

Analysis of Large Scale Social Networks

Exercise Session 2: Neo4J - Graph Database

Software

- <http://neo4j.org>

Cypher Reference

- <http://neo4j.com/docs/developer-manual/current/cypher>

Data Sets

- <https://www.yelp.com/dataset/challenge>

Preprocessing

See: <https://neo4j.com/docs/graph-algorithms/current/yelp-example/>

Access Neo4J Database

Running an a EC2 instance:

<https://ec2-18-197-32-13.eu-central-1.compute.amazonaws.com:7473/browser/>
<https://www.bartthijs.net:7473/browser/>

These servers don't have a SSL certificate. You'll get a warning about the self-signed certificate. Grant access to the website

User: alssn

Passwod: alssn

Exercise 1:

Run these queries and describe the output:

```
MATCH p=()-[r:FRIEND]->() RETURN p LIMIT 25
```

.....

```
MATCH p=()-[r:FRIEND]->() RETURN count(r)
```

.....

```
MATCH (n:User) RETURN count(n)
```

.....

```
MATCH (n) WHERE EXISTS(n.useful) RETURN DISTINCT "node" as entity, n.useful  
AS useful LIMIT 25 UNION ALL MATCH ()-[r]-() WHERE EXISTS(r.useful) RETURN  
DISTINCT "relationship" AS entity, r.useful AS useful LIMIT 25
```

.....

.....

Exercise 2:

Complete and draw the Graph Data Model

Label Source Node	Relation Type	Label Target Node
	FRIEND	
	REVIEWS	

Exercise 3:

Write Cypher queries to calculate or find this information or statistics

1. *Number of nodes and relations.*

.....

2. *Number of closed/open 'friend-of-friend' triangles*

.....

3. *The number of reviews written for businesses 'Cantine' and 'Emerald Chinese Restaurant'*

.....

.....

4. *The number of businesses by State*

.....

.....

5. *The number of reviews written in 2017*

.....

.....

6. *A list of users who wrote the 10 best and most useful reviews for 'Hilton Toronto'*

.....

.....

.....

.....

7. *A list of friends who wrote a review for the same business*

.....

.....

.....

.....

8. *Average stars given by users reviewing 'Arizona Steamers'*

.....

.....

.....

9. *Are there friendships between user who have given the same score to the 'Arizona Steamers'*

.....

.....

.....

10. *Similarity between categories 'Sandwiches, Salad, Restaurants, Burgers, Comfort Food' and 'Restaurants, Italian' based on number of users who write reviews for both categories.*

.....

.....

.....

.....

.....

Find other interesting questions related to this graph database, try to solve it and post the question (not yet the answer) on Toledo for the other students.

You can post your answer later this week.