What Make Long Term Contributors

Willingness and Opportunity in Open Source Community

Minghui Zhou   Audris Mockus

Peking University   Avaya Labs Research
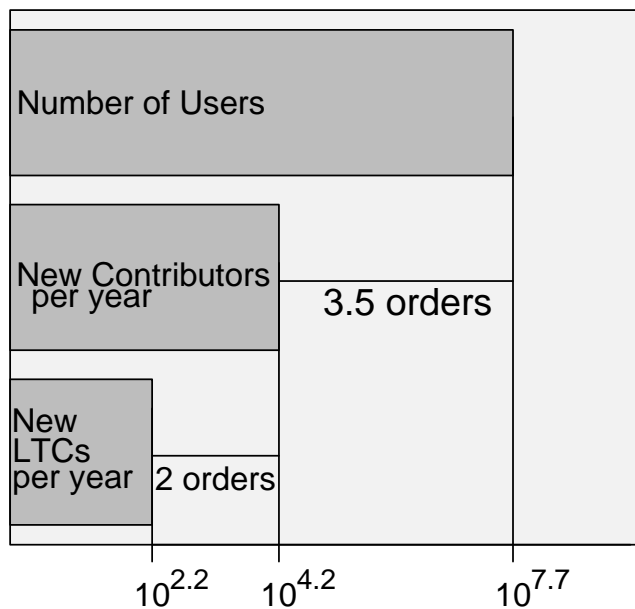
zhmh@pku.edu.cn   *audris@avaya.com*

Outline

❖ Long-term contributors (LTCs) are crucial to project success

❖ Context: million+ issues reported for Gnome and Mozilla

❖ Questions

  – Why some become LTCs and others don't?
  – Can we tell during their first month?

❖ Answers

  – Because of their ability, willingness, and environment
  – Yes

❖ Implications

  – Projects: take care of newcomers
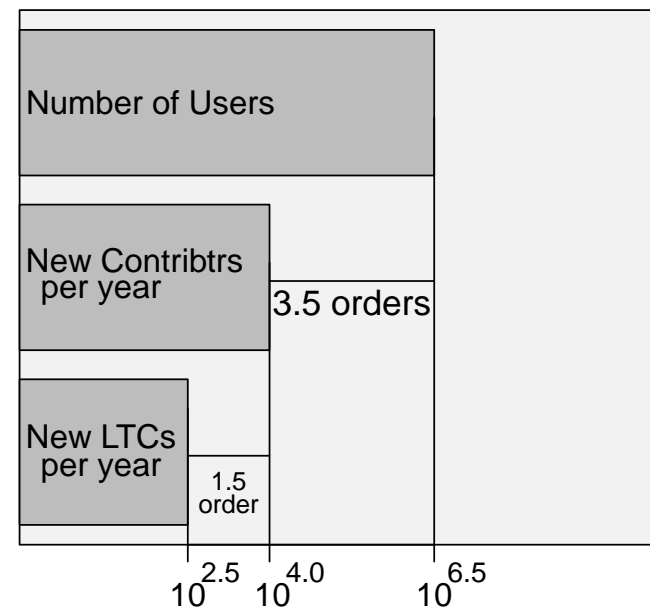 – Newcomers: be more community-oriented

"OSS doesn't work without contributions from the community"

