Analyzing information from versioning systems to detect logical dependencies in software systems

Adelina Diana Stana, Ioana Șora

Department of Computer and Information Technology
Politehnica University Timisoara, Romania

SACI, 2019

Structural dependencies

Definition

Structural dependencies are the result of *source code analysis and can be extracted from : members, call parameters, local variables.*

```
class ClassA{
    ...
    public void doSomething(ClassB b){
        b.doWork();
    }
}
```

```
class ClassB{
    ...
    public void doWork(){
        ....
    }
}
```

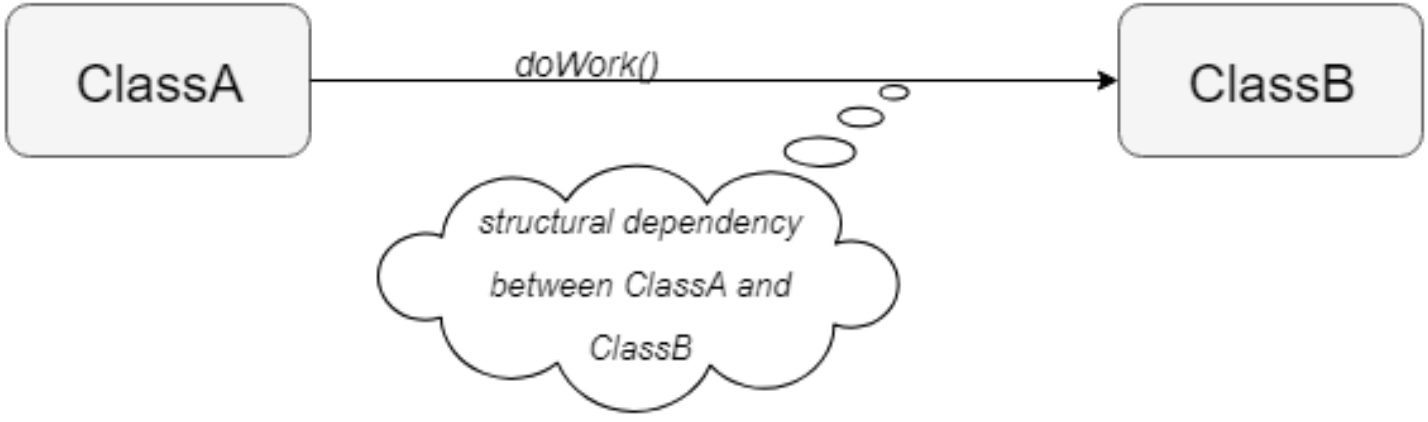


Figure 1: Example of structural dependency between two classes

Logical dependencies

Definition

Logical dependencies are the result of software history analysis and can reveal relationships that are not present in the source code code (structural dependencies).

