Homogeneous temporal activity patterns in a large online communication space

A. Kaltenbrunner[1,2], V. Gómez[1,2], A. Moghnieh[1], R. Meza[1,2], J. Blat[1,2] and V. López[1,2]

[1]Department of Information and Communication Technologies
Universitat Pompeu Fabra

[2]Barcelona Media Centre d'Innovació

Workshop on Social Aspects of the Web (SAW 2007)

in conjunction with

10th International Conference on Business Information Systems (BIS 2007)

Outline

Outline

Motivation

Time patterns of human activity:

- Human behavior is expected to be heterogeneous
- Are there some underlying common patterns (e.g. in the inter-event or waiting times)?
- Several studies report heavy tailed distributions
- What form have such homogeneous patterns?

Controversy: Power-law or log-normal?

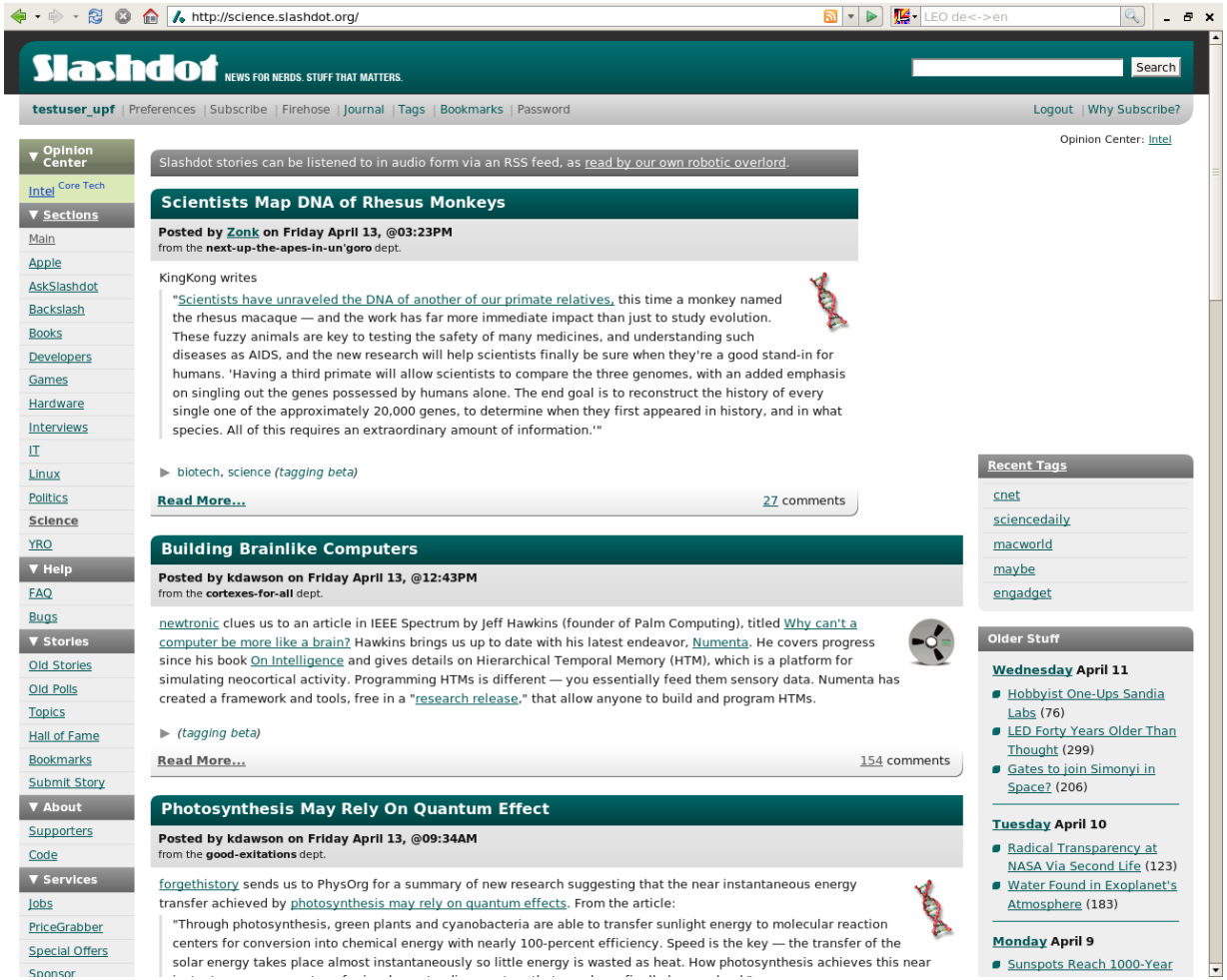
- e.g. Waiting time in e-mail (one-to-one) communication
- Barabasi, 2005 vs. Stouffer et al. 2006

And many-to-many communication?

