

CENG 790: Big Data Analytics, Spring 2019

Assignment 4: Introduction to Neo4j

This assignment is part of the evaluation of CENG 790 course and it should be done individually. The deadline is 26 May 2019 23:59. The late submissions will receive a penalty of %25 for each day, starting from 27 May 2019 00:00.

You should submit a report containing the answers of the questions for each part. You may provide the screenshots of the output where necessary. You should also submit the code you produced to answer the questions. Make sure that your code is understandable, use comments extensively to explain what code piece is written for which question.

Make sure you follow the [METU Academic Integrity guide](#). Copying code from internet is also considered as cheating. Make sure the code you produced is original.

Nowadays, the links between the data are of high importance to enterprises. Many applications need real-time insights into these relationships, and not data alone. Graph databases use graph structures with nodes, edges, and properties to store and retrieve data along with the relationships. In this assignment, we will explore graph databases. As in the lectures, the database covered is Neo4j. To get some hands on experience with this NoSQL database, we'll start by installing Neo4j and then use its query language Cypher to explore an interesting dataset.

Dataset

If you follow the popular culture you already know about Game of Thrones (GoT). Let's keep up with the trends and explore how Neo4j capabilities can help in analyzing relationships between GoT characters before they are gone forever from our lives.

Andrew Beveridge and Jie Shan published "Network of Thrones" where they analyze a network of character interactions from the novel "A Storm of Swords", the third book in "A Song of Ice and Fire" and the basis for the Game of Thrones TV series. In their [paper](#) they detail how they constructed the network of character interactions by using text analysis and entity extraction to find characters mentioned together in the text. The csv file we will use has the following structure:

Source, Target, Weight where Source and Target represent characters from the books, while the weight quantifies the number of interactions between the them two.

Getting Started

Download, install and start Neo4j. <https://neo4j.com/download/>

Next, you can open the provided URL (by default: <http://localhost:7474/browser/>) in a local web browser. You can use the existing 'neo4j' account. With this setup you can use the browser for importing data, executing [Cypher queries](#), and getting preview of results in tabular or graph formats. You can follow the option "Write Code" that takes you to some sample datasets to play with. You can start by getting familiar with Cypher exploring the movie dataset.

