

Harry Potter Character Network

Project Topic

I decided to implement the Six Degrees of Separation theory for my project, using a dataset of every character in the Harry Potter universe and all other characters they've directly interacted with throughout the series. I thought this would be an interesting topic to work on, especially for such a vast character network like Harry Potter. The Six Degrees of Separation theory says that any two people on Earth are six or fewer links apart. With a world like Harry Potter, I think this to be very true and possibly on average, less than six links apart. My program aims to randomly select a character, analyze all its interactions and deliver an output which shows its level of degree of separation with every other character.

Dataset:

The choice of Harry Potter characters for this project was due to the large number + network of them. Unfortunately I couldn't find a dataset that met my needs but, I found this JSON :

https://github.com/dpmartin42/Networks/blob/master/Harry%20Potter/data/HP_network_clean.json

I manually made it into a table (with vertices and nodes) for every character and their interactions, then converted it into a CSV which I used for the project.

Code:

`csv_to_txt`

This module starts the data processing stage by reading data from a csv file. It also cleans the data by removing duplicates then writes the cleaned data into a txt file. In the main function, this module will output the new txt file and "CSV to TXT successful!" sentence.

`read_csv`

This module opens the csv and iterates through each line, turning it into nodes. It also filters out lines that don't contain at least two elements. Then, it returns a vector of (String, String) tuples representing edges extracted from the CSV file.

`remove_duplicates` and `write_txt`

This module uses a hash map (`unique_edges`) to store tuples (`node1`, `node2`) taken from edges, which it then inserts each edge tuple into the `HashMap`, overwriting duplicates. Its results are returned as a vector of edges.

`load_data`

This module is responsible for reading through the data and uses `node_map` to track node indices for each character. It also adds edges between each node for characters depending on their relationships. Its output in the main function is text that says “Graph loaded successfully!”

`random_sample`

This module takes a slice of items and a `sample_size` as input then uses `rand::thread_rng()` to initialize a random number generator. Its purpose is to randomly select a specified number of characters.

`six_degrees_to_all`

This is the core of the Six Degrees of Separation theory and uses the breadth-first search (BFS) algorithm to calculate the degree of separation from a starting character to all the other characters in the graph.

`main`

The main function is a compilation of the modules above, leaving the space for me to manually input the name/file path of my csv file as well as what I want the txt.file to be named as. It outputs the name of the randomly selected character and its relationships with the rest. It also outputs the time it took for the algorithm to run (in microseconds). In general, my results indicate that all characters in the Harry Potter network are indeed linked in some way. One way I would improve on this algorithm is to write a module that calculates the average degree of separation for all the characters as well as which character has the highest - it would be interesting for statistical purposes. Here's what my output looks like:

main.rs — ds210_finalproject

EXPLORER

- DS210_FINALPROJECT
 - ds_210_final_...
 - src
 - main.rs M
 - tests.rs
 - target
 - Cargo.lock
 - Cargo.toml
 - hp_character_netw...
 - hp_character... M
 - README.md

main.rs

```
1 use std::time::Instant;
```

TERMINAL

```
(base) crc-dot1x-nat-10-239-236-8:ds_210_final_proj tobioyinyoye$ cargo run
Compiling ds_210_final_proj v0.1.0 (/Users/tobioyinyoye/Desktop/ds_210_final_proj/ds210_finalproject/ds_210_final_proj)
Finished dev [unoptimized + debuginfo] target(s) in 0.68s
Running `target/debug/ds_210_final_proj`
CSV to TXT successful!
Graph loaded successfully!
Randomly selected start character: Neville Longbottom
Degrees of separation from 'Neville Longbottom':
Augusta Longbottom: 1 degrees
Angelina Johnson: 1 degrees
Rolanda Hooch: 1 degrees
Fluffy: 1 degrees
Kingsley Shacklebolt: 1 degrees
Alice Longbottom: 1 degrees
Moaning Myrtle: 2 degrees
Percy Weasley: 2 degrees
Ariana Dumbledore: 2 degrees
Kendra Dumbledore: 2 degrees
Percival Dumbledore: 2 degrees
Firenze: 1 degrees
```

Ln 1, Col 24 Spaces: 4 UTF-8 LF Rust

main.rs — ds210_finalproject

EXPLORER

- DS210_FINALPROJECT
 - ds_210_final_...
 - src
 - main.rs M
 - tests.rs
 - target
 - Cargo.lock
 - Cargo.toml
 - hp_character_netw...
 - hp_character... M
 - README.md

