

Code Review

Investigating Code Review Quality: Do People and Participation Matter?

Oleksii Kononenko, Olga Baysal, Latifa Guerrouj, Yaxin Cao,
Michael W. Godfrey

Presented by: Zhaoyi Ge

- ▶ A standard in modern software engineering workflow
- ▶ Goals:
 - ▶ Ensure maintainability
 - ▶ **Finding defects**



Code Review

- ▶ A standard in modern software engineering workflow
- ▶ Goals:
 - ▶ Ensure maintainability
 - ▶ **Finding defects**
- ▶ Factors:
 - ▶ Personal factors
 - ▶ Participation
- ▶ Research Questions
 - ▶ Do code reviewers miss many bugs?
 - ▶ Do personal factors affect the quality of code reviews?
 - ▶ Does participation in code review influence its quality?



Mozilla

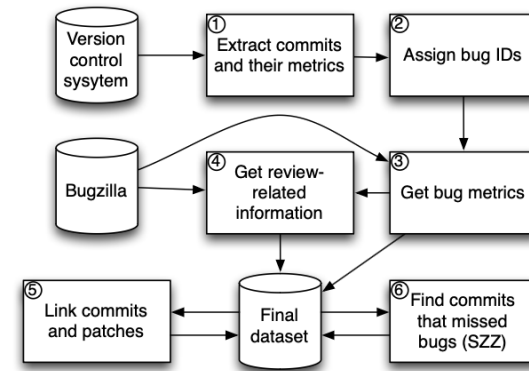
- ▶ Core, Firefox, and Firefox for Android



Mozilla

- ▶ Core, Firefox, and Firefox for Android
- ▶ Two-tiered code review process
 - ▶ *Review* (162, 25, 11)
 - ▶ *Super-review* (30)
- ▶ Bugzilla

Data Extraction



Explanatory Factors

- ▶ Technical
 - ▶ LOC
 - ▶ Module
 - ▶ Super-review

Explanatory Factors

- ▶ Technical
 - ▶ LOC
 - ▶ Module
 - ▶ Super-review
- ▶ Personal
 - ▶ Reviewer experience
 - ▶ Writer experience

Explanatory Factors

- ▶ Technical
 - ▶ LOC
 - ▶ Module
 - ▶ Super-review
- ▶ Personal
 - ▶ Reviewer experience
 - ▶ Writer experience
- ▶ Participation
 - ▶ Number of comments
 - ▶ Number of developers on CC

Model Construction and Analysis

- ▶ Multiple Linear Regression (MLR):
 $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$
- ▶ y : Code review quality (Buggy or not)
- ▶ x_i : Explanatory Factors

Model Construction and Analysis

- ▶ Multiple Linear Regression (MLR):
 $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$
- ▶ y : Code review quality (Buggy or not)
- ▶ x_i : Explanatory Factors
- ▶ Transformations:
 - ▶ Outliers
 - ▶ Categorical variables

Model Construction and Analysis

- ▶ Multiple Linear Regression (MLR):
 $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$
- ▶ y : Code review quality (Buggy or not)
- ▶ x_i : Explanatory Factors
- ▶ Transformations:
 - ▶ Outliers
 - ▶ Categorical variables
- ▶ Collinearity
 - ▶ Use VIF Score

Result

RQ1: Do code reviewers miss many bugs?

System	# Reviews	# Buggy Reviews	% Buggy Reviews
Mozilla-all	28,127	15,188	54.0 %
Core	18,759	10,184	54.3 %
Firefox	2,668	1,447	54.2 %
Firefox4Android	2,160	1,210	56.0 %

Result

RQ2: Do personal factors affect the quality of code reviews?

(a) Technical and personal factors.

	Mozilla	Core	Firefox	FF4A
Adjusted R^2	0.128	0.123	0.173	0.138
Size (LOC)	0.102***	0.098 ***	0.108***	0.115***
Chunks	†	†	†	†
Number of files	0.058***	0.059 ***	0.109***	0.062*
Module	*	n/a	n/a	n/a
Priority	*	*	†	.
Severity	†	†	.	†
Super review	-0.139**	-0.177***	.	n/a
Review queue	0.017***	0.0204***	0.038**	0.045**
Reviewer exp.	-0.013***	-0.012***	-0.029***	-0.041***
Reviewer exp. (mod.)	†	†	†	0.018*
Writer exp.	.	-0.004*	†	†
Writer exp. (module)	†	†	†	.
# prev patches	†	†	†	-0.045***
# writer patches	-0.012***	.	.	†

†Disregarded during VIF analysis (VIF coefficient > 5).

* "It's complicated": for categorical variables see explanation of the results in Section IV.

FF4A = Firefox for Android.

- Review Queue
- Reviewer Exp.

Result

RQ3: Does participation in code review influence its quality?

(b) Technical and participation metrics.

	Mozilla	Core	Firefox	FF4A
Adjusted R^2	0.134	0.128	0.173	0.147
Size (LOC)	0.105***	0.103***	0.105***	0.117***
Chunks	†	†	†	†
Number of files	0.060***	0.059***	0.090***	0.067***
Module	*	n/a	n/a	n/a
Priority	†	*	†	*
Severity	*	†	*	†
Super review	-0.124***	-0.160***	†	n/a
# of devs on CC	0.053***	0.056***	†	0.049*
# comments	†	†	†	†
# commenting devs	-0.124***	-0.102***	-0.075***	-0.176***
# comments/ # dev	-0.039***	-0.029**	†	†
# reviewer comments	0.010**	†	0.026*	†
# writer comments	.	.	.	-0.047**

†Disregarded during stepwise selection.

Statistical significance: '***' $p < 0.001$; '**' $p < 0.01$; '*' $p < 0.05$; '.' $p \geq 0.05$.

- # Devs on CC
- # Reviewer Comment

Positive Points

- Comprehensive empirical methodology
- Insights for everyone

Negative Points

- ▶ Multiple Linear Regression
- ▶ Lack of actionable items
- ▶ Is catching bug a good indicator of the quality of code reviews?



Rating

3.5/5. Very approachable paper for outsiders.



Future Work

Investigations that aims to capture the social context, communication patterns, and interpersonal dynamics that influence review effectiveness.



Discussions

- ▶ What are your explanations for the two surprising factors? Do you agree with the author or not?
- ▶ What are your takeaways from this paper? Would you do anything differently when doing code reviews after this paper?
- ▶ What are some of the factors that go beyond technical, people and participation?

