Abstract
SAGA, the Substructure Indexed-based Approximate Graph Alignment tool and TALE, a Tool for Approximate Subgraph Matching of Large Queries Efficiently, allow users to match query graphs against a large database of graphs. The biological application of SAGA/TALE allows users to query and compare biological pathways against the KEGG pathway database. Here we describe a Cytoscape plugin that sends query graphs to SAGA/TALE and retrieves the approximate matching graphs. Cytoscape, an open source platform for visualizing molecular networks is an ideal input and display framework for SAGA/TALE.

Motivation
- Why a Cytoscape Plugin Interface to SAGA/TALE?
  - Eliminates the context shift between Cytoscape and Web page diagrams
  - Can automatically send to SAGA or TALE depending on the node/edge count of the input graph
  - No need for human formatting of the input graph
  - Greater user interactivity
  - Input and output graphs can be easily manipulated in Cytoscape
  - Multiple layout styles for the output graph
  - Link out to other knowledge-bases from the output graph nodes and edges

SAGA – A fast and flexible subgraph matching tool
- Motivation
  - Graphs provide a powerful primitive for modeling biological data
  - Most real life data sets are noisy and incomplete in nature: so exact matching does not produce useful results
  - Need approximate graph matching
- Index-based Matching Algorithm
  - Build and index on small graph substructures in the database
  - Use the index to match fragments of the query with fragments in the database, allowing for various types of mismatches
  - Assemble larger matches using a graph clique detection algorithm
- Cytoscape support: Modified output from DOT to XGMML format

TALE – A tool for approximate large graph matching
- Motivation
  - SAGA is very efficient for querying relatively small graphs, but becomes prohibitively expensive for querying large graphs
  - Biological graphs are becoming larger (100s ~ 1000s of nodes & edges), need approximate large graph matching — TALE is able to handle very large graph queries
- The Novel Matching Paradigm
  - Distinguish nodes by their relative importance in the graph structure
  - Match the important nodes in the query graph
  - Extend the matches progressively by enclosing nearby nodes of already matched nodes
- Cytoscape support: Modified output from GML to XGMML format

Cytoscape – A tool for visualizing networks
- What is Cytoscape?
  - An open source bioinformatics software platform for visualizing molecular interaction networks and biological pathways
  - Cytoscape core distribution provides a basic set of features for data integration and visualization
  - Additional features are available as plugins
  - Provides an open API based on Java for third-party plugin development
- We have developed a plugin that uses Cytoscape as the input mechanism and visualization framework for SAGA/TALE graph matching queries

Integrating SAGE, TALE, and Cytoscape through a Web Service

Screenshots of Cytoscape Interface
- Screenshot of Query Graph Input in Cytoscape
- Screenshot of Match Graph Display in Cytoscape

Challenges and Future Directions
- Graph Layout
  - How best to layout the graphs in a biologically meaningful way
  - There are dozens of layout algorithms available in Cytoscape usually selected through “trial and error”
  - Solution: create a dedicated SAGA/TALE layout based on User feedback and preferences, but can still use the built-in layouts
- A Large Number of Resulting Match Graphs
  - How best to allow the user to choose relevant Match Graphs; there can be hundreds of resulting Match Graphs
  - Solution: Cluster the resulting Match Graphs and display in a way that lets the user easily filter out the uninteresting ones and focus on those that are relevant; use input parameters to limit/narrow results

Acknowledgements
This work was supported by National Institutes of Health: Grant #U54 DA021519