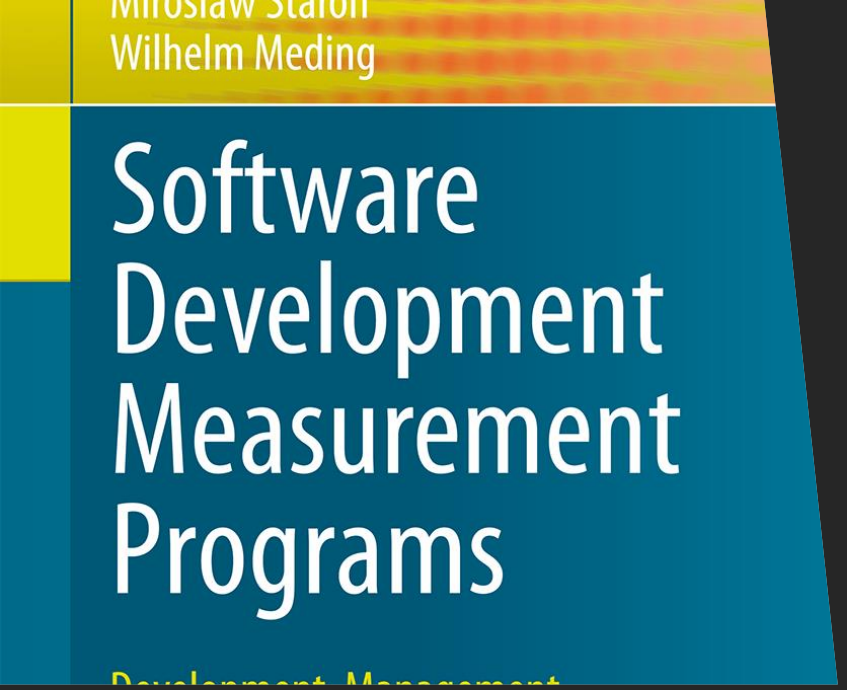
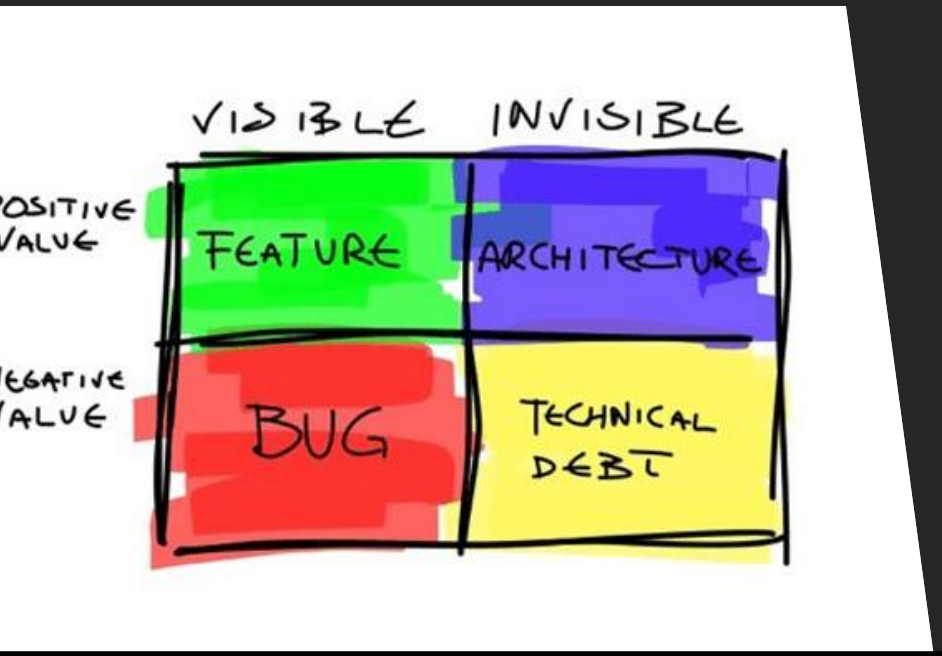
EVOLUTION OF TECHNICAL DEBT: AN EXPLORATORY STUDY

ABDULLAH MAMUN
ANTONIO MARTINI
MIROSLAW STARON
CHRISTIAN BERGER
JÖRGEN HANSSON





Research goal

- To understand how the newly introduced metric *technical debt density trend* explains evolution of technical debt

Study design

- 4013 Java source files
- 11,822 Git commits
- 21 open source Java projects
- SonarQube tool with 138 code smells

