

Patterns of complexity: From the correspondence of Darwin and Einstein to the occurrence of numbers in the World Wide Web

João Gama Oliveira (University of Aveiro / Notre Dame)

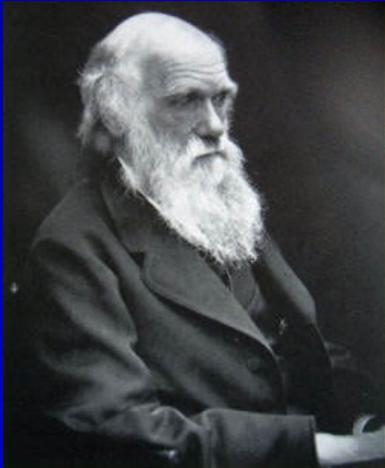
José Fernando Mendes (University of Aveiro)

Albert-László Barabási (University of Notre Dame, USA)

Sergey Dorogovtsev (University of Aveiro)



Darwin and Einstein correspondence patterns



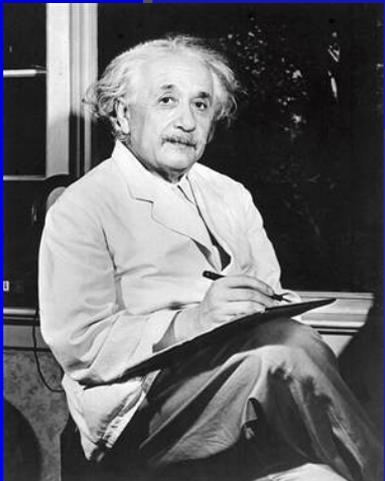
1809-82

Sent:

7,591

Received:

6,530



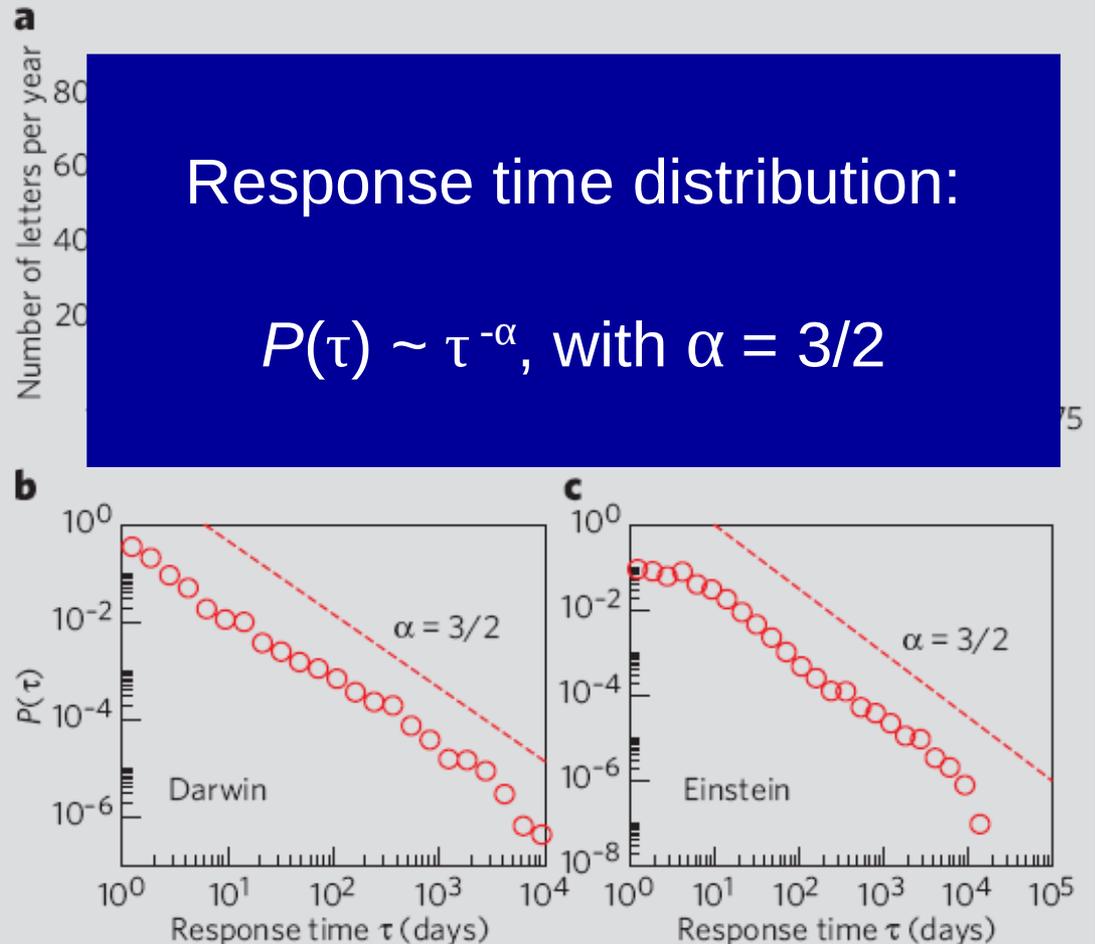
1879-1955

Sent:

