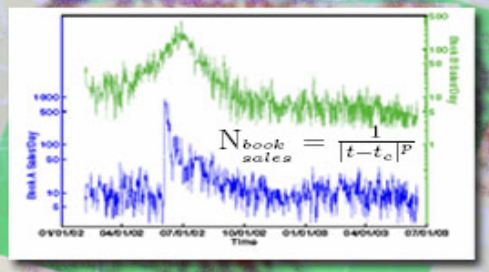
Endogenous versus exogenous dynamics and scaling laws in YouTube, Open Source Softwares and Cyber-risks

Didier Sornette
Chair of Entrepreneurial Risks
ETH Zurich
(Swiss Federal Institute of Technology, Zurich)
Department of Management, Technology and Economics
http://www.er.ethz.ch/

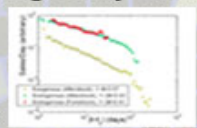
Collaborators:
Riley Crane (post-doc)
Thomas Maillart (PhD student)

Social networks

Endogenous vs Exogenous Shocks



The response of the bookbuyers can be characterized with power laws and rigorously described using theories of epidemic propagation with memory in a network of acquaintances. This reflects a self-similar fractal-like response of human decision making.



Complex Interacting Social Networks

The organization of social networks displays many interesting scaling properties from "small-world" to "scale-free" which have consequences for the dynamical behavior of the system. The mathematical characterization of the topology is currently a very active area of research.

Publishers face the problem of optimal allocation of resources

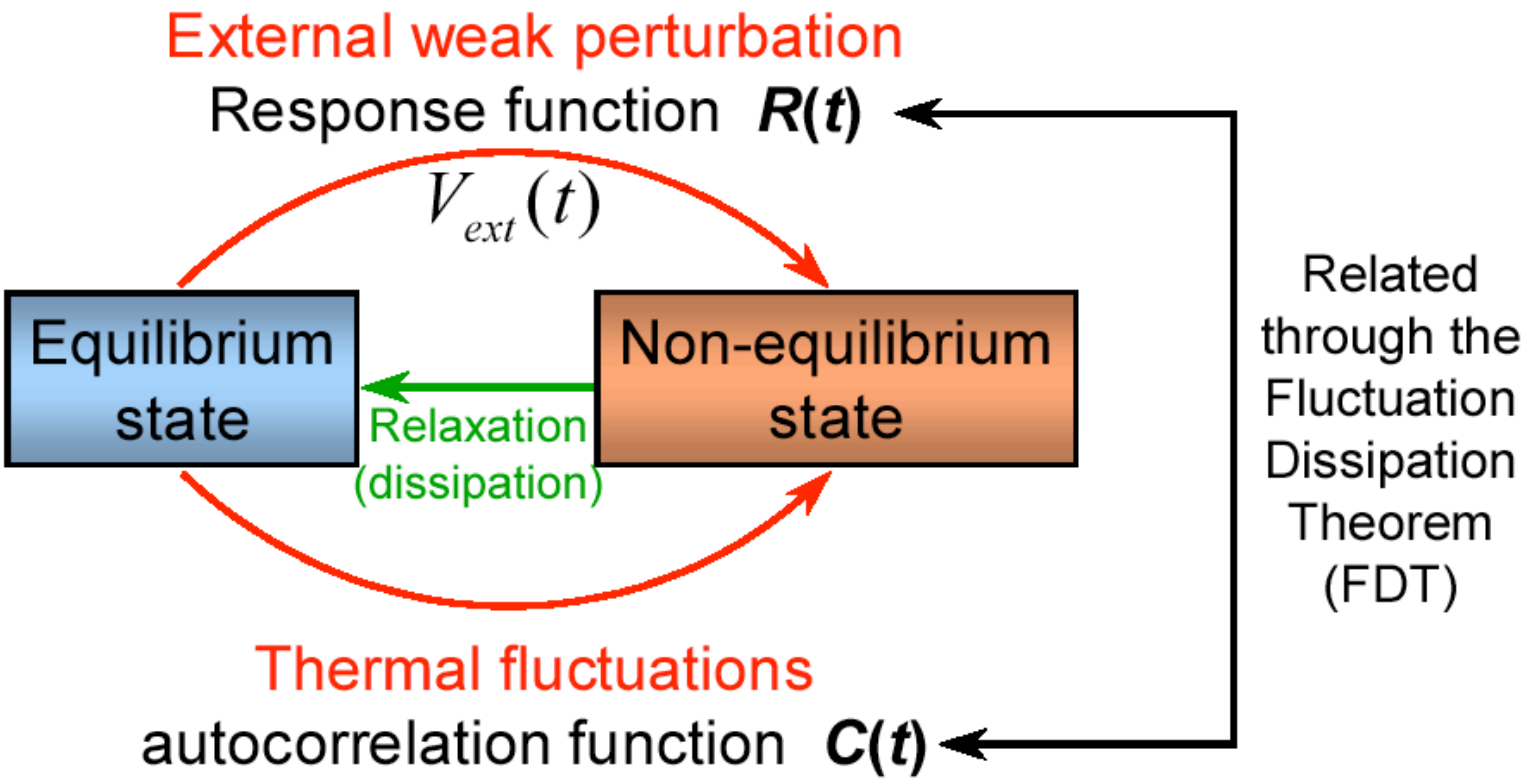
This research will give publishers the tools they need to optimally allocate their marketing and publishing resources allowing them to modify their business model in a constantly evolving marketplace.

Other Applications

Commercial sales (Books, CDs, DVDs, etc.), Theatre Attendance of Hollywood Movies Open Source Projects, Search Engines, Real Time Measure of the Social Climate.

Guidelines from Physics: perturb and study the response

Linear Response Theory



AMAZON BOOK SALES

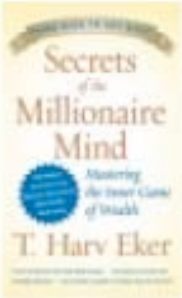
Updated every hour

1. 

Harry Potter and the Half-Blood Prince (Book 6)
by J. K. Rowling, Mary GrandPrÈ (Illustrator)
Publication Date: July 16, 2005
Not Yet Published (Rate this item)
This is the 6th item in The Harry Potter Series.

List Price: $29.99
Price: **$17.99**
You Save: $12.00 (40%)

Pre-order this item

2. 

Secrets of the Millionaire Mind: Mastering the Inner Game of Wealth
by T. Harv Eker
Average Customer Review:
Usually ships in 24 hours
(Rate this item)

List Price: $19.95
Price: **$13.57**
You Save: $6.38 (32%)

Add to cart | Add to Wish List

Used & new from $12.62

3. 

Blood Brother: 33 Reasons My Brother Scott Peterson Is Guilty
by Anne Bird
Average Customer Review:
Usually ships in 24 hours
(Rate this item)

List Price: $25.95
Price: **$17.13**
You Save: $8.82 (34%)

Add to cart | Add to Wish List

Used & new from $15.48

4. 

See related items

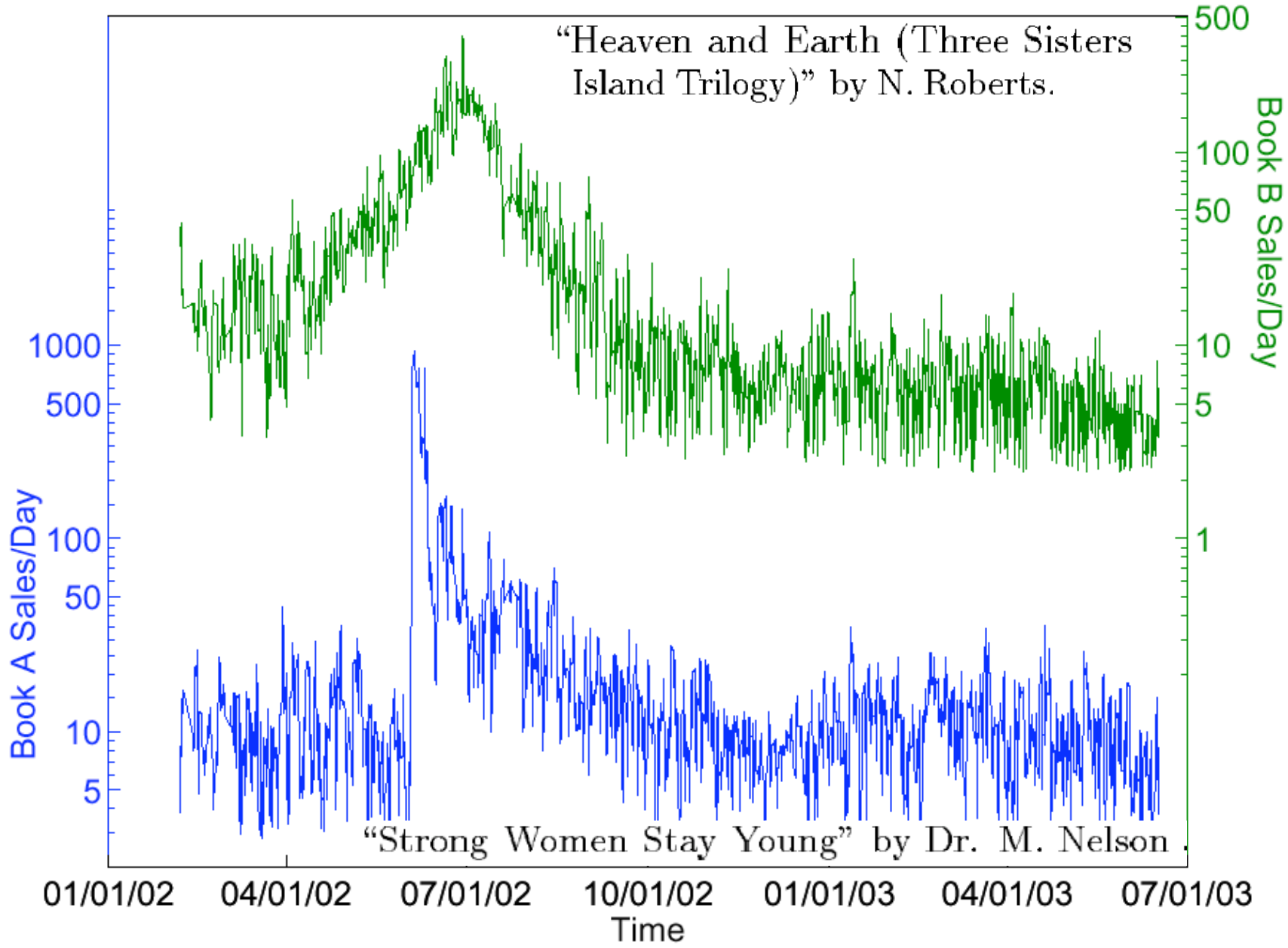
The Kite Runner
by Khaled Hosseini
Average Customer Review:
Usually ships in 24 hours
(Rate this item)

List Price: $14.00
Price: **$9.80**
You Save: $4.20 (30%)

Add to cart

Used & new from $8.87

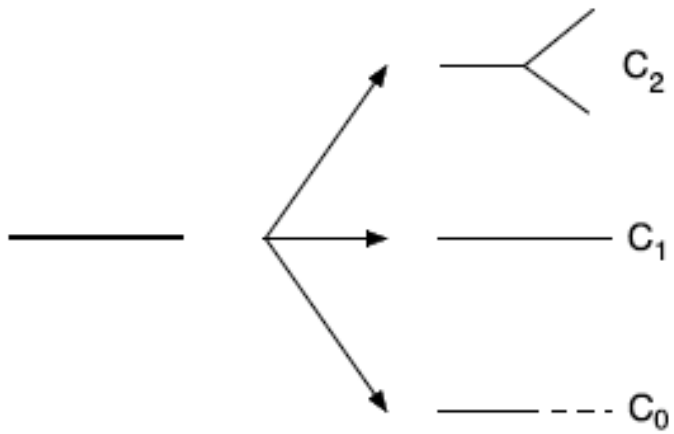
20:00 4 March. 2005



D. Sornette et al., Phys. Rev. Letts. 93 (22), 228701 (2004)

Epidemic processes by word-of-mouth

Simplest example of branching



Definition of the branching model:
starting from an existing branch, with probability C_0 the branch stops at the next step; with probability C_1, the branch continues to grow at the next step; with probability C_2, it develops two branches

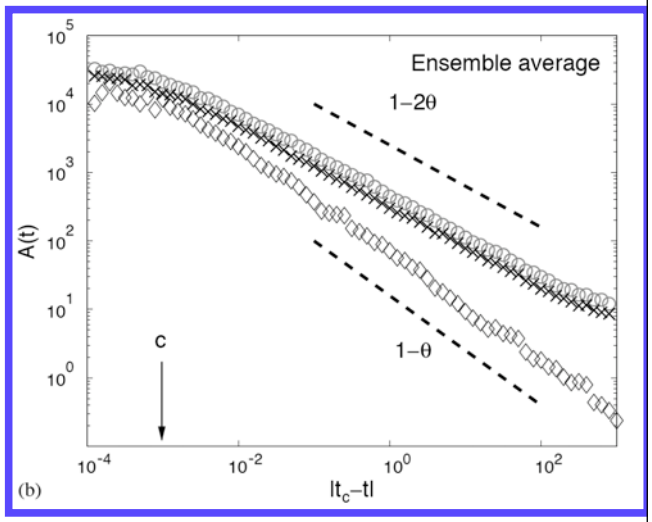
1. if $C_1 + 2C_2 < 1$, the average number of generations is finite and all earthquakes are finite. Their size distribution is an exponential.
2. if $C_1 + 2C_2 > 1$, the probability to generate a "run away" (i.e. an event of infinite size) becomes nonzero. This is similar to being above the threshold in the percolation model presented in Chap. 12.
3. if $C_1 + 2C_2 = 1$, the system is critical and the size distribution of events is a power law as we show below.

