How Does Contributors Involvement Influence Open Source Systems

Reem Alfayez, Pooyan Behnamghader, Kamonphop Srisopha, Barry Boehm

Open Source Systems

- A company open sourcing a system.
- A single developer or a group of developer.

Core vs. Peripheral Contributors

Core	Peripheral
• Contribute more to the system. • Make major decisions regrading the system.	• Contribute less to the system. • Have less power when making decisions.

Motivation

- Peripheral contributors represent a large part of the OSS community (48.98%)
- They:
  - –Fix bugs (30.20%).
  - – Fix typos and grammar issues (28.64%).
  - – Add new features (18.75%).
  - –Refactor code (8.85%).

Research Questions

- **RQ1:** Do core developers change LOC more than peripheral developers?
- **RQ2:** Do core developers increase LOC more than peripheral developers?
- **RQ3:** Do core developers change TD more than peripheral developers?
- **RQ4:** Do peripheral developers introduce TD more than core developers?

What is Technical Debt

- In 1992, Ward Cunningham described technical debt as writing immature or "not quite right" code in order to ship a new product to market faster.
- Technical Debt consists of:
  - **Principle: measures the cost or effort for eliminating technical debt .**
  - **Interest: measures the extra cost or effort over some period of time incurred for NOT eliminating the technical debt.**

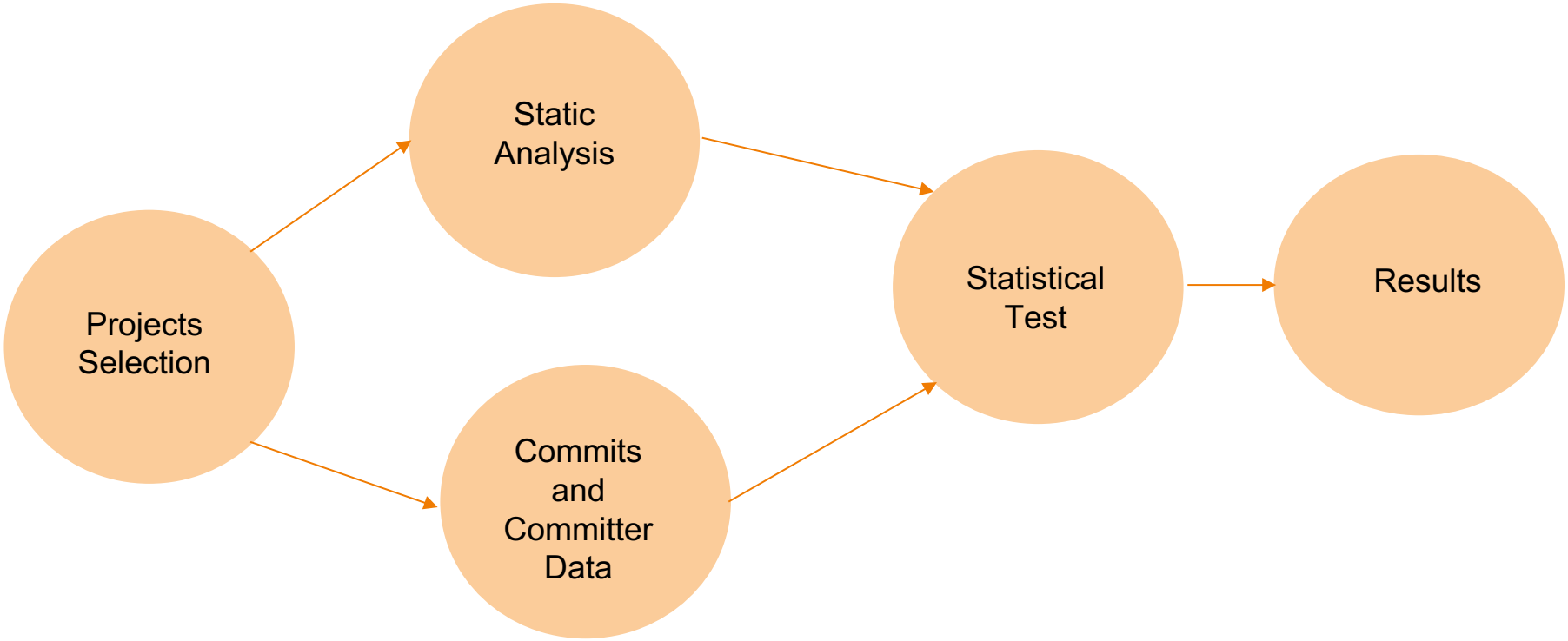
Why Do We Take Technical Debt

- Release faster.
- Decrease current release cost.
- Gather more information.
- Delay decisions.

Technical Debt Consequences

- Increased time to delivery.
- Increased number of defects.
- Raising maintainability cost.
- Decreased customer satisfaction.

