

## **SAGA: A FAST AND FLEXIBLE GRAPH MATCHING TOOL**

Tian, Yuanyuan<sup>1</sup>; McEachin, Richard C.<sup>2</sup>; Patel, Jignesh M.<sup>1</sup>

<sup>1</sup>Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI;

<sup>2</sup>National Center for Integrative Biomedical Informatics, University of Michigan, Ann Arbor, MI

**Keywords:** Subgraph matching, indexing, node/edge

Graph comparison queries provide a powerful method for querying and mining biological graph datasets, such as pathways and protein interaction networks. However, due to the noisy and incomplete characteristics of these datasets, instead of exact subgraph matching, approximate subgraph matching is more useful. Approximate matching queries require a powerful and flexible model that allows for approximate matching based on node labels, edge weights, and graph structural components. Existing graph matching models are too restrictive as they only allow exact subgraph matching, or have limited approximate match models (such as only allowing node insertions or deletions). We have developed a novel approximate subgraph matching technique called SAGA, which employs a flexible model for computing graph and subgraph similarity. In addition, we have designed an indexing method for efficiently evaluating these approximate subgraph queries. Using a prototype implementation of our algorithm on top of the PostgreSQL database engine, we show that our model can produce biological meaningful matches, while existing methods fail. Our prototype implementation also demonstrates the efficiency of our method.

This research was supported by the National Institutes of Health Grant # U54-DA021519, National Center for Integrative Biomedical Informatics