Justin Klein
David Mendelssohn
Zhi Xuan Ong

CS 122 Proposal

If you have ever gone to see a Marvel movie, you know that a whole lot of characters seem to just jump into the film out of practically nowhere. Vision, Black Panther, Hawkeye, Scarlett Witch, Quicksilver: these are all characters who seem to have popped onto the screen with very little background or explanation. After watching comic book movies, one often is left wondering just who these people are: are they important? Are you supposed to know who they are? What relationship do they have with the characters you recognize? Do they usually appear alongside the main characters? Not only can this distract from a film while watching it, but it can cause hours of Wikipedia searching afterward. First you search the characters you know, then the ones you do not; then you see links of people you do not recognize so you open them, and follow even more blue links until finally you realize just how long it has been since you started.

We aim to create an application that will mitigate this situation. The application will allow you to search the name of a character and it will return information about this character. Fivethirtyeight, Marvel.com, and Wikipedia all have a lot of information about each character, from eye and hair color to their current comic book status (living or dead), all of the stories, events, and issues the character has appeared in, their first appearance, and more. Beyond this, Amazon Web Services points to a dataset of Marvel appearances that can be used to create a social network for each character. For each character, there is a ranked list of other characters in descending order of concurrent appearances. This can be passed into the program Gephi in order to visualize the network of social connections for this character.

We will use the Fivethirtyeight database via sqlite3, the Marvel.com API, and webscraping on Wikipedia in order to gather enough information to create an informative and detailed report about any character's input. We can also potentially add to this if necessary by giving a list of all of the shared events, stories, or issues in which any pairing of characters appeared. One possible extension to the project includes crawling through the description of stories via the Marvel API for text analysis. We can look for the most common phrases or N-grams associated with stories of a character, or deduce common themes and trends from the words that show up frequently (e.g. war comics post WWII, LGBT themes only after 1990s).