The critical condition $C_1 + 2C_2 = 1$ together with the normalization $C_0+C_1+C_2 = 1$ yields the condition $C_0 = C_2$ at criticality.

$$P(E) \simeq A e^{-aE} E^{-(1+\mu)} \qquad a \simeq \frac{(C_0 - C_2)^2}{4C_0} \qquad \mu = \frac{1}{2}$$

Mean field theory of Hawkes self-exciting conditional Poisson Process

$$A(t) = \int_{-\infty}^{t} d\tau \; \eta(\tau) \; K(t-\tau)$$



Exogeneous shock

$$\boxed{\mathrm{E}_{\mathrm{exo}}[A(t)] = A_0 \; K(t) + n\langle\eta\rangle}$$

$$n = \int_0^{+\infty} \tau \; K(\tau)$$

Endogeneous shock

$$\mathrm{E}[X(t)|Y = A_0] - \mathrm{E}[X(t)] = (A_0 - \mathrm{E}[Y]) \; \frac{\mathrm{Cov}(X(t), Y)}{\mathrm{E}[Y^2]}$$

$$\mathrm{Cov}(A(t), A(0)) = \int_{-\infty}^{0} d\tau \; K(t-\tau) \; K(-\tau)$$

$$\boxed{\mathrm{E}_{\mathrm{endo}}[A(t)|A(0) = A_0] \propto A_0 \int_0^{+\infty} du \; K(t+u) \; K(u)}$$

Theory: Null Hypotheses

- The tests are about the slopes of the response functions, conditional on the class of peak determined by the slope of the growth AT CRITICALITY n=1

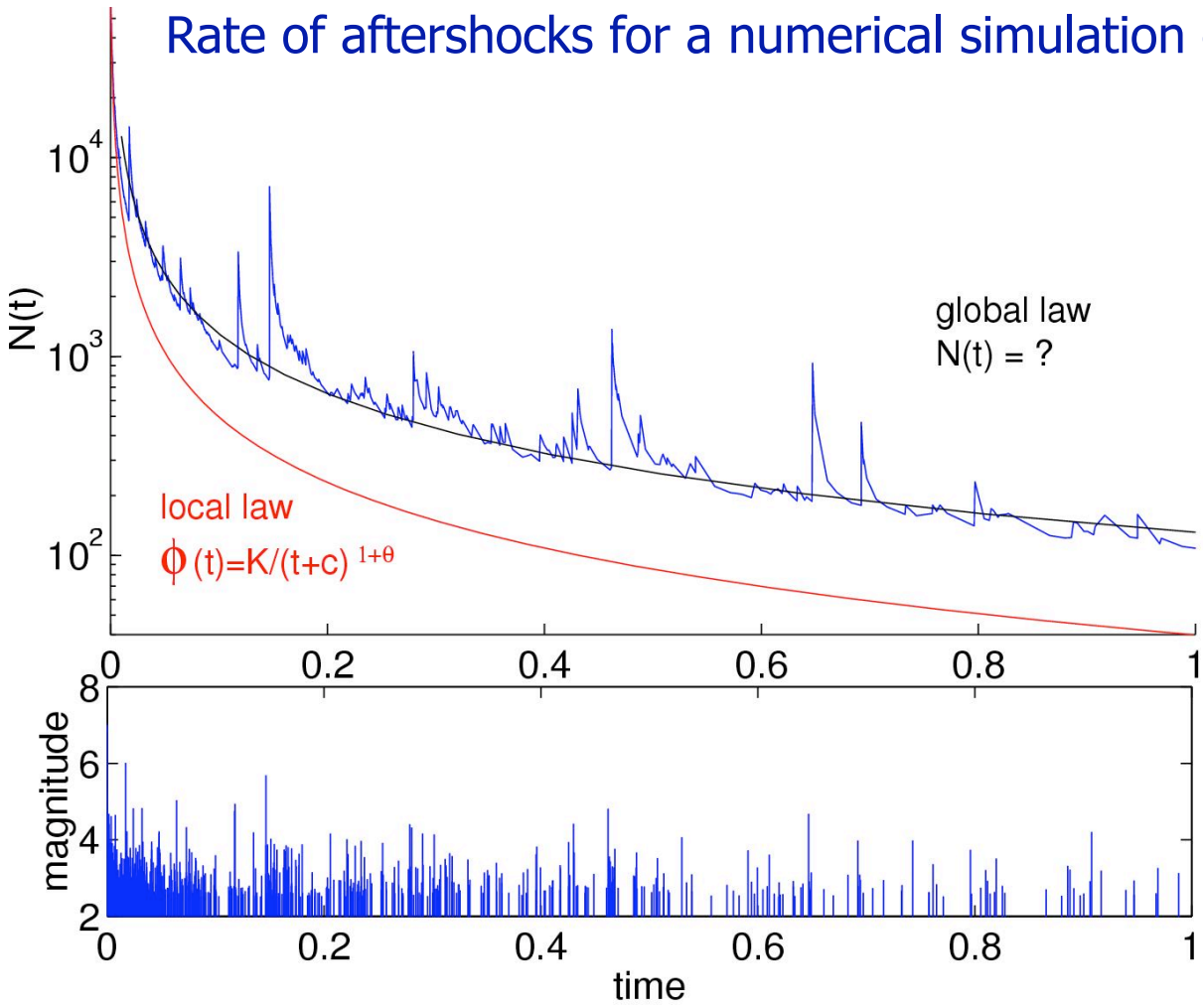
	Endogenous	Exogenous
Foreshock (or growth)	$S(t) \propto \frac{1}{\lvert t \rvert^{1-2\theta}}$	Abrupt peak
Aftershock (or decay)	$S(t) \propto \frac{1}{t^{1-2\theta}}$	$S(t) \propto \frac{1}{t^{1-\theta}}$

Non-critical: $S(t) \propto \frac{1}{t^{1+\theta}}$

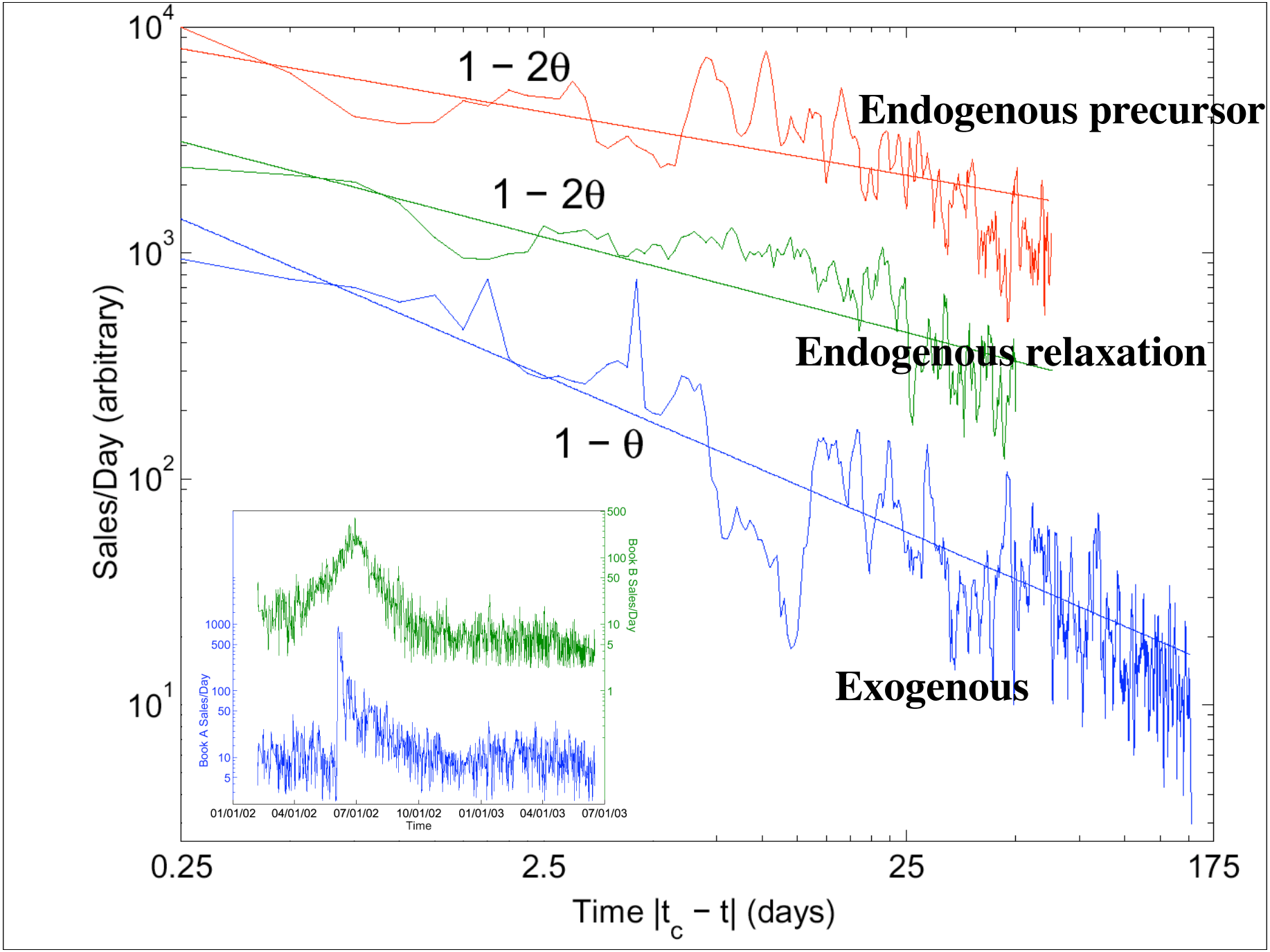
Hawkes ETAS model and numerical simulations

The impact of cascades of generations

Rate of aftershocks for a numerical simulation of the ETAS model



- is the global law also an Omori law ?
- p_{global} # p_{local} ?


Sales/Day (arbitrary)
Time |t_c − t| (days)
0.25
2.5
25
175
1 − 2θ
1 − 2θ
1 − θ
Endogenous precursor
Endogenous relaxation
Exogenous
Book A Sales/Day
Book B Sales/Day
Time
01/01/02
04/01/02
07/01/02
10/01/02
01/01/03
04/01/03
07/01/03

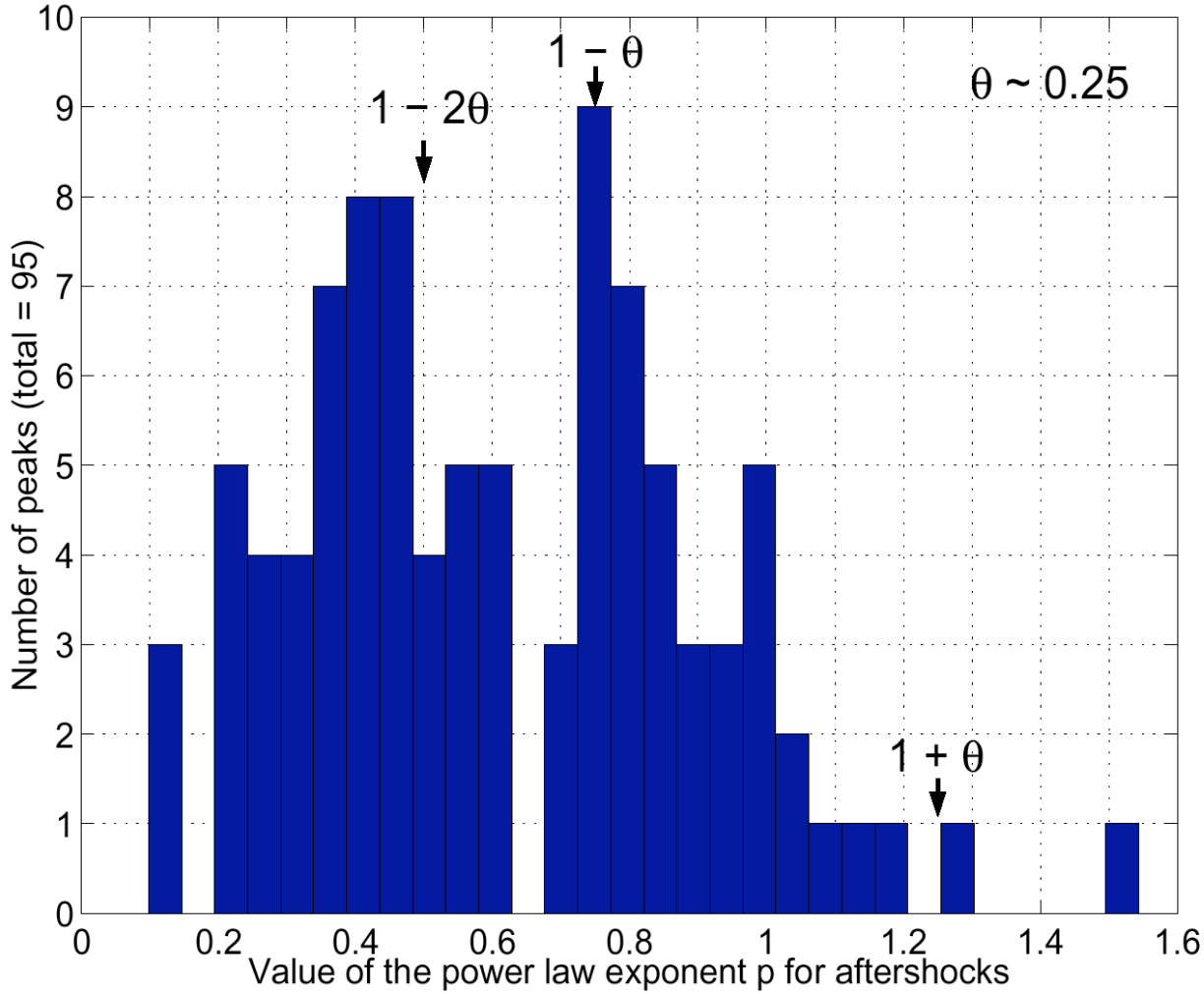
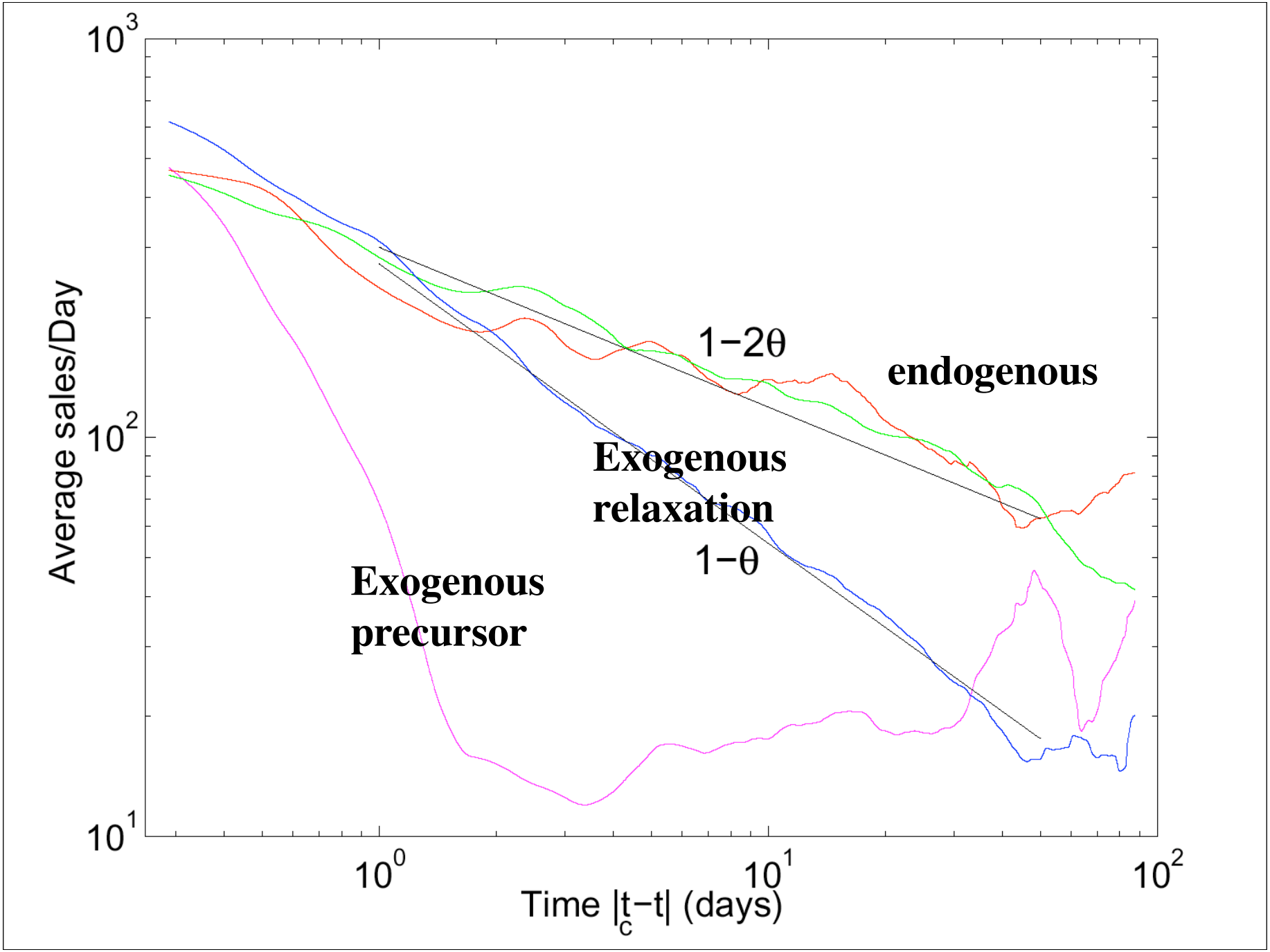


Figure 4: Histogram of the estimates of the power law exponents p of the relaxations of the sales of books following the largest peaks. The sample is obtained from 77 books, which yield 95 major peaks. One can clearly identify two classes of exponents: endogenous cluster for $1-2\theta$ close to 0.45 and exogenous cluster for $1-\theta$ close to 0.75, compatible with the estimation $\theta = 0.3 \pm 0.1$. The tail of the distribution of exponents extending up to at $1+\theta$ represents an exogenous crossover due to deviations from criticality (see text).


