lab-soft
software engineering laboratory

Unraveling the Impact of Code Smell Agglomerations on Code Stability

AMANDA SANTANA, EDUARDO FIGUEIREDO, JULIANA ALVES PEREIRA

lab-soft
software engineering laboratory

Agenda

INTRODUCTION

BACKGROUND

STUDY DESIGN

RESULTS

THREATS TO VALIDITY

CONCLUSION

FUTURE WORKS

lab-soft
software engineering laboratory

Code Smell

- Symptoms of developer's decisions that may lead to code quality degradation:
  - Complexity;
  - Cohesion;
  - Coupling;
  - Modularity;
  - Size;
  - Faults.

lab-soft
software engineering laboratory

Code Smell Agglomerations

- When two or more code smell occurs on the same piece of code.
- Heterogeneous
  - Two or more smells of different types
- Homogeneous
  - Two or more smells of the same type
- Isolated
  - Only one smell
- Clean
  - No smell

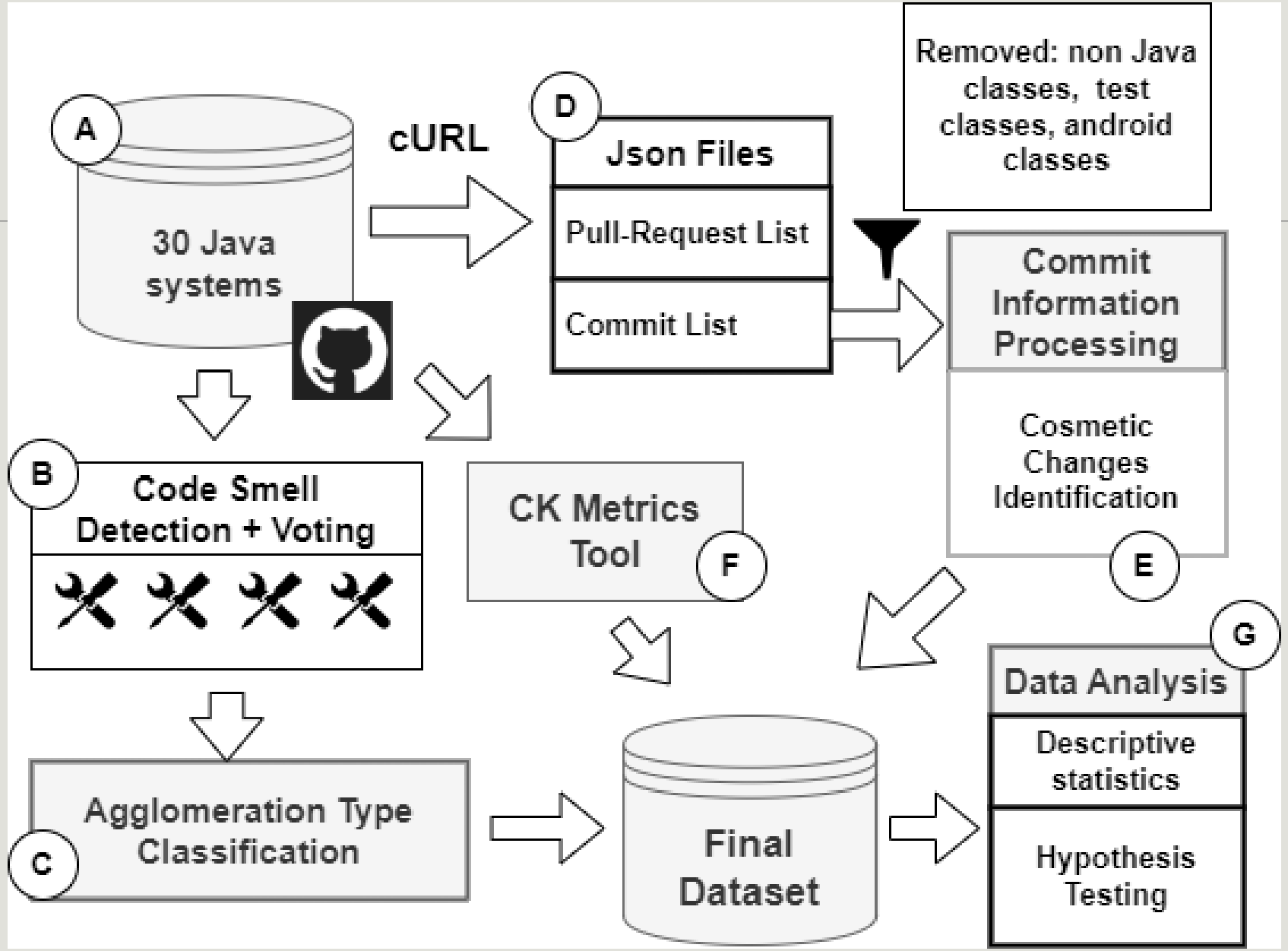
Goal: Provide Evidences of Code Smell Agglomeration Stability

