

A Case Study of Open Source Software

Measuring the Effect of Social Communications on Individual Working Rhythms

Qi Xuan, Mohammad Gharehyazie, Premkumar T Devanbu, Vladimir Filkov
Department of Computer Science, University of California, Davis
Email: qxuan@ucdavis.edu



SocialInformatics2012

UC DAVIS
UNIVERSITY OF CALIFORNIA

Rhythm in Music



Fryderyk Fanciszek Chopin
1810-1849

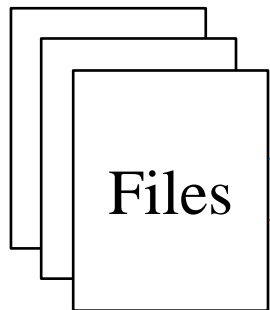
Slow



Maksim Mrvica
1975-

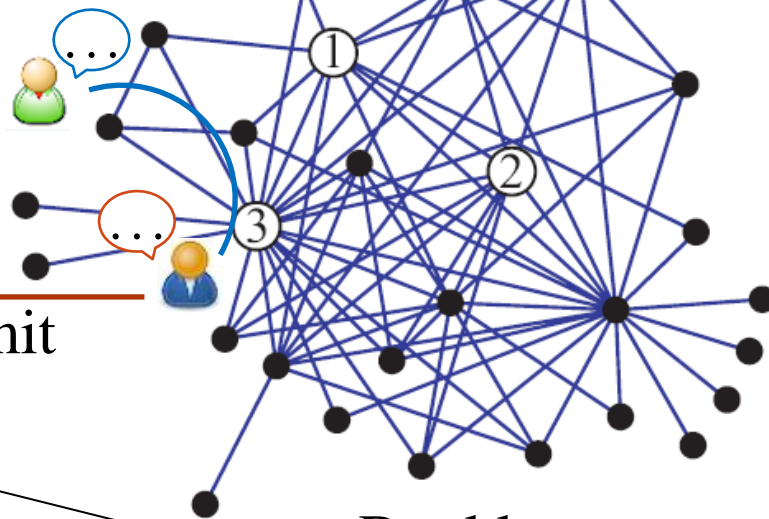
Fast

Rhythm in OSS



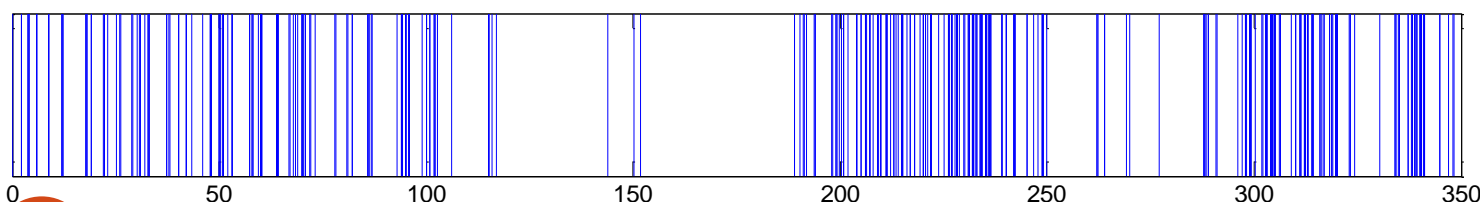
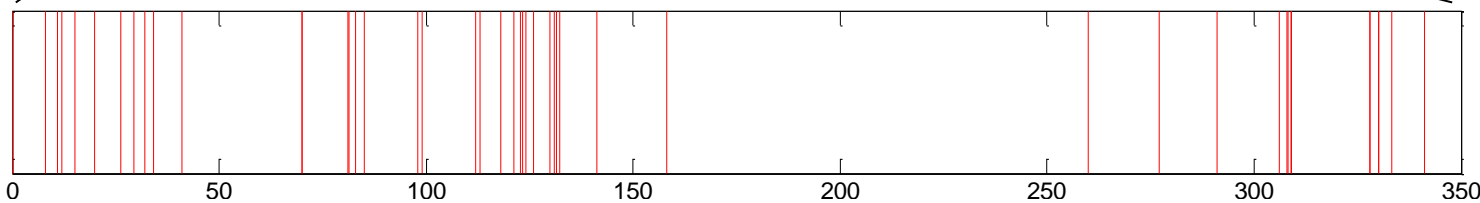
Use/Evaluate

Commit

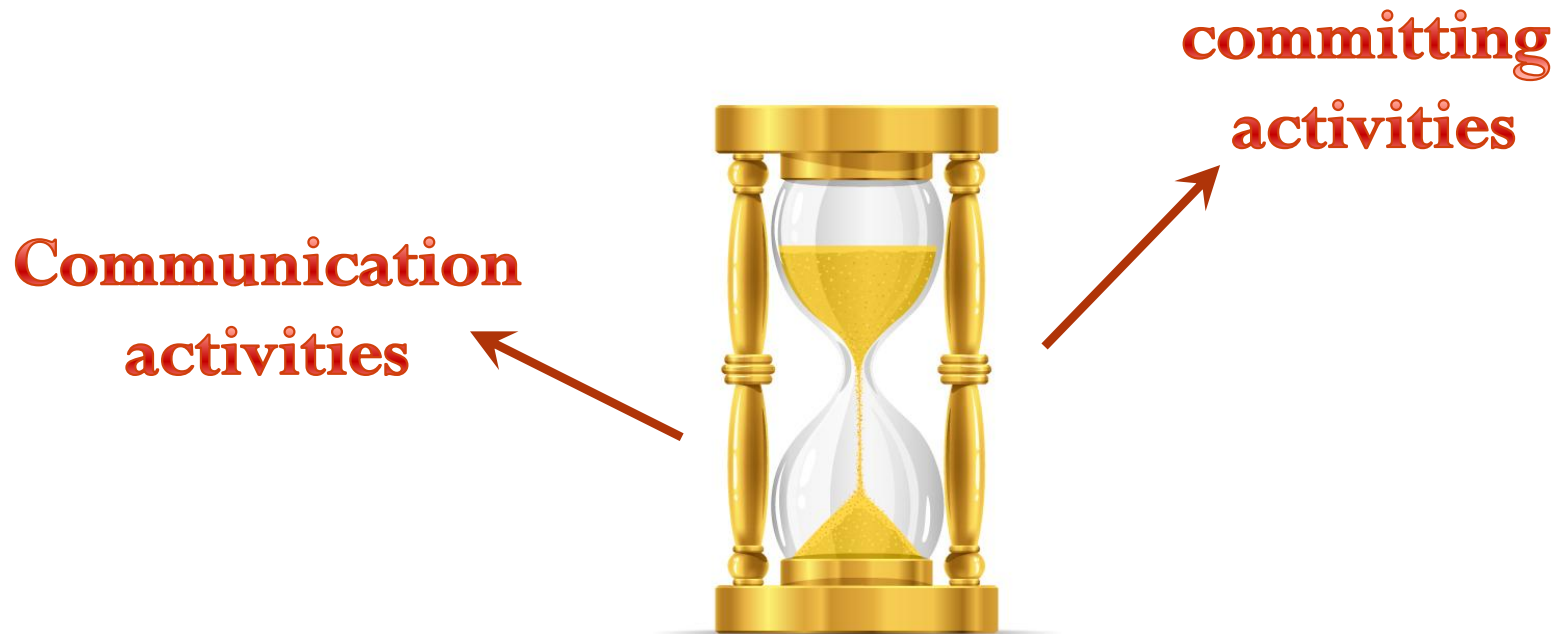


Rhythms

Productivities, Responsibilities, etc



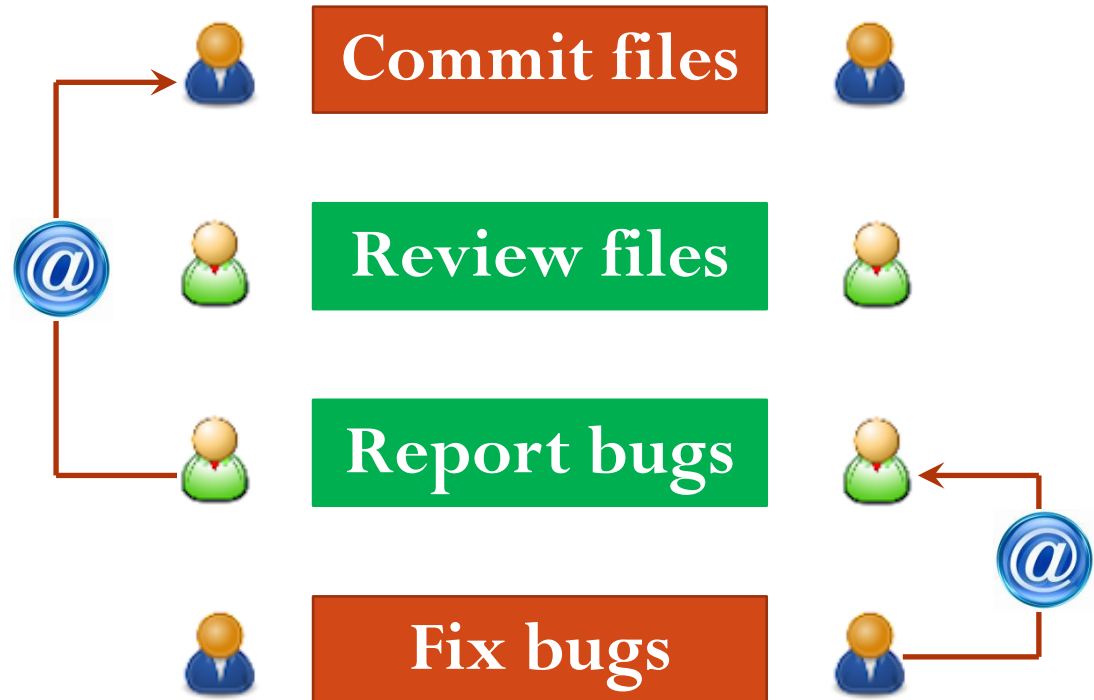
Work Vs. Talk: Negative effects



Few words, many deeds.

Work VS Talk: Positive effect

Evaluation-Response Mechanism



With great power comes great responsibility.

Which is true?

Do the communication activities impede committing activities or do they accelerate them?



Our contribution

Propose two methods:

Method #1: Network based method

Method #2: Time-series based method