We analyze online debates on Slashdot (`www.slashdot.org`)

Outline

Slashdot: Start Page



Facts about Slashdot:

- tech-news website
- created 1997
- allows **comments** to short news **posts**
- during $\approx$ 14 days
- moderation system ensures quality of communication

Example of a post with comments

Building Brainlike Computers

Posted by kdawson on Friday April 13, @12:43PM
from the **cortexes-for-all** dept.

newtronic clues us to an article in IEEE Spectrum by Jeff Hawkins (founder of Palm Computing), titled Why can't a computer be more like a brain? Hawkins brings us up to date with his latest endeavor, Numenta. He covers progress since his book On Intelligence and gives details on Hierarchical Temporal Memory (HTM), which is a platform for simulating neocortical activity. Programming HTMs is different — you essentially feed them sensory data. Numenta has created a framework and tools, free in a "research release," that allow anyone to build and program HTMs.

post

Sad news...John Corzine dead at 60 (Score:-1, Troll)

by Anonymous Coward on Friday April 13, @12:50PM (#18720205)

I just heard some sad new on Talk Radio, John Corzine corrupt billionaire governor of the Socialist state of New Jersey was found dead in his New Jersey automobile. There were no further details. Even if you were not a fan of this criminal sleazebag, there is no denying his contribution to the politics of waste and fraud. Truly an American Icon!

[Reply to This]

I built a brainlike computer, but it wasn't useful (Score:0, Funny)

by Anonymous Coward on Friday April 13, @12:51PM (#18720233)

It spent most of the time watching TV all stoned. Too many receptors.

[Reply to This]

this is stupid (Score:0, Troll)

by corynthian_dude (1087973) on Friday April 13, @12:52PM (#18720245)

to believe that man could create a brain is absurd. Only god could create a brain. Computer programmers seem to have a delusion that they can make something in the image of gods creation. You can make games, wordprocessors, email programs, no problem, those are easy, but to believe you can make something capable of understanding the world in all its god given glory is heresy.

[Reply to This]

End of civilization (Score:4, Funny)

by Anonymous Coward on Friday April 13, @12:52PM (#18720251)

Because it would signal the end of civilization...if computers can look like women (porn), feel like women (Realdolls), and think like women (have a brain, at least in some cases), then all procreation would cease and humans would suffer the same fate as the dinosaurs.

[Reply to This]

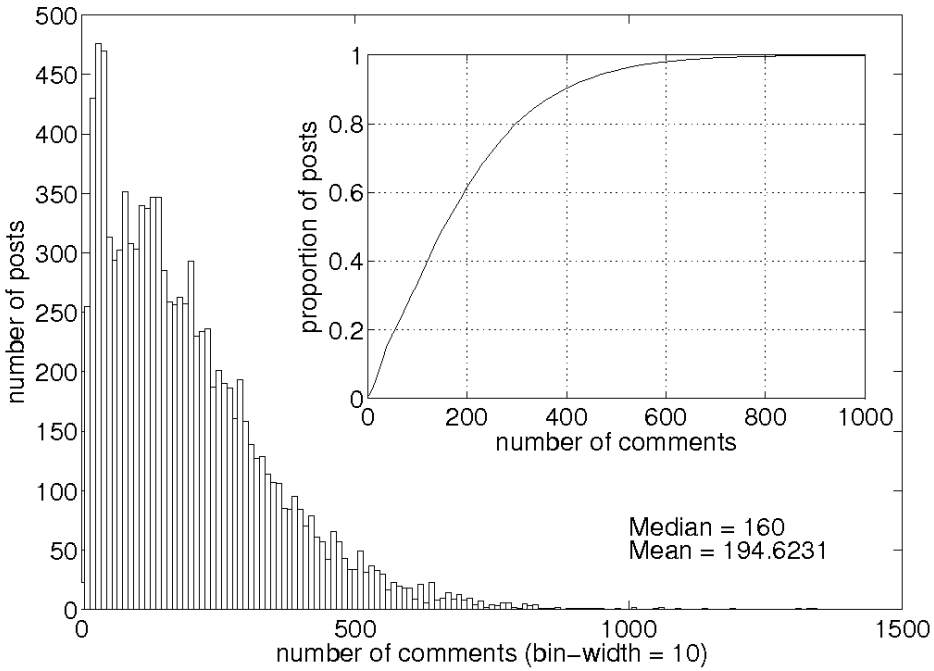
comments

Outline

Data collected

Period covered:

- We collected the data of one year of debates on Slashdot
- Time-span: August 26th 2005 – August 31st 2006.



Data contains:

- $\approx 10^4$ Posts
- $\approx 2 \cdot 10^6$ Comments
- $\approx 10^5$ Commentators
- 18.6% Anonym. comments

← number of com. per post

Outline

Post Comment Interval (PCI)

Building Brainlike Computers

Posted by kdawson on Friday April 13, @12:43PM
from the **cortexes-for-all** dept.

newtronic clues us to an article in IEEE Spectrum by Jeff Hawkins (founder of Palm Computing), titled Why can't a computer be more like a brain? Hawkins brings us up to date with his latest endeavor, Numenta. He covers progress since his book On Intelligence and gives details on Hierarchical Temporal Memory (HTM), which is a platform for simulating neocortical activity. Programming HTMs is different — you essentially feed them sensory data. Numenta has created a framework and tools, free in a "research release," that allow anyone to build and program HTMs.

Sad news...John Corzine dead at 60 (Score:-1, Troll)

by Anonymous Coward on Friday April 13, @12:50PM (#18720205)

I just heard some sad new on Talk Radio, John Corzine corrupt billionaire governor of the Socialist state of New Jersey was found dead in his New Jersey automobile. There were no further details. Even if you were not a fan of this criminal sleazebag, there is no denying his contribution to the politics of waste and fraud. Truly an American Icon!

[Reply to This]

I built a brainlike computer, but it wasn't useful (Score:0, Funny)

by Anonymous Coward on Friday April 13, @12:51PM (#18720233)

It spent most of the time watching TV all stoned. Too many receptors.

