On Isomorphic Matching of Large Disk-Resident Graphs using an XQuery Engine

Carlos R. Rivero and Hasan M. Jamil
University of Idaho
Dept. of Computer Science
Postdoctoral researcher (Moscow, ID)
Roadmap

1. Introduction
2. Graphlets and Minimum hub covers
3. Graph matching using XQuery
4. Experimental results
5. Conclusions
Increasing interest on graphs: domains
Increasing interest on graphs: projects
Our focus: graph matching

Query graph

Data graph

\[ x_1 = u_1, x_2 = u_2, x_3 = u_3, x_4 = u_4, x_5 = u_5, x_6 = u_6, x_7 = u_7, x_8 = u_8 \]
Unit of query processing

- VF2, GraphQL, QuickSI, GADDI, TurboISO
- SPath
- STwig
Question: more complex unit?
Roadmap

1. Introduction
2. Graphlets and Minimum hub covers
3. Graph matching using XQuery
4. Experimental results
5. Conclusions
Graphlets

<graphlet numberOfNeighbors="2"
    numberOfBoundaries="1">
    <node id="x_1" />
    <neighbor id="x_3" />
    <neighbor id="x_4" />
    <boundary u="x_3" v="x_4" />
    <label id="x_1" value="d" />
    <label id="x_3" value="b" />
    <label id="x_4" value="c" />
</graphlet>

Graphlet $x_1$
Graphlets

```
<graphlet numberOfNeighbors="4"
          numberOfBoundaries="2">
  <node id="x_3" />
  <neighbor id="x_1" />
  <neighbor id="x_2" />
  <neighbor id="x_4" />
  <neighbor id="x_5" />
  <boundary u="x_1" v="x_4" />
  <boundary u="x_2" v="x_5" />
  <label id="x_1" value="d" />
  <label id="x_2" value="c" />
  <label id="x_3" value="b" />
  ...
</graphlet>
```
Minimum hub covers

Query graph
Roadmap

1. Introduction
2. Graphlets and Minimum hub covers
3. Graph matching using XQuery
4. Experimental results
5. Conclusions
Computing the search space

The number of neighbors of $x_2$ is 2 and the number of neighbors of $u_{11}$ is 1.
Computing the search space

<searchSpace id="x_2"> {
  for $x$ in doc("DataGraph.xml")/graph/graphlet
  where
  count($x/neighbor) >= 2 and
  count($x/boundary) >= 1 and
  some $l$ in $x$/label satisfies
    $l/@id=$x/node/@id and $l/@value="c"
  return <value id="{$x/node/@id}" />
} </searchSpace>
Select and order a minimum hub cover
Computing substitutions

for $x$ in doc("DataGraph.xml")/graph/graphlet, $x1$ in $x/neighbor$, $x2$ in $x/neighbor$, … where (some $ssX3$ in doc("x3.xml")/searchSpace satisfies $ssX3/value/@id=x/node/@id) and … count(distinct-values(($x/node/@id, $x1/@id, ...)) = 5 and (some $b0$ in $x/boundary satisfies (($b0/@u=x1/@id and $b0/@v=x4/@id) or ($b0/@v=x1/@id and $b0/@u=x4/@id))) and … return <substitution> <item var="x3" value="{$x/node/@id}/"/> … </substitution>
Computing substitutions

\[ x_3 \bowtie x_6 \bowtie x_7 \]

Data graph
Roadmap

1. Introduction
2. Graphlets and Minimum hub covers
3. Graph matching using XQuery
4. Experimental results
5. Conclusions
XQuery implementation: Yeast cliques

![Graph](image1.png)

![Graph](image2.png)

![Graph](image3.png)

![Graph](image4.png)
In-memory implementation (updated)
In-memory implementation (updated)
In-memory implementation (updated)
Roadmap

1. Introduction
2. Graphlets and Minimum hub covers
3. Graph matching using XQuery
4. Experimental results
5. Conclusions
Benefits

Databases

Graphlets

Minimum hub covers

$x_3 \bowtie x_6 \bowtie x_7$
Optimizations to XQuery engines

Push conditions

Different nodes

Indexes
Other future work

Table A  Table B

Joins

RDF
Thanks!

crivero@uidaho.edu

CARLOS R. RIVERO AND HASAN M. JAMIL
UNIVERSITY OF IDAHO
DEPT. OF COMPUTER SCIENCE