CSE 538 Graph Database and Graph Analytics

Term Project

Octavio Villalaz

27/7/2024

Problem Statement

The goal of this project is to find out how influential is the Eleventh Doctor and what is the role of his companions in the British Tv Series Doctor Who. By analyzing the dataset, we want to understand the following:

Character Influence: Determine the influence of key characters within the dataset. Characters such as the Doctor and his companions are expected to have a significant impact on the series. This influence will be tested using centrality metrics, including degree centrality and betweenness centrality. The primary research question is: "How much influence does the Eleventh Doctor has in the Doctor Who network?"

Role of Companion: Assess what impact do the Eleventh Doctor's Companions have on him? A comparison between centrality metrics with and without the companions will be used to determine the role of the companions. As well as finding out the impact on the Doctors relationship by analyzing the Eleventh Doctors direct connection before and after removing the companions.

Dataset Analysis and Graph Data Model Design

We will be using a csv file containing Source, Target, Weight, and Type columns, which represent the relationship between the characters and their score. Neo4j will be used to construct the graph database making sure the undirected relationships are made correctly. Centrality measures will be calculated to identify the most influential characters and then Centrality measure will be calculated without the companions to find the role of the companions as well as the Louvain method to get the communities with and without the companions to get information on how important they are to the Eleventh Doctor.

The expected outcomes of this study include:

- High influence of the Eleventh Doctor’s based on his centrality measures due to him being one of the main characters of the series.
- High importance of the companions to the Eleventh Doctor because they are always with him through the series.

By conducting this analysis, we aim to gain a deeper understanding of the character dynamics and structural properties of the Doctor Who network

Graph Database Implementation and Data Import

Download the dataset from Kaggle and store it in the database import folder. Then run the following Cipher script. This will load a graph containing nodes named ‘Character’ joined by edges named ‘RELATED’ with a property named ‘weight', that indicates how many times do the characters meet through the series, and ‘type’ that shows if the relationship is undirected or directed. Since Neo4j does not allow undirected graphs, we added this property to the edges although it is not used.

```
LOAD CSV WITH HEADERS FROM 'file:///doctorwho.csv' AS row
MERGE (source:Character {name: row.Source})
MERGE (target:Character {name: row.Target})
MERGE (source)-[r:RELATED {type: row.Type}]->(target)
ON CREATE SET r.weight = toInteger(row.Weight)
ON MATCH SET r.weight = r.weight + toInteger(row.Weight);
```

```
Added 694 labels, created 694 nodes, set 14824 properties, created 7065 relationships, completed after 1840 ms.
```

Exploration of Graph Database

Display the entire Doctor Who Network by running this Cipher script to make sure the data is loaded correctly.

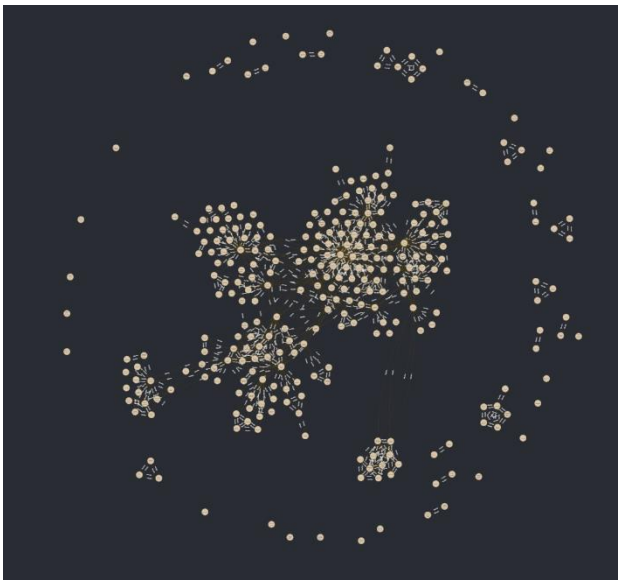
Full:

```
MATCH (n:Character) RETURN n
```