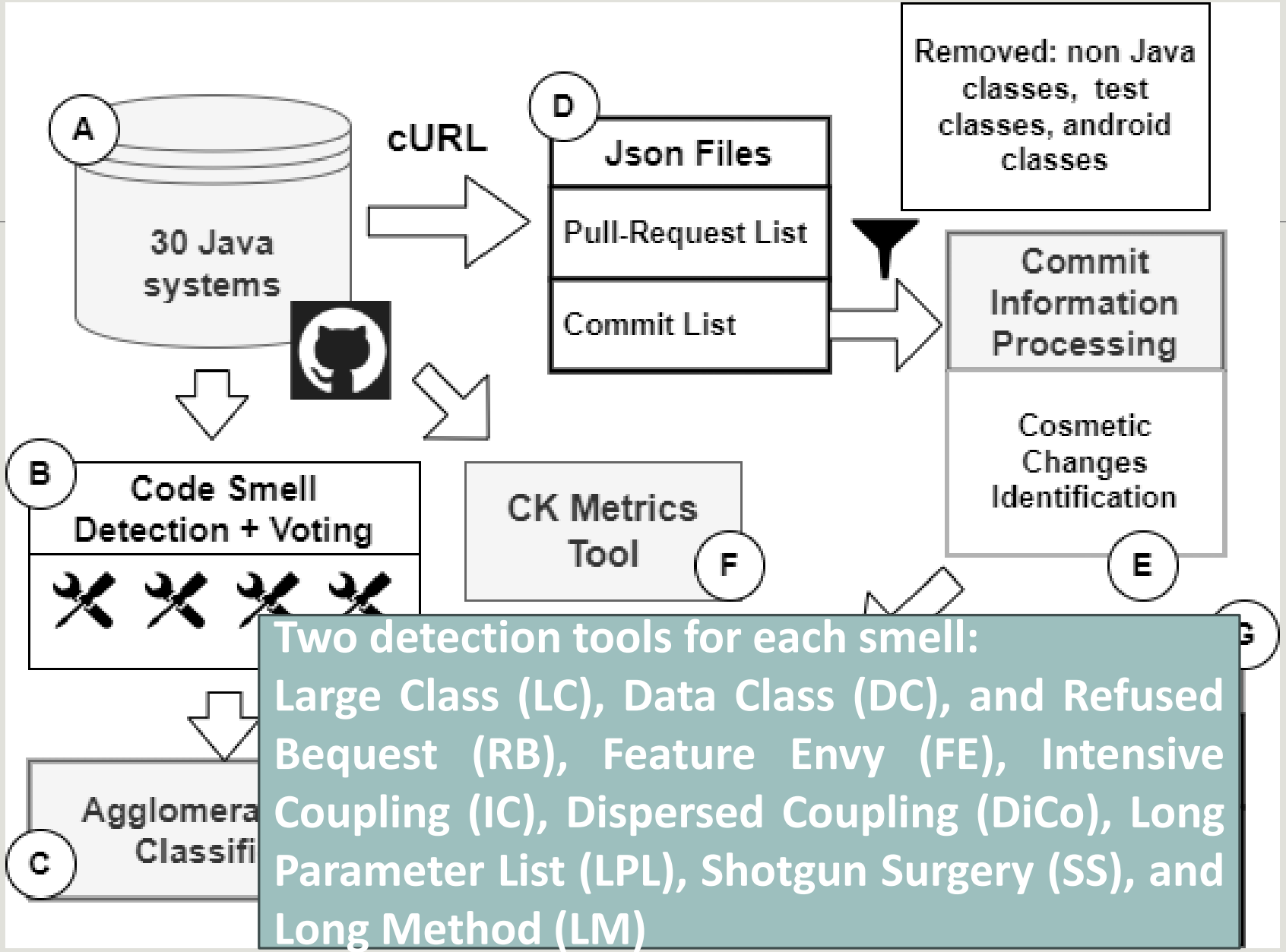
Some concepts

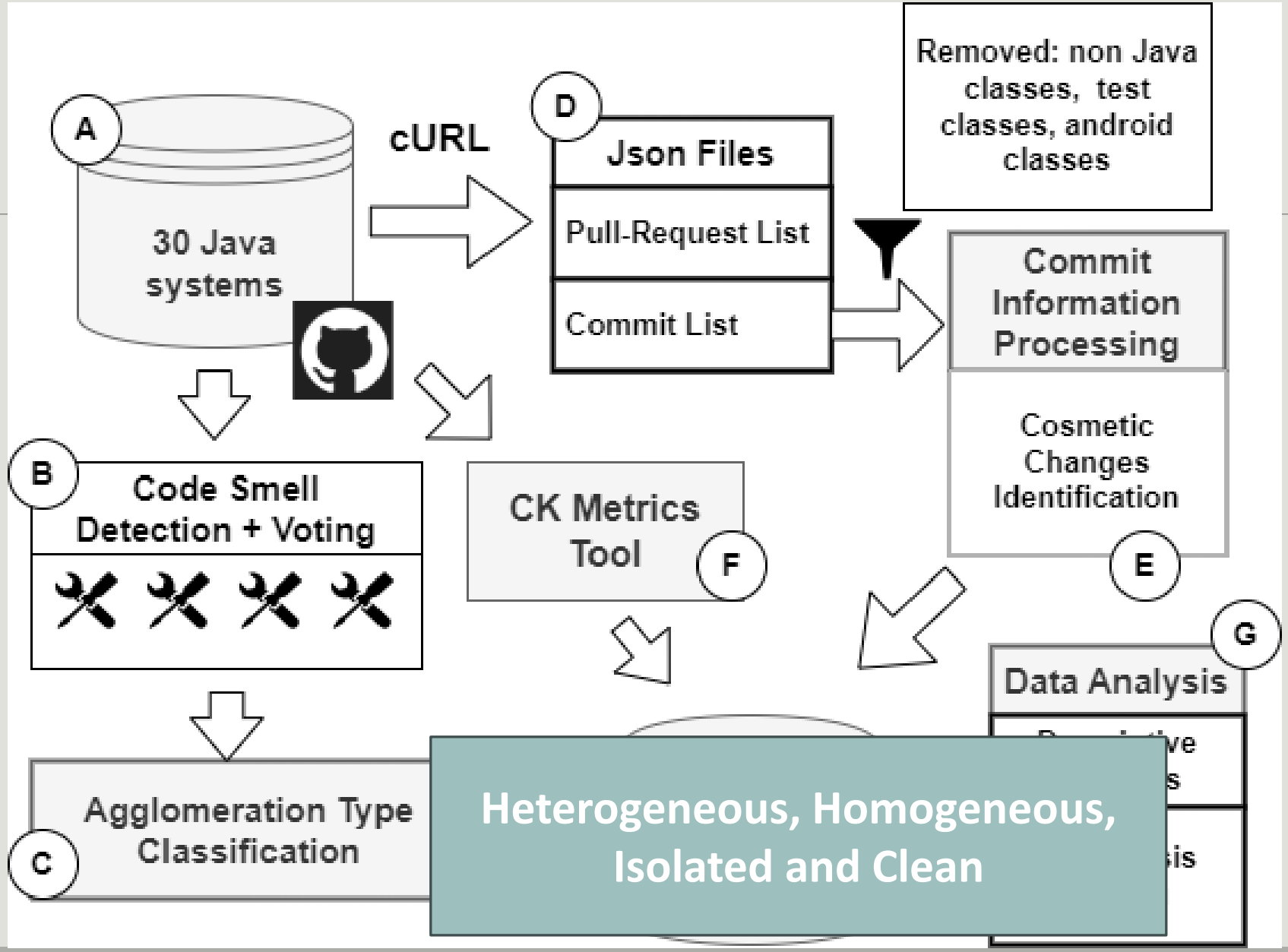
- General: all 30 systems are considered as our dataset;
- Class History Type: Modified, New, Dead;
- Modification Type: #Add, #Del, Churn.

