ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Categorizing Bugs with Social Networks

A Case Study on Four OSS Communities

Ingo Scholtes
Chair of Systems Design
ETH Zürich

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Chair of Systems Design www.sg.ethz.ch

Handling bug reports is laborious!

- **Most contributed bug reports are …**
  - … actually not bugs …
  - … not reproducible …
  - … duplicates of known bugs
- **Mozilla Firefox**
  - Community has processed 64,000 bug reports
  - 50,000 (~ 79 %) of those were faulty
- **Tool support appreciated!**
  - Automated priorization of valid reports
  - Can decrease response and fix time
  - Can increase productivity

What can we learn from the social layer?

Social awareness in bug tracking tools

Categorizing Bugs with Social Networks: A Case Study on Four Open Source Software Communities

Marcelo Serrano Zanetti, Ingo Scholtes, Claudio Juan Tessone and Frank Schweitzer
Chair of Systems Design – ETH Zurich
Zurich, Switzerland
http://www.sg.ethz.ch/

Abstract—Efficient bug triaging procedures are an important precondition for successful collaborative software engineering projects. Triaging bugs can become a laborious task particularly in open source software (OSS) projects with a large base of comparably inexperienced part-time contributors. In this paper, we propose an efficient and practical method to identify *valid* bug reports which a) refer to an actual software bug, b) are not duplicates and c) contain enough information to be processed right away. Our classification is based on nine measures to quantify the social embeddedness of bug reporters in the collaboration network. We demonstrate its applicability in a case study, using a comprehensive data set of more than 700,000 bug reports obtained from the BUGZILLA installation of four major OSS communities, for a period of more than ten years. For those projects that exhibit the lowest fraction of valid bug reports, we find that the bug reporters' position in the collaboration network is a strong indicator for the quality of bug reports. Based on

calls for (semi-)automated techniques that assist bug handling communities in the triaging and prioritization of bug reports. The provision of methods which are able to automatically identify *valid* bug reports with high precision can have huge implications for practitioners of distributed software engineering: Being able to filter, assign and prioritize those bug reports that likely result in a bug fix can significantly improve the responsiveness of support communities. Furthermore, a temporary deferral of those bug reports that are likely to be duplicates, invalid or incomplete to a moderation queue can considerably alleviate the effort required for bug triaging. It can also be used to automatically enforce the adherence to community guidelines, e.g. by automatically asking original reporters to reconfirm that reported bugs are neither duplicates

Marcelo S. Zanetti, Ingo Scholtes, Claudio J. Tessone and Frank Schweitzer: **Categorizing Bugs with Social Networks: A Case Study on Four Open Source Software Communities**, *In Proceedings of the 35th International Conference on Software Engineering (ICSE 2013), SEIP track, San Francisco, CA, USA, 2013, http://arxiv.org/abs/1302.6764*

ETH Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Chair of Systems Design www.sg.ethz.ch