Average sales/Day
Time |t_c−t| (days)
1−2θ
endogenous
Exogenous
relaxation
1−θ
Exogenous
precursor

A Shocking Look At...

YouTube
Broadcast Yourself™

Videos

Browse
Most Recent
Most Viewed
Top Rated
Most Discussed
Top Favorites
Most Linked
Recently Featured

Riley Crane, Didier Sornette
ETH Zurich, D-MTEC
Chair of Entrepreneurial Risks

The Front Page

http://youtube.com/

Google

YouTube
Broadcast Yourself™

Sign Up | My Account | History | Help | Log In | Site:

Videos | Categories | Channels | Community | Upload

Search

powered by Google

Videos being watched right now...

01:20 | 00:24 | 01:30

Promoted Videos

- Dear Body.. — whataboutadam
- Got Halo 3? — UltraNeko
- Worcester Love Epis... — worcesterlove
- Woody's sneak peak ... — DatelineHollywood

Featured Videos

See More Featured Videos

Featured | Most Viewed | Most Discussed | Top Favorites

Final Words...Thank you YouTube!
This is it guys...
From: desiprince786
Views: 22,069
Time: 03:30
More in People & Blogs

The Cacaman - Ain't Got Much of ...
The Cacaman's Music Video for "Ain't got much of a smile." Original music and Lyrics by The Cacaman. Check out more at: http://(more)
From: CacamanMusic
Views: 143,521
Time: 03:42
More in Music

Rescue children living in povert...
I believe children, no matter where they are born, deserve to live free of danger, abuse and poverty. I Stir for a better world! (more)
From: WorldVisionStir
Views: 51,614
Time: 01:21
More in News & Politics

Ezekiel's Story
The Taylors shared their baby story, now go tell yours by visiting www.marchofdimes.com/everybaby Following the premature birth (more)
From: MarchofDimes
Views: 36,002
Time: 06:31
More in People & Blogs

Share Your Story at the 92nd Str...
Sharing Stories about the 92nd Street Y in New York City. To learn more about the 92nd Street Y, visit http://www.92Y.org
From: 92ndStreetY
Views: 18,122
Time: 03:08
More in People & Blogs

MOTORAZR2 ESSENCE OF SHARPER TRAILER
From: MOTORAZR2
Comments: 0

Login to rate video

Login | Sign Up | Help
Username:
Password:
Login
Login with your Google account

Broadcast Your Cause
The YouTube Nonprofit Program
Apply now!

What's New

YouTube Nonprofit Program
Broadcast your cause! Apply for the Nonprofit program and use video to get your organization's message out.

YouTube Mobile
Watch and upload YouTube videos on your mobile device.

Create Custom Players
Now you can choose the design and color of your player, and fill it with any of your playlists!

Broadcast Your Cause: The YouTube Nonprofit Program
Does your nonprofit organization have a great story to tell? Do you want to connect with supporters online but don't have the money for expensive outreach campaigns? Or are you just looking for way...

Read more in our Blog

Open Your Eyes - Sean Delaney
Sean Delaney is musician as well as a Direct Support Professional with the Special Needs Program in upstate New York. After spendi (more)
From: AutismSpeaksVids
Views: 52,531
Time: 03:48
More in Music

WOMEN OF DARFUR
IRC communications officer Emily Holland recently traveled to Darfur to document the lives of displaced Sudanese. In this video, (more)
From: theirc
Views: 84,468
Time: 01:22
More in News & Politics

Polar Bears
Two polar bears muse on the causes of global warming. An entry for our green film competition. http://www.foe.co.uk for info.
From: friendsoftheearth
Views: 235,760
Time: 01:00
More in Film & Animation

E-Waste: Dumping on the Poor ?? ?...
Asia Society's multimedia look at electronic waste shipped overseas and the toxic effect it has on places such as Guiyu, China--kn (more)
From: asiasociety
Views: 102,646
Time: 04:35
More in News & Politics

24 Hours for Darfur -- Numbers &...
This video montage highlights people's efforts to grapple with the scope and depth of violence and suffering wrought by the genoci (more)
From: 24hoursfordarfur
Views: 21,139
Time: 04:05
More in News & Politics

Kanye West: Education needs to b...
http://www.edin08.com - Join Kanye and ED in '08 ED in '08 is a campaign to make America's schools a top issue in the 2008 pres (more)
From: EDin08
Views: 241,230
Time: 00:29
More in News & Politics

Contest to End Youth Homelessness
A video explanation of the "Out Of The Shadows" Contest to end Youth Homelessness. Find out what you can do to address the problem (more)
From: YouthNoise
Views: 52,505
Time: 01:30
More in News & Politics

See More Featured Videos

Compare candidates on the issues!

Popular Videos for Mobile Devices

Wainy Days Ep#8 - "Plugged"
From: MyDamnChannel
Views: 1642

American Eskimo puppies
From: invisioning
Views: 7495

Search

Face The Candidates

Your Account: Videos, Favorites, Playlists, Inbox, Subscriptions, more...

Help & Info: Help Center, Video Toolbox, Developer APIs, Safety Tips, Copyright Notices, Code of Conduct

YouTube: Company Info, TestTube, Terms of Use, Privacy Policy, Press, Contact, Blog, Jobs

Overview

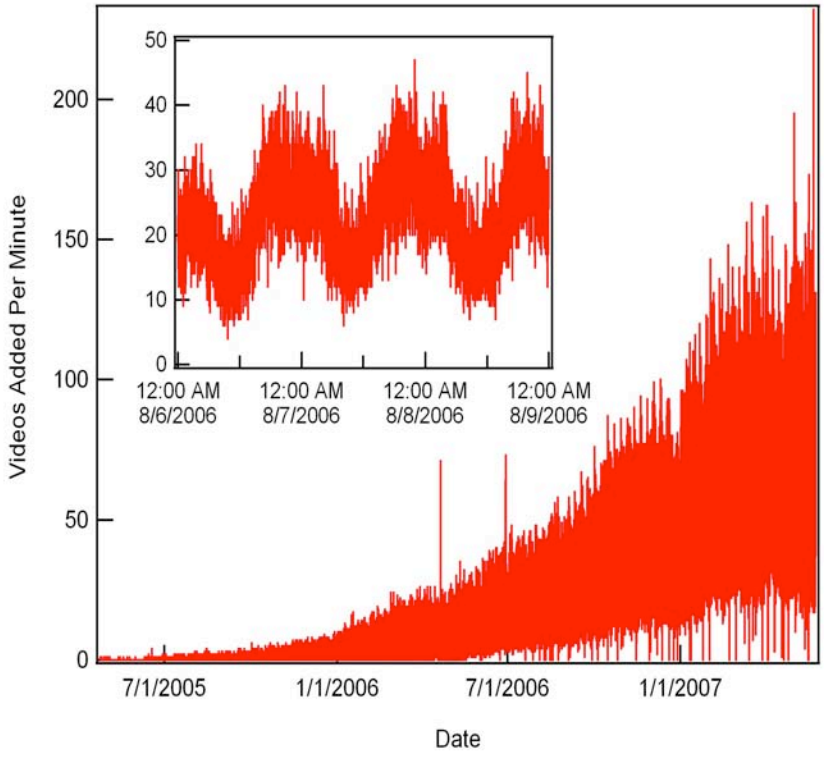
- Video Arrival and Site Growth
- Featuring – Endogenous/Exogenous Shocks
- Dynamical Relaxation Following Shocks

Perl script, via application programming interface (API) for the automated request of data. Stored in MySQL database

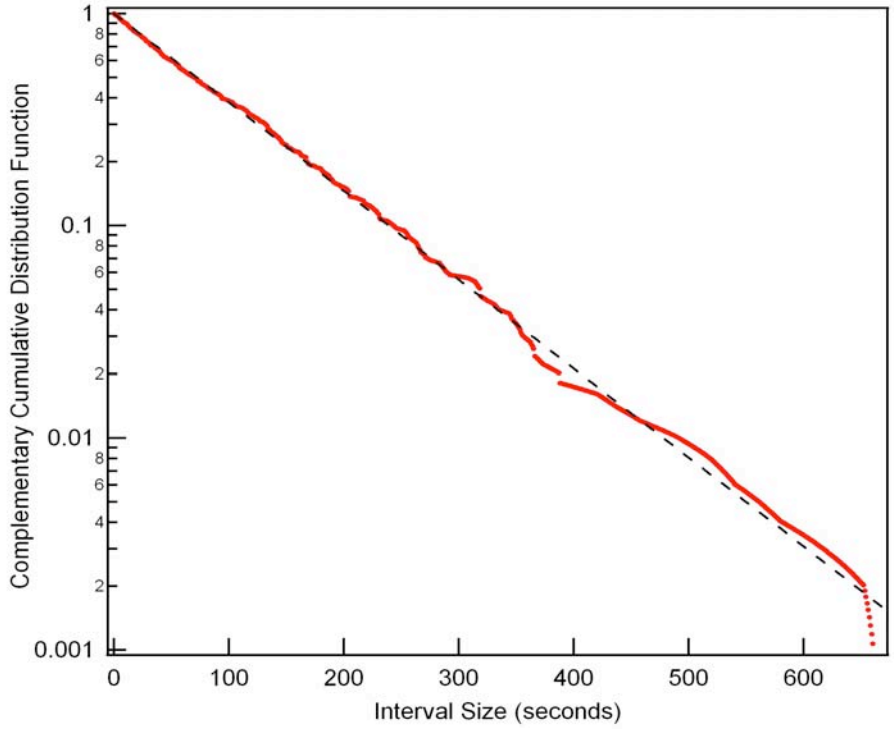
YouTube responds with a structured (XML) document containing information such as the cumulative number of times a video has been viewed (dynamic), along with descriptive information (static) concerning the user who posted the video, the title, tags, length, category, rating, comments, etc

Birth of a Video

Videos Added per Minute



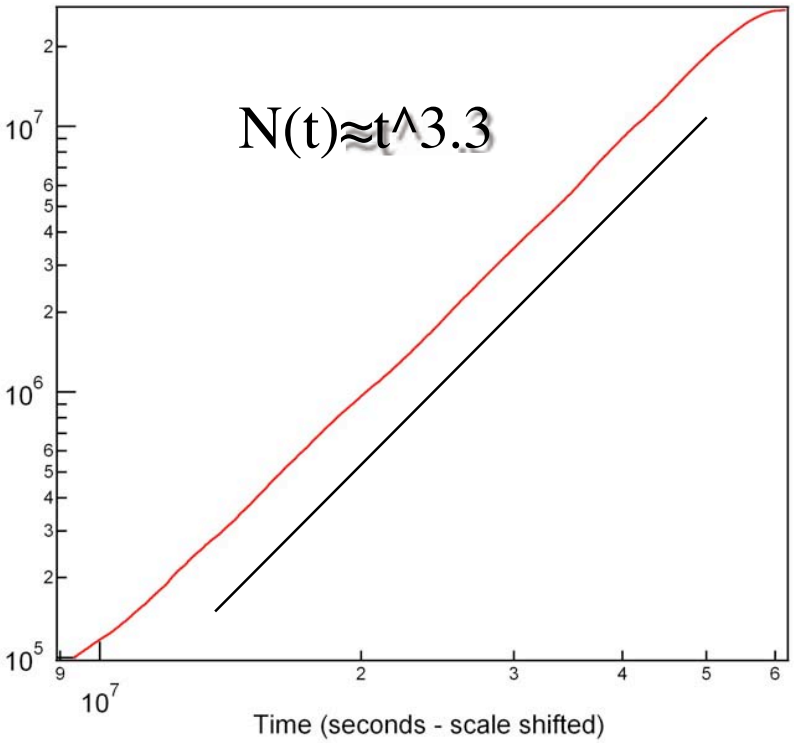
Locally - distribution of Waiting times is Exponential



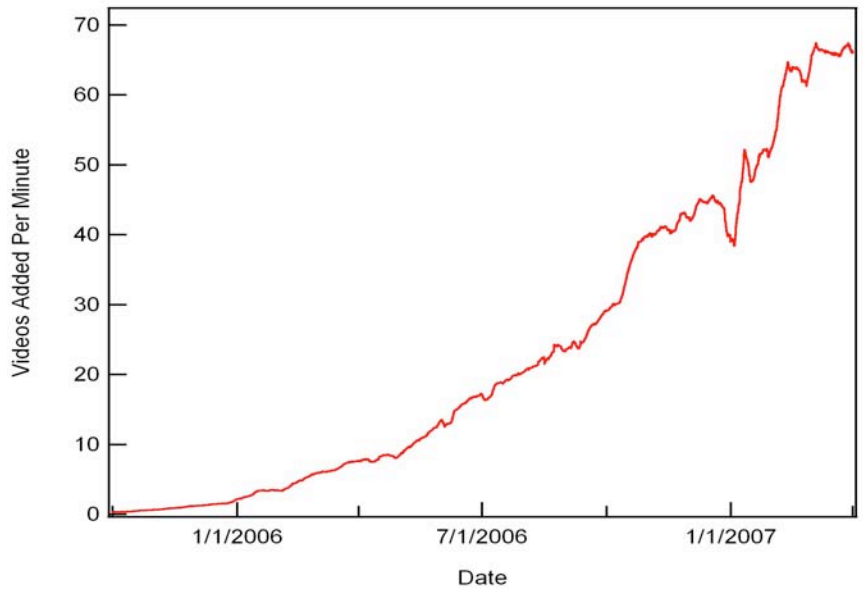
Non-Stationary Poisson Process

Exponent 3.3 reveals the average nb of "friends"

Cumulative Number of Videos



Non-Stationary Poisson Rate



Most Viewed Page

http://youtube.com/browse?s=mp

unix time

YouTube
Broadcast Yourself™

Sign Up | My Account | History | Help | Log In | Site:

Videos | Categories | Channels | Community | Upload

Search

powered by Google

Videos

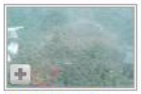
Most Viewed : (Today) : All

Browse
- Most Recent
- Most Viewed
- Top Rated
- Most Discussed
- Top Favorites
- Most Linked
- Recently Featured
- Most Responded
- Watch on Mobile

Time
- Today
- This Week
- This Month
- All Time

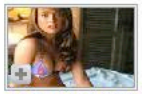
Category
- All
- Autos & Vehicles
- Comedy
- Entertainment
- Film & Animation
- Gadgets & Games
- Howto & DIY
- Music
- News & Politics
- People & Blogs
- Pets & Animals
- Sports
- Travel & Places

Justin Timberlake Beyonce- Until The End of Time [REMIX]!!
05:29
Added: 1 day ago
From: ParadoxModelMinority
Views: 532,685
679 ratings



RAF VIDEO 5
00:58
Added: 1 day ago
From: RAFFeed
Views: 405,876
122 ratings

||-- Tvr Tamora --||
00:49
Added: 18 hours ago
From: number1position
Views: 251,290
58 ratings

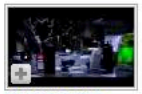


Headshots For Real
01:54
Added: 1 day ago
From: actingforreal
Views: 215,901
321 ratings