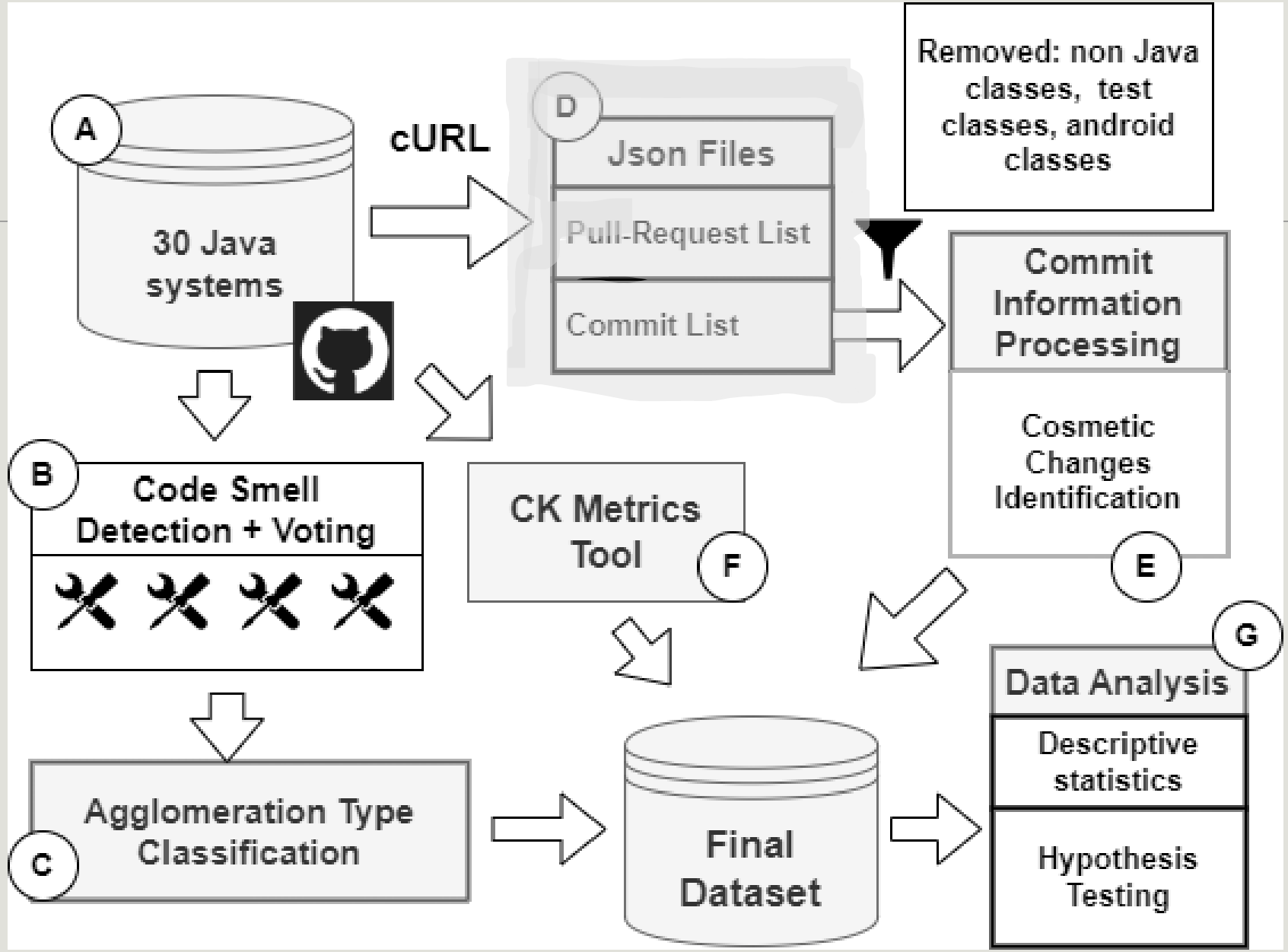
Research Questions

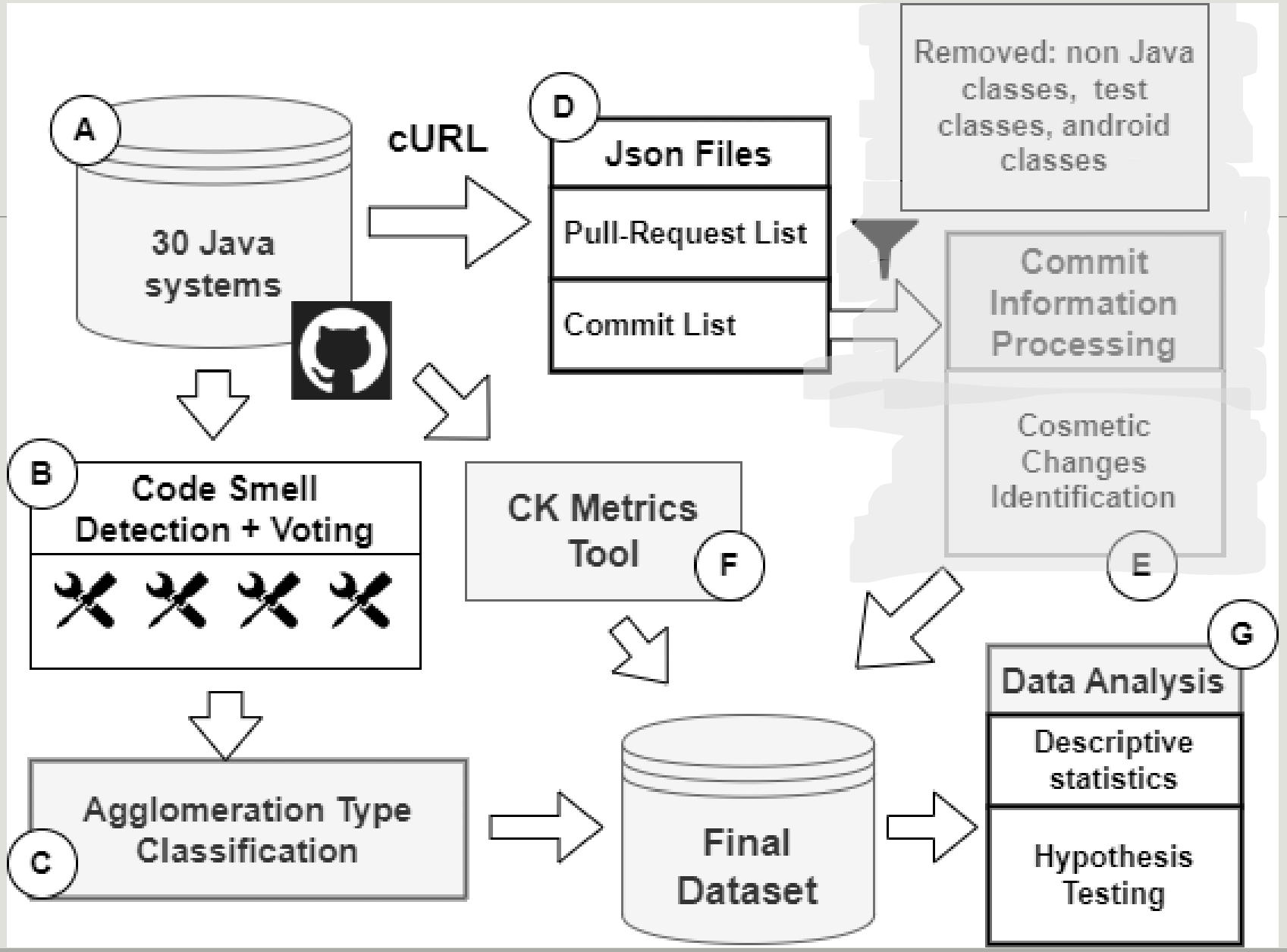
- **RQ1:** Do Heterogeneous and Homogeneous Agglomerations undergo more frequent changes compared to Isolated and Clean types?
  - #Commits, Lines Added/Deleted/Churn, Number of Agglomerations Changed
- **RQ2:** Do Heterogeneous and Homogeneous Agglomerations undergo changes in more intensity compared to Isolated and Clean types?
  - proportion of the lines of code added/deleted/churn with respect to the number of lines of code (LOC) of its respective class