14,500

Received:

16,200



Modeling the Dynamics of Human Behavior

Poisson Processes

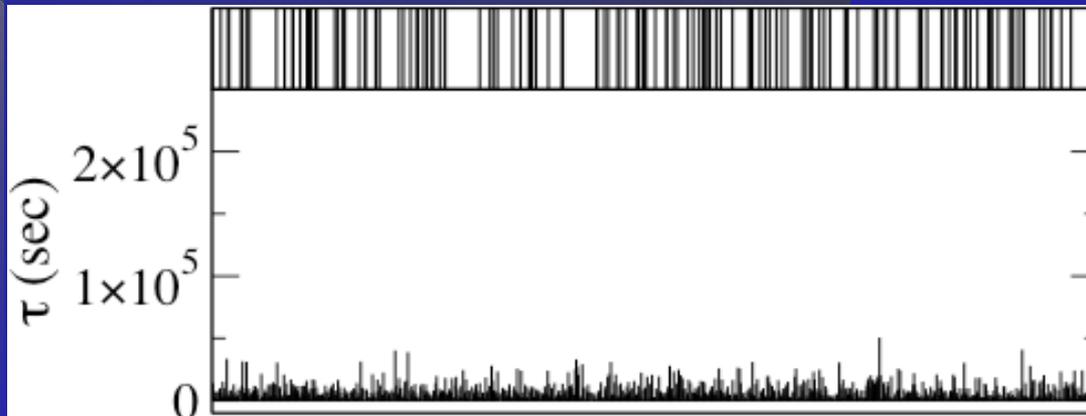
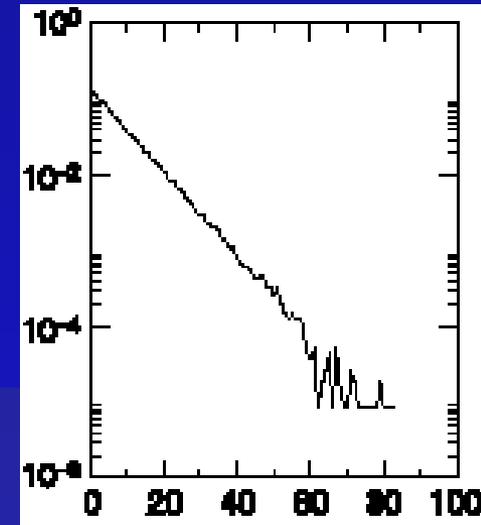
Current models assume that the timing of human activities is well approximated by a Poisson process

Basic assumption: Events take place randomly, at a constant rate λ

Consequence: The timing of events follows a Poisson process.

τ : time interval between two consecutive events
(*waiting or inter-event time*)

$$P(\tau) = \lambda \exp(-\lambda \tau)$$



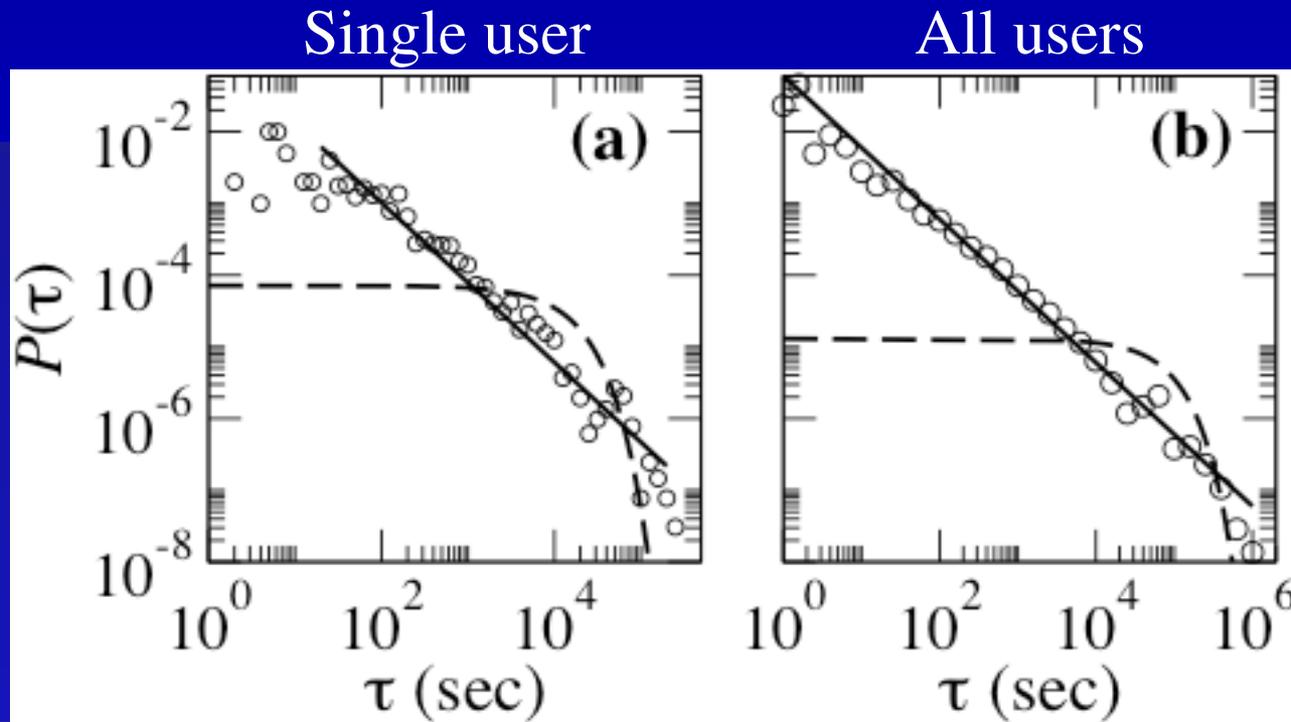
[Moivre (1718)]