HOLLYWOOD.TV - Britney Spears Becomes Paparazzi in Burbank!
02:12
Added: 1 day ago
From: hollywoodtv
Views: 202,082
130 ratings

Mac Daddy
02:00
Added: 1 day ago
From: HappySlip
Views: 177,061
2259 ratings

Santana Lopes versus José Mourinho versus SIC Notícias
01:58
Added: 1 day ago
From: FBEmd
Views: 138,726
388 ratings



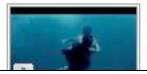
MOTORAZR2 ESSENCE OF SHARPER TRAILER
01:04
Added: 17 hours ago
From: MOTORAZR2
Views: 102,699

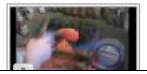
Diet.com Weight Loss Challenger II: Week 6 Tonya Vision
01:32
Added: 1 day ago
From: diethealth
Views: 95,165
34 ratings

Tutti in consiglio comunale
01:04
Added: 1 day ago
From: StaffGrillo
Views: 94,991
192 ratings

Barcelona 4 - 1 Real Zaragoza - La Liga Highlights
03:24
Added: 1 day ago
From: footyleak
Views: 87,708
88 ratings

????????!!???nakedNEWS????
01:31
Added: 1 day ago
From: hohten
Views: 85,463
52 ratings





Most Recent Page

http://youtube.com/browse?s=mr&t=t&c=0&l=

unix time

Sign Up | My Account | History | Help | Log In | Site:

Broadcast Yourself™

Videos | Categories | Channels | Community | Upload

Search

powered by Google

Videos

Most Recent : All

Browse

- Most Recent
- Most Viewed
- Top Rated
- Most Discussed
- Top Favorites
- Most Linked
- Recently Featured
- Most Responded
- Watch on Mobile

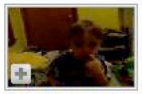
Category

- All
- Autos & Vehicles
- Comedy
- Entertainment
- Film & Animation
- Gadgets & Games
- Howto & DIY
- Music
- News & Politics
- People & Blogs
- Pets & Animals
- Sports
- Travel & Places

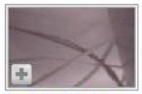
Unwetter Slowenien
01:24
Added: 25 minutes ago
From: willem1946
Views: 1

Interview du créateur de Heroes
08:19
Added: 25 minutes ago
From: Dewey63
Views: 0

un tanar batut de politie
00:23
Added: 25 minutes ago
From: mcmadrid
Views: 0

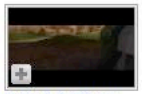


Joey Bean Walking 3
00:35
Added: 25 minutes ago
From: happyheatherly
Views: 0



beautiful lady
03:01
Added: 25 minutes ago
From: shyqal34
Views: 1

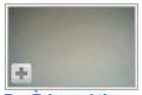
thyra
01:03
Added: 25 minutes ago
From: ruffBITCH
Views: 2



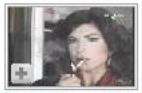
Halo 2 Artillery Bombardment
00:36
Added: 25 minutes ago
From: bondboy8
Views: 0

Se7en on STAR real story I AM [Arirang TV 4-7-2007] Part 5
05:00
Added: 25 minutes ago
From: AkaKelvin
Views: 0

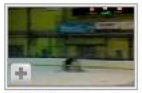
"When you walk by" - Live Song @ wedding (HEBREW)
05:15
Added: 25 minutes ago
From: eshedesh
Views: 0



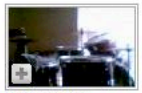
Re: Eric and the Army of the Phoenix (1/5)
09:59
Added: 25 minutes ago
From: ximubayu
Views: 0



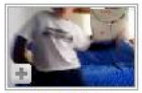
Ida Di Benedetto Light up cigarette
00:14
Added: 25 minutes ago
From: MDL72
Views: 0



hczvole
03:11
Added: 25 minutes ago
From: Hodnak
Views: 0

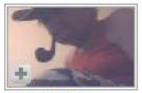


jealous guy
09:59
Added: 25 minutes ago
From: bigbreadeaterellis
Views: 0



im a sucker
03:12
Added: 25 minutes ago
From: alexcityhero
Views: 1

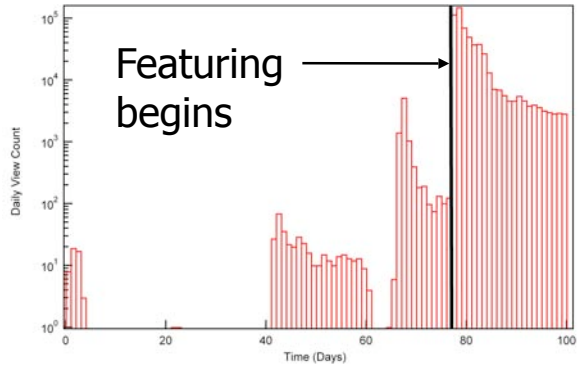
9-6 Friends Forever 2004-07
08:16
Added: 25 minutes ago
From: shbita92
Views: 0



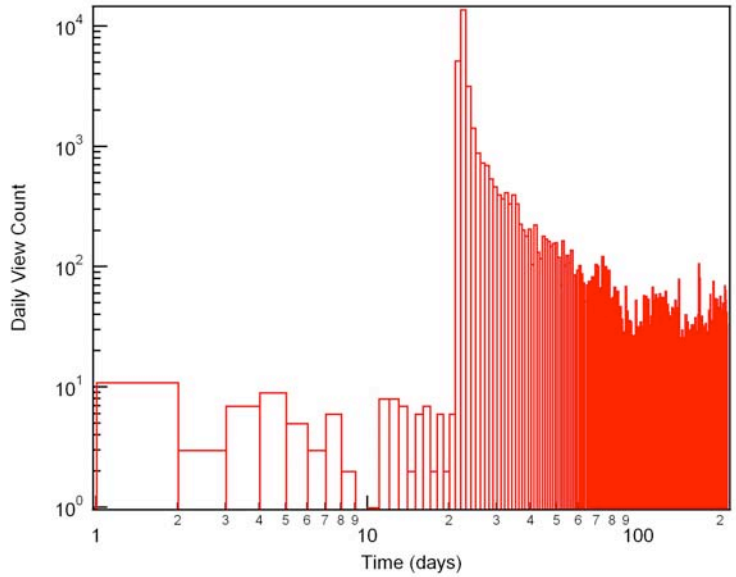
CONTINUED, SMOKING:MAGNUM PIPE TOBACCOS.
03:13
Added: 25 minutes ago
From: MARIANUSPIPESTOBACCO
Views: 0

The Effect of Featuring

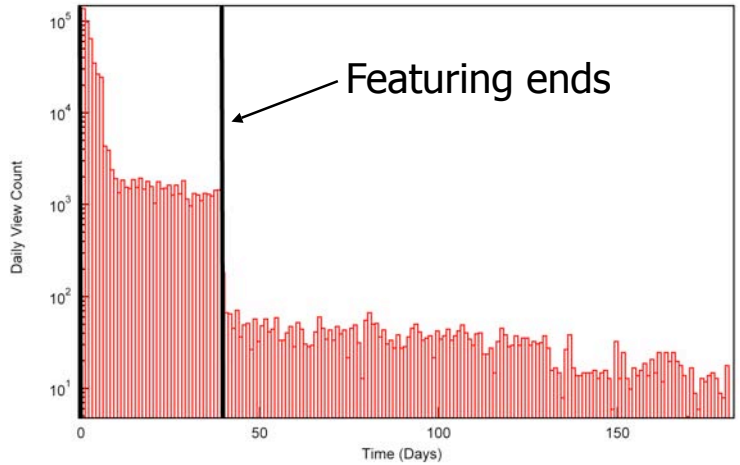
Growth of a Video before being featured



Editorial Featuring
(arbitrary and random)

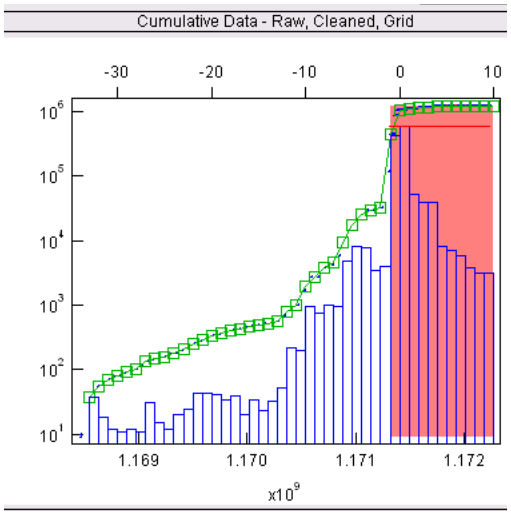


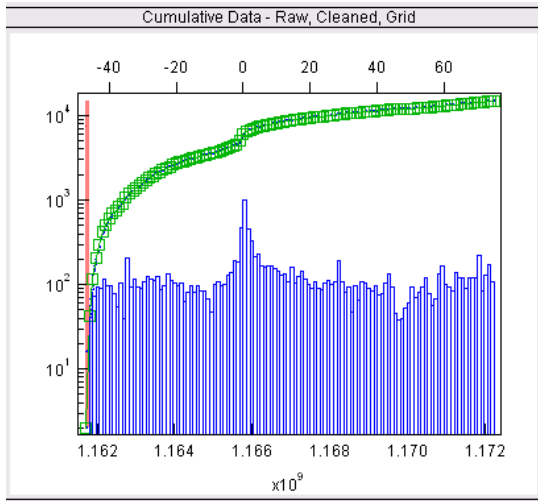
Decline of a video after being featured

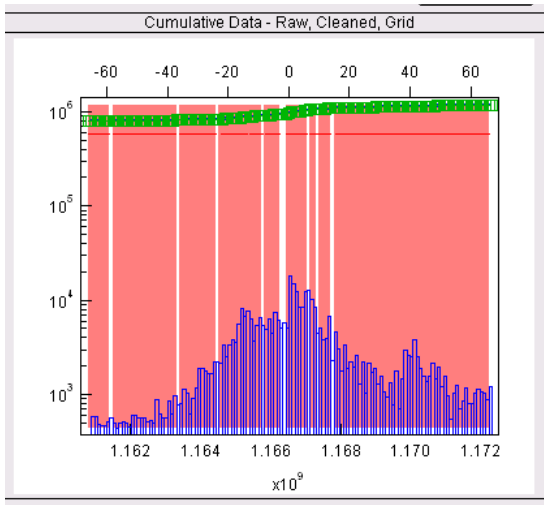


Shocks in YouTube

“Endogenous”

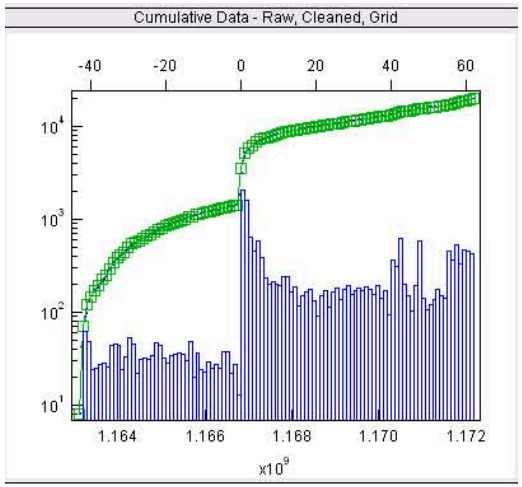


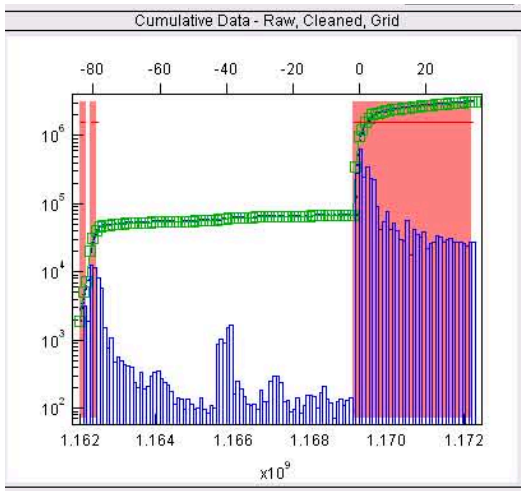


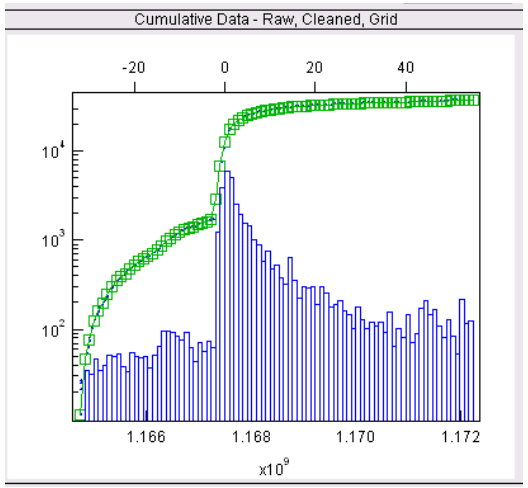


Shocks in YouTube

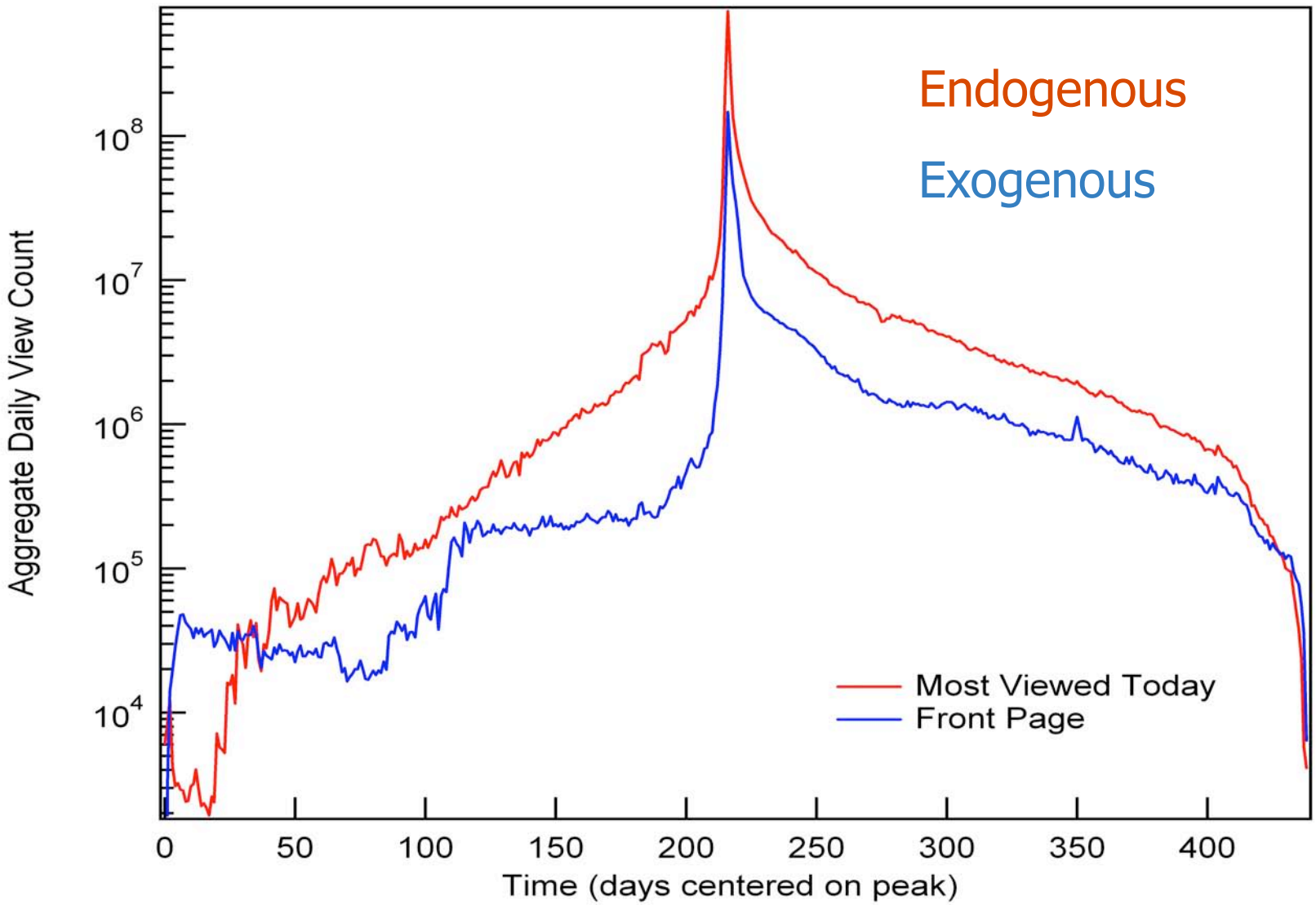
“Exogenous”





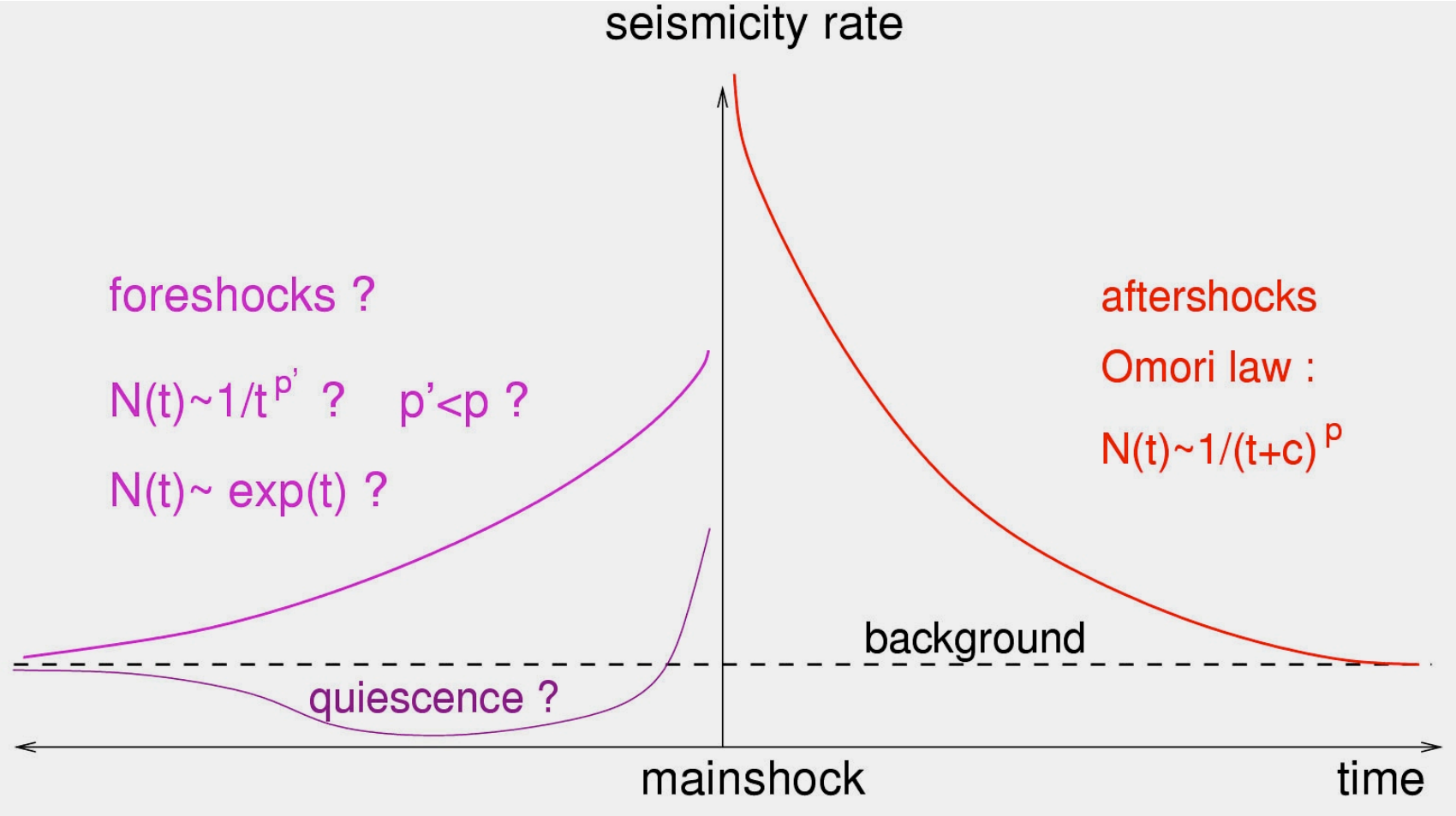


Non-Parametric Superposition

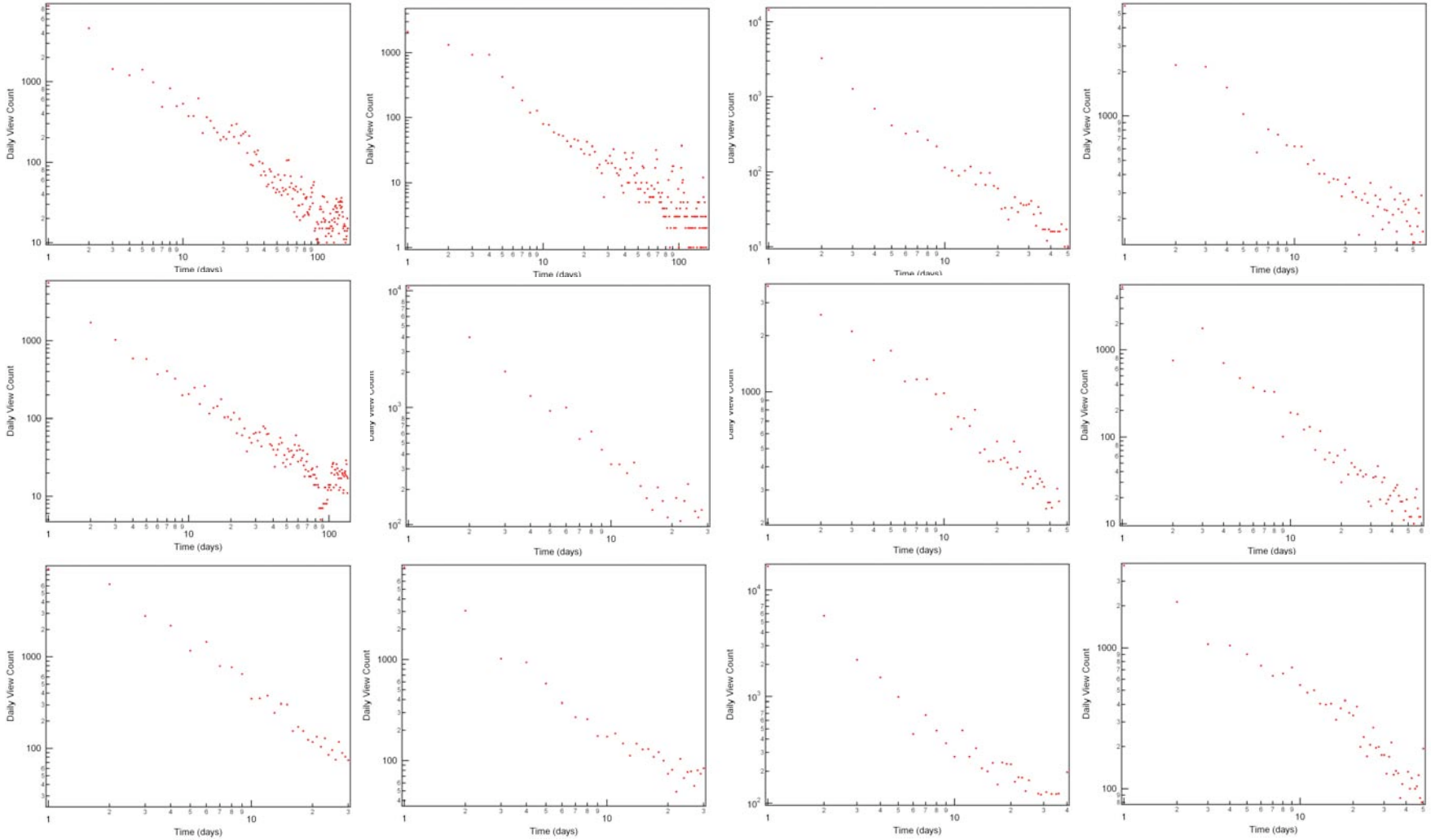


Temporal variation of seismicity

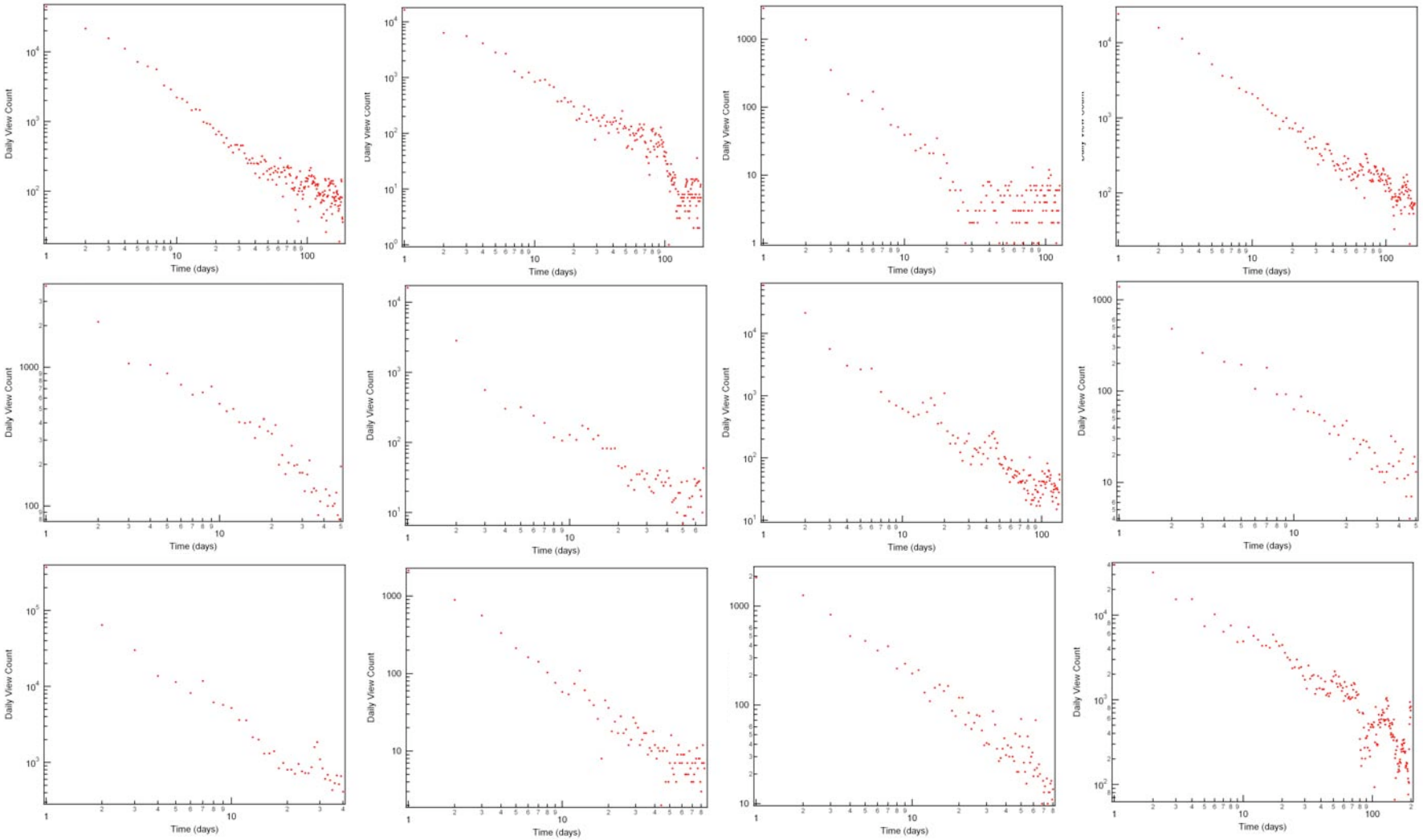
Observations :



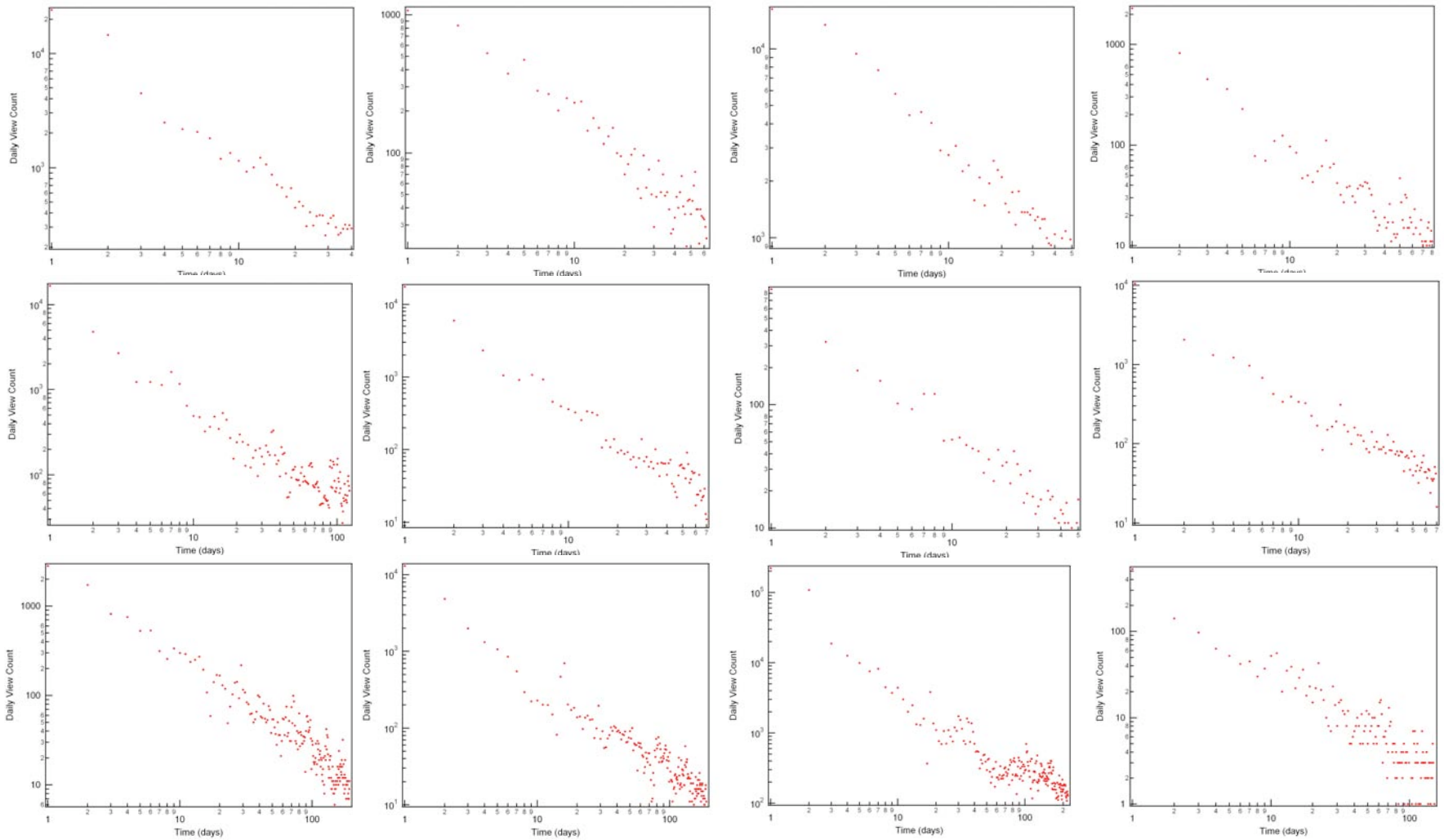
Typical Relaxation Following Peak



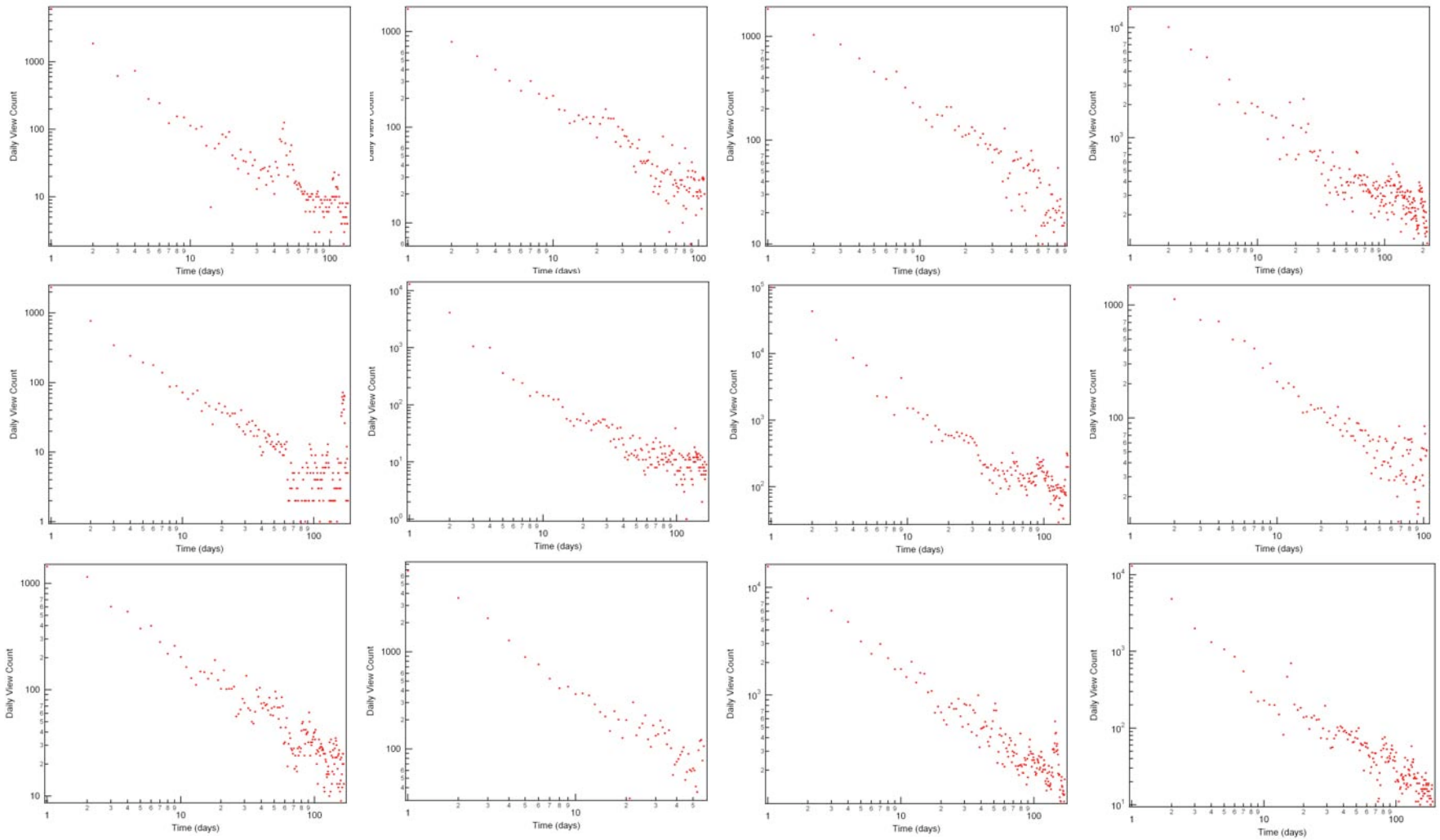
Typical Relaxation Following Peak



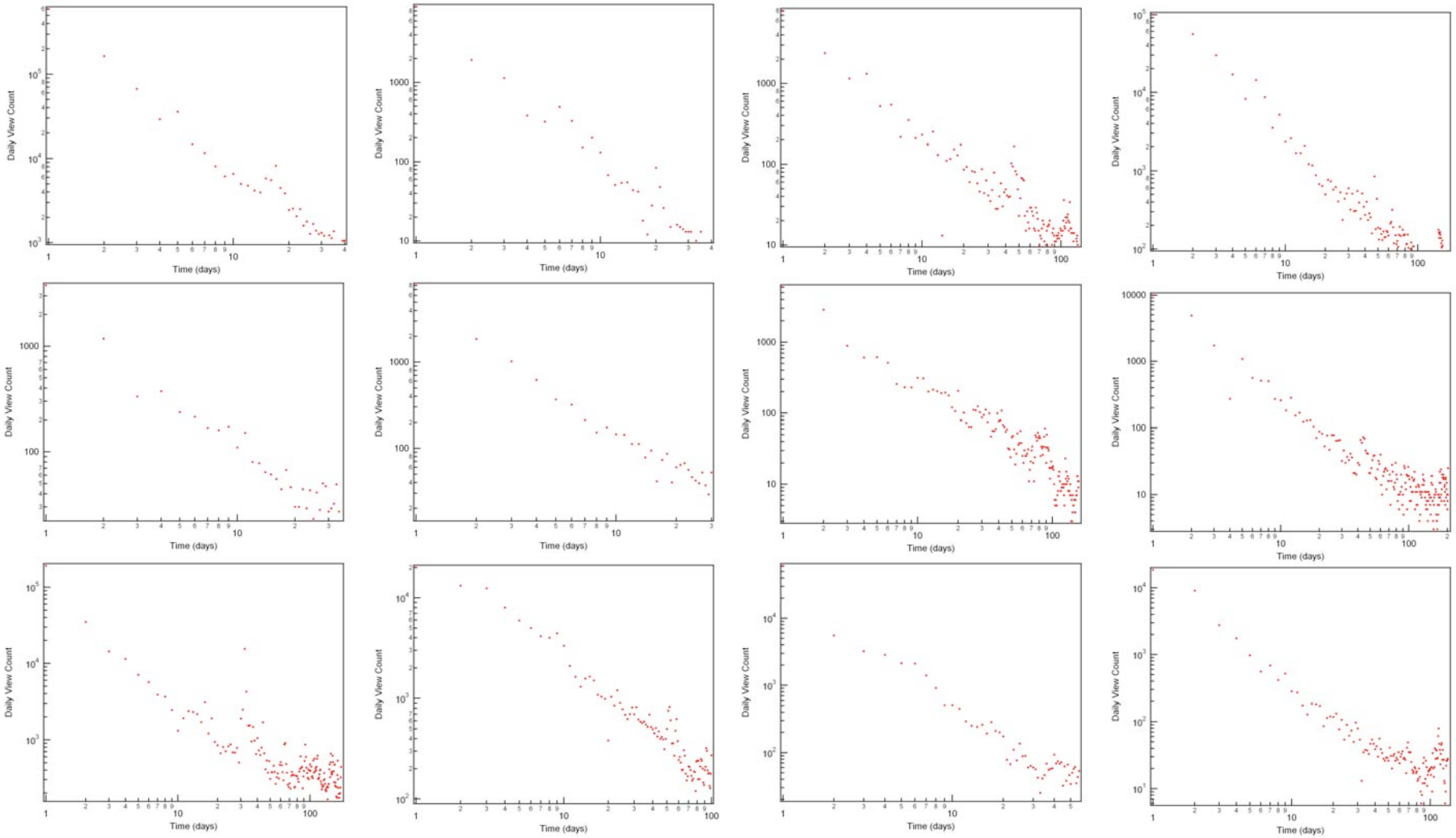
Typical Relaxation Following Peak



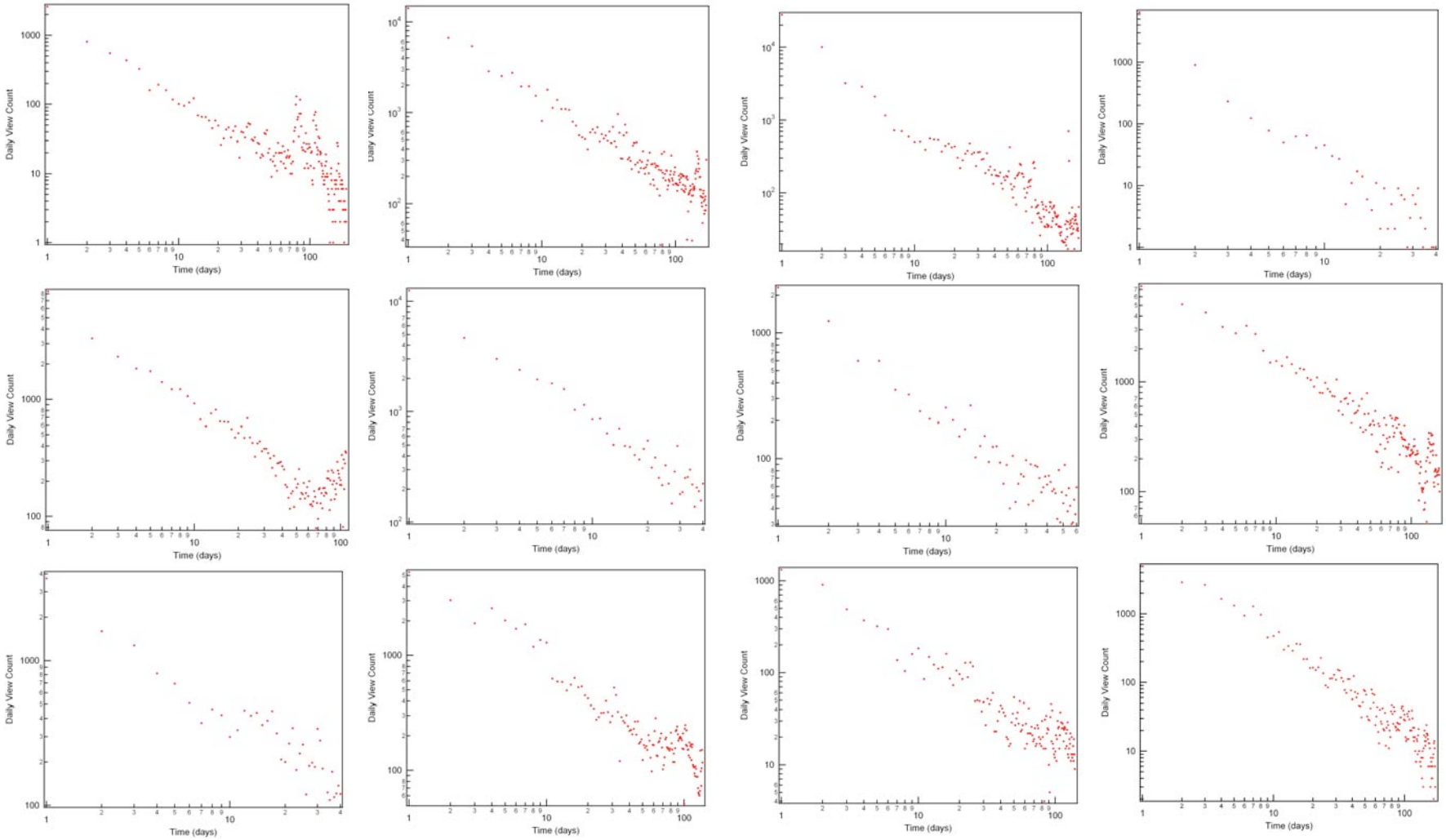
Typical Relaxation Following Peak



Typical Relaxation Following Peak

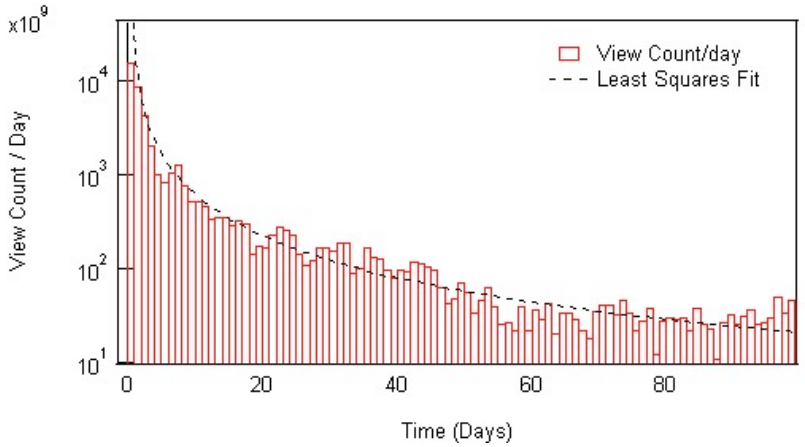


Typical Relaxation Following Peak



Typical Response

Shock: more than 100 views on a single day, and has at least 10 days following this peak. Of the 5 million videos we are tracking, 76% do not receive 100 views on any given day. Furthermore, 15% either don't have 10 days worth of data, or don't have 10 days following a qualified peak. This leaves us with roughly 9% (=421,487 videos).

