Statistics of innovation: how statistics appears in search processes -all of it- from Gauss to Zipf

Stefan Thurner

MEDICAL UNIVERSITY OF VIENNA

SANTA FE INSTITUTE

IIASA

COMPLEXITY SCIENCE HUB VIENNA

www.complex-systems.meduniwien.ac.at www.santafe.edu

with Bernat Corominas-Murtra and Rudolf Hanel

BCM, RH, ST, PNAS 112 (2015) 5348-5353

BCM, RH, ST, New J Physics 18 (2016) 093010

BCM, RH, ST, J Roy Soc Interface 12 (2016) 20150330

ST, BCM, RH, Phys Rev E (2017) in press

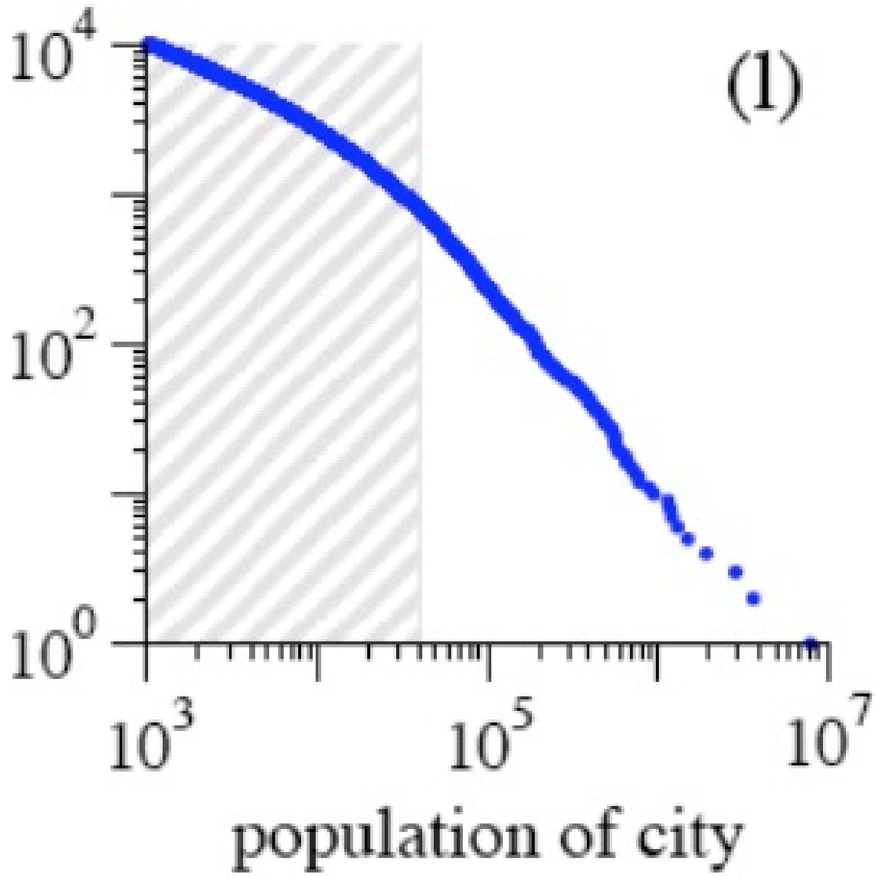
BCM, RH ST, Sci Rep (2017) in press

BCM, RH, ST, arxiv.org/1706.10202

Power laws are pests

- they are everywhere
- its hard to control them
- you never get rid of them

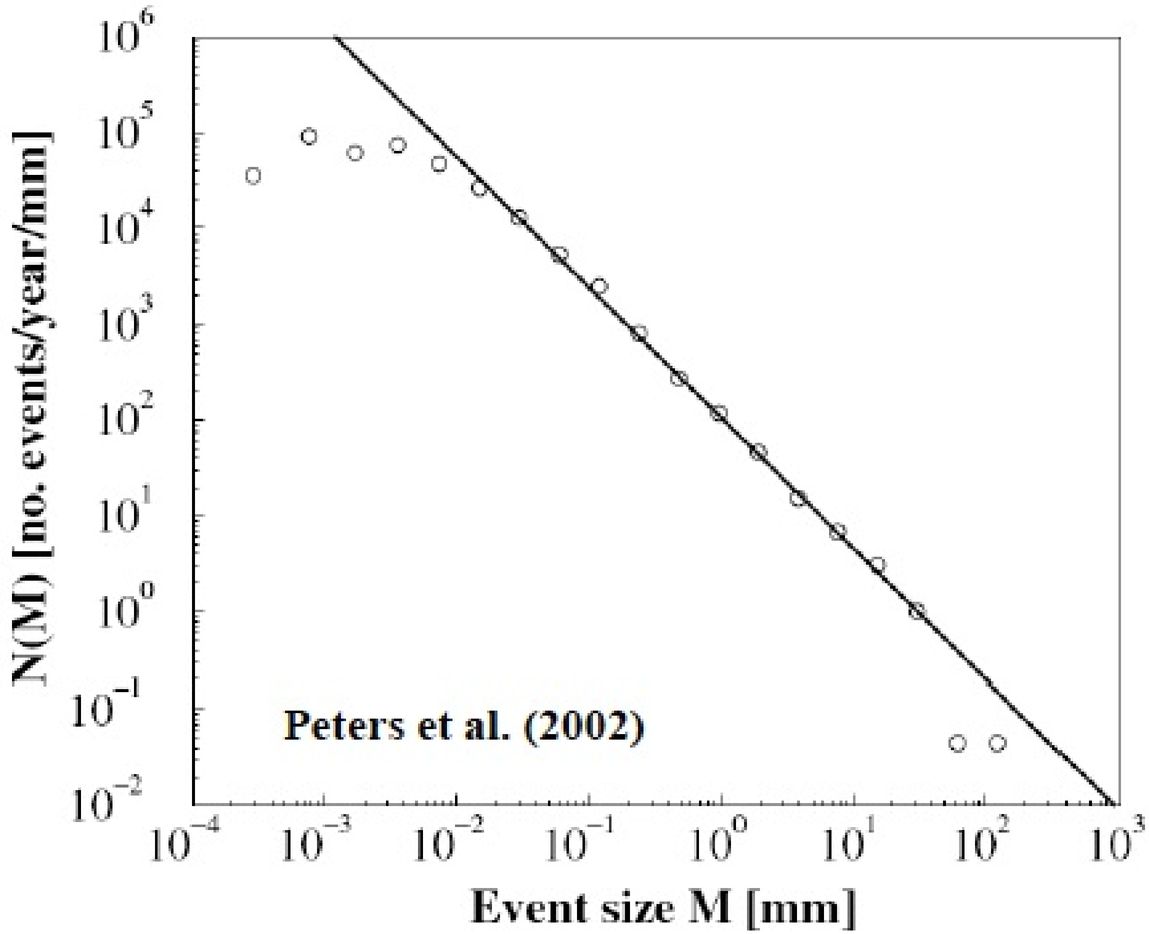
City size



MEJ Newman (2005)

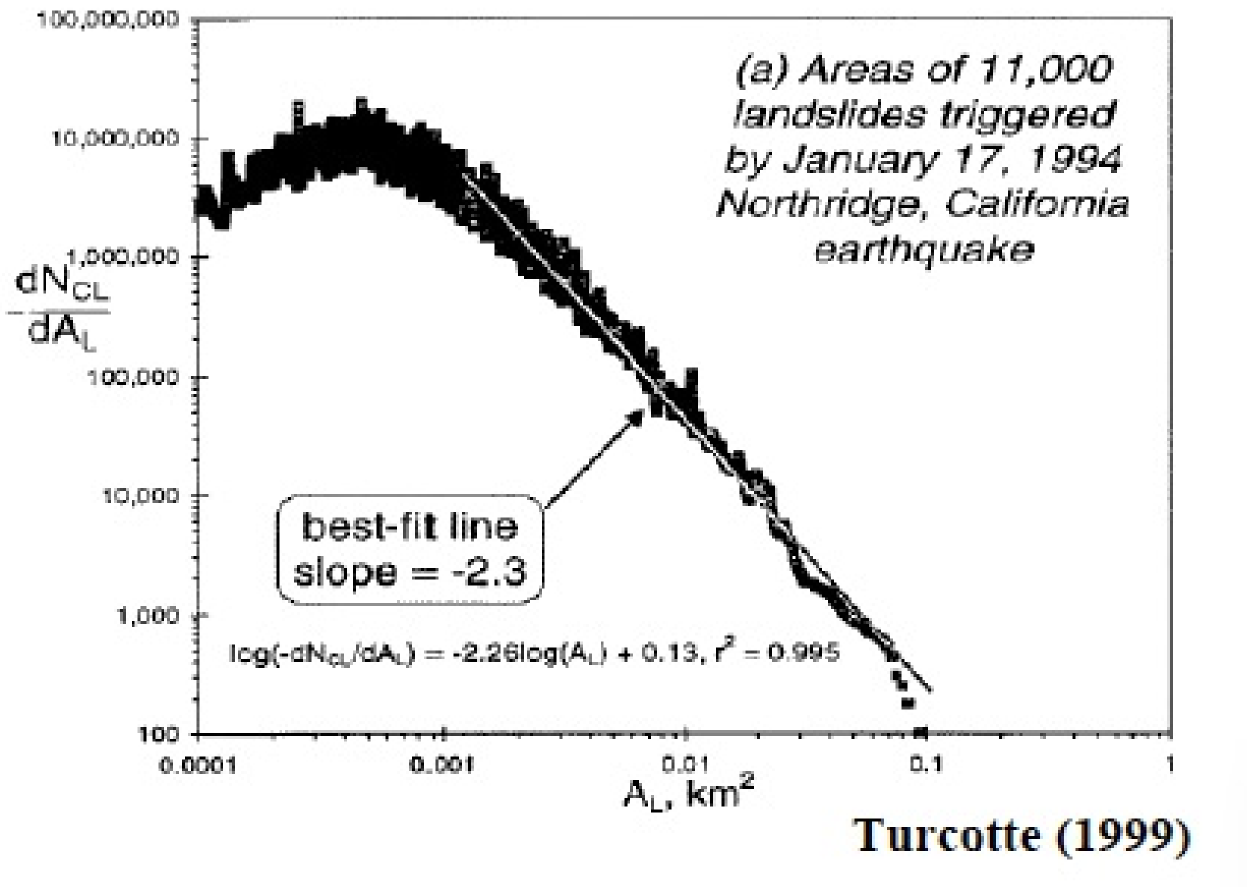
multiplicative

Rainfall



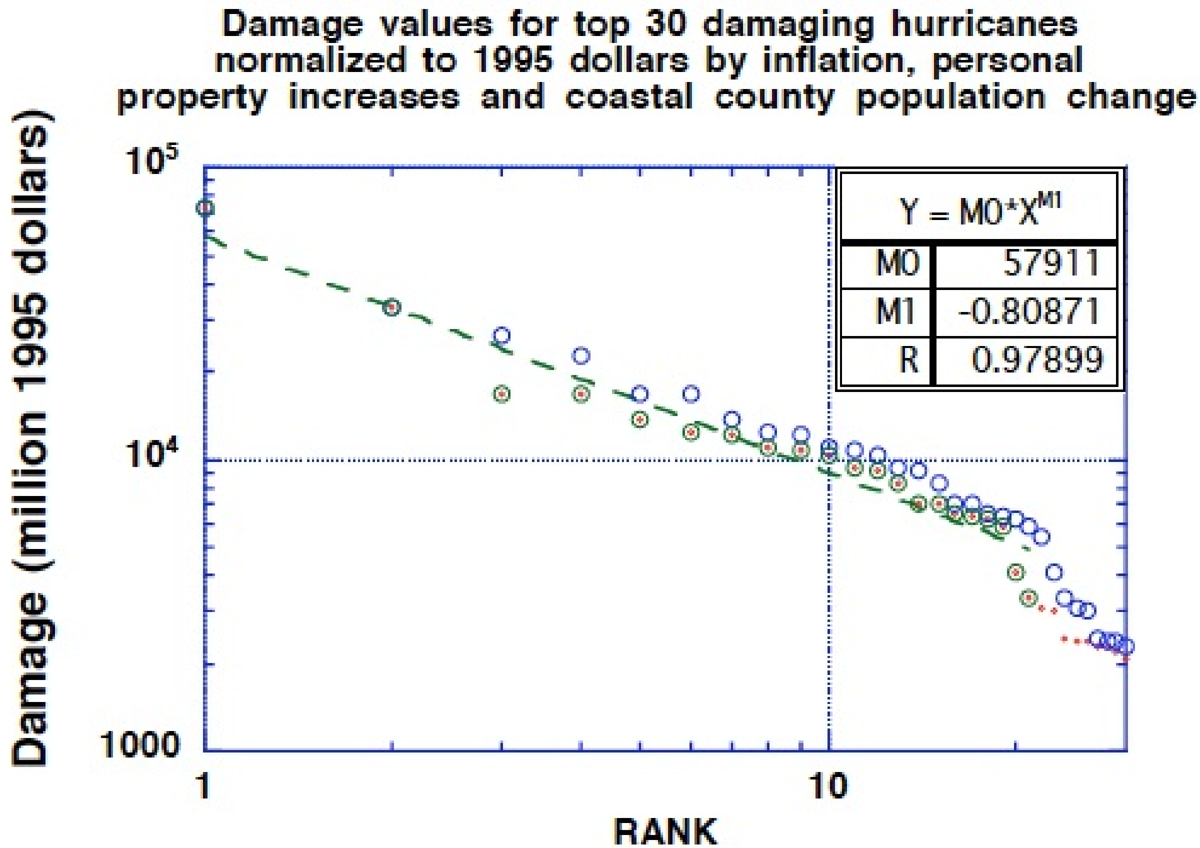
SOC

Landslides



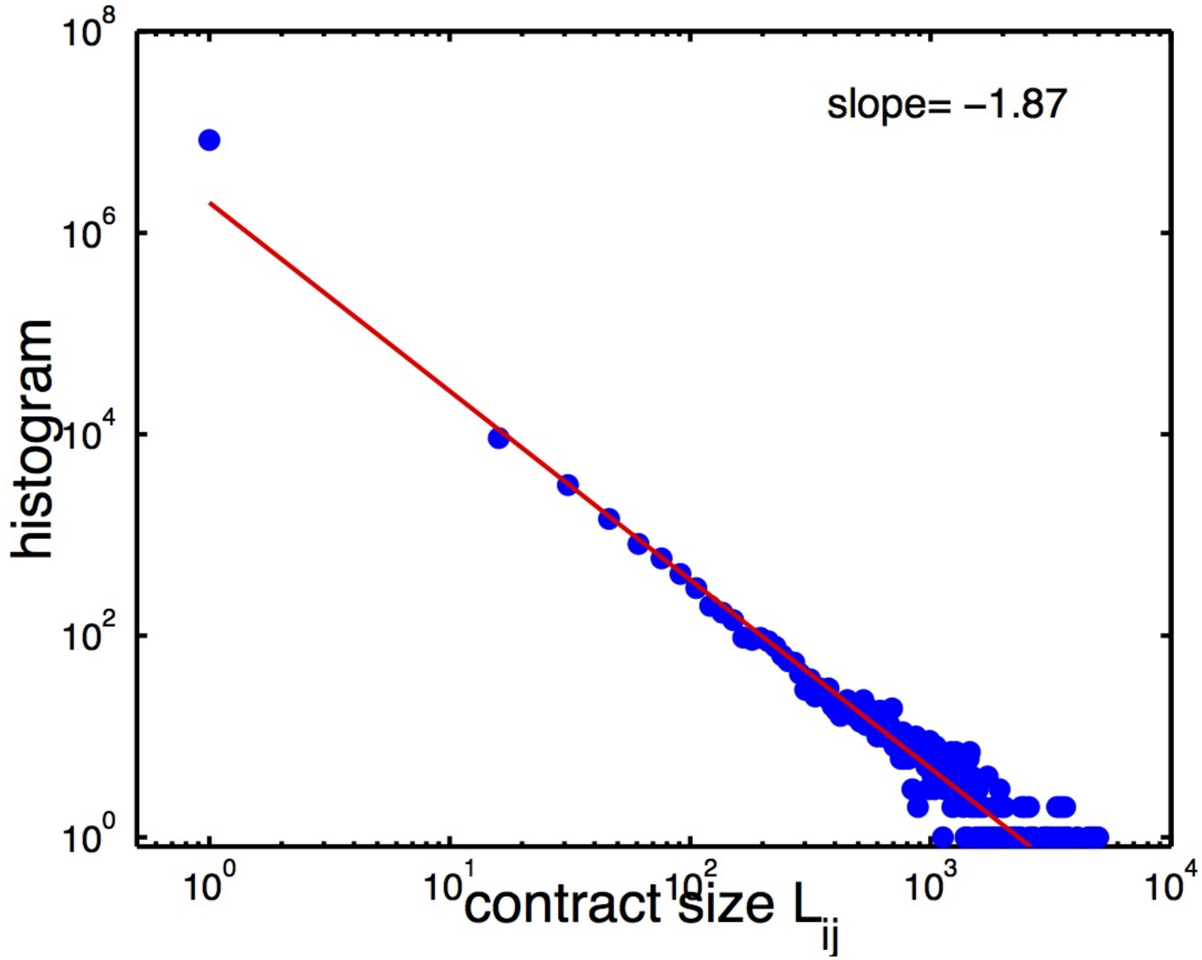
SOC

Hurrican damages



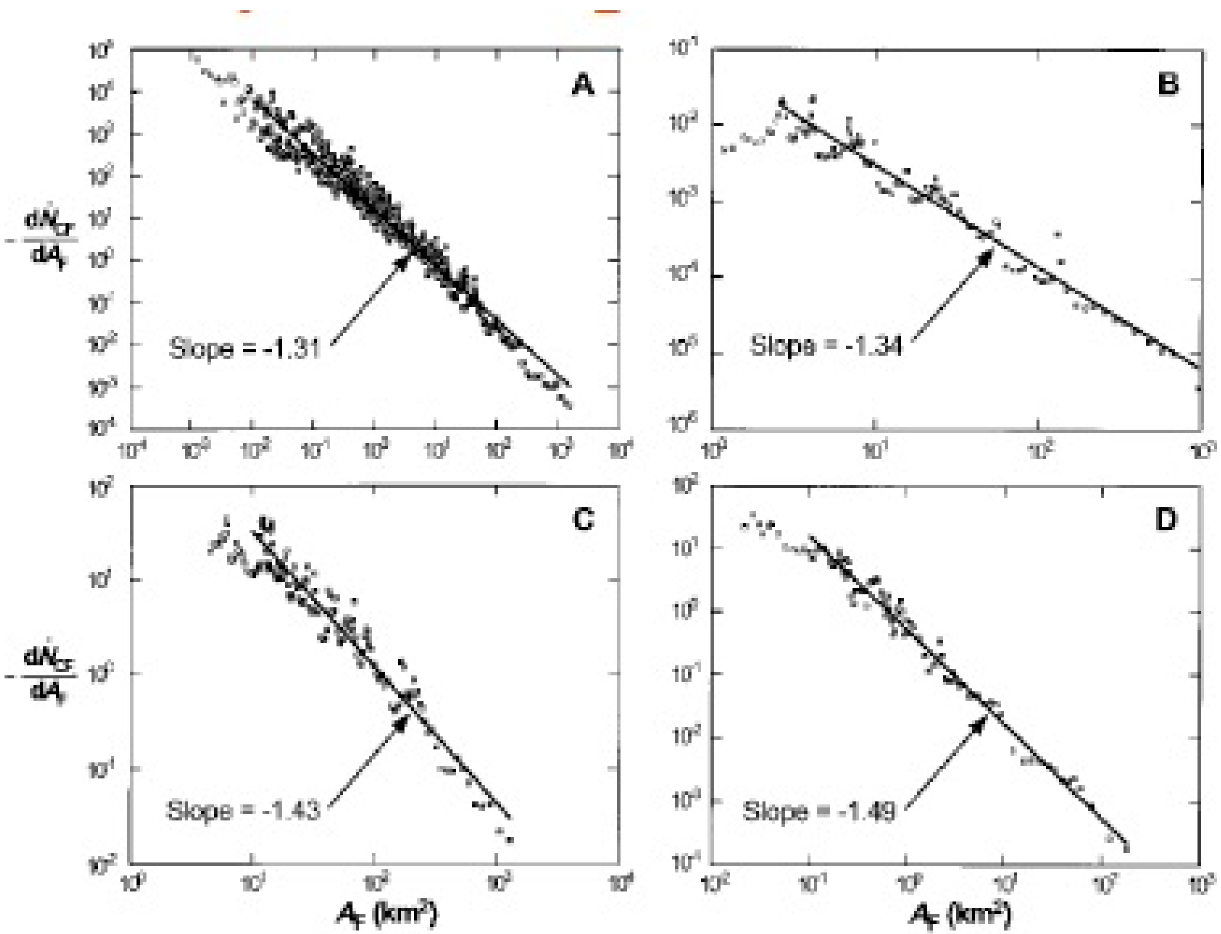
secondary (multiplicative) ???