- ❖ Only long-term contributors can accomplish critical tasks
  - ⬦ Developers take at least three years to become fluent [FSE'10]
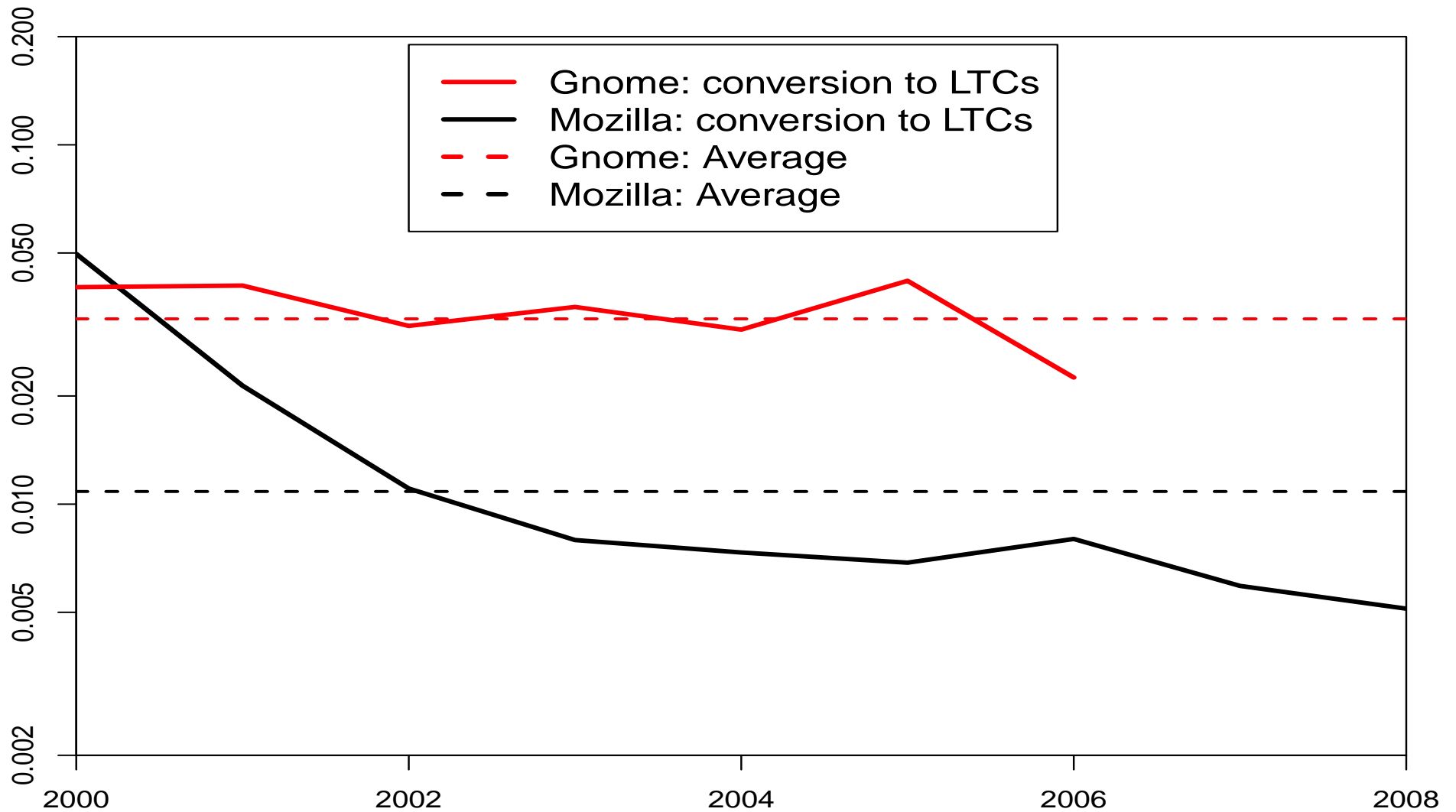- ❖ Few newcomers become Long-Term Contributors (LTCs)

Mozilla (average over 2000–2008)

Number of Users

New Contributors per year

3.5 orders

New LTCs per year | 2 orders

$10^{2.2}$ $10^{4.2}$ $10^{7.7}$

Gnome (average over 1999–2007)

Number of Users

New Contribtrs per year

3.5 orders

New LTCs per year | 1.5 order

$10^{2.5}$ $10^{4.0}$ $10^{6.5}$

Newcomer to LTC conversion drops!

