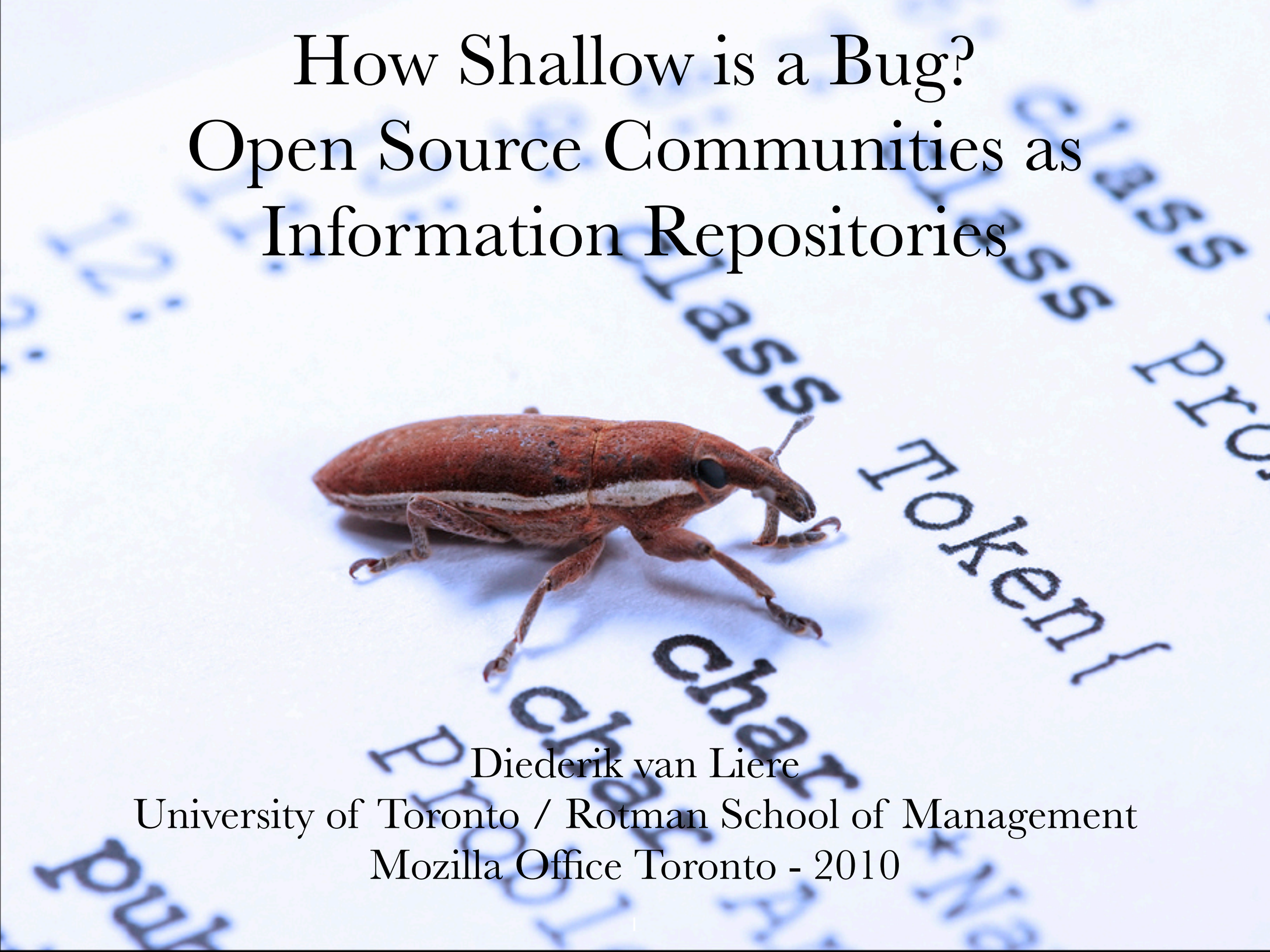
How Shallow is a Bug?
Open Source Communities as Information Repositories



Diederik van Liere
University of Toronto / Rotman School of Management
Mozilla Office Toronto - 2010

Who am I?

- Post-doc researcher at the strategy department Rotman School of Management / University of Toronto
- PhD information & decision sciences department at the Rotterdam School of Management
- "Networkophile"
- My research focuses on the intersection of social & digital networks and open source software
- Mosaic / Netscape / Firefox user since 1994
- Blog: www.network-labs.org

My Coding Credentials

Python:

```
print “Hello Mozilla Toronto”
```

PHP:

```
print “Hello Mozilla Toronto”;
```

jQuery:

```
$(document).ready(function(){

  $(“body”).append(“<p>Hello Mozilla
  Toronto</p>”);

});
```

So…it feels like being thrown in front of the lions

“All Bugs are Shallow Given Enough Eyeballs”

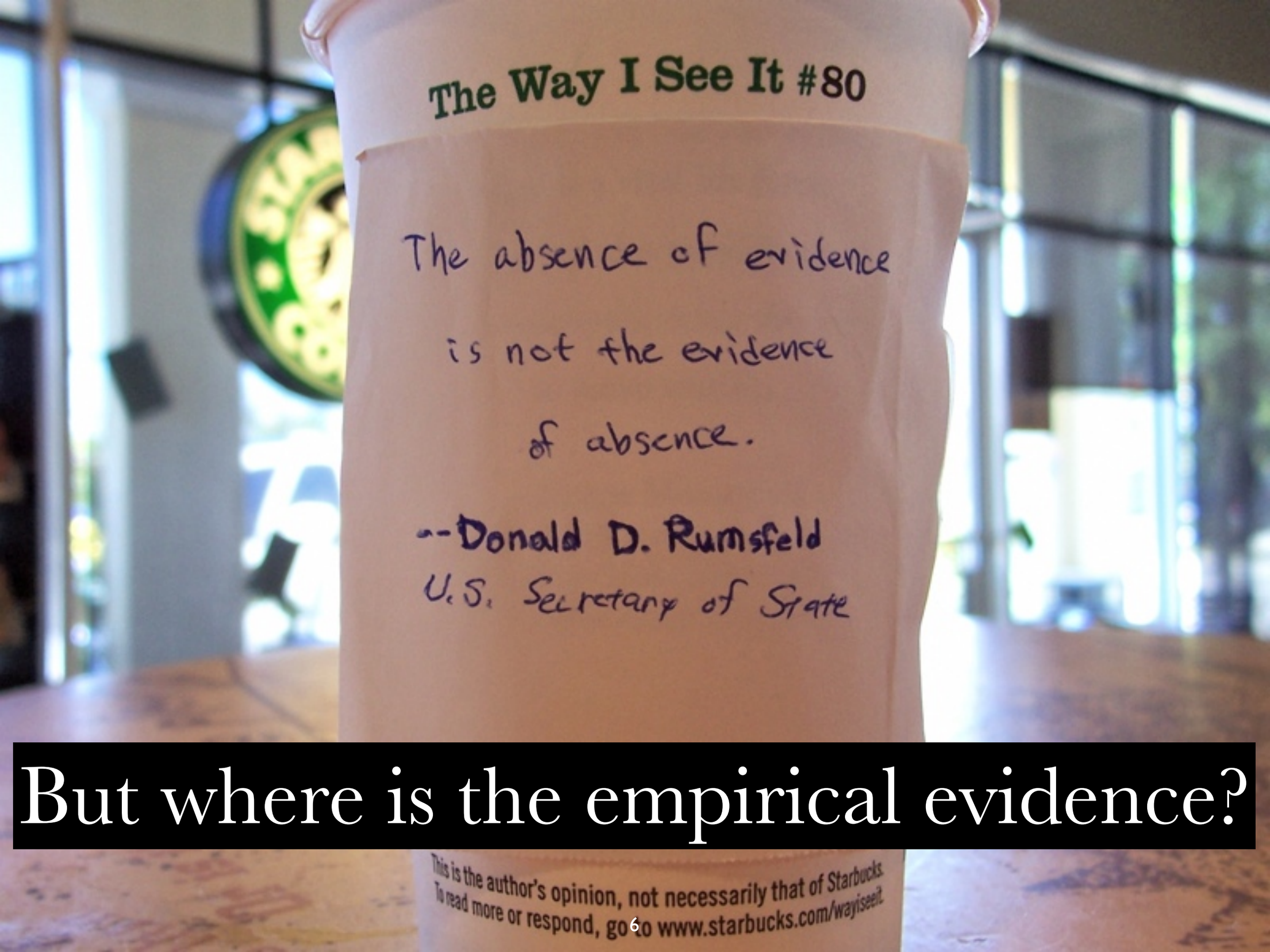
Linus Torvalds as quoted by Eric S. Raymond

The Way I See It #80

The absence of evidence
is not the evidence
of absence.

--Donald D. Rumsfeld
U.S. Secretary of State

This is the author's opinion, not necessarily that of Starbucks.
To read more or respond, go to www.starbucks.com/wayiseeit



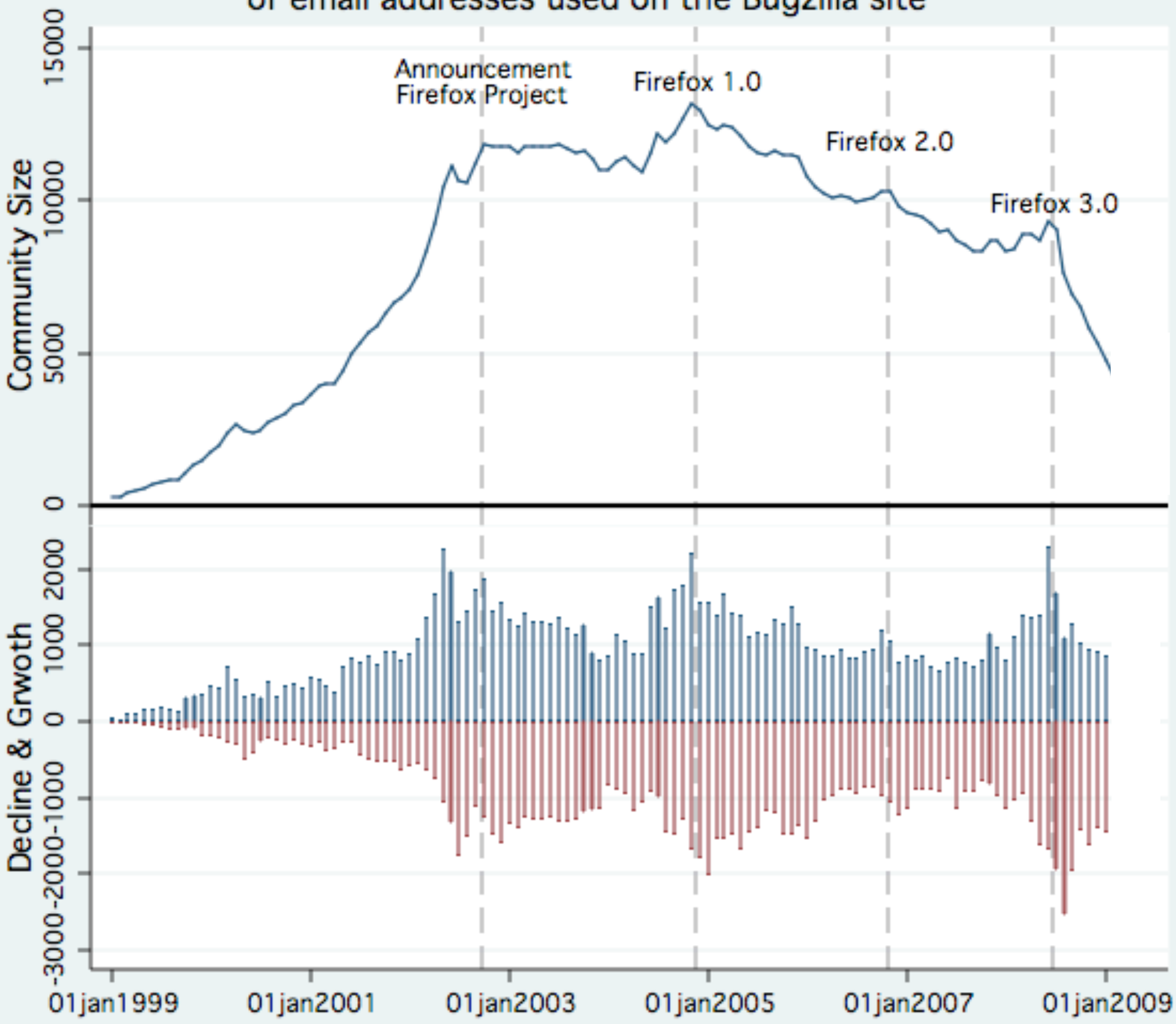
But where is the empirical evidence?

Data Collection - Firefox

- A community member is someone who posted at least 1 message at bugzilla.mozilla.org
  - Date of entry: first time message posted
  - Date of exit: a month after last message posted
- Collected bug reports filed at bugzilla.mozilla.org with id 1 - 480.000 if product is Firefox / Core / Seamonkey, in total 320.655 bug reports
- This covers the period late 1998 to December 31st, 2008
- Tried to match different email addresses to single developer (more about this later)

Firefox Community Dynamics: 1999 - 2009

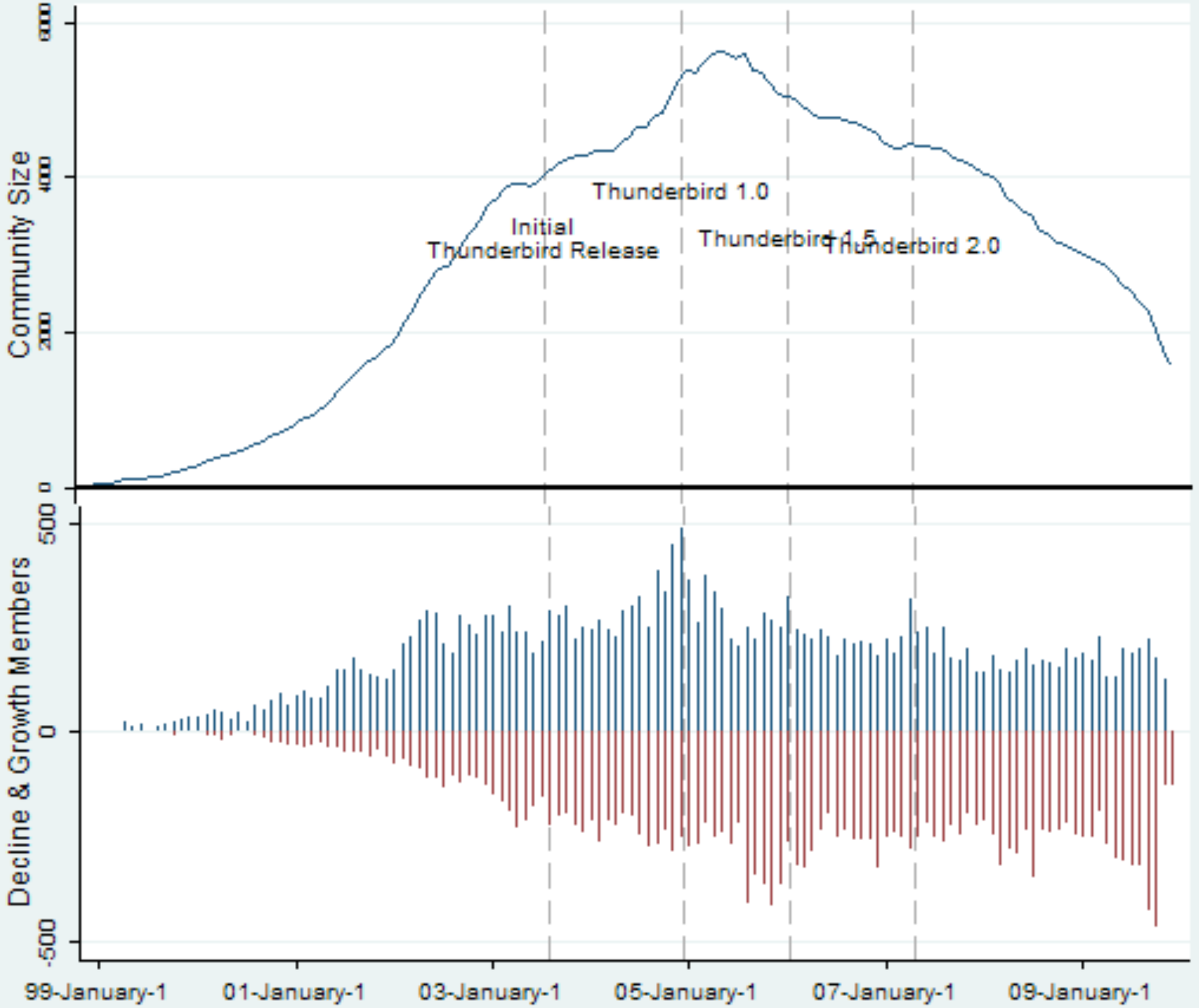
Community size is measured by counting the unique number of email addresses used on the Bugzilla site



Data analysis and visualization by Diederik van Liere | www.network-labs.org

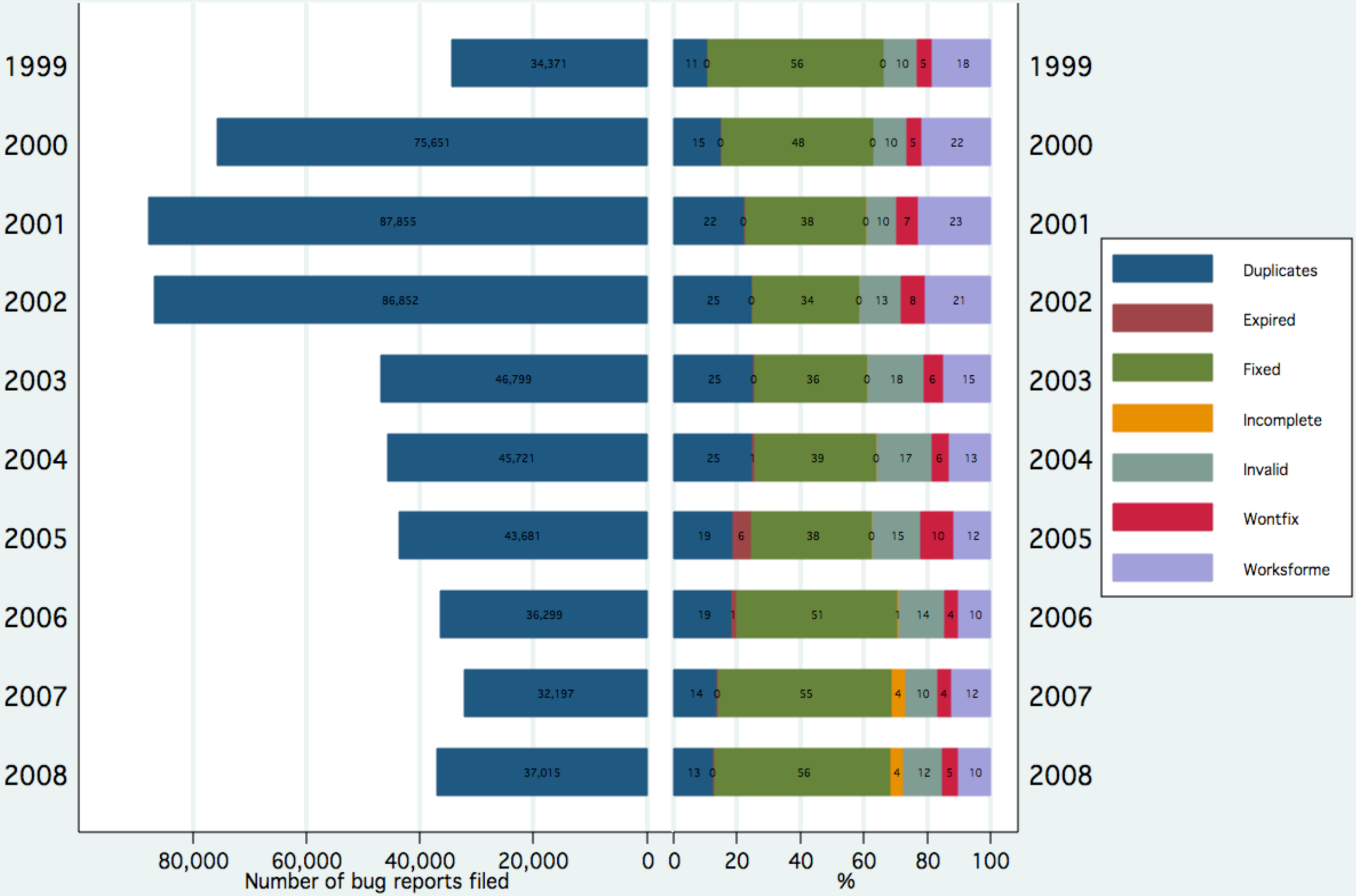
Thunderbird Community Dynamics: 1999 - 2009

Community size is measured by counting the unique number of email addresses used on the Bugzilla site



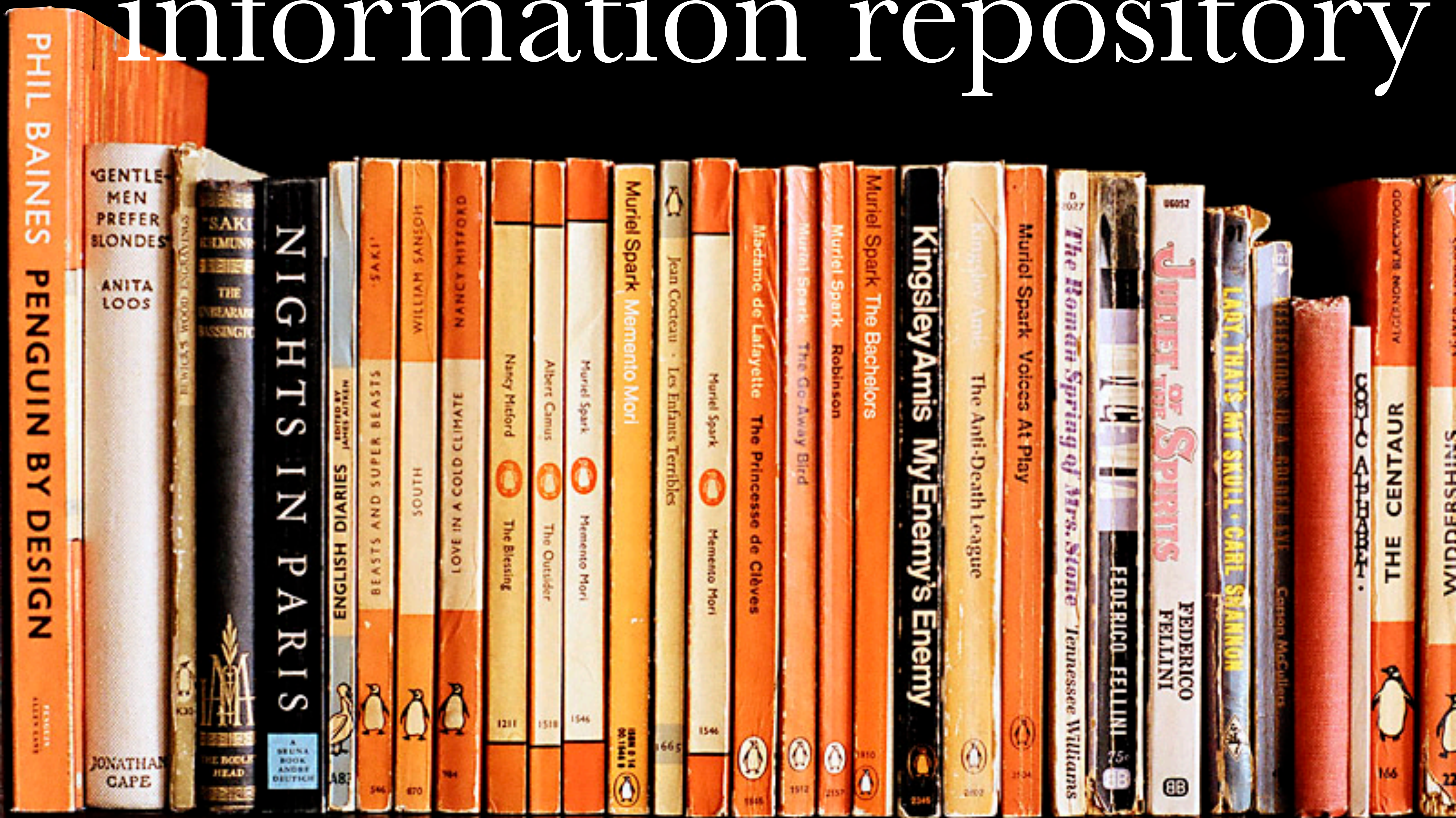
Bug Reports Filed by Community Members: 1999 - 2008

Breakdown of resolutions of bug reports submitted



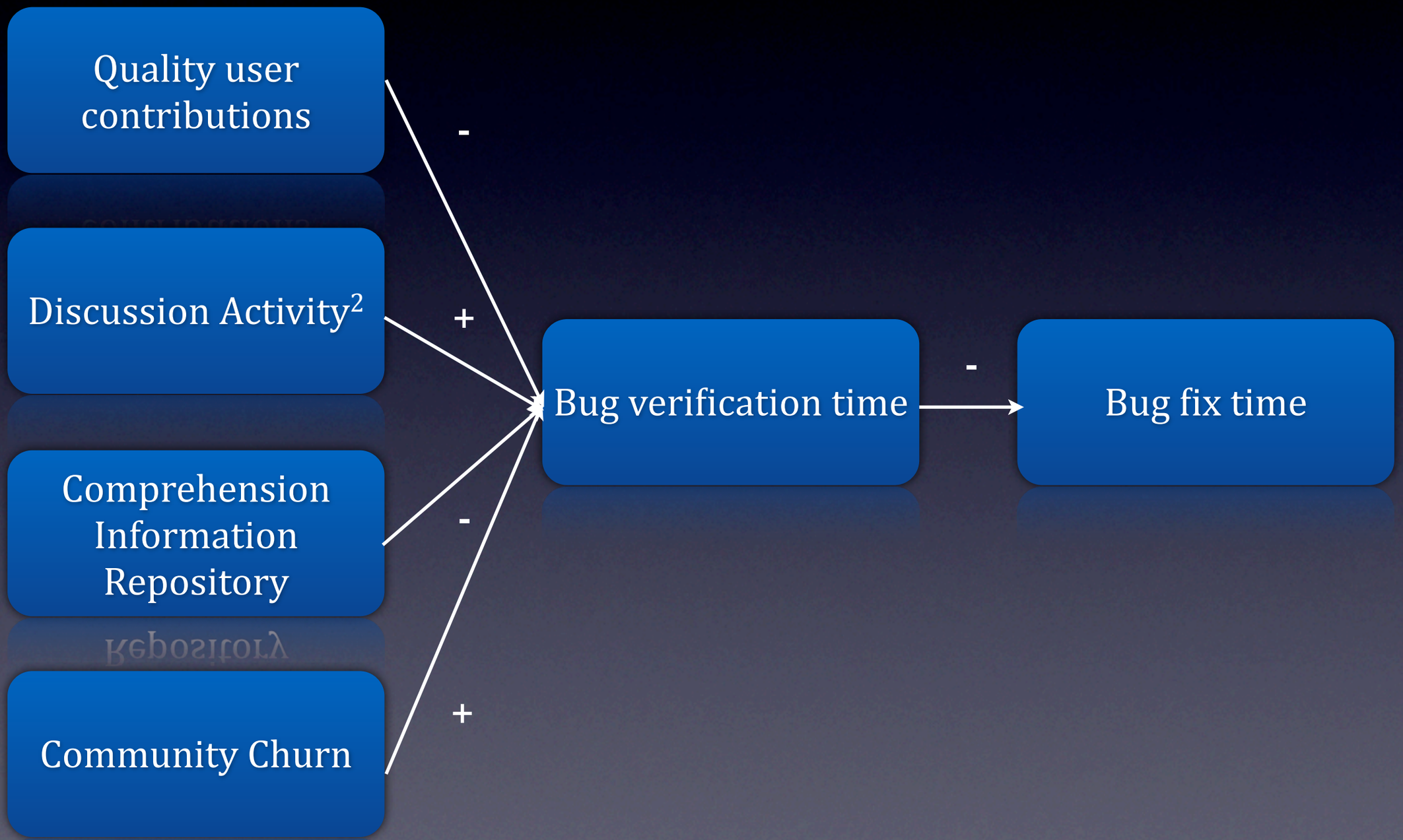
An open source community is also an

information repository



Community members build, discuss & update the information repository

Research Model



Measurements

- Quality of bug report ranges from 0 to 4:
  - Sum the presence of ‘steps to reproduce’, ‘stack trace’, ‘screenshot’ & ‘version information’
- Discussion activity: count the number of unique e-mail addresses for each software defect
- Understanding of information repository: average experience bug discussants marking bugs duplicate
- Churn rate community: $\frac{outflow}{size + inflow}$

Dependent variable: time to verify software defect	Hypotheses	Model 1 Coefficient	Std. Err.		Model 2 Coefficient	Std. Err.		Model 3 Coefficient	Std. Err.	
Firefox initial beta release		0.1213	[0.0412]	**	0.2128	[0.0414]	***	0.1354	[0.0402]	***
Firefox 1.0 release		0.4680	[0.0706]	***	0.4475	[0.0708]	***	0.3613	[0.0697]	***
Firefox 1.5 release		0.5218	[0.1060]	***	0.5137	[0.1064]	***	0.3680	[0.1047]	***
Firefox 2.0 release		0.7581	[0.1259]	***	0.7427	[0.1265]	***	0.4751	[0.1241]	***
Firefox 3.0 release		1.0576	[0.1442]	***	1.0371	[0.1448]	***	0.6849	[0.1422]	***
Number of community members		0.0002	[0.0000]	***	0.0002	[0.0000]	***	0.0002	[0.0000]	***
Community workload		-0.0001	[0.0000]	***	-0.0001	[0.0000]	***	-0.0001	[0.0000]	***
Number of words of initial report (log)		0.0637	[0.0077]	***	0.1953	[0.0081]	***	0.0263	[0.0076]	***
Number of words of title initial report		0.0190	[0.0016]	***	0.0217	[0.0016]	***	0.0133	[0.0015]	***
Software defect is labeled for milestone (dummy)		1.6471	[0.0201]	***	1.6238	[0.0201]	***	1.6294	[0.0199]	***
Software defect causes crash (dummy)		-0.0391	[0.0228]		0.0285	[0.0229]		0.0958	[0.0228]	***
Software defect has been flagged as regression (dummy)		-0.6107	[0.0356]	***	-0.6167	[0.0358]	***	-0.4475	[0.0349]	***
Community member experience (reports)		0.0010	[0.0000]	***	0.0009	[0.0000]	***	0.0009	[0.0000]	***
Developer is employed by Mozilla.org (dummy)		1.4628	[0.0506]	***	1.4195	[0.0500]	***	1.2262	[0.0497]	***
Developer workload (log)		0.4107	[0.0062]	***	0.3350	[0.0065]	***	0.3552	[0.0062]	***
Software defect dependencies		0.7556	[0.0329]	***	0.7265	[0.0327]	***	0.7083	[0.0325]	***
Software defect complexity		-0.0006	[0.0017]		-0.0005	[0.0018]		-0.0008	[0.0015]	
Number of comments (month)		-0.0730	[0.0019]	***	-0.0733	[0.0020]	***	0.0798	[0.0030]	***
Information Repository - Quality of user contribution	H1: -				-0.3889	[0.0077]	***			
Information Repository - Number of discussants								-1.0751	[0.0252]	***
Information Repository - Number of discussants (squared)	H2: +							0.0493	[0.0033]	***
Information Repository - Understanding by community members	H3: -									
Information Repository - Churn rate of community (%)	H4: +									
Year dummies included		Yes			Yes			Yes		
Severity dummies included		Yes			Yes			Yes		
Platform dummies included		Yes			Yes			Yes		
Priority dummies included		Yes			Yes			Yes		
Constant		2.4951	[0.0792]	***	2.2734	[0.0796]	***	4.6404	[0.0848]	***
Number of observations		396,307			396,307			396,307		
Number of submitted bug reports		317,041			317,041			317,041		
Degrees of freedom		59			60.00			61.00		
Shape parameter (p)		0.32			0.32			0.33		
Wald Chi2		58,113***			62,000***			74,825***		
Log-Likelihood		-724,452			-723,169			-716,515		
Log-Likelihood Improvement (compared to Model 1)					2,566.42***			15,873.54***		

* p<0.05, ** p<0.01, *** p<0.001

Robust standard errors, clustered by software defect

High quality contributions shorten time to verify

Quality user contribution → Less time needed to verify bug report

Understanding what's in the information repository shortens time to verify

Understanding of the information repository → Less time needed to verify bug report

