Jingwan Li, Yuyang Wang, Branden Hidalgo

CSCI 490

Maoyuan Sun

Project 2

1. How do you build the network from this given dataset?

We randomly chose 150 movies from the original movie data set.

2. How do you find two subnetworks?

From the 150 movies in our network, we created two subnetworks on the basis of genres.

Particularly, we chose 20 movies from the action genre for subnetwork A and 20 movies from the comedy genre for our second subnetwork B. Nodes within the subnetworks are labeled with an actor name. Links within a subnetwork are created based on similar movies between two or more actors (i.e. if they played in the same movie title).

3. What is/are your metric(s) to compute similarity between the two subnetworks?

We used the plot\_keywords attribute as the metric to compare between the two data sets. The more keywords shared between the subnetworks, the greater the similarity will be. The number of similar keywords between the two subnetworks and the total number of keywords associated with the actors involved is computed. The total number of similar keywords is first divided by the total number of keywords in subnetwork A, then it is divided again by the total number of keywords in subnetwork B. These two numbers are then used to calculate an average similarity between the tables based on the nodes selected and the keywords associated with those nodes.

4. What is your design to show the similarity?

Based on the users selections,

The subnetworks each have nodes with an actor name on each one. Similarities within a subnetwork are shown by lines drawn between actor names. These lines indicate that the actors have played together in the same movie. The user clicks on the actors he/she wishes to include in the comparison. The node will change color from grey to yellow, indicating it is chosen. The nodes that contain similar information change from yellow to green. After nodes are chosen from each subnetwork, the user then clicks “SUBMIT” located at the top of the page. The page then generates two tables ONLY IF there are similarities, one for subnetwork A and one for Subnetwork B. The tables indicate the actor name, the movie title which the node is referencing, and the plot\_keywords that are associate with that movie title. If there are any similarities, the similarity score will appear at the top of the page as a percentage after “Similarity Result: “ along with the keywords that were similar, listed after “Similarity Keywords: “. Otherwise, a 0% similarity score will be shown along with a NULL output for Similarity Keywords.

5. How to run your program?

To run the program, the user will choose actor names he wishes to compare across the genres. The user can drag and drop any of the nodes and rearrange the table however he/she wishes in order to see actor-movie similarities within a subnetwork more clearly. The user may choose at least one actor from both subnetworks. If the uses makes an error (i.e. clicks an actor he/she does not wish to include in the comparison), there is a “Reselect” button above each subnetwork which will reset the user’s selections for that subnetwork whenever it is pressed. Once the user is done choosing actors to compare in both subnetworks, the user simply clicks “SUBMIT” to submit his/her selection of actors to compare. If there are any similarities, the page will show the comparison data at the top of the page. Otherwise, the page will inform the user that no similarities were found given the actor names chosen. The user can then choose to edit the selections to calculate a new similarity score, he/she can do so either by refreshing the page or by clicking any of the nodes in the subnetwork. This click will refresh the page just as the refresh button does. The user then repeats these steps as many times as he/she desires.