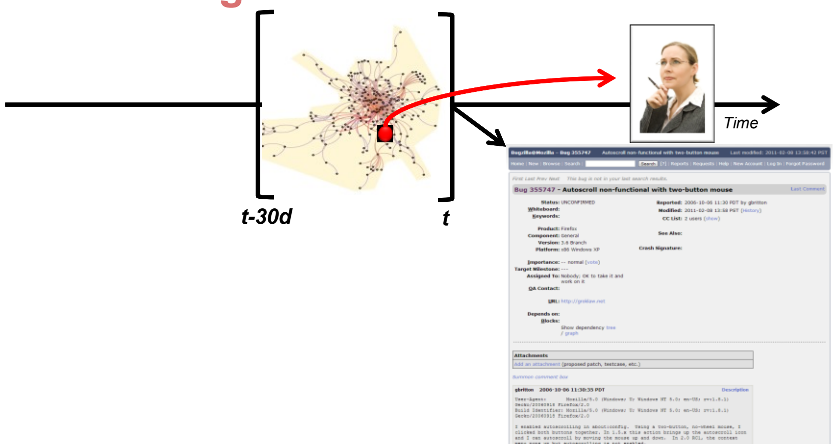
Open Source Communities

- **Valid**
  - *FIXED*
  - *WONTFIX*
- **Faulty**
  - *DUPLICATE*
  - *INVALID*
  - *INCOMPLETE/NEEDINFO*

	FIREFOX	THUNDERBIRD	ECLIPSE	NETBEANS	Total
Start date	April 2002	January 2000	October 2001	January 1999	–
Total bug reports	112,968	35,388	356,415	210,921	715,692
Change events	1,068,070	313,957	2,594,385	1,875,878	5,852,290
Changes / report	9.45	8.87	7.28	8.89	8.18
Resolved bugs (resolved/total)	64,088 (0.57)	21,644 (0.61)	158,957 (0.45)	42,851 (0.19)	287,540 (0.40)
FIX (FIX / resolved)	10,856 (0.17)	4,508 (0.21)	103,453 (0.65)	21,442 (0.50)	140,259 (0.49)
DUP (DUP / resolved)	24,263 (0.38)	10,336 (0.48)	28,227 (0.18)	9,328 (0.22)	72,154 (0.25)
INV (INV /resolved)	11,785 (0.18)	2,829 (0.13)	12,601 (0.08)	4,082 (0.10)	31,297 (0.11)
WOF (WOF / resolved)	2,708 (0.04)	581 (0.03)	14,676 (0.09)	5,515 (0.13)	23,480 (0.08)
INC (INC / resolved)	14,476 (0.23)	3,390 (0.16)	-	2484 (0.06)	20,350 (0.07)

Marcelo S. Zanetti, Ingo Scholtes, Claudio J. Tessone and Frank Schweitzer: **Categorizing Bugs with Social Networks: A Case Study on Four Open Source Software Communities**, *In Proceedings of the 35th International Conference on Software Engineering (ICSE 2013), SEIP track, San Francisco, CA, USA, 2013, http://arxiv.org/abs/1302.6764*

Evolving collaboration networks



H3: The position of a bug reporter in the monthly collaboration network preceding the time of the report is indicative for the eventual outcome of the bug handling process.

Hypothesis

Hypothesis	Comparison of Distrib.	FIREFOX	THUNDERBIRD	ECLIPSE	NETBEANS
H1	$FIX_1 \sim FIX_2$	$<, p = 0.0026$, (*) (5847, 6140)	$>, p = 0.0351$, (*) (2139, 2377)	$\neq, p = 0.1453$ (66208, 69026)	$\neq, p = 0.6435$ (13930, 14668)
H2	$DUP_1 \sim DUP_2$	$>, p = 0.0349$, (*) (6799, 8697)	$>, p < 2.22e-16$, (*) (973, 3027)	$>, p < 2.22e-16$, (*) (17600, 22215)	$>, p < 2.22e-16$, (*) (3984, 5470)
H2	$INV_1 \sim INV_2$	$\neq, p = 0.7268$ (1321, 1394)	$>, p = 0.0449$, (*) (242, 297)	$\neq, p = 0.8489$ (5313, 5958)	$\neq, p = 0.1266$ (1906, 2066)
H3	$FIX_1 \sim WOF_1$	$>, p = 1.81$[illegible], (*) (5847, 1022)	$>, p =$ [illegible] (2139, 1[illegible])	$<, p < 2.22e-16$, (*) (66208, 7769)	$>, p < 2.22e-16$, (*) (13930, 2847)
H3	$FIX_1 \sim DUP_1$	$>, p < 2.22e-16$, (*) (5847, 6799)	$>, p < 2.22e-16$, (*) (2139, 973)	$<, p < 2.22e-16$, (*) (66208, 17600)	$>, p < 2.22e-16$, (*) (13930, 3984)
H3	$FIX_1 \sim INV_1$	$>, p < 2.22e-16$, (*) (5847, 1321)	$>, p = 4.93e-10$, (*) (2139, 242)	$<, p < 2.22e-16$, (*) (66208, 5313)	$>, p < 2.22e-16$, (*) (13930, 1906)
H3	$FIX_1 \sim INC_1$	$>, p < 2.22e-16$, (*) (5847, 587)	$>, p < 2.22e-16$, (*) (2139, 159)	(-)(-) (66208, 0)	$>, p < 2.22e-16$, (*) (13930, 661)

