HW4 - Text Analytics Project

Final Write-Up

Aditya Venkataraman, Kevin Zhai, Ik-Hwan Kim

**Dialogue Text of Choice: Hamilton Musical**

The dialogue we chose to do for our final project is the lyrics from Hamilton. Hamilton is a musical written by Lin-Manuel Miranda, chronicling the life of Alexander Hamilton, one of the founding fathers of the US. The musical is unique in that it combines traditional musical themes and show tunes with the language of hip-hop to represent the American Revolution.

Our corpus spans all the lyrics of the musical, save for a few things that we note in the key assumptions below. We chose this body of text because there seems to be a lot of subtle wordplay and evolution of topics that we hope to capture in our analysis. One question we would love to answer is looking at how characters’ sentiments might change throughout the musical or perhaps analyzing which characters are more like Burr or Hamilton. Another interesting analysis would be to use topic modeling to see which topics are associated with which characters. We all love the musical, so we are hoping to delve deeper into what makes Hamilton so special.

Before we jump into our analysis, we should note the key assumptions we made about our corpus. The first thing is that we excluded some of the very minor characters, such as Samuel Seabury and Phillip Hamilton. Many of these characters only have a couple of lines of lyrics, so it would have been too little to analyze. Another thing that we are wary about is the size of the corpus for other characters -- it may turn out that we do not have enough for proper topic modeling or sentiment analