

COMP3008 - Big Data Analytics

Workshop 3 - Queries in Neo4j

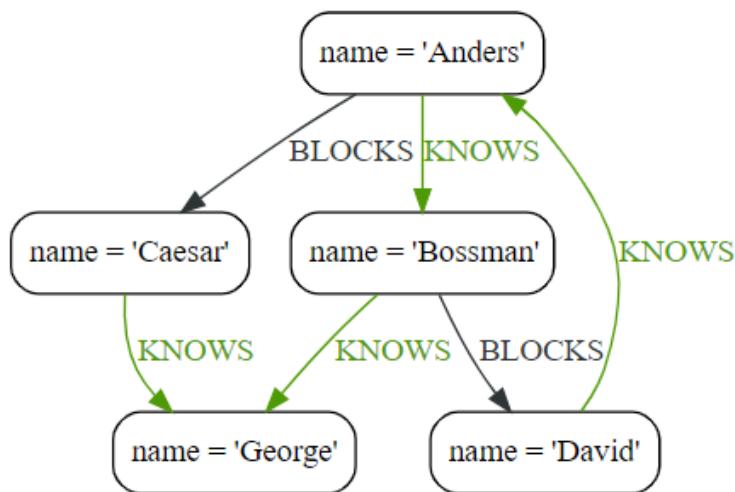
Exercise 1 - Query the dataset using MATCH()

For this first exercise, you will need to use the graph of social media data created at the end of the last workshop. Now, you will query the dataset to find the following information using the MATCH function.

1. Find the most prolific users in the collection that is, the users that published the largest number of tweets.
Make sure that you discover the precise number of tweets published by each user in the list.
2. List the five most prolific users in the collection users, and make sure that you discover the precise number of tweets published by each of them.
3. Find all the tweets published by the @EHFoundation237 Twitter account, and all the hashtags tagging such tweets.

Exercise 2

For the first exercise, you will need to recreate the figure below:



1. Find out all the people connected to Bossman.
2. Run a query using the wildcard (*) to find names and relationships between all the people.

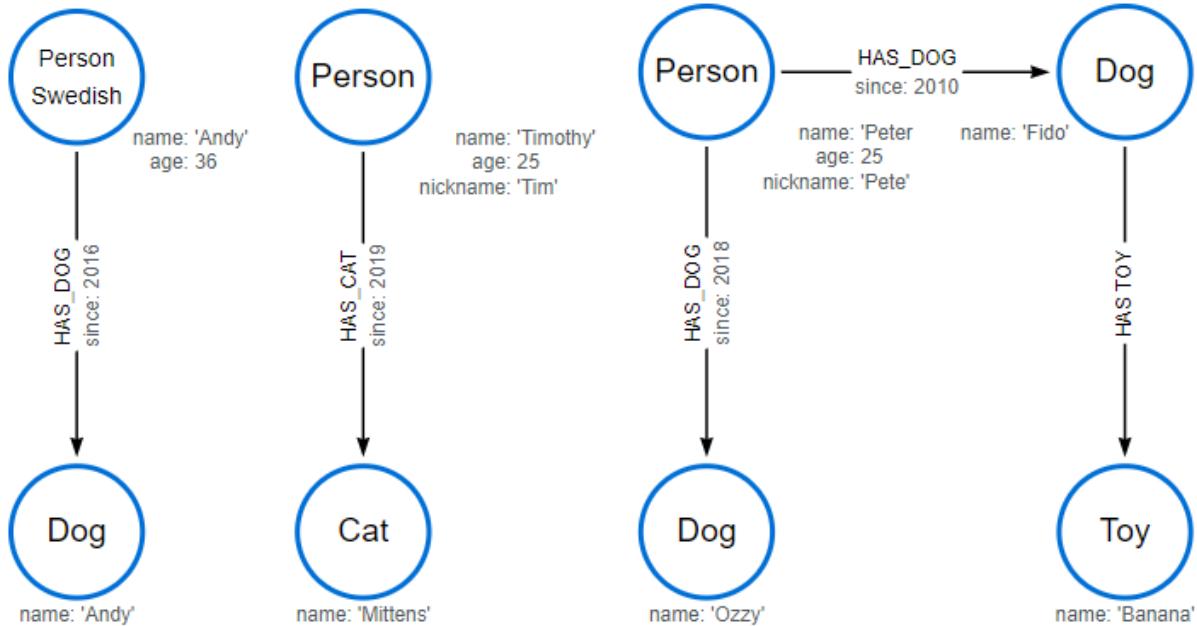
Exercise 3

Use the following code to create a graph:

```
CREATE  
(andy:Swedish:Person {name: 'Andy', age: 36}),  
(timothy:Person {name: 'Timothy', nickname: 'Tim', age: 25}),  
(peter:Person {name: 'Peter', nickname: 'Pete', age: 35}),  
(andy)-[:HAS_DOG {since: 2016}]->(:Dog {name:'Andy'}),
```

```
(timothy)-[:HAS_CAT {since: 2019}]->(:Cat {name:'Mittens'}),
(fido:Dog {name:'Fido'})-<[:HAS_DOG {since: 2010}]->(:Dog {name:'Ozzy'}),
(fido)-[:HAS_TOY]->(:Toy{name:'Banana'})
```

Be careful with the arrows if you are copying and pasting. You should get the following graph:



1. Use COLLECT to find the owner of a cat called Mittens.
2. Use COLLECT to find all the pet names.

Exercise 4

Using the Twitter datasets from last week, query the dataset to find the following information (some are the same as the first exercise, however for this exercise you need to include the WITH and COLLECT functions).

1. Find out all the users that have published more than 100 tweets.
2. Find out all the users that have published more than 900 tweets, and the tweets that each of them has published.
3. Find out the most common hashtags.
4. Find out the hashtags most used by the user @EHTFoundation237.
5. Find out the top 20 most common hashtags and the tweets that contain them.