

PART II: “THE PROBLEM OF COMPANY GROWTH”

Question: Are there “laws” quantifying how companies grow/shrink?

Answer: Economists know much, dating back to Gibrat (1930’s)

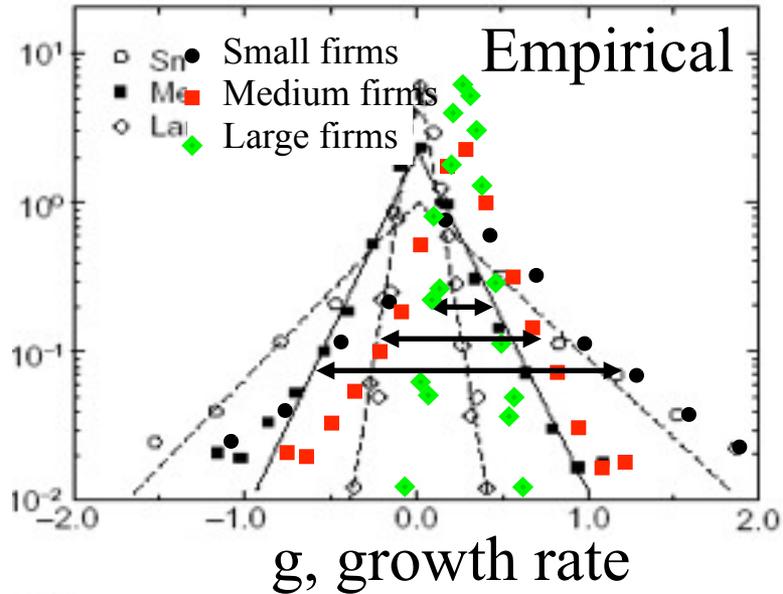
Take home message

- P(growth rate) Laplace in Center: universal
- Width decreases as $-1/6$ power of size bin
- P(growth rate) crosses over to power law in wings
- No theory for $-1/6$ power law for width
- Theory (Buldyrev et al) for growth rate power law

Collaborators: Salinger, Buldyrev, Canning, Havlin, Amaral, Fu, Pammoll
Yamasaki, Matia, Ponta, Riccaboni (also: Jeffrey Sachs!!!!!!)

(surprising) Empirical Observations (before 1999)

Probability density

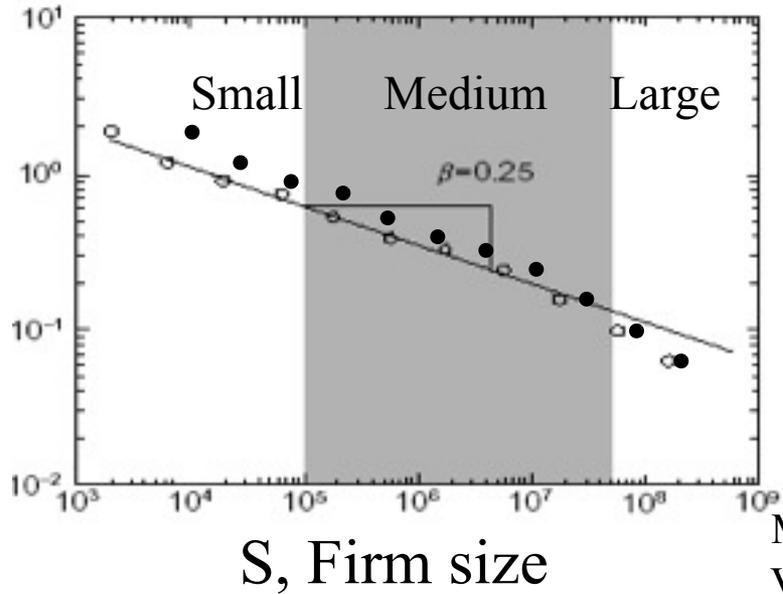


Reality: it is "tent-shaped"!

$$\text{pdf}(g|S) \sim e^{-\frac{|g|}{\sigma(S)}}$$

[[NOT log-normal (Gibrat theory)]]

Standard deviation of g

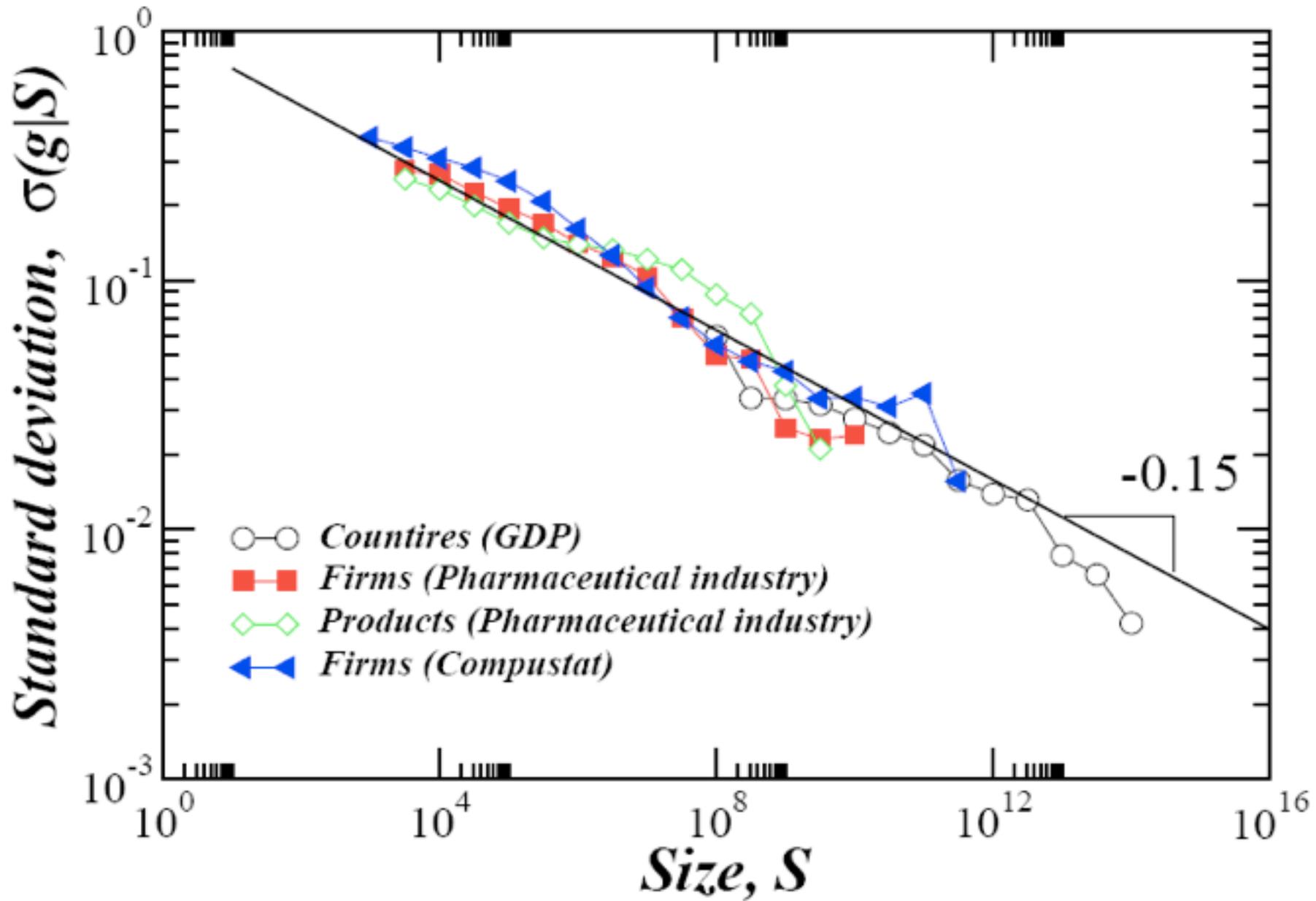


$$\sigma_g(S) \sim S^{-\beta}, \quad \beta \approx 0.2$$

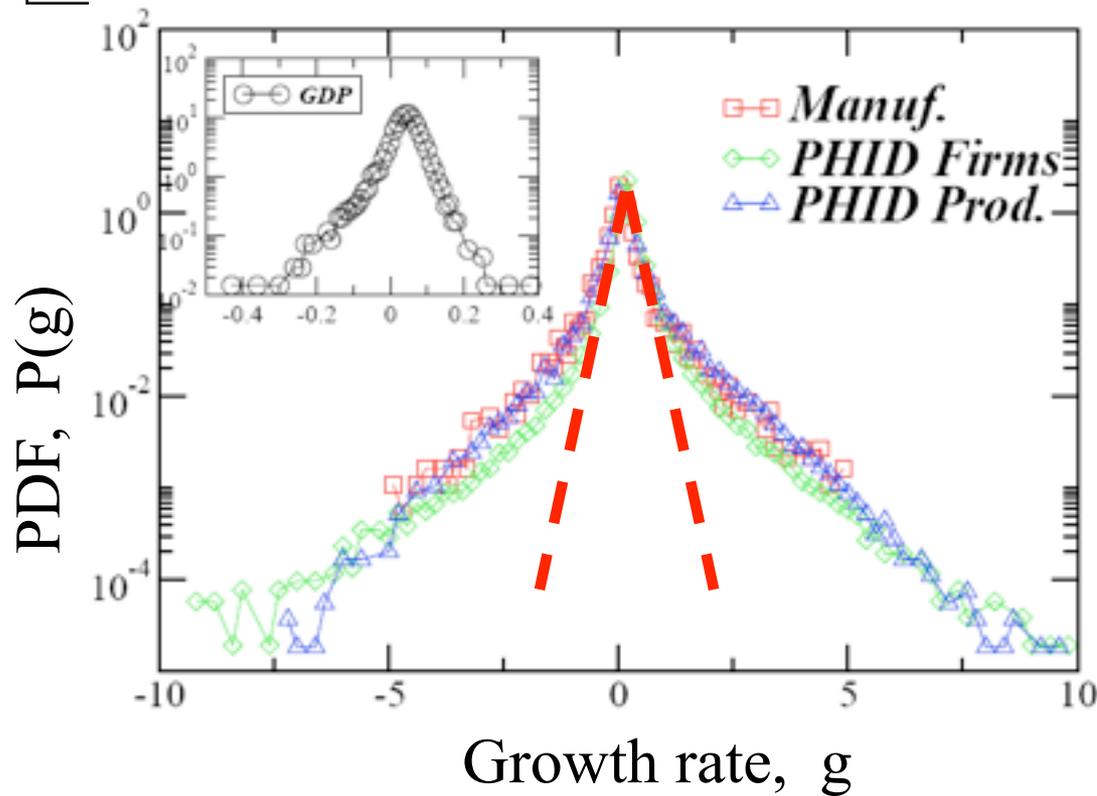
Universal for different economies (Takayasu) and organizations (university budgets, bird populations)

Michael H. R. Stanley, *et.al.* Nature **379**, 804-806 (1996).
 V. Plerou, *et.al.* Nature **400**, 433-437 (1999).

Size-Variance Relation Universality



Empirical Findings: $P(g)$



Laplace fitting function:

$$P(g) \sim \exp(-|\bar{g}-g|/\sigma)$$

[Nature 379, (1996)]

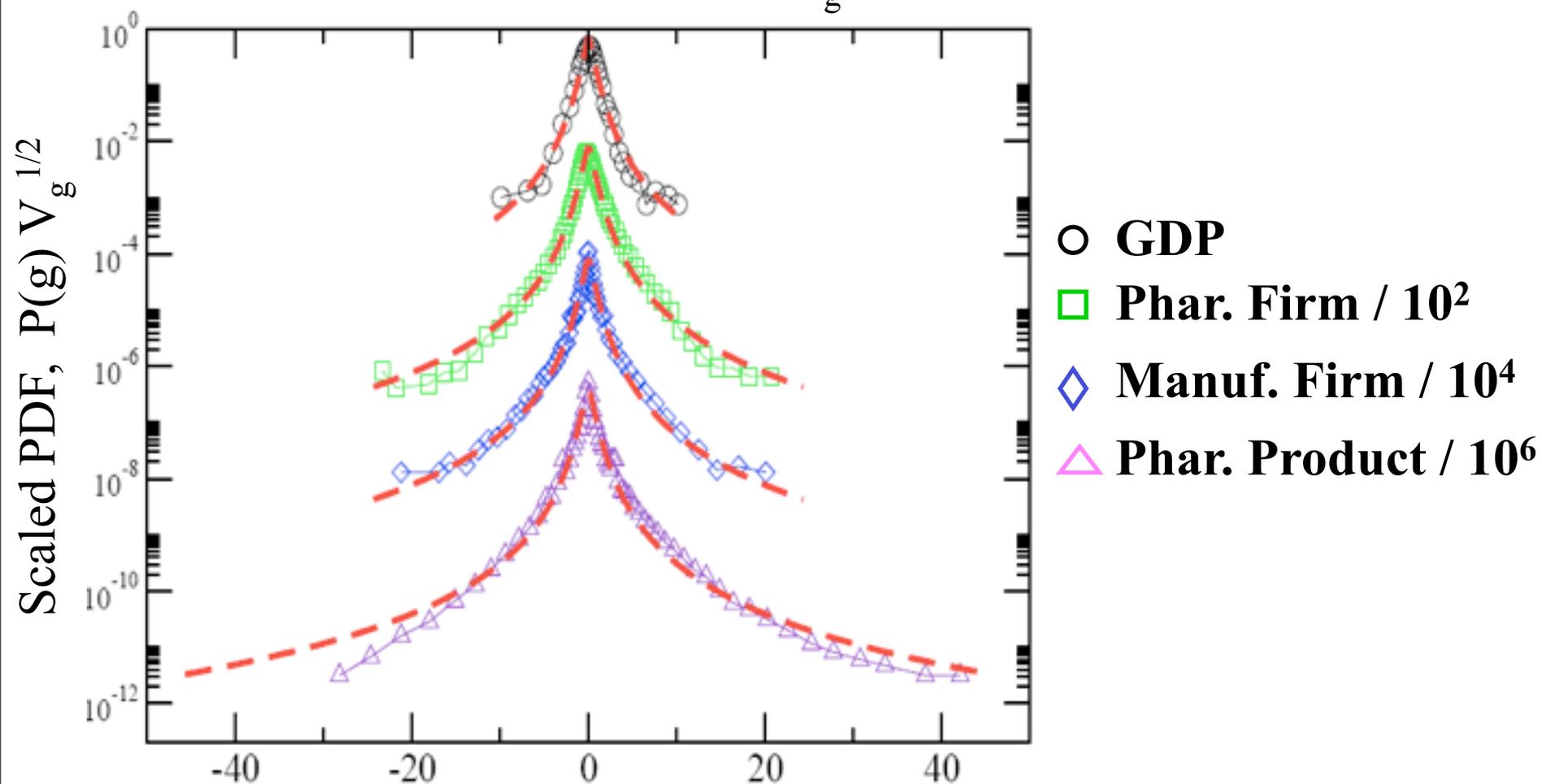
Our question: What is the function $P(g)$, the PDF of growth rate?

Answer: Not Gaussian, [Gibrat (1930)].

Not Laplace, [M.H.R. Stanley, et al (1996)].

Test variance scaling with Empirical Data

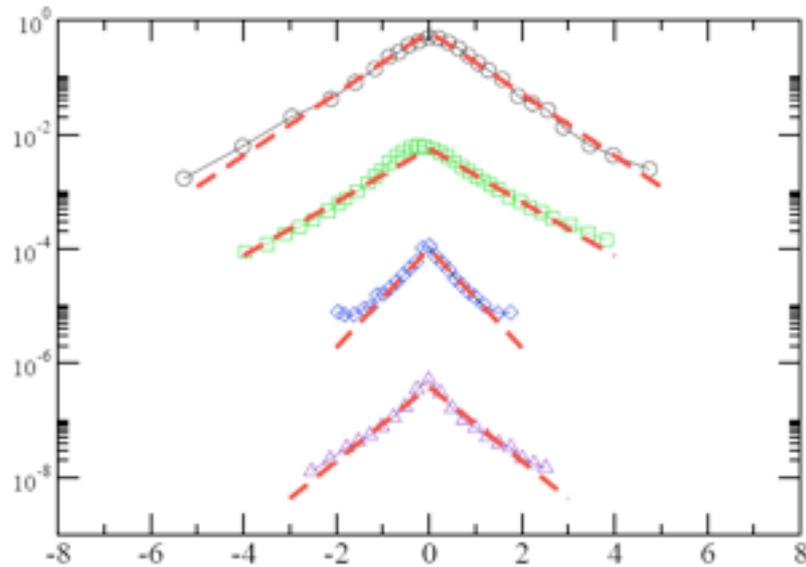
ONLY **One** Parameter: V_g



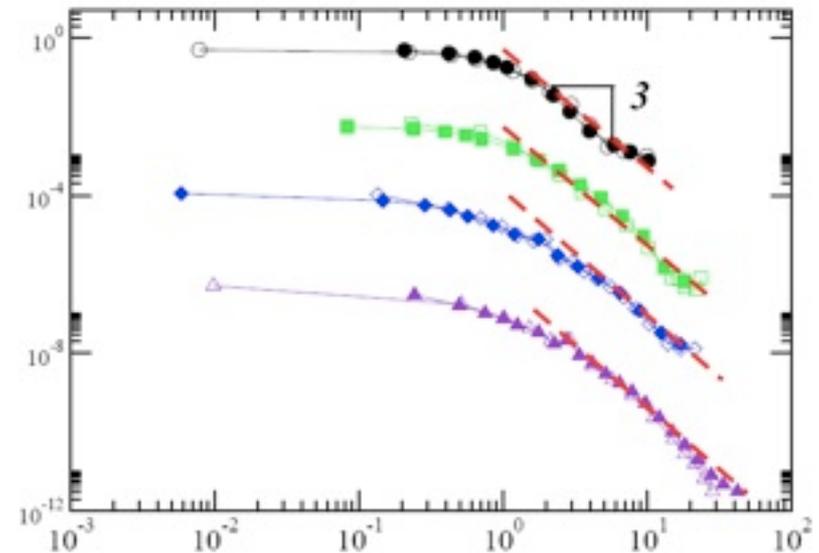
Scaled growth rate, $(g - \bar{g}) / V_g^{1/2}$

The Test of Central & Tail Parts of $P(g)$

Scaled PDF, $P(g) V_g^{1/2}$



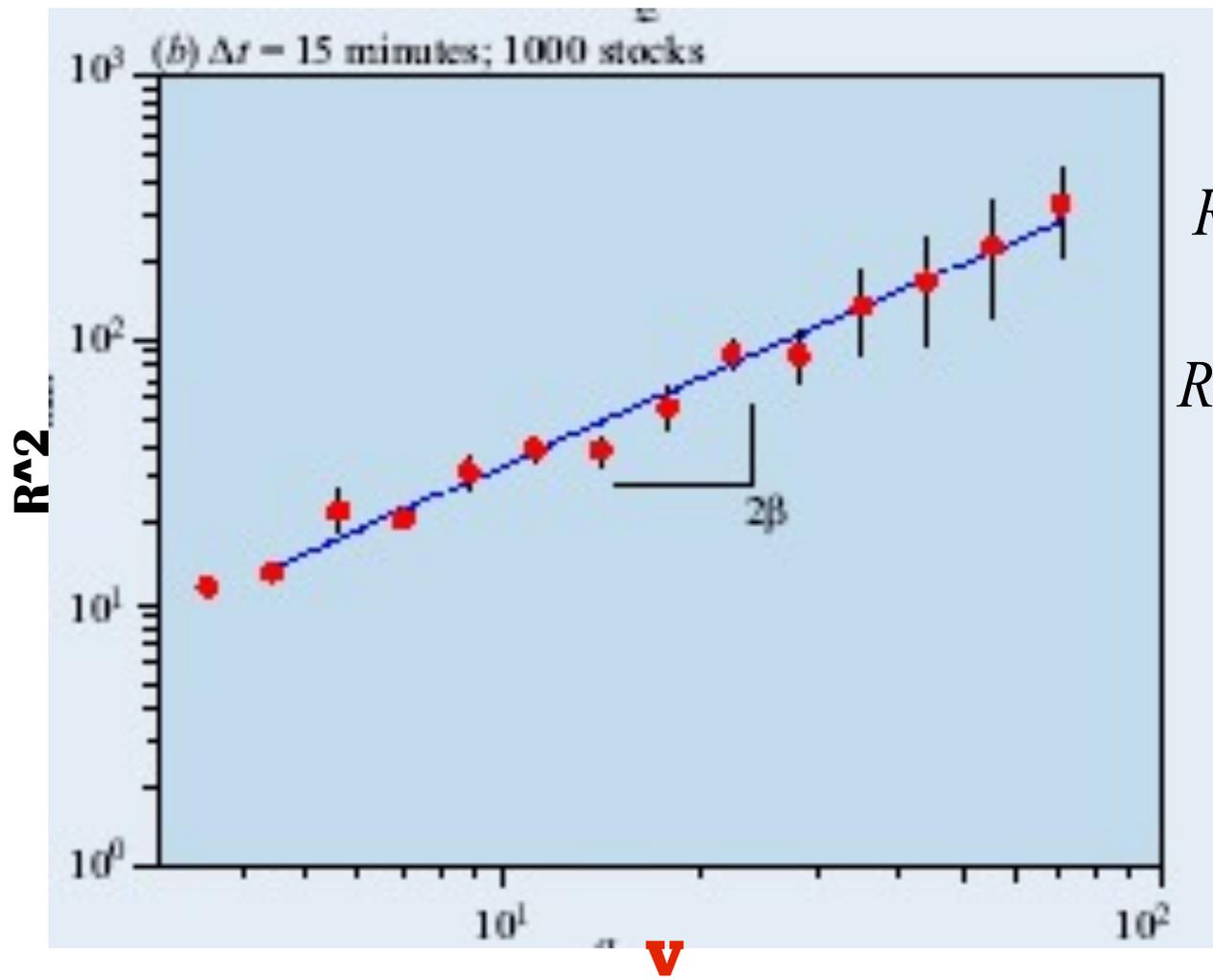
Scaled growth rate, $(g - \bar{g}) / V_g^{1/2}$



Central part is Laplace.

Tail part is power-law
with exponent -3.

Returns & Volume: Square-root form of Price Impact



$$R^2 \sim V^{2\beta} \quad 2\beta \approx 1$$

$$R^2 \sim V \Rightarrow |R| \sim \sqrt{V}$$

$$\Rightarrow \zeta_R = 2\zeta_V$$

“Volume drives price”

Summary of results

- Distribution of returns consistent with a power-law functional form

$$P(R > x) \sim x^{-\zeta_R} \quad \zeta_R \approx 3$$

- Two more power-law relationships for market activity (N) and the volume traded:

$$P(N > x) \sim x^{-\zeta_N} \quad \zeta_N \approx 3$$

$$P(V > x) \sim x^{-\zeta_V} \quad \zeta_V \approx \frac{3}{2}$$

- Square-root form of price impact

$$R^2 \sim V \Rightarrow |R| \sim \sqrt{V} \quad \zeta_R = 2\zeta_V$$

- Power-law tails of returns arise from volume
- Long-memory of $|R|$ arises from N

Goals:

- Explain the abundance of Pareto law in economics and other branches of science and humanities.
- A man should look for what **is**, and not for what he thinks **should be** (A. Einstein)
- Investigate how few simple assumptions give rise to various statistical patterns in socio-economics systems.

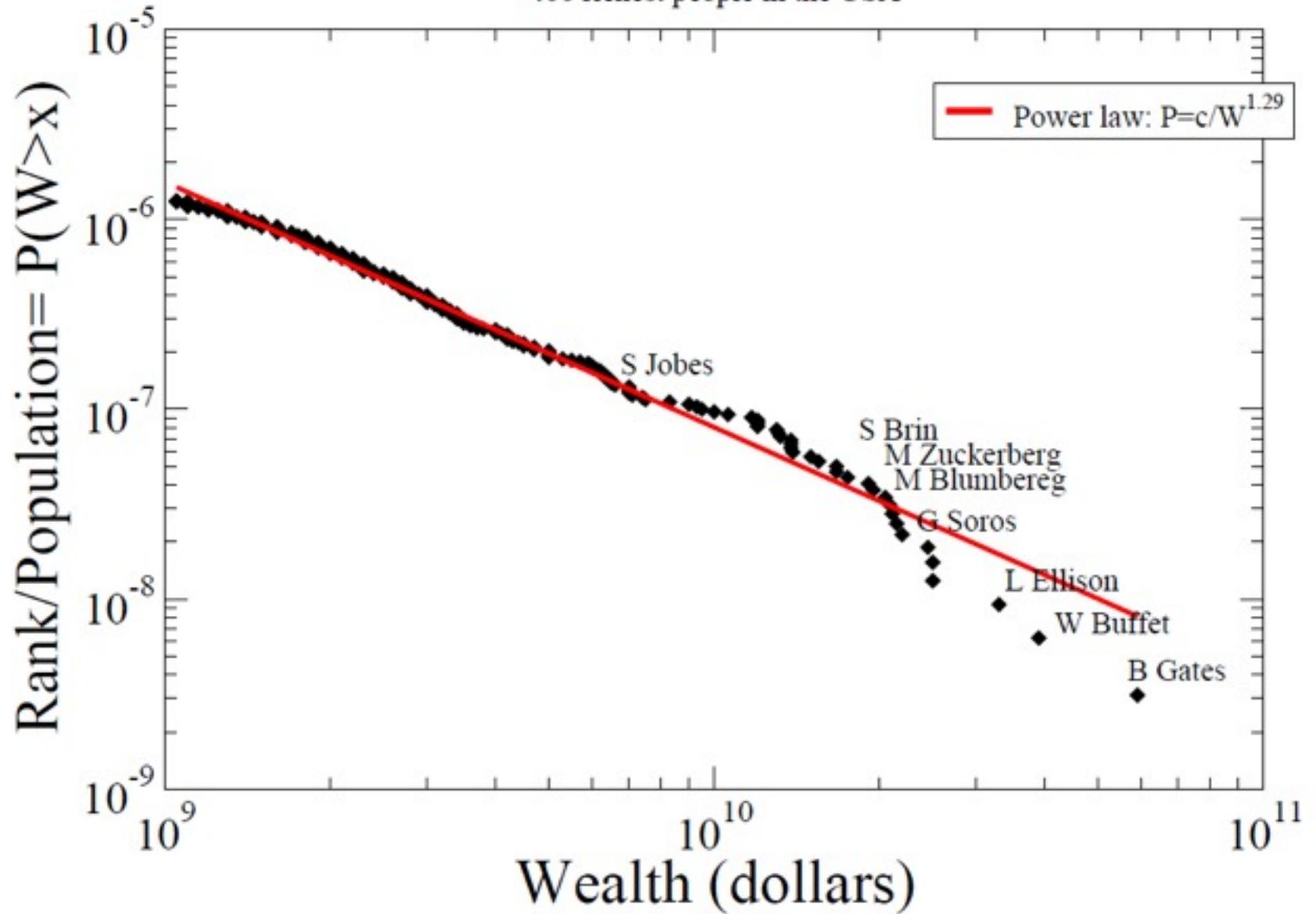
The grand aim of all science is to cover the greatest number of empirical facts by logical deduction from the smallest number of hypotheses or axioms (Einstein)

- Learn how to simulate models of socio-economic phenomena on the computer
- Build a comprehensive theory of firm growth.

1		Bill Gates	\$59 B	55	Medina, Washington	Microsoft
2		Warren Buffett	\$39 B	81	Omaha, Nebraska	Berkshire Hathaway
3		Larry Ellison	\$33 B	67	Woodside, California	Oracle
4		Charles Koch	\$25 B	75	Wichita, Kansas	diversified
4		David Koch	\$25 B	71	New York, New York	diversified
6		Christy Walton	\$24.5 B	56	Jackson, Wyoming	Wal-Mart
7		George Soros	\$22 B	81	Katonah, New York	hedge funds
8		Sheldon Adelson	\$21.5 B	78	Las Vegas, Nevada	casinos
9		Jim Walton	\$21.1 B	63	Bentonville, Arkansas	Wal-Mart
10		Alice Walton	\$20.9 B	61	Fort Worth, Texas	Wal-Mart