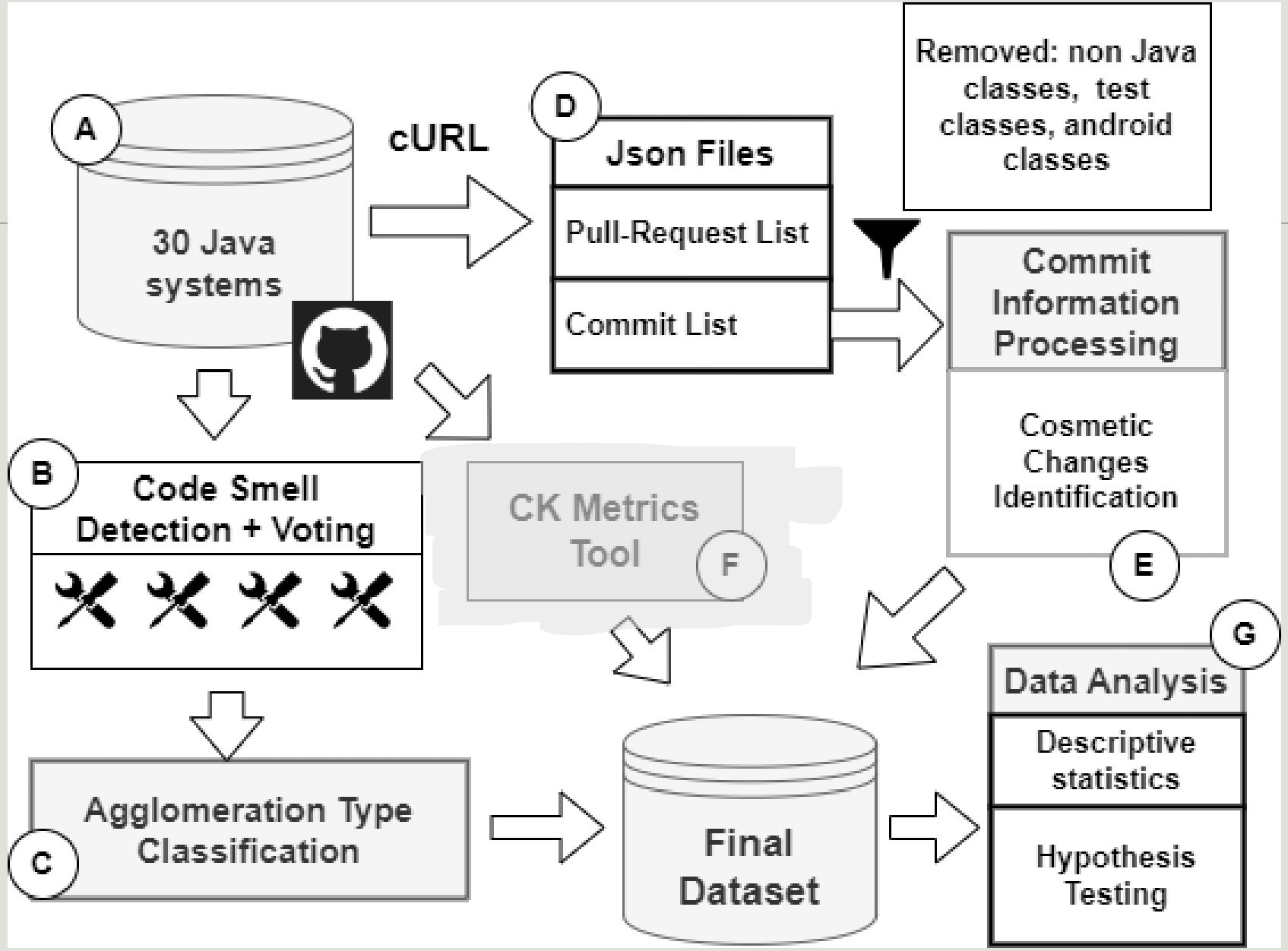

A
30 Java systems
cURL
D
Json Files
Pull-Request List
Commit List
Removed: non Java classes, test classes, android classes
Commit Information Processing
Cosmetic Changes Identification
E
B
Code Smell Detection + Voting
CK Metrics Tool
F
G
Data Analysis
Descriptive statistics
Hypothesis Testing
C
Agglomeration Type Classification
Final Dataset

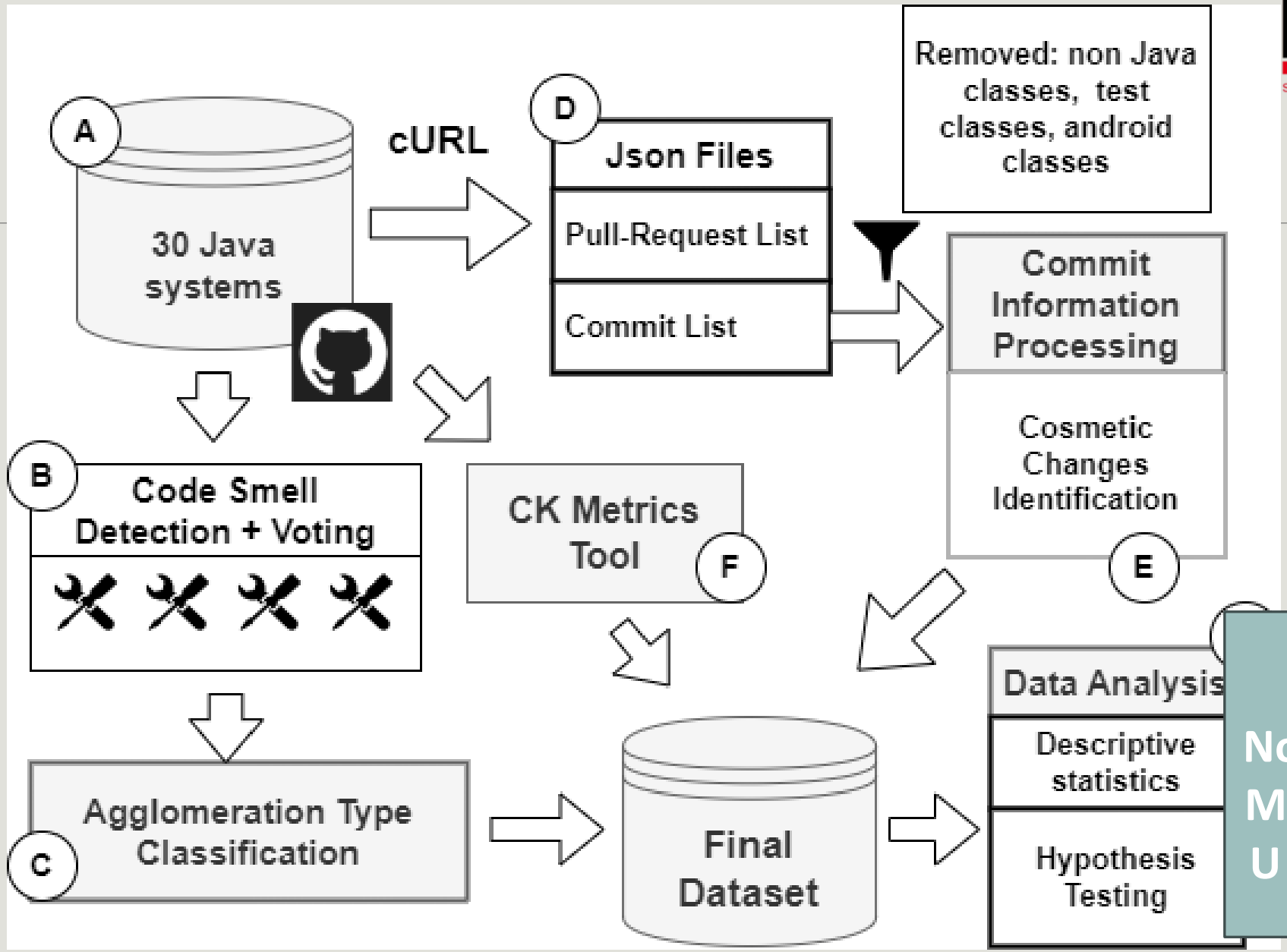

A
30 Java systems
cURL
D
Json Files
Pull-Request List
Commit List
Removed: non Java classes, test classes, android classes
Commit Information Processing
Cosmetic Changes Identification
E
B
Code Smell Detection + Voting
CK Metrics Tool
F
C
Agglomera
Classifi
Two detection tools for each smell:
Large Class (LC), Data Class (DC), and Refused Bequest (RB), Feature Envy (FE), Intensive Coupling (IC), Dispersed Coupling (DiCo), Long Parameter List (LPL), Shotgun Surgery (SS), and Long Method (LM)

