GEPHI ASSIGNMENT REPORT

Author: Ignacio Amaya de la Peña
Tutor: Mariano Rico

Description of the problem

The objective of this work was visualize information about the fictional characters and their authors to extract some useful information from them. We would like to easily find if several fictional characters were created in the same city or in the same country to know whether or not they share similar features. This could be very useful when studying literature similarities among authors. Genres are also important, so information about which fictional characters are present in different writing styles is something we also want to know.

Due to the huge amount of fictional characters it is difficult to keep track of all those things. The graph constructed saves time to those literature researchers. Finding new authors related very fast can make them save a lot of time.

Dataset

We've used DBpedia to extract the data using a SPARQL query. The key features we wanted to extract were:

- Fictional characters
- Authors
- Nationalities of the authors
- Birthplaces of the authors
- Genre

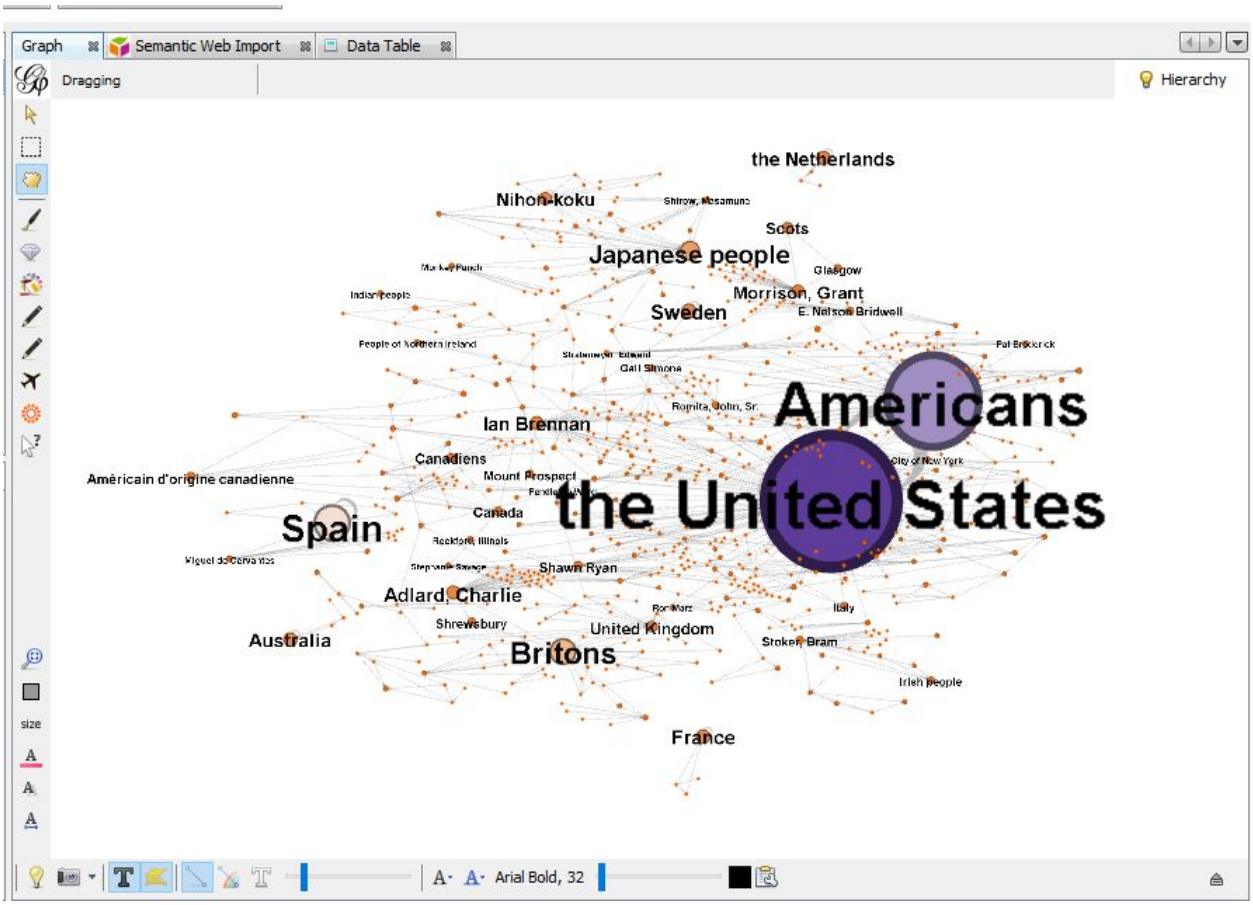
Each of the fictional characters is linked to one author. Each author was born in a city and is from a nationality. Also, each author has a genre associated with him. Each city is also linked with its own nationality. Hence, there are 5 kinds of edges and 5 kind of nodes. In total the dataset has 674 nodes and 858 edges. Not all the authors have a birthplace or a genre associated because those edges are considered as optionals.

