The first phase of the project began with preprocessing the given dataset (IMDb 5000 Movies) wherein, it was analyzed, and data cleanup was carried out. The networks that the project aims to build are comic book-based movies. Data cleanup involved dropping the rows (movies) which are not related to that subject and the columns (attributes) that didn’t contribute much to the analyses that the project intended to carry out. Consequently, a dataset containing only relevant information to build our network was obtained.

The dataset was split into two sets under the categories – movies with 7.0 or below on IMDb and the ones above 7.0. These two datasets were then converted into JSON files from which the subnetworks were built. Each node in the subnetwork symbolizes a movie from the datasets under consideration and the edges (links) that are provided between the nodes are done taking into consideration the production house and the comic-book universe from which the movies were derived.

The similarity score was determined by comparing the similarity found among the two subnetworks against the total possible similarities that could be obtained. The result is normalized to a scale of 10 to obtain the final similarity score.

The interactive design to compute and visualize the similarity between two subnetworks was done by allowing the user to select nodes (which essentially symbolize the movies) from the two subnetworks individually and then click the ‘Similarity’ button in order to find the commonality. This results in production of a table that presents the movies from two subnetworks being compared along with the common attribute. Additionally, this also results in computation and display of the similarity score once the common attributes are identified.

The program is submitted along with the supporting JSON files from which the subnetworks were built. It is necessary to place these JSON files in the same directory from which the source code runs. It is required to host the source file on a local web server in order to see complete execution of the project. The application could be run at the localhost:8080/HCI/index.html on the user’s browser. The project submitted comes as a bundle comprising of the source Html, CSS, JavaScript files along with some other supporting files that include the two JSON files as well.