What Make Long Term Contributors Zürich, 2012

Approach

- Learn what was going on

  - Transcribe recurring themes associated with future LTCs

    - Read issues of 40 contributors (20 non-LTCs/20 LTCs)
    - Survey 56 (36 non-LTCs and 20 LTCs)
    - Extract practices published on project web sites
    - Review other research on Gnome and Mozilla

- Measure discovered factors via activity in Bugzilla

- Fit models of future LTCs

- Validate

  - Predict future LTCs
  - Investigate stability and data quality

- Interpret, consider practical implications, future

Ability/Willingness distinguishes LTCs

- Numbers and types of tasks

  - Non-LTC: "I don't have enough time/knowledge to resolve issues by myself", provide minimum information necessary to report, don't respond to requests for information
  - LTC: "Patch to get access attributes for nested class/struct/union"
  - LTCs had higher response rate (Fisher's-test p-value=0.07)

- Willing to spend more effort on tasks

  - "If I want the bugs to go away, I have to be willing to note the bugs."
  - "If you have faced a bug, you need to spend effort to describe it... to check for duplicates... to create report... to wait until response."
  - "All time you are waiting you must keep an issue in mind."
  - "After [the] initial response there is [a] good possibility that devs can't or don't want to reproduce the issue and you must know how to [do] diagnostics and how to prove that issue really exists."

Environment determines people's fate

- Macro-climate: *popularity*:
  - "GNOME is something which you can show to your friends and family members"

- Micro-climate: *attention, number of peers, performance of peers*
  - "With bugzilla, ... the feedback from the developers shows that they care, and appreciate the effort I made, and actively work to solve the bug in a way that I can see progress."
  - "As I met a lot of nice people at GUADECs who became friends there was also a personal component involved in the motivation."
  - "I learned a lot from this leading open source project while working with other contributors"

Measures of Ability/Willingness and Environment

✦ Observation I: Ability/Willingness can be measured via

  ✧ The volume and the type of tasks
  ✧ The effort spent on tasks

✦ Observation II: Environment can be measured via

  ✧ Macro-climate (shared among participants)
    ✧ Project's popularity
    ✧ Project's relative sociality
  ✧ Micro-climate (unique for each person)
    ✧ Number of peers
    ✧ Peers' productivity
    ✧ Peers' social clustering
    ✧ The attention received from peers

