Oluwaseun Ademiloye

DS210

Report for Final Project

Marvel Heroes Network Analysis

Project Overview

Data Selection

The choice of Marvel hero interactions as the dataset for this project provides a rich and dynamic network to analyze. Marvel's extensive universe and interconnected characters make it an interesting domain to explore. The dataset, loaded from a CSV file, represents interactions between heroes, capturing the complexity of relationships within the Marvel universe.

Project Rationale

The decision to implement the Six Degrees of Separation concept on the Marvel hero network stems from the intrigue of exploring how closely connected or distant these iconic characters are. The Six Degrees of Separation theory suggests that any two people on Earth are six or fewer acquaintance links apart. Applying this theory to Marvel heroes allows us to uncover interesting insights into the connectivity of the superhero community.

Code Structure

csv\_to\_txt.rs

The csv\_to\_txt module initiates the data preprocessing stage. It includes functions for reading CSV data, removing duplicates of heroes, and writing cleaned edges to a TXT file. This module sets the foundation for subsequent analyses.

data\_loader.rs

Responsible for loading and transforming data, the data\_loader module ensures each hero is uniquely identified. It returns a tuple containing edges and a mapping of hero names to their corresponding identifiers (HeroId).

six\_degrees.rs

The core of the Six Degrees of Separation analysis resides in this module. The six\_degrees\_of\_separation function employs the breadth-first search algorithm to explore the degrees of separation for each hero in the network.

main.rs

The entry point for the program, main.rs orchestrates the entire process. It includes functions to build the hero network, choose a random starting node, and execute the Six Degrees of Separation analysis.

Output Explanation

Loading Data

The statement "Loading data from: hero\_network.csv" is just to show the data loading from the correct CSV file .

Loaded Edges

The vector of edges represents interactions between heroes, reflecting the complexity of relationships within the Marvel universe.

Start Node

"Start Node" indicates the randomly selected hero from which the Six Degrees of Separation analysis starts with. Starting the analysis from a random node avoids bias and provides a holistic view of the network's structure and connectivity.

Highest Degree of Separation

Identification of the hero with the highest degree of separation and the associated list of heroes reveals central figures within the network. I believe the highest degree of separation of any hero is 4, so any given Marvel superhero is at most 4 heroes away from another one.

Screenshot of Output of Function :

A computer screen with white text

Description automatically generated  
  
Screenshot of Tests Output:

A screen shot of a computer program

Description automatically generated

A screenshot of a computer program

Description automatically generated

To note:  
Something else I tried to do was print the names of the heroes associated with the IDs so we know what hero it is that has the degree of separation. However, I did not have enough time to figure out how to do this (without causing a million errors that is), so for now we only have the IDs of the heros.

Testing Approach

The inclusion of a testing suite in the `tests` directory demonstrates a commitment to ensuring code correctness. The tests are designed to validate key functionalities, including CSV parsing, data cleaning, and the Six Degrees of Separation algorithm.

---

Challenges Faced and Resolutions

Test Failures and Resolution

The challenges faced during testing, particularly the unresolved names and types, can be attributed to discrepancies in module paths or missing imports. Careful review and refinement of the `tests.rs` file, focusing on module paths and necessary imports, will likely resolve these issues. The goal is to align the test suite with the project's structure.

Code Review and Cleanup

While the code is well-organized, some unnecessary comments could be removed for cleaner code. Code comments are beneficial for understanding complex logic, but I had some unnecessary ones to debug and see what each part of my code was producing. A review of comments will enhance code readability.

---

Overall Significance

The output shows heroes with high degrees of separation, shedding light on their importance or centrality in the Marvel hero network. The Six Degrees of Separation analysis proves effective in unraveling the intricate web of connections within this complex network.

Future Improvements

Fixing the tests to work for all types of datasets is an obvious fix I want to work on. I am not sure how to implement my code to work with all types of datasets, but I will work on that later. I will also refine my code comments by removing redundant or unnecessary ones and adding explanatory comments where the code logic might not be evident.