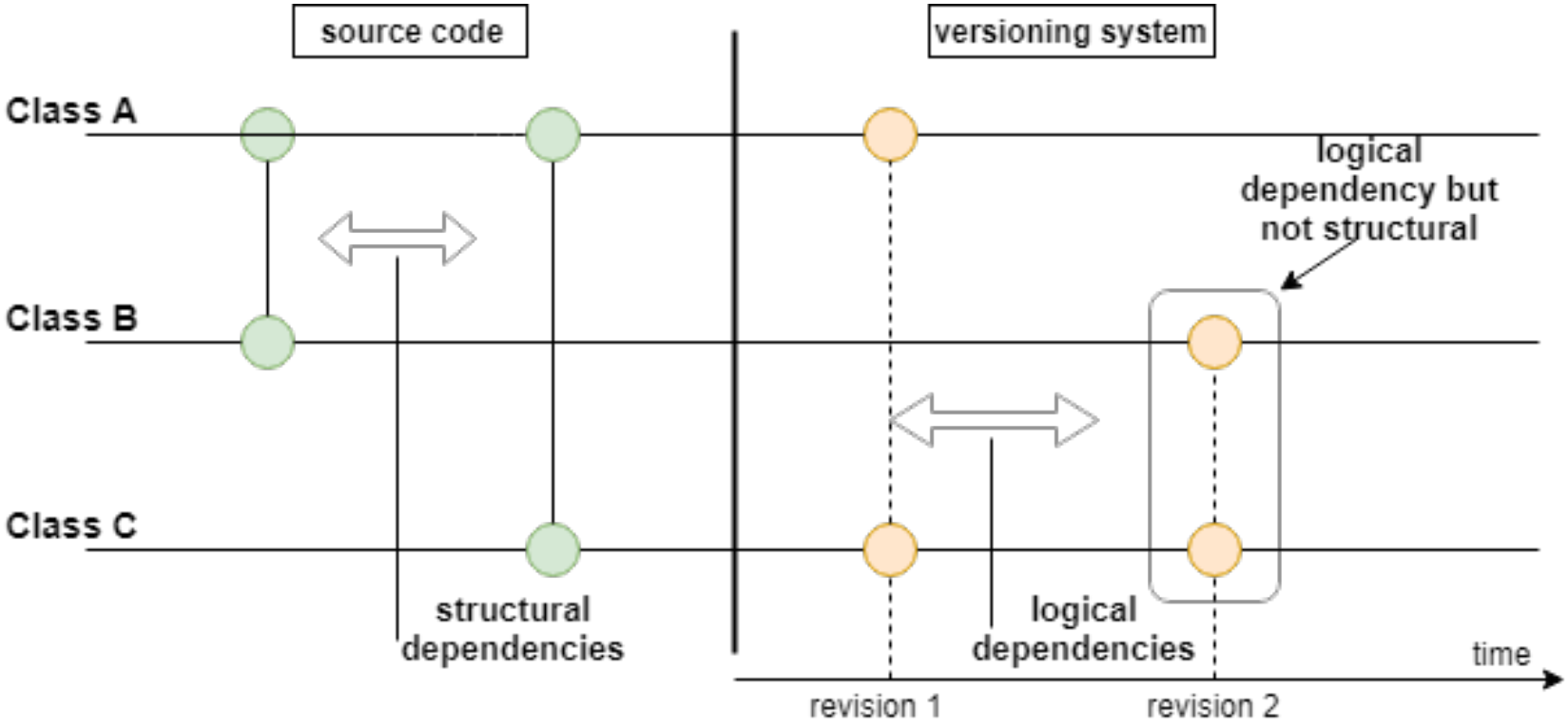


Figure 2: Example of logical and structural dependencies

Tool for measuring software dependencies

We performed analysis of 27 cpp and java systems.

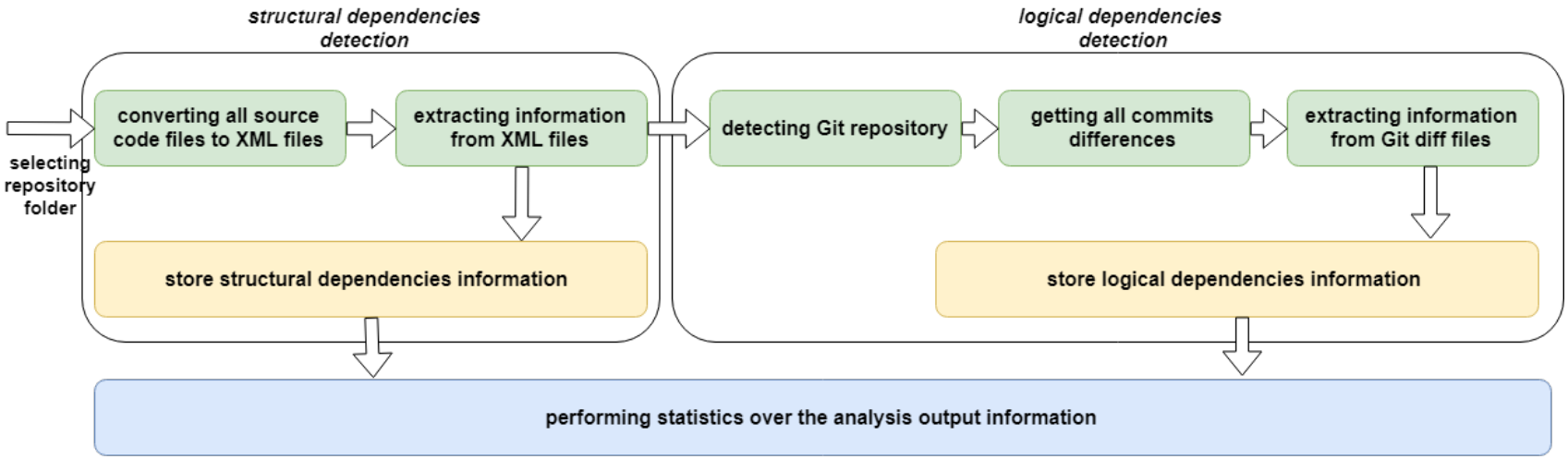
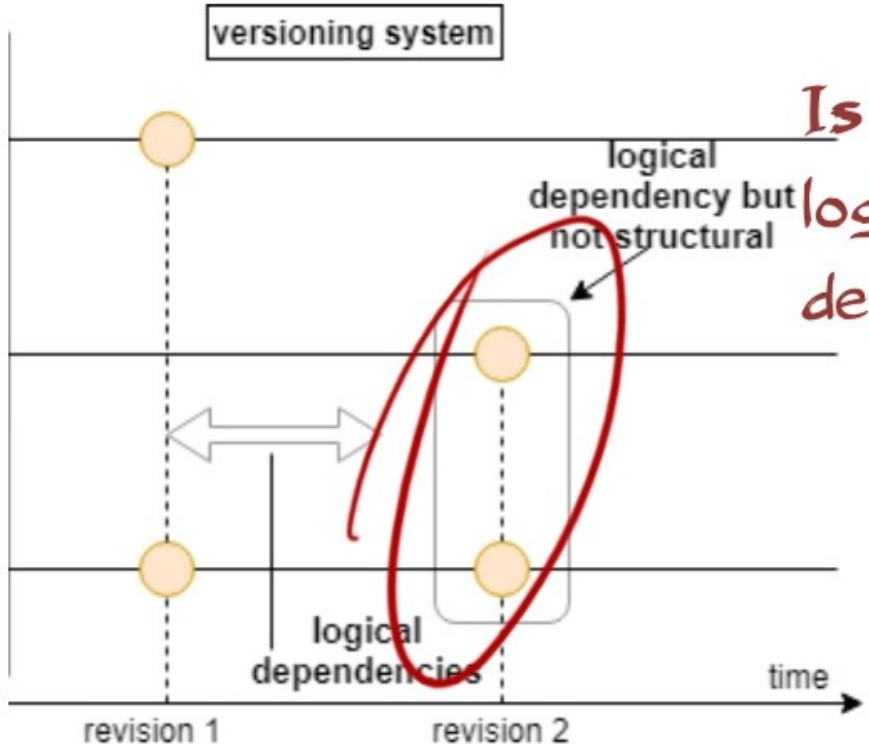


Figure 3: Workflow diagram of the tool

Logical dependencies



Is this a logical dependency??

co-changes can happen from various reasons, not all software related

Co-changing classes = logical dependencies ?

Biggest number of commits from our studied systems

Biggest number of commits from Github

10 000 commits

54 000 commits

304 000 pairs of co-changing classes

can generate aprox 1.5 million pairs of co-changing classes

Co-change  Logical dependency

Filter co-changing classes, how?

commit size

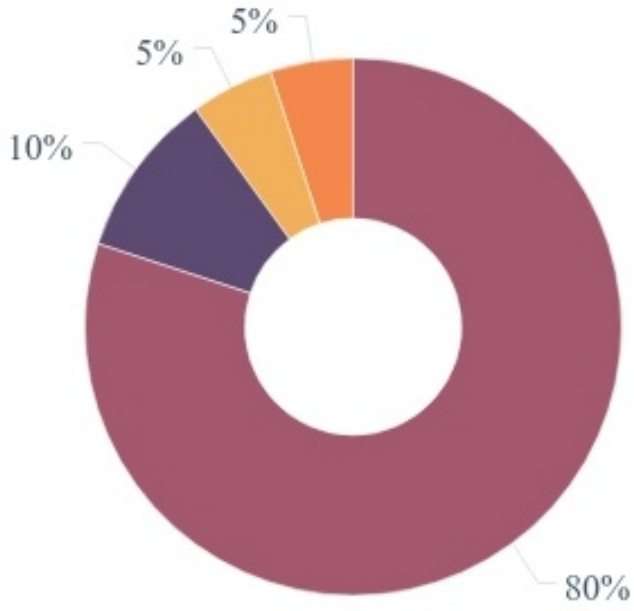
code/comment change

number of co-change occurrence

Filter Thresholds

- commit size (cs): the maximum size of commit transactions which are accepted to generate logical dependencies. The values for this threshold were 5, 10, 20 and no threshold (infinity).
- number of occurrences (occ): the minimum number of repeated occurrences for a co-change to be counted as logical dependency. The values for this threshold were 1, 2, 3 and 4.
- with/without taking comments into consideration as valid change.