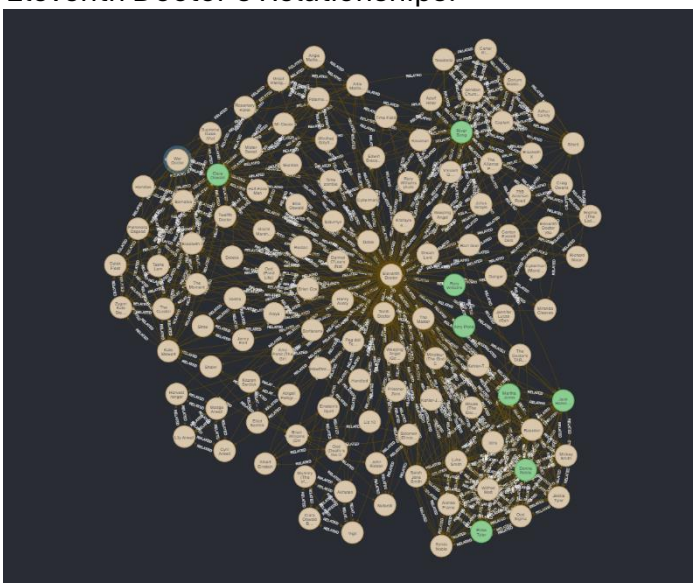
Related:

```
MATCH (doctor:Character {name: 'Eleventh Doctor'})-[r:RELATED]-(related:Character)
Return doctor, related
```

Full Network:



Eleventh Doctor’s Relationships:



Before running the algorithm scripts, we have to create an in-memory graph projection in GDS

```
CALL gds.graph.project( 'DocWho', 'Character', 'RELATED' )
```

How much influence does the Eleventh Doctor has in the Doctor Who network?"

Degree Centrality:

```
CALL gds.degree.stream('DocWho') YIELD nodeId, score
RETURN gds.util.asNode(nodeId).name AS Character_Name, score
ORDER BY score DESC LIMIT 10;
```

	Character_Name	score
1	"Fourth Doctor"	122.0
2	"The Master"	122.0
3	"Eleventh Doctor"	121.0
4	"Tenth Doctor"	115.0
5	"Third Doctor"	103.0

Betweenness Centrality:

```
CALL gds.betweenness.stream('DocWho') YIELD nodeId, score
RETURN gds.util.asNode(nodeId).name AS Character_Name, score
ORDER BY score DESC LIMIT 10;
```

	Character_Name	score
1	"The Master"	107018.36912944539
2	"Eleventh Doctor"	85436.84415046974
3	"Fourth Doctor"	63114.31115197205
4	"First Doctor"	61404.28067635812
5	"Tenth Doctor"	60930.85916315996
6	"Sarah Jane Smith"	57130.7407460754

Closeness Centrality:

```
CALL gds.closeness.stream('DocWho') YIELD nodeId, score
RETURN gds.util.asNode(nodeId).name AS Character_Name, score
ORDER BY score DESC LIMIT 10;
```

	Character_Name	score
1	"Tannis (Death Comes to Time)"	1.0
2	"Minister of Chance"	1.0
3	"Antimony (Death Comes to Time)"	1.0
4	"The Brigadier (Death Comes to Time)"	1.0
5	"Ace (Death Comes to Time)"	1.0

Analysis:

- Degree Centrality Test: The test shows that the Eleventh Doctor has one of the highest numbers of connections in the entire network. Meaning that the doctor is one of the most influential and central nodes in the network.

- Betweenness Centrality Test: The test show that the Eleventh Doctor has the second highest score of the entire network making the doctor work as a bridge to the other nodes.

- Closeness Centrality Test: The Test shows that the Eleventh Doctor, ranked #16 with a score of 0.451, is not one of the best nodes to spread information through the network.

Taking into consideration the three tests we can that the Eleventh Doctor has a lot of influence on the network, even though he is not the most central node in the network.

What impact do the Eleventh Doctor's Companions have on him?

To Test Centrality without companions we first need to remove the companions for the eleventh doctor.

```
MATCH (c:Character)-[r:RELATED]-()
WHERE c.name IN ["Amy Pond","Rory Williams"]
DELETE r;
```

And then create an in-memory graph projection in GDS without the companions

```
CALL gds.graph.project( 'DocWhoNoCcompanion', 'Character', 'RELATED' )
```

Degree Centrality (Without Companion):

```
CALL gds.degree.stream('DocWhoNoCompanion') YIELD nodeId, score
RETURN gds.util.asNode(nodeId).name AS Character_Name, score
ORDER BY score DESC LIMIT 10;
```

	Character_Name	score
1	"The Master"	122.0
2	"Fourth Doctor"	122.0
3	"Eleventh Doctor"	119.0
4	"Tenth Doctor"	115.0
5	"Third Doctor"	103.0

Betweenness Centrality (Without Companion):

```
CALL gds.betweenness.stream('DocWhoNoCompanion') YIELD nodeId, score
RETURN gds.util.asNode(nodeId).name AS Character_Name, score
ORDER BY score DESC LIMIT 10;
```