Financial interbank loans



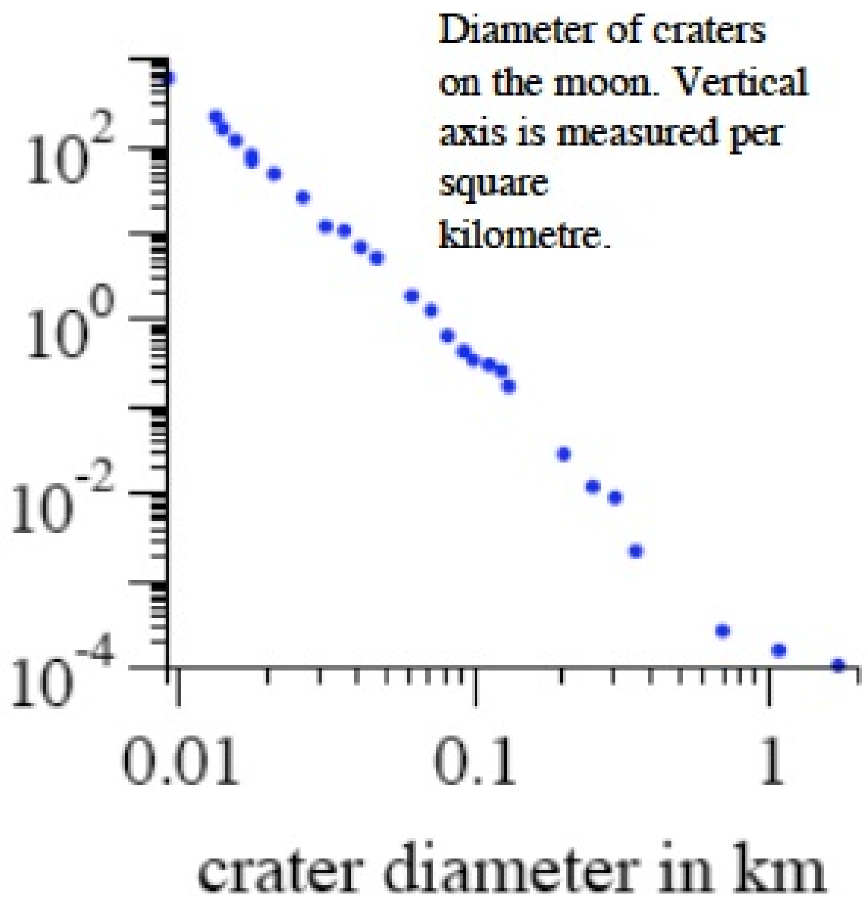
multiplicative / preferential

Forrest fires in various regions



SOC ?

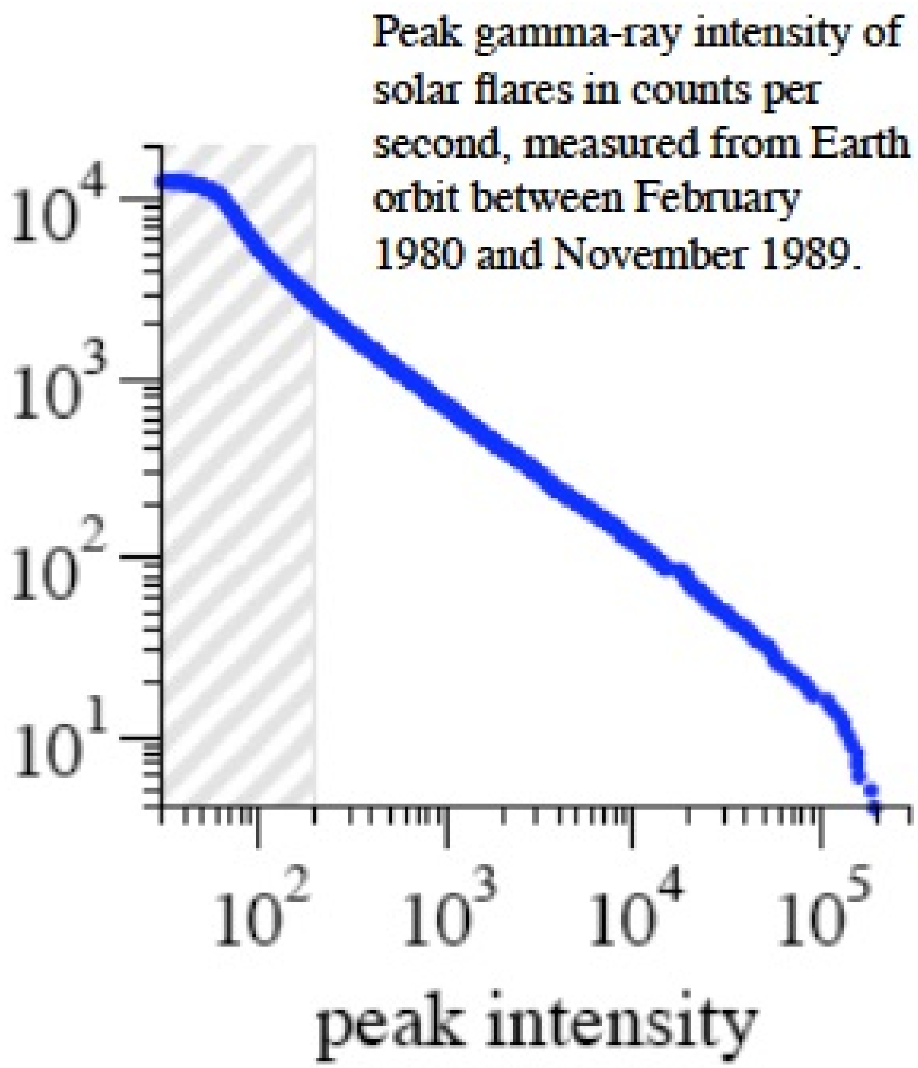
Moon crater diameters



MEJ Newman (2005)

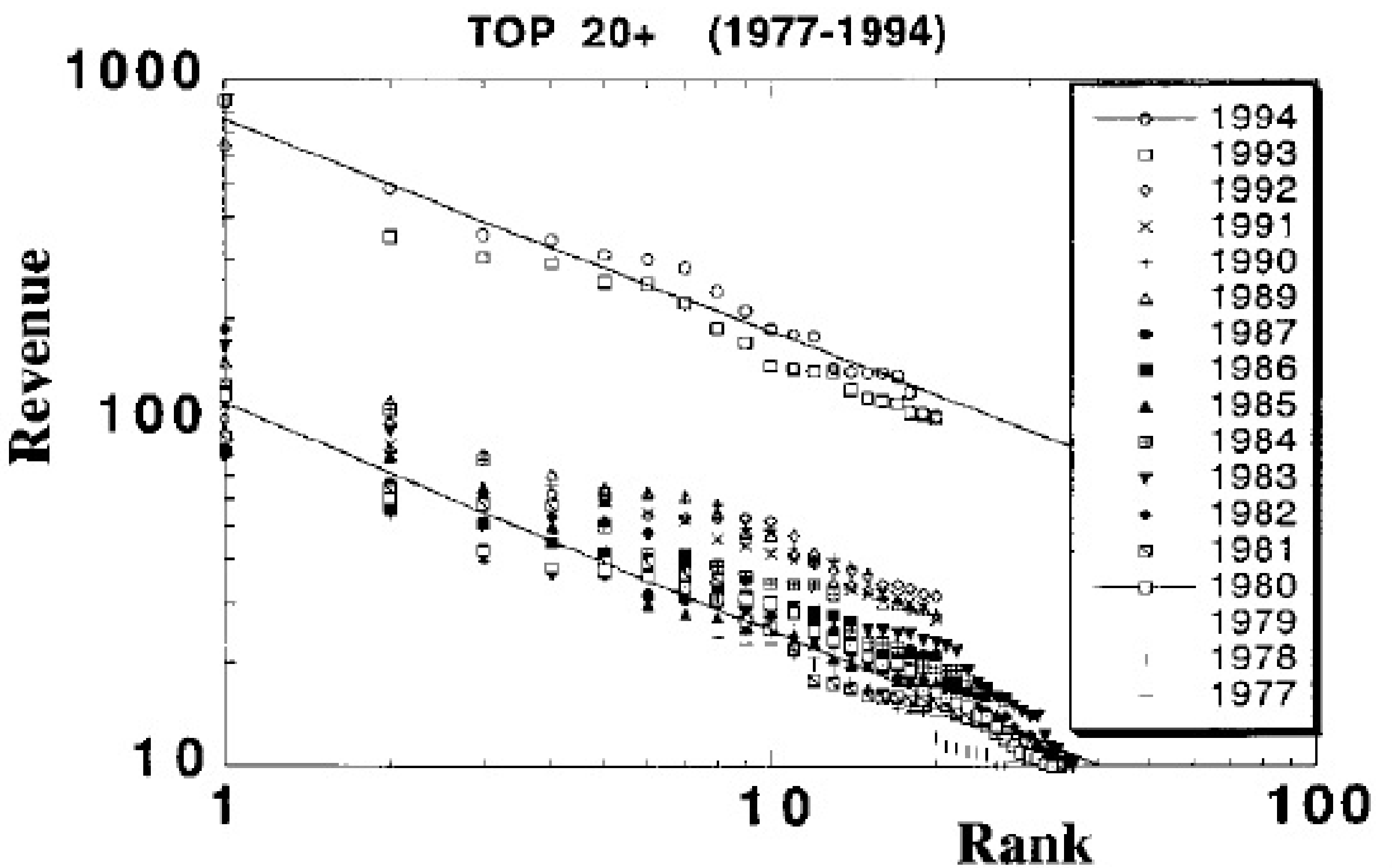
???

Gamma rays from solar wind



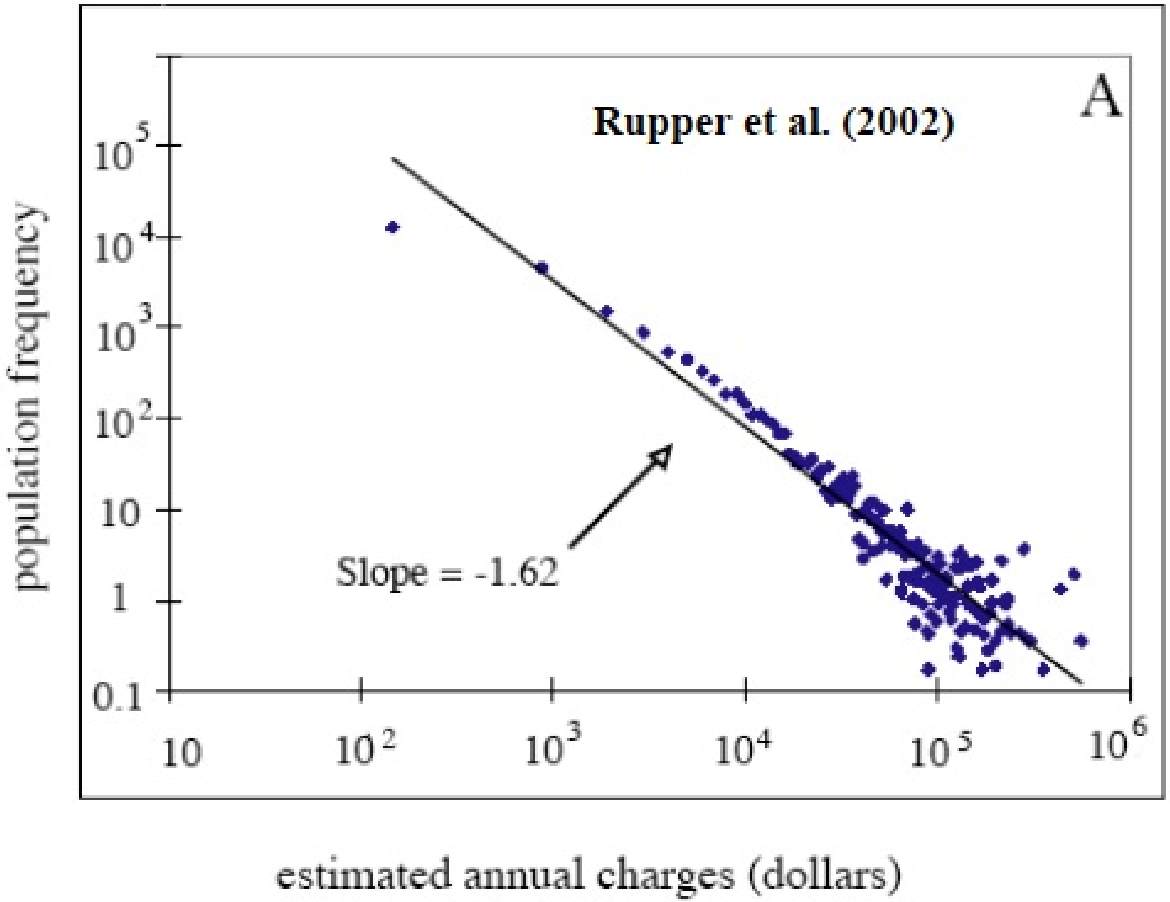
MEJ Newman (2005)

Movie sales



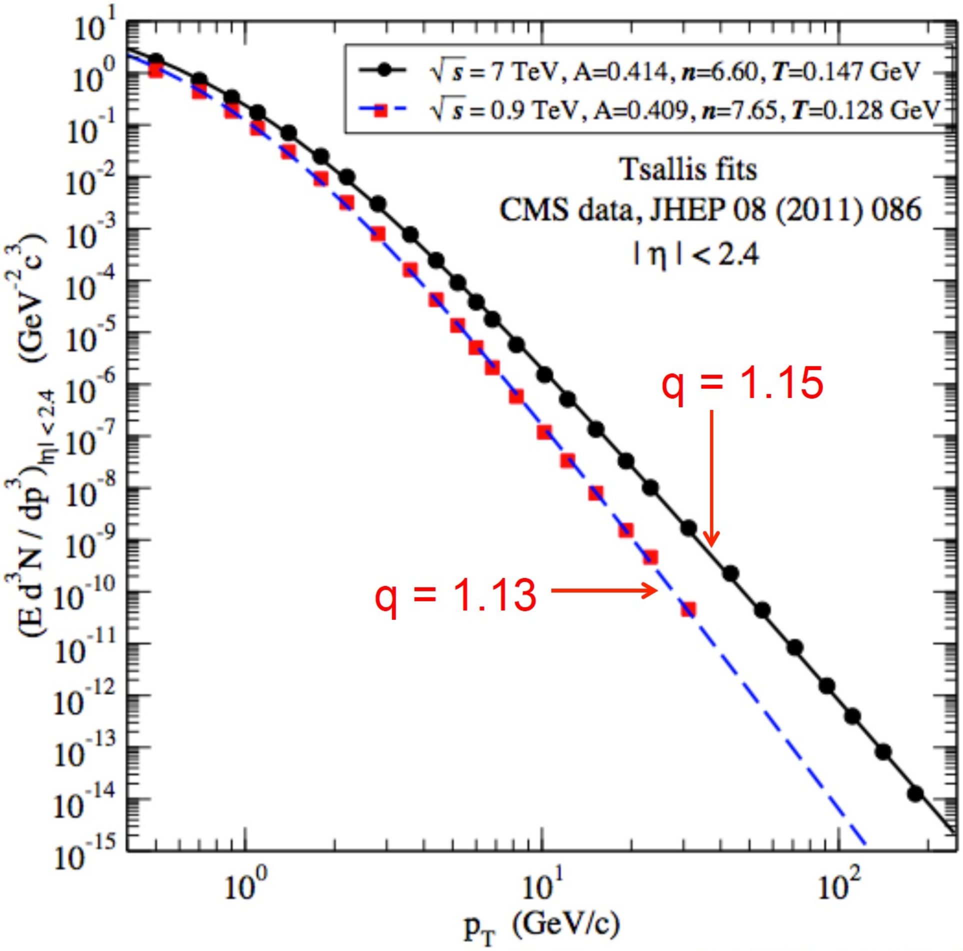
SOC

Healthcare costs



multiplicative ???

