Implementing GEX: A Framework for approximate matching of complex queries on graph data and evaluating with other graph algorithms

Vishwas Tantry (vxt1640@g.rit.edu)
Advisor: Dr. Carlos Rivero
Rochester Institute of Technology

INTRODUCTION
➢ Graph database is an unique solution in managing rich, highly connected and high volume data.
➢ Storage and representation is just half of the work done, the other half is finding effective ways to extract knowledge out of this stored data corpus.
➢ Need for algorithms that extract relevant data efficiently.

GOAL
➢ To implement Gex, a graph algorithm framework that is used for approximating complex queries.
➢ Implementing the scoring function and Top-K algorithm.
➢ Identify queries that can be processed efficiently with Gex.
➢ Compare Gex with Nema for various scenarios.

IMPLEMENTATION
The following modules were implemented
➢ Edge matching algorithm to generate candidate edges using Levenshtein and Jaccard distance
➢ Cursor generation module to assign candidate edges to appropriate query edges
➢ Cursor selection module to select the next appropriate cursor for answer generation
➢ Compute Answer module to generate all possible solution given a candidate edge
➢ Scoring function module to assign score to every solution generated
➢ Top - k module to select top K best solutions based on the scores

SAMPLE SCENARIOS

<table>
<thead>
<tr>
<th>Data Graph</th>
<th>Query Graph</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Sample Scenario 1" /></td>
<td><img src="image2" alt="Sample Scenario 1" /></td>
<td>Nema Output: {0=0, 1=1, 2=2} -- 0.0</td>
</tr>
<tr>
<td><img src="image3" alt="Sample Scenario 2" /></td>
<td><img src="image4" alt="Sample Scenario 2" /></td>
<td>Gex Output: {0=0, 1=1, 2=3, 3=3, 4=2} -- 1.44</td>
</tr>
<tr>
<td><img src="image5" alt="Sample Scenario 3" /></td>
<td><img src="image6" alt="Sample Scenario 3" /></td>
<td>Gex Output: No solution</td>
</tr>
</tbody>
</table>

CONCLUSION AND FUTURE WORK
➢ Gex provides better solution in most of the scenarios compared to Nema.
➢ Gex is more suited for graphs which conform to RDF structure. For other types of graphs the time complexity increases.
➢ Future work would involve testing and comparing algorithms with different types of graph database.

REFERENCES