Community churn reduces understanding of the information repository

Time to verify software defect

Multiplier time to verify software defect

80.0
60.0
40.0
20.0
0.0
-20.0
-40.0
-60.0

0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100

Information repository comprehension

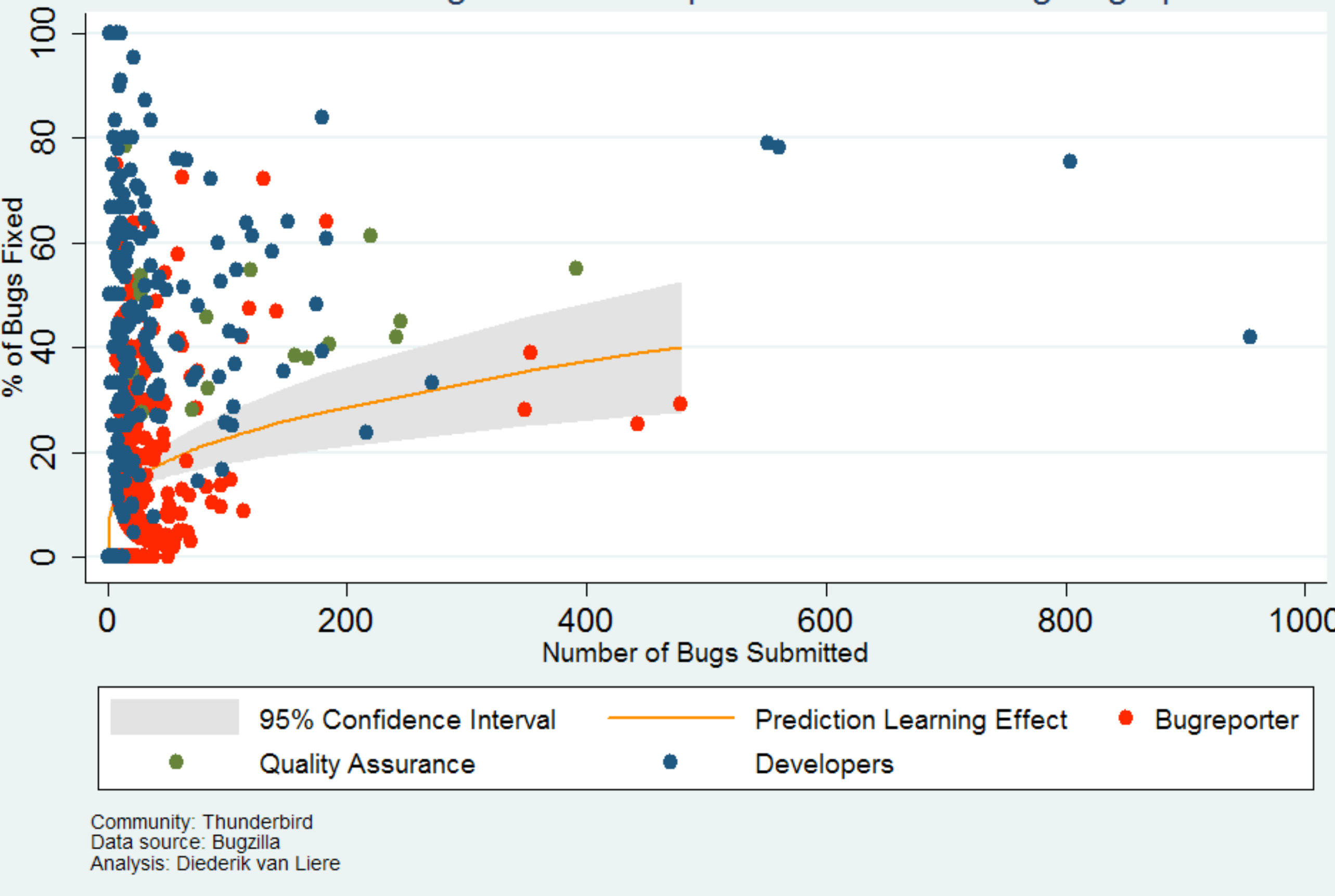
Number of discussants

0 15 30 45 60 75 90

Data Collection - Thunderbird

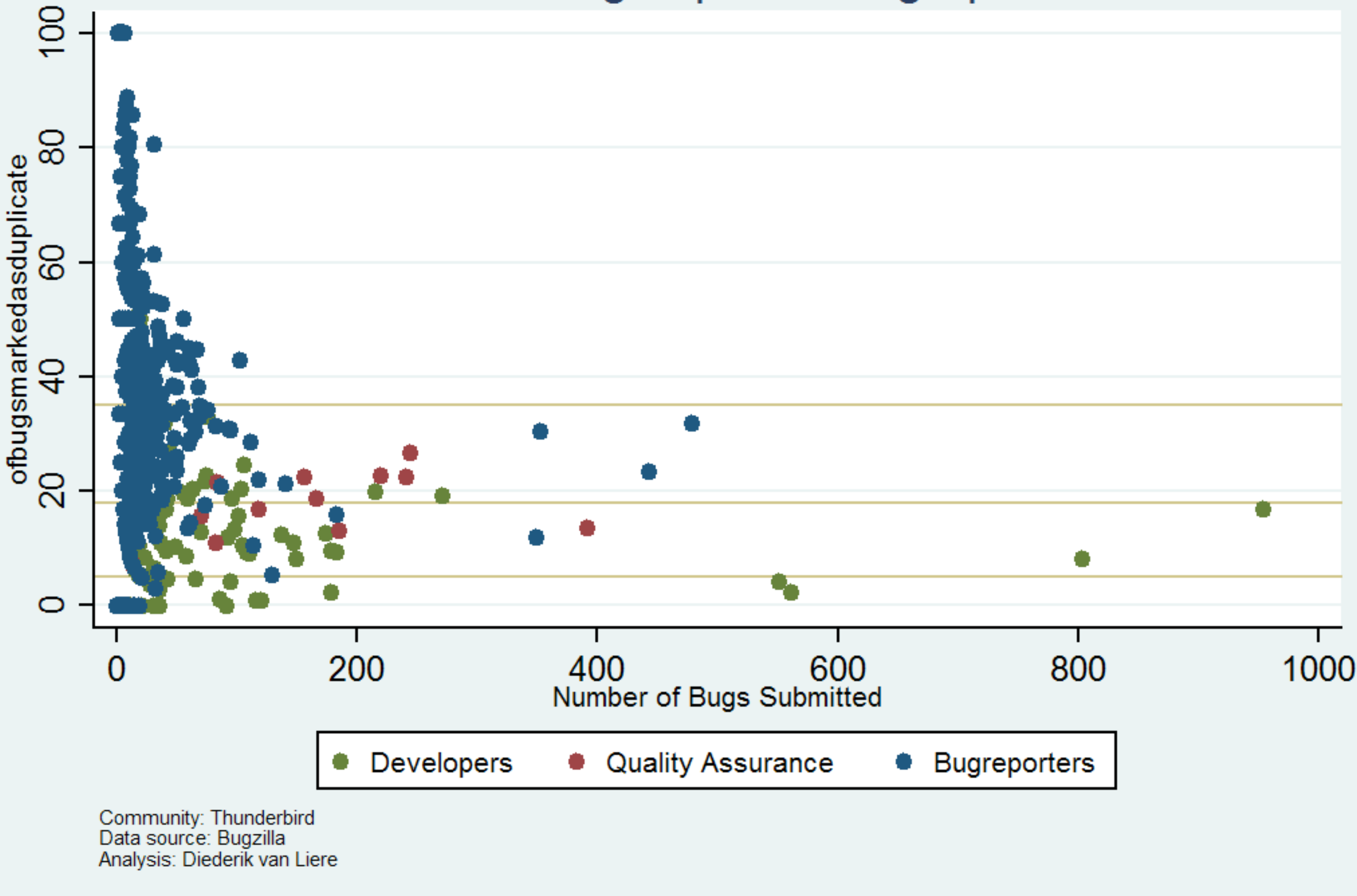
- Collected bug reports filed at bugzilla.mozilla.org with id 1 - 520.000 if product is MailNews / Thunderbird /, in total 51.421 bug reports
- This covers the period late 1998 to December 1st, 2009
- Three different roles:
  - Developer (if you submitted >4 patches through Bugzilla)
  - Quality Assurance (if you are listed as 'qa_contact')
  - Bugreporter (if you are neither a Developer or Quality Assurance but have submitted at least 1 bug report)
- Gender:
  - Using the program gender.c written by Michael Jörg, I determine gender by parsing the first name. Gender.c recognizes more than 40.000 names from 54 countries.

Success Rate Bug Fixes and Experience in Submitting Bugreports



Community: Thunderbird
Data source: Bugzilla
Analysis: Diederik van Liere

Who is writing duplicate bugreports?