[Poisson (1837)]

[Erlang (1909)]

**Is the timing of human
initiated events
purely random?**

Email communications



τ : time between two consecutive emails sent by a user

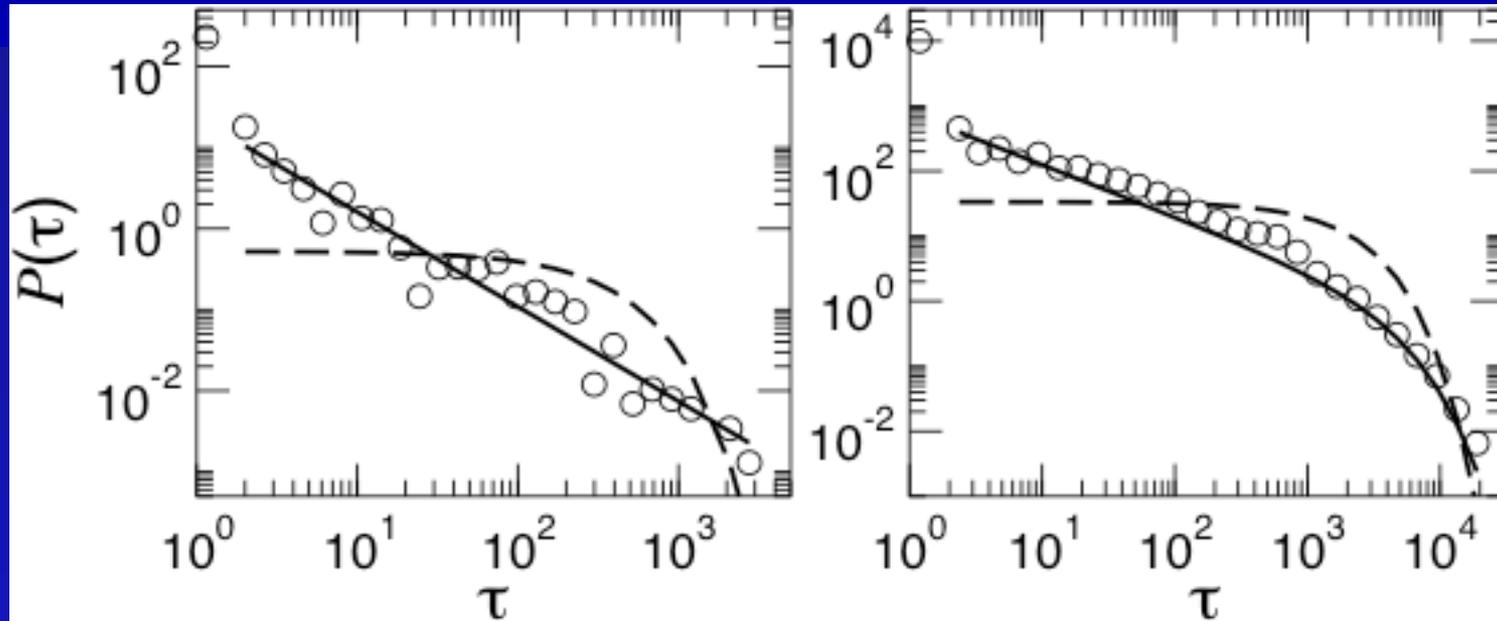
- Number of users: 3,180
- Number of events: 129,135
- Time period: 3 months
- Time resolution: 1 second

- [Barabási, Nature (2005)]
- [Eckmann et al., PNAS (2004)]
- [Ebel et al., Phys. Rev. E (2002)]

Library visitation

Single user

All users



τ : time between two consecutive library visits (checkout records)

- Number of users: 2,247
- Number of events: 48,408
- Time period: 3 years
- Time resolution: 1 minute