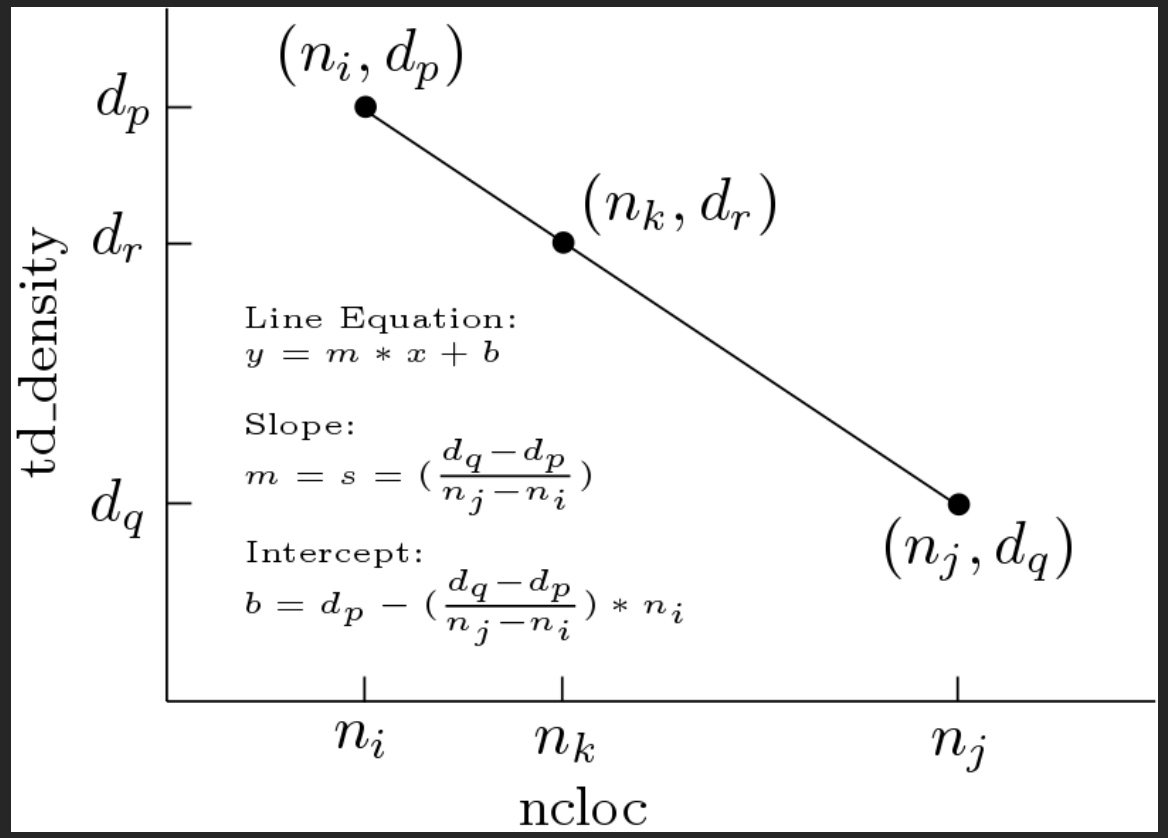
Table 2: Overview of the Selected Projects

Organization	Project Name [Number of Analyzed Revisions (commits), Project Duration (months), Analyzed Java Code (KLOC)]
Microsoft	malmo [295, 5, 14]; oauth2-useragent [171, 12, 3]; Vso-intellij [305, 12, 64]; thrifty [242, 12, 44]; Git-Credential-Manager-for-Mac-and-Linux [141, 13, 5]
Twitter	Ambrose [167, 48, 5]; cloudhopper-smpp [94, 57, 12]; elephant-bird [449, 76, 23]
Netflix	Fenzo [98, 20, 11]; ribbon [223, 46, 22]; astyanax [549, 55, 55]
square	Dagger [306, 46, 9]; retrofit [776, 72, 13]; picasso [518, 42, 10]
Esri	Solutions-geoevent-java [218, 38, 35]; geometry-api-java [100, 43, 76]
Shopify	Nokogiri [1,788, 75, 26]
SAP	Cloud-sfsf-benefits-ext [52, 24, 3]
Apache	Kafka [2,302, 64, 89]; zookeeper [1,474, 109, 73]; zeppelin [1,606, 39, 65]
Total	[11,874, 907, 658]