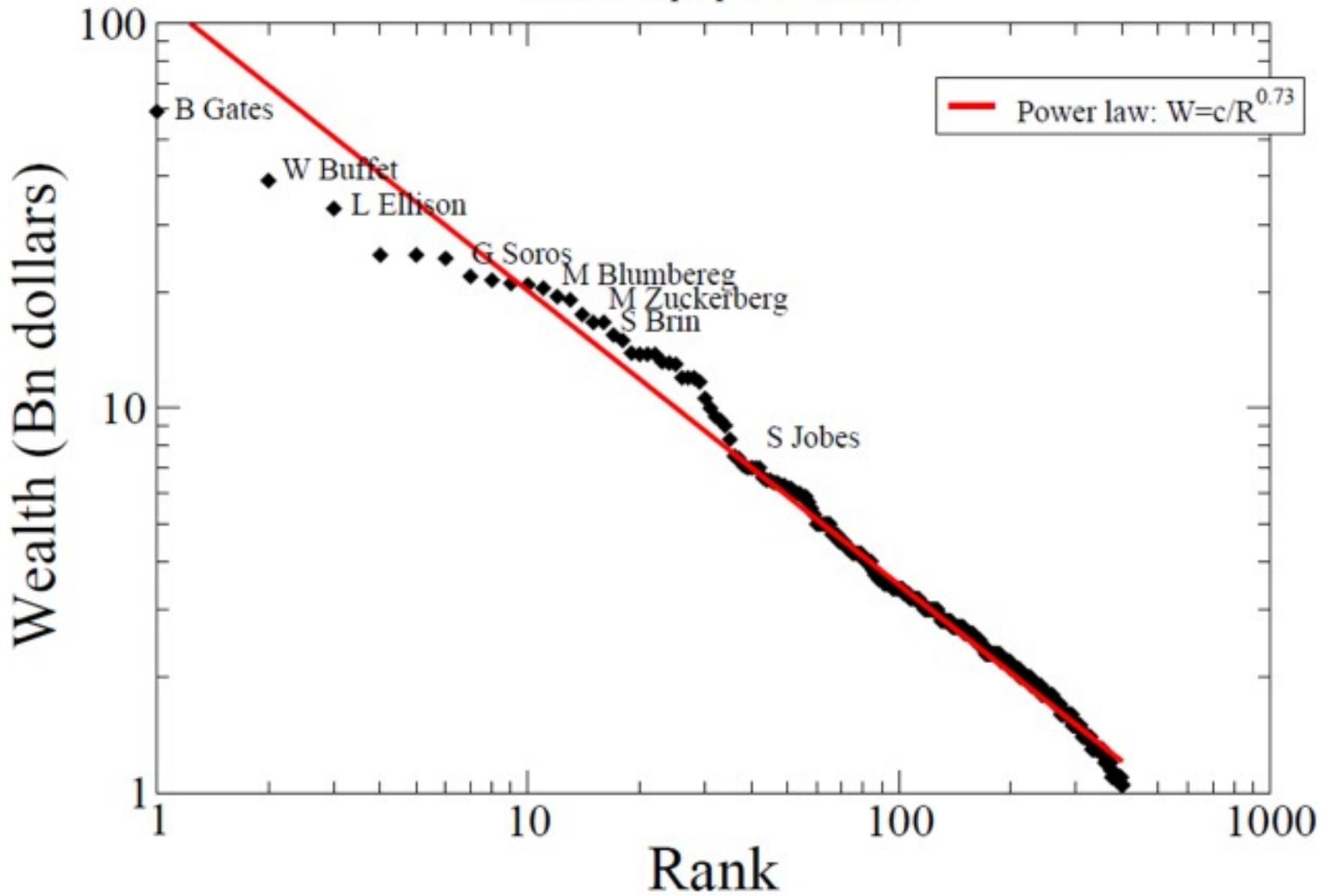
Forbes

400 richest people in the USA

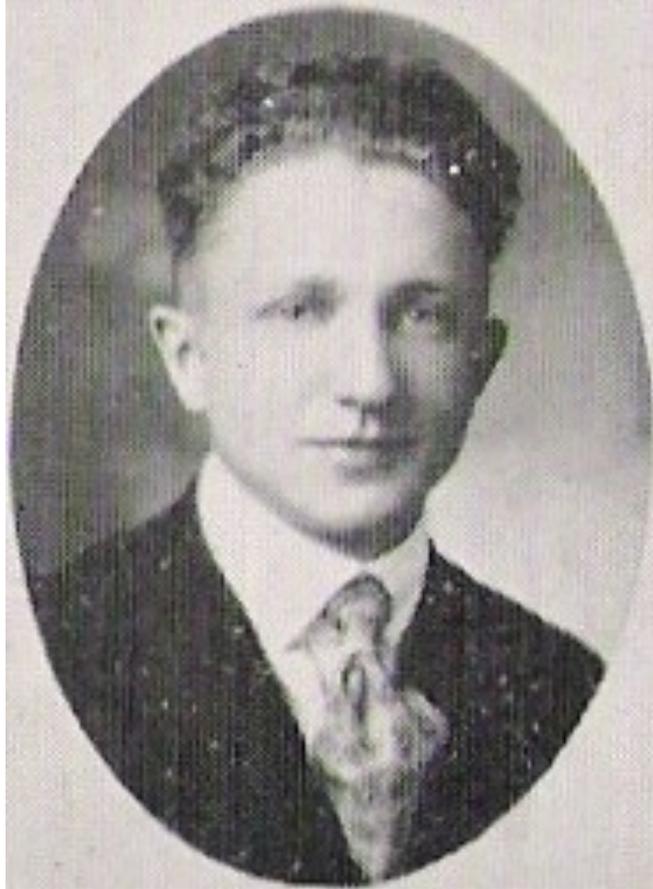


Forbes

400 richest people in the USA

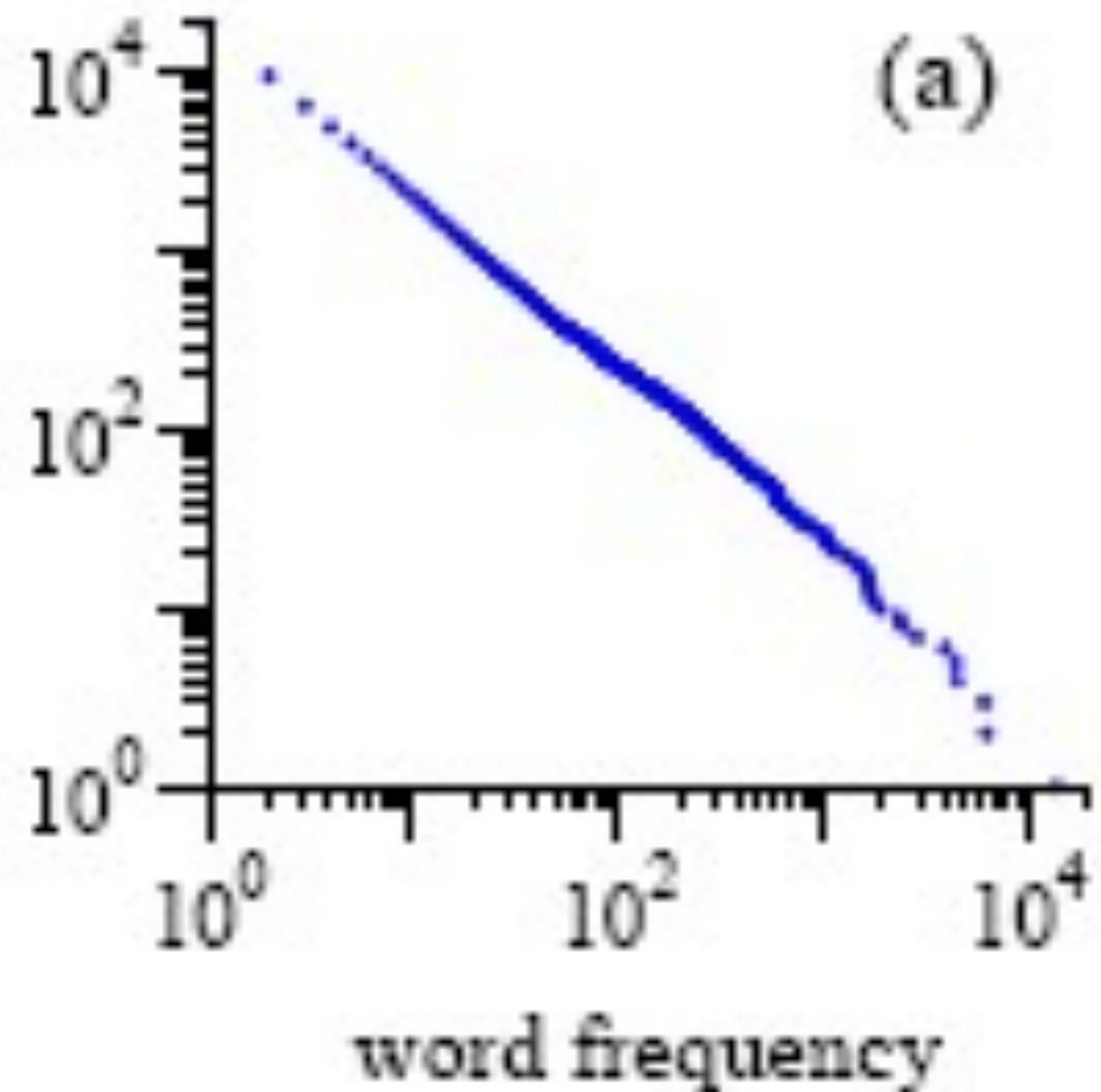


George Kingsley Zipf (1902-1950)



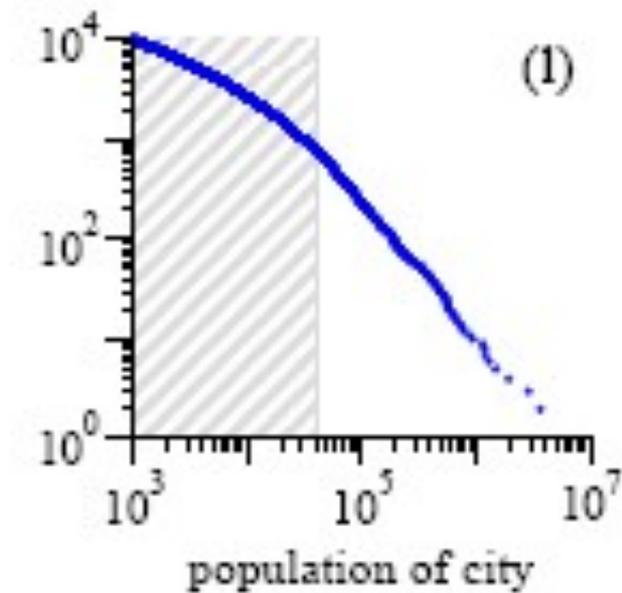
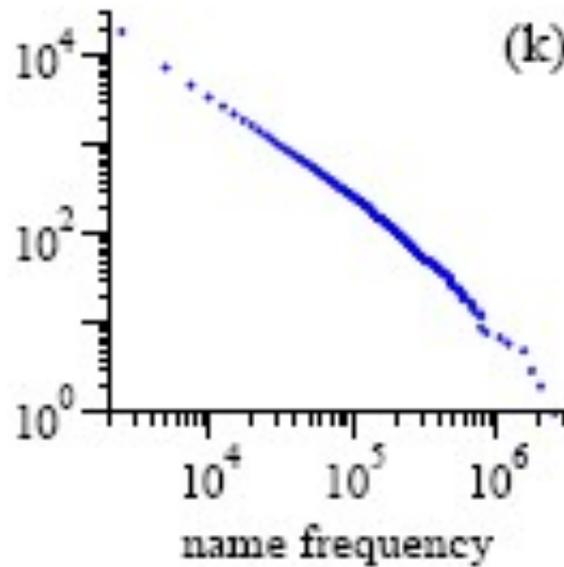
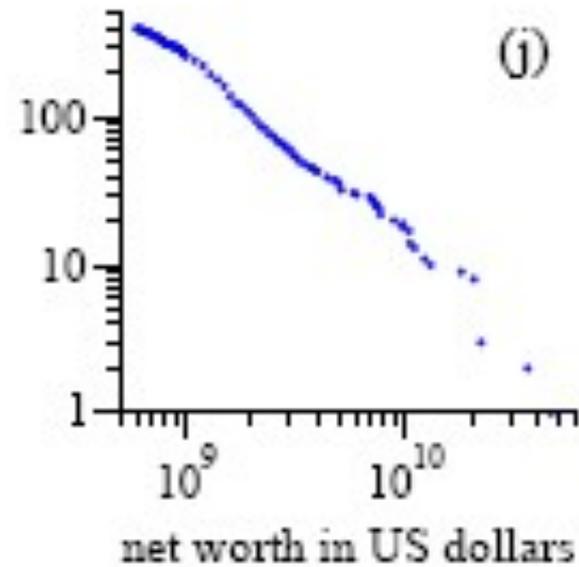
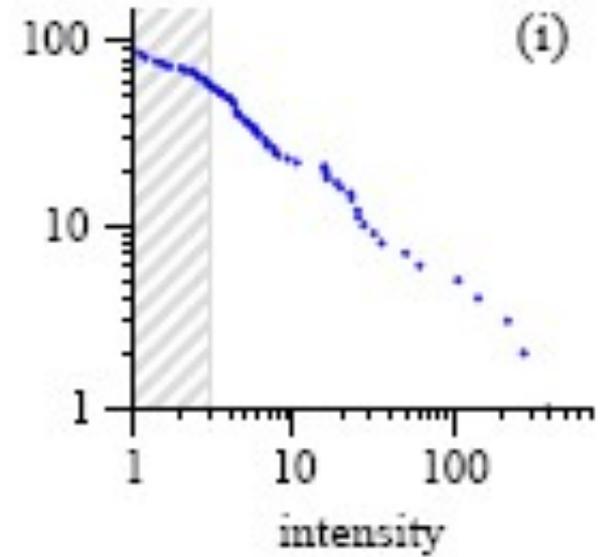
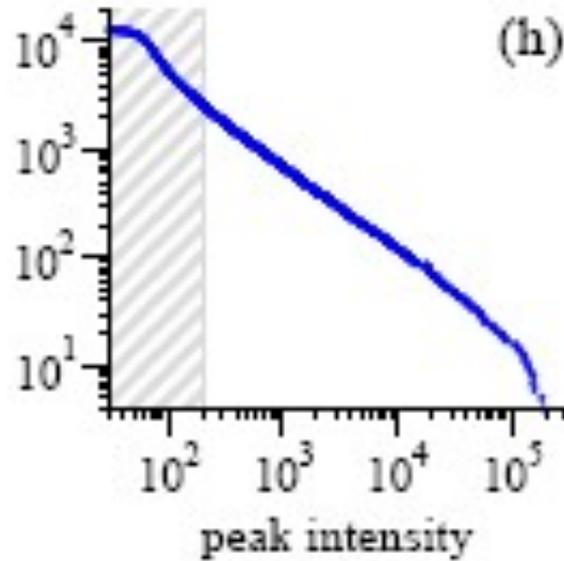
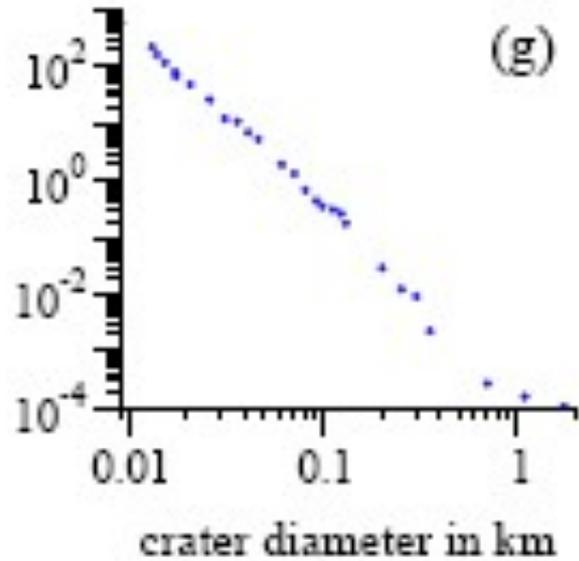
(1949): *Human behavior and the principle of least effort*

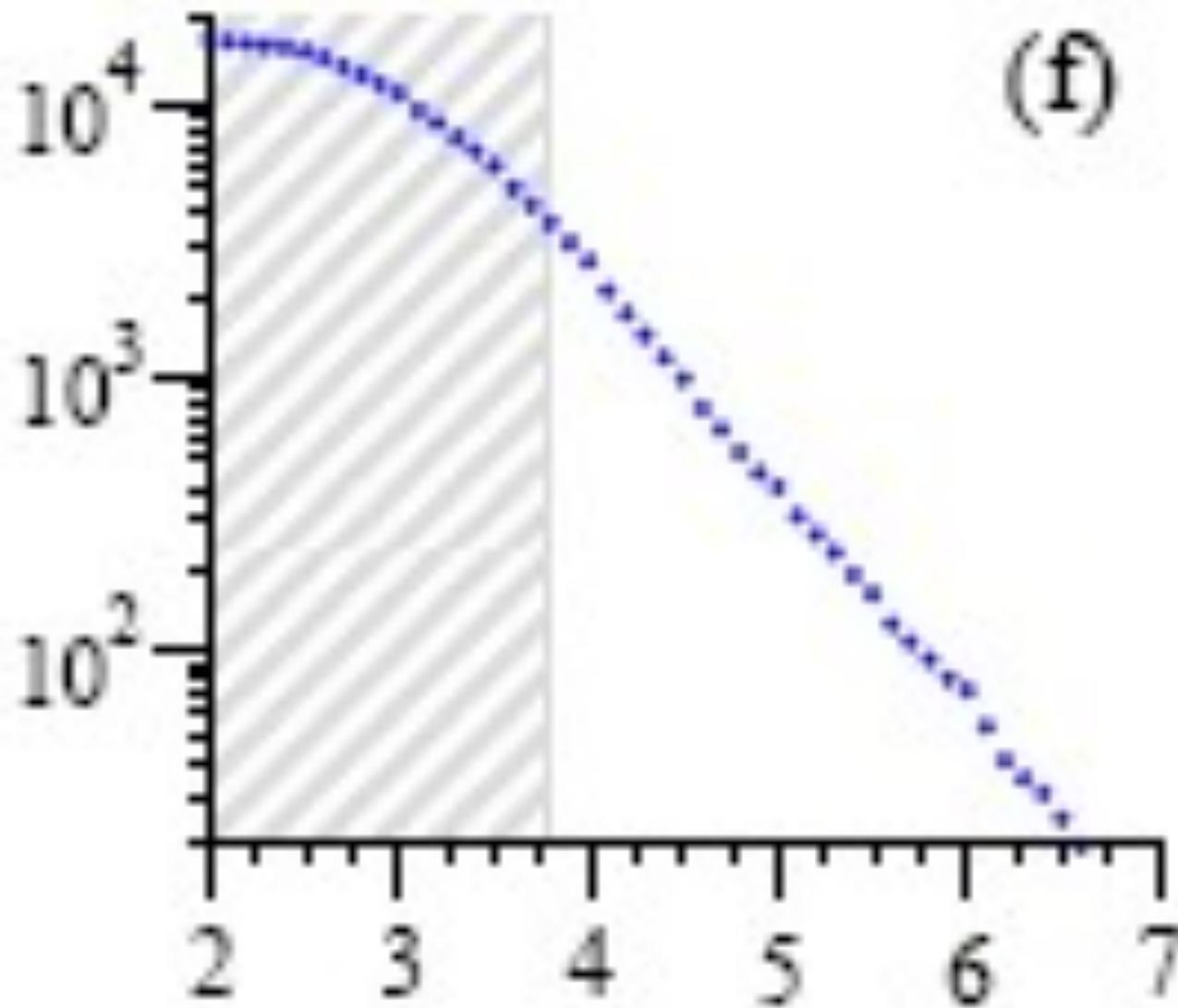
occurrences of words in the novel *Moby Dick*



