Background Research

Graph Mining Domains:
1. Frequent subgraph mining
2. Approximate graph pattern mining
3. Graph pattern summarization
4. Graph classification
5. Graph clustering
6. Graph indexing
7. Graph searching
8. Correlated graph pattern mining
9. Optimal graph pattern mining
10. Graph kernels
11. Link mining
12. Web structure mining
13. Workflow mining
14. Biological network mining

Algorithm (Contd.)

3. Lexicographical ordering of DFS codes

4. Minimum DFS Code

5. Perform dfs on DFS code tree

Example

Graphs=2, Frequency=2, closed graph

Figure 2: A simple example of patterns mined from 2 graphs

RESULTS

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Gspan vs Gaston (dense edges)</th>
<th>Gspan vs Gaston (sparse edges)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CONCLUSIONS

It is very clear from the results that gSpan works faster than the other branch and bound candidate graph generation algorithm due to DFS codes introduced. It is also clear that gspan mines more relevant subgraph patterns as compared to gaston as it allows performing closed mining.

FUTURE WORK

Building approximate graph mining on top of frequent subgraph mining to add approximation to the mined patterns which is required due to the noise and the diversity of the data. Handle complex data such as programs data where each node is a complex structure.

REFERENCES

1. X. Yan and J. Han. gSpan: Graph-based substructure pattern mining, UIUC CS Tech. Rep. R-2002-2296, a 4-page short version in...