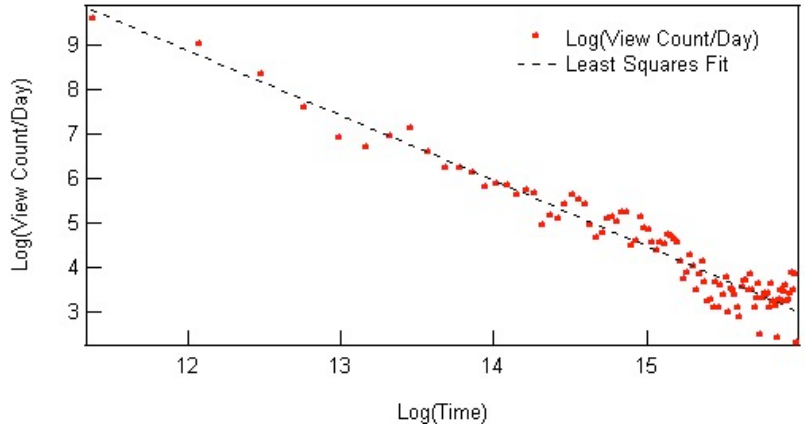


View Count per Day = $A(t - t_c)^p$

A = Amplitude

p = exponent governing decay

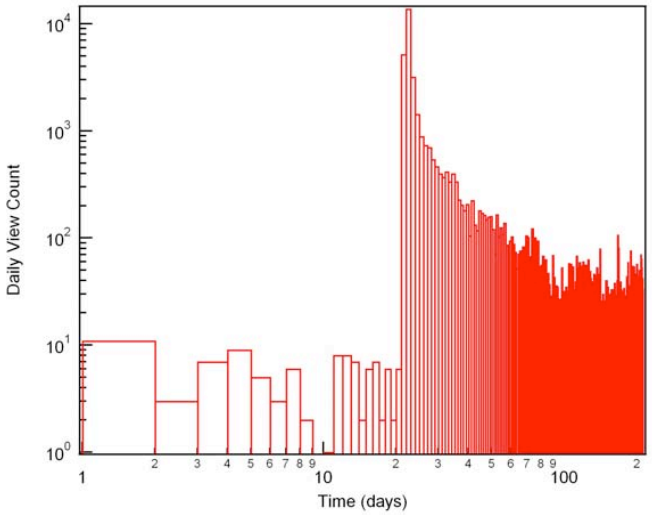
- A Least-Squares Fit is performed on the log-log data over the largest possible range.
- The exponent "p" is extracted



Sorting Out the data: Peak Height Fraction

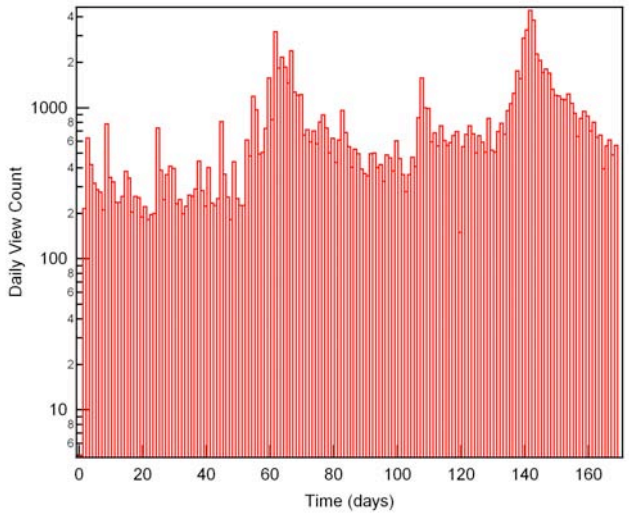
"Not Critical/Exo"

Fraction ~ 40%

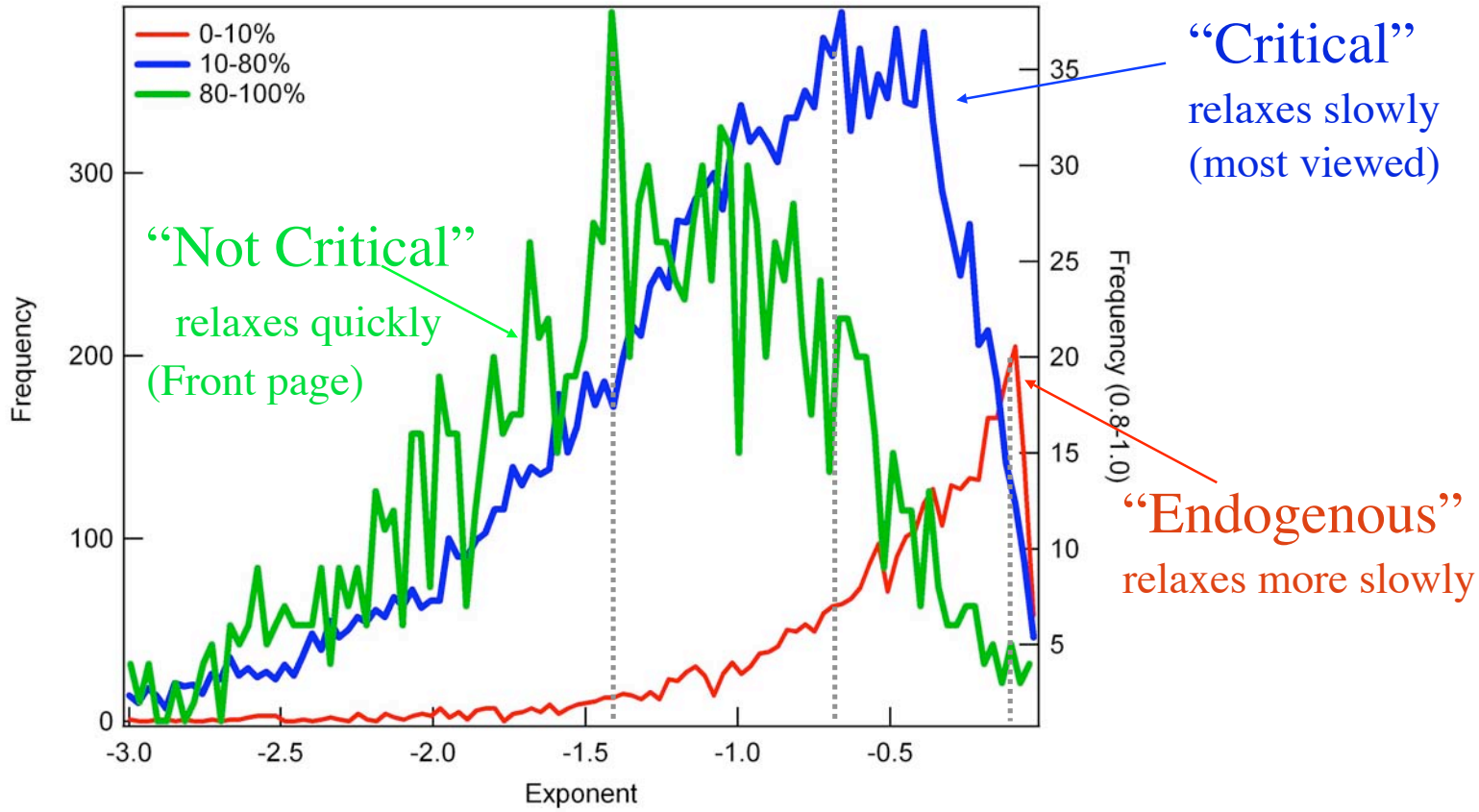


"Critical/Endo"

Fraction ~ 1%



Exponent – Shock as Fraction of Total Peak Height



	Endogenous	Exogenous
Foreshock (or growth)	$S(t) \propto \frac{1}{\lvert t \rvert^{1-2\theta}}$	Abrupt peak
Aftershock (or decay)	$S(t) \propto \frac{1}{t^{1-2\theta}}$	$S(t) \propto \frac{1}{t^{1-\theta}}$

Non-critical: $S(t) \propto \frac{1}{t^{1+\theta}}$

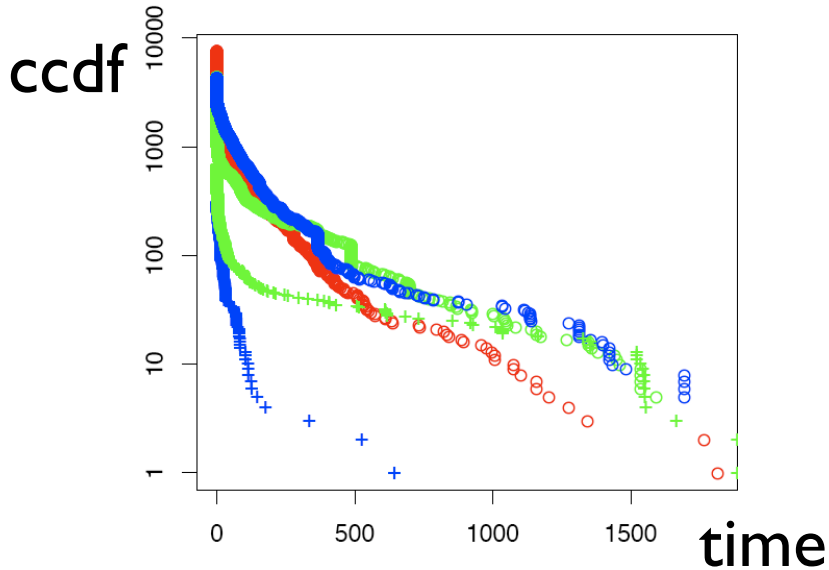
Software vulnerability dynamics

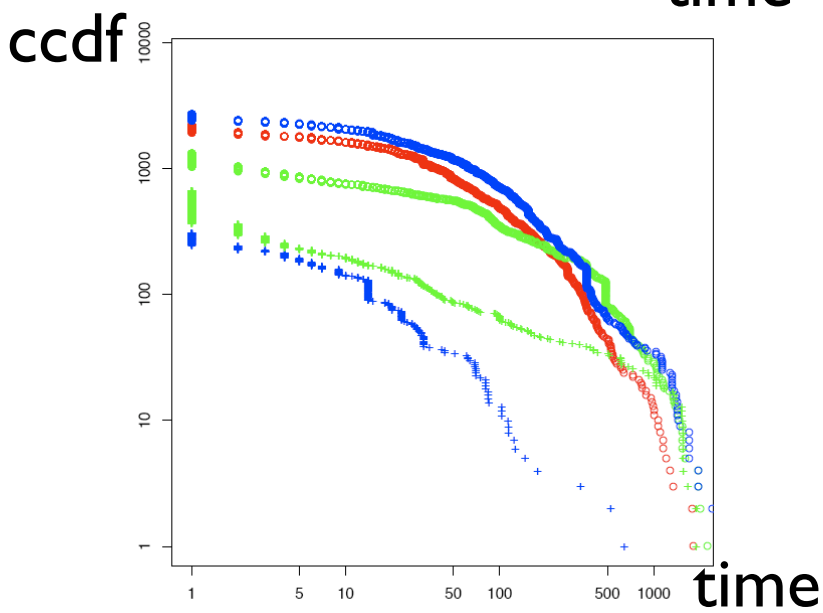
with S. Frei (ETH Zurich)

- vulnerability process is a good proxy of software resilience to bugs
- we identify 4 steps in vulnerability process:

  1. discovery (red)
  2. exploit (green)
  3. public disclosure (time reference)
  4. patch release (blue)

- exploits and patch can appear before disclosure (crosses) or after (circles)
- once again, response distribution in this process is heavily tailed
- very characteristic is the distribution of exploits (before disclosure, green crosses) which shows some patterns of power-law with phase transition, in lower tail





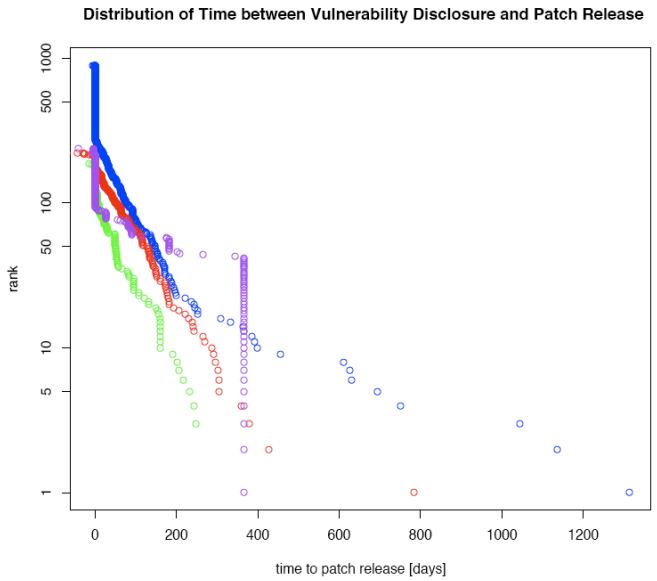
Software vulnerability dynamics

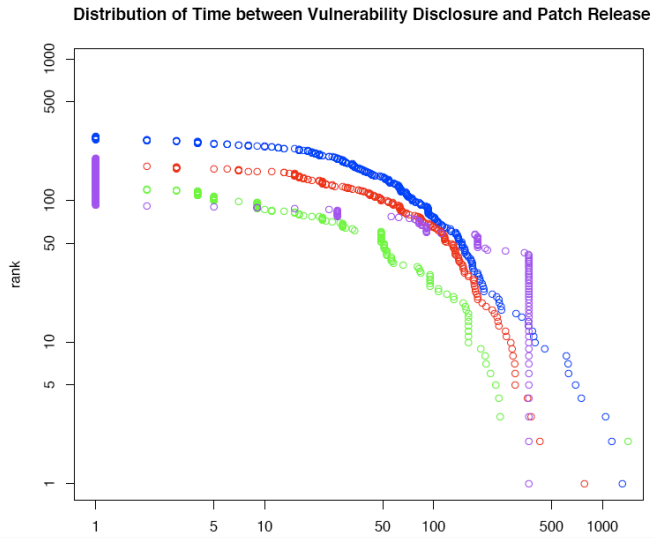
- Here we show comparison between types of softwares:

  - Microsoft (blue)
  - Linux (red)
  - Oracle (purple)
  - Mozilla (green)

- We can see that time to patch distribution is also heavily tailed.

- While it varies differently according to considered software the allure remains somehow the similar, especially when we consider Microsoft (blue) and Linux (red).





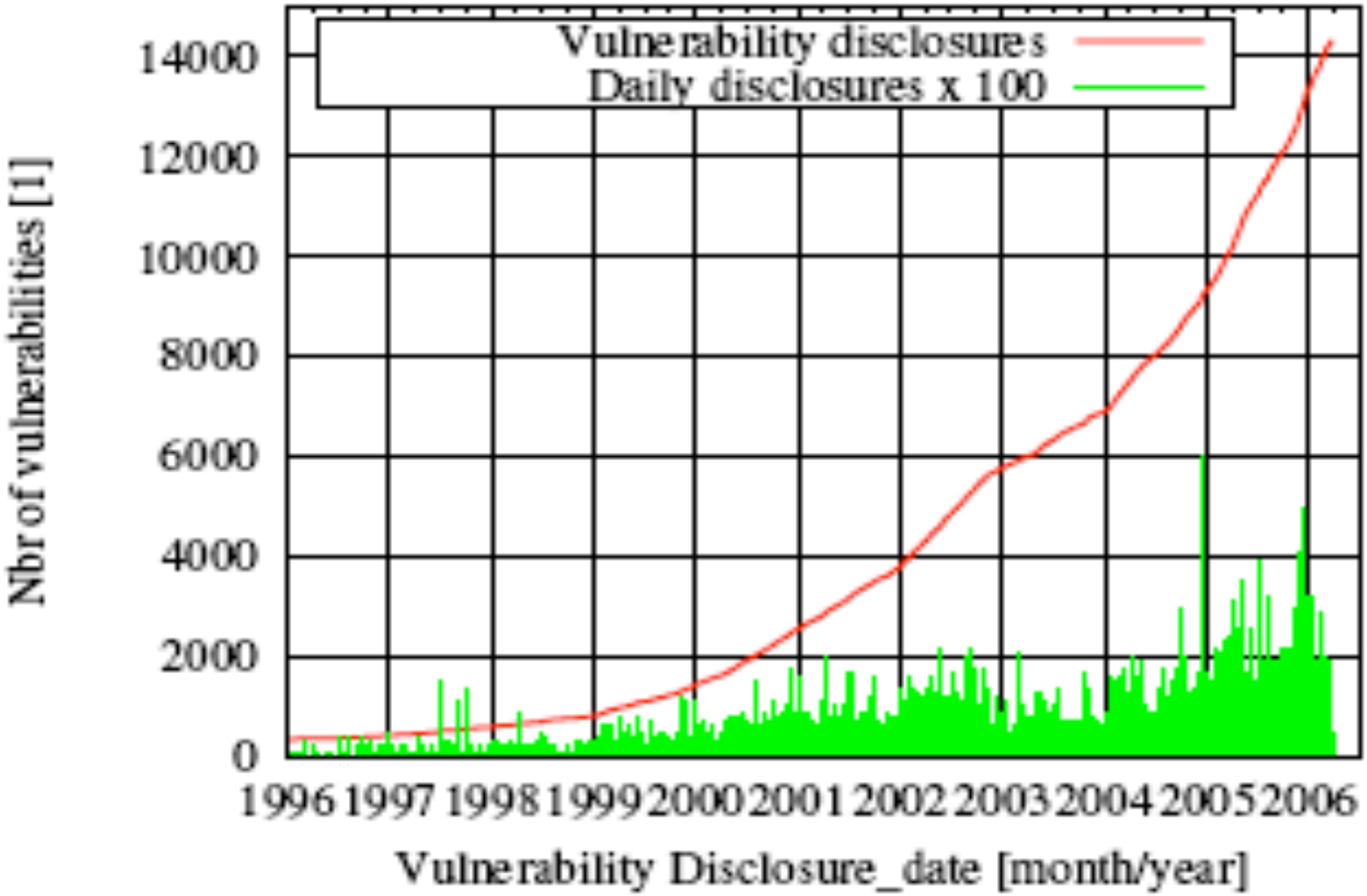
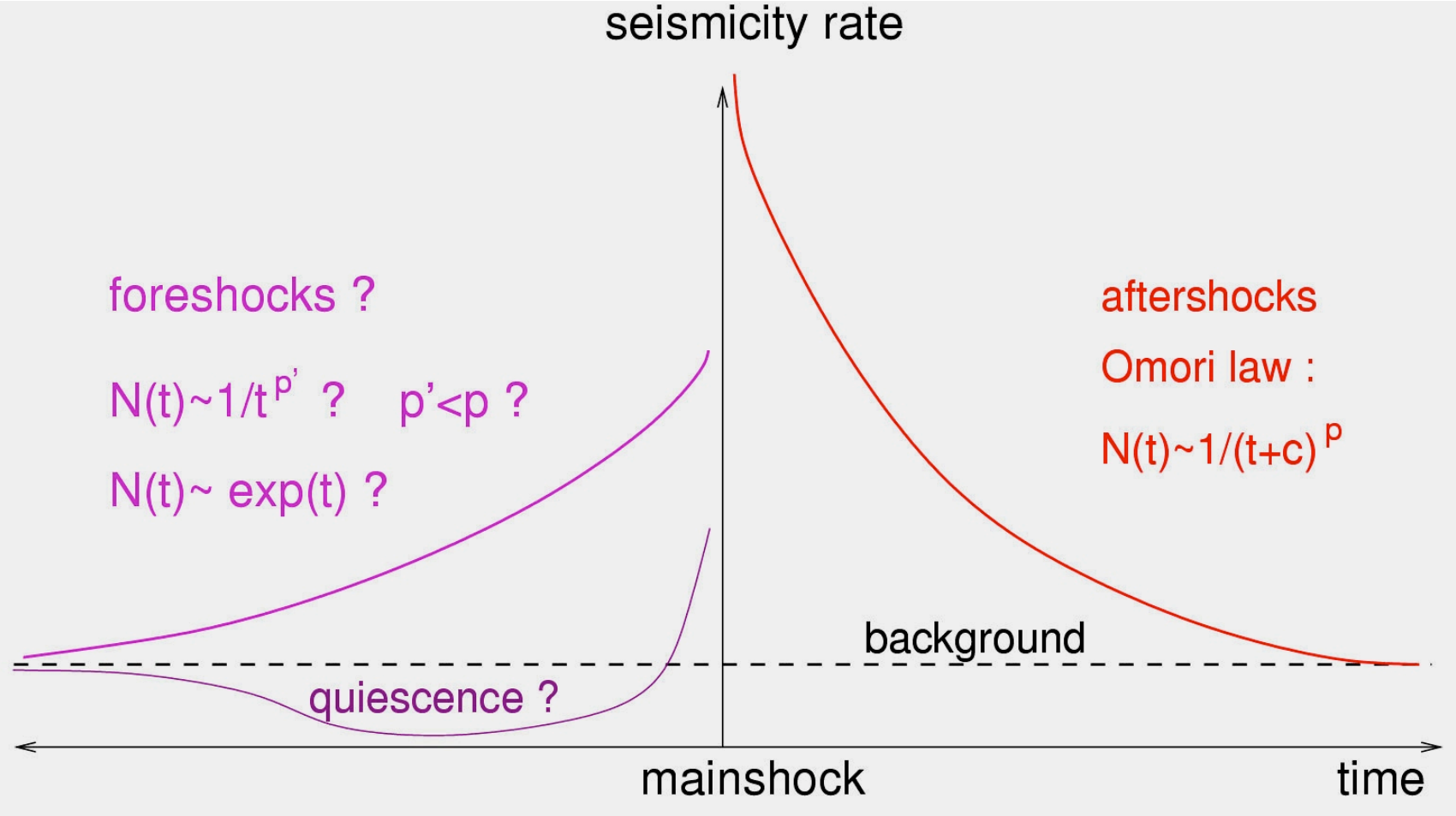


Figure 1: Cumulated number and daily rates of disclosed vulnerabilities between 1996 and 2006

(Frei et al. 2006)

Towards Predictions

Observations :



Predicting the rise and fall of social and economic interactions by monitoring and modeling internet activities and commercial sales

- Books, Music, DVD,
- Electronics (audio and video, cameras and photography, software, computers and video games, cell phones. . .)
- Office
- Children and Babies
- Home and Garden (which includes pets)
- Gifts, Registries, Jewellery and Watches
- Apparel and Accessories
- Food
- Health, Personal Care, Beauty
- Sports and Outdoors
- Services (movies, restaurants, travel, cars, . . .)
- Arts and Hobbies
- Friends and Favourites

Endogenous versus Exogenous

Extinctions
- -meteorite at the Cretaceous/Tertiary KT boundary
- -volcanic eruptions (Deccan traps)
- -self-organized critical events

Discoveries
- -serendipity
- -maturation

Volatility bursts in financial time series
- -external shock
- -cumulative effect of "small" news

Financial crashes
- -external shock
- -self-organized instability

Earthquakes
- -tectonic driving
- -triggering

Immune system
- -external viral or bacterial attack
- - " internal" (dis-)organization

Parturition
- -mother/foetus triggered?
- -mother-foetus complex?

Brain (learning)
- -external inputs
- -internal self-organization and reinforcements (role of sleep)

Aviation industry recession
- -September 11, 2001
- -structural endogenous problems

Commercial success and sales
- -Ads
- -epidemic network

Recovery after wars?
- -internally generated (civil wars)
- -externally generated

Social unrests
- -triggering factors
- -rotting of social tissue

Motivation, Effort, Production in Open Source Software

Open Source Software (OSS) Contribution in Mozilla

- data mining in Concurrent Versioning System (CVS)
- analyze software (Mozilla Project) since its start (1998)
- focus on developers, debuggers, contribution

Open Source Software (OSS) Network

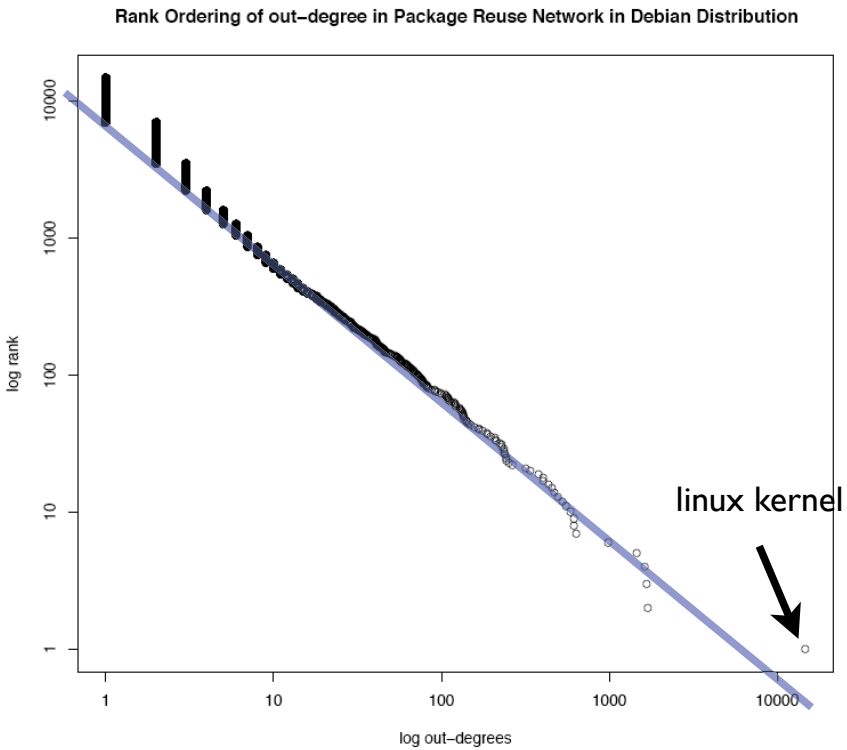
(with Thomas Maillart)

- One Key feature of OSS is the capability to reuse pieces of source code wherever they are useful or needed.
- Programs call other Programs, allowing development time savings, and long range updates
- We can define Links between programs (nodes) as Edges of an OSS Network. These Links are directed.
- We study the connectivity of nodes (out degrees) distribution among a particular OSS subset: packages included in Debian Linux Distribution

Open Source Software (OSS) Network

- exponent = 1
  Zipf Law
- on four orders of magnitude
- this is the typical pattern of a scale free network

Debian software

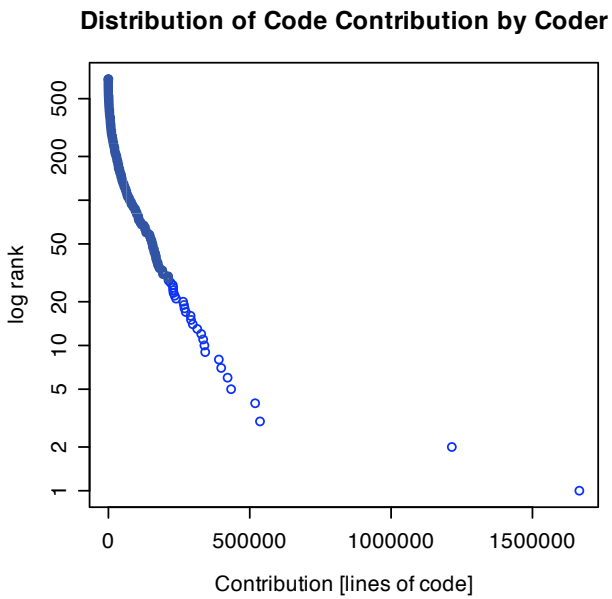


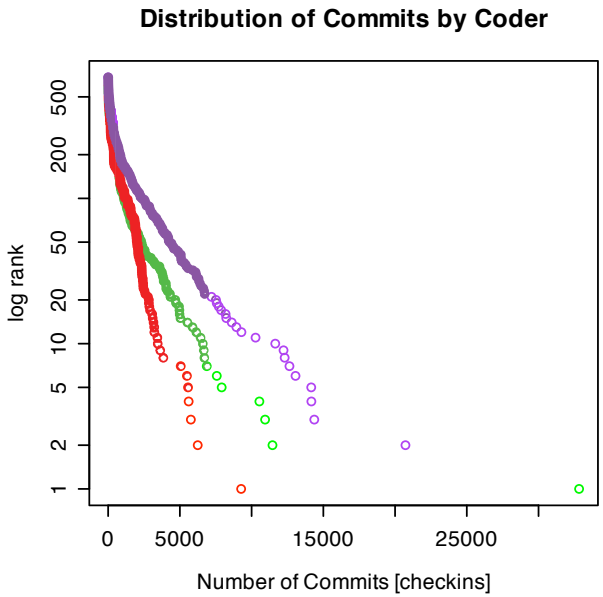
Open Source Software (OSS) Contribution in Mozilla

- Since no formal organisation occurs in OSS, we determine the activity of source code committers in time series
- We differentiate by developers (adding features) and debugger (adding robustness)
- We can clearly see that clusters of activity occur in time and space (source code tree)

Open Source Software (OSS) Contribution in Mozilla

- distribution of contributions (both development and debugging) is heavy tailed
- we analyze distribution of deviations between development and debugging per committer to find coder's profiles.
- Many are developing as much as debugging
- But some develop far much more than debugging
- This distribution is also heavy tailed

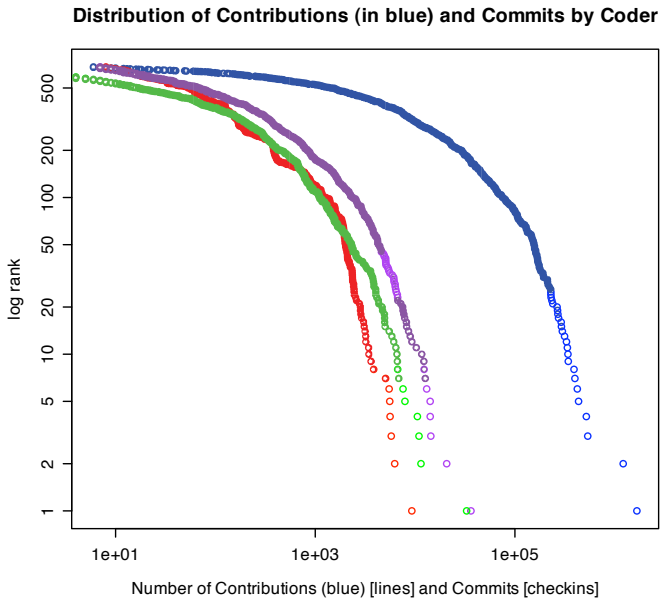




- The graphs show 4 distributions:

1. checkins in the source code repository (purple)
2. contribution in lines of code (blue)
3. bugs treated (red)
4. difference between checking and bugs (green)

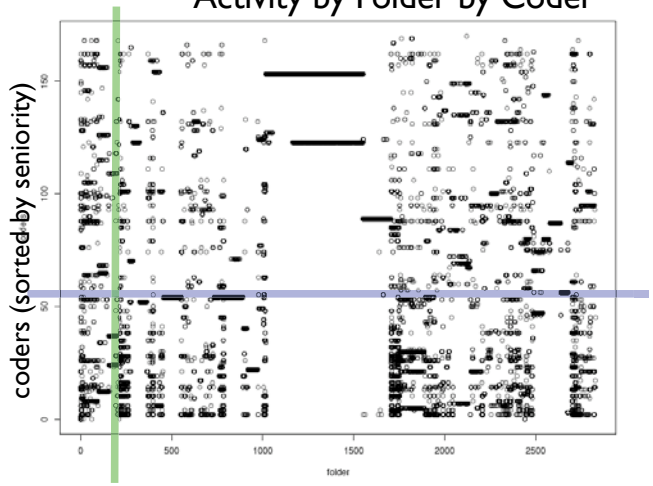
- We can see that all distributions are heavy tailed, denoting a wide dispersion in coders contributions as well as in their specialization (developers or debuggers)
- From the green graph, we assert that developers mainly also debug (lower tail) while have a tendency to only develop (upper tail)



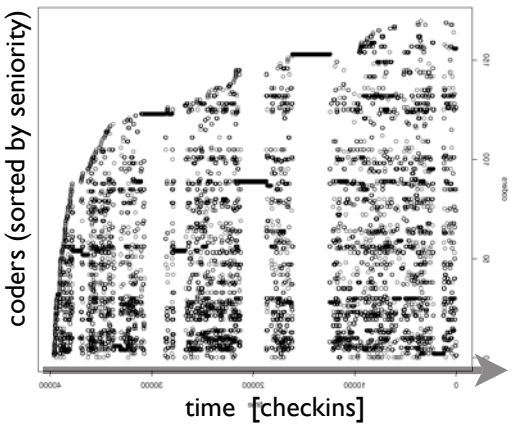
Open Source Software (OSS) Contribution in Mozilla - Activity Maps -

- Clusters of activity appear by visual inspection
- Coders tend to work in localized space (source code tree) and time
- Open question: are there cascading effects, source code development ? Intuitevely Yes!
- What are the sources of these cascading effects? exogenous? endogenous? How does source code development process evolves in the life of OSS project?

Activity by Folder by Coder



Activity by Coder vs Time



Activity by Folder vs Time