

Pre-processing Steps

Normalize tokens

Convert to Source to Graph
(AST, CFG or DDG)

Build good set using unit tests

Remove comments

0

20

40

60

80

100

Proportion (%)