[Reply to This]

this is stupid (Score:0, Troll)

by coryntHian_dude (1087973) on Friday April 13, @12:52PM (#18720245)

to believe that man could create a brain is absurd. Only god could create a brain. Computer programmers seem to have a delusion that they can make something in the image of gods creation. You can make games, wordprocessors, email programs, no problem, those are easy, but to believe you can make something capable of understanding the world in all its god given glory is heresy.

[Reply to This]

End of civilization (Score:4, Funny)

by Anonymous Coward on Friday April 13, @12:52PM (#18720251)

Because it would signal the end of civilization...if computers can look like women (porn), feel like women (Realdolls), and think like women (have a brain, at least in some cases), then all procreation would cease and humans would suffer the same fate as the dinosaurs.

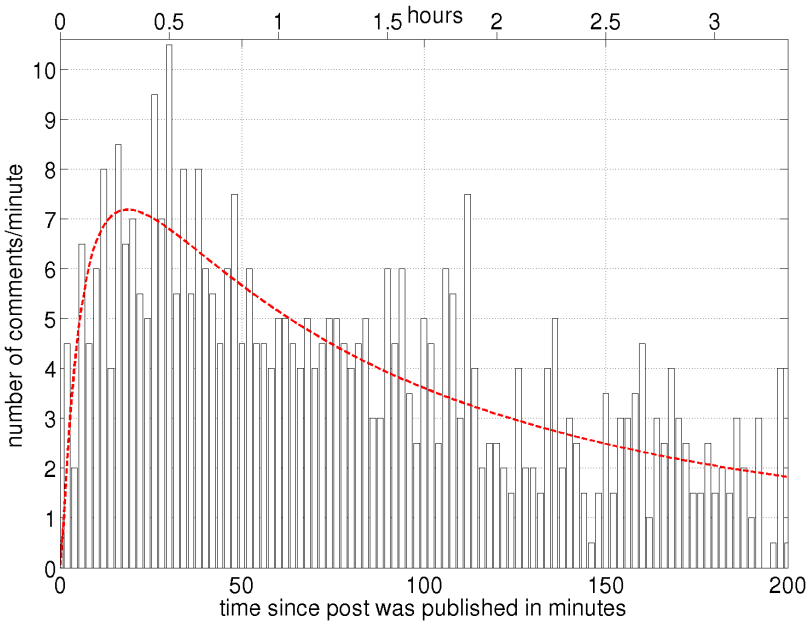
[Reply to This]

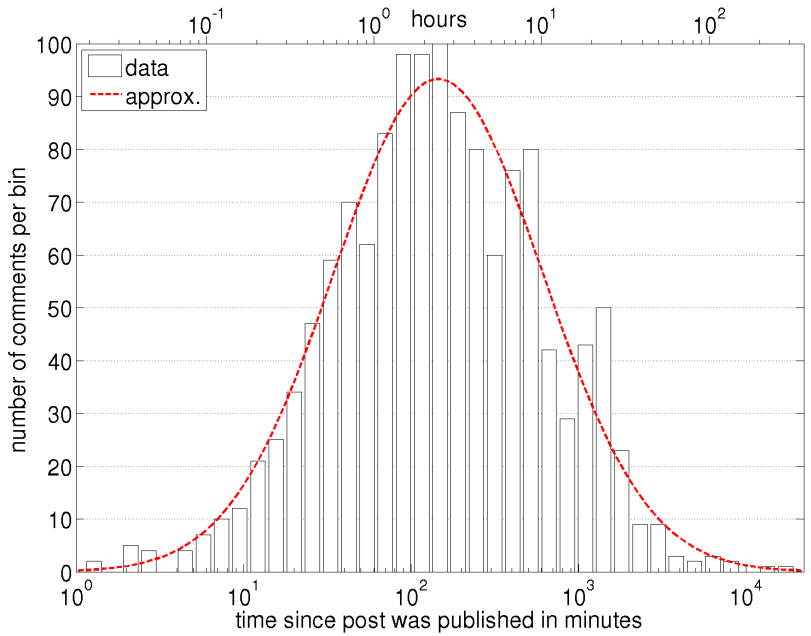
Time difference between Post and Comment = PCI

PCI-Distribution of a single post

How many new comments receives a post after *x* minutes?

- Or: Probability of receiving a comment at time *x*?
- Approximately given by a log-normal distribution (LN)





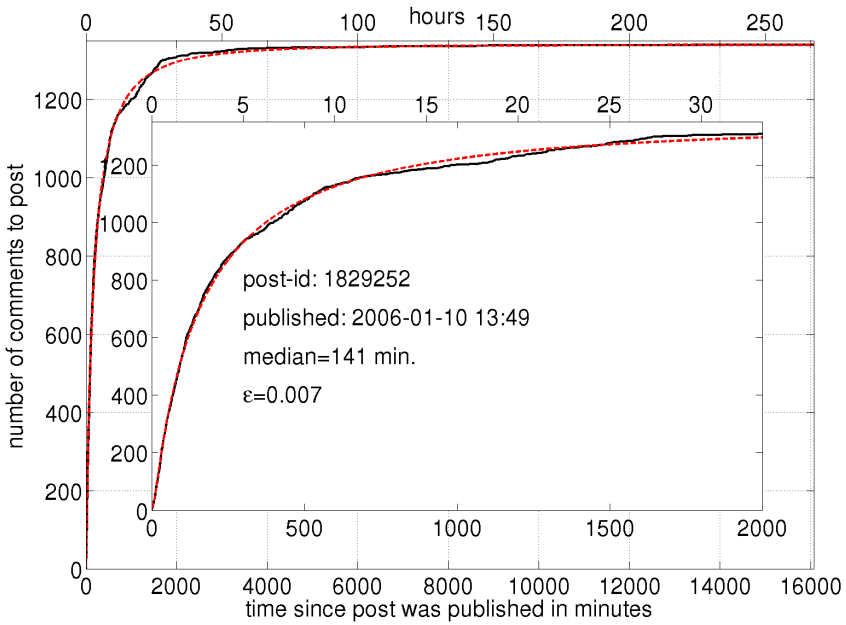
Cumulative PCI-Distribution of a single post