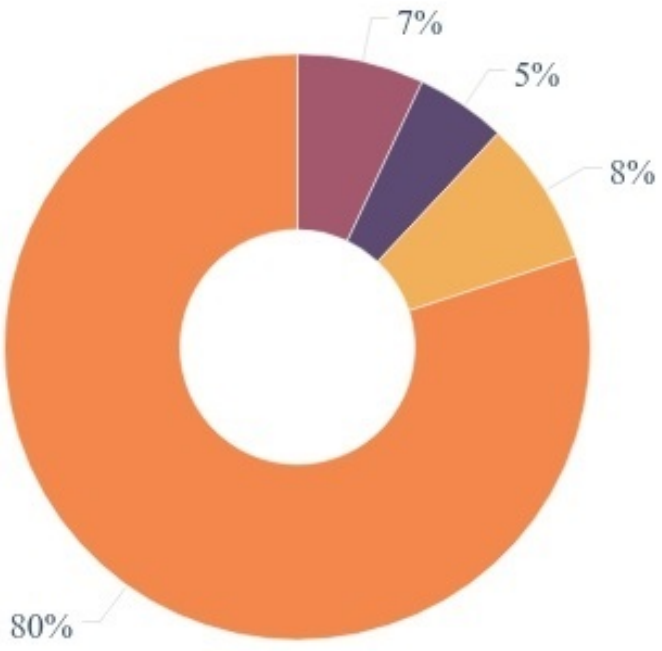
Commit transactions size - overview in percentages



less than 5 files	80%
between 5 - 10 files	10%
between 10 - 20 files	5%
more than 20 files	5%

⚠ **only code files where counted**

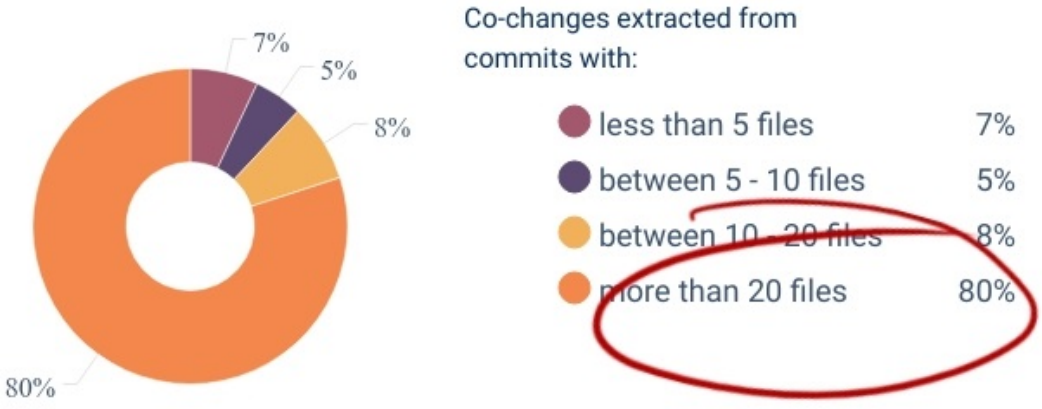
Pairs of co-changes extracted



Co-changes extracted from commits with:

less than 5 files	7%
between 5 - 10 files	5%
between 10 - 20 files	8%
more than 20 files	80%

