**STRUCTURAL METRICS MINING**

**SAMPLING**

**MODEL VALIDATION**

**PROJECT QUALITATIVE AND QUANTITATVE ANALYSIS**

**PROJECT SELECTION**

**SECTION**

**SECTION**

**SECTION**

**DATA MINING**

**SAMPLE ANALYSIS**

**UNIVARIATE TEAM SIZE ANALYSIS**

**MULTIVARIATE TEAM SIZE ANALYSIS**

**TEAM SIZE**

**STRUCTURAL METRICS**

**TEAM SIZE**

**STRUCTURAL METRICS**

**STRUCTURAL METRICS**

**TEAM SIZE**

**CONFOUNDING FACTORS**

* **Select project sample**
* **Retrieve source code for each project revision and conduct static code analysis to mine structural metrics**
* **Analyse sample to ensure it is representative of forge**
* **Analyse team size trends**
* **Analyse structural metric distributions and correlations**
* **Bucket project-level metrics by committer count**
* **Carry out bucket comparison analysis (Mann-Whitney tests)**
* **Analyse relationship between the team size and metric values**
* **Analyse code volume and revision counts as confounding factors**
* **Bucket class-level metrics by team size and revision count to mitigate impact of functional complexity**
* **Carry out bucket comparison analysis**

**4.2**

**4.3**

**4.4**

**4.5**

**SECTION**

**SECTION**

**PROJECT-LEVEL ANALYSIS**

* **Data set dimensionality reduction and visualisation to support project selection**
* **Quantitatively compare metrics. Qualitatively compare committer activity and source code**
* **Perform linear regression on individual projects**

**4.6**