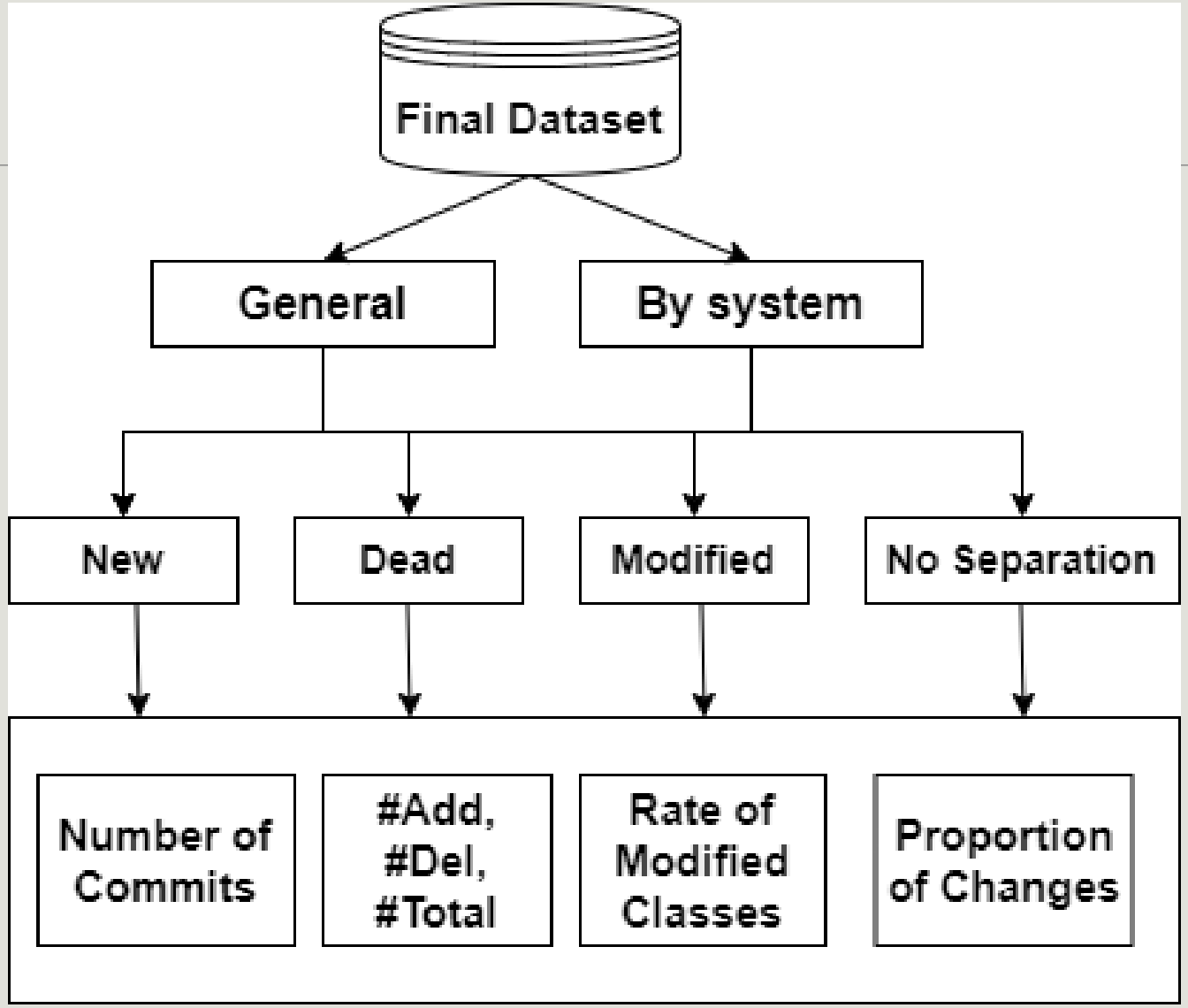

A
30 Java systems
cURL
D
Json Files
Pull-Request List
Commit List
Removed: non Java classes, test classes, android classes
Commit Information Processing
Cosmetic Changes Identification
E
B
Code Smell Detection + Voting
CK Metrics Tool
F
G
Data Analysis
C
Agglomeration Type Classification
Heterogeneous, Homogeneous, Isolated and Clean


A
30 Java systems
cURL
D
Json Files
Pull-Request List
Commit List
Removed: non Java classes, test classes, android classes
Commit Information Processing
Cosmetic Changes Identification
E
B
Code Smell Detection + Voting
CK Metrics Tool
F
G
Data Analysis
Descriptive statistics
Hypothesis Testing
C
Agglomeration Type Classification
Final Dataset


A
30 Java systems
cURL
D
Json Files
Pull-Request List
Commit List
Removed: non Java classes, test classes, android classes
Commit Information Processing
Cosmetic Changes Identification
E
B
Code Smell Detection + Voting
CK Metrics Tool
F
G
Data Analysis
Descriptive statistics
Hypothesis Testing
C
Agglomeration Type Classification
Final Dataset


A
30 Java systems
cURL
D
Json Files
Pull-Request List
Commit List
Removed: non Java classes, test classes, android classes
Commit Information Processing
Cosmetic Changes Identification
E
B
Code Smell Detection + Voting
CK Metrics Tool
F
G
Data Analysis
Descriptive statistics
Hypothesis Testing
C
Agglomeration Type Classification
Final Dataset


A
30 Java systems
cURL
D
Json Files
Pull-Request List
Commit List
Removed: non Java classes, test classes, android classes
Commit Information Processing
Cosmetic Changes Identification
E
B
Code Smell Detection + Voting
CK Metrics Tool
F
C
Agglomeration Type Classification
Final Dataset
Data Analysis
Descriptive statistics
Hypothesis Testing
Min-Max Normalization + Mann Whittney U + Cliff's Delta


Final Dataset
General
By system
New
Dead
Modified
No Separation
Number of Commits
#Add, #Del, #Total
Rate of Modified Classes
Proportion of Changes

lab-soft
software engineering laboratory

Results

Commits

- Het=Clean (-0.58)
- Hom=Clean (-0.58)
- Isol=Clean (-0.584).

Clean classes receive statistically more commits than smelly classes, and the difference is Large.

Lines of Modified Code – General Perspective – Class Type History

GENERAL PERSPECTIVE OF MODIFICATION TYPES BY CLASS TYPE HISTORY

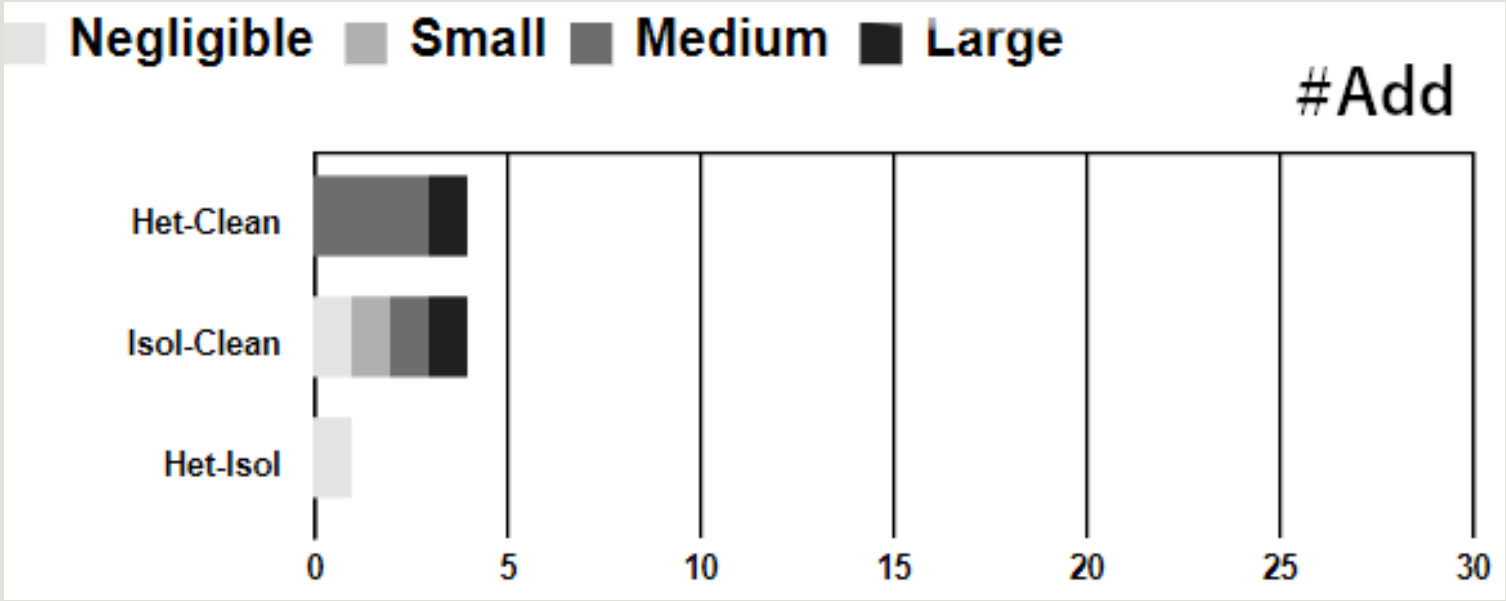
Dataset	Mod. Type	Agg. Types	Cliff's Delta
No Separation	#Add	Het-Hom	0.37
		Hom-Clean	-0.47
Modified	#Add	Het-Isol	0.11
		Het-Clean	0.27
		Isol-Clean	0.13
	#Del	Het-Isol	0.12
		Het-Clean	0.17
		Hom-Isol	0.2
		Hom-Clean	0.26
	Churn	Het-Clean	0.12
		Het-Clean	0.27
		Isol-Clean	0.13
Dead	#Add	Het-Isol	0.73
		Het-Clean	0.81
		Isol-Clean	0.47
	Churn	Het-Isol	0.73
		Het-Clean	0.81
		Isol-Clean	0.47

Lines of Modified Code – General Perspective – Class Type History

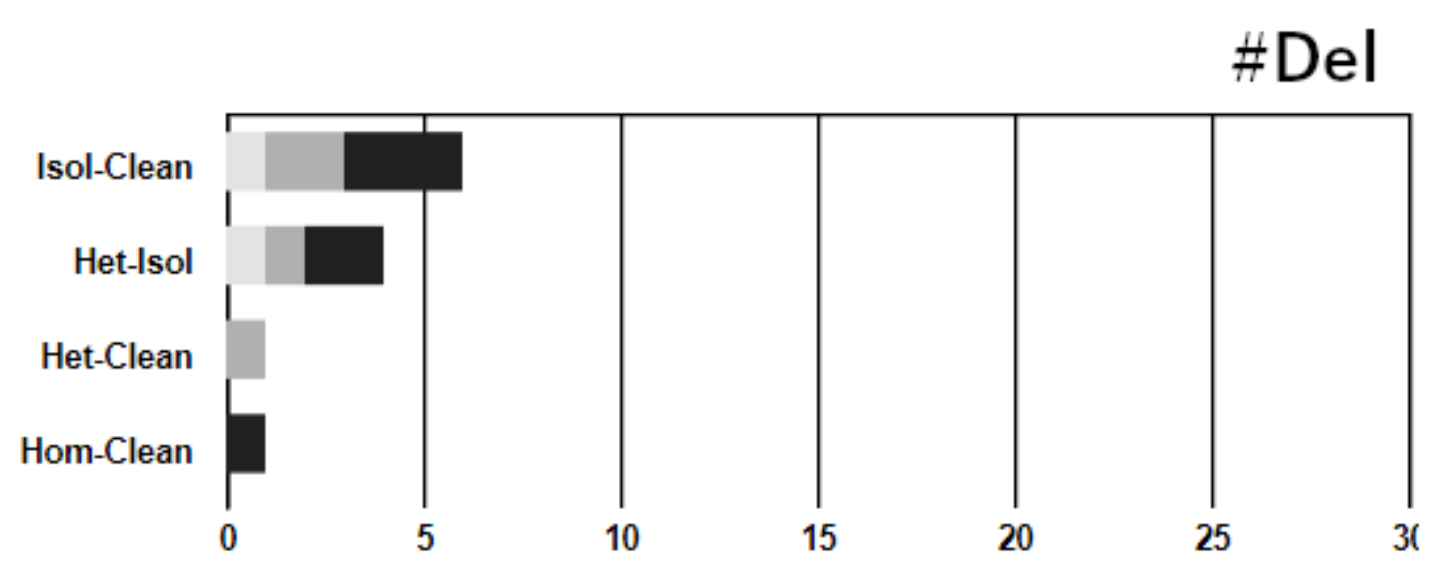
GENERAL PERSPECTIVE OF MODIFICATION TYPES BY CLASS TYPE HISTORY

Dataset	Mod. Type	Agg. Types	Cliff's Delta
No Separation	#Add	Het-Hom	0.37
		Hom-Clean	-0.47
Modified	#Add	Het-Isol	0.11
		Het-Clean	0.27
		Isol-Clean	0.13
	#Del	Het-Isol	0.12
		Het-Clean	0.17
		Hom-Isol	0.2
		Hom-Clean	0.26
	Churn	Het-Clean	0.12
		Het-Clean	0.27
		Isol-Clean	0.13
Dead	#Add	Het-Isol	0.73
		Het-Clean	0.81
		Isol-Clean	0.47
	Churn	Het-Isol	0.73
		Het-Clean	0.81
		Isol-Clean	0.47