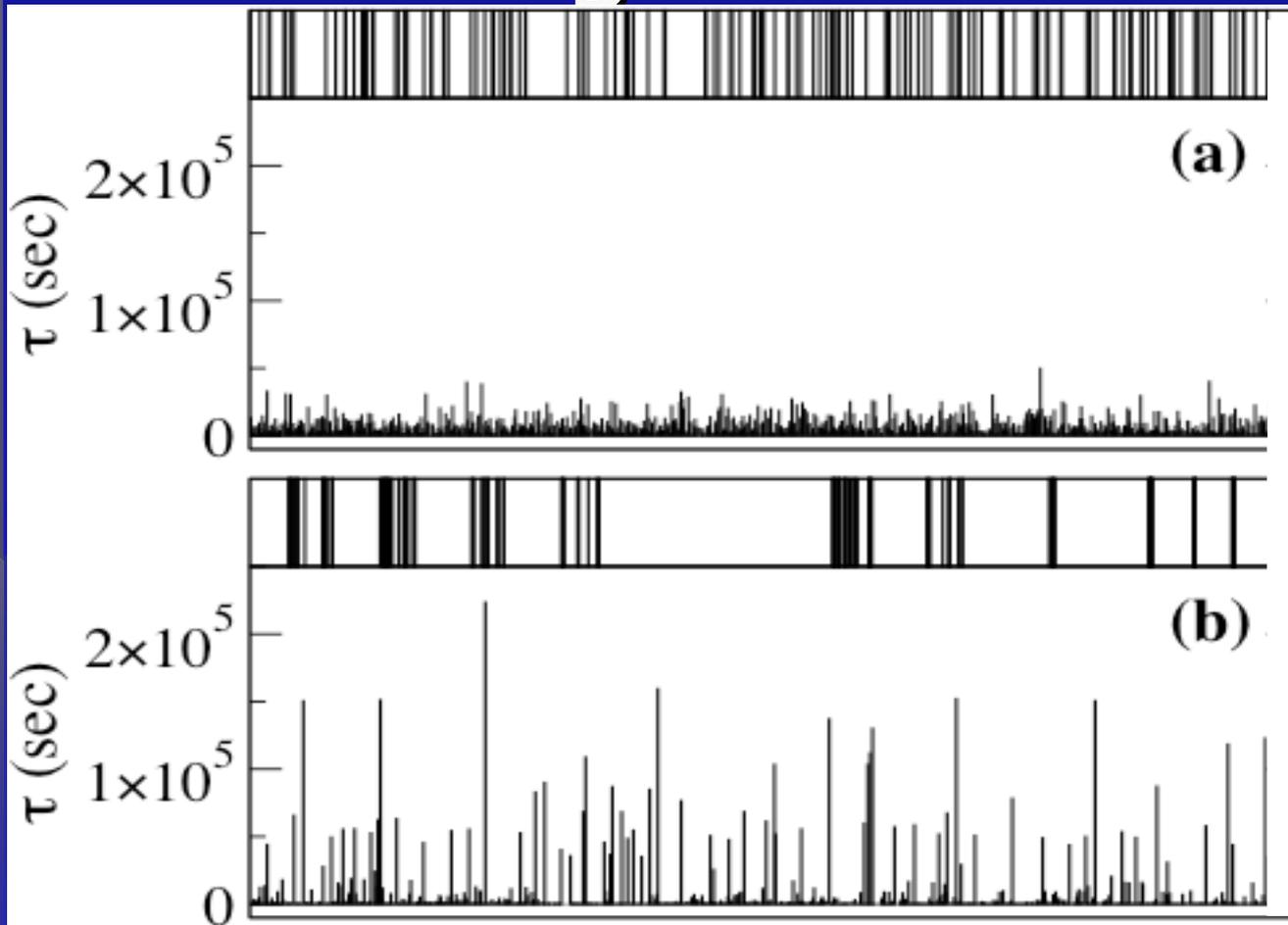
[Vázquez et al., Phys. Rev. E (in press)
arXiv: physics/0510117]

Other Human Driven Events

- instant messages during online chats [Dewes, C. *et al.*, in Proc. 2003 ACM SIGCOMM Conf. on Internet Measurement (IMC-03) (ACM Press, New York, 2003)]
- job submissions on a supercomputer [Kleban, S.D. & Clearwater, S.H. Proc. SC2003 <http://www.sc-conference.org/sc2003/paperpdfs/pap222.pdf> (2003).]
- directory listings [Paxson, V. & Floyd, S. IEEE/ACM Trans. Netw. **3**, 226 (1996)]
- file transfers (FTP requests) [Paxson, V. & Floyd, S. IEEE/ACM Trans. Netw. **3**, 226 (1996)]
- printing jobs [Harder, U. & Paczuski, M. <http://xxx.lanl.gov/abs/cs.PF/0412027> (2004)]
- individual trades in currency futures [Masoliver, J., Montero, M. & Weiss, G. H. *Phys. Rev. E* **67**, 021112 (2003)]
- online games played by a user [Henderson, T. & Nhatti, S. *Proc. 9th ACM Int. Conf. on Multimedia* 212–220 (ACM Press, New York, 2001)]

Poisson vs. heavy tailed processes

Summarizing...