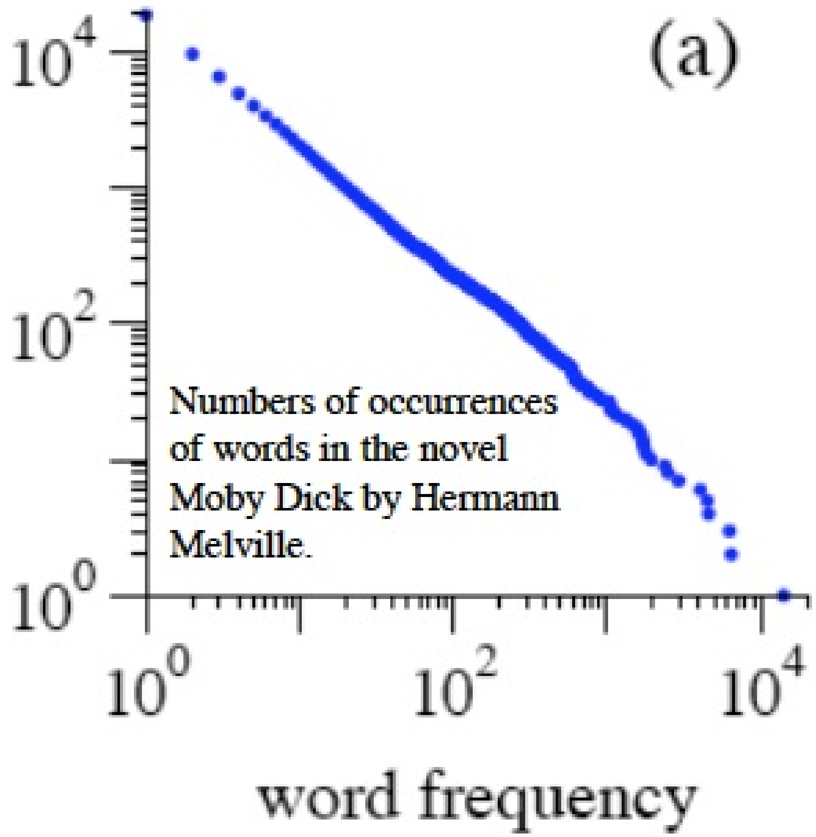
Particle physics



C.Y. Wong and G. Wilk (2012), 1210.3661 [hep-ph]

???

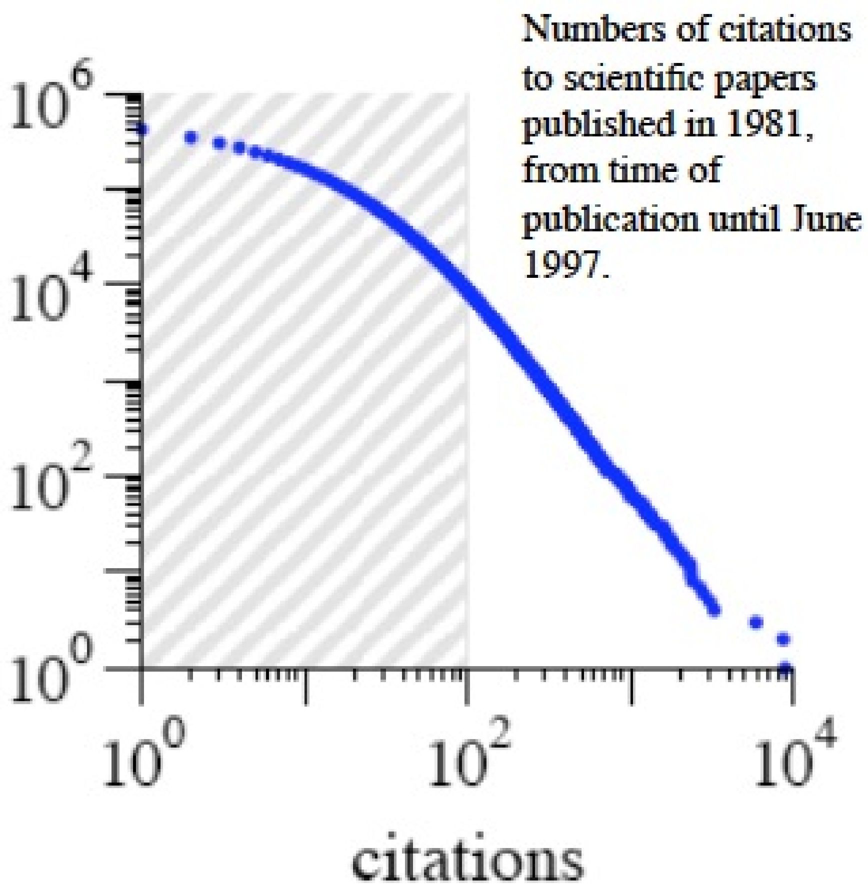
Words in books



MEJ Newman (2005)

preferential / random / optimization

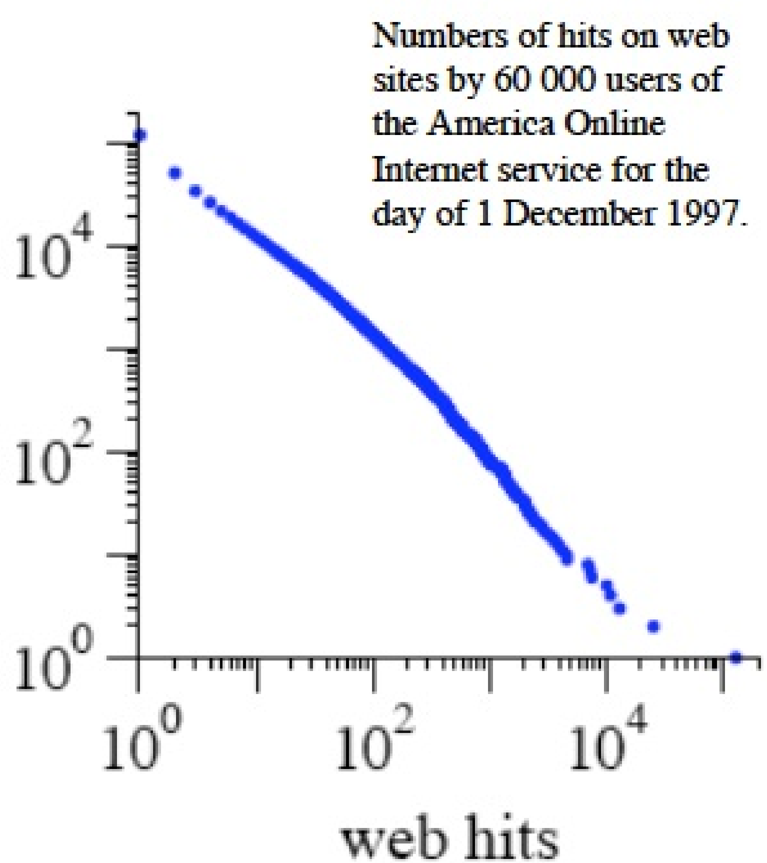
Citations of scientific articles



MEJ Newman (2005)

preferential

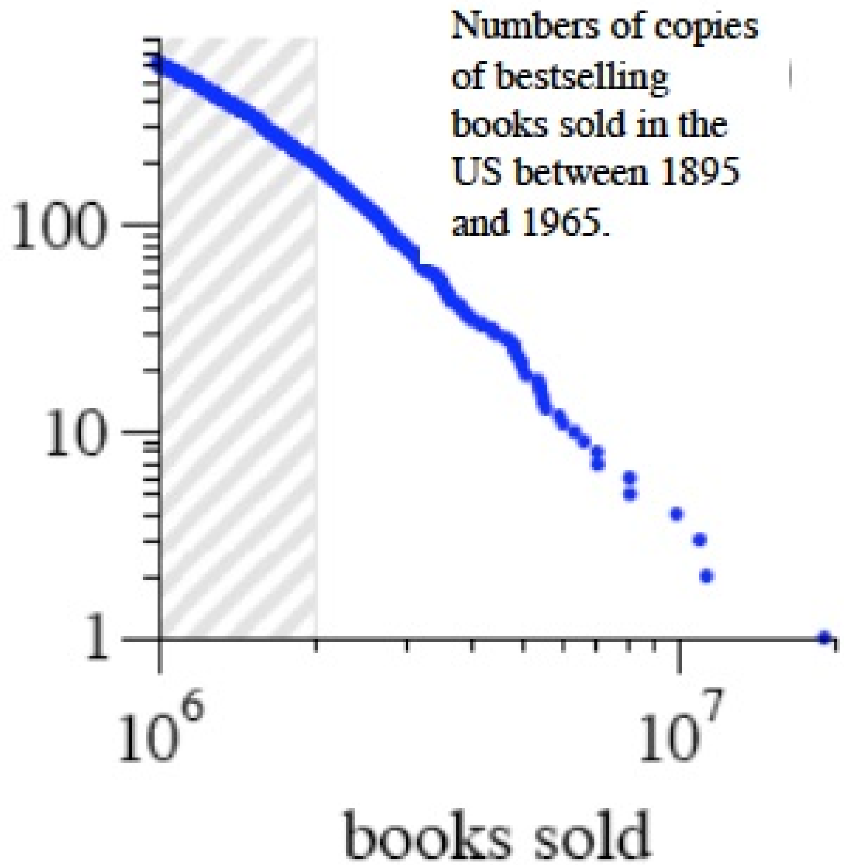
Website hits



MEJ Newman (2005)

preferential

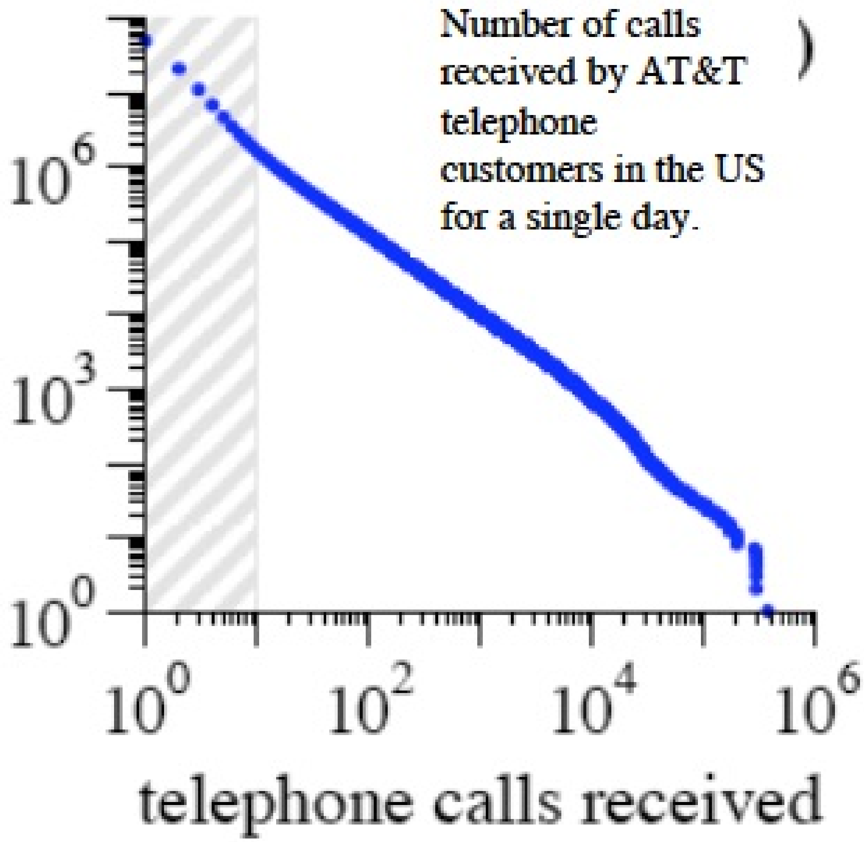
Book sales



MEJ Newman (2005)

preferential

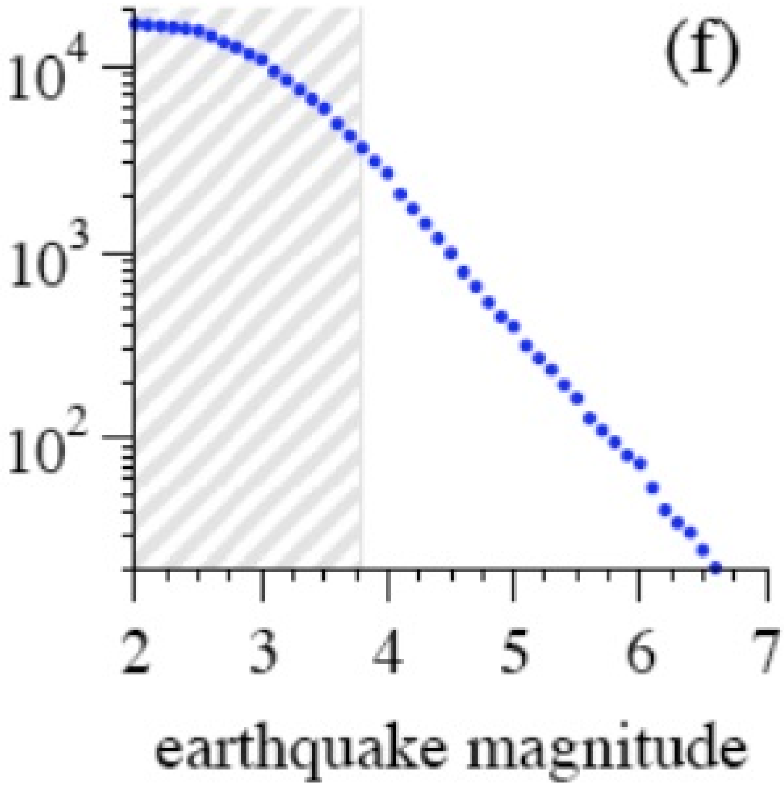
Telephone calls



MEJ Newman (2005)

preferential

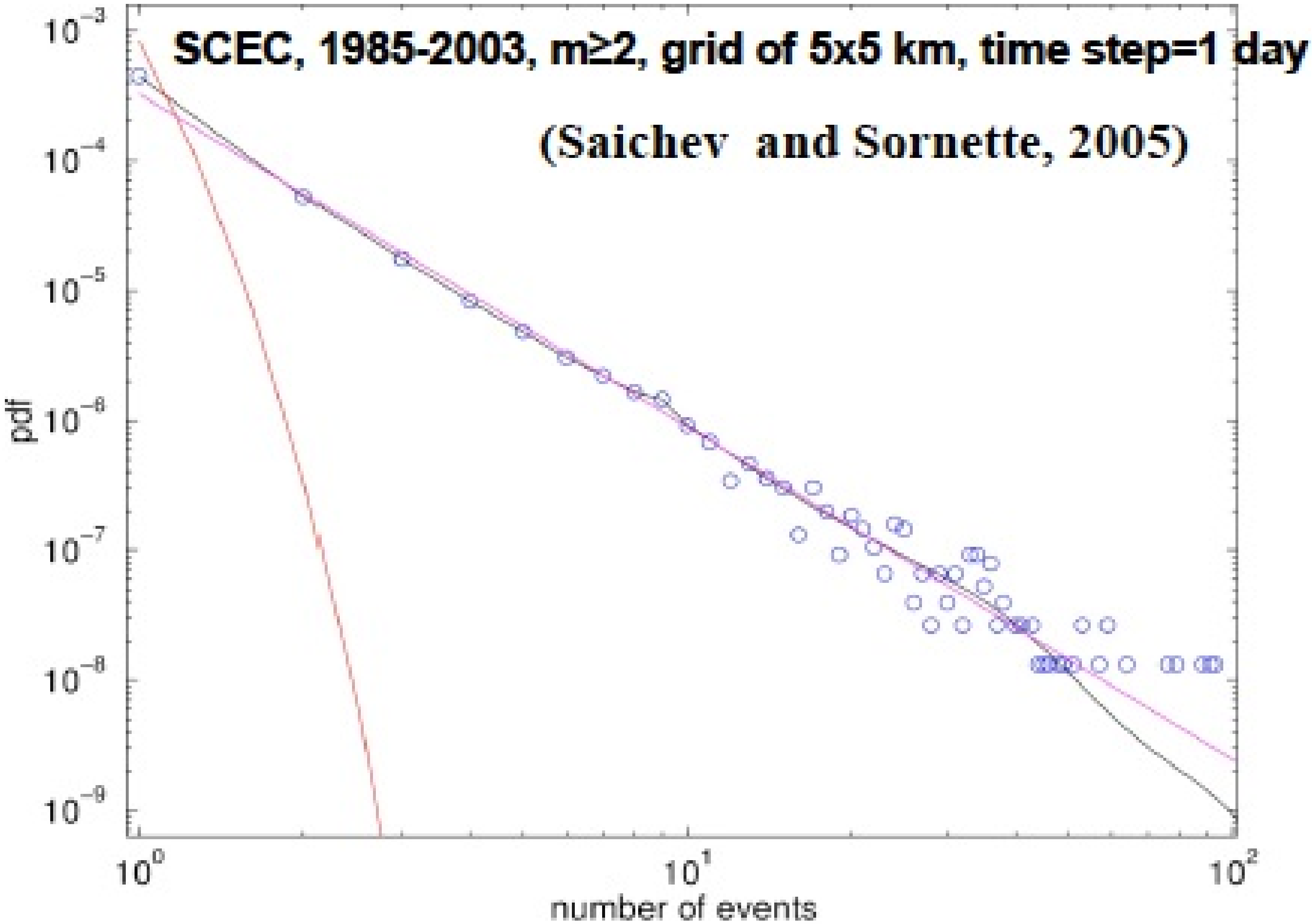
Earth quake magnitude



MEJ Newman (2005)