Metrics used in this study

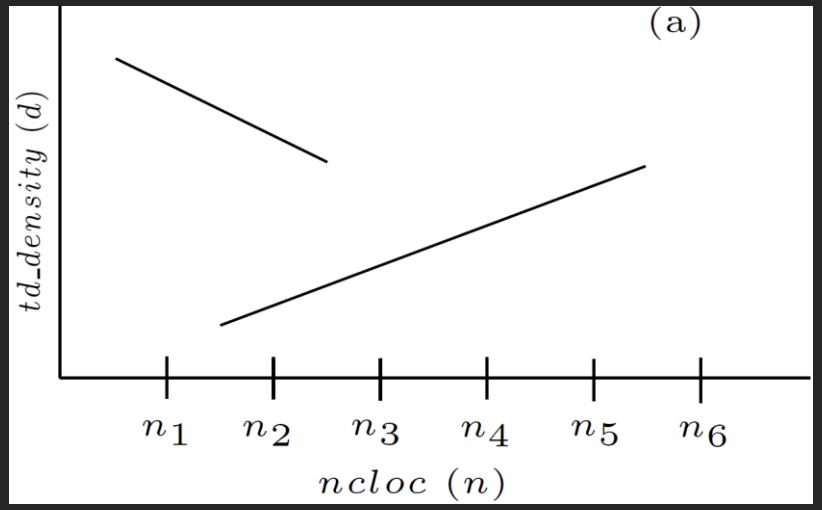
Metric Name	Description
ncloc	Number of physical lines of code that are not comments (line only containing space, tab, and carriage return are ignored)
tech_debt	Effort (time in minutes) to resolve identified code smells
td_density	Density of *tech_debt* indicating *tech_debt* per 100 *ncloc*. Calculated as: *td_density* = (*tech_debt*/*ncloc*)*100
td_density_trend	Slope of the line of two points ($ncloc_p$, $td\_density_p$) & ($ncloc_q$, $td\_density_q$), where p and q are two successive commits.

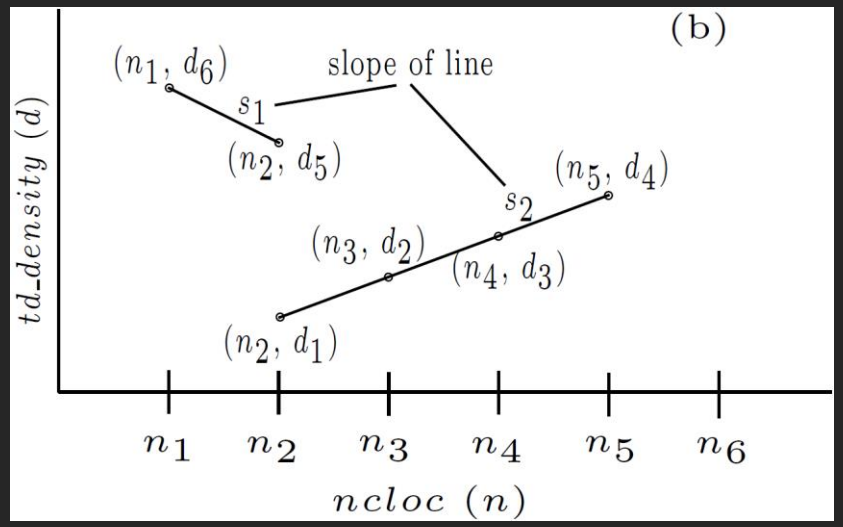
Technical debt density

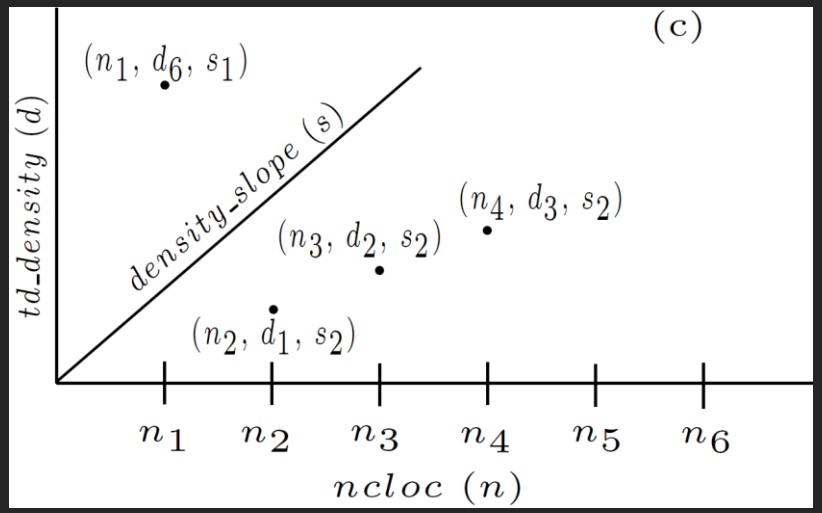


Technical debt density trend

Data processing for two revisions of two files.