Rank	Name	Population
1	New York City, New York	8,143,197
2	Los Angeles, California	3,844,829
3	Chicago, Illinois	2,842,518
4	Houston, Texas	2,016,582
5	Philadelphia, Pennsylvania	1,463,281
6	Phoenix, Arizona	1,461,575
7	San Antonio, Texas	1,256,509
8	San Diego, California	1,255,540
9	Dallas, Texas	1,213,825
10	San Jose, California	912,332
11	Detroit, Michigan	886,671
12	Indianapolis, Indiana	784,118
13	Jacksonville, Florida	782,623
14	San Francisco, California	739,426
15	Columbus, Ohio	730,657
16	Austin, Texas	690,252
17	Memphis, Tennessee	672,277
18	Baltimore, Maryland	635,815
19	Fort Worth, Texas	624,067
20	Charlotte, North Carolina	610,949

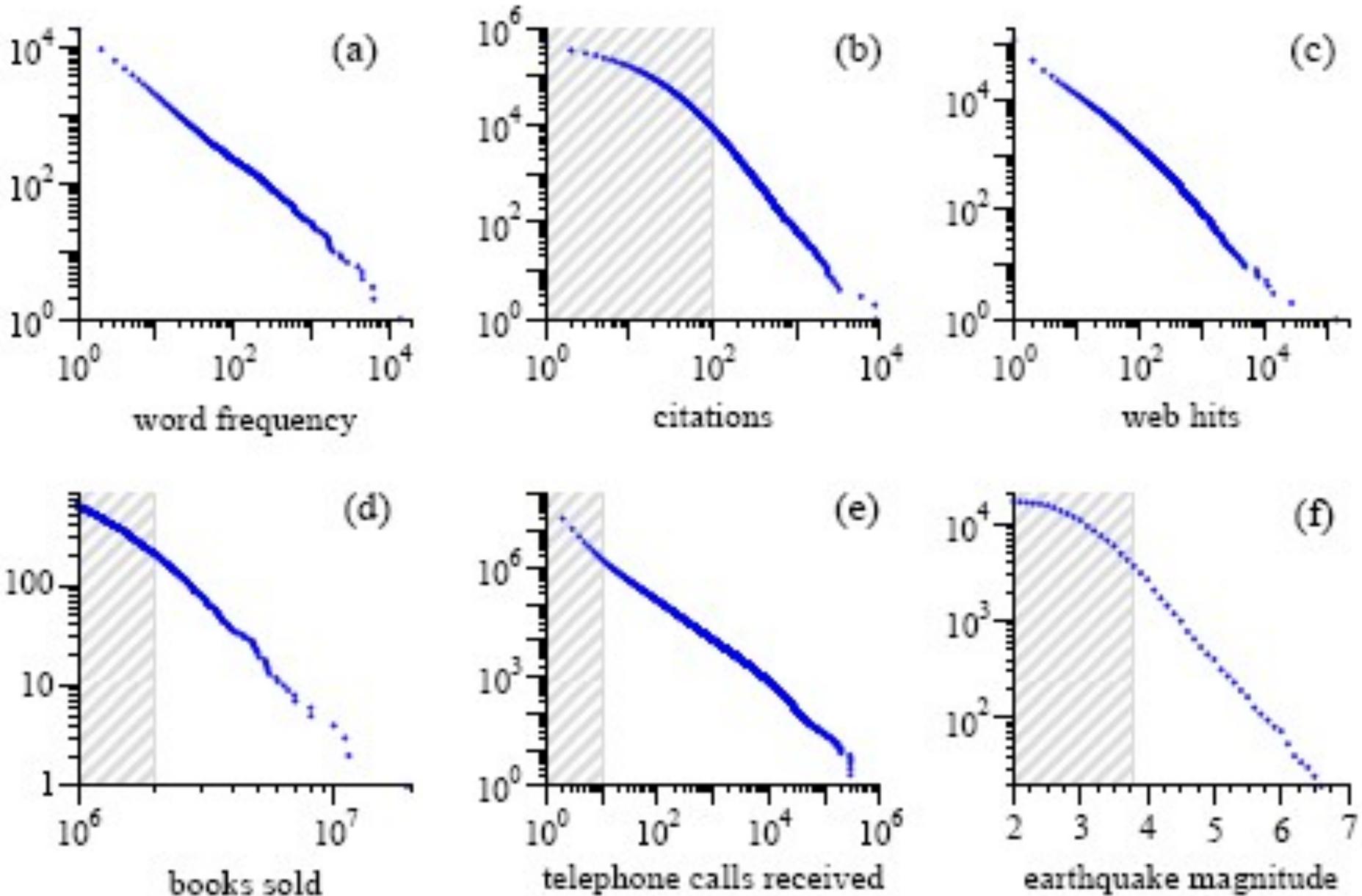
Universality---6 more examples





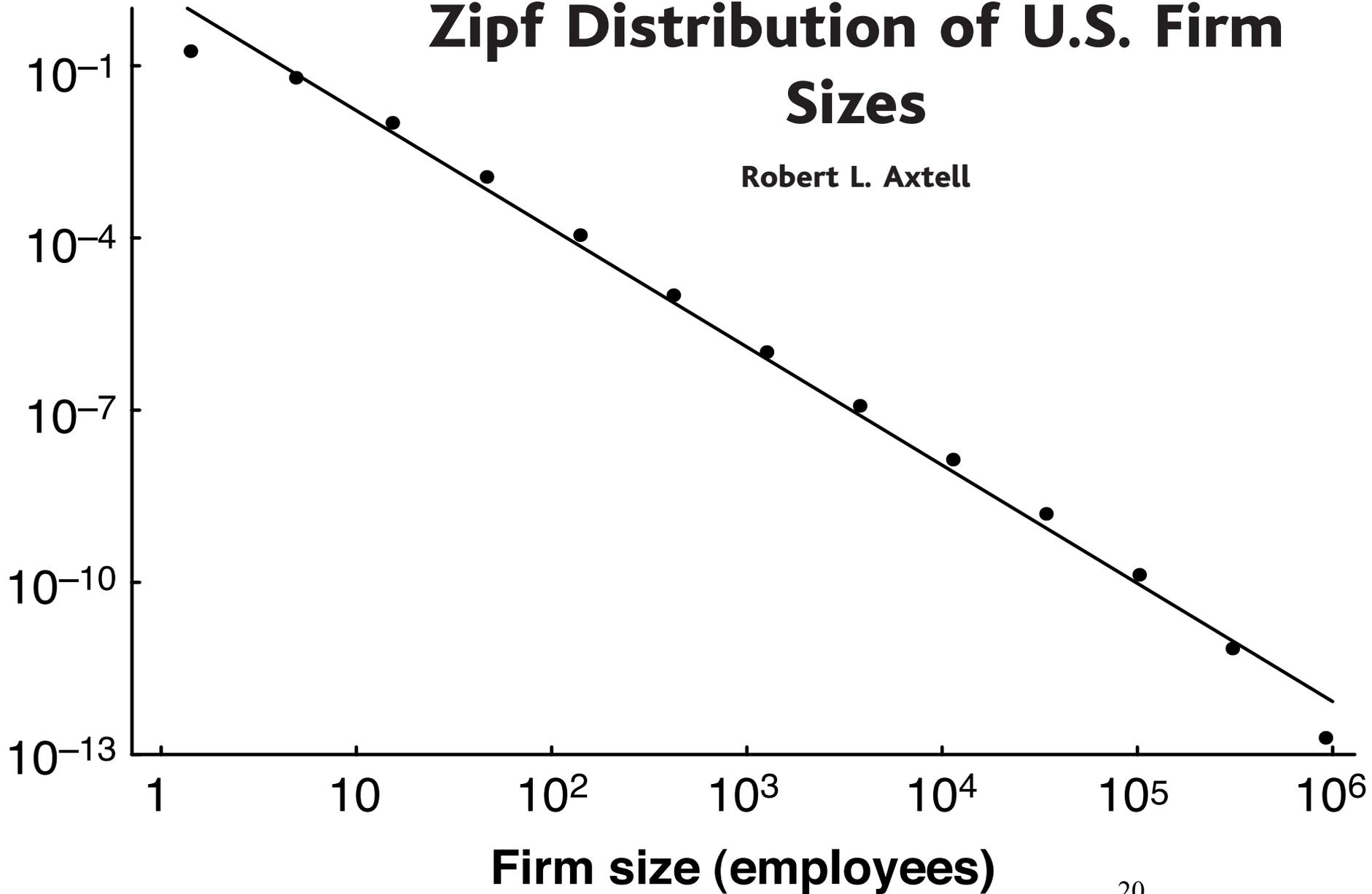
California earthquake magnitude

Universality—6 examples



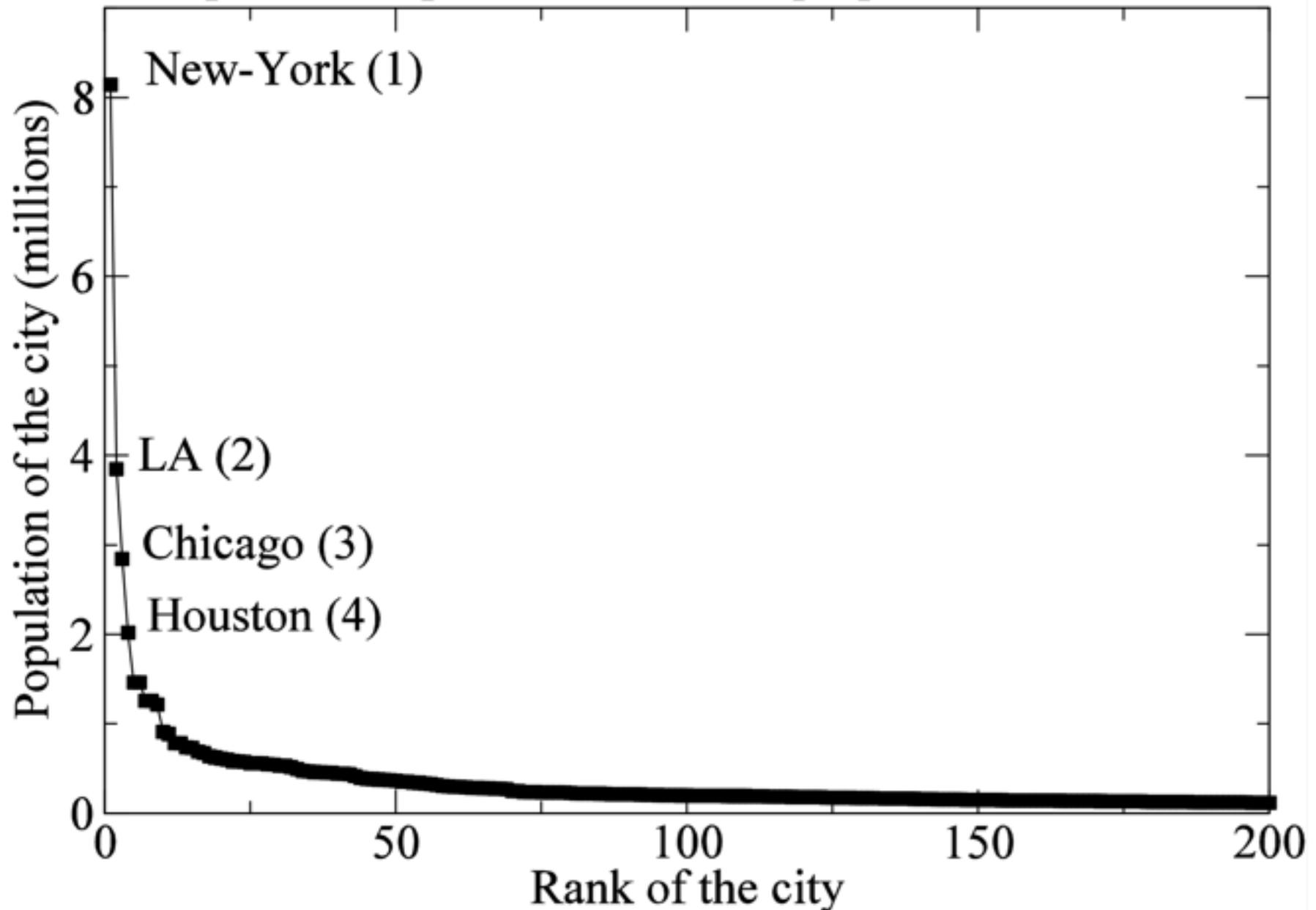
Zipf Distribution of U.S. Firm Sizes

Robert L. Axtell



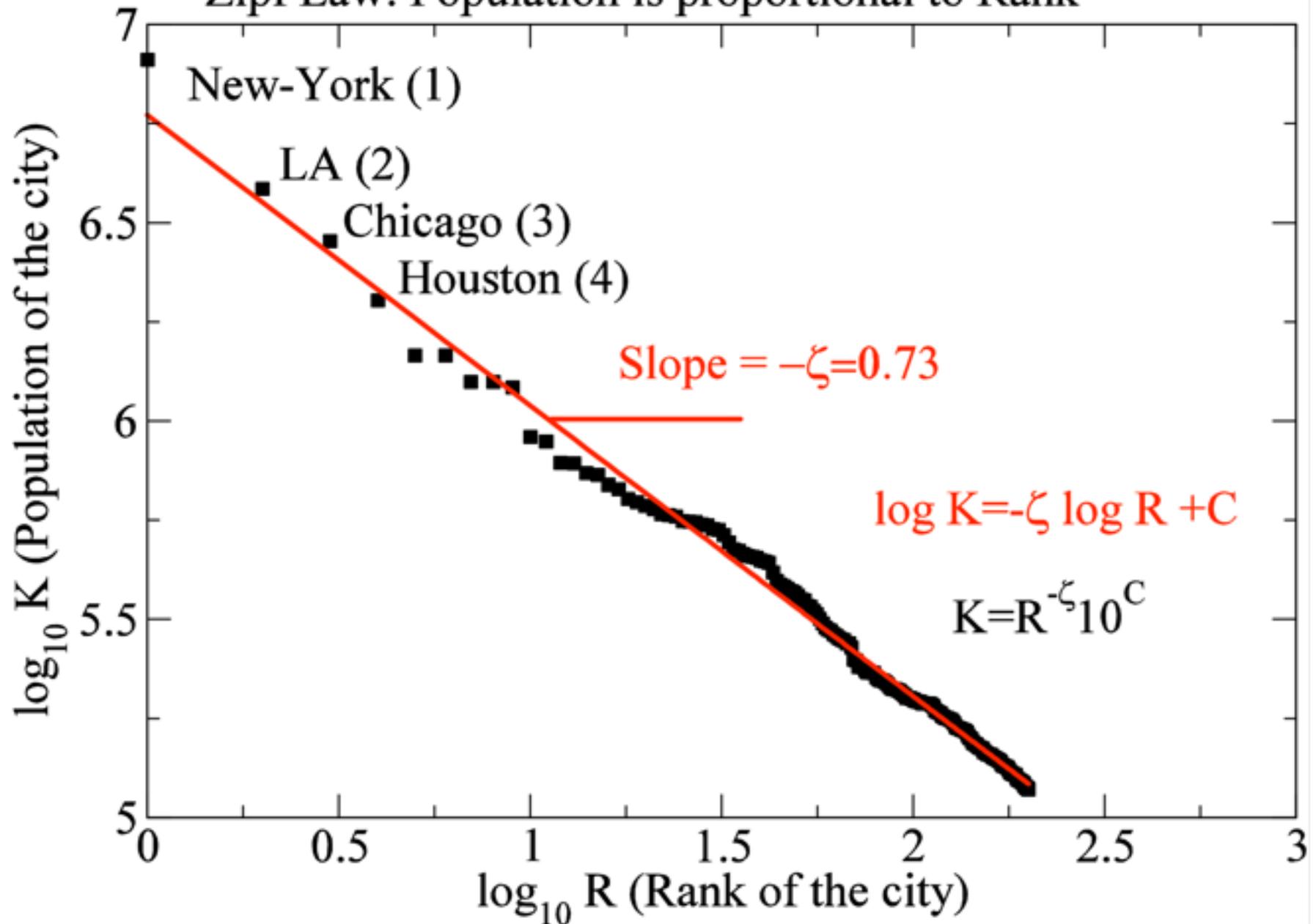
Population of 200 largest American cities

Zipf Law: Population is inverse proportional to Rank



Population of 200 largest American cities

Zipf Law: Population is proportional to Rank^{- ζ}



Elementary derivation of the Zipf law

Rules of the model:

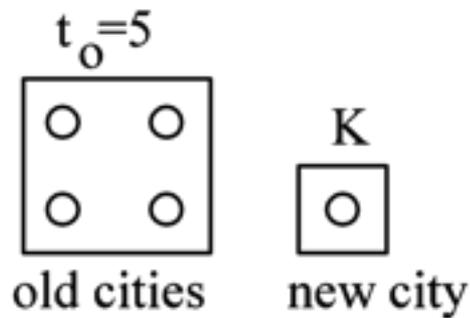
- At each time step a person is born in a city.
- All cities have approximately the same birth rate.
- With very small probability a person creates a new city.