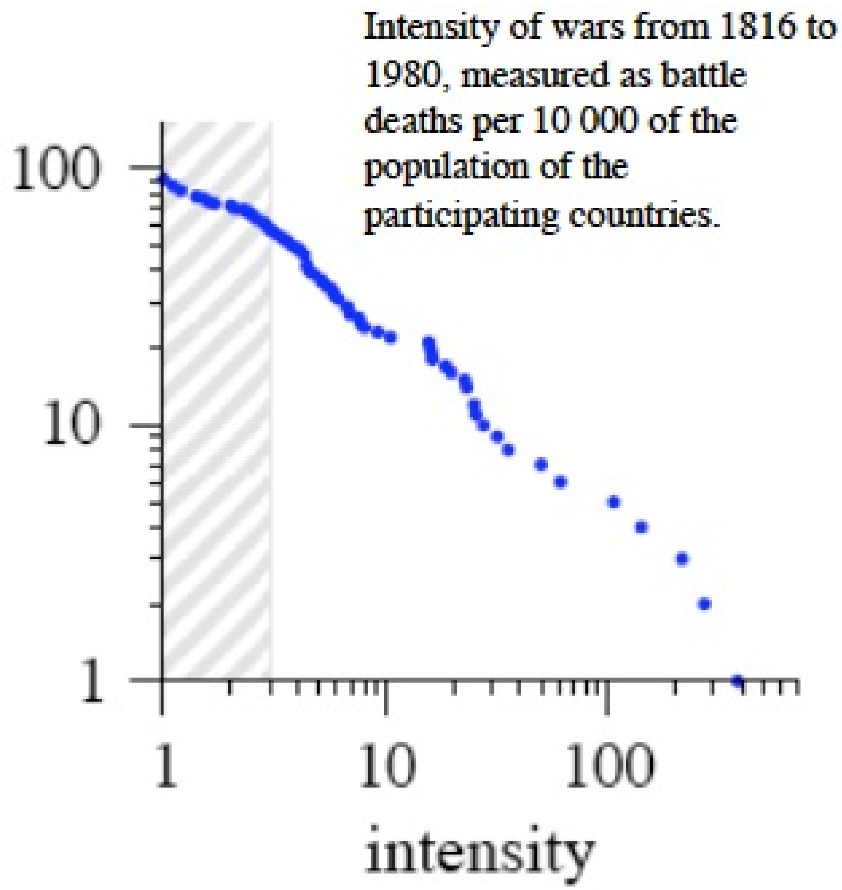
SOC

Seismic events



SOC

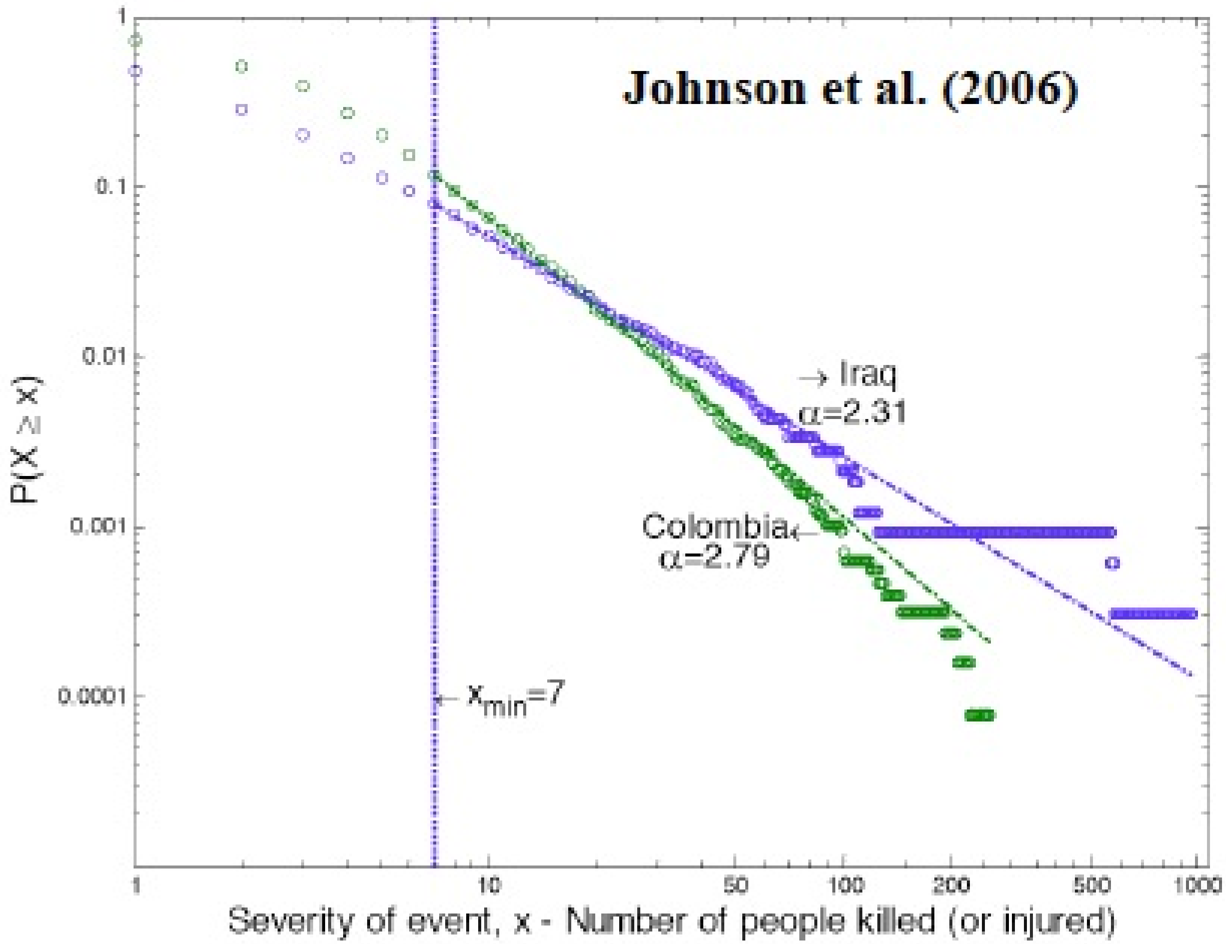
War intensity



MEJ Newman (2005)

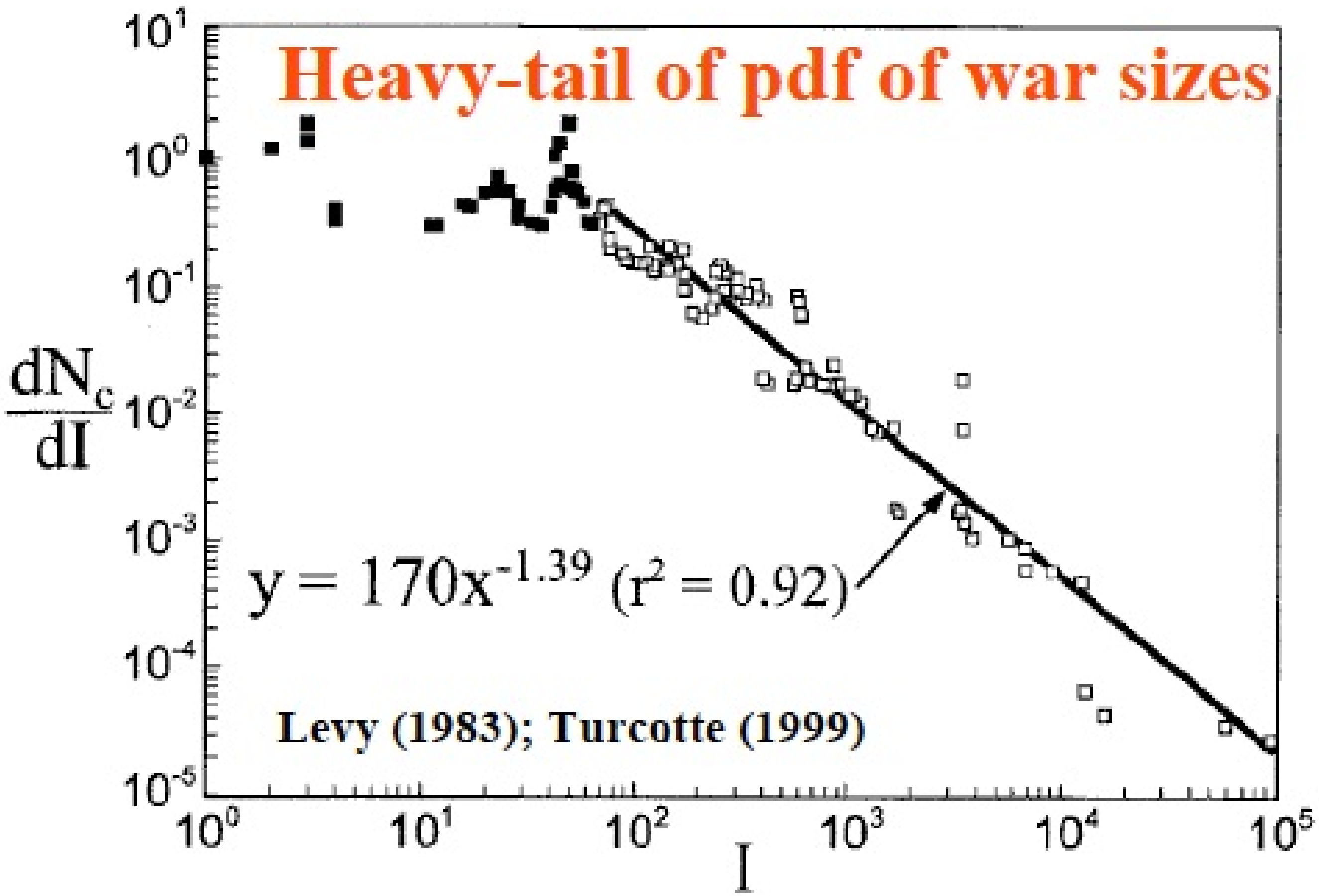
???

Killings in wars



???

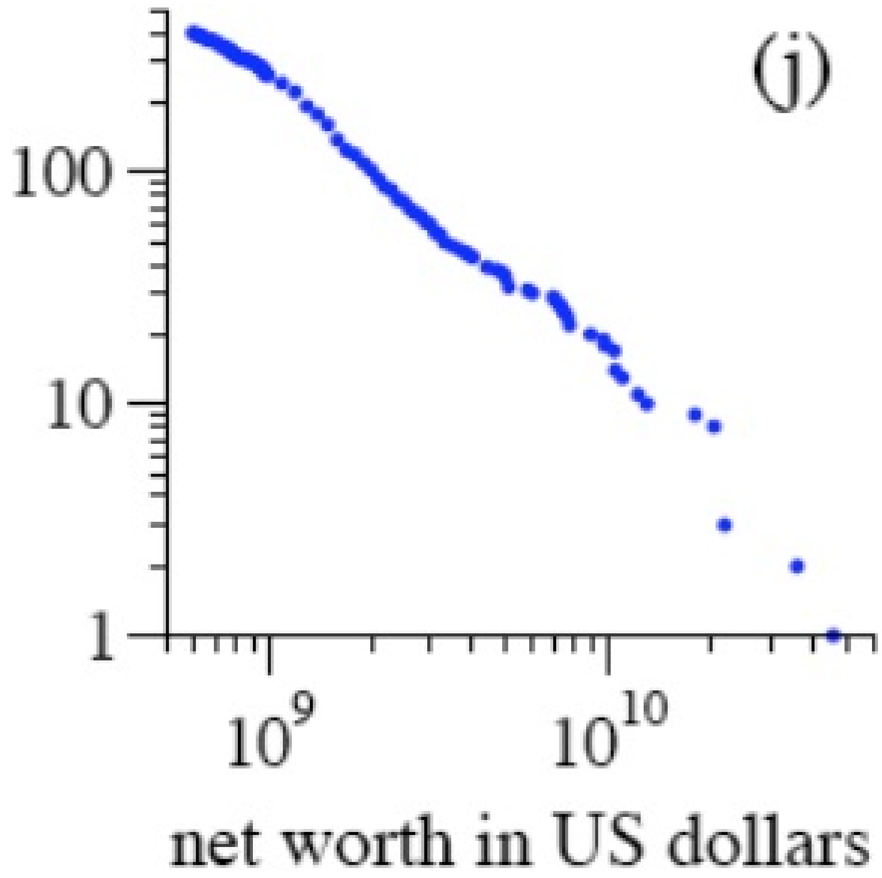
Size of war



Heavy-tail of pdf of war sizes
y = 170x^-1.39 (r^2 = 0.92)
Levy (1983); Turcotte (1999)
dNc/dI
I
10^1
10^0
10^-1
10^-2
10^-3
10^-4
10^-5
10^0
10^1
10^2
10^3
10^4
10^5

???

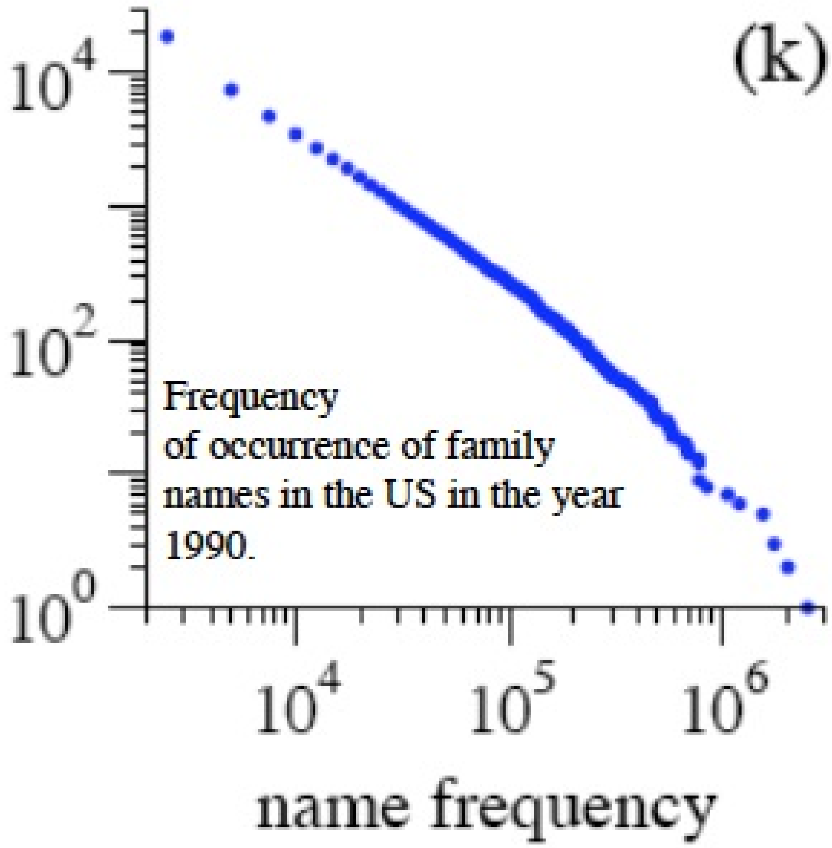
Wealth distribution



MEJ Newman (2005)

multiplicative

Family names



MEJ Newman (2005)

???

More power laws ...

- networks: literally thousands of scale-free networks
- allometric scaling in biology
- dynamics in cities
- fragmentation processes
- random walks
- crackling noise
- growth with random times of observation
- blackouts
- fossil record
- bird sightings
- terrorist attacks
- fluvial discharge, contact processes
- anomalous diffusion ...

Where do they come from?

Classical routes to understand power laws

- statistical mechanics: at phase transitions
- self-organised criticality
- multiplicative processes with constraints
- preferential processes

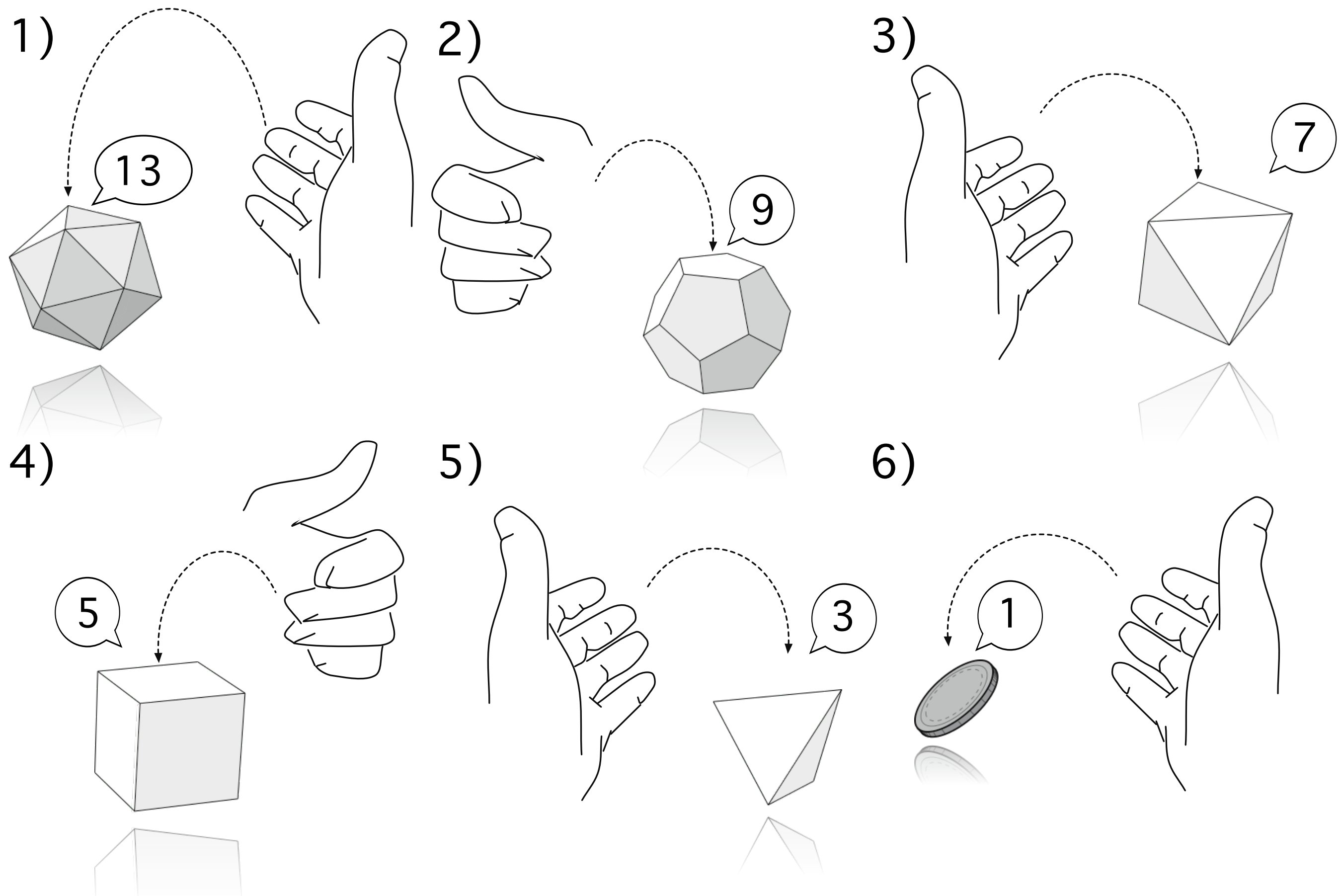
did we miss something ?