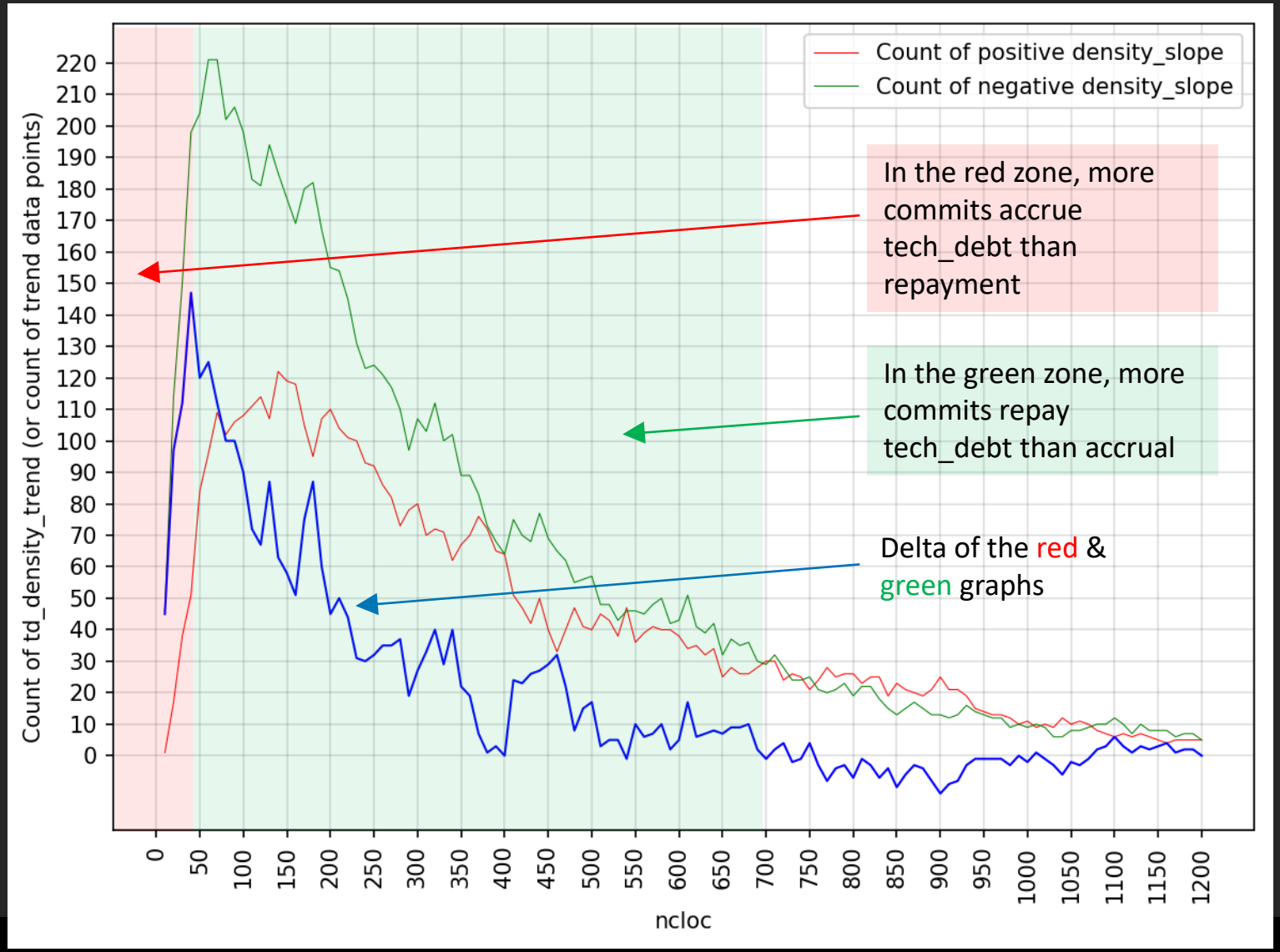


Data collection

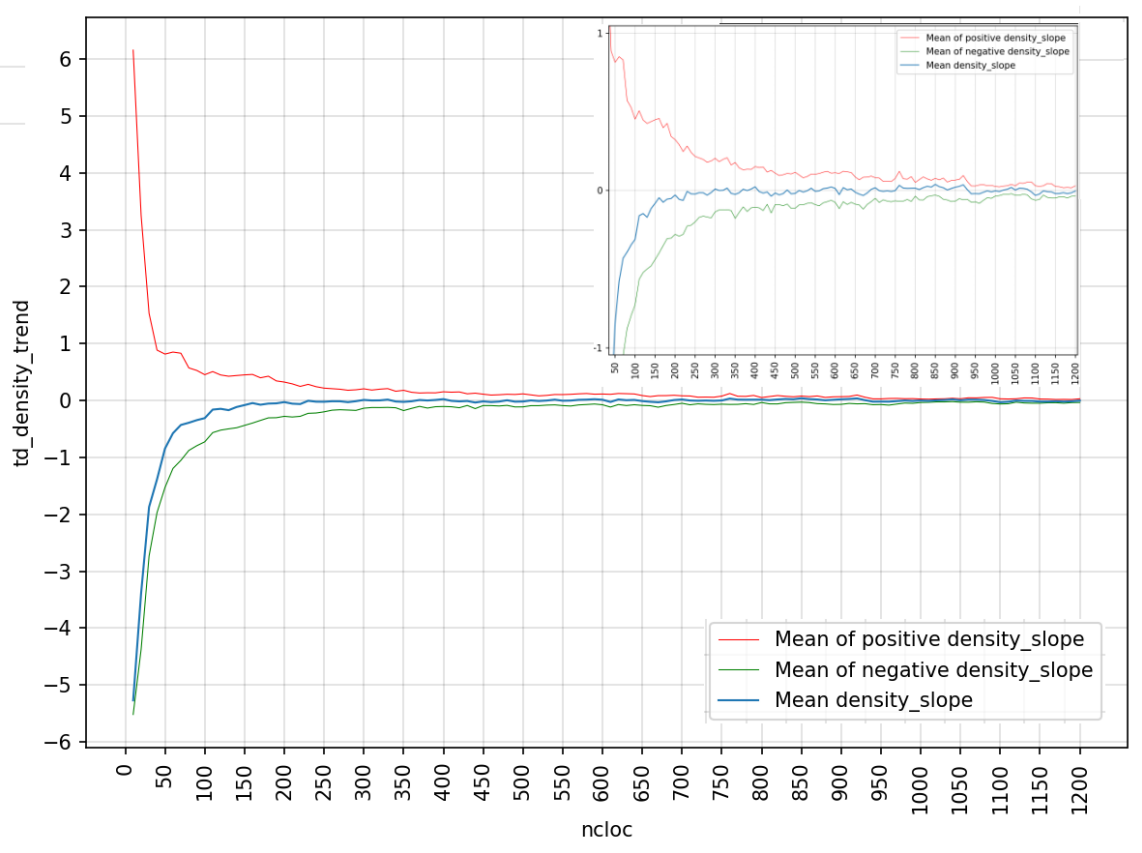
Table 3: Chronological steps for data processing.

Step	Operation	Number of files	Number of file segments	Number of data points
1	Raw data collected from database	11,358		2,527,990
2	Data points with NONE value removed	11,358		2,527,990
3	Remove successive duplicated data points	11,358		24,976
4	Files with single data points removed	4,013		18,371
5	Split every two successive data points into pairs		14,358	14,358*2 = 28,716
6	Transform every pair of data points so that we have data at specific *ncloc* points, i.e., n_1, n_2, n_3, etc. like in Fig. 2(b)		14,358	28,716
7	Delete any file segment that has a single data points		3,665	17,584
8	Create new pairs/segments by taking every two successive data points of the existing file segments. A pair is of the form (n_1, d_6), (n_2, d_5) as shown in Fig. 2(b)		13,919	13,919*2 = 27,838
9	Calculate slope s_1 from (n_1, d_6), (n_2, d_5) using Equ. 1 as shown in Fig. 2(b) and transform the pair into a single data point with three elements (n_1, d_6, s_1) as seen in Fig. 2(c)			13,919
10	Remove 782 data points associated to any *ncloc* >1200 because of lower number of samples for higher *ncloc* values			13,137
11	Remove 17 outliers related to *td_density*			13,120

Result: Overall Accrual & Repayment of TD throughout the whole development period

