**Final Project Proposal**

## **Gabrielle Bartomeo**

For the Final Project, I will be using sentiment analysis and network analysis to examine how the mythological characters in the Elder and Younger Eddas are portrayed and interact with one another. These books are essentially the source material for Marvel's Thor, Odin, Loki, so on and so forth.

This will be accomplished primarily using NLTK, pandas, networkx, and matplotlib. Some questions we're looking to answer are:

1. How are the characters of this text aligned based on average scoring?
2. Did these alignments meet expectations (i.e., versus how the characters that are based off Gods are portrayed in the Marvel movies)? How did they do compare to the Marvel Wikia Data provided by 538?
3. Are characters more likely to interact with their own score-based alignment than another, based on the data?

**Datasets:**

* The Elder Eddas of Saemund Sigfusson; and the Younger Eddas of Snorre Sturleson; [available for free](http://www.gutenberg.org/cache/epub/14726/pg14726.txt) on Project Gutenberg
* [Marvel Wikia Data CSV](https://github.com/fivethirtyeight/data/blob/master/comic-characters/marvel-wikia-data.csv) on Github, provided by 538

**Plan:**

1. Bring in the necessary libraries
   * Requests (URLs)
   * NLTK (text analysis)
   * Re (regular expressions)
   * Pandas (data frames)
   * Networkx (graphing)
   * Maplotlib's Pyplot (more graphing)
   * Numpy (math)
2. Bring the CSV file in from 538 using pandas
3. Get the necessary rows out of it – ALIGN is the most important column
4. Bring in the books (single text file) from Project Gutenberg
5. Format the text as needed
6. Get the sentences from the books as lists
   * One list for the Elder Edda
   * One list for the Younger Edda
   * One list with them combined
7. Use the Sentiment Intensity Analyzer to determine polarity scores for each sentence
8. Check each sentence for the presence of a deity and if so, assign the sentence itself and its polarity score to that deity
9. Create a data frame where:
   * Each row has the name of the deity, whether or not they're a deity, a given sentence, the scoring for that sentence
   * No exact duplicates are present
   * No rows without a polarity score are present
10. Create another data frame that calculates the average polarity scores and the polarity z-scores
11. Add the Earth-616 character names and alignment (from that original CSV) on as columns to this new data frame
12. Create a network graph where:
    * Nodes are either Gods or Alignment
    * Edges are present when Gods share sentences (God to God) and from God to Alignment
13. In addition to the analysis written with each graph and each step of the way, a written section going over the results and conclusions that can be inferred as a result of the analysis
14. Conclusion

**Concerns:**

1. That the professor won't find this nearly interesting enough and require me to pick another project