Pairs of co-changes extracted



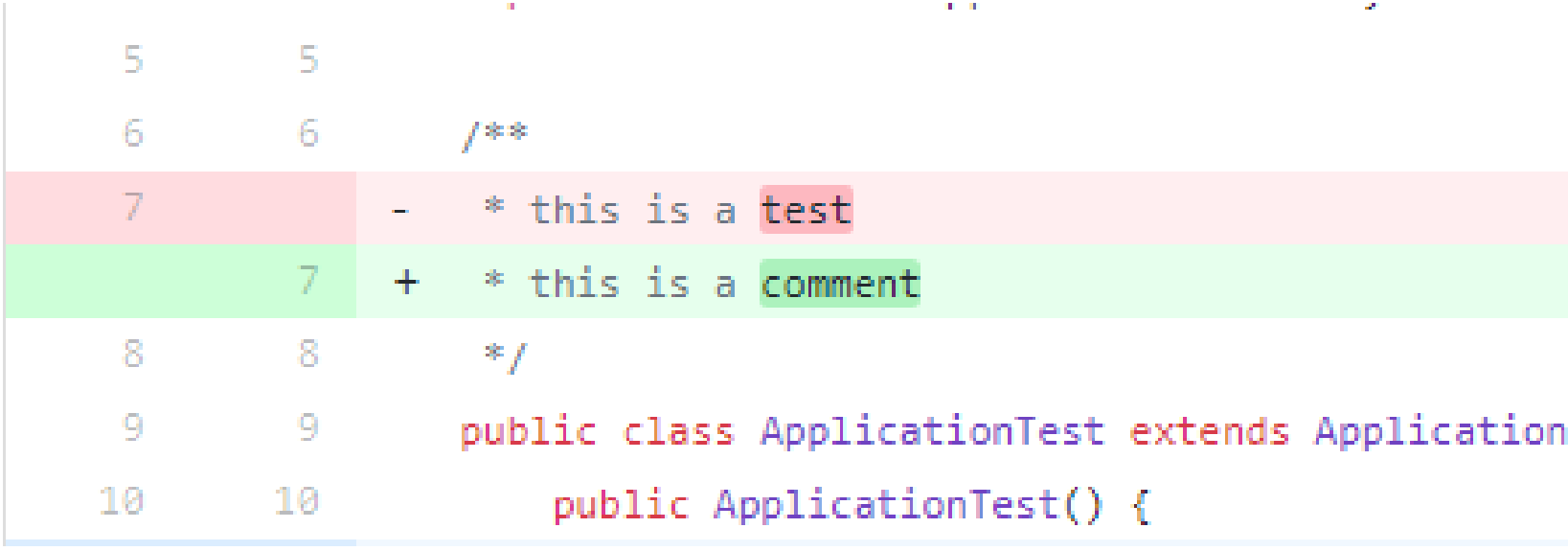
5% of total commits generate 80% from total co-changes extracted

Filter on commit size

During our research we found commit transactions with 1030 source code files, this means that **one single commit** can generate ${}^{n}C_{k} = \frac{n!}{k!(n-k)!} = \frac{1030!}{2!(1028)!} = 529935$ co-changes.