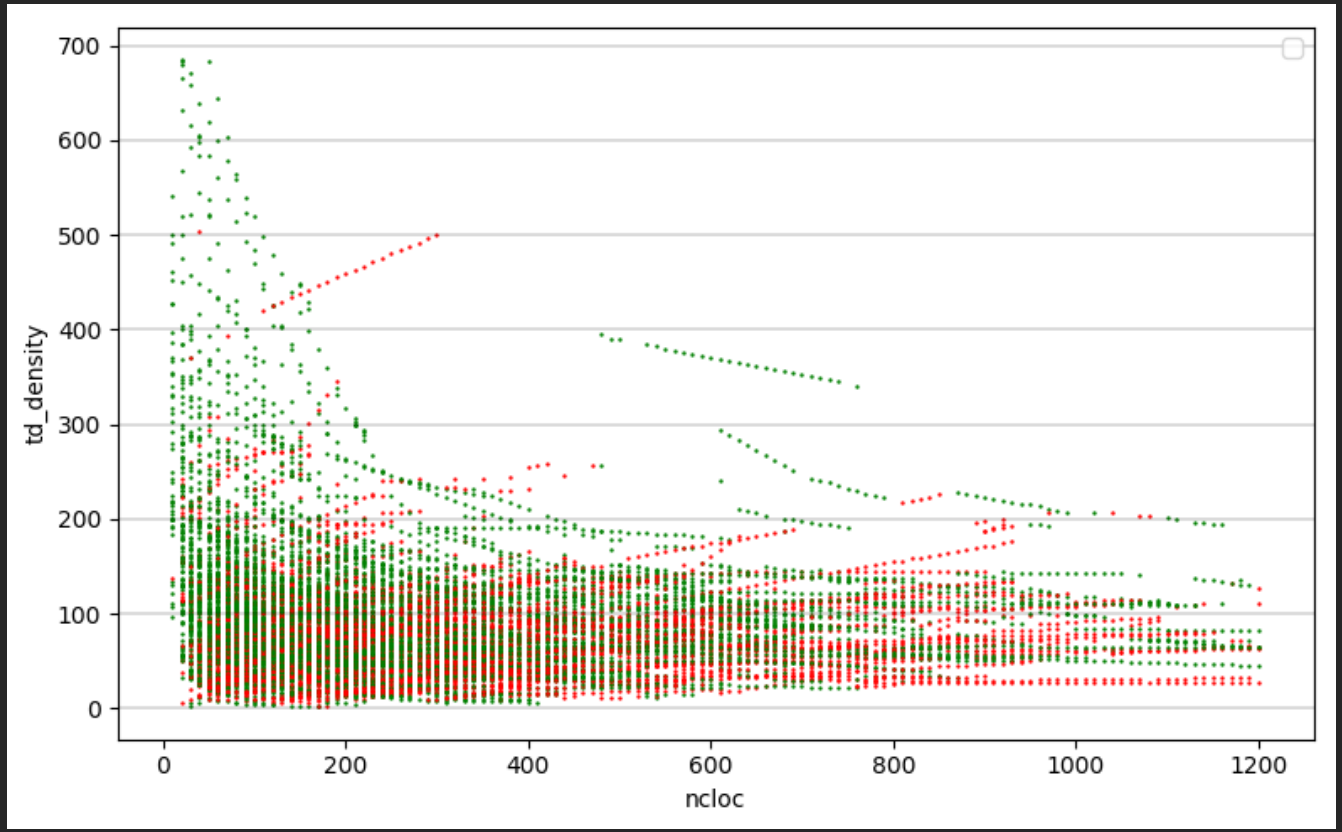


Result: Be Careful of the Measurement Type

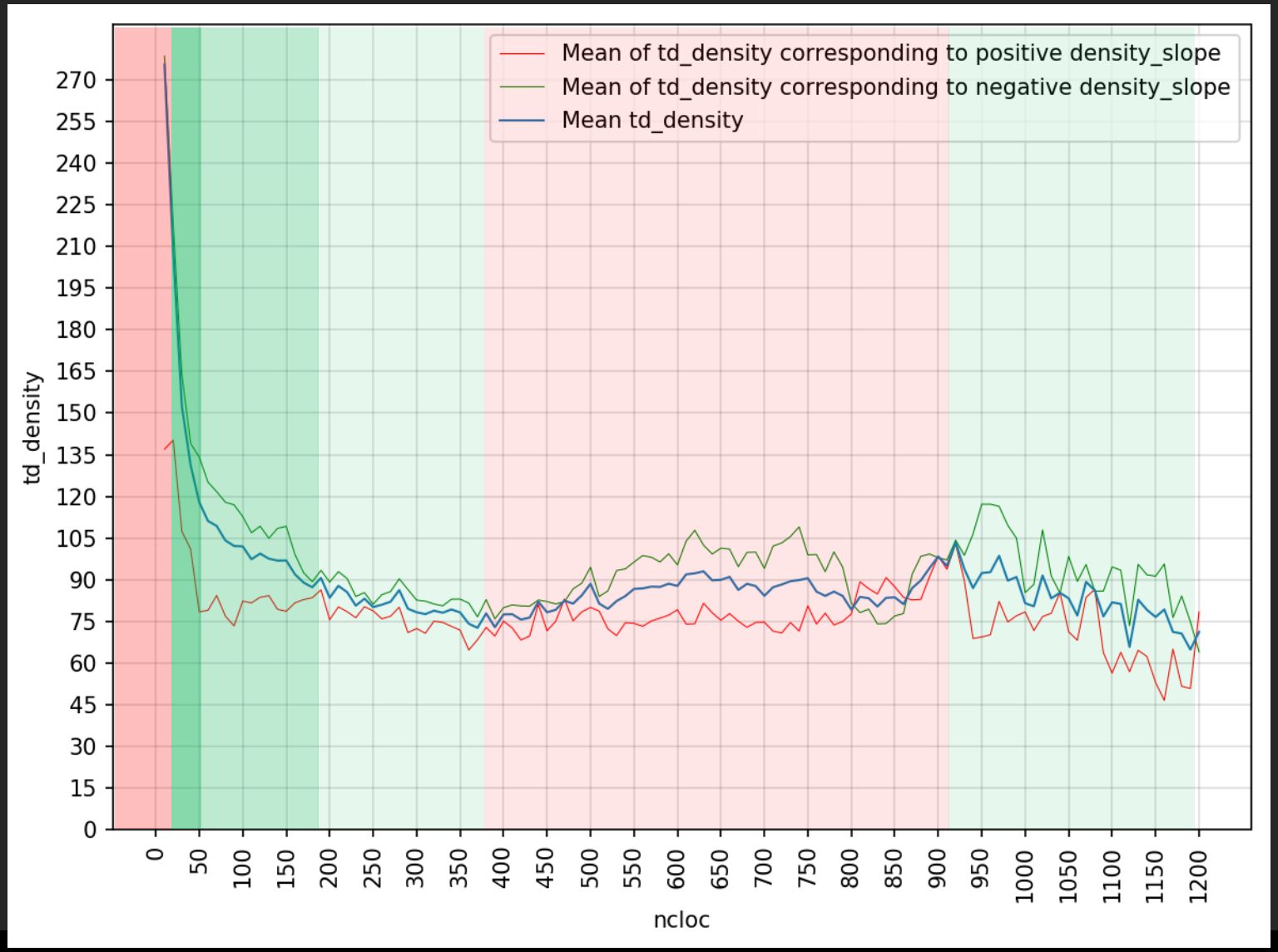


- The tech_debt metric is cumulative
- The td_density metric is normalized but still it carries effects of cumulatively measureming tech_debt
- The td_density_trend metric also carries effects of cumulatively measureming tech_debt