to macroscopically and microscopically quantify the relationship between committing rhythm and communication activities of developers based on the real OSS data collected from Apache Software Foundation.

Data description

Project	N_T	N_D	N_F	$\langle k_{out} \rangle$	$\langle k_{in} \rangle$
Accumulo	75	5	1622	3.5	3.5
Mahout	552	16	5123	4.4	4.4
Lucene	2148	41	6674	4.2	4.2
Nutch	862	16	3072	3.4	3.4
Derby	1128	35	6563	5.6	5.6
Ode	377	18	11006	3.8	3.8
Openejb	179	38	43960	5.8	5.8
Log4php	88	9	1409	2.1	2.1
Wicket	540	24	48045	5.5	5.5
Log4j	540	19	5519	2.3	2.3
Bookkeeper	32	3	407	3.0	3.0

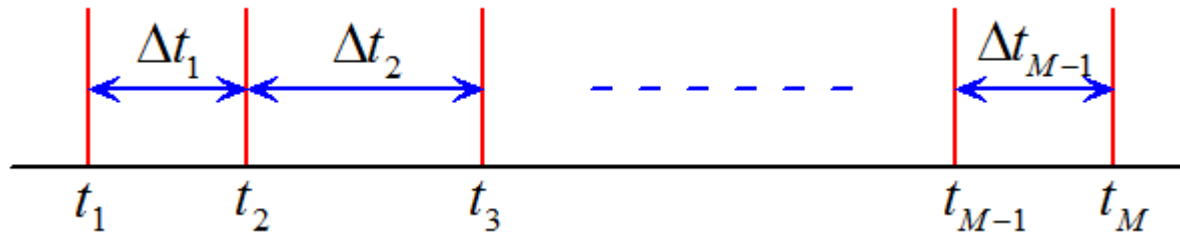
.....