Modified Dataset – By System - #Add

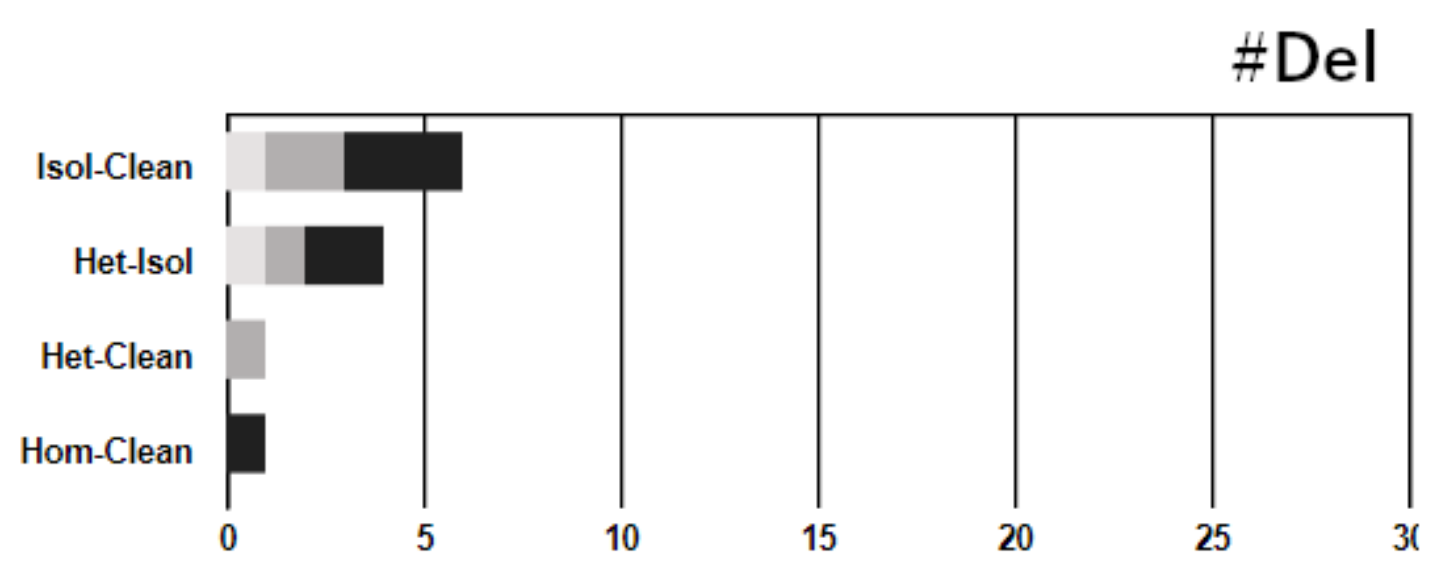


Modified Dataset – By System - #Del



lab-soft
software engineering laboratory

Modified Dataset – By System - Churn



Answering RQ1

RQ1: Do Heterogeneous and Homogeneous Agglomerations undergo more frequent changes compared to Isolated and Clean types? For the General dataset, smelly classes change more frequently than Clean ones. We also found that Heterogeneous agglomerations change more frequently than other agglomeration types. For number of commits, we found evidence favorable to the Clean classes being unstable.

Intensity of Changes – General Dataset

INTENSITY OF CHANGES - NO SEPARATION OF GENERAL DATASET

Agg. Types	Cliff's D		
	#Add	#Del	Churn
Het-Hom	-0.27		-0.08
Het-Isol	0.26	-0.11	
Het-Clean	0.99	0.89	0.95
Hom-Isol	0.55		
Hom-Clean	1.0	0.75	0.9
Isol-Clean	1.0	0.75	0.9

Intensity of Changes – General Dataset

INTENSITY OF CHANGES - NO SEPARATION OF GENERAL DATASET

Agg. Types	Cliff's D		
	#Add	#Del	Churn
Het-Hom	-0.27		-0.08
Het-Isol	0.26	-0.11	
Het-Clean	0.99	0.89	0.95
Hom-Isol	0.55		
Hom-Clean	1.0	0.75	0.9
Isol-Clean	1.0	0.75	0.9

Answering RQ2

RQ2: Do Heterogeneous and Homogeneous Agglomerations undergo changes in more intensity compared to Isolated and Clean types? In the General perspective, we could observe that smelly classes tend to change in more intensity in the three modification types, with the presence of Large Effects. We could also observe similar results for the Modified and No Separation dataset. Meanwhile, when observing the results by system, for all three modification types we could not reject H0 for more than 50% of the systems. We provide evidence that, in general, smelly classes change in more intensity than clean ones.

Threats to Validity

Detection of smells at only one system release;

Two year mining timespan;

Systems selected;

Effect of system size.

Conclusions

- We provide evidences that Agglomerations, in general, change more frequently and in more intensity than Isolated and Clean classes;
- Agglomerations should be prioritized when refactoring the code.

Future Works

- Identify which agglomerations of the Heterogeneous and Homogeneous changes with more frequency and intensity
- Further explore other measurements of stability;
- Explore the impact of agglomerations on coupled classes;
- Explore changes in terms of smelly methods.

Thank you! :)

AMANDADS@DCC.UFMG.BR