- the threshold for the commit size can be 5, 10 or even 20

Filter on comment changes



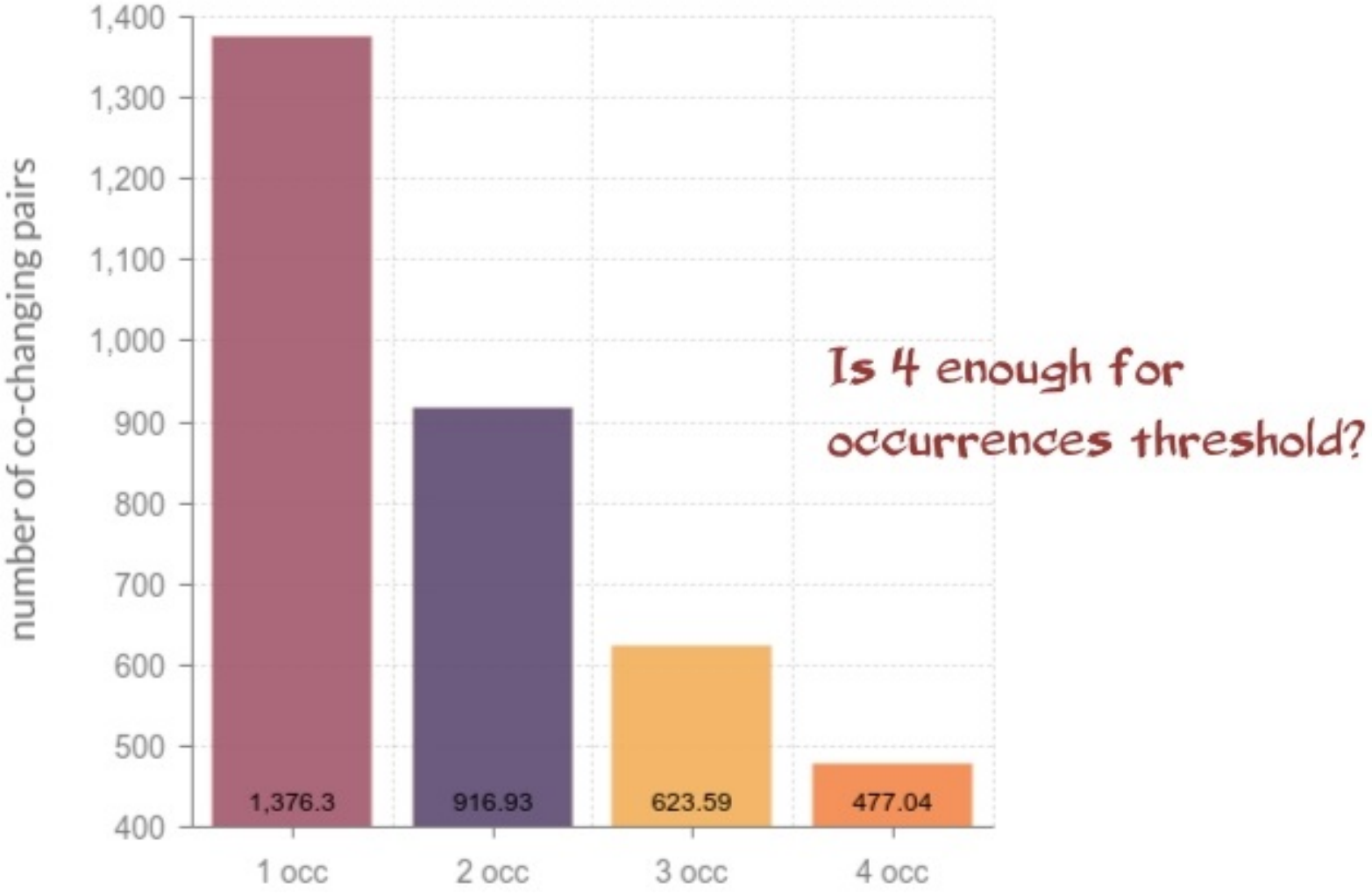
```
 5   5
 6   6      /**
 7       -   * this is a test
     7   +   * this is a comment
 8   8      */
 9   9     public class ApplicationTest extends Application
10  10         public ApplicationTest() {
```

- approx -5% from co-changes extracted from all commit sizes
- approx -1% from co-changes extracted from commits with less than 10 files

Occurrence of co-changing classes

The more occurrences of a co-changing pair the highest chance to be a truly logical dependency.