702 developers in 31 projects

From Apache Software Foundation on March 24th, 2012

Definition of Rhythm

Time-series of committing activities:



List of inter-activity time intervals:

$$\Delta t_i = t_{i+1} - t_i, i = 1, 2, \dots, M - 1$$

Average inter-activity time interval:

$$\langle \Delta t \rangle = \frac{\sum_{i=1}^{M-1} \Delta t_i}{M - 1} = \frac{t_M - t_1}{M - 1}$$

Network Based Method

RQ #1: Whether the developers with higher social status have faster or slower committing rhythms than those with lower social status?



Two Cases

Case I: The developers are divided into two groups according to their *incoming degrees*.

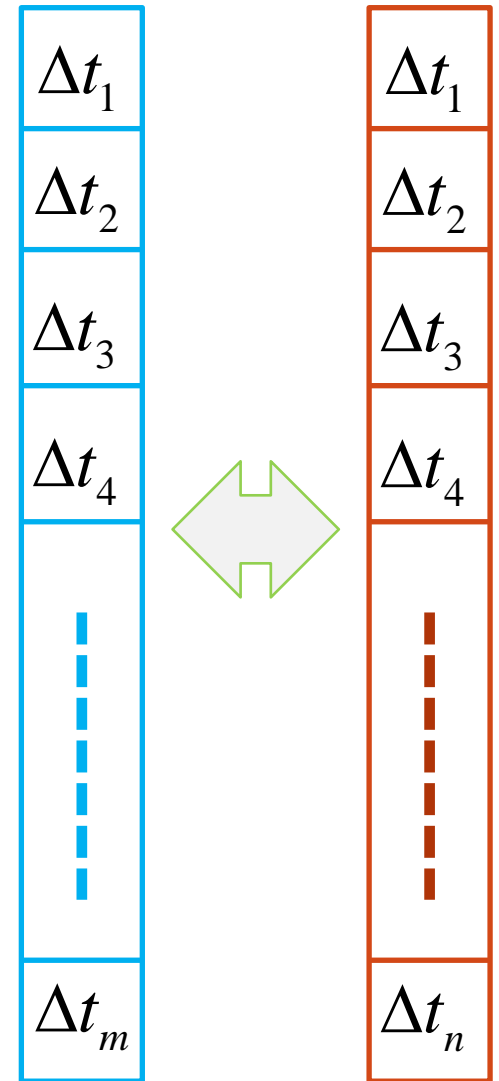
$$k_{in} \leq 50$$

$$k_{in} > 50$$

Case II: The developers are divided into two groups according to their *outgoing degrees*.