Properties:

- The total population, n_0 , of the cities existing at time t_0 is proportional to t_0 : $n_0 \sim t_0$
- The rank of the city created at time t is proportional to t : $R \sim t_0$
- The ratio of the size of this city to the total population remains the same $K/n \sim 1/n_0 \Rightarrow K \sim 1/n_0 \sim 1/t_0$
- Finally: $K \sim 1/t_0 \sim 1/R \Rightarrow K \sim 1/R$

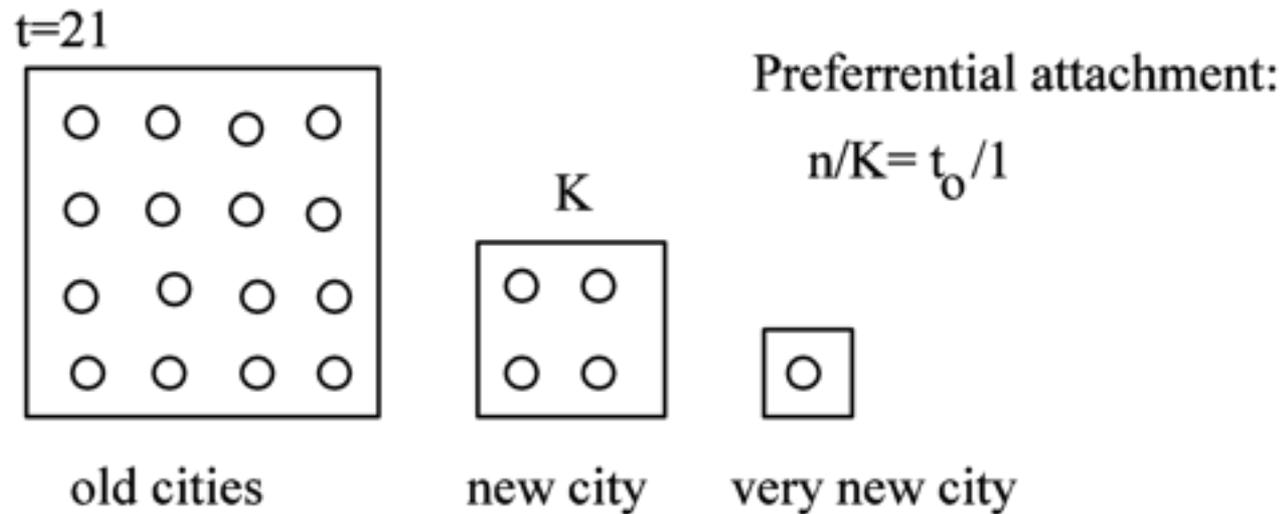
Conclusion:

- Size is inversely proportional to its rank.



t - time
 n - total population, $n=t$
 N - number of cities: $N=t$

In this example $b=1/5$, but we assume that b is very very small

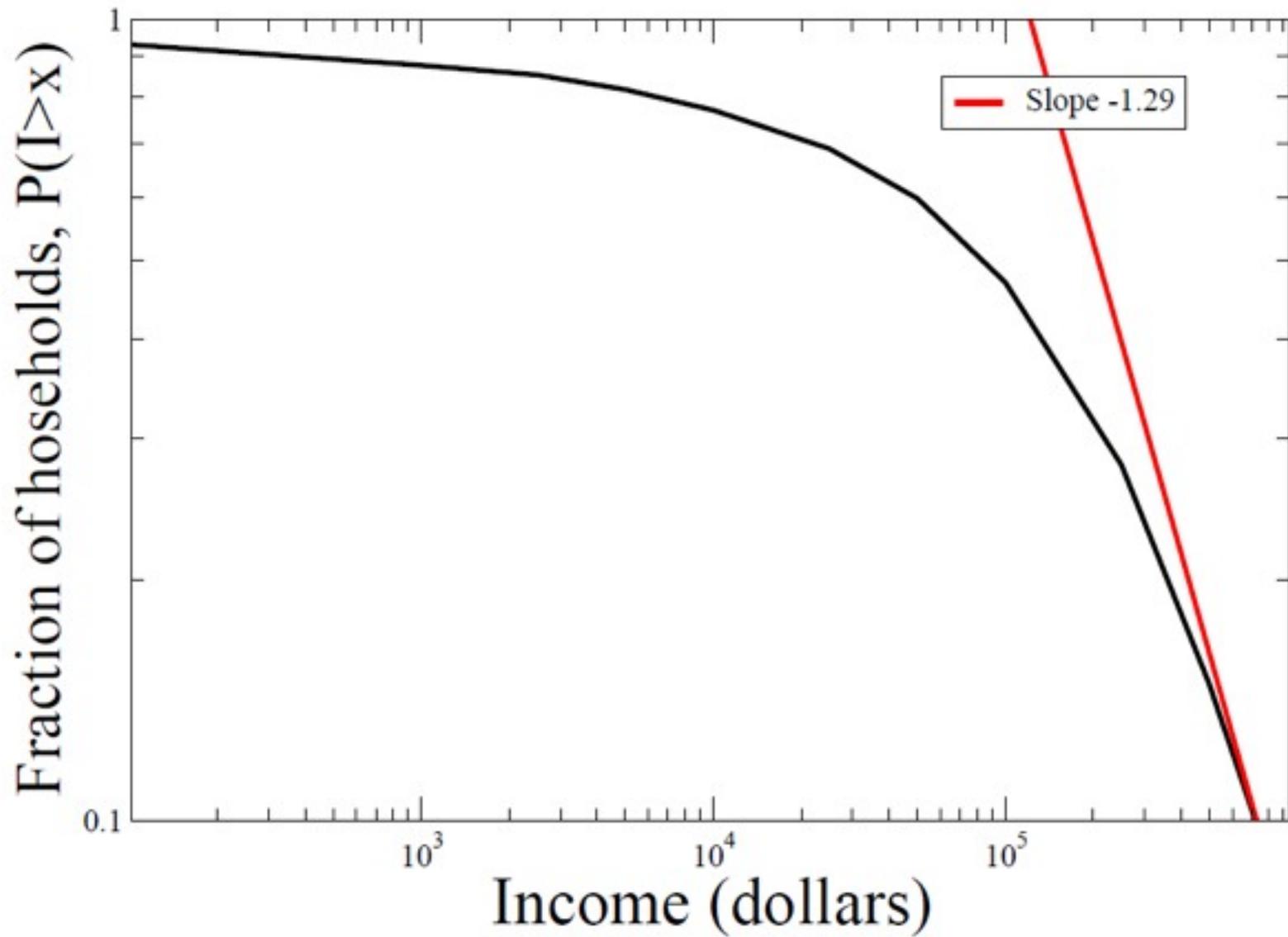


A city that born at time t_0 has rank $R=t_0 b$, it remains the same for the entire history
 Due to preferntial attachment its population remains $K=n/t_0$ (if we neglect newer cities)

Thus $t_0=R/b$ and $K=nb/R \Rightarrow$

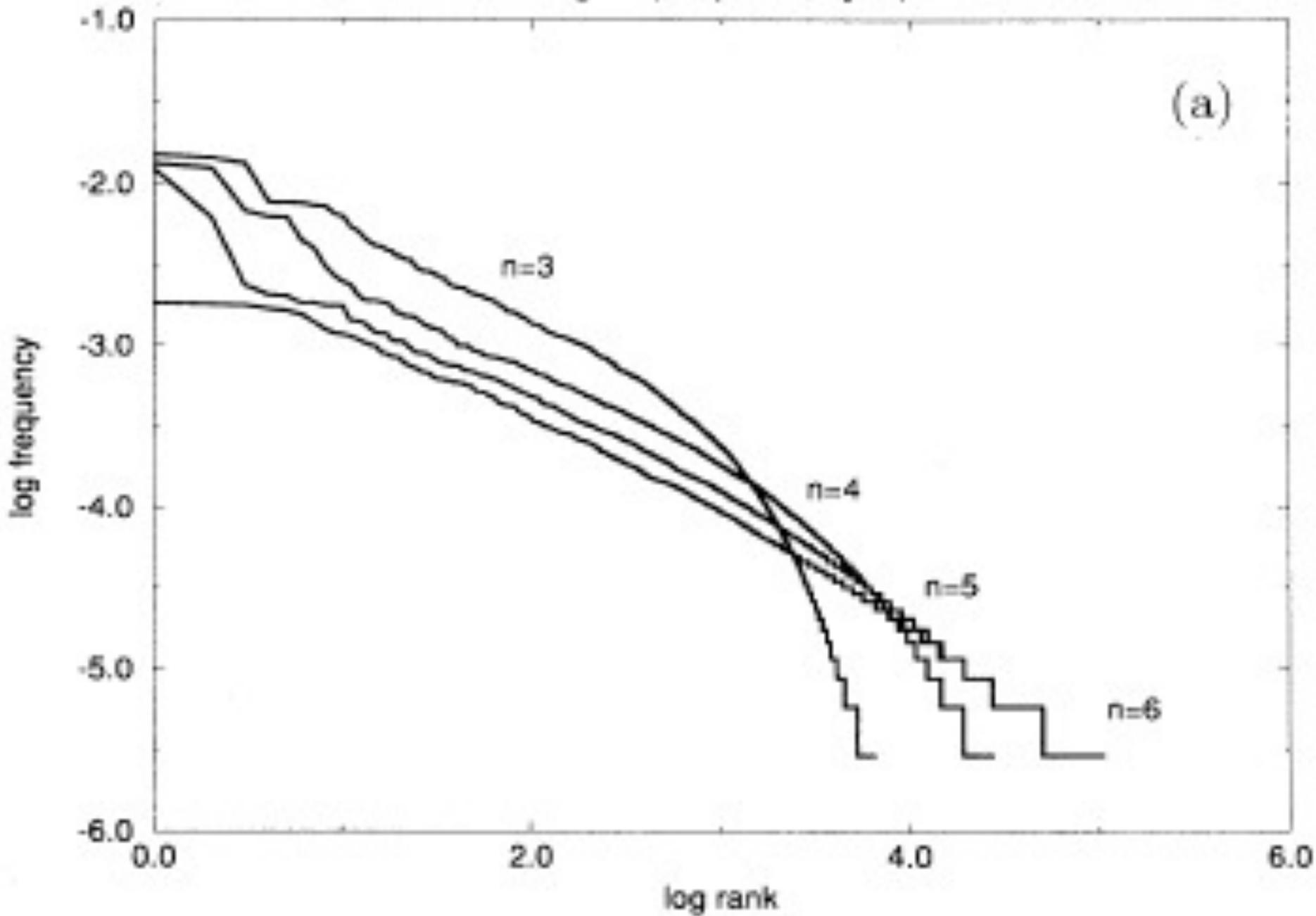
Population is inverse proportional to Rank

