Final Project Proposal

For my final project, I used a dataset describing a list of characters from a fantasy novel series named The Witcher by Andrzej Sapkowski. The dataset involves the relationships between several of the characters and the number of interactions between the two characters throughout the novel series. The dataset involved 4 columns: Source, Target, Weight, and Book. The Source and Target describe which character interacted with another character within the series and the connections for this dataset were utilized as a directed graph which means that each starting node connects to another node one way. The Weight describes the number of interactions each character set had. The Book column describes which book this interaction occurred in; however, I did not utilize this column too much since I was more interested in the overall network of characters rather than focusing on the smaller networks in each of the 7 books.

Using this dataset, I was interested in analyzing which character is the most important in terms of the overall Series which is analyzed through degree centrality. Thus, I had to measure the number of outgoing relationships from a node which meant traversing each node's neighbors and computing the sum of the adjacent relationships of each node. This was computed by iterating through the edges of each node and collecting a list of how many neighbors, or in this case, character interactions, each node had. I created a subset of data that organized all the nodes' neighbors by the amount of times they have interacted with another character. I also specified in which book it took place to display clearly why there may be a repeat of character node neighbors for each character. The Degree centrality was calculated based solely on the number of neighbors connected to each node. It relays how many direct, 'one-hop' connections each node has to other nodes in the network and condones which node had the max amount of connections.

After implementing the algorithm, I discovered that the most influential node happened to be under the name Geralt who happens to be the main character in the series.

In addition, going through the output of data can show the overall flow of the dataset displaying which characters interacted with each other the most overall and confirms in which book these interactions occurred.