Many processes are history- or path dependent

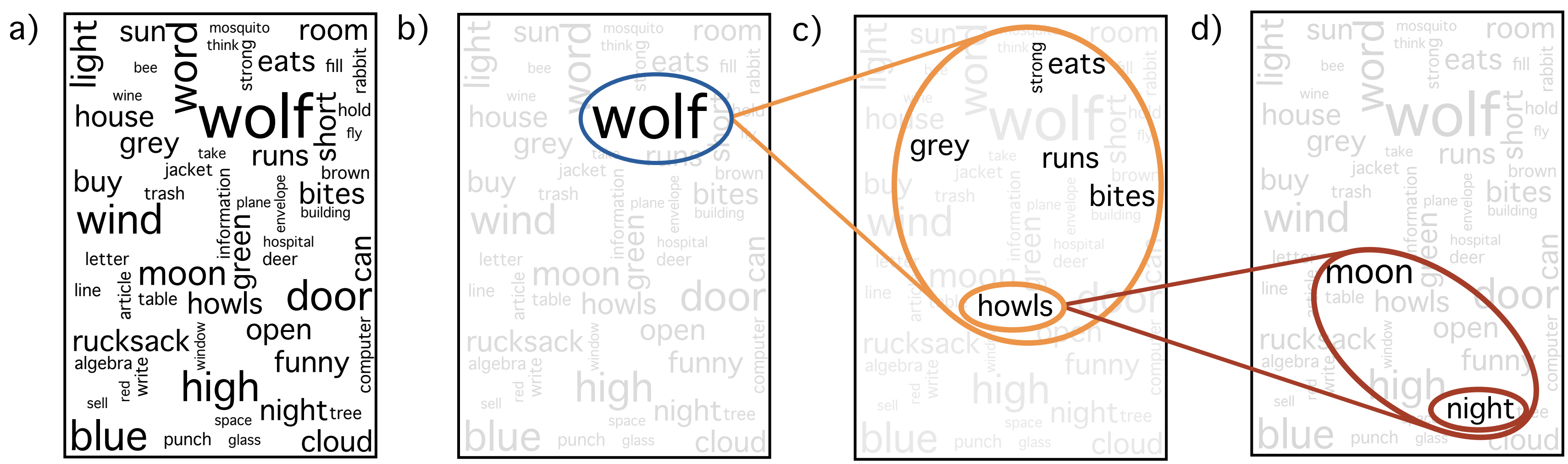
- future events depend on history of past events
- often past events constrain possibilities for future

$\rightarrow$ sample-space of these processes reduces as they unfold

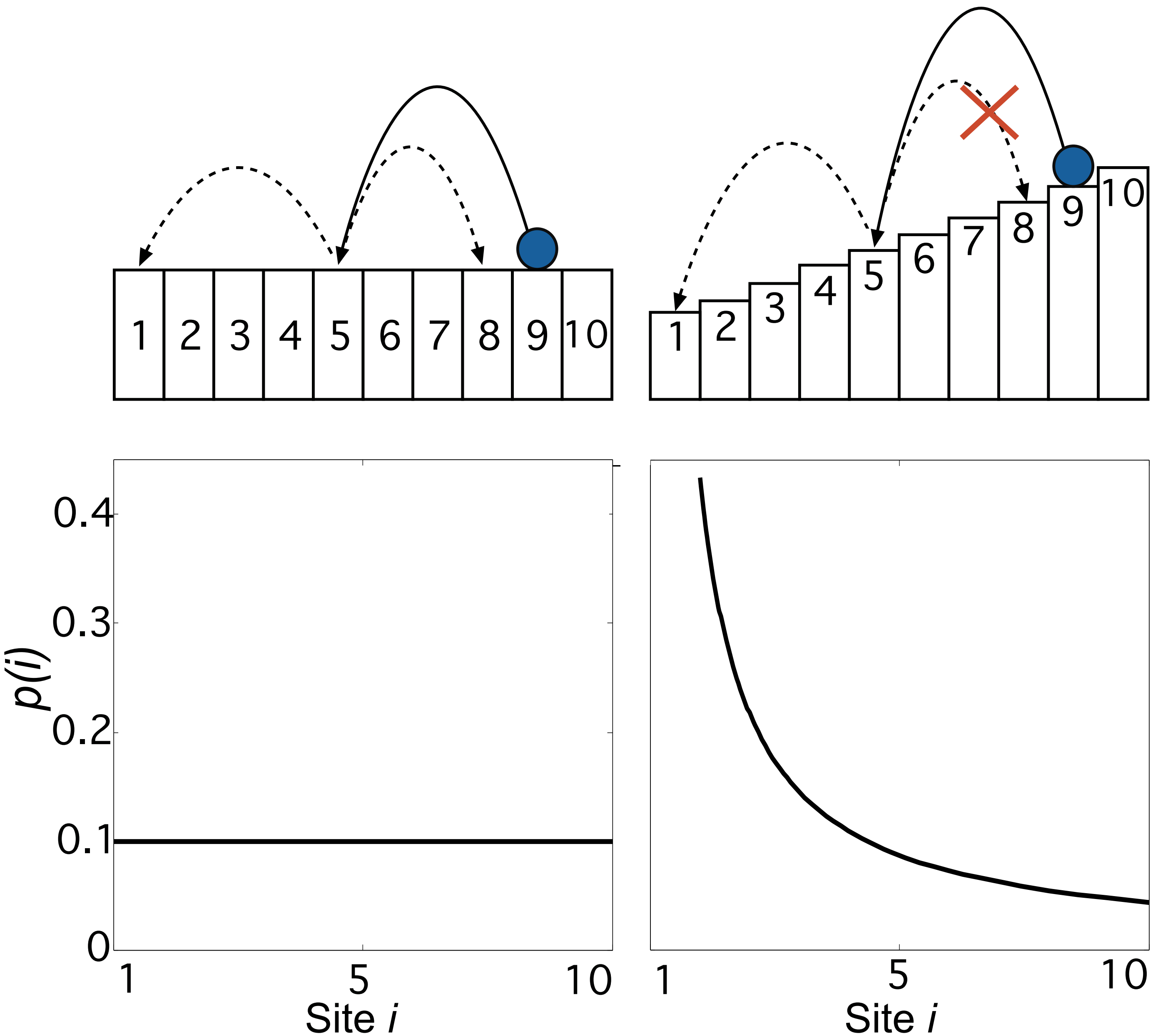
Example: History-dependent processes



Sentence-formation is SSR



Sample-Space Reducing Processes (SSR)



1
2
3
4
5
6
7
8
9
10
p(i)
0.4
0.3
0.2
0.1
0
Site i

SSR lead to exact Zipf law!

$$p(i) = \frac{1}{i}$$

$p(i)$ is probability to visit site i

Proof by induction

Let $N = 2$. There are two sequences ϕ: either ϕ directly generates a 1 with $p = 1/2$, or first generates 2 with $p = 1/2$, and then a 1 with certainty. Both sequences visit 1 but only one visits 2. As a consequence, $P_2(2) = 1/2$ and $P_2(1) = 1$.

Now suppose $P_{N-1}(i) = 1/i$ holds. Process starts with dice N, and probability to hit i in the first step is $1/N$. Also, any other j, $N \geq j > i$, is reached with probability $1/N$. If we get $j > i$, we get i in the next step with probability $P_{j-1}(i)$, which leads to a recursive scheme for $i < N$, $P_N(i) = \frac{1}{N}\left(1 + \sum_{i<j\leq N} P_{j-1}(i)\right)$. Since by assumption $P_{j-1}(i) = 1/i$, with $i < j \leq N$ holds, some algebra yields $P_N(i) = 1/i$.

True for all systems with an adjacent possible that shrinks over time

What if not strictly SSR

ϕ ... Sample Space Reducing process (SSR)

ϕ_R ... Random walk

Mix both processes

$$\Phi^{(\lambda)} = \lambda\phi + (1-\lambda)\phi_R \quad , \quad \lambda \in [0,1]$$

Add noise with strength $(1-\lambda) \to$ any power becomes possible

$$p(i) = i^{-\lambda}$$

noise $(1-\lambda)$ is a **surprise factor** for SSR process

The role of noise – result is exact too

Clearly $p^{(\lambda)}(i) = \sum_{j=1}^{N} P(i|j)\, p^{(\lambda)}(j)$ holds, with

$$P(i|j) = \begin{cases} \frac{\lambda}{j-1} + \frac{1-\lambda}{N} & \text{for } i < j \quad (SSR) \\ \frac{1-\lambda}{N} & \text{for } i \geq j > 1 \quad (RW) \\ \frac{1}{N} & \text{for } i \geq j = 1 \quad (restart) \end{cases}$$

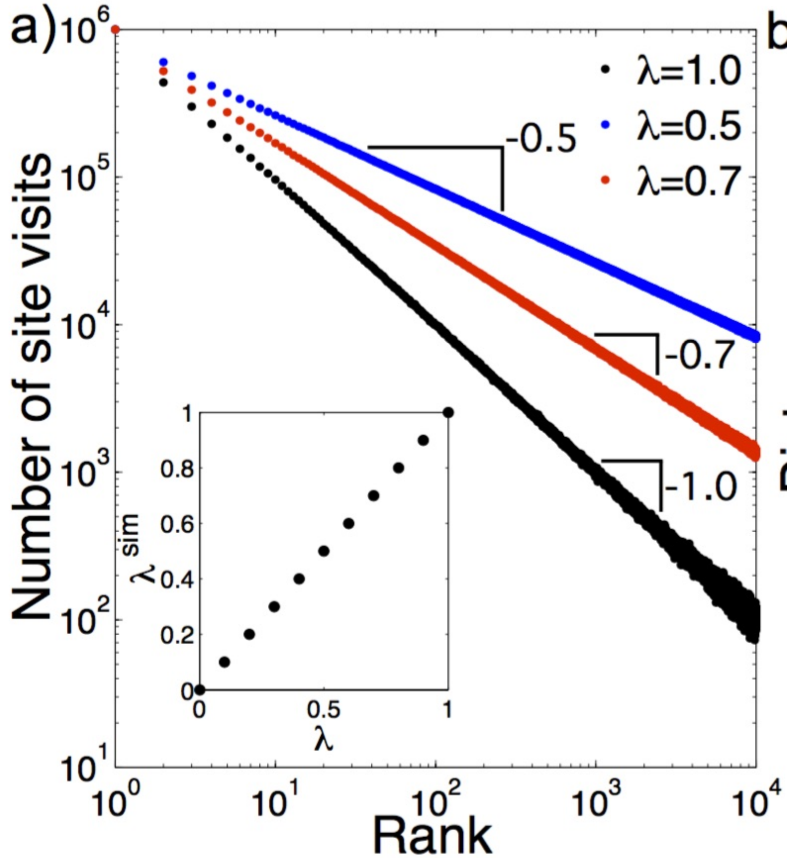
We get $p^{(\lambda)}(i) = \frac{1-\lambda}{N} + \frac{1}{N} p^{(\lambda)}(1) + \sum_{j=i+1}^{N} \frac{\lambda}{j-1} p^{(\lambda)}(j)$

to recursive relation $p^{(\lambda)}(i+1) - p^{(\lambda)}(i) = -\lambda \frac{1}{i} p^{(\lambda)}(i+1)$

$$\begin{aligned} \frac{p^{(\lambda)}(i)}{p^{(\lambda)}(1)} &= \prod_{j=1}^{i-1} \left(1 + \frac{\lambda}{j}\right)^{-1} = \exp\left[-\sum_{j=1}^{i-1} \log\left(1 + \frac{\lambda}{j}\right)\right] \\ &\sim \exp\left(-\sum_{j=1}^{i-1} \frac{\lambda}{j}\right) \sim \exp\left(-\lambda \log(i)\right) = i^{-\lambda} \end{aligned}$$

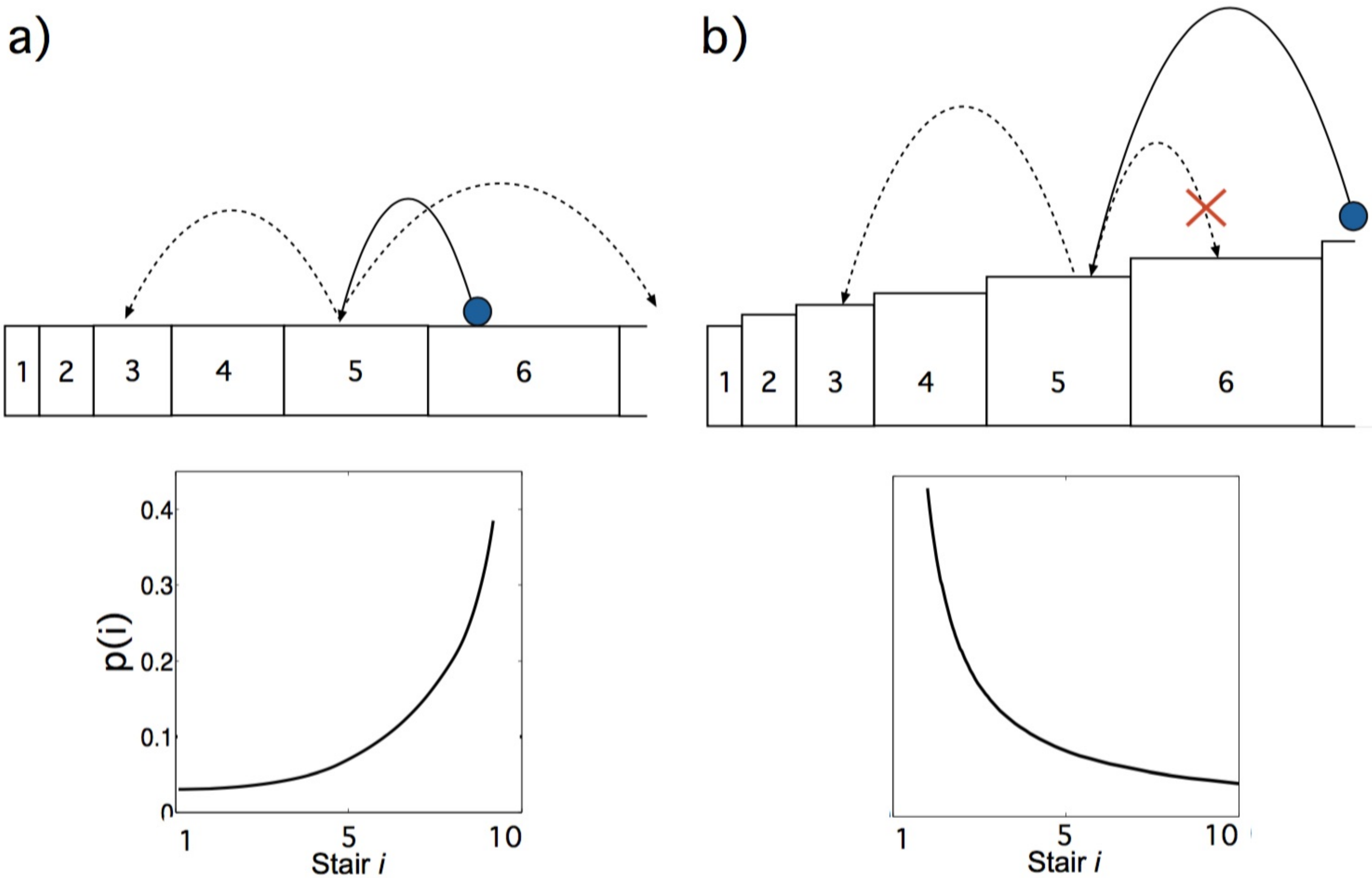
True for all systems with an adjacent possible that shrinks over time with probability λ

History-dependent processes with noise



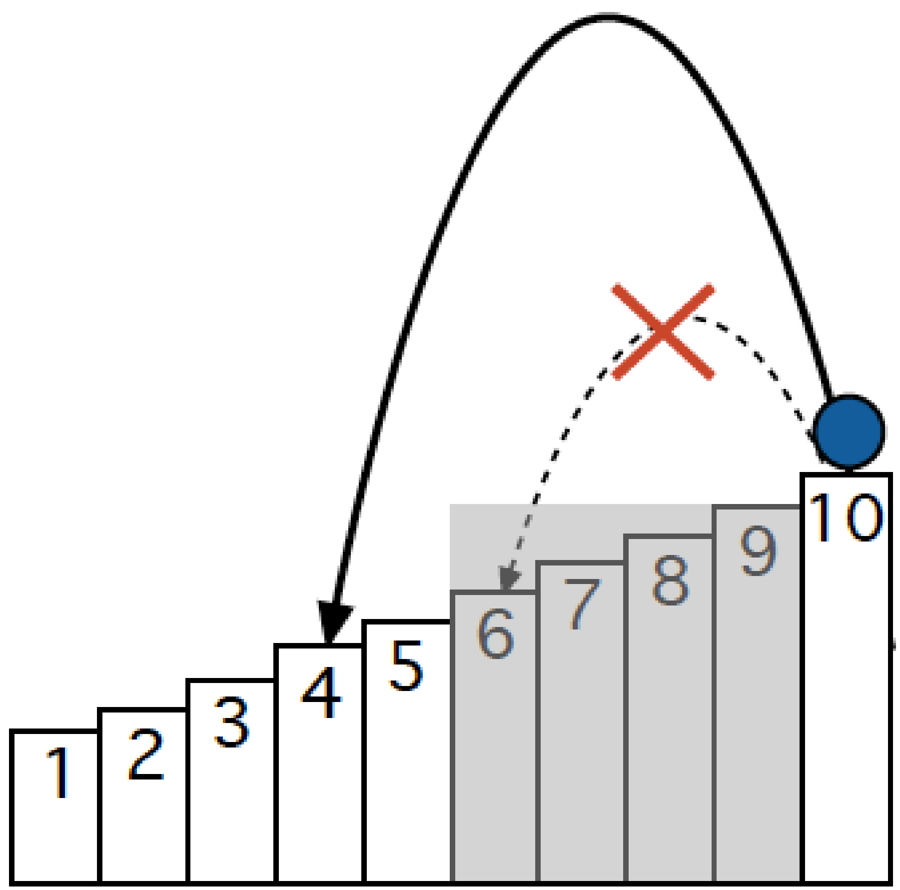
same convergence speed as CLT for iid processes

SSR based Zipf law is extremely robust



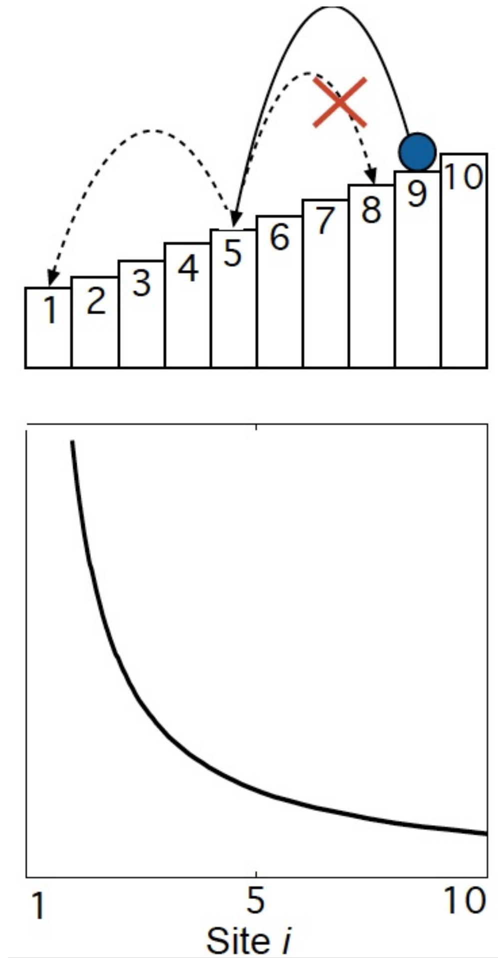
prior probabilities are practically irrelevant!

