

- **Academia**, researchers tend to cite well-known scholars
- **Consumer activity** follows the apparent popularity: books and music from bestseller lists are sold more
- **Career progress**: more experienced and more reputable individuals posess more competitive advantage in obtaining new career opportunities
- **Internet, social networks**

The Price model

A General Theory of Bibliometric and Other Cumulative Advantage Processes*

A Cumulative Advantage Distribution is proposed which models statistically the situation in which success breeds success. It differs from the Negative Binomial Distribution in that lack of success, being a non-event, is not punished by increased chance of failure. It is shown that such a stochastic law is governed by the Beta Function, containing only one free parameter, and this is approximated by a skew or hyperbolic distribution of the type that is widespread in bibliometrics and diverse social science phenomena. In particular, this is shown to be an appropriate underlying probabilistic theory for the Bradford Law, the Lotka Law, the Pareto and Zipf Distributions, and for all the empirical results of citation frequency analysis. As side results one may derive also the obsolescence factor for literature use. The Beta Function is peculiarly elegant for these manifold purposes because it yields both the actual and the cumulative distributions in simple form, and contains a limiting case of an inverse square law to which many empirical distributions conform.

Derek de Solla Price
Department of History of Science and Medicine
Yale University
New Haven, CT 06520

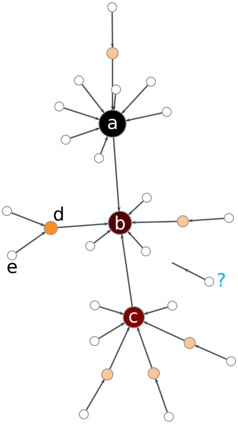
[De Solla Price, *1976* (J. Assoc. Inf. Sci.)]

The Price model

- **Nodes:** scientific papers
- **Links:** citations
- Probability of a new node i points to an existing one j:

$$p(i \to j) \propto k_j^{(in)} + a$$

- $p(\to d) = 2p(\to e)$,
  $p(\to a) = 4p(\to d)$,

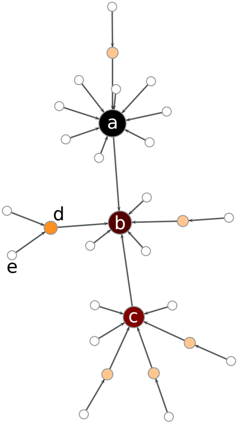


The Price model

- **Nodes:** scientific papers
- **Links:** citations
- Probability of a new node i points to an existing one j:

$$p(i \to j) \propto k_j^{(in)} + a$$

- $p(\to d) = 2p(\to e)$,
  $p(\to a) = 4p(\to d)$,



The Price Model explained

- Let p_k to denote the fraction of nodes with degree k
  ... so $\sum_k p_k = 1$
- Each paper cites on average c papers
- This value is constant over time, and:
  $c = \sum_k k p_k$
- Probability to link to a specific node is:

$$Prob(\text{attach to } i, t) = \frac{q_i + a}{\sum_j q_j + a} = \frac{q_i + a}{t(c + a)} \qquad (3)$$

where $q_i = k_i^{(in)}$ - in-degree of node i

The Price Model explained

Let $P(q,t)$ be the degree distribution of nodes with in-degree q at time t

At time t, there are $tP(q,t)$ nodes with in-degree q

$$(t+1)P(q,t+1) = tP(q,t) - c\frac{t(q+a)}{t(c+a)}P(q,t) + \\ + c\frac{t(q-1+a)}{t(c+a)}P(q-1,t) \quad (4)$$

$$(t+1)P(0,t+1) = tP(0,t) - c\frac{ta}{t(c+a)}P(0,t) + 1 \quad (5)$$

The Price Model properties

The stationary degree distribution is given by:

$$P(q, t \to \infty) = \frac{B(q + a, 2 + \frac{a}{c})}{B(a, 1 + \frac{a}{c})} \quad (6)$$

For large degree, this reduces to:

$$P(q, t \to \infty) \propto q^{2+\frac{a}{c}} \quad (7)$$

Barabási-Albert model is a special case with $a = c = m$

Many real-world networks exhibit exponents in the range $\alpha \in (2, 3)$

Results of the model

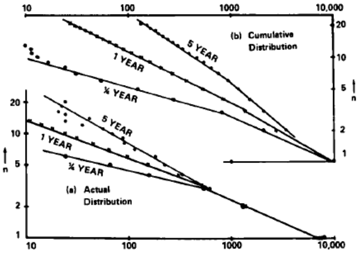
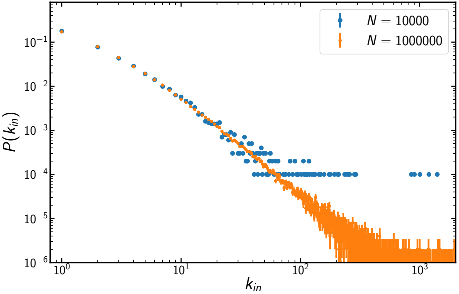


Fig. 1. Number of papers with (a) exactly and (b) at least n citations in ¼, 1, and 5-year indexes.

Cummulative distribution of number of citations

Notice that both scales are logarithmic!

Properties


$N = 10000$
$N = 1000000$
$P(k_{in})$
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{0}
10^{1}
10^{2}
10^{3}
k_{in}

Model extensions

Fitness model

[Bianconi, Ginestra, Barabási, Albert-László, *Bose-Einstein condensation in complex networks* (2001)]

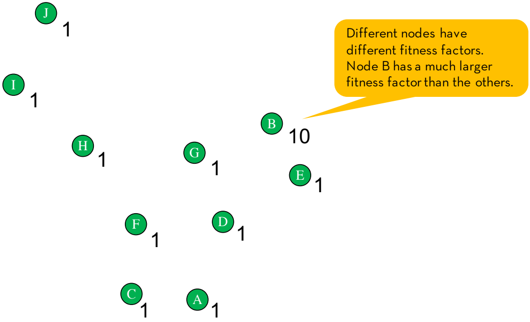
- Extension of BA model
- **Fitness** *is an inherent time independent factor* that makes particular nodes more attractive
- Fitness works together with the degree of nodes creating overall attractiveness of a node

Fitness model explained

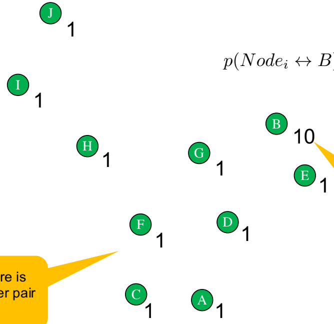
- Consider a network that grows in discrete time, with initial size m_0
- **Each node in endowed with a fitness factor** η_i
- At each time step, a new node is added to the network
- The new node i is connected to $m \leq m_0$ existing nodes
- The probability that the new node i establishes a link to an existing one j is:

$$p(i \leftrightarrow j) = \frac{\eta_j k_j}{\sum_\ell (\eta_\ell k_\ell)} \tag{8}$$

Fitness model illustration



Fitness model illustration



$$p(Node_i \leftrightarrow B) = \frac{0.1 * 10 * 1}{1 + 0.1 * 10 * 1} = 50\%$$

The probability that there is a link between node B and any other node is 50%.

The probability that there is a link between any other pair of nodes is 9%.

$$p(Node_i \leftrightarrow Node_j) = \frac{0.1 * 1 * 1}{1 + 0.1 * 1 * 1} = 9\%$$

Some properties of fitness model

- Reflects basic properties of many real systems;
- A node with strong competitive power (higher fitness factor) may enter the network at a later time and overcome 'older' nodes (unlike in plain BA model);
- The network growth process depends on the choice of fitness factor.
- The choice of fitness is somewhat arbitrary;
- The actual fitness factors might be unobservable or change over time.

Intrinsic fitness introduciton

[G. Caldarelli, *et al.*, *Scale-Free Networks from Varying Vertex Intrinsic Fitness* (2002)]

- Proposed as an **alternative mechanism** leading to scale-free networks
- Power-law distribution is neither related to network growth nor to preferential attachment

Intrinsic fitness explained

[Diego Garlaschelli, Maria I Loffredo, *Fitness-dependent topological properties of the world trade web* (2004)]

- Consider network with N nodes
- Each node is endowed with a fitness z_i
- Every link is created with a probability p_{ij}

$$p_{ij} = f(z_i, z_j) = \frac{z_i z_j}{1 + z_i z_j} \tag{9}$$

- Probability of existence of an edge between two nodes depends uniquely on the actual nodes characteristics!

