Towards Scalable Identification of Motifs Representing Non-Determinism in HPC Simulations

Ali Y Khan¹, Sanjukta Bhowmick¹ (advisor), Michela Taufer (advisor)²

¹University of North Texas. ²University of Tennessee Knoxville

Contributions:
Scalable Motif based Network Alignment
Compute Multiple Motifs per Execution
Sensitive to Changes in Network
Adapted to Graph Structure

Network Alignment
Event graphs representing large HPC simulations can differ slightly across executions, even when the execution parameters remain unchanged. Identifying these slight differences can help in identifying the regions of non-determinism in the code, and subsequently lead to more reliable and reproducible software.

Current network alignment algorithms cannot effectively identify the difference in nearly similar graphs.

We present a motif based network alignment algorithm that focuses on identifying subtle differences in graphs.

Algorithm for Computing GDV
1. Create Initial Subtrees
   - Pick root node from graph and create root-neighbor subtrees

2. Subtree Enumeration
   - Create copies of subtree and expand all combinations of end nodes

3. Motif Generation
   - Add all combinations of valid backedges to generate motifs
     a) subtree is valid if neighbor of root > end node
     b) back edge is valid if connects 2 lowest nodes in cycle

4. Calculate GDV
   - Use distance and degree vectors to identify corresponding orbit and update GDV

Graphlet Degree Vector
Graphlet (or motifs): structures formed of small number of vertices.
Orbit: The position of a vertex in the graphlet

Graphlet Degree Vector (GDV): Number of times a vertex appears in a given set of orbits

Acknowledgements
We are grateful for the assistance from Krishna Sai Ujwal Kambhumpati, Nigel Tan, and Patrick Bell. Support of XSEDE and IBM through a Shared University Research Award is also acknowledged.

This work is supported by NSF grants 1900765 and 1900888.

References