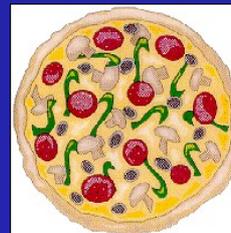
Poisson Process
Heavy tailed process

What is going on?

Queuing theory

Priority queues

To do list	Random order	First In First Out	With priorities
• Clean the house	1	1	0.3
• Give a talk	6	2	0.9999
• Call the dentist	3	3	0.1
• Get a date	4	4	0.7
• Check emails	5	5	0.6
• Read a book	2	6	0.8



Back to the Darwin and Einstein correspondence... The model



Waiting time distribution:

$$P(\tau) \sim \tau^{-3/2} \exp(-\tau / \tau_0)$$

If $\lambda \geq \mu$, $\tau_0 \rightarrow \infty$ and $P(\tau) \sim \tau^{-3/2}$

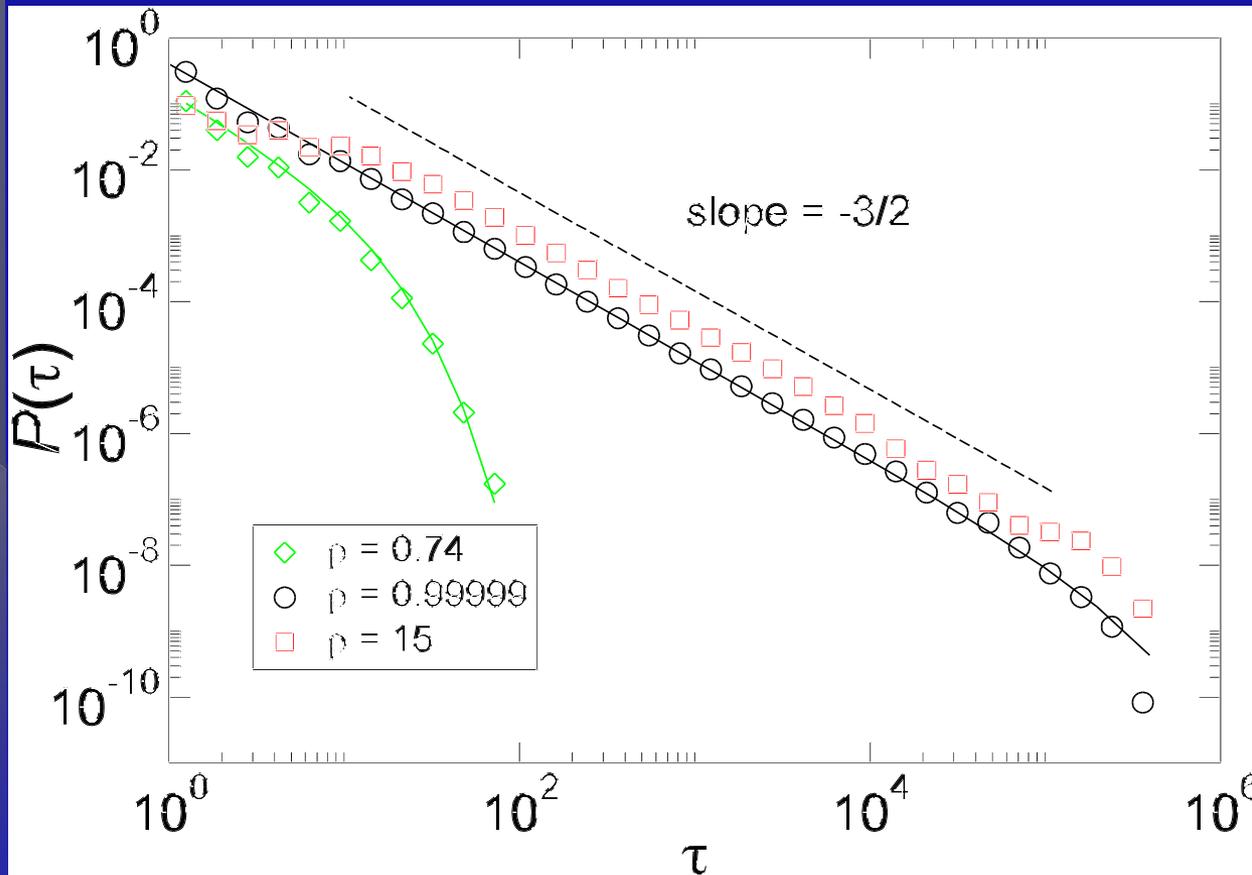
L-1

L

L+1

Numerical simulation of the model

$$P(\tau) \sim \tau^{-3/2} \exp(-\tau / \tau_0)$$



$$\rho = \lambda / \mu$$

Some open questions

- What is the time elapsed between two consecutive sent/received letters?
- Correspondence with specific persons
- Is the observed pattern really global?
- Why does email communication ($\alpha = 1$) belong to a different universality class than surface mail communication ($\alpha = 3/2$) ?