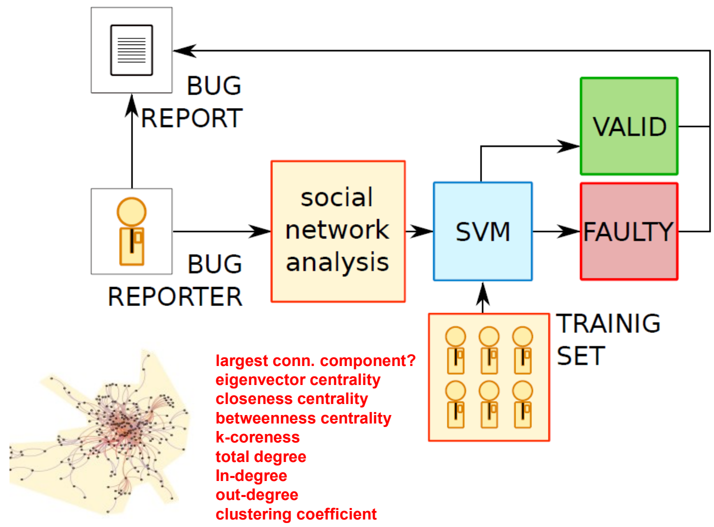
YES, it is!

H3: The position of a bug reporter in the monthly collaboration network preceding the time of the report is indicative for the eventual outcome of the bug handling process.

ETH Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Chair of Systems Design www.sg.ethz.ch

Support Vector Machine



ETH Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Chair of Systems Design www.sg.ethz.ch

Support Vector Machine

PRECISION (p), RECALL (r) AND F-SCORE OF FILTERING VALID BUG REPORTS BASED ONLY ON MEASURES OF SOCIAL EMBEDDEDNESS.

	FIREFOX	THUNDERBIRD	ECLIPSE	NETBEANS
Valid	21.0%	23.3%	74.3%	62.4%
p (LCC)	44.1%	62.1%	76.3%	71.9%
r (LCC)	50.9%	44.5%	62.6%	62.4%
F (LCC)	0.47	0.52	0.69	0.67
p (evcent)	60.4%	68.6%	76.3%	76.7%
r (evcent)	30.5%	5.4%	62.6%	38.8%
F (evcent)	0.41	0.10	0.69	0.52
p (SVM)	82.5%	90.3%	88.7%	78.9%
r (SVM)	44.5%	38.9%	91.0%	87.0%
F (SVM)	0.58	0.54	0.89	0.83

Social networks can be used to automatically categorize bug reports

Precision and recall higher than for competing methods

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Chair of Systems Design www.sg.ethz.ch

Thank you!

Marcelo S. Zanetti, Ingo Scholtes, Claudio J. Tessone and Frank Schweitzer: **Categorizing Bugs with Social Networks: A Case Study on Four Open Source Software Communities**, *In Proceedings of the 35th International Conference on Software Engineering (ICSE 2013), SEIP track, San Francisco, CA, USA, 2013,* http://arxiv.org/abs/1302.6764

Marcelo S. Zanetti, Ingo Scholtes, Claudio J. Tessone and Frank Schweitzer: **The Rise and Fall of a Central Contributor: Centralization and Performance in the Gentoo Community**, *In Proceedings of the 6th International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE) held at ICSE 2013, San Francisco, CA, USA, 2013,* http://arxiv.org/abs/1302.7191

Marcelo S. Zanetti, Emre Sarigol, Ingo Scholtes, Claudio J. Tessone, Frank Schweitzer: **A Quantitative Study of Social Organization in Open Source Software Communities,** *In Proceedings of Imperial College Computing Student Workshop (ICCSW), Open Access Series in Informatics, pp. 116-122, London, UK, September 27-28, 2012,* http://arxiv.org/abs/1208.4289

More information on our work

http://www.sg.ethz.ch/
http://www.sg.ethz.ch/research/topics/social-se/

http://www.ingoscholtes.net