Three dimensions

What Make Long Term Contributors
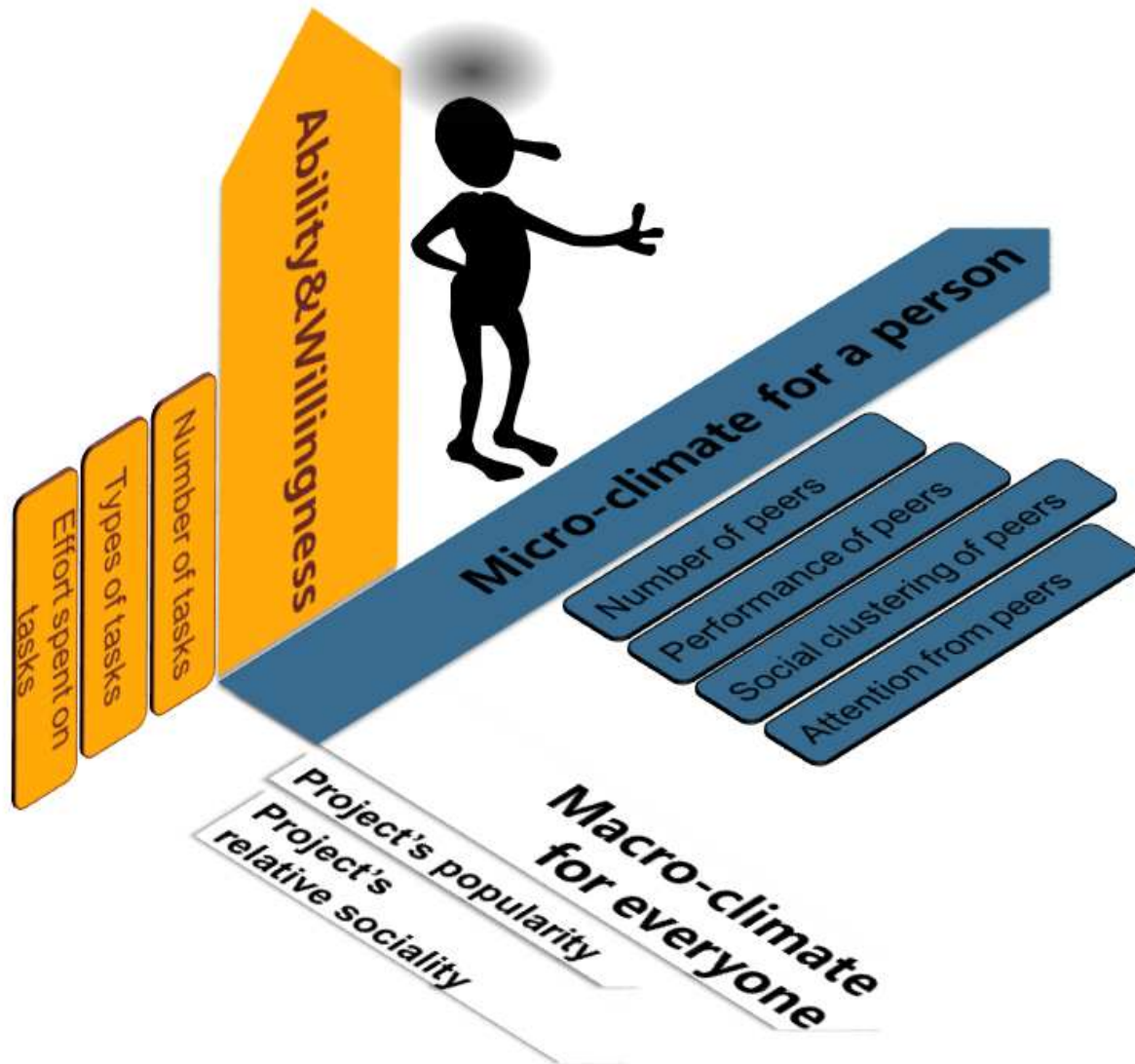
Logistic regression model for LTCs

Measure	Predictor	Odds Ratio		Direction
		Mozilla	Gnome	
Ability & Willingness	got at least one fix	2	2	⇑
	comment/not BB	1.5	3	⇑
	number of comments	2	1.5	⇑
Micro env	lack of attention	$\frac{2}{3}$	$\frac{2}{3}$	⇓
	peers' productivity	1.2	2	⇑
	peers' soc. clust.	1.5	1.2	⇑
	number of peers	1.14	0.94	⇕
Macro env.	number of users	0.85	$\frac{1}{2}$	⇓
	relative sociality	1.07	0.73	⇕

Response: {not-LTC, LTC} for Mozilla/Gnome (130,472/125,665 observations)

What Make Long Term Contributors Zürich, 2012

Who will become an LTC?

- ✦ Actions in the first month predict LTCs
  - ✧ Pro-community attitude has the greatest positive effect
    - ✧ The choice to start by a comment for an existing issue
    - ✧ Effort spent to improve the quality of issue reporting
  - ✧ Bad environment deters via
    - ✧ Macro-climate of high project popularity
    - ✧ Micro-climate of low attention
  - ✧ Good environment attracts via
    - ✧ Micro-climate of peer performance and
    - ✧ Micro-climate of peer social clustering

Can we predict future LTCs?