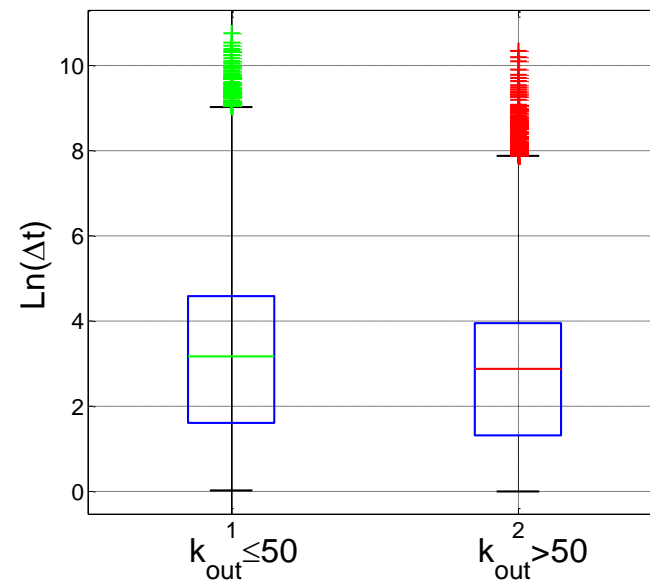
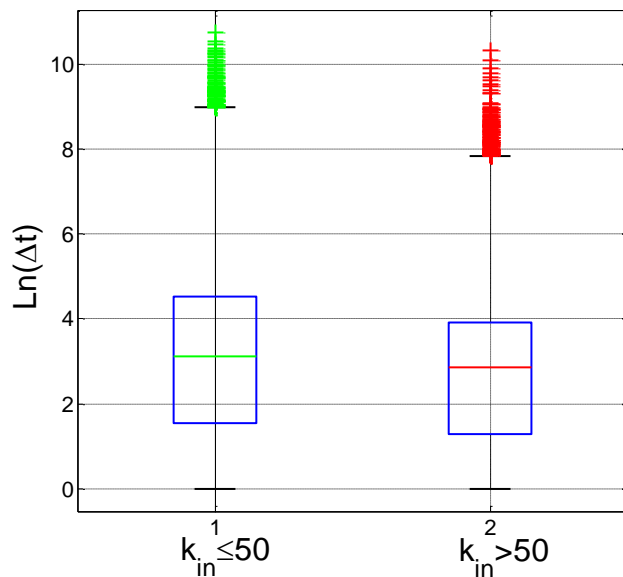
$$k_{out} \leq 50$$

$$k_{out} > 50$$



Results

T-test	$k \leq 50$	$k > 50$	T-value	Significance
Case I	103.0 (h)	45.5 (h)	20.4	$p < 10^{-6}$
Case II	111.2 (h)	48.5 (h)	22.3	$p < 10^{-6}$



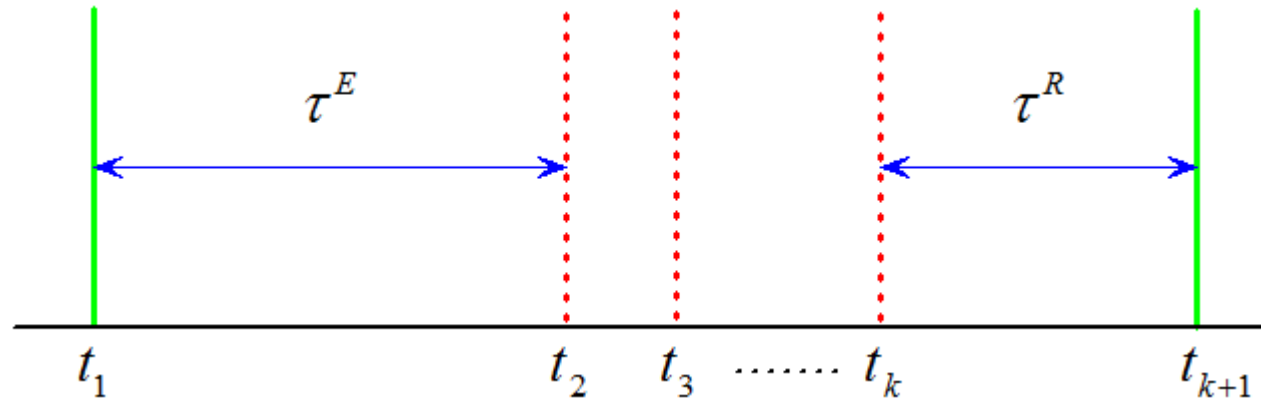
Higher social status indicate faster committing rhythms.

Time-Series Based Method

RQ #2: Is there an evaluation-response mechanism in OSS projects so that committing and communication activities accelerate each other?