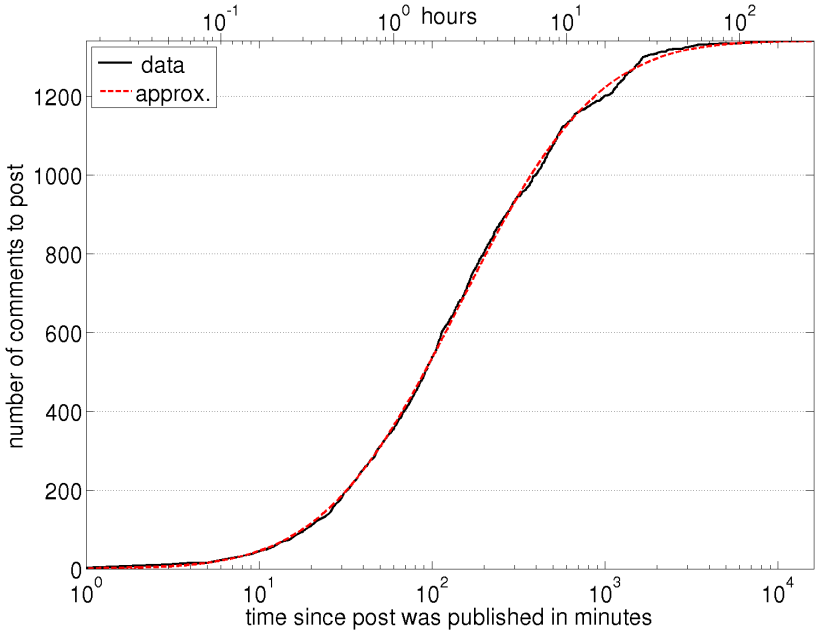
How good is the approximation?

- Fluctuations averaged out in cumulative distribution (cdf)

$$\Downarrow$$

- Quality of approximation becomes better visible.



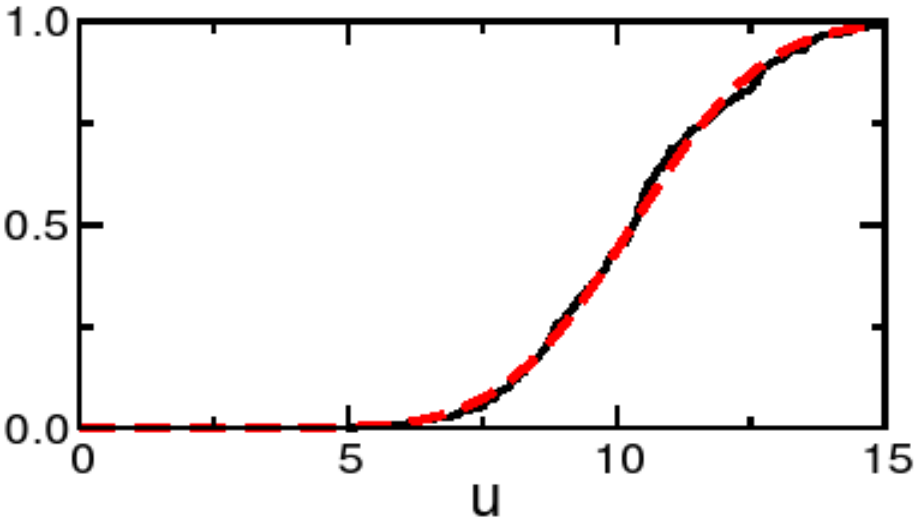


Comparison with one-to-one communication

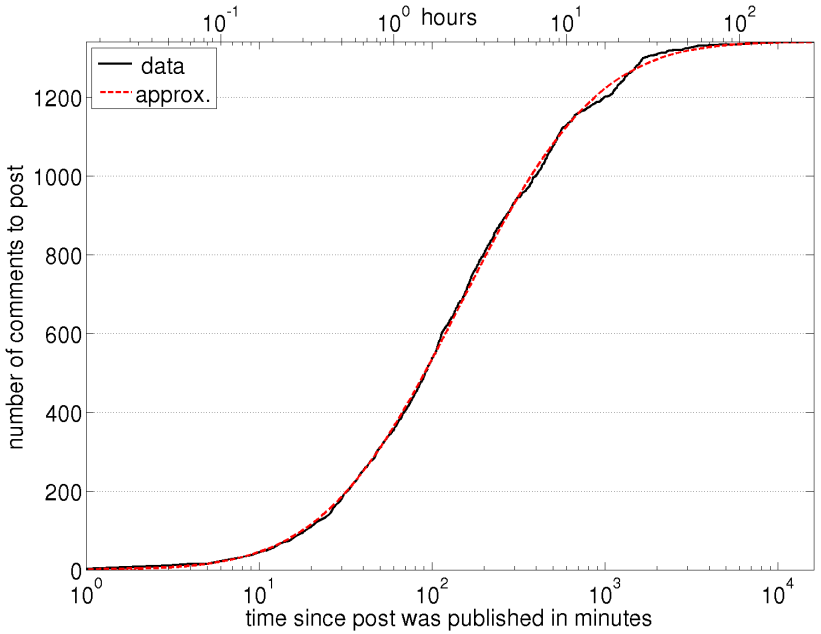
Similar patterns:

Left figure: Response time of a single user to e-mails (Stouffer et al. 2006)

Right figure: Response time to a news-post on Slashdot



$u = \ln(\text{time})$

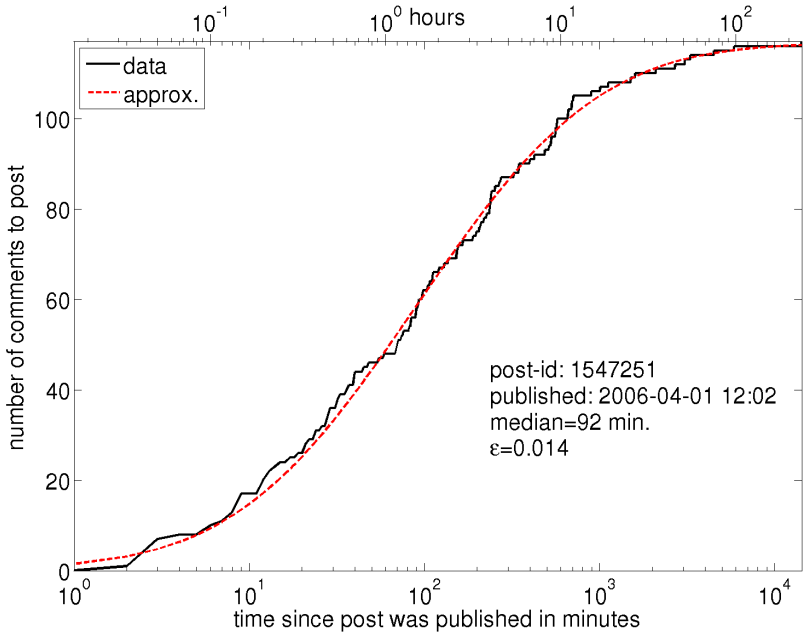


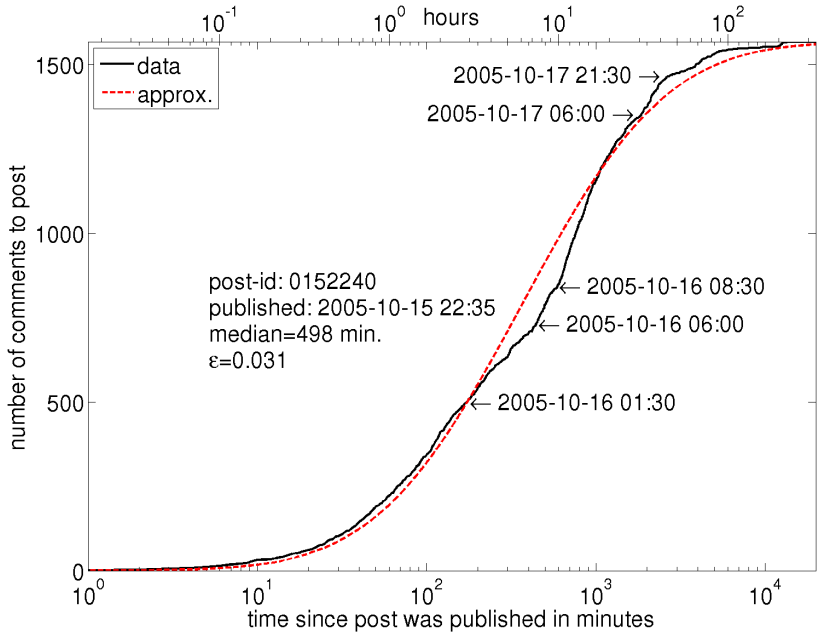
Log-normal distribution of PCI: more examples

2 more examples:

Left figure: Low number of comments

Right figure: Late Post $\Rightarrow$ Bad fit
Caused by circadian rhythm?