2008/1   2009/5   2011/1   2012/5
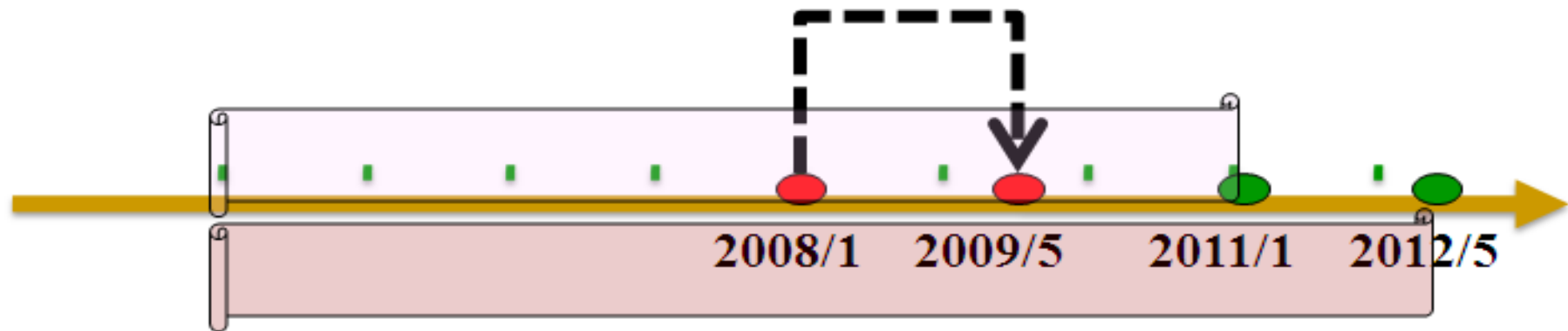
- ❖ Created prediction using 2011 snapshot:
    - ⬦ 25,406 joiners during 2008.01-2009.05
- ❖ Determine LTCs from a new Mozilla snapshot on 2012.05
- ❖ Prediction performance
    - ⬦ 24% recall (32 out of 131 LTCs were predicted)
    - ⬦ 37% precision (32 of 86 predictions were LTCs)
    - ⬦ 72 times higher than a random choice

Limitations

- Four snapshots for Gnome data and two for Mozilla

- Sensitivity analysis using various operationalizations

  - Full email was not available for post-2008 Gnome

  - Person to ID (email) changes over time

- Variation in operationalizations

  - BugBuddy in Gnome vs start from a bug report in Mozilla

- Do measures capture the right concepts: e.g., peer clustering

- Should relationships be in the observed direction: e.g. project popularity is bad?

- Are Gnome and Mozilla projects representative?

Summary of Contributions

- Methodology
  - Measure individuals' attitudes and emotional dispositions from digital traces of their activity

- Science
  - Models of project success show largest effects brought by soft qualities, such as willingness

- Software practice
  - Projects: particular attention for new contributors
  - Newcomers: deeds matter, not intentions, limit expectations

- Future and Reproducibility
  - Implications for OSS and commercial development practices and non-software domains
  - http://www.passionlab.org/projects/developerfluency.html

Reading Issues

- ❖ non-LTC: Alice reported 2 issues: 435220 and 450656

  - ◇ Provided only minimal information needed to report the bug according to a template
  - ◇ Didn't respond to request "Could you please help fixing this by installing some debugging packages..."
  - ◇ The issue was resolved as INCOMPLETE

- ❖ LTC: Bob's first issue report

  - ◇ "Patch to get access attributes for nested class/struct/union"
  - ◇ Gnome developer responded "I'll include it in the first CVS release"
  - ◇ The issue was resolved as FIXED

Examples of survey responses

- What motivated you to start contributing?

  - "When I was a college student I was dreaming to be a hacker"
  - "It is kind of like making the world a better place in small steps"

- What caused you to continue your contributions?

  - "I learned a lot from this leading open source project while working with other contributors"
  - "When I installed Linux for the first time I was fascinated by the names of individuals in those boxes. So, basically, I wanted to have my name there"

- LTCs had higher response rate (Fisher's-test p-value=0.07)