Changing the subject...

Frequency of occurrence of numbers in World Wide Web

- The first digit problem [Newcomb 1881, Benford 1938]

In databases the first digit of numbers is not uniformly distributed...

1	→	30%	Benford's law!
2	→	18%	
3	→	12%	
...			

- How frequently do numbers occur in databases?

Using the Web as a database...

Google Web Images Groups News Froogle Local more »

2005 Search [Advanced Search](#) [Preferences](#)

Web Results 1 - 10 of about 2,540,000,000 for 2005. (0.04 seconds)

[The 14th International World Wide Web Conference 2005](#)
13 May, 2005: WWW2006 in Edinburgh Scotland will be held from May 22, 2006 to May 26, 2006. [All the news]. General questions about WWW2005 may be sent to ...
[www2005.org/](#) - 12k - [Cached](#) - [Similar pages](#)

[World Year of Physics 2005](#)
Plans to bring the excitement of physics to the public and inspire a new generation of scientists. Includes information about projects, events and Einstein, ...
[www.physics2005.org/](#) - [Similar pages](#)

[2005 - Wikipedia, the free encyclopedia](#)
2005 is a common year starting on Saturday of the Gregorian calendar and is the ... 2005 is the World Year of Physics, the Year of the Rooster in the ...
[en.wikipedia.org/wiki/2005](#) - 101k - 26 Jul 2005 - [Cached](#) - [Similar pages](#)

Number of web pages containing "2005" as reported by Google in July 2005

Typical numbers with high frequencies of occurrence

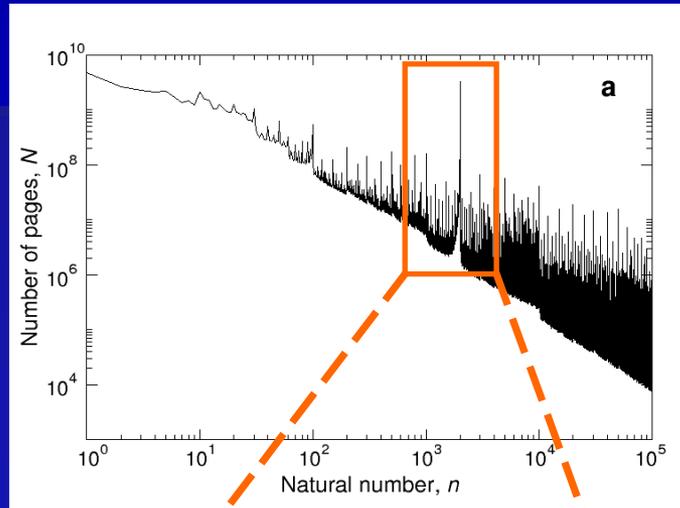
TABLE I: Typical numbers with high frequencies of occurrence

Example	Description
1000	powers of 10
2460, 2465	'round' numbers: multiples of 10 and 5
666, ^a 131313	numbers easy to remember or symmetric
$512 = 2^9$	powers of 2
666, ^a 777	numbers with strong associations
78701	popular zip codes
866, 877	toll free telephone numbers
1812	important historical dates
747, 8086	serial numbers of popular products
314159	beginning parts of mathematical constants

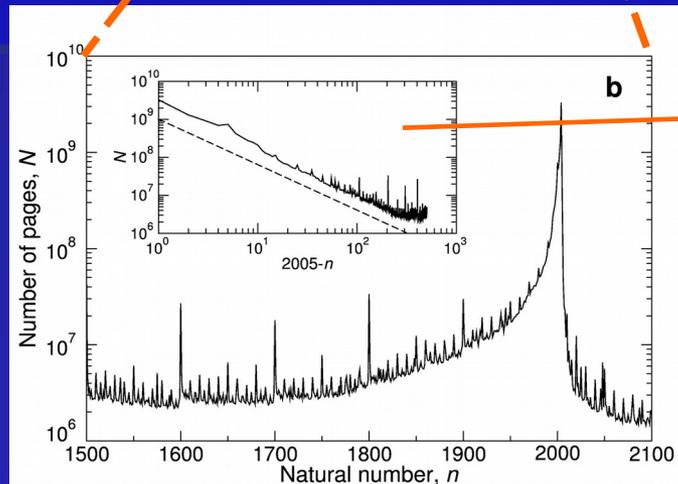
^aA number may occur simultaneously in several lines of the table.