Analysis

The data has been imported to Gephi using the SPARQL plugin that allows to use queries. That way we can easily get data from DBpedia and other linked data sources. In this case we've only

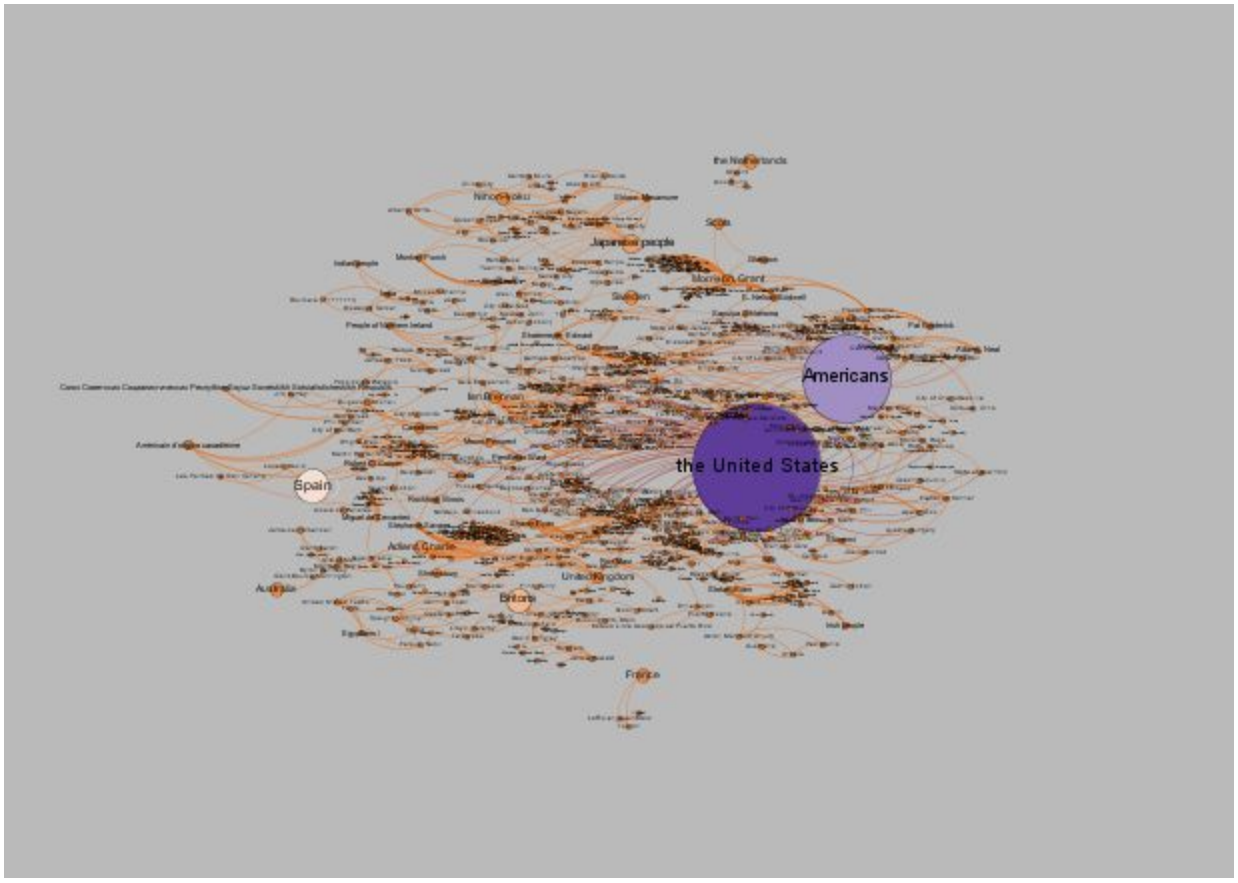
used DBpedia but it could be easily extended using other linked data sources and that is what makes it powerful.

After importing the data, we've used a combination of Layout options to make it look nicer. We also ran PageRank to score the importance of the nodes in order to color them and scale them accordingly. That way the biggest nodes have a different color. The nodes labels have also been scaled to reflect their importance. The label adjust option in the layout panel have been used to separate all the labels and make them readable.