Distribution of annual income in the USA

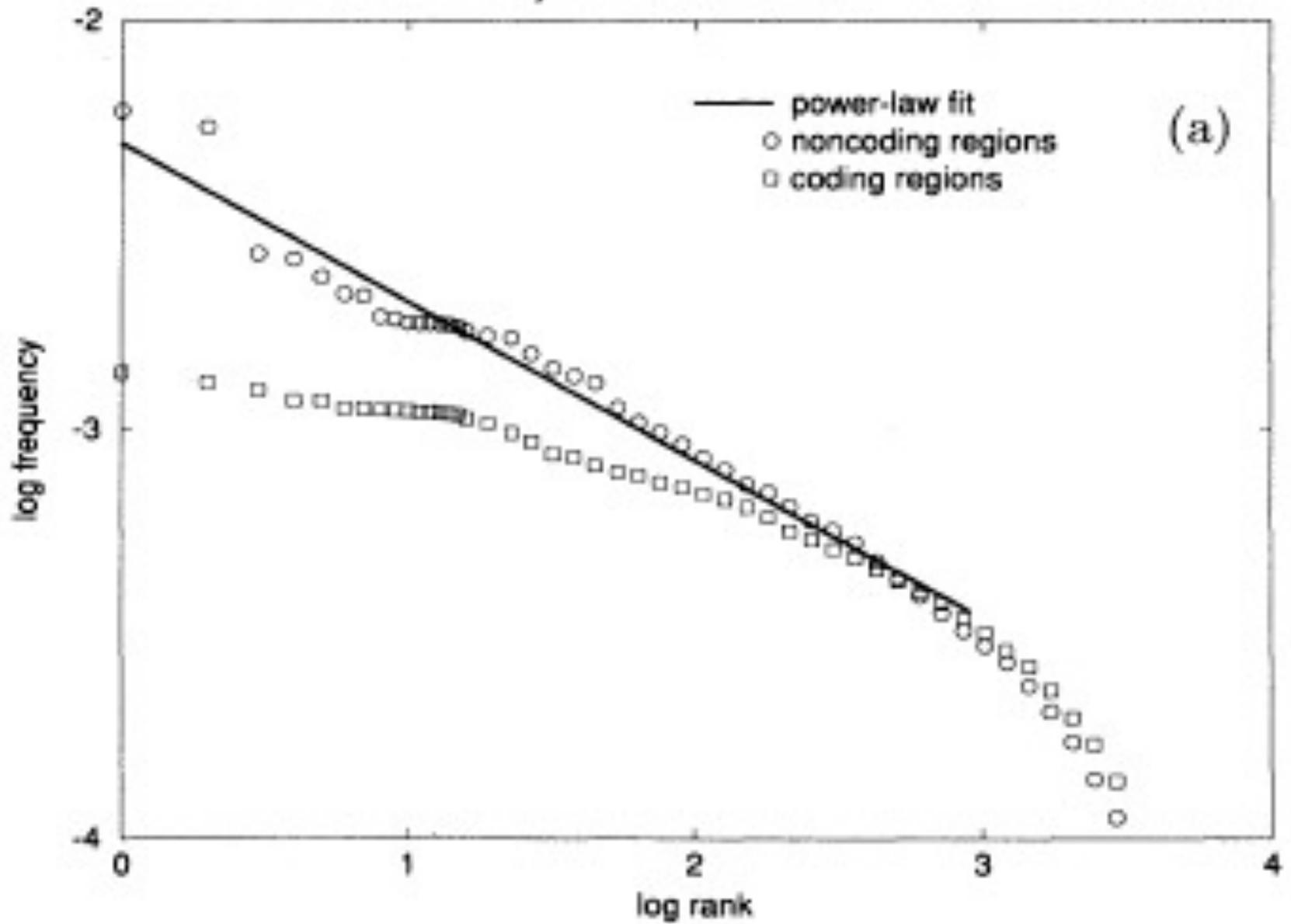


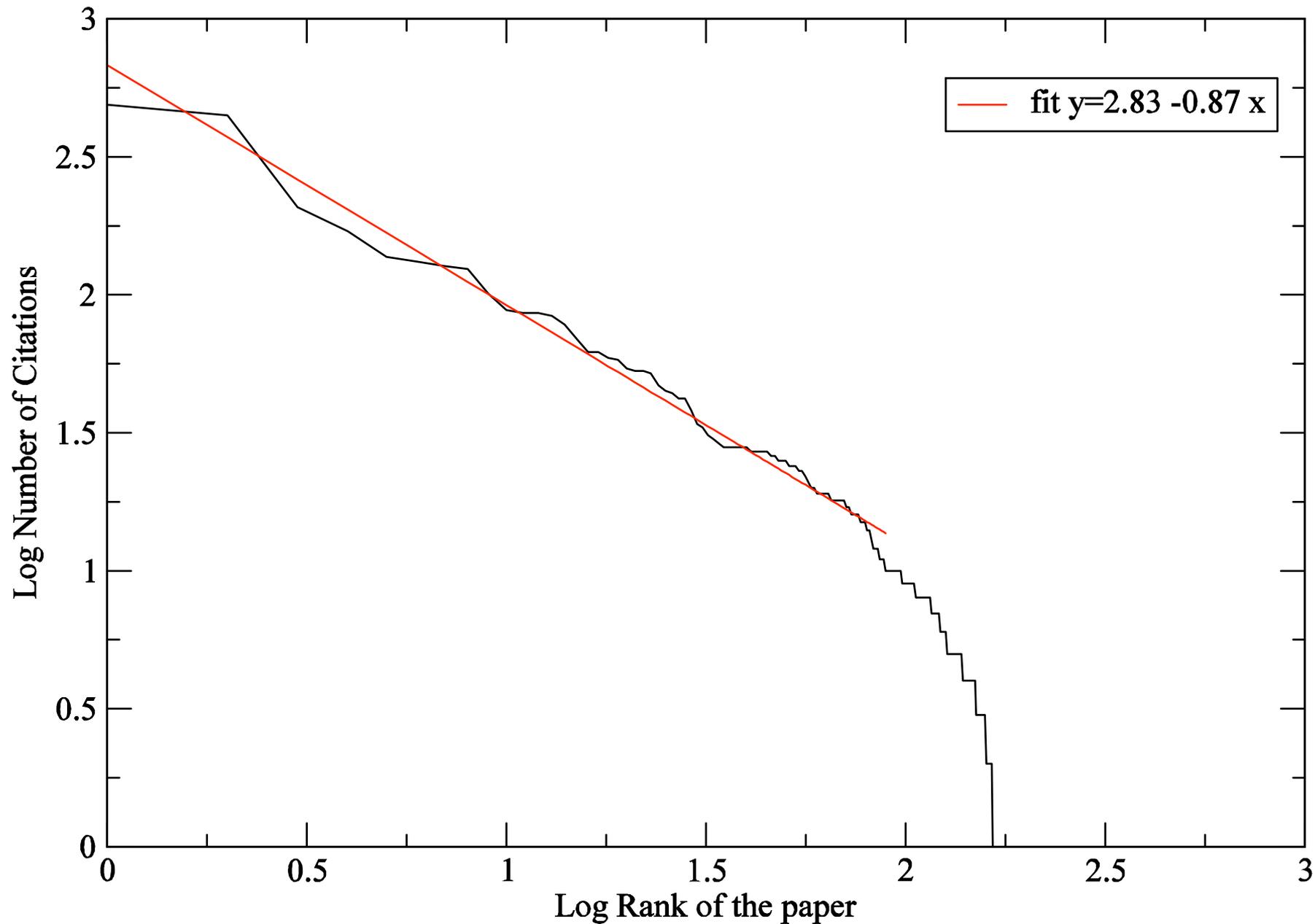
English (n-tuples analysis)

(a)

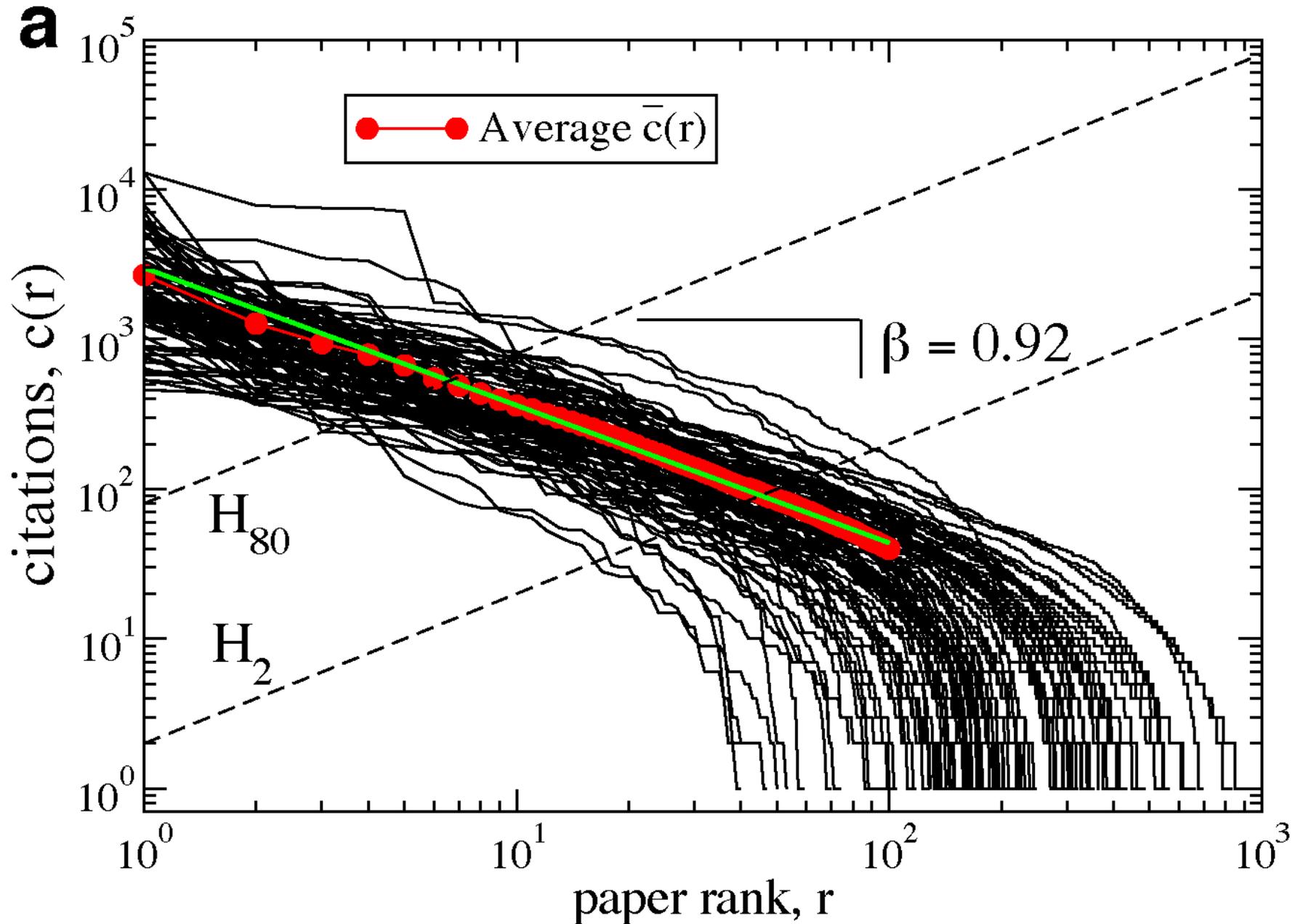


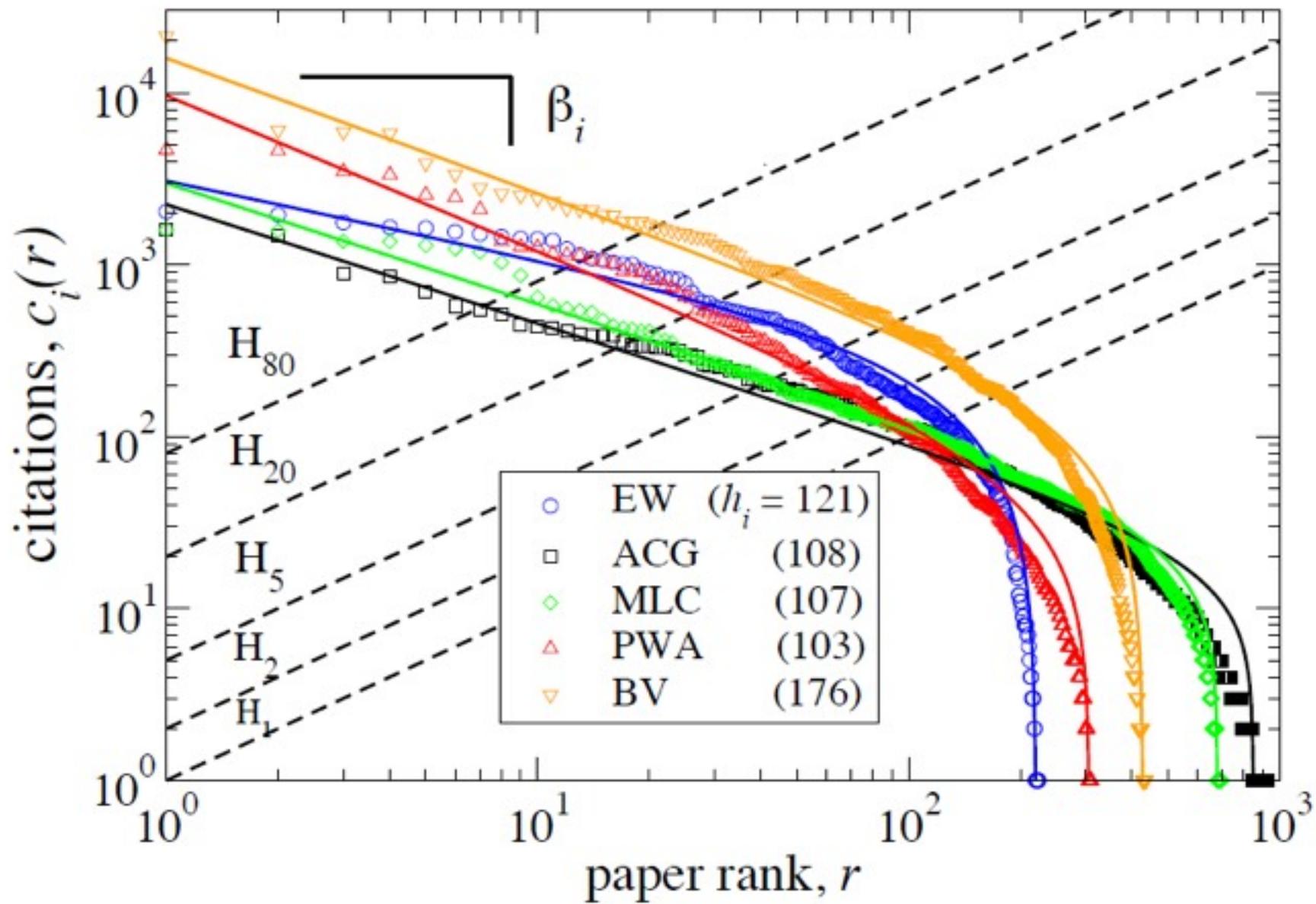
yeast chromosome XI



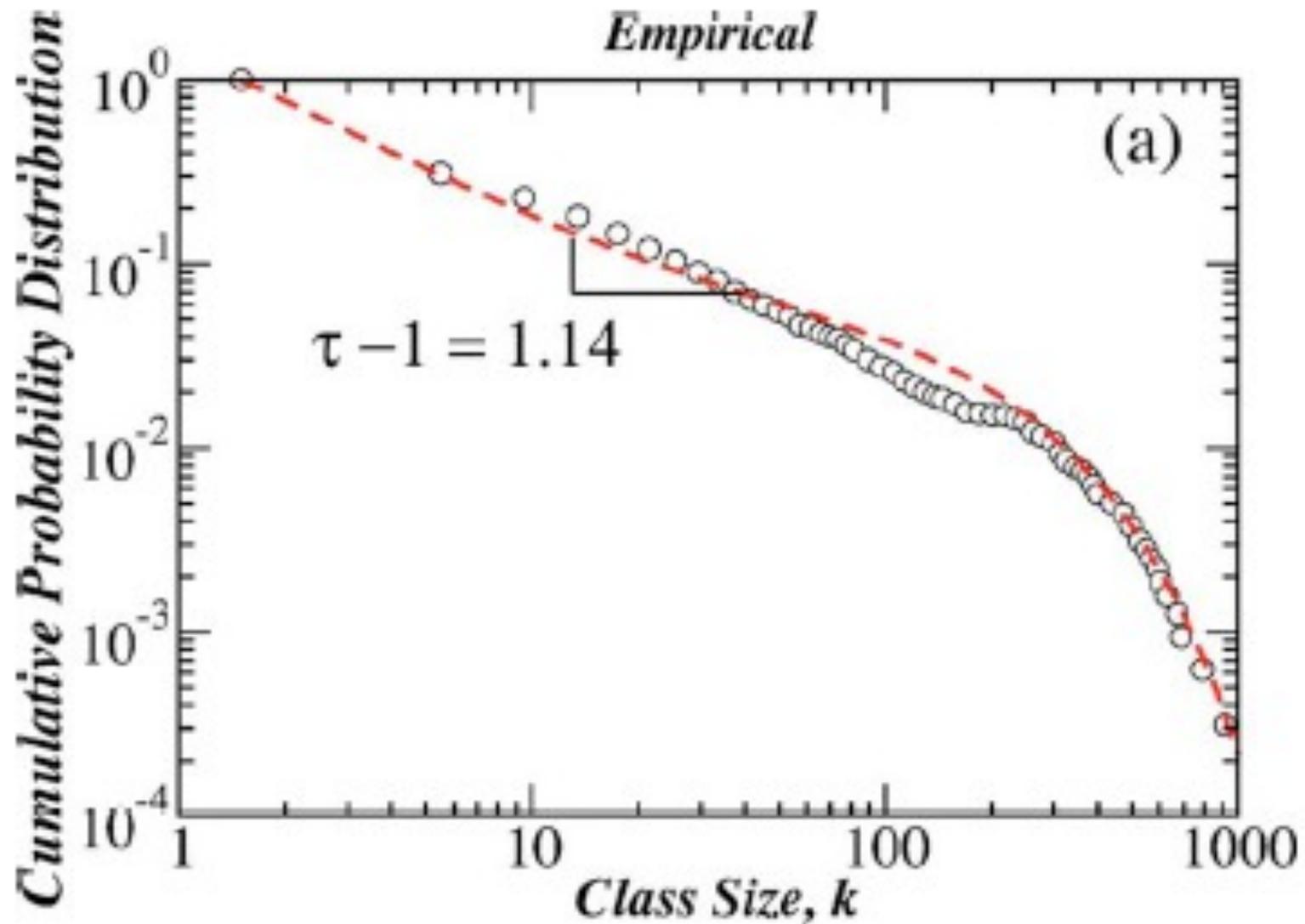


Rank-ordered citations of 100 physicists





Our goal is to build a simple model which would explain this graph



Preferential Attachment Model

Let us take a look on the distribution of citations of papers of a given author or population of cities. In our model we will call cities or papers classes and we will call people and citations units.

Let us assume that in a unit of time

(1) existing classes get λ new units, which are distributed to the existing classes in proportion to their existing size measured in number of units.
and

(2) β new classes, each of unit size are created. We introduce

$N(t)$ - number of classes as function of time;

$n(t)$ - number of units as function of time;

$N_0 = N(0)$ - initial number of classes at $t = 0$.

$n_0 = n(0)$ - initial number of units at $t = 0$.

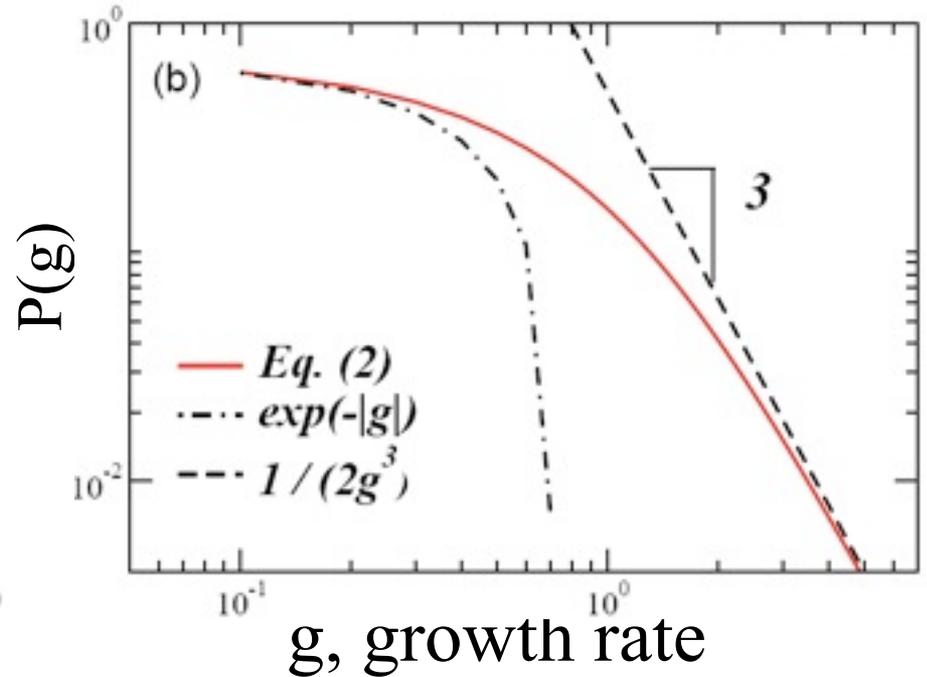
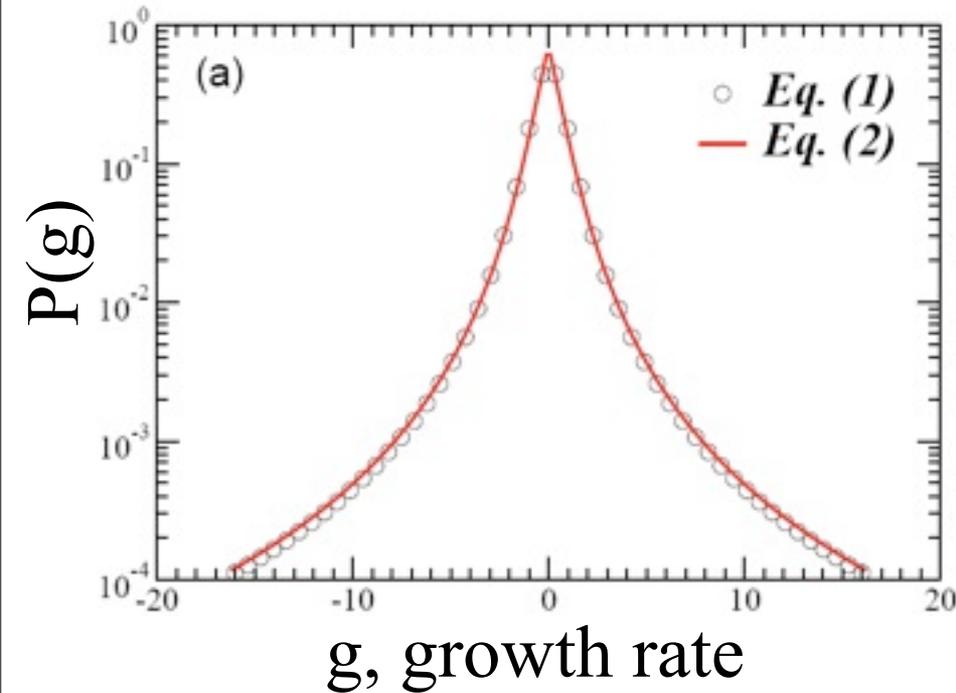
Then

$$\frac{dN}{dt} = \beta \quad \text{and} \quad \frac{dn}{dt} = \beta + \lambda. \quad (1)$$

Integration gives:

$$N(t) = \beta t + N_0 \quad \text{and} \quad n(t) = (\beta + \lambda)t + n_0. \quad (2)$$

Crossover in $P(g)$ from Exp. to Power Law



$P(g)$ same as $P_{\text{old}}(n)$ and $P_{\text{new}}(n)$.

1. for small g ,
 $P(g) \approx \exp[-|g| (2 / V_g)^{1/2}]$.
2. for large g , $P(g) \sim g^{-3}$.

Behavior of the distribution for large time intervals

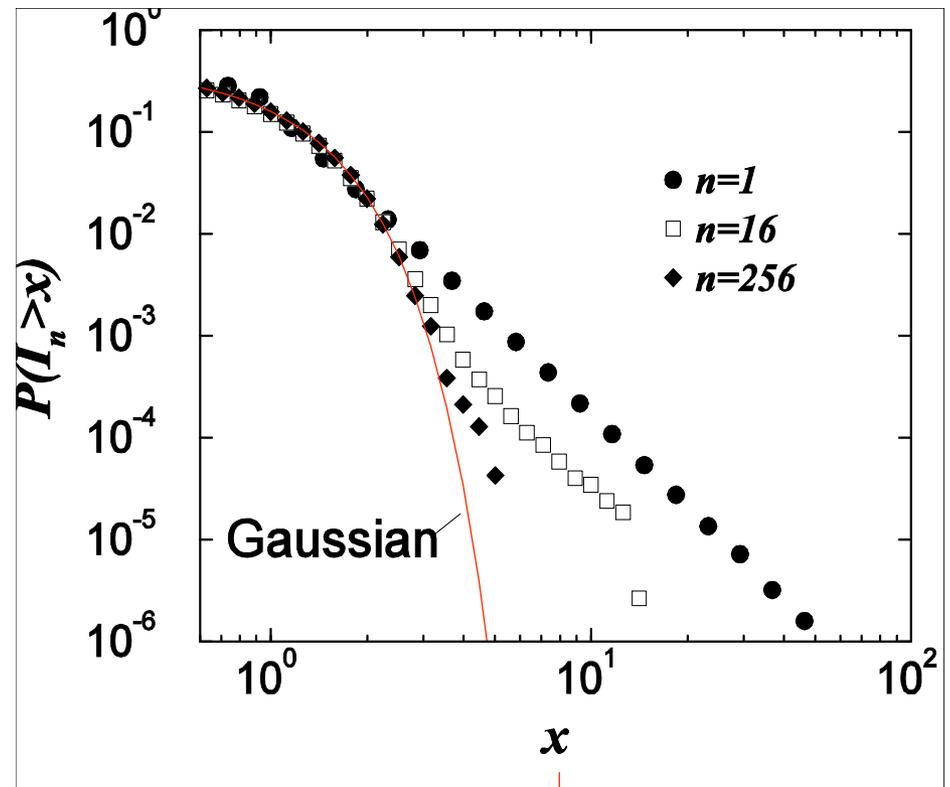
- Since

$$r_{N\Delta t} = \sum_{i=1}^N r_{\Delta t}^i$$

- We expect by Central Limit Theorem that $P(R)$ for larger times to converge to a Gaussian
- Indeed a generated power law distribution with the same exponent ζ_R converges quickly to Gaussian under aggregation. Consider

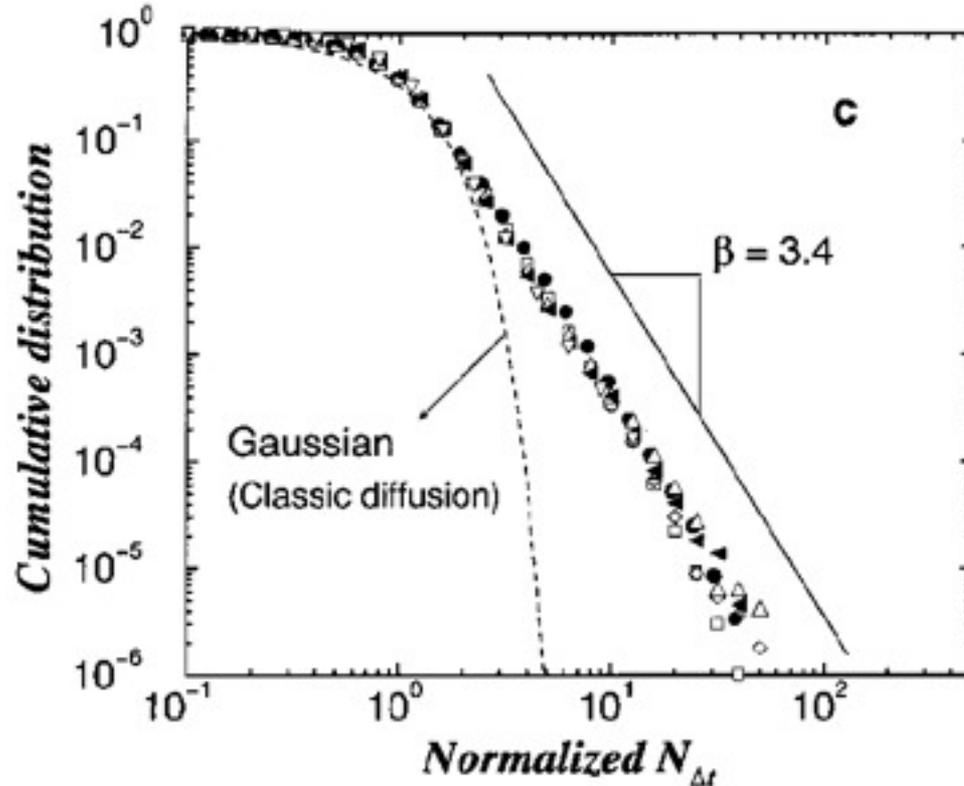
$$I_n = \sum_{i=1}^n x_i$$

Convergence of generated iid variables



Do large returns arise from large market activity ?

For this to be possible, since $R \sim \varepsilon \sigma \sqrt{N}$ we expect $\zeta_N = \frac{\zeta_R}{2}$



In sharp contrast, we find:

$$\zeta_N \approx 3$$

too large to explain

$$\zeta_R = 3$$

Fluctuations in market activity too mild to explain fat tails of returns.

Take home message

- P(growth rate) Laplace in Center: universal
- Width decreases as $-1/6$ power of size bin
- P(growth rate) crosses over to power law in wings
- No theory for $-1/6$ power law for width
- Theory (Buldyrev et al) for growth rate power law

<http://polymer.bu.edu/hes> (PDF of published papers)

REST OF SLIDES NOT USED

Data analyzed (Gopikrisnan/Plerou/Liu/...)

Trades and Quotes (TAQ) database

- 2 years 1994-95
- 1000 stocks largest by market cap on Jan 1, '94 (200 million records)

To test “universality”, also analyze other databases, including:

Center for Research in Security Prices (CRSP) database

- 35 years 1962-96
- approximately 6000 stocks

Tick data for the London Stock Exchange

- 2 yrs 2000-01
- 250 stocks.

Transactions data from the Paris Bourse

- 30 stocks; 1994-95

After-Dinner Drink: theory/model?

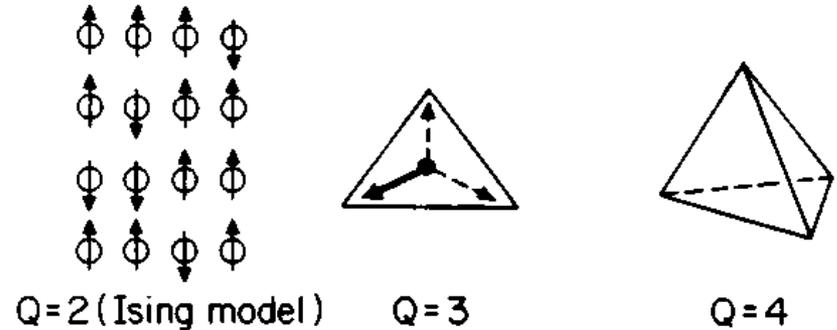
Each stock is a unit, interacting with other stocks (units). This type of model studied in statistical physics.

Typical models:

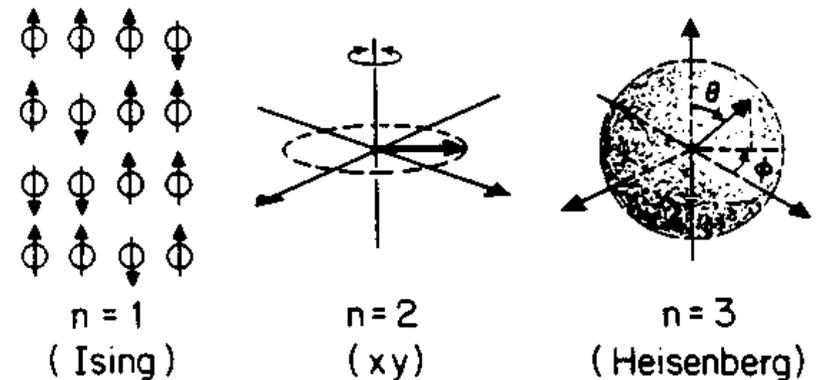
1. set of units, each of which can be in Q different states (POTTS MODEL).

2. set of n -dimensional units, each of which can be in a continuum of states (n-VECTOR MODEL)

(a) Potts Model:



(b) n-Vector model:



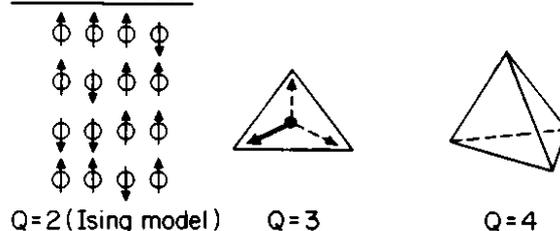
(c)

Pillar 2, Universality (Universality classes):

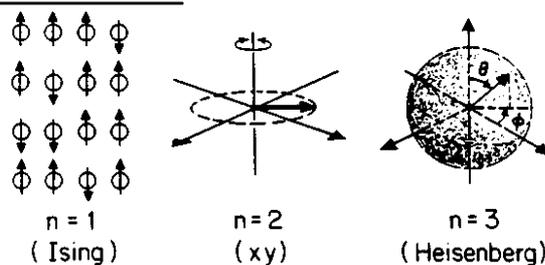
Experimental fact:

A wide range of magnetic materials belong to one of two families of “Universality classes”: the Q -state Potts model (Potts 1952) and the n -vector model (HES 1968). The purely geometric phase transition “percolation” corresponds to the limit $Q=1$, while the self-avoiding random walk corresponds to $n=0$.