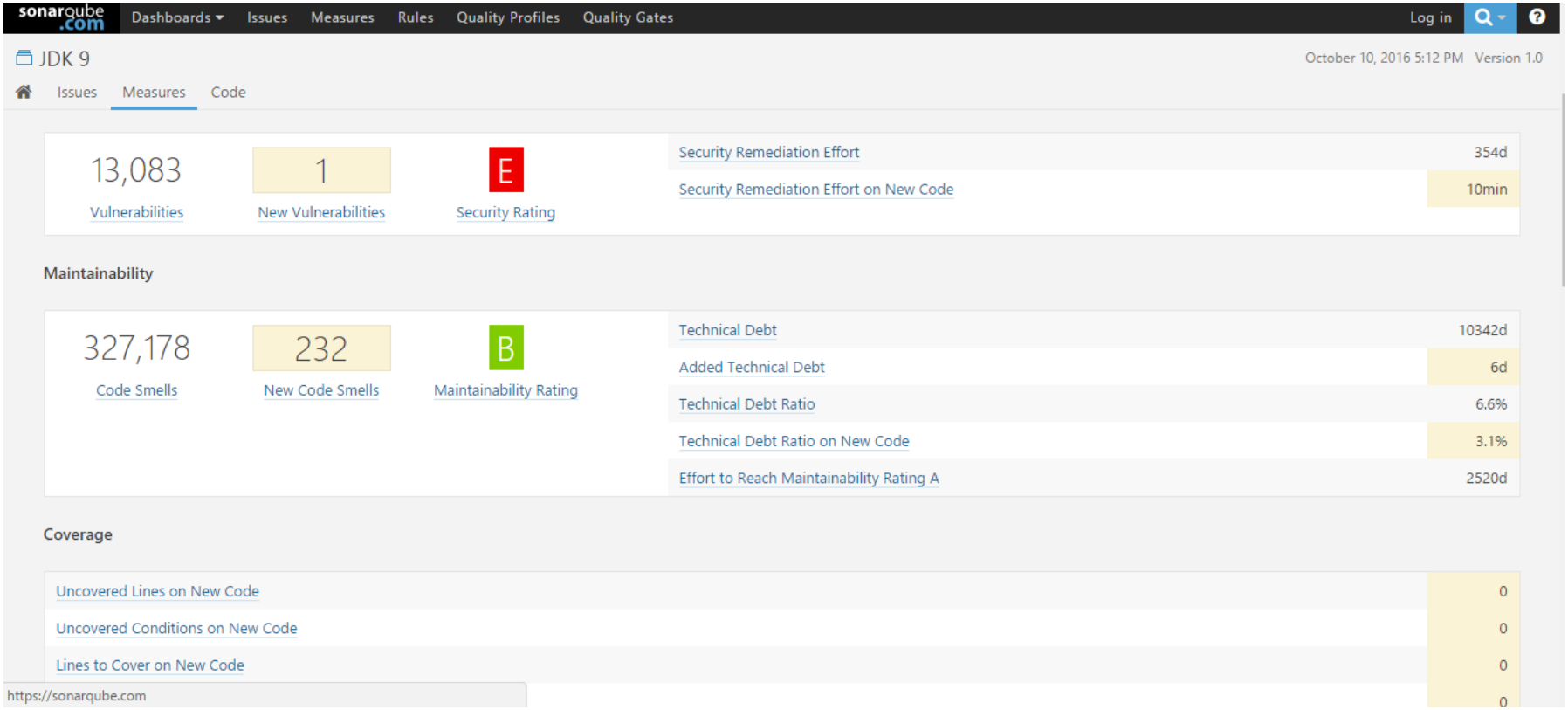
Methodology



Data Selection

- 38 Apache Software Foundation Systems.
- Selection Criteria:
  - There are more than one official releases.
  - The programming language is Java.
  - The Git repository is active and has an update in 2017.
  - The system has less than 3,000 commits.
  - The system has one module that contains most of the source code.

Technical Debt Calculation



Technical Debt Calculation

- **Debt(in man days) =**

cost_to_fix_duplications

+cost_to_fix_violations

+ cost_to_comment_public_API

+cost_to_fix_uncovered_complexity

+cost_to_bring_complexity_below_threshold

Contributors Categorization Perspectives

- Social perspective: study developers communication and collaboration.
- Technical perspective: study system artifacts.

Contributors Categorization Approaches

- Count-based techniques .
- The most used metrics are:
  - Commit count.
  - LOC count.
  - Bug tracking messages count.
  - Mail count.

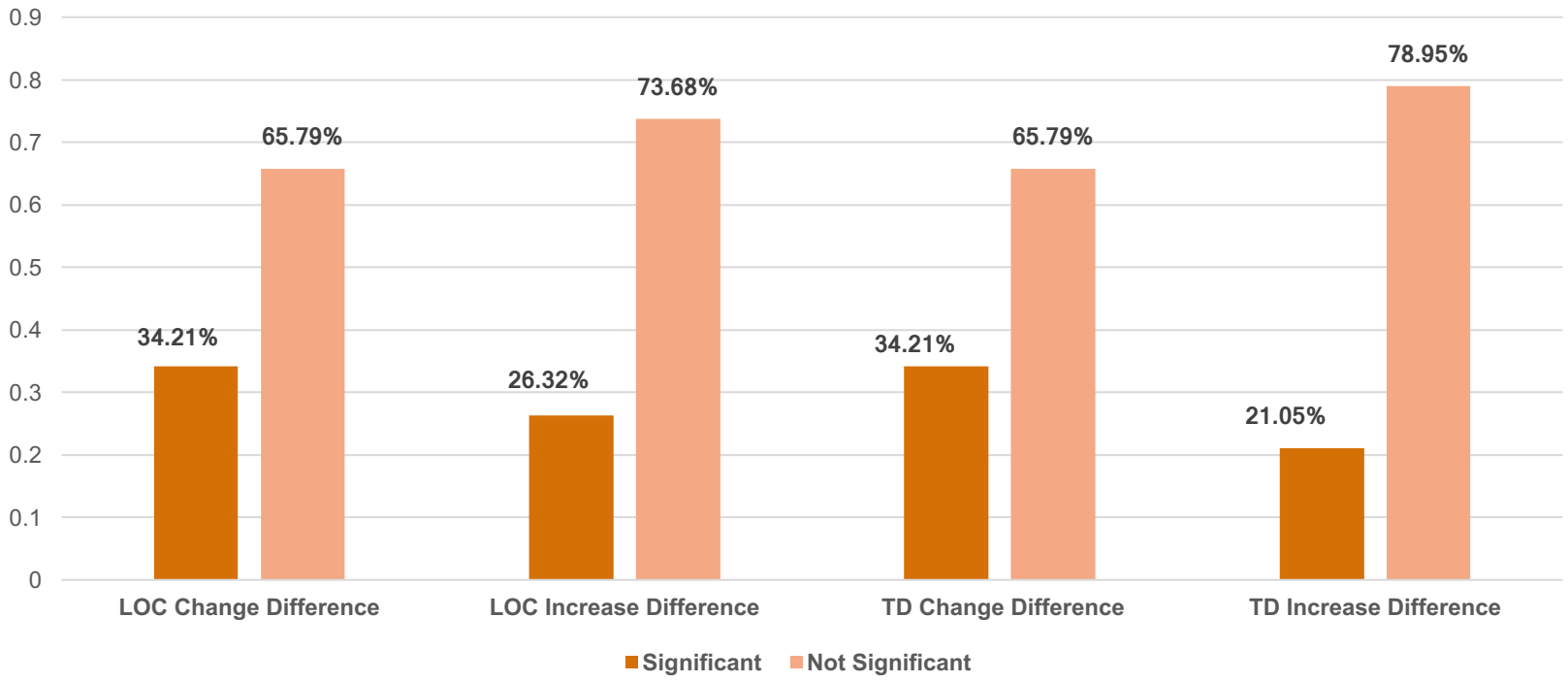
Statistical Test

- K-means clustering.
- We clustered the developers in each system based on the total number of commits they authored.

Statistical Test

- Fisher's Exact Test.
  - –Determines whether the differences in the values between the two categories are statically significant or random.
  - –Significance level is 0.05.
  - –Confidence level is 95%.

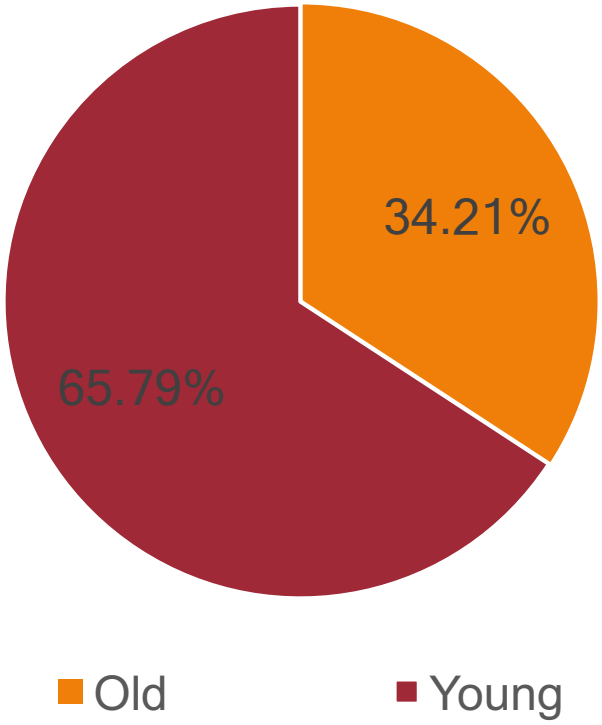
Results


0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
34.21%
65.79%
26.32%
73.68%
34.21%
65.79%
21.05%
78.95%
LOC Change Difference
LOC Increase Difference
TD Change Difference
TD Increase Difference
Significant
Not Significant

