Assignment 5

Techniques for large scale data

Authors

Yossra Gharbi(gusghayo@student.gu.se)
AlejandroCorrochano(guscorral@student.gu.se)

Problem a:

1. **Install Neosemantics library**

We've used the standard APOC and GDS libraries, which we can install with a single click in the Neo4j Desktop application. On top of that, we've added the **Neosemantics** library to our stack. It is used to interact with RDF data in the Neo4j environment. We can either import RDF data to Neo4j or export property graph models in RDF format.

To install the Neosemantics library, we've downloaded the latest release and saved it to the Neo4j plugins folder.
We also need to add the following line to the Neo4j configuration file:
`dbms.unmanaged_extension_classes=n10s.endpoint=/rdf`

2. **Create uniqueness constraint**

All methods that persist data into Neo4j have a schema level prerequisite: this is the existence of a uniqueness constraint on the property uri of nodes with the label Resource. If the constraint is not present yet, all we need to do is run the following command on our DB, otherwise all RDF-importing procedures will throw an error message indicating this has to be done first.

```
CREATE CONSTRAINT n10s_unique_uri ON (r:Resource)
ASSERT r.uri IS UNIQUE;
```

3. **Setting the configuration of the graph**

Here, we've initiated the Neosemantics configuration with the following cypher procedure:

```
CALL n10s.graphconfig.init({handleVocabUris: "IGNORE"})
```

*'IGNORE': uris are ignored and only local names are kept.

4. **Construct WikiData SPARQL query**

We've constructed SPARQL queries by using the WikiData query editor. It also helps with query debugging.

5. **Choosing between Inline and Fetch**

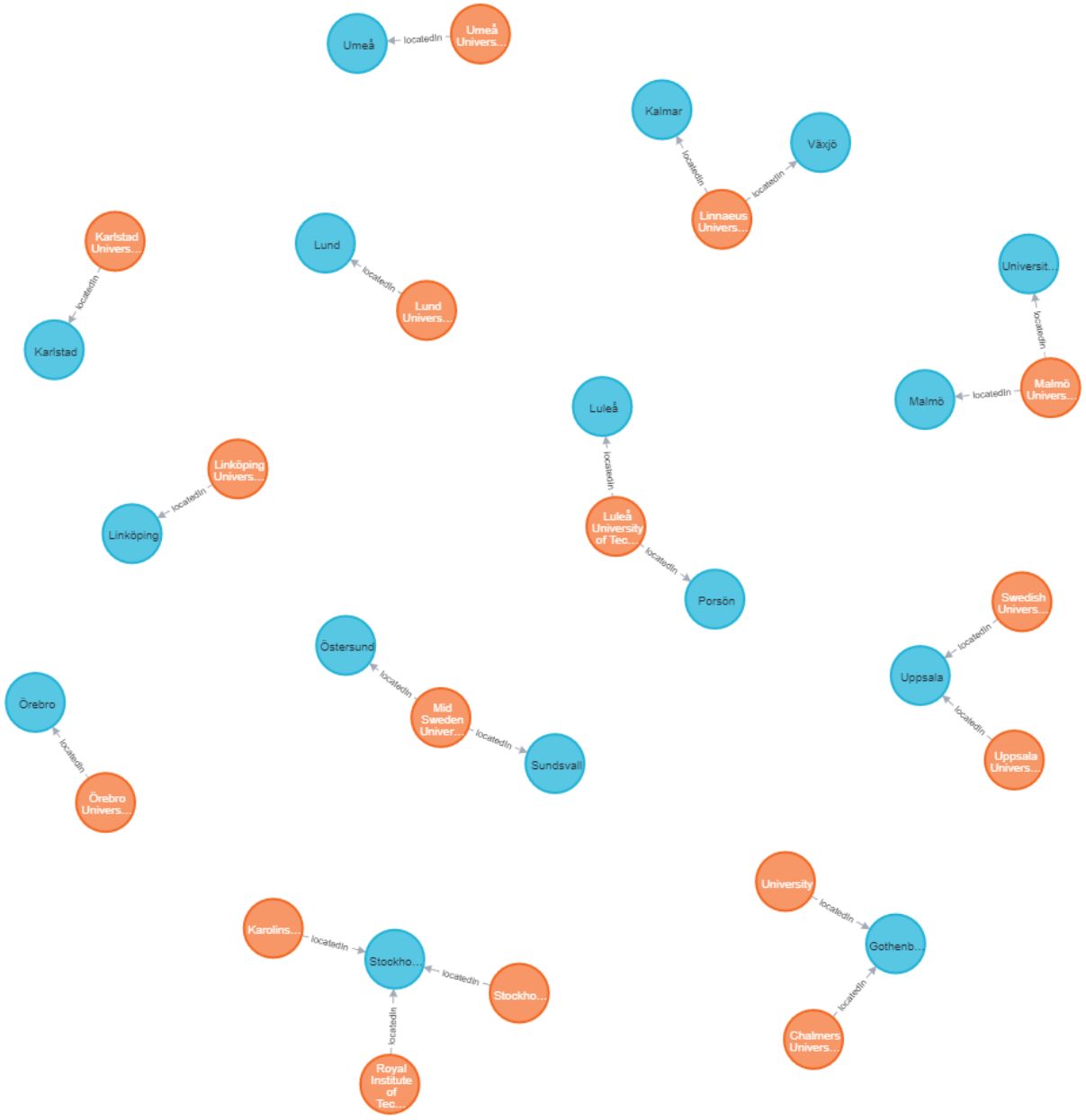
Depending on the source of the RDF query, we can run suffix each procedure with `.fetch` or `.inline`. Our RDF is hosted at a remote URL, we call the `.fetch` procedure.