- The above figure/plot is not normalized even though td_density_trend comes from the normalized td_density metric
- Constructing this figure using a non-cumulative tech_debt metric will give us a normalized plot.

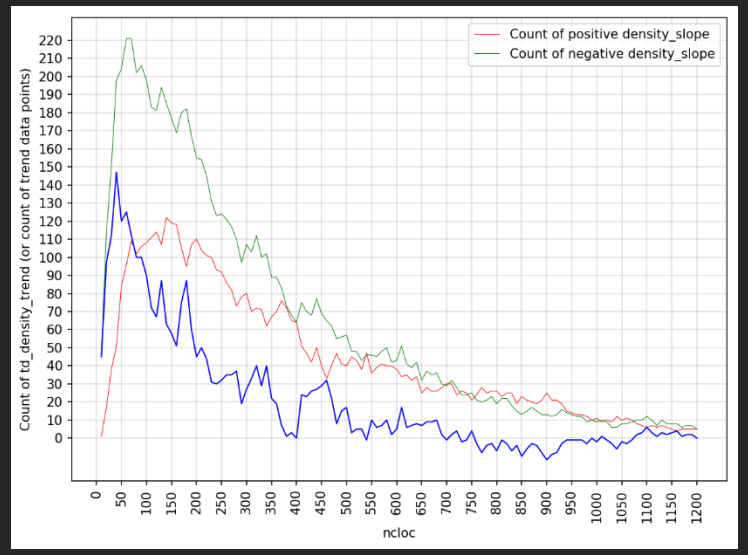
Result: Evolution of td_density explained with td_debt_trend

