DS 210 Final Report — Minghui Peng

About Dataset — Marvel Universe Social Network dataset
Available on Kaggle at https://www.kaggle.com/datasets/csanhueza/the-marvel-universe-social-network?select=edges.csv

This dataset includes a list of edges representing relationships between different Marvel characters, as well as the number of comics each character has appeared in. The dataset has a total of 93193 rows and contains two columns: "hero" and "comic". The "hero" column indicates which hero appears in the comic, and the "comic" column indicates which comic the hero appears in.

The result of the program is:

- The top 5 heroes in terms of percentage of comics they have appeared in are:

1. Spider Man
2. Hulk
3. Cyclop
4. Iceman
5. Iron Man

- The hero who has the most connection is: Captain America

- The number of degrees of separation between ‘SPIDER-MAN/PETER PARKER' & ‘ANT-MAN/DR HENRY J’ = 3

To achieve these results, the program does the following:

- Defines several structs to represent the heroes and the graph of the Marvel universe. The Hero struct includes a name, the number of comics the hero has appeared in, and their centrality score. The **MarvelGraph struct** represents the graph of the Marvel universe and includes a directed graph object **from petgraph crate** with heroes as nodes and relationships between heroes as edges.

- Uses the **from_edges_data() method** of the MarvelGraph struct to build a graph representation of the Marvel universe from the provided list of edges.
- Uses the **most_connected_hero() method** of the MarvelGraph struct to identify the hero who has appeared in the most comics with other heroes.
- Uses the **betweenness_centrality()** function from the **rustworkx_core crate** to calculate the betweenness centrality of each hero in the Marvel universe.
- Uses the **shortest_path()** method of the MarvelGraph struct to find the shortest path between two given heroes

In summary, this program analyzes the Marvel universe social network to identify:

- The appearance of the top 5 heroes among all Marvel comics
- Out of all those comics that have been published who was the one who appeared the most
- How many “degrees of separation” apart two Marvel characters are

It does this by building a graph representation of the Marvel universe from a provided list of edges and using centrality measures and shortest path algorithms to analyze the connectedness of the different heroes in the network. This allows the program to examine the relationships between different Marvel characters and determine how many "degrees of separation" apart they are. The program also calculates the betweenness centrality of each hero, which is a measure of the number of shortest paths that pass through a given hero, and uses this information to identify the top 5 heroes in terms of the percentage of comics they have appeared in. Overall, it can be used to understand the social structure of the Marvel universe and the relationships between its characters.