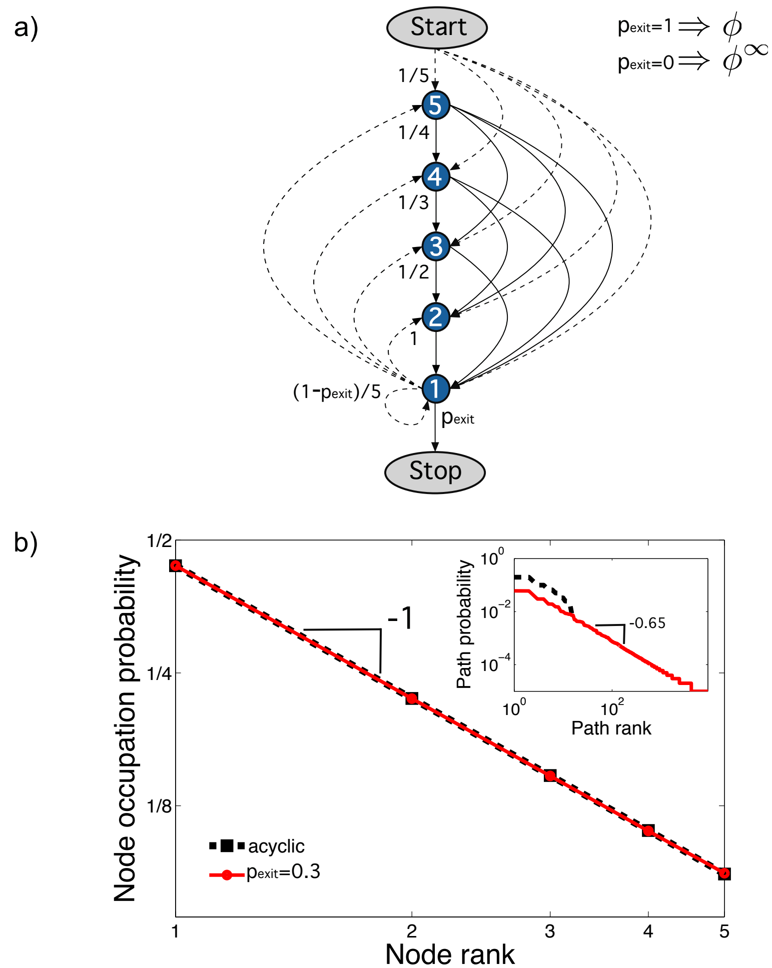
Zipf law is remarkably robust – accelerated SSR



What does this have to do with networks?

SSR is a random walk on directed ordered NW

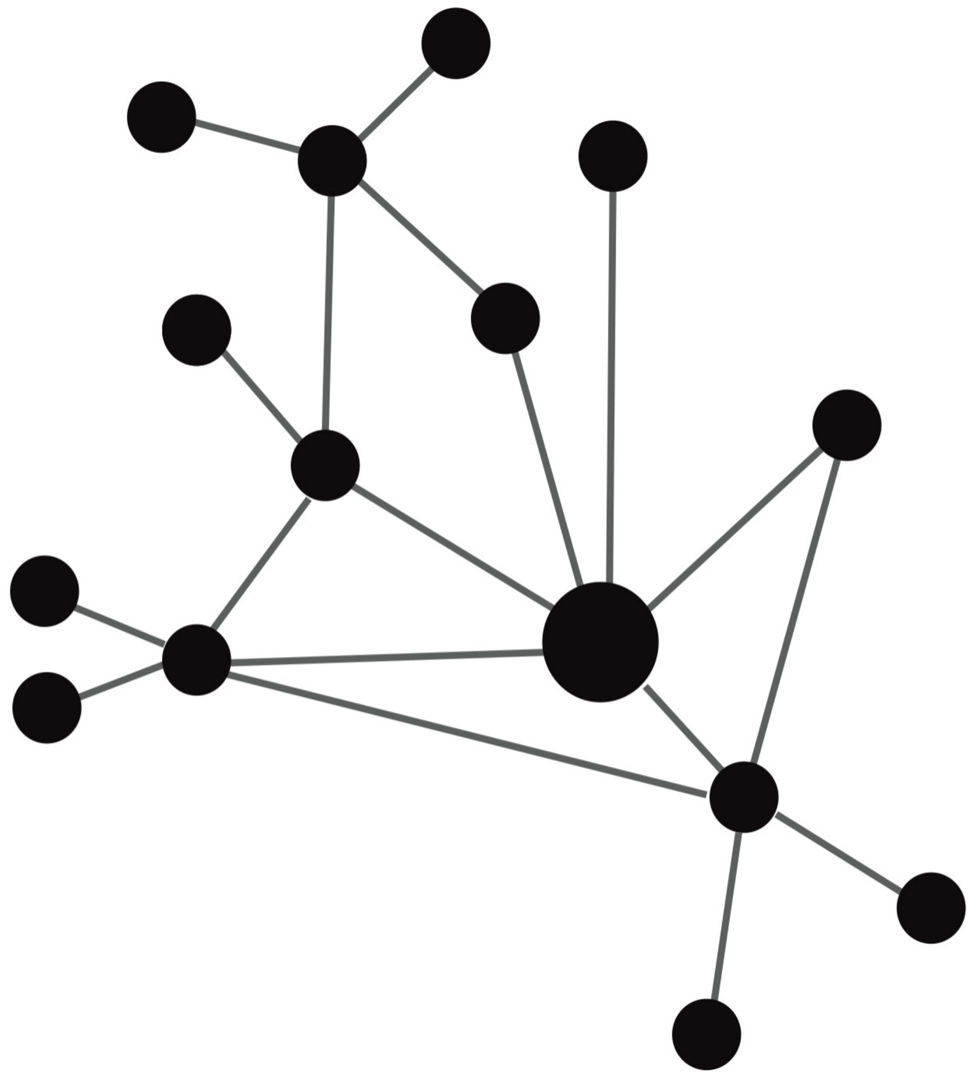


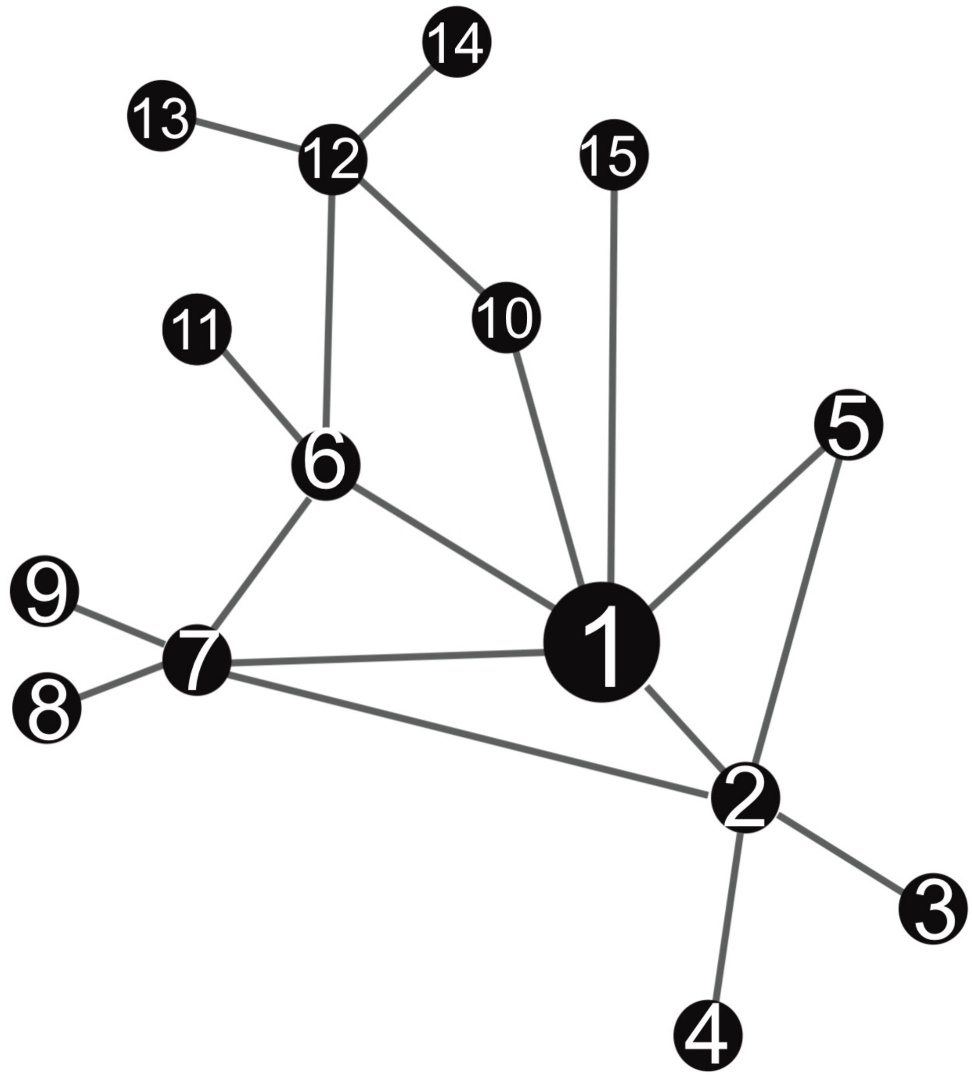


note fully connected

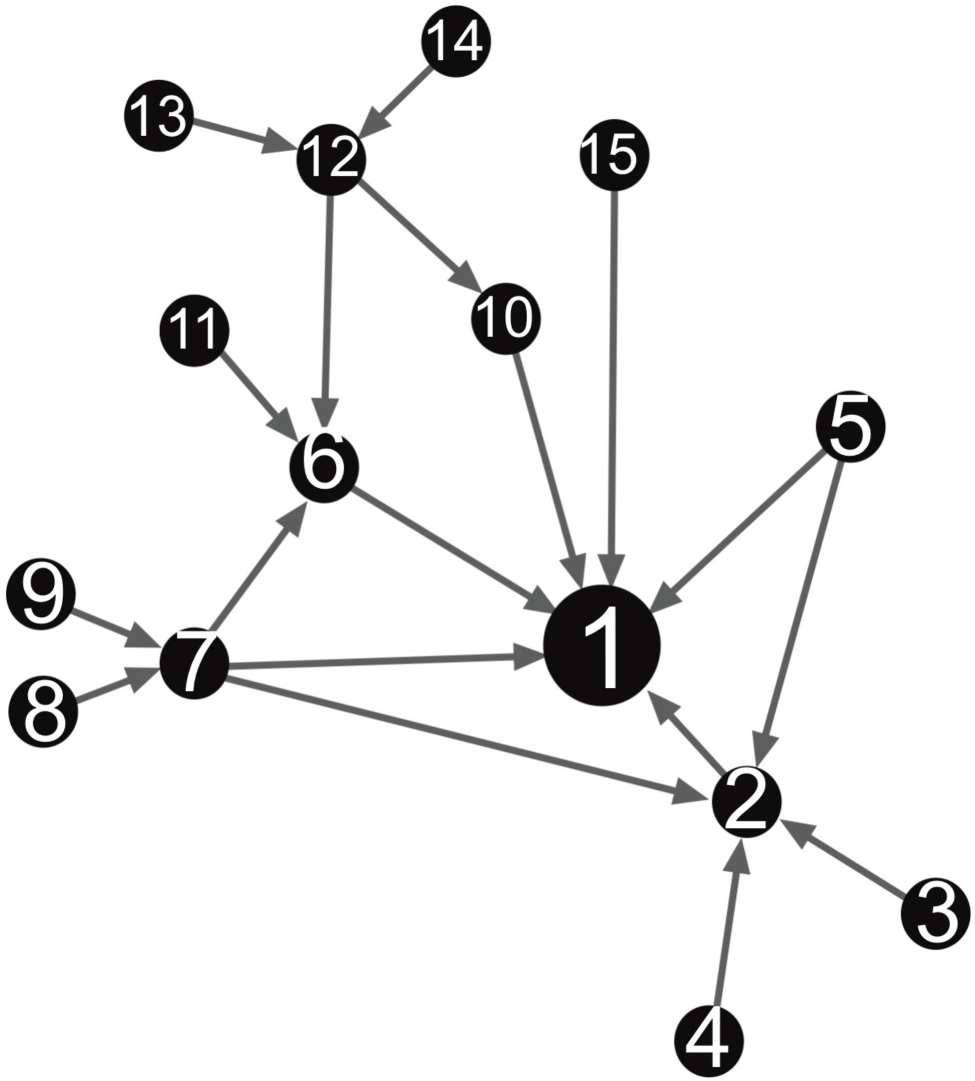
SSR = targeted random walk on networks

- for targeting need routing strategy
- simple choise **Directed Acyclic Graph** (no cycles)





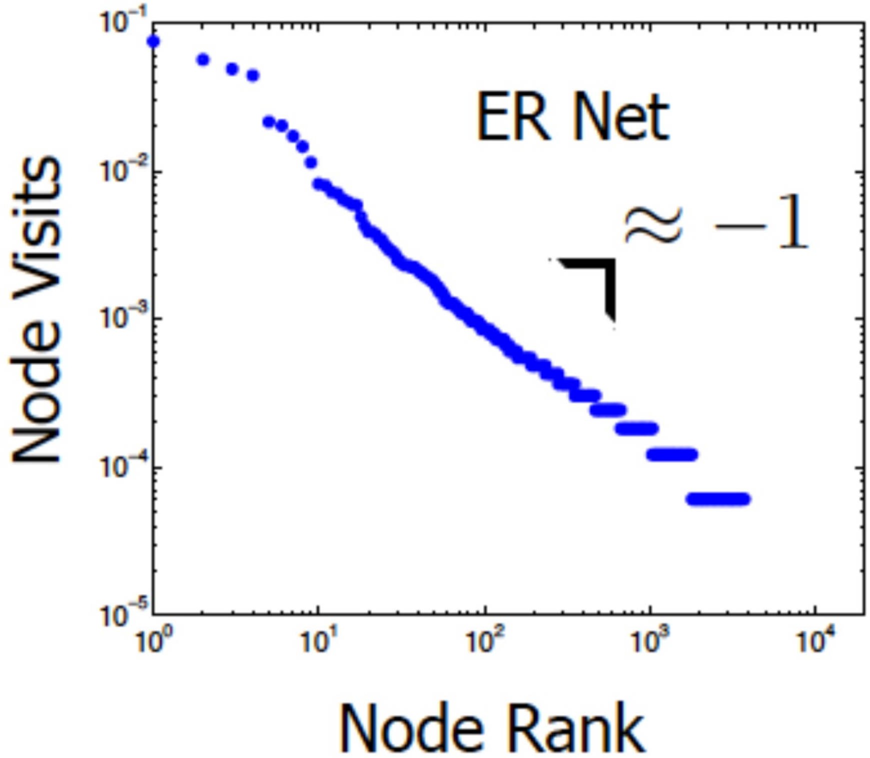
MEDICAL UNIVERSITY OF VIENNA SANTA FE INSTITUTE IIASA COMPLEXITY SCIENCE HUB VIENNA