Evaluation & Response Latency



Submit codes

Evaluations

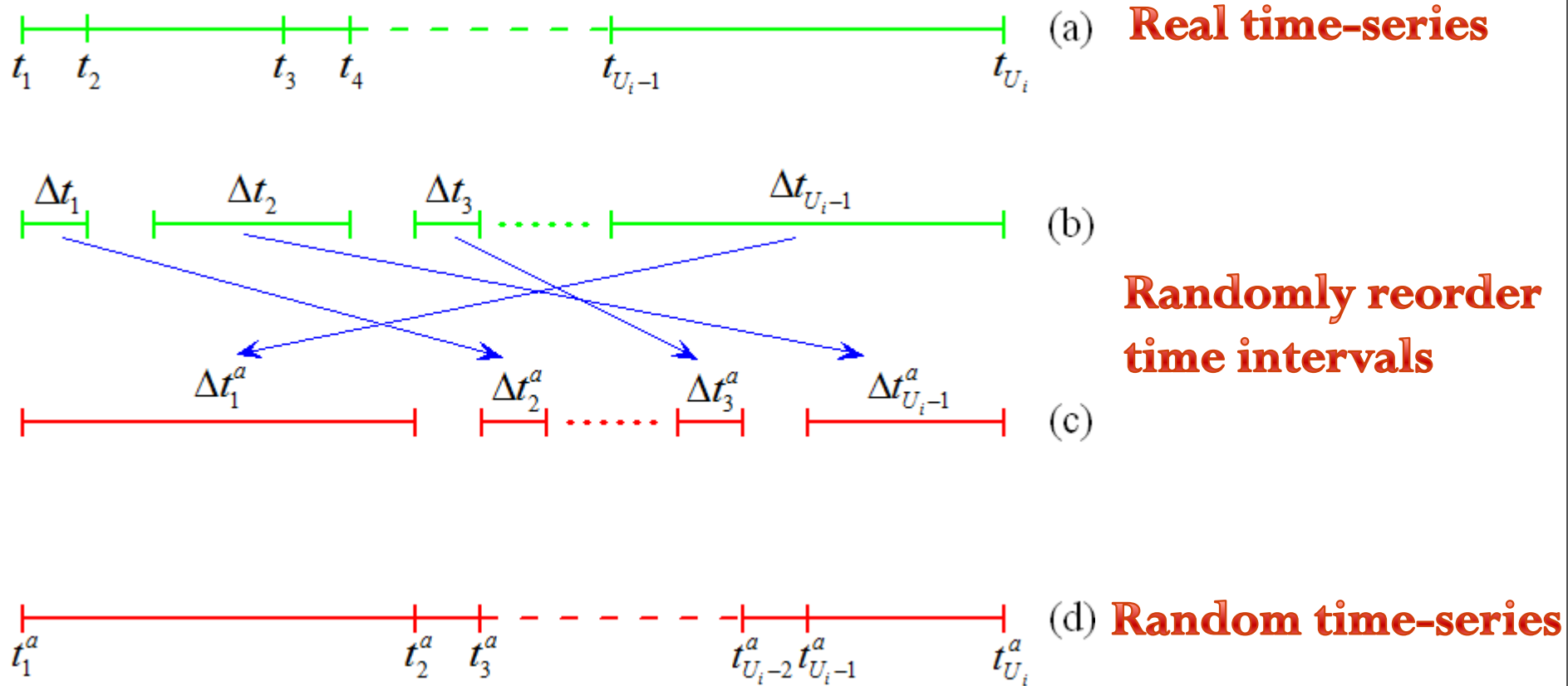
Response

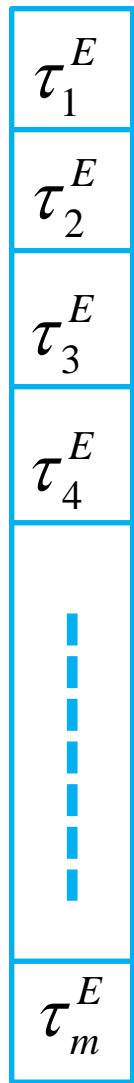
Code reviews and problem reports

Evaluation latency: $\tau^E = t_2 - t_1$

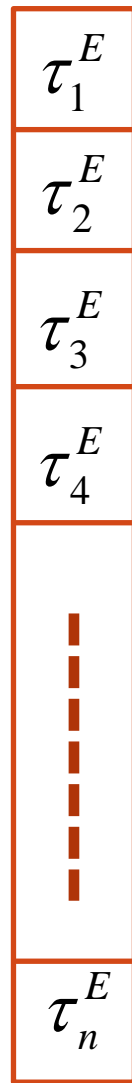
Response latency: $\tau^R = t_{k+1} - t_k$