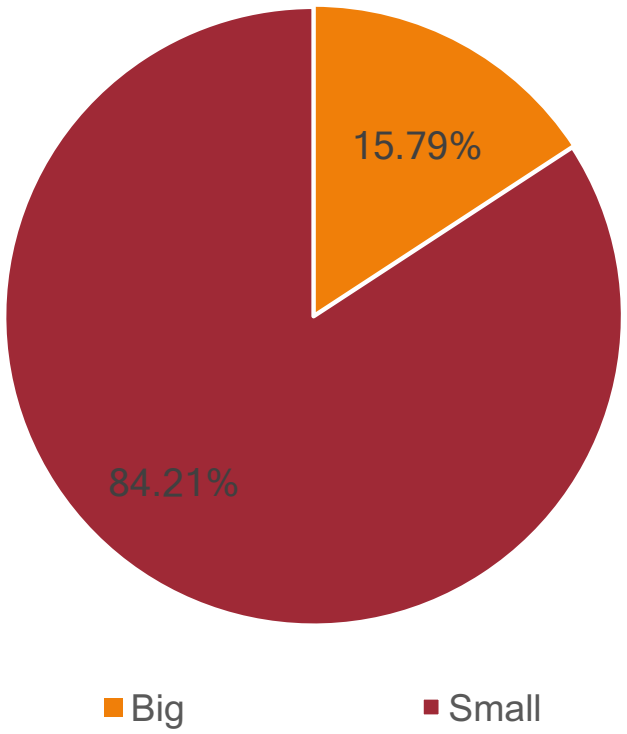
Results

- System age.
- System size.
- System number of releases.

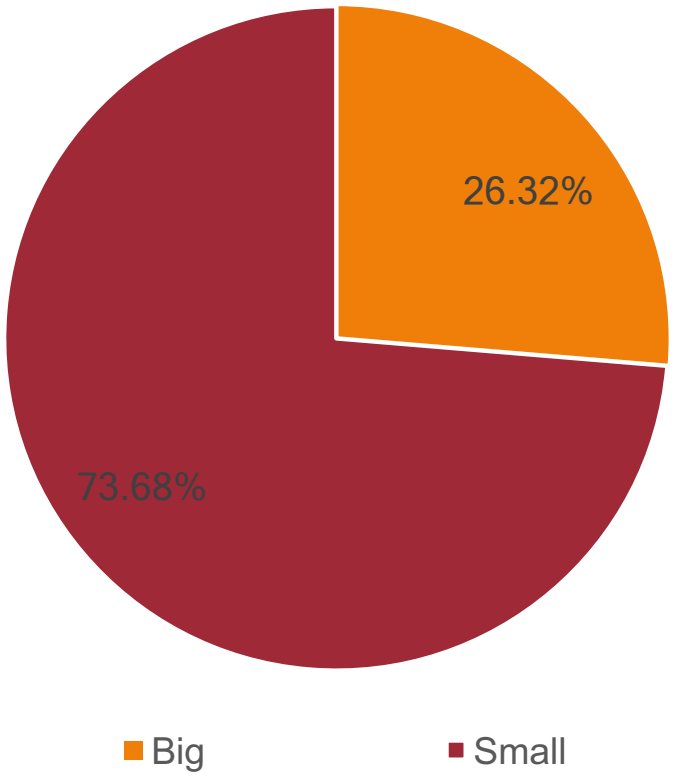
Systems Break Down based on Age



Systems Break Down based on Size



Systems Break Down based on Number of Releases



Results

	AGE Old vs. Young	Releases Large vs. Small	Size Big vs. Small
LOC change	.148	.594	.076
LOC increase	.262	.278	.642
TD change	.473	.594	.644
TD increase	.221	**.023**	1.0

Implications

- Peripheral contributors:
  - Overcome their fear of joining new OSS systems.
  - Contribute regardless of their inability for long time commitment.
  - Increase the number of new volunteers and insure a constant influx of new volunteers .

Implications

- CS educators:
  - The learners of today are the practitioners of tomorrow.
  - Win-win for the OSS community and the students.
  - Encourage educators and students.

Implications

- Project owners:
  - Encourage project owners to accept and welcome new-comer

Implications

- Researchers:
  - What other factors relate to the differences.
  - Pair programming matching.