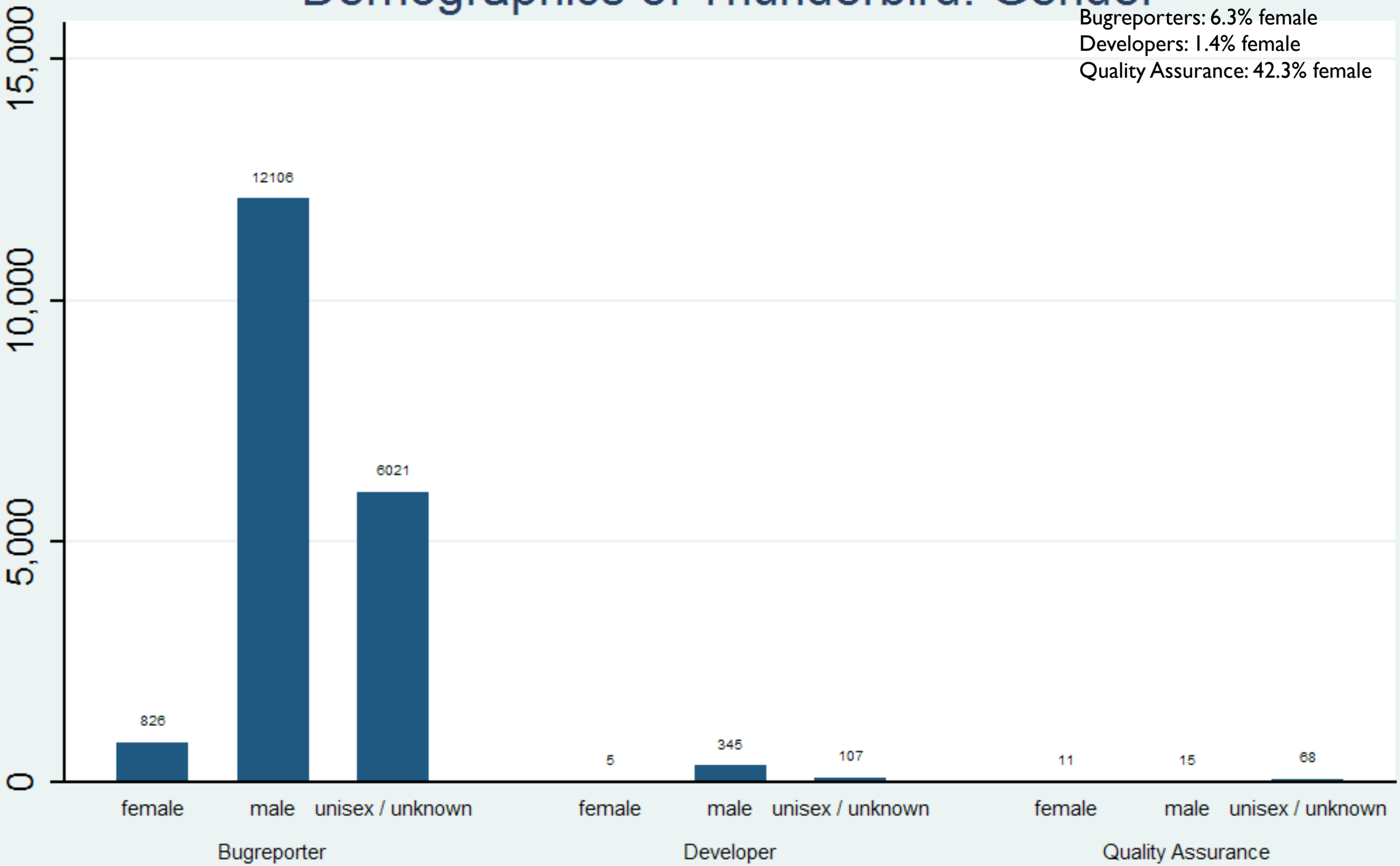


Community: Thunderbird
Data source: Bugzilla
Analysis: Diederik van Liere

‘Bleeding Edge’

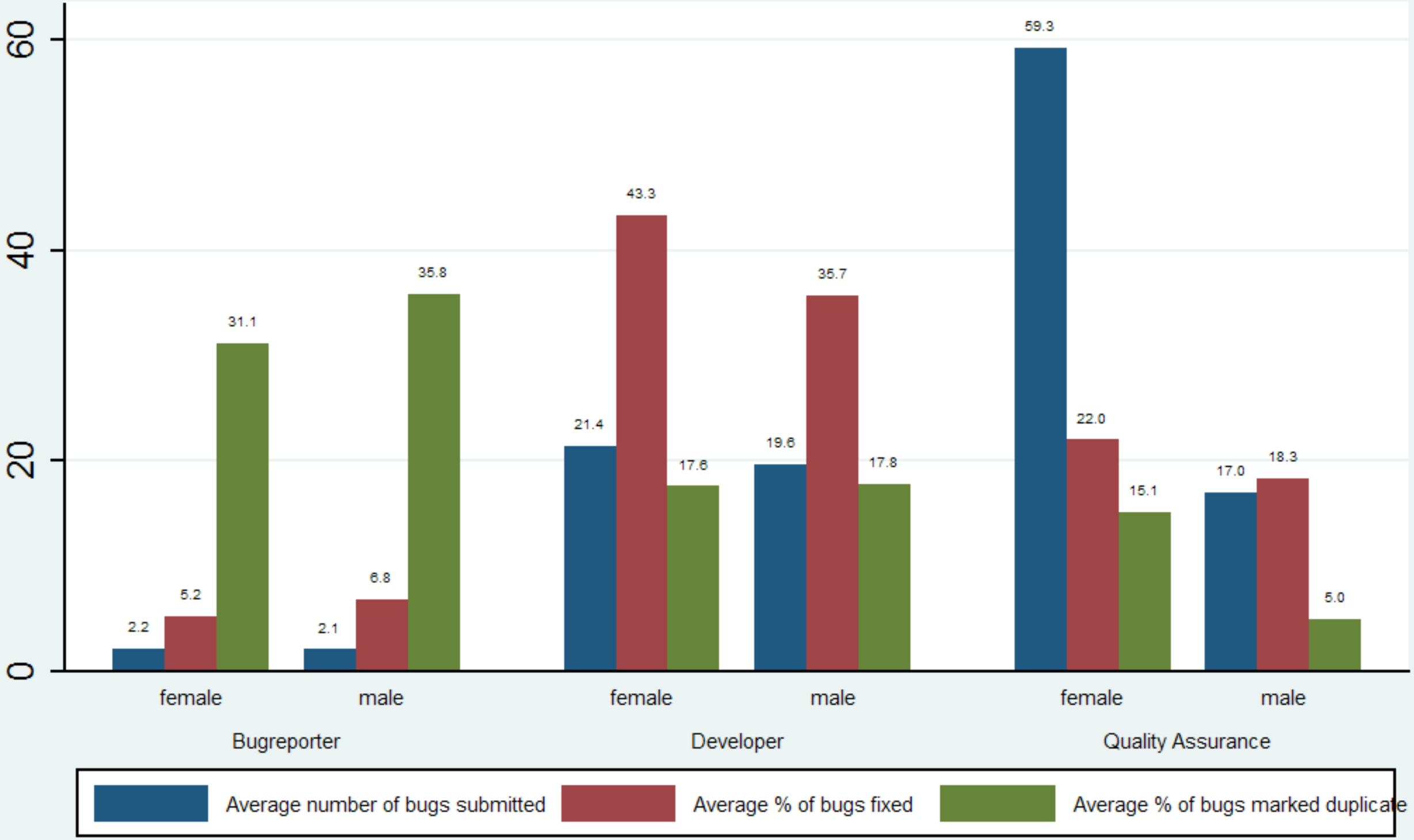
Demographics of Thunderbird: Gender

Bugreporters: 6.3% female
Developers: 1.4% female
Quality Assurance: 42.3% female



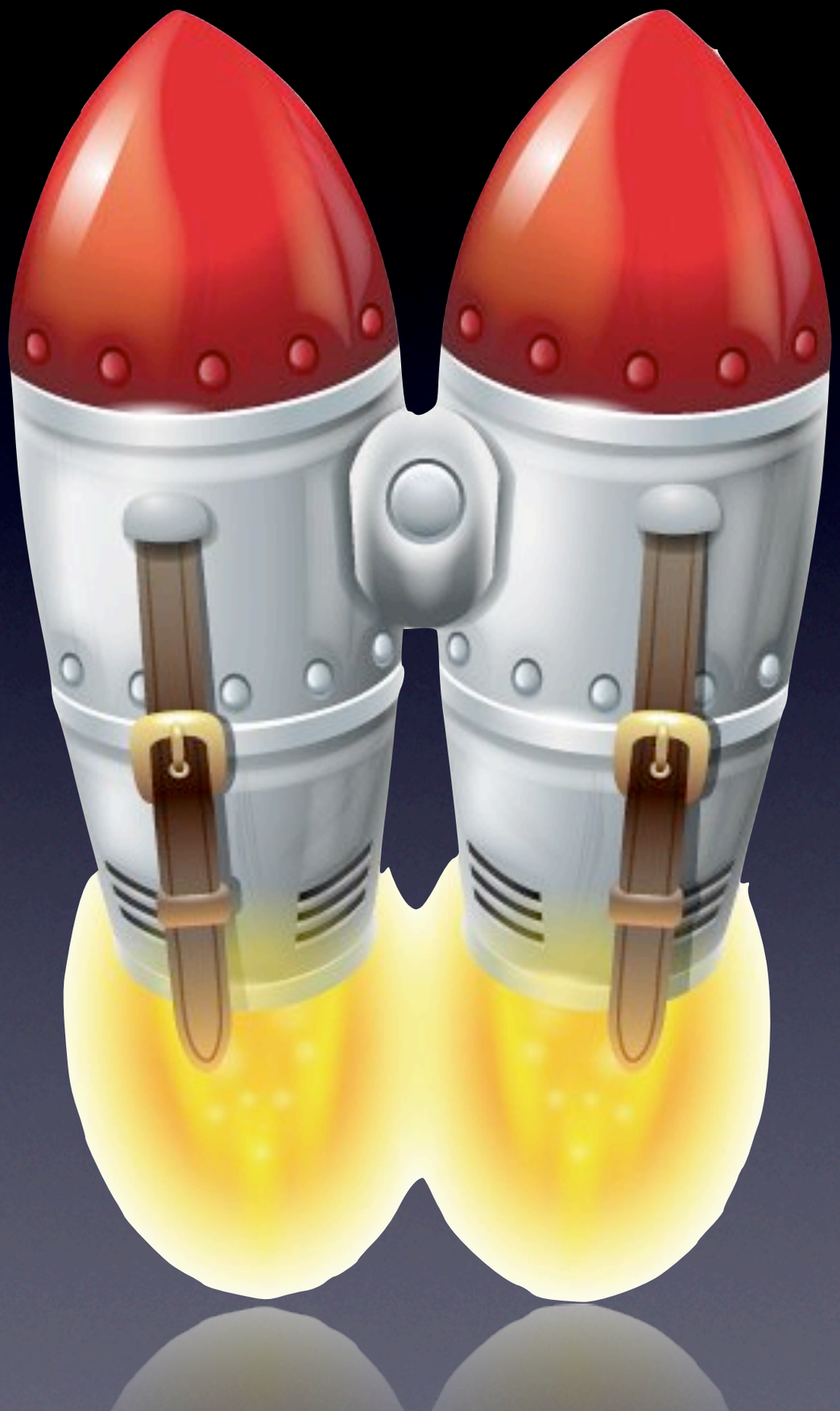
Community: Thunderbird
Source: Bugzilla
Analysis: Diederik van Liere