Some results

- Frequency in the range from 1 to 100,000:



- Numbers associated to western calendar years:

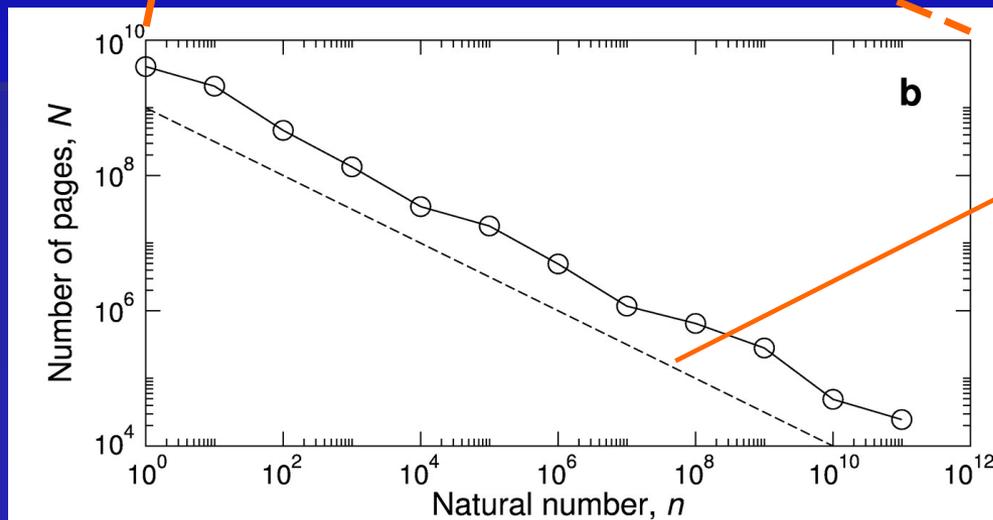
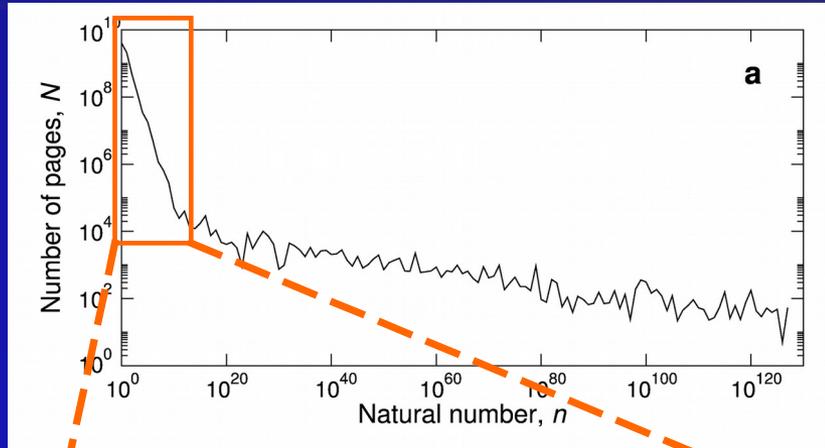


Power law decay, as we move to the past:

$$N \sim (2005-n)^{-1.2}$$

Some more results...

- Frequency of powers of 10 (up to the huge 10^{127}):



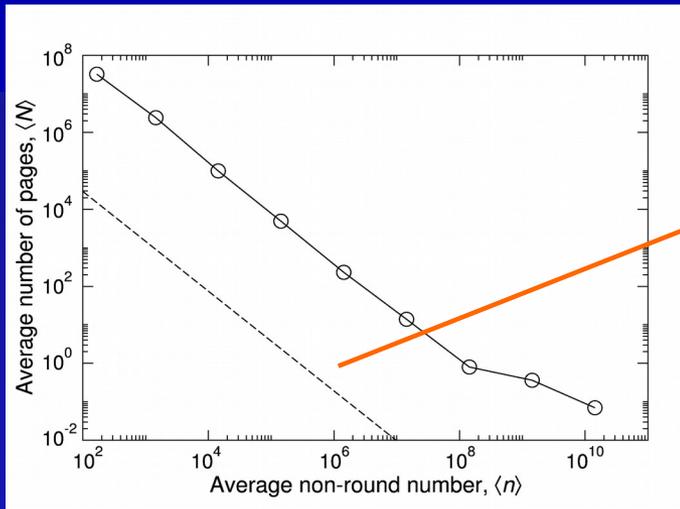
Slope = $-1/2$



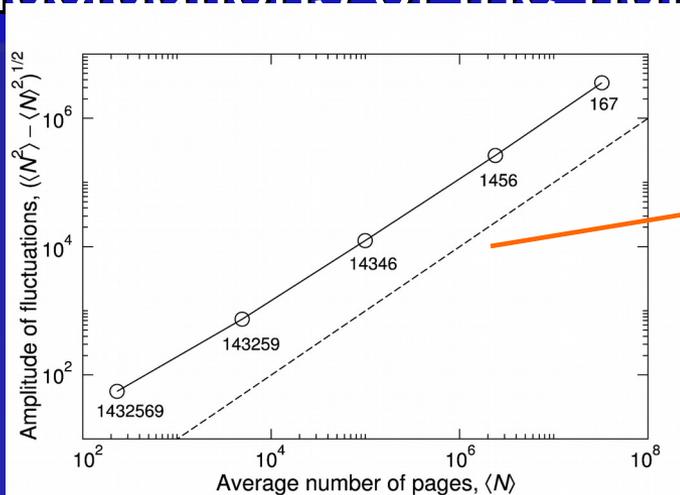
$$N \sim n^{-1/2} = 1 / \sqrt{n}$$

Yet more results...