The good thing about Gephi is that allows interaction with the graph. You can move nodes and if you position your pointer on top of a node then only its associated edges are highlighted. This allows the users an easy way for discovering information in the graph.

It is also possible to create a higher quality image. We've done that to show how it would look like, but in the results section we have used the low resolution ones to show highlighted edges.



Results

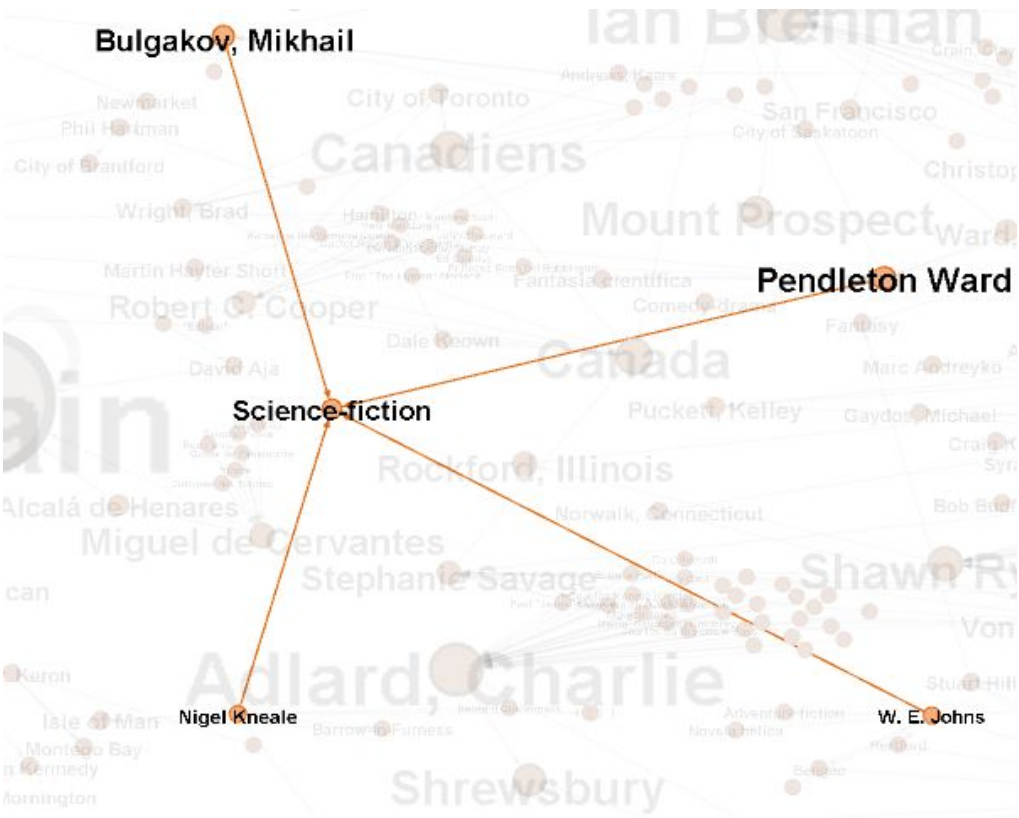
Due the high amount of data we are going to present some examples of how the information in the graph can be analyzed. In the following image we can see that in our database we have three spanish writers and two birthplaces from those writers.

If we focus on one writer, such as Miguel de Cervantes we can tell the characters he created and the place he was born.

We can also select a fictional character and instantly know who created it.

It we have a place in mind, such as Alcala de Henares we can check which authors were born there and afterwards check which characters they created. In the following example we see that in our database there are no other authors from Alcala de Henares apart from Cervantes.

We can also find a genre we want to analyze, such as sci-fi and take a look at the authors linked with it. Then we could see where are they from and maybe discover that all of them come from the same place. This way we could infer that some genres are more popular among author from some nationalities or that were born in the same places.



Conclusions

We have constructed a graph that is useful for studying literature characters and authors according to their nationalities and genres. We've shown how powerful can be working with open linked data and how can that data be easily visualized and presented, so we can extract information fast and easily. Most of the times working with these huge graphs can be challenging, but in Gephi in some minutes applying the transformations we've mentioned before we obtain beautiful graph that we can show to other people.

As it has been mentioned before the graph presented here can be augmented using other data sources other than DBpedia. This way, the number of fictional characters would increase. Adding other relationships or nodes that would enrich our graph could also be possible. In order to do that only a few modifications in the SPARQL query would be needed.