Do Women and Men Differ in Their Contributions?



Community: Thunderbird
Source: Bugzilla
Analysis: Diederik van Liere

SOLUNAR



Tools to improve bug fixing process:

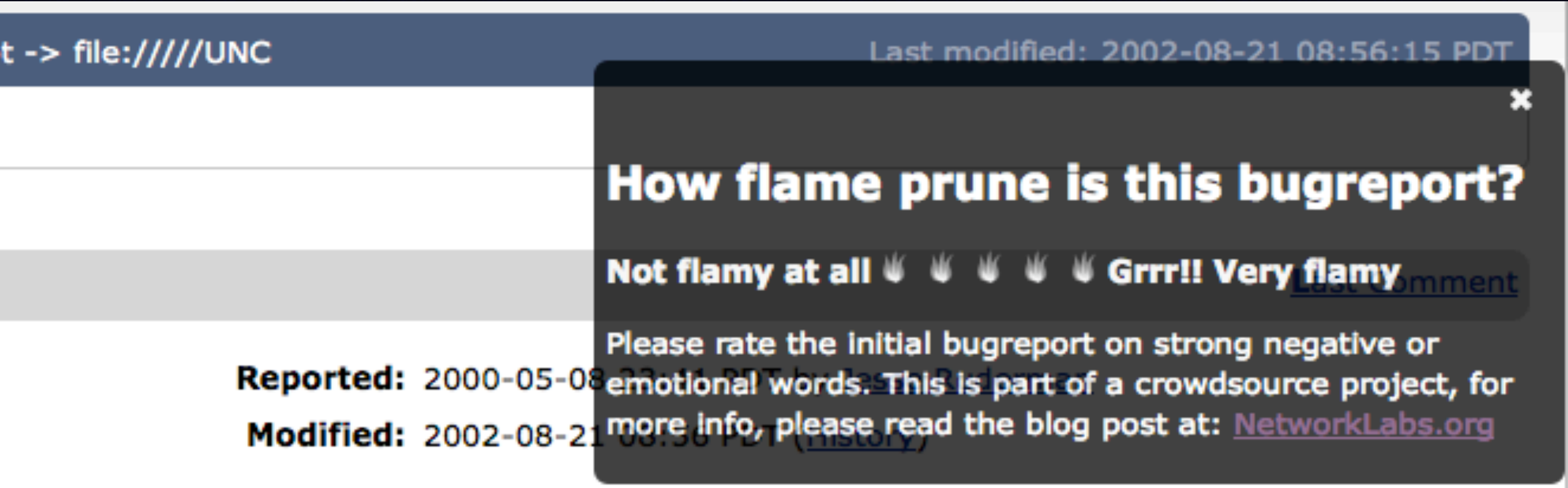
1. Predict Bugfix (Jetpack Add-on)
2. Crowdsource ‘flaming’ reports (Jetpack Add-on)
3. Crowdsource dev handles (network-labs.org/mozilla/)

available at:

www.network-labs.org

Predicting bug fixes

Crowdsourcing flamy bug reports



www.network-labs.org/mozilla/

Sometimes it's hard for us to track a particular Mozilla developer as the community of Mozilla developers continues to grow and developers change their email addresses or misspelled email addresses are entered in the Bugzilla / CVS / HQ databases.

So, if you are a Mozilla developer and changed your email address then please have a look in our database to see if we are aware of this change. Just enter your **current** email address that you use for your Mozilla development activities. If you are not yet in the database, please add your current email address.

If you want to find the current email address and you only an prior email address, just enter the prior email address and the current address will be returned.

We will only use your information to contact you for the purpose of Mozilla development activities, this will probably very rare. We will not share this information with anybody outside of the Mozilla organization.

We currently have **11111** handles (email addresses, nicknames, irc handles) in our database and **5642** matched old aliases and expired email addresses.

Help us with staying in touch with the Mozilla developers!

Search for handle: Search for Mozilla Developer

The search engine will only return exact matches to prevent people from harvesting email addresses.

Initial database development and data collection by Diederik van Liere.

This work by Diederik van Liere is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 Netherlands License.

This research is partially funded by The Netherlands Organisation for Scientific Research (NWO).

Conclusion

Conclusions

- Retention of community members is key
- Get community members through the learning curve asap
- Extend Bugzilla with prediction functionality to assist in allocating resources
- Funnel 1st time bug reporters to ‘user assistance area’

www.network-labs.org

diederik.vanliere@rotman.utoronto.ca

drdee_is_wired

References

- http://www.flickr.com/photos/eriknielsen/2233837359/sizes/l/
- http://www.flickr.com/photos/fastjack/282707058/sizes/o/
- http://www.flickr.com/photos/59303791@N00/3398708956/sizes/o/
- http://www.flickr.com/photos/_ilkin_/3890567460/sizes/o/
- http://www.flickr.com/photos/wondertubs/2152108411/
- http://www.flickr.com/photos/asurroca/136223817/sizes/o/
- http://www.flickr.com/photos/thinmints137/452928157/sizes/o/

References

- http://www.flickr.com/photos/hanneorla/4032004209/sizes/o/
- http://www.flickr.com/photos/97629199@N00/3254201501/in/pool-bookshelf
- http://www.flickr.com/photos/gadl/740994053/
- http://www.flickr.com/photos/blu_blue/262096844/sizes/o/
- http://www.flickr.com/photos/kurisuuu/3227744533/sizes/o/
- http://www.flickr.com/photos/astrid/2330867426/sizes/o/
- http://www.flickr.com/photos/sweetone/2666516868/sizes/o/
- http://www.flickr.com/photos/caveman_92223/2982595969/sizes/o/
- http://www.flickr.com/photos/mrtea/2114891329/sizes/o/