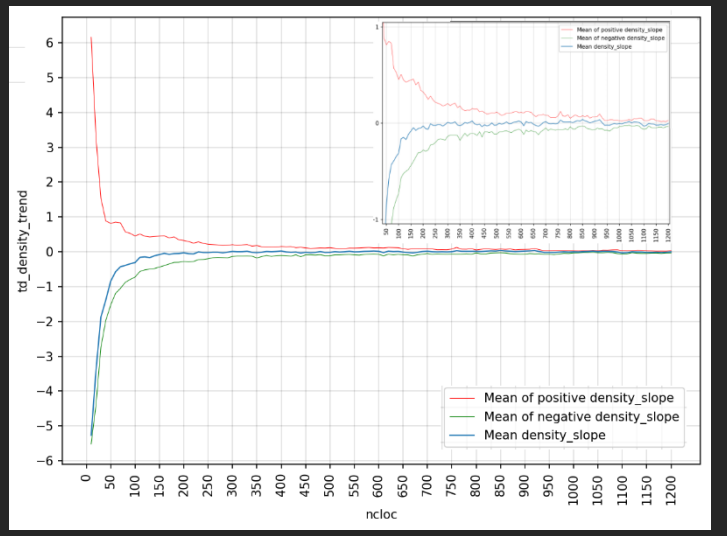


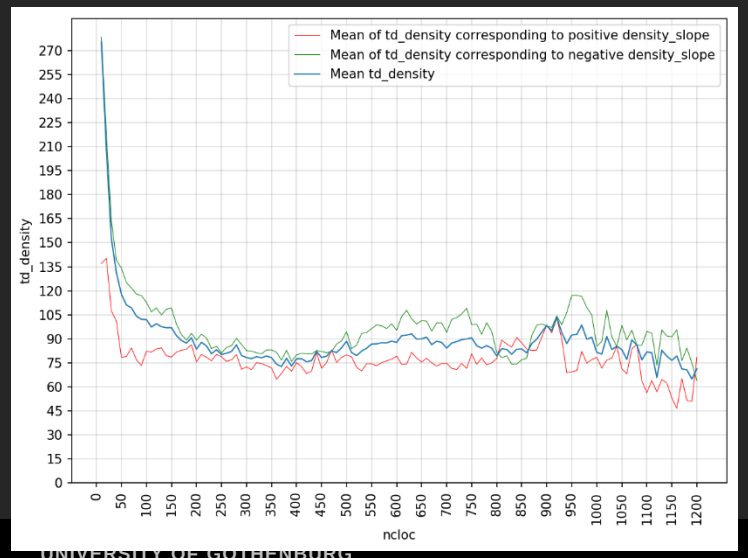
Result: Evolution of td_density explained with td_debt_trend (cont.)

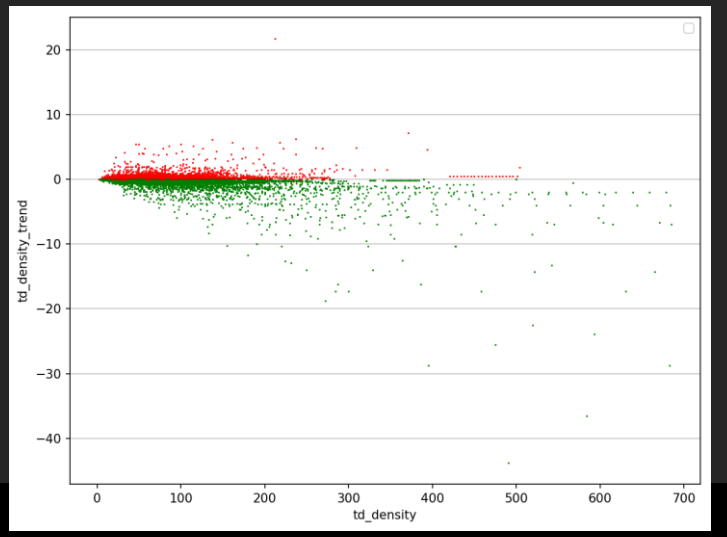


Result: Componentization of tech_debt









Conclusion and future work

- *Technical debt density trend* metric
  - Allows explaining the evolution of technical debt in terms of accrual and repayment (componentization of technical debt) in multiple ways

- We should be careful about the basic measurement types: Cumulative vs. non-cumulative

- Future Work
  - Investigate evolution of technical debt incorporating non-cumulative (organic) measurement of technical debt


UNIVERSITAS GOTHOBURGENSIS
1891
UNIVERSITY OF
GOTHENBURG