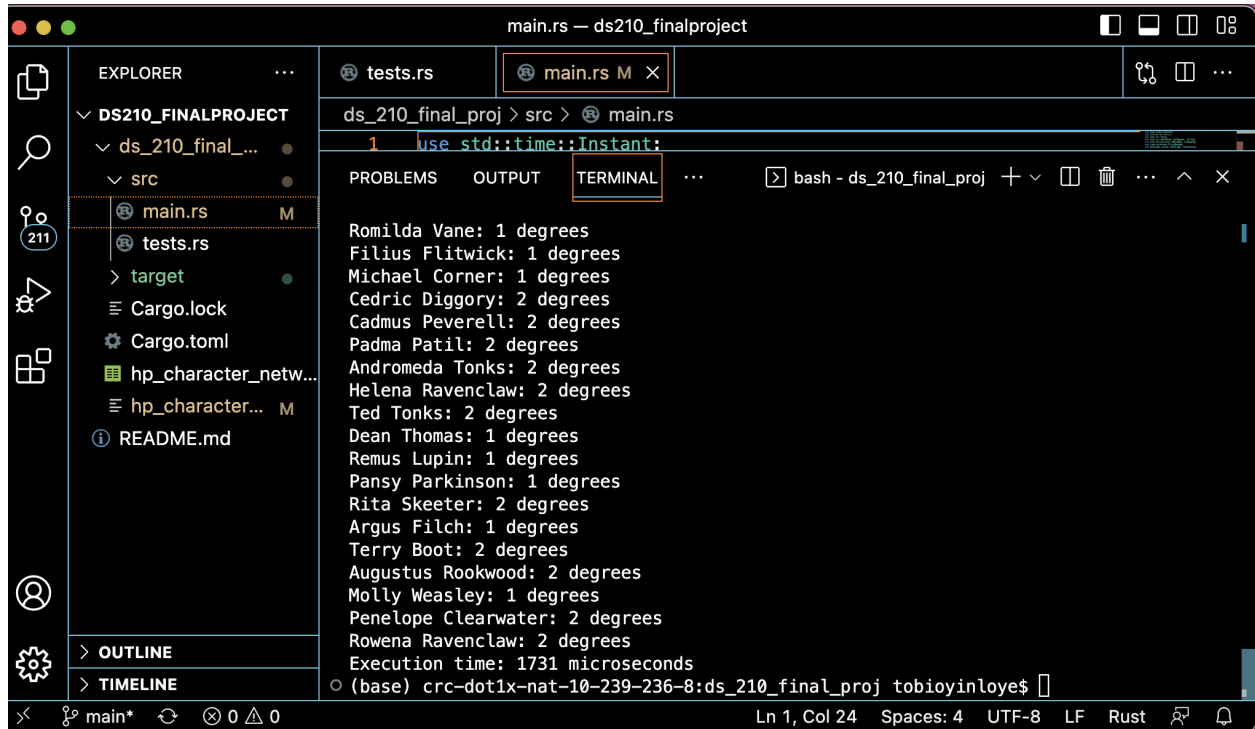
main.rs

```
1 use std::time::Instant;
```

TERMINAL

```
Wormtail: 2 degrees
Teddy Lupin: 2 degrees
Peeves: 1 degrees
Mrs. Norris: 2 degrees
Lucius Malfoy: 2 degrees
Cho Chang: 1 degrees
Irma Pince: 2 degrees
Frank Longbottom: 1 degrees
Nymphadora Tonks: 1 degrees
Zacharias Smith: 1 degrees
Nicolas Flamel: 1 degrees
Barty Crouch Sr.: 2 degrees
Stan Shunpike: 1 degrees
Charlie Weasley: 1 degrees
Rufus Scrimgeour: 2 degrees
Tom Riddle Sr.: 2 degrees
Marjorie Dursley: 3 degrees
Fang: 1 degrees
Kreacher: 2 degrees
Quirinus Quirrell: 1 degrees
Draco Malfoy: 1 degrees
```

Ln 1, Col 24 Spaces: 4 UTF-8 LF Rust



main.rs — ds210_finalproject

EXPLORER

- DS210_FINALPROJECT
 - ds_210_final_...
 - src
 - main.rs M
 - tests.rs
 - target
 - Cargo.lock
 - Cargo.toml
 - hp_character_netw...
 - hp_character... M
 - README.md

main.rs

```
1 use std::time::Instant;
```

PROBLEMS OUTPUT TERMINAL

bash - ds_210_final_proj

```
Romilda Vane: 1 degrees
Filius Flitwick: 1 degrees
Michael Corner: 1 degrees
Cedric Diggory: 2 degrees
Cadmus Peverell: 2 degrees
Padma Patil: 2 degrees
Andromeda Tonks: 2 degrees
Helena Ravenclaw: 2 degrees
Ted Tonks: 2 degrees
Dean Thomas: 1 degrees
Remus Lupin: 1 degrees
Pansy Parkinson: 1 degrees
Rita Skeeter: 2 degrees
Argus Filch: 1 degrees
Terry Boot: 2 degrees
Augustus Rookwood: 2 degrees
Molly Weasley: 1 degrees
Penelope Clearwater: 2 degrees
Rowena Ravenclaw: 2 degrees
Execution time: 1731 microseconds
(base) crc-dot1x-nat-10-239-236-8:ds_210_final_proj tobiyoinloye$
```

main* 0 0 0 Ln 1, Col 24 Spaces: 4 UTF-8 LF Rust