6. **Final queries**

The file **problem_a** contains the final queries.

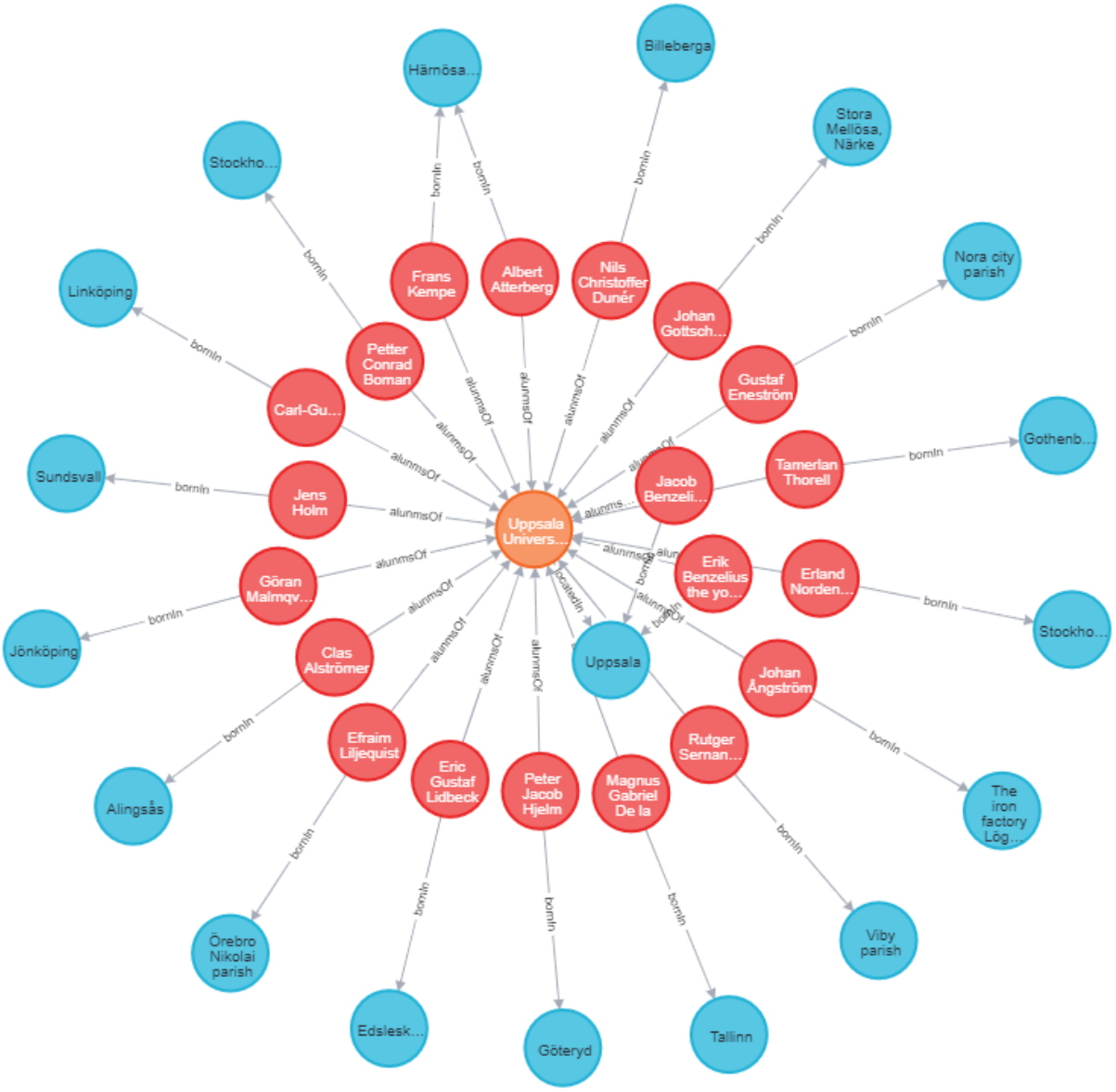
7. Previewing Data

a. *All Swedish universities*



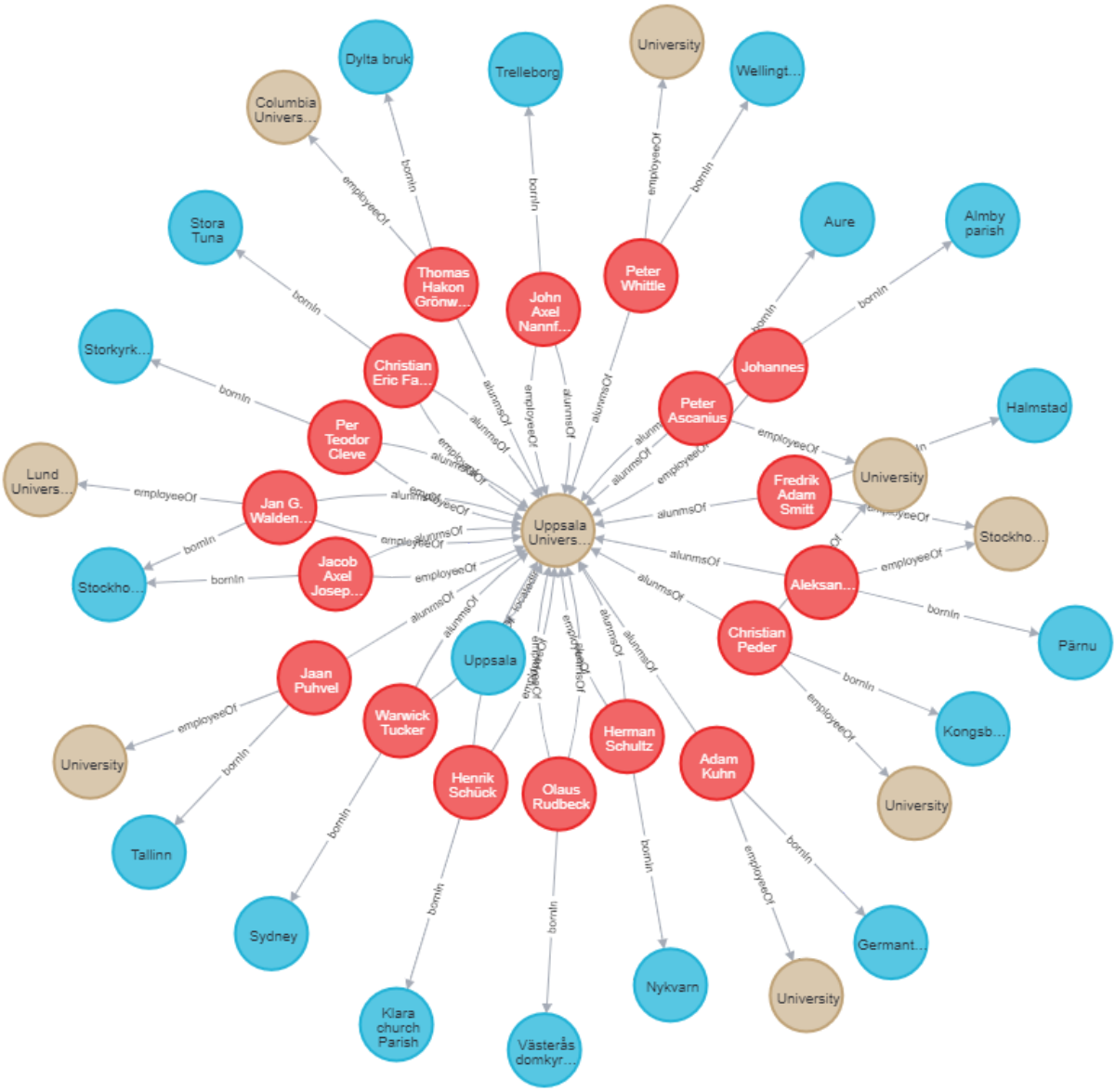
b. *All alumni of Swedish universities*

For the purpose of the graph visualization, we've limited the number of results returned from the query.



c. *All organisations (universities and others) where alumni of Swedish universities are employed.*

For the purpose of the graph visualization, we've limited the number of results returned from the query.



Problem b:

The file **problem_b** contains Cypher queries that explore the data in Neo4j.