	Character_Name	score
1	"The Master"	108163.61322321011
2	"Eleventh Doctor"	95324.68701819518
3	"Fourth Doctor"	63030.136564437475
4	"First Doctor"	61235.58833681626
5	"Tenth Doctor"	60784.96513552069

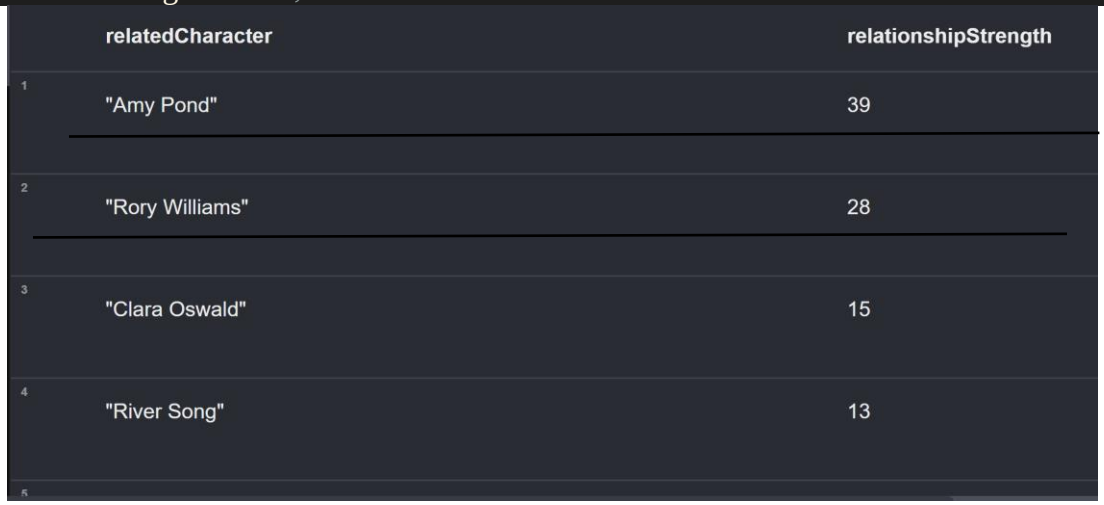
Closeness Centrality (Without Companion):

```
CALL gds.closeness.stream('DocWhoNoCompanion') YIELD nodeId, score
RETURN gds.util.asNode(nodeId).name AS Character_Name, score
ORDER BY score DESC LIMIT 10;
```

	Character_Name	score
1	"Tannis (Death Comes to Time)"	1.0
2	"Minister of Chance"	1.0
3	"Antimony (Death Comes to Time)"	1.0
4	"The Brigadier (Death Comes to Time)"	1.0
5	"Ace (Death Comes to Time)"	1.0

Eleventh Doctor's Relationships:

```
MATCH (doctor:Character {name: 'Eleventh Doctor'})-[r:RELATED]-(related:Character)
RETURN DISTINCT related.name AS relatedCharacter, r.weight AS relationshipStrength
ORDER BY r.weight DESC;
```



	relatedCharacter	relationshipStrength
1	"Amy Pond"	39
2	"Rory Williams"	28
3	"Clara Oswald"	15
4	"River Song"	13

Analysis:

- Degree Centrality Test: The Doctor's score lowers 1 point but stays in the same rank as before, implying that his companions do not lower his influence on the network.
- Betweenness Centrality Test: The Doctor's score lower but he still at the top of the list not changing his capabilities of being a bridge to other group of characters.
- Closeness centrality test also doesn't not affect the doctor's score, he still remain low on being able to spread information to other nodes.

Comparing the results from the previous centrality metrics with the metrics without the companions we can see that the Eleventh Doctor's score does lower a small amount but not enough to make a difference. Which means that even though his companions have the best relationship strength with him they are not impactful on the Doctors influence in the network. This may be because in the series the companions are always traveling with the doctor and meet the same people as him.

Conclusion

The results of this project show that the Eleventh Doctor, from the British TV series Doctor Who, is one of the most influential characters in the entire series. He manages to meet a lot of characters while traveling through space-time and thus making him a bridge for many groups of characters in the network. He travels with two Companions, Amy Pond and Rory Williams, that don't seem to impact how influential he is to the series even though they are with him through most of his time in it.

References:

Doctor Who dataset. (2021, April 19). Kaggle.

https://www.kaggle.com/datasets/manueldileo/doctor-who-dataset/data