Filter on number of occurrences

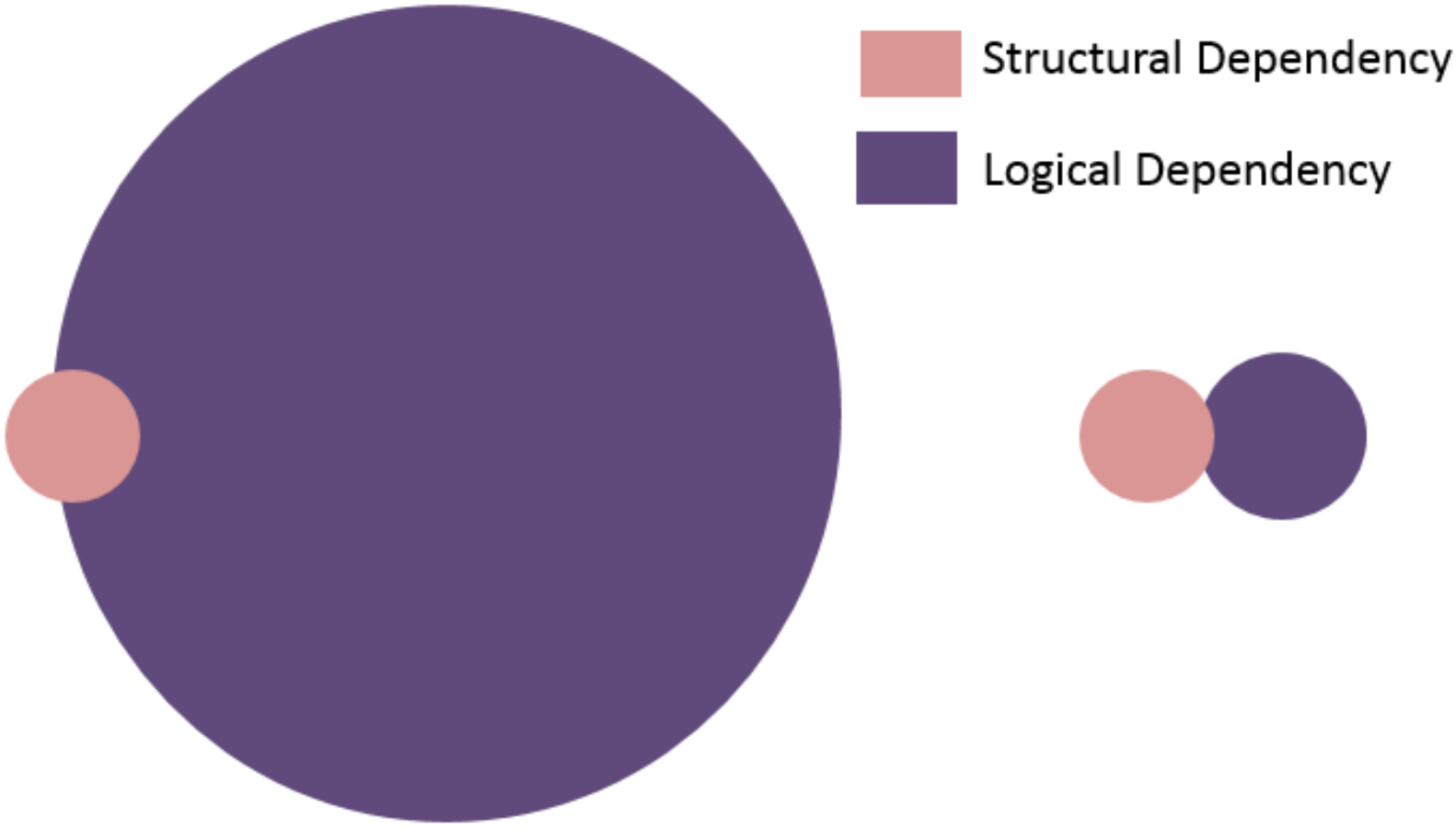


Filter on number of occurrences

	$occ \geq 1$	$occ \geq 2$	$occ \geq 3$	$occ \geq 4$
restfb	92979	78434	29824	29031
rxjava	14987	9842	3000	2237
metro-jax-ws	1621	793	480	431

Table 1: Filtered co-changing pairs from commits with less than 5 files

Impact of co-change filtering



Case a.)
cs $\leq \infty$
occ ≥ 1
LD/SD=314.59

Case b.)
cs ≤ 5
occ ≥ 4
LD/SD=1.81

Impact of co-change filtering - observations

Only few of structural dependencies are doubled by logical dependencies.

	$cs \leq 5$	$cs \leq 10$	$cs \leq 20$	$cs < \infty$
$occ \geq 1$	19,75	29,86	39,29	76,59
$occ \geq 2$	12,50	20,20	27,68	66,11
$occ \geq 3$	8,49	14,22	19,94	55,99
$occ \geq 4$	6,58	10,95	15,76	47,12

Table 2: Percentage of SD that are also LD

Future work

- validation of extracted logical dependencies by using them to enhance dependency models
- extract structural dependencies from all the revisions of the system to filter out the old logical dependencies

Conclusions

- small commit transactions are the most frequent kind of transactions (80% of all commit transactions)
- increasing the threshold for the minimum number of repeated occurrences for a co-change to be counted as a logical dependency reduces significantly the number of co-changing pairs of classes
- filter thresholds shall be calculated according to some variables such as: total number of commits, total number of entities.