Random Time-Series

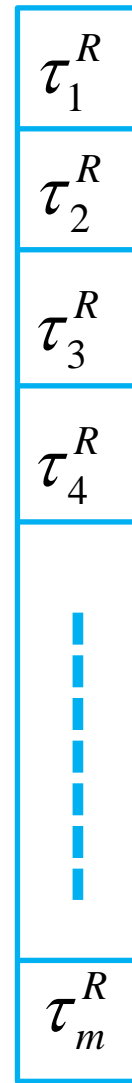




Real



Simulated



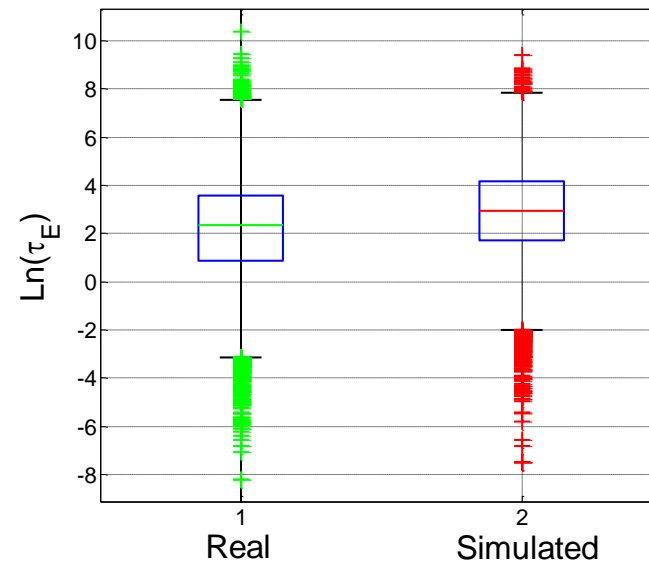
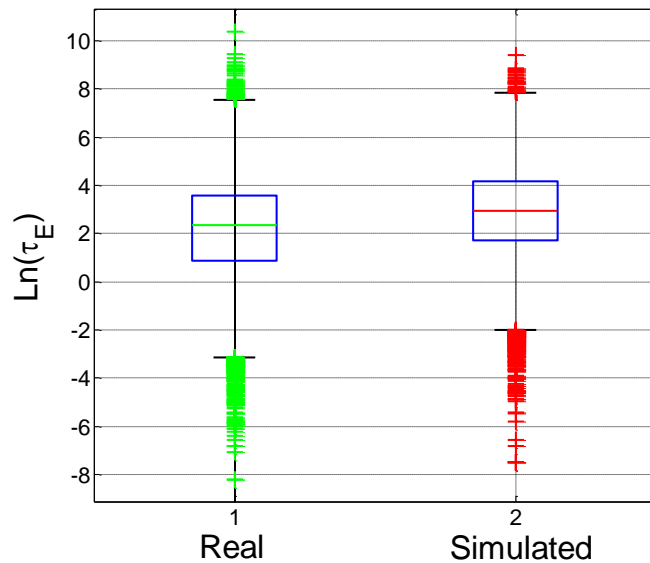
Real



Simulated

Results

T-test	simulated	Real	T-value	Significance
Evaluation	89.9 (h)	58.5 (h)	8.21	$p < 10^{-6}$
Response	97.0 (h)	57.2 (h)	9.92	$p < 10^{-6}$



Communication and committing activities accelerate each other.

Limitations & Future Works

- **Limitations:** In this work, we just use the time-series of committing and communication activities to estimate the evaluation and response latencies, which needs to be further checked by textural analysis of emails and codes.
- **Future works:** the network and time-series methods can be used together in order to reveal whether the individuals response at different rhythms for the evaluations from others of different social status.

Thanks!