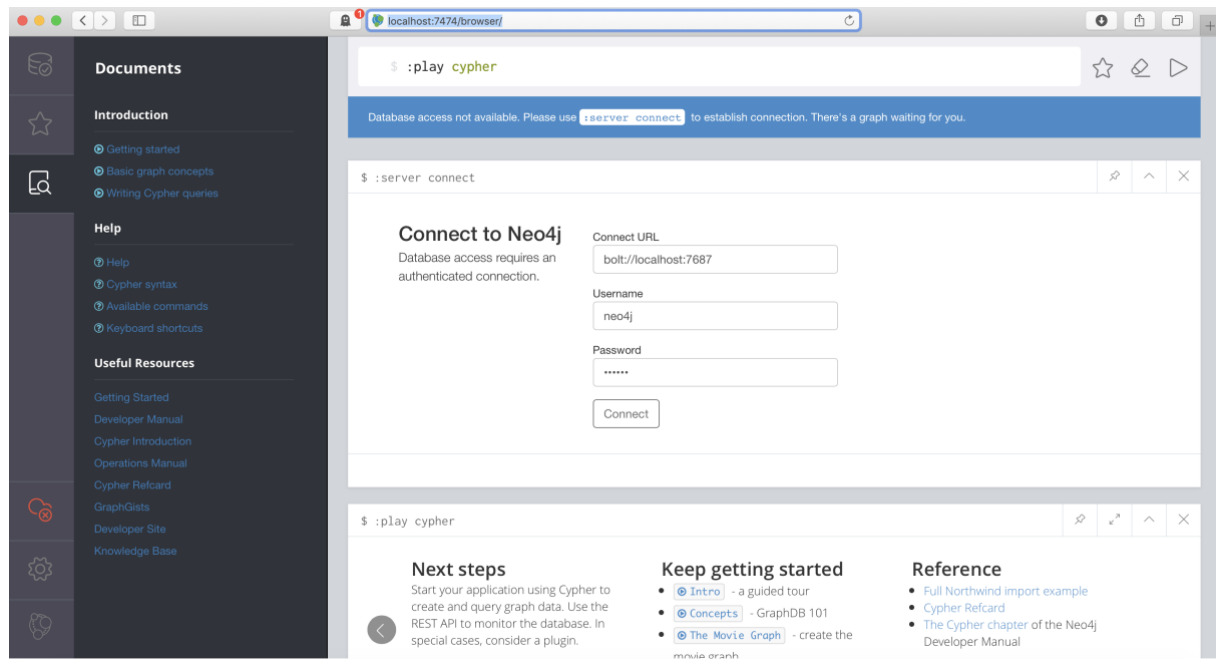


Figure 1 neo4j Browser

Task 1 Analyzing the Graph of Thrones

Report Format

There are a total of 9 questions in this assignment. In your report, please provide the query to find the answer together with the output of the neo4j screen (a partial screenshot is ok).

Import dataset

The first part of the data for this exercise will be directly imported from the authors' website. We will convert the adjacency list from the file into a simple data model (Figure 2):

`(:Character {name})-[:INTERACTS {weight}]->(:Character {name})`



Figure 2 Data model for Network of thrones

Nodes with label 'Character' represent characters from the text, and we have a single relationship type 'INTERACTS' which connects characters who have interacted in the text. We'll store the character's name as property-name on the node and the number of interactions between two characters as a property-weight on the relationships.

To assure integrity of our schema, we first create a constraint for uniqueness of the character names:

`CREATE CONSTRAINT ON (c:Character) ASSERT c.name IS UNIQUE;`

1) Does imposing this constraint improve the performance in a way other than integrity?

Once the constraint is created, we can use the Cypher LOAD CSV statement to import the data:

```
LOAD CSV WITH HEADERS FROM "https://www.maclester.edu/~abeverid/data/stormofwords.csv" AS row
MERGE (src:Character {name: row.Source})
MERGE (tgt:Character {name: row.Target})
MERGE (src)-[r:INTERACTS]->(tgt)
ON CREATE SET r.weight = toInt(row.Weight)
```

Analyzing the network

With the dataset loaded you can start exploring the graph. Start by visualizing the dataset. Next, write Cypher queries to answer the following questions (you can use the Cypher Reference Card for quick help on syntax):

2) How many characters appear in the graph?

3) Get summary statistics for the minimum, maximum and average number of characters each character has interacted with.

4) Find the shortest path from Arya to Ramsay. Consider the possibility of having few paths of same length. List all of them in the result. You can use the functions 'shortestPath' and 'allShortestPaths' provided in Cypher.

5) What is the longest shortest-path distance between any two characters?

6) How many characters are there with interaction distance 4 to Cersei Lannister?

Task 2 Expanding the Graph

Next, we'll enrich our graph by adding information on the family ties between the characters. Import the second data set using the following command:

```
LOAD CSV WITH HEADERS FROM "file:///family_ties.csv" AS row
MERGE (src:Character {name: row.character1})
MERGE (tgt:Character {name: row.character2})
MERGE (tgt)-[r:RELATIONSHIP]->(src)
ON CREATE SET r.tie = row.tie
```

With the new dataset included we extend out model to the following (Figure 3):

```
(:Character {name})-[:RELATIONSHIP {tie}]->(:Character {name})
```



Figure 3. Data model for Network of thrones and Family ties

If you visualize the graph, when you hover over links named 'relationships', you should be able to see the ties.

Write the queries to find the answers to those questions:

7) Who are Jon Snows' parents?

8) You can notice that sibling relationships are missing. Create them by finding common parents for the ties we have already in the graph.

9) Are there any children of incestuous relationships?

HAVE FUN and GOOD LUCK !