

Knowledge Graph Analysis

Solutions to Exercise Sheet 2

Dr. Asja Fischer, Prof. Jens Lehmann

January 30, 2017

1 IN CLASS

1.
 - ▷ Labels are simple strings and assigned in order to group nodes (and query them efficiently). Properties are key value pairs and assigned to a single node.
 - ▷ Disadvantages of relational databases:
 - a) Relationships exist only between tables. / Relationships are not first class citizens in the relational data model.
 - b) Relationship traversal can become very expensive.
 - c) It is bounded by a previous schema.
 - ▷ Disadvantages of graph databases:
 - a) Not optimised for queries with no clear starting point.
 - b) Often not optimised for large batch analytics in which expensive joins (as an RDBMS does it) need to be performed for all entities anyway.
 - ▷ For creating a database in Neo4j we must assign direction to relationships, but it's not mandatory in querying in Cypher.
 - ▷ Model example:
 - Labels: Person, Cat (Programmer)
 - Nodes: Mark, Fifi, Maria
 - Properties: Name, Age, (Job)

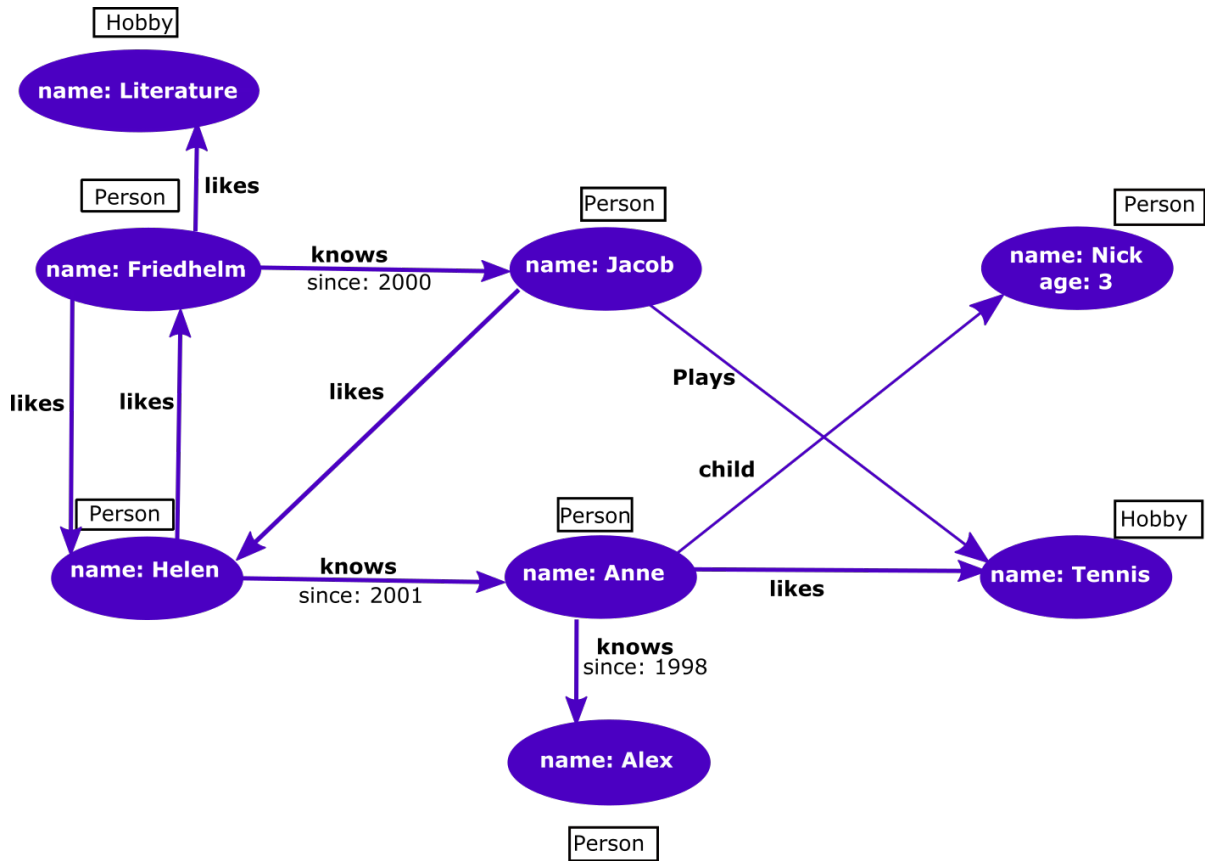


Figure 1.1: Graph of Q3.

- Relationships: Has, Married to, (Job)
2. The second model would be better. The first model has the disadvantage that almost all nodes would be assigned different labels, which is not a meaningful way to group them. In the second model, the grouping of nodes is more suitable and, as a consequence, it also allows faster queries because of the indexes over labels.
 3. See Figure 1.1.
 4.
 - ▷ **MATCH** (d)-[:DIRECTED]->(m)
WHERE m.title = "The Green Mile"
RETURN d.name
 - ▷ **MATCH** (a)-[:ACTED_IN]->(m)
WHERE m.title = "The Green Mile"
RETURN a.name ORDER BY a.name ASC
 - ▷ **MATCH** (a)-[:ACTED_IN]->(m)<-[:ACTED_IN]-(t {name : "Tom Hanks"})
WHERE m.title = "The Green Mile"

RETURN COUNT(a.name)

▷ **MATCH** (a)-[:ACTED_IN]->(m)<-[:ACTED_IN]-(t {name : "Tom Hanks"})
RETURN a.name, m.title

▷ **MATCH** (a)-[:ACTED_IN]->(m)<-[:ACTED_IN]-(t {name : "Tom Hanks"})
RETURN COUNT(DISTINCT(a.name))

Note that the query is incorrect without **DISTINCT** as “Meg Ryan” may be listed twice.

5. See Figure 1.2.

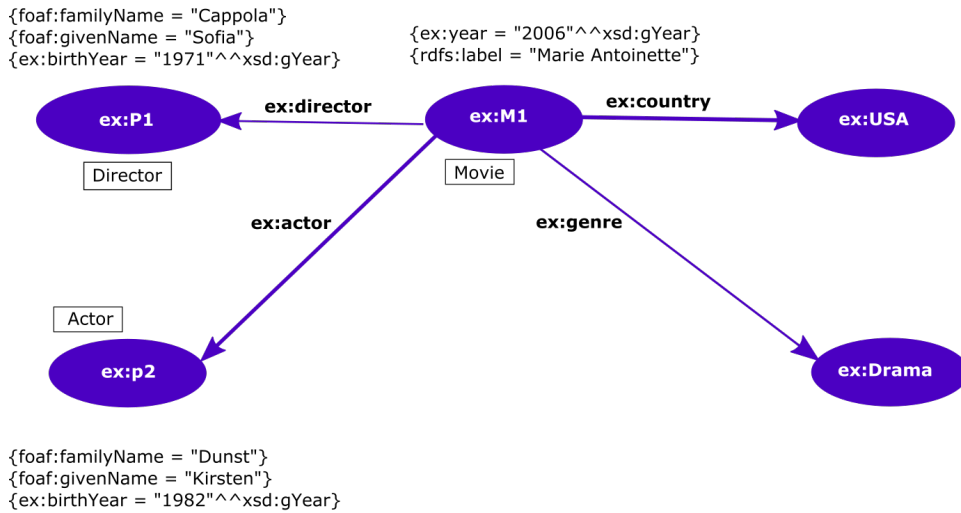


Figure 1.2: Graph of Q5.

2 AT HOME

1. No solution necessary.
2. Will be provided on request.
3. Will be provided on request.