Intrinsic fitness model in real data

+ *Network modelling*: aimed at understanding mechanisms yielding global propeties
+ *Intrinsic fitness model*: aimed to relate node intrinsic properties of the observed nework

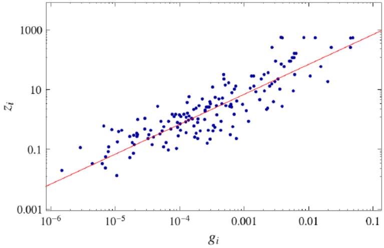
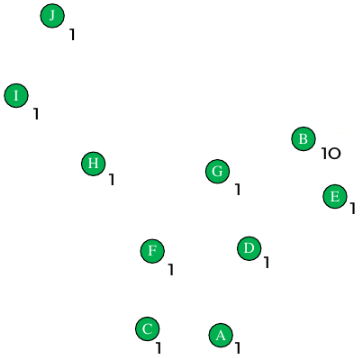


Figure: GDP vs. fitness in the World Trade Web

Intrinsic fitness visualisation

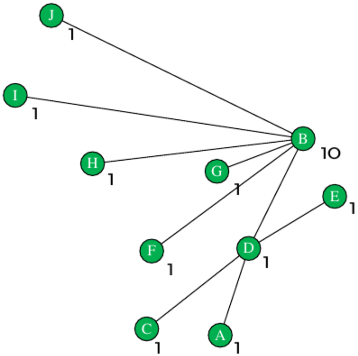


We start with a graph with no edges.
For every pair of nodes, there is a probability of connection:

$$P(Node_i \leftrightarrow B) = \frac{0.1*10*1}{1+0.1*10*1} = \frac{1}{2}$$

$$P(Node_i \leftrightarrow Node_j) = \frac{0.1*1*1}{1+0.1*1*1} = \frac{1}{11}$$

Intrinsic fitness visualisation



One possible network might look like this

Intrinsic fitness model: properties

+ Can successfuly reproduce power-law degree ditribution
+ Shows the *good-get-richer* rather than the rich get richer
+ Intrinsic fitness factor can be obtained from other sources (e.g. GDP), but can also be unobserved or change over time
+ Cannot be used to model network evolution

Intrinsic fitness model: interpretation

- We start with a network $G^*, \{a_{ij}\}, \{k_i^*\}$
- Each node is endowed with a fitness x_i
- Probability of creating a link between two nodes is:
  $p_{ij} = f(x_i, x_j) = \frac{x_i x_j}{1+x_i x_j}$
- The probability to oberve a network G^*, given these fitness values, is then:
  $P(G^*|\{x_i\}) = \prod_{i<j} p_{ij}^{a_{ij}} (1 - p_{ij})^{1-a_i j}$
- Log-likelihood of this observation:
  $L(G^*|\{x_i\}) = \sum_{i<j} \left[a_{ij} \log(p_{ij}) + (1 - a_{ij}) \log(1 - p_{ij})\right]$

Intrinsic fitness model: interpretation

Maximisation of the log-likelihood w.r.t. x_i yields:

$$\langle k_i \rangle = \sum_{i \neq j} \frac{x_i x_j}{1 + x_i x_j} = k_i^*, \forall i \tag{10}$$

Interpretation of x_i: expected degrees $\langle k_i \rangle$ match the observed ones $k_i^* = k_i(G^*)$

This is equivalent to maximising enthropy under constraint: $k_i^* = \langle k_i \rangle$, and enthropy is given by:

$$\mathbb{H}(G^*) = -\sum_{i=1}^{N} \log(x_i) k_i^* \tag{11}$$

References I

- Ravi Srinivasan, Chih-Hung Chen *M375T/M396C: Topics in Complex Networks*, Lecture notes, 2013.
- Ginestra Bianconi, Albert-Laszlo Barabási, *Bose-Einstein condensation in complex networks*, in Physical Review Letters, 2001.
- Albert- Laszlo Barabási, *Network Science* Cambridge University Press, 2015.