- Non-round numbers:



- Fluctuations of the number of Web pages:

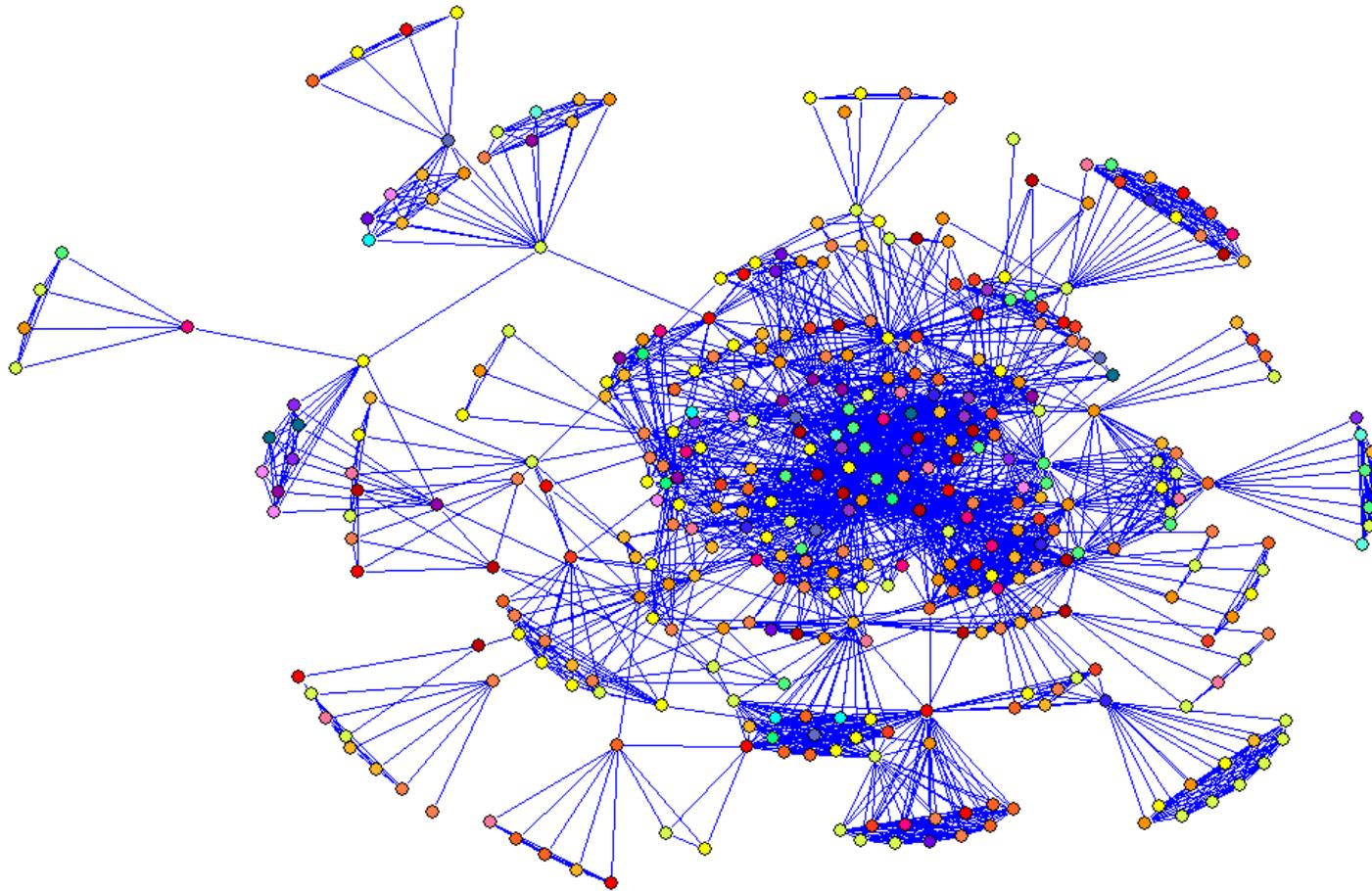


So...

The distribution of numbers in the Web follows a complex and scale-free pattern: the numbers separate in different classes that occur with different frequencies following power laws, whose exponents identify the class.

Some open questions...

- How do the occurrence frequencies of specific numbers vary in time?
- How do different numbers correlate and co-occur in Web documents?
- How does the statistics of numbers relate to the organization of human memory and to semantics?
- How do the various numeral systems influence the general statistics of numbers?
- Is the Benford law fulfilled for the numbers contained in the Web?



More informations:

<http://sweet.ua.pt/~a29583>