Program Similarity Measurement for Evaluation of Students' Programs
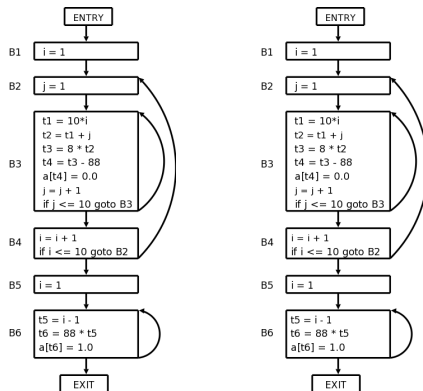
Mladen Nikolić   Milena Vujošević-Janičić

Fifth Workshop on Formal and Automated Theorem Proving and Applications, Belgrade, 2012.

Program evaluation tasks

- Testing and verification check for functional correctness and bugs
- Modularity — form rather than correctness
- Structural simplicity

Control flow graphs

- Control flow graph represents the structure of a program
- Nodes of the graph contain chunks of code
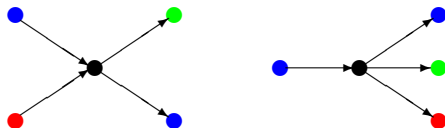
- Program similarity = content similarity of nodes + topological similarity of nodes/CFGs

- Edit distance
- Similarity of instruction influence graphs

- Iterative calculation according to (Nikolić, 2013)

- In similarity $= 1/2$
- Out similarity $= 2/3$
- Total similarity $= (1/2+2/3)/2=7/12$
- How to define similarity of whole graphs?

- How to combine two types of similarities?

- Similarity of CFG nodes enables matching of parts of programs
- Similarity of CFGs reflects the overall similarity of programs

- We wished to confirm that our CFG similarity measure reflects some intuitive similarity of programs
- 15 problems — more than 200 solutions
- Classification of solutions to corresponding problems
- Classification accuracy 80%

- Richer corpus
- Different content and topological similarity measures
- Web tool