(a) Potts Model:



(b) n -Vector model:



(c)

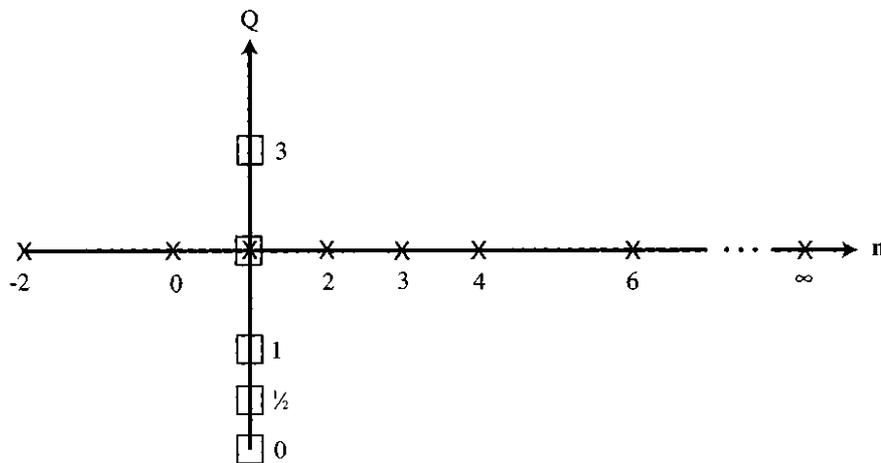
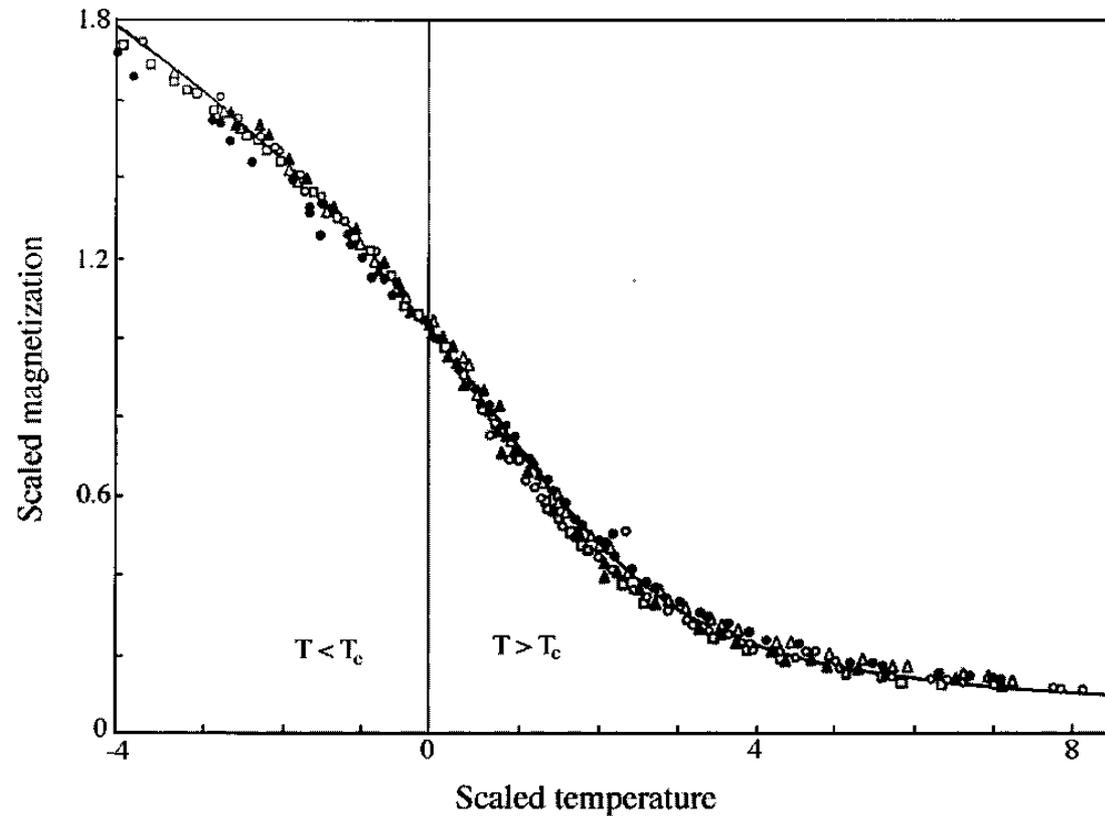


FIG. 2. Schematic illustrations of the possible orientations of the spins in (a) the s -state Potts model, and (b) the n -vector model. Note that the two models coincide when $Q=2$ and $n=1$. (c) North-south and east-west “Metro lines.”

Pillar 1 (continued):

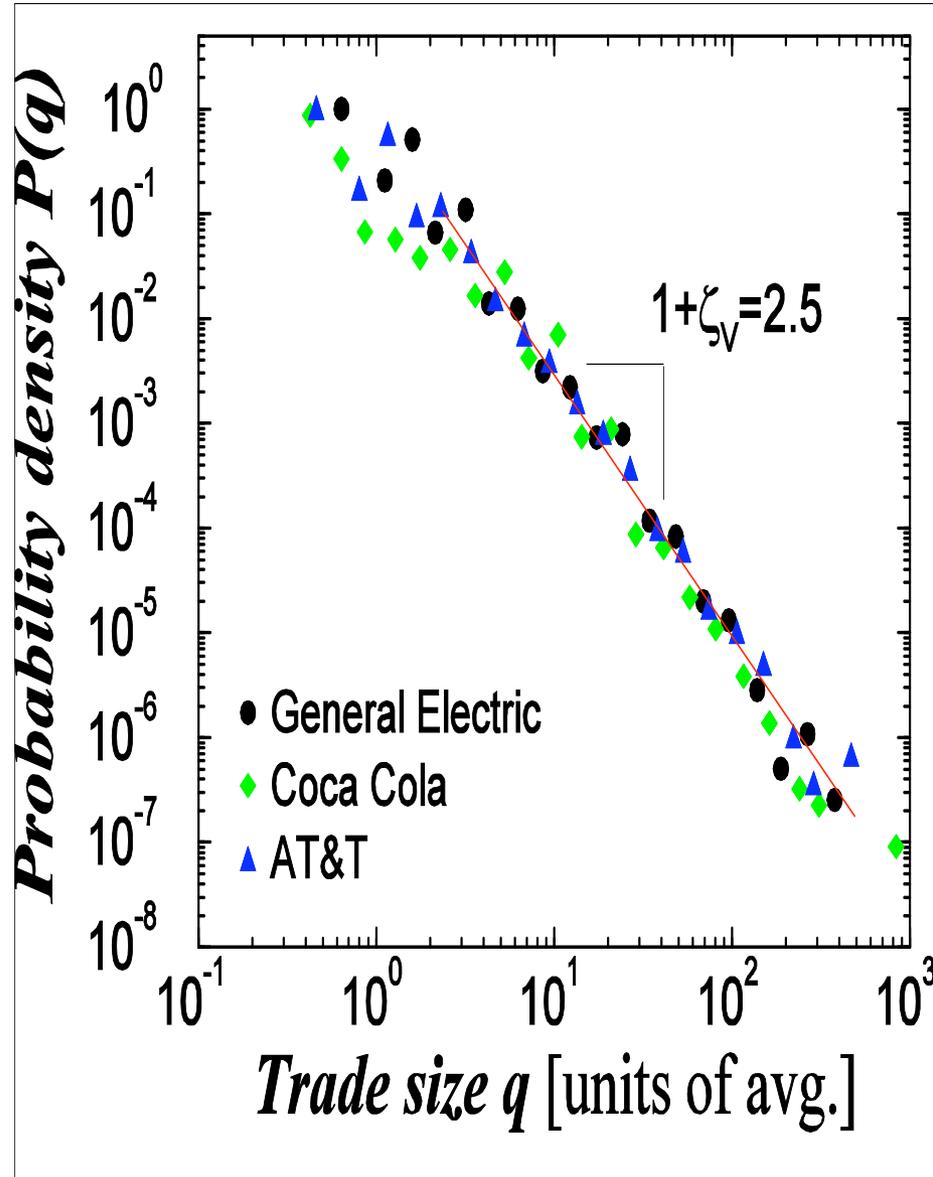
Experimental test of data collapse (Pillar 1): Equation of State for 5 different magnets near their respective critical points.



Pillar 2: Universality

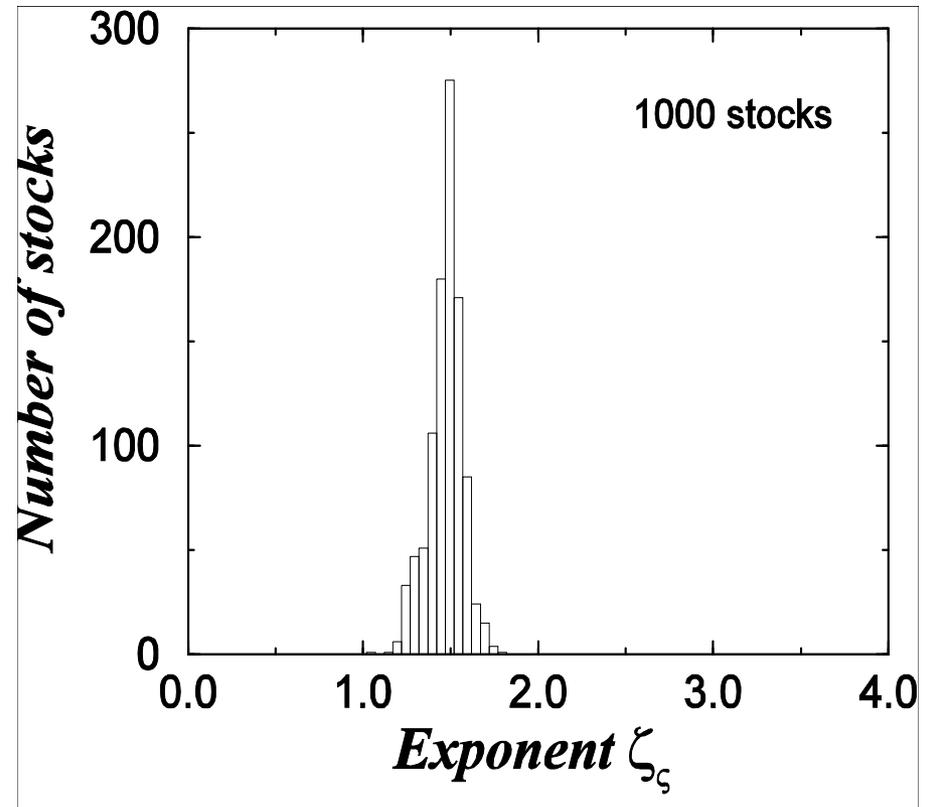
First hint: all 5 magnets have same scaled equation of state.

FIG. 1. Experimental *MHT* data on five different magnetic materials plotted in scaled form. The five materials are CrBr_3 , EuO , Ni , YIG , and Pd_3Fe . None of these materials is an idealized ferromagnet: CrBr_3 has considerable lattice anisotropy, EuO has significant second-neighbor interactions. Ni is an itinerant-electron ferromagnet, YIG is a ferrimagnet, and Pd_3Fe is a ferromagnetic alloy. Nonetheless, the data for all materials collapse onto a single scaling function, which is that calculated for the $d=3$ Heisenberg model [after Milošević and Stanley (1976)].



$$P(V > x) \sim x^{-\zeta_V}$$

$$\zeta_V \approx \frac{3}{2}$$



Pillar 3: RENORMALIZATION GROUP

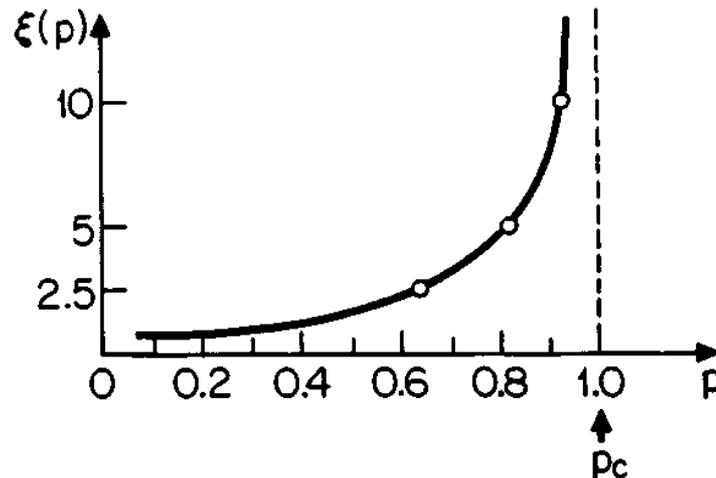
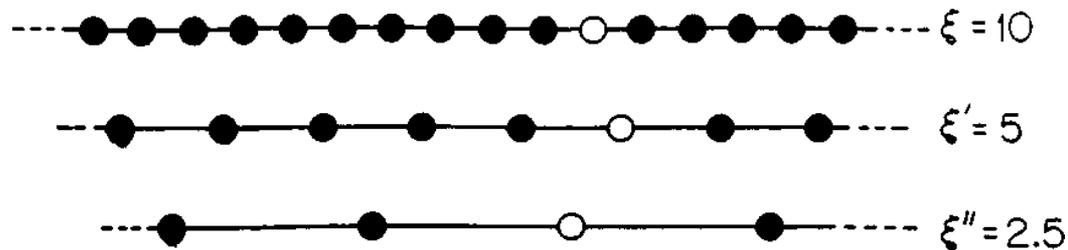
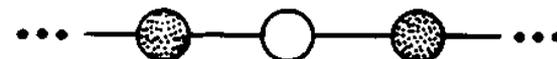
Kadanoff site-to-cell coarse-graining successively tames the problem of an infinite correlation length.

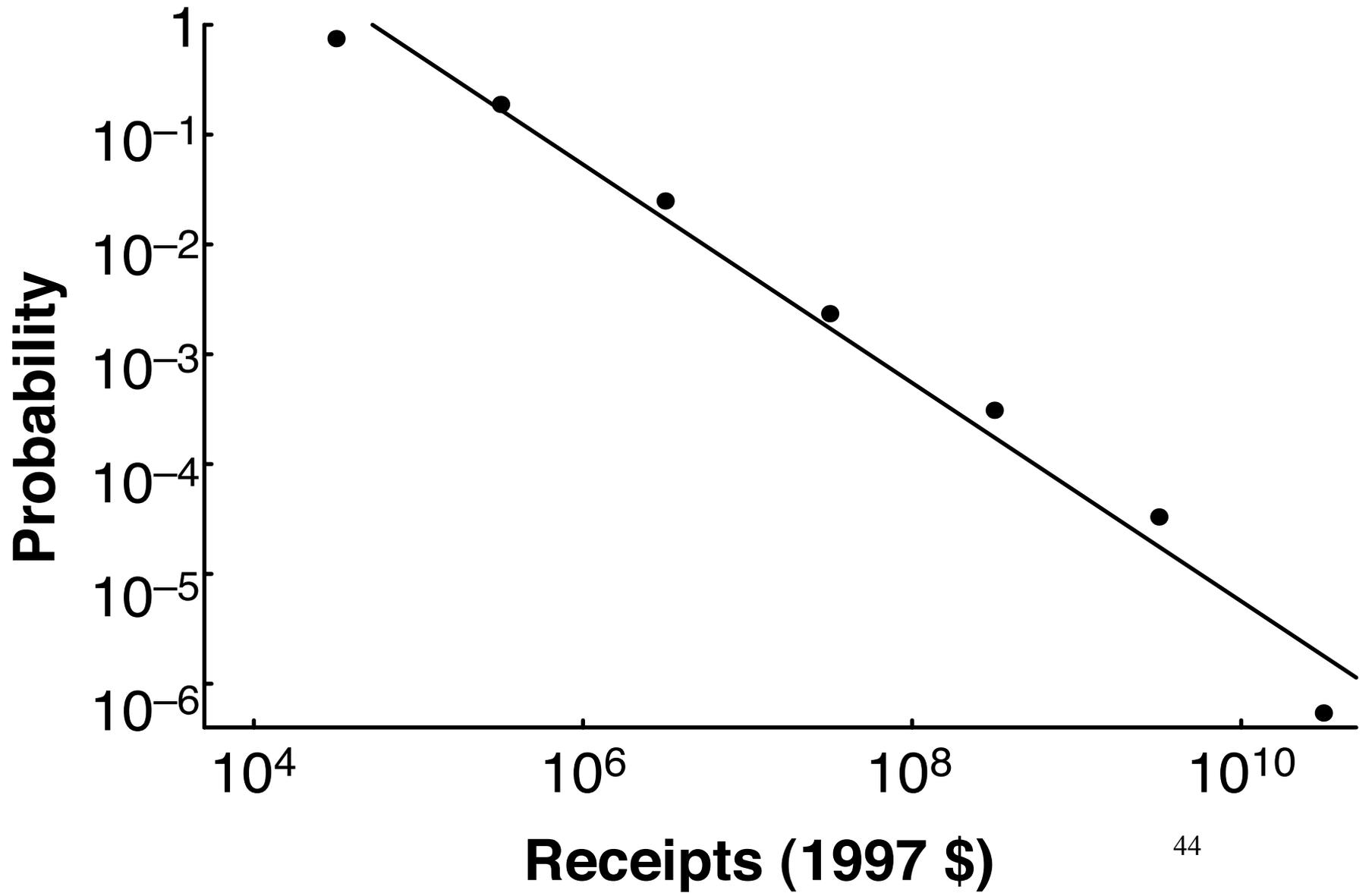
Example: $Q=1$ Potts model for $d=1$ (1d percolation)

(a) Site level [occupation probability = p]



(b) Cell level [occupation probability p']





Population of N=200 largest American cities Relation between rank and probability distribution