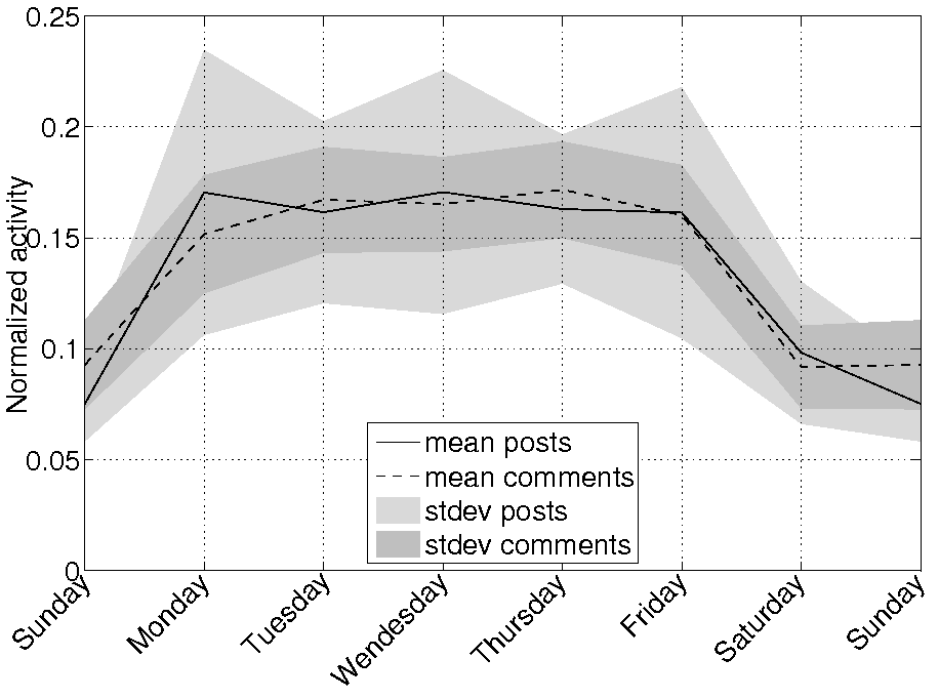




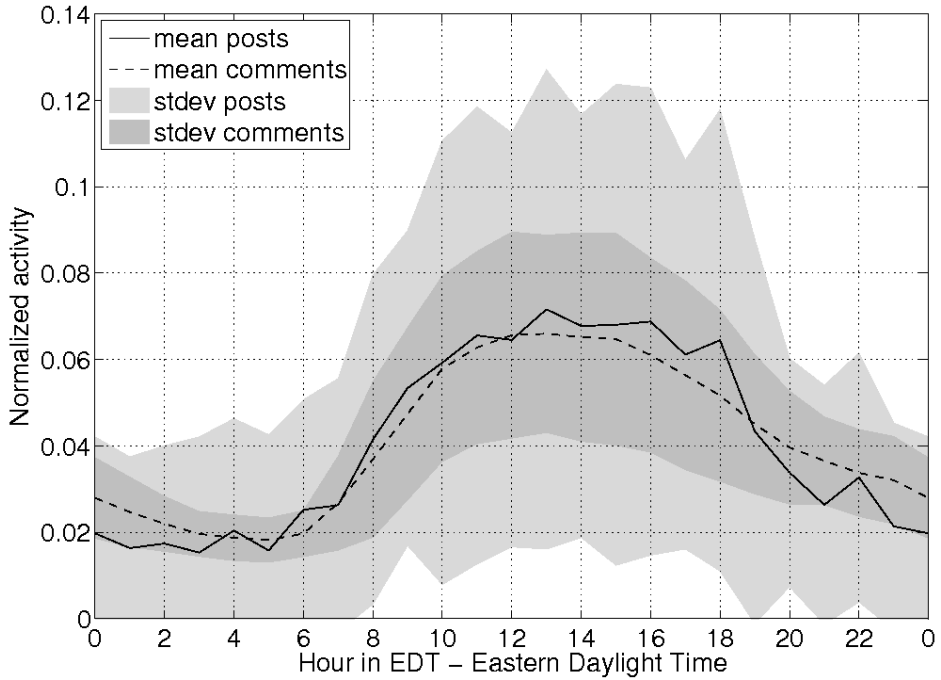
Daily and Weekly Activity cycle

Activity peaks during working hours

Does it influence the log-normal behavior?