14
13
12
15
11
10
5
6
9
7
1
8
2
3
4

Simple routing algorithm

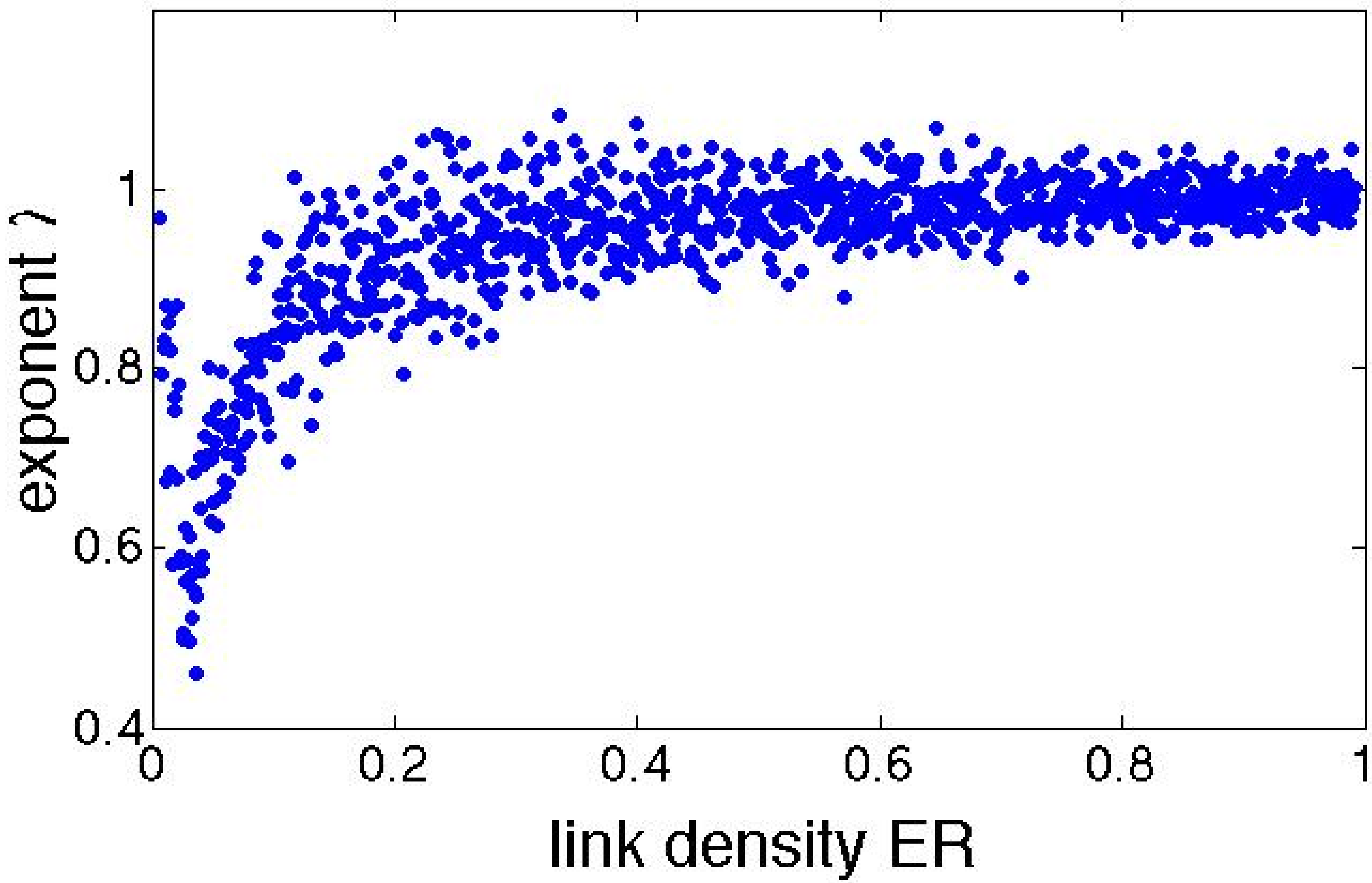
- take directed acyclic network fix it
- pick start-node
- perform a random walk from start-node to end-node (1)
- repeat many times from other start-nodes
- **prediction** visiting frequency of nodes follows **Zipf law**

All diffusion processes on DAG are SSR



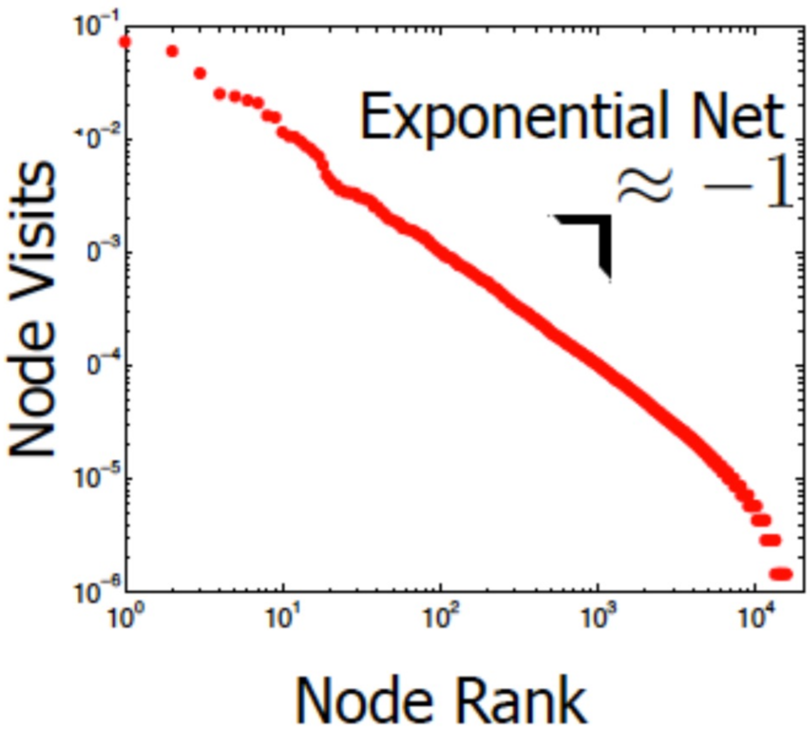
sample ER graph $\rightarrow$ direct it $\rightarrow$ pick start and end $\rightarrow$ diffuse

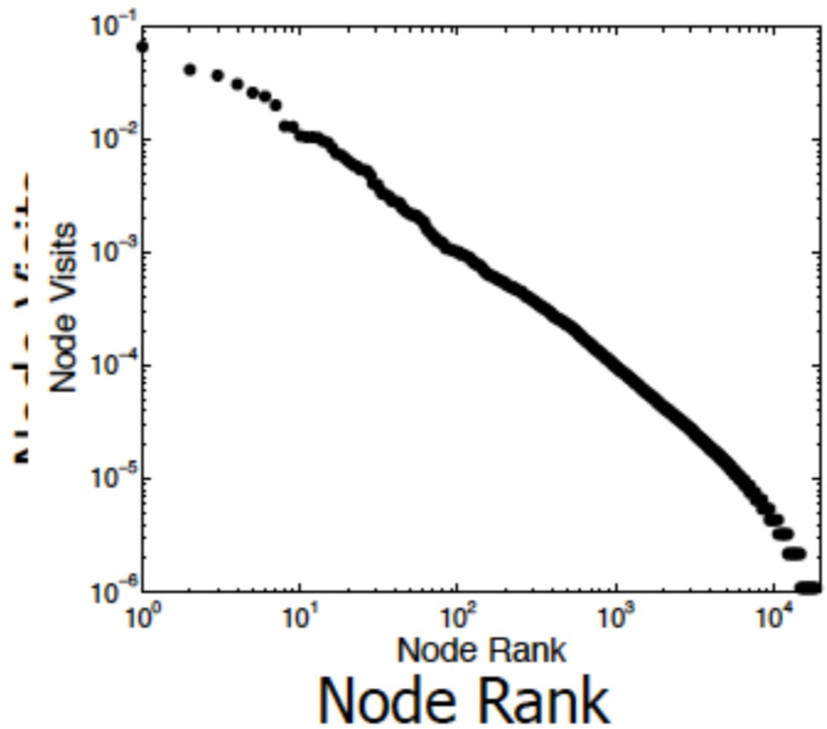
Zipf holds for any link probability



Exponential NW

HEP Co-authors

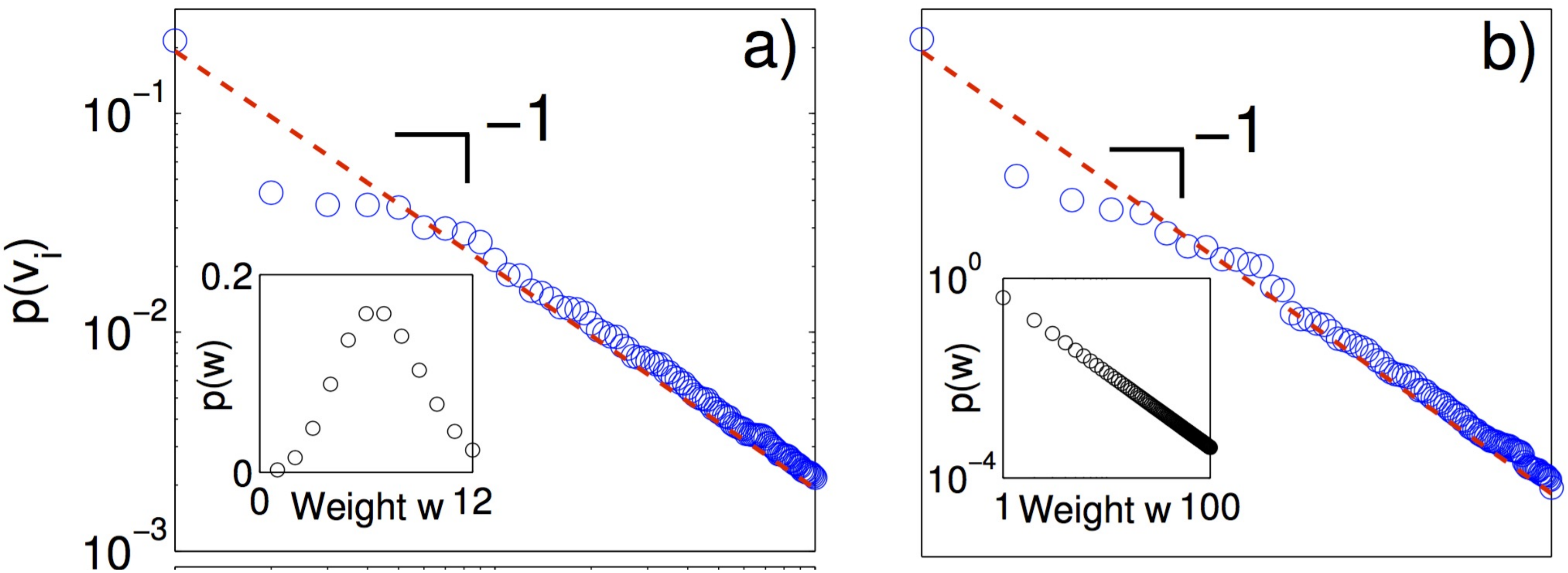




prior probabilities are practically irrelevant!

What happens if introduce weights on links?

ER Graph

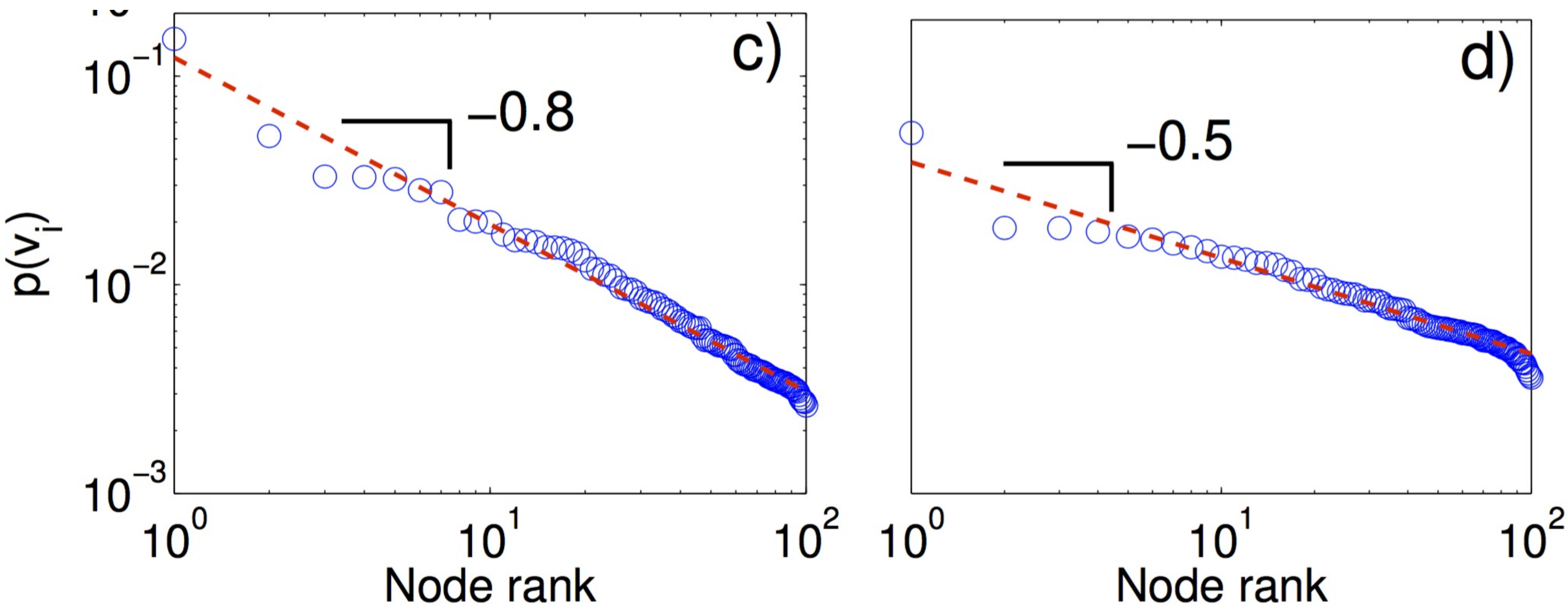


poisson weights   power weights

prior probabilities are practically irrelevant!

What happens if introduce cycles?

ER $\rightarrow$ direct it $\rightarrow$ change link to random direction with $1-\lambda$



noise level $\lambda = 0.8$ $\qquad$ $\lambda = 0.5$

Zipf's law is an immense attractor!

Zipf's law is an attractor

- no matter what the network topology is $\rightarrow$ Zipf
- no matter what the link weights are $\rightarrow$ Zipf
- if have cycles $\rightarrow$ exponent is less than one

And reality?

Every good search process is SSR!

What is good search?

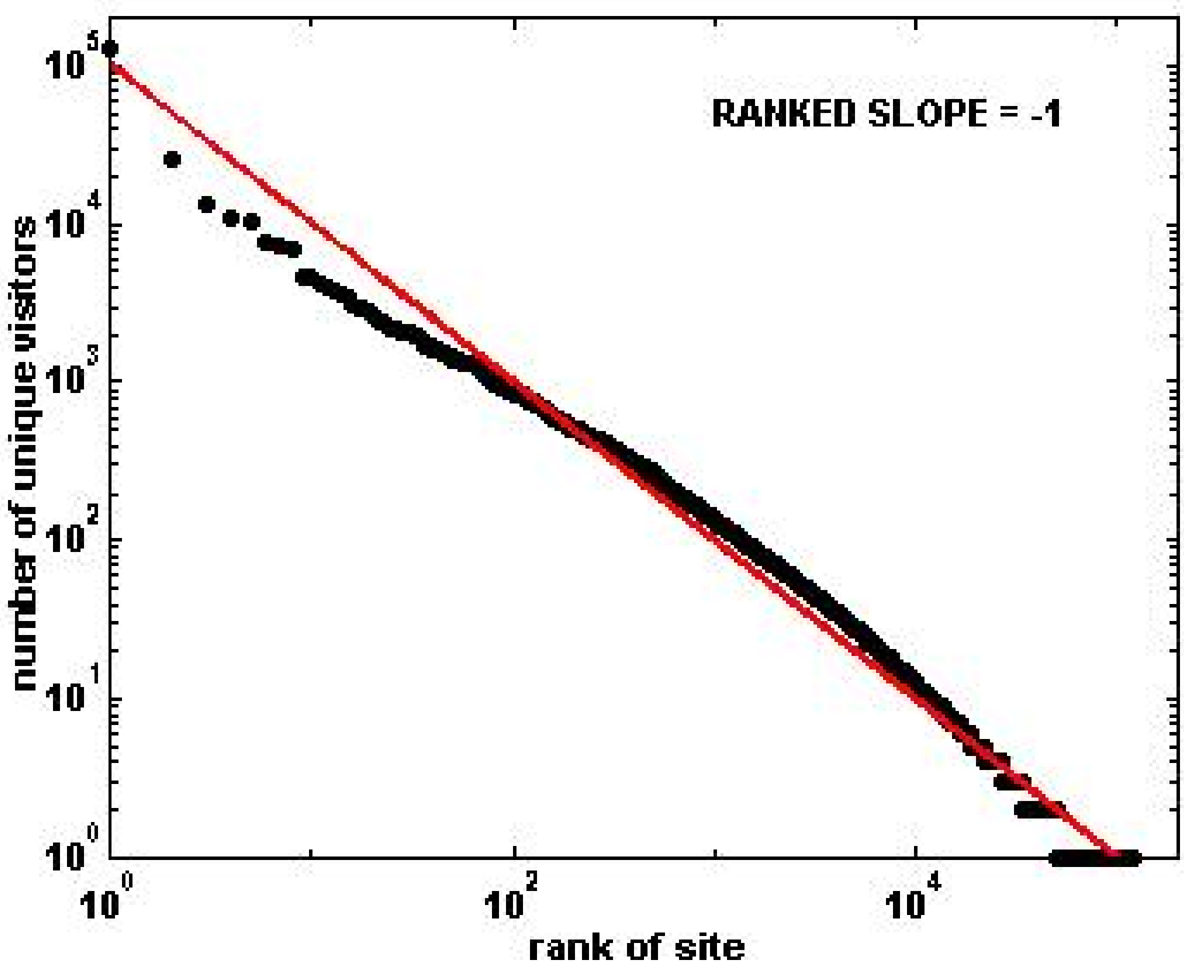
Search is a SSR process. Good search is ...

- ... if at every step you eliminate more possibilities than you actually sample
- ... every step you take eliminates branches of possibilities

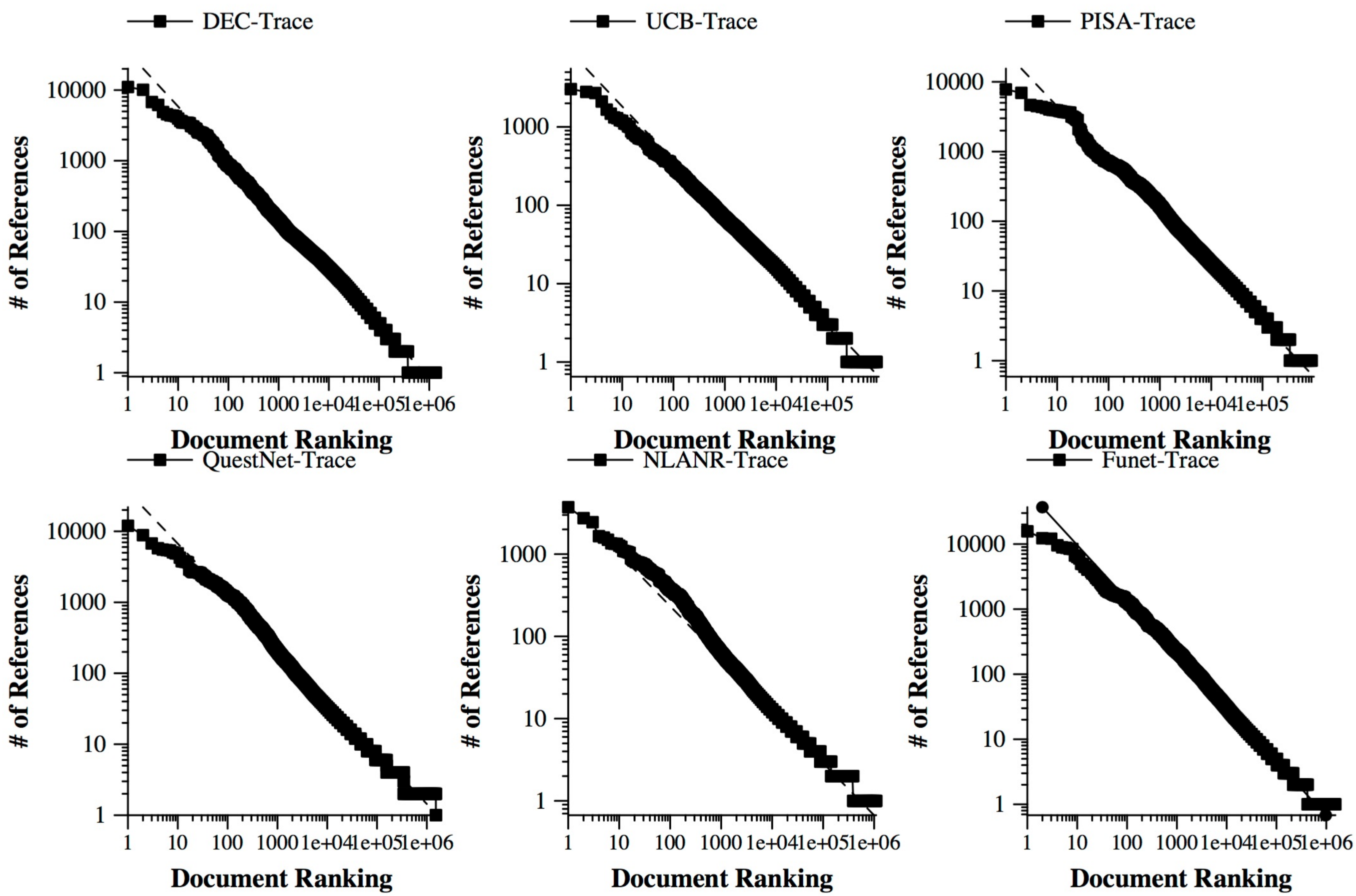
if eliminate fast enough $\rightarrow$ power law in visiting times

if eliminate too little $\rightarrow$ sample entire space (exhaustive search)

Clicking on web page is often result of search process

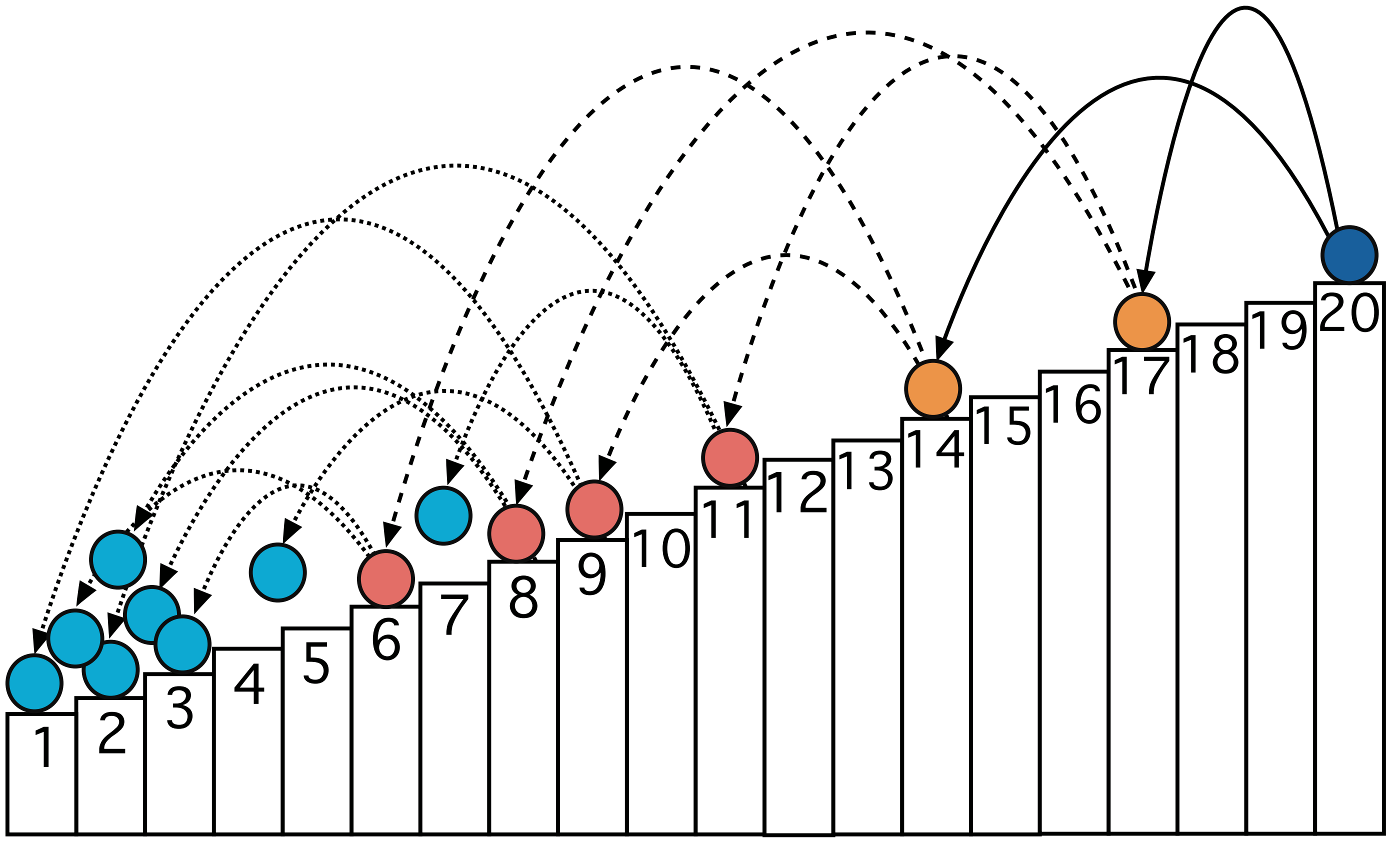


adamic & hubermann 2002



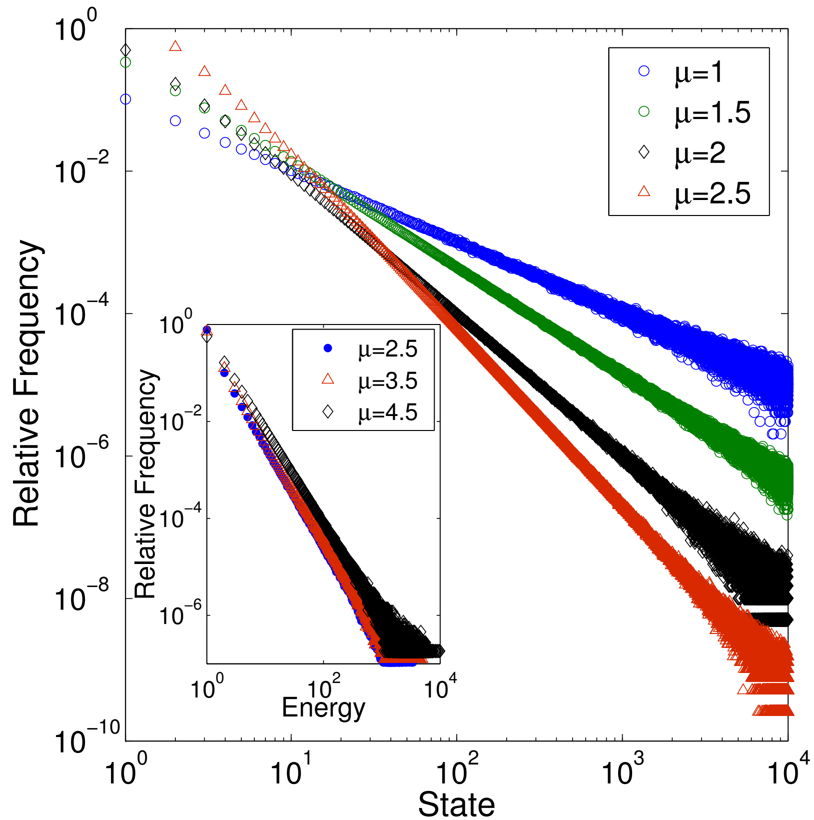
breslau et al 99

What about exponents > 1?



Multiplication factor μ

$$\rightarrow p(i) = i^{-\mu}$$


Relative Frequency
State
μ=1
μ=1.5
μ=2
μ=2.5
Energy
μ=2.5
μ=3.5
μ=4.5

What if we introduce conservation laws?

Conservation laws in SSR processes

Assume that you have duplication at every jump $\mu = 2$

If you are at i $\rightarrow$ duplicate $\rightarrow$ one jumps to j, the other to k

conservation means: $i = j + k$.

For any μ, conservation means:

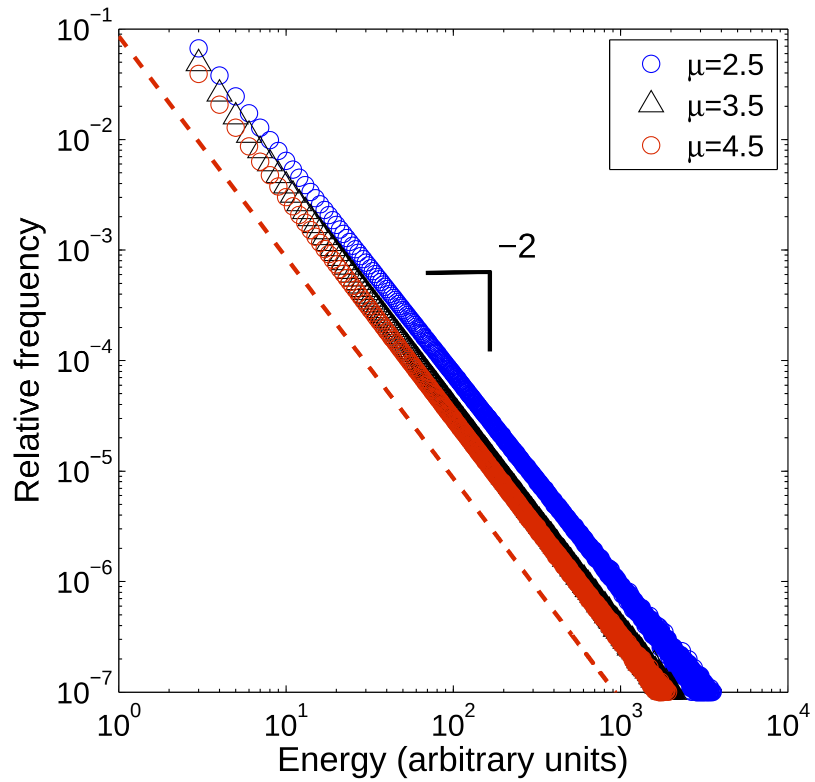
$i = \text{state}_1 + \text{state}_2 \; + \cdots + \text{state}_\mu$

$$\rightarrow p(i) = i^{-2} \quad \text{for all} \quad \mu$$

This was found by E. Fermi for particle cascades

Example for conservation: fragmentation in 2D

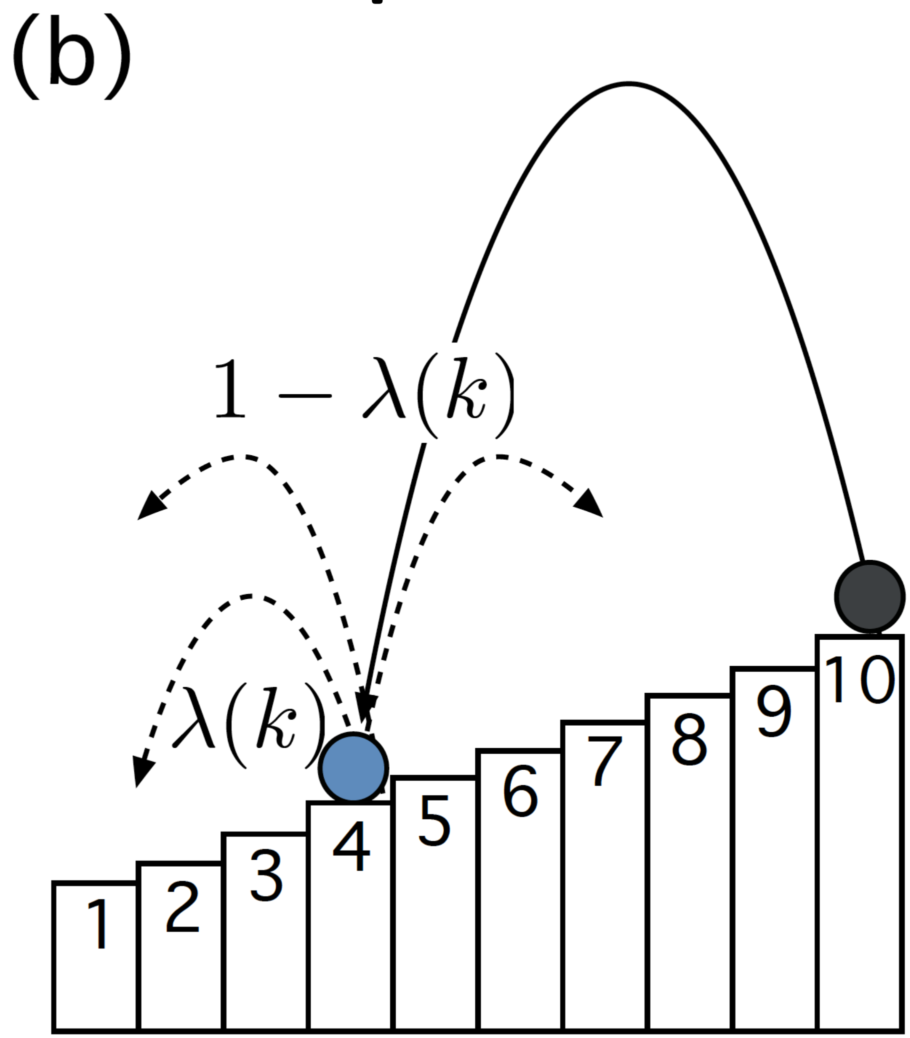
Click to start



MEDICAL UNIVERSITY OF VIENNA SANTA FE INSTITUTE IIASA COMPLEXITY SCIENCE HUB VIENNA

What if AP depends on age of process?

Assume that noise depends on the state $\lambda(i)$



Can derive the relation

$$\rightarrow \lambda(x) = -x\frac{d}{dx}\log p_{\lambda}(x)$$

That can be proved as a theorem.

Proof

The transition probabilities from state k to i are

$$p_{\mathrm{SSR}}(i|k) = \begin{cases} \lambda(k)\frac{q_i}{g(k-1)} + (1-\lambda(k))q_i & \text{if} \quad i<k \\ (1-\lambda(k))q_k \text{ otherwise} \end{cases} ,$$

$g(k)$ is the cdf of q_i, $g(k) = \sum_{i\leq k} q_i$. Observing that

$$\frac{p_{\lambda,q}(i+1)}{q_{i+1}}\left(1+\lambda(i+1)\frac{q_{i+1}}{g(i)}\right) = \frac{p_{\lambda,q}(i)}{q_i}$$

we get

$$p_{\lambda,q}(i) = \frac{q_i}{Z_{\lambda,q}}\prod_{1<j\leq i}\left(1+\lambda(j)\frac{q_j}{g(j-1)}\right)^{-1} \sim \frac{q(i)}{Z_{\lambda,q}}e^{-\sum_{j\leq i}\lambda(j)\frac{q(j)}{g(j-1)}}$$

$Z_{\lambda,q}$ is the normalisation constant. For uniform priors, taking logs and going to continuous variables gives the result $\lambda(x) = -x\frac{d}{dx}\log p_\lambda(x)$.

Special cases $\lambda(x) = -x\frac{d}{dx}\log p_\lambda(x)$

- Zipf: no noise $\rightarrow$ $p(x) = x^{-1}$
- Power-law: $\lambda(x) = \alpha$ $\rightarrow$ $p(x) = x^{-\alpha}$
- Exponential: $\lambda(x) = \beta x$ $\rightarrow$ $p(x) = e^{-\beta(x-1)}$
- Power-law + cut-off: $\lambda(x) = \alpha + \beta x$ $\rightarrow$ $p(x) = x^{-\alpha}e^{-\beta x}$
- Gamma: $\lambda(x) = 1 - \alpha + \beta x$ $\rightarrow$ $p(x) = x^{\alpha-1}e^{-\beta x}$

Special cases $\lambda(x) = -x\frac{d}{dx}\log p_\lambda(x)$

- Normal: $\lambda(x) = 2\beta x^2 \quad \rightarrow \quad p(x) = e^{-\frac{\beta}{2}(x-1)^2}$
- Stretched exp: $\lambda(x) = \alpha\beta|x|^\alpha \quad \rightarrow \quad p(x) = e^{-\frac{\beta}{\alpha}(x-1)^\alpha}$
- Gompertz: $\lambda(x) = (\beta e^{\alpha x} - 1)\beta x \quad \rightarrow \quad p(x) = e^{\beta x - \alpha e^{\beta x}}$
- Weibull: $\lambda(x) = \beta^{-\alpha}\alpha x^\alpha + \alpha - 1 \rightarrow p(x) = \left(\frac{x}{\beta}\right)^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^\alpha}$
- Tsallis: $\lambda(x) = \frac{\beta x}{1-\beta x(1-Q)} \quad \rightarrow \quad p(x) = (1-(1-Q)\beta x)^{\frac{1}{1-Q}}$

Problems that are of SSR nature

- search, e.g. targeted diffusion
- language: sentence formation
- fragmentation: break spaghetti
- sequences of human behavior
- games: go
- internet communication

Conclusions

- many history dependent processes are SSR
- SSR offers new route to scaling – huge applicability
- SSR is an extremely robust Zipf – priors don't matter
- targeted diffusion on networks leads to Zipf's law,
no matter what NW looks like $\rightarrow$ **attractor**
- all good search – has mechanism to generate power laws
- noise level determines power exponent
- if noise is state dependent – get practically all distribution
- Noise and SSR explain practically every statistics