Weekly cycle

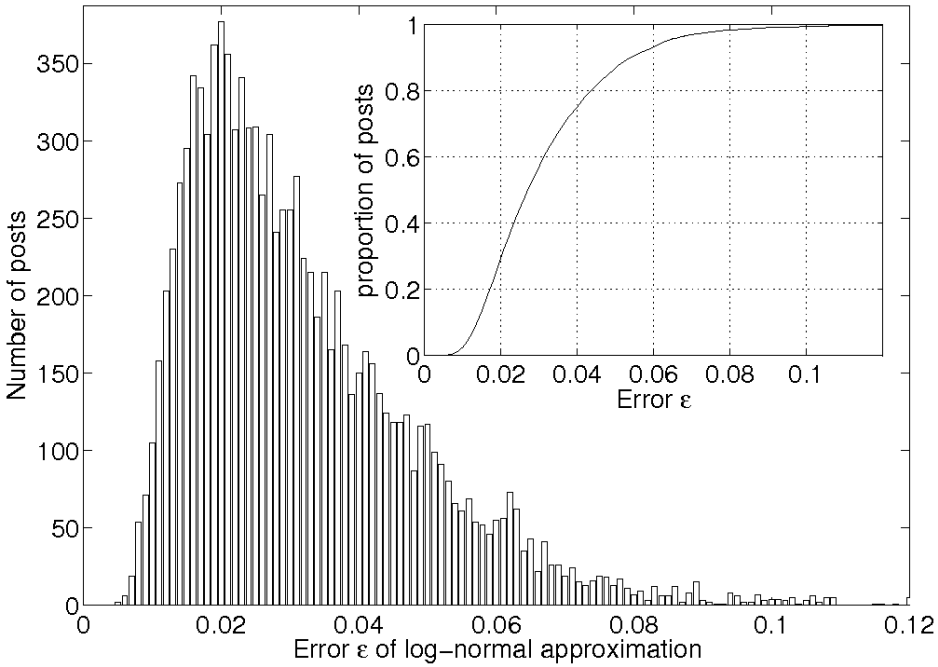


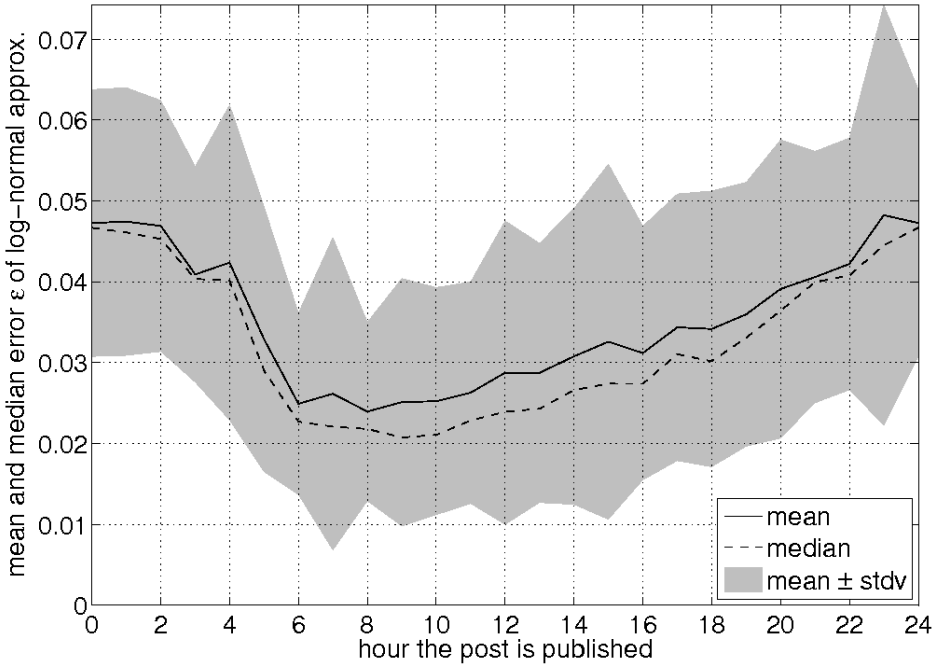
Daily cycle (EDT = GMT − 4h)

Approximation Error

Circadian cycle influences log-normal behavior

- Error ϵ = mean distance between LN-approx. and data
- Quality of LN-approx. depends on publishing-hour of post



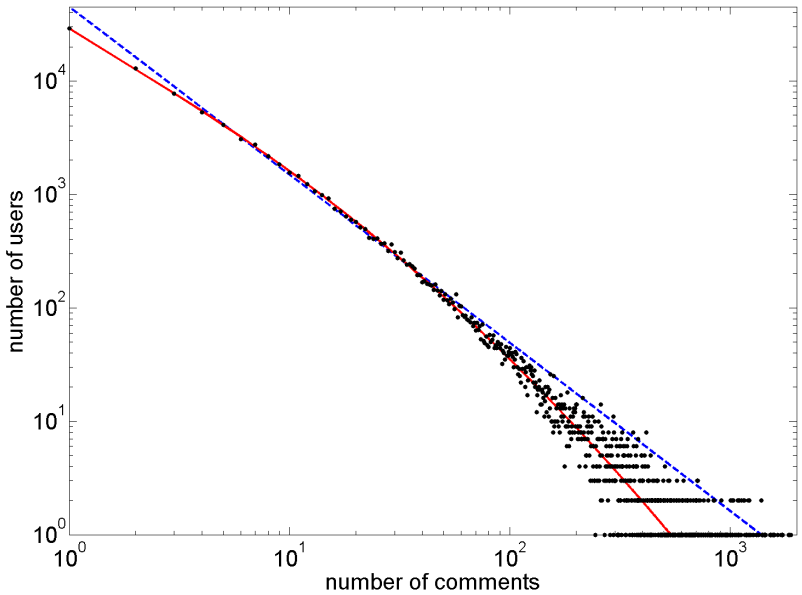


Outline

Number of comments per user I

We analyze:

- How heterogeneous is the population?
- How many comments do the users write?



dots: data

dashed: power-law

continuous: truncated log-normal

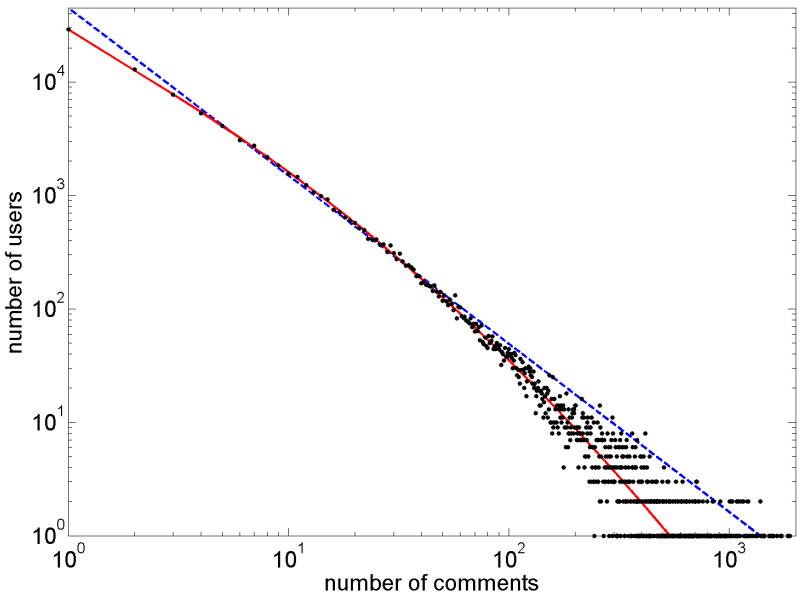
Question

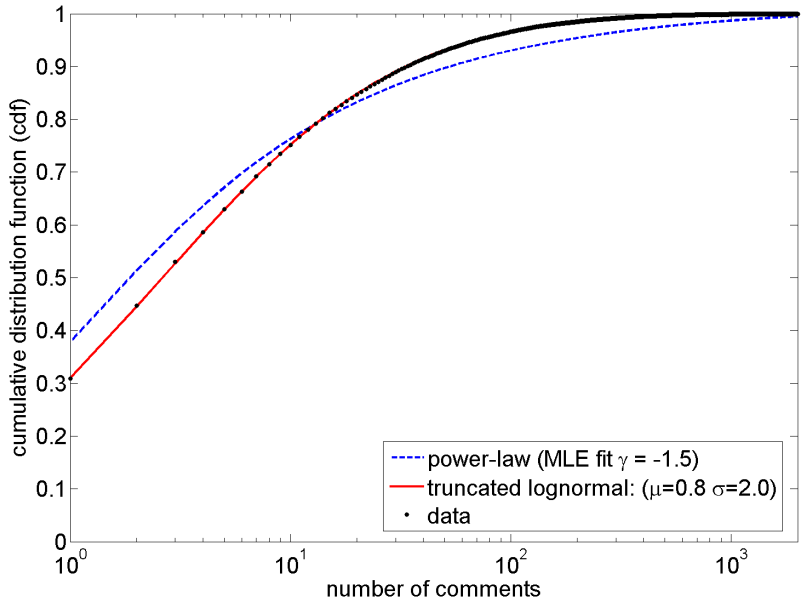
- Log-normal or power-law?

Number of comments per user II

Cumulative distribution gives the answer

- KS-test forces to reject power-law hypothesis
- LN-approximation can characterize entire dataset





Outline

Temporal patterns of single users I

Change of focus

So far: post-comment intervals (PCIs) of single posts

Now: PCIs of single user's comments (to several posts)

Analysis of activity of the two most active users

- Contributions of the two most active users

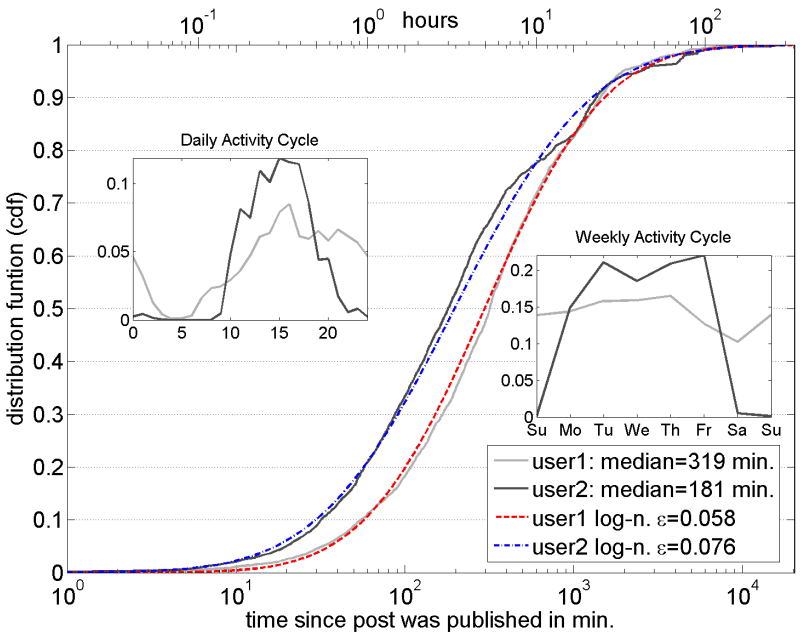
	user1	user2
commented posts	1189	1306
comments	3642	3350

- Both users comment $\approx$ 10% of all posts.
- $\approx$ 3 comments per post

Temporal patterns of single users II

PCIs of single users

- Daily and weekly activity patterns are quite different.
- Nevertheless PCI-distribution resembles a LN.



- Circadian rhythm more pronounced in user2

⇓

- Bumps in PCI after 8 and 16h

Summary

Conclusions

- Activity on Slashdot shows homogeneous patterns.
- Patterns fit log-normal distributions.
- Circadian cycle influences the quality of the fit.
- Power-law models cannot explain the data.

Outlook

- Similar results are found in other datasets

  $\Downarrow$

  Need for model to explain LN-patterns.
- Use results for prediction of expected user activity?

Thank you

Some Formulas

Log-Normal Density

$$f(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(\frac{-(\ln(x)-\mu)^2}{2\sigma^2}\right)$$

Approximation Error ϵ

$\mathbb{T}$: Set of time-bins where a post receives a comment

T: The cardinality of $\mathbb{T}$

$f(t)$: Function approximating $g(t)$ (defined for all $t \in \mathbb{T}$)

Error of $f(t)$: $\epsilon = \frac{1}{T}\sum_{t\in\mathbb{T}} |f(t) - g(t)|$

For Further Reading


Barabási, AL.
The origin of bursts and heavy tails in human dynamics.
Nature **435**:207–211, 2005.


Stouffer, DB, Malmgren RD & Amaral LAN.
Log-normal statistics in e-mail communication patterns.
e-print physics/0605027, 2006.


Vázquez, A, Oliveira, JG, Dezso, Z, Goh, KI, Kondor, I & Barabási, AL.
Modeling bursts and heavy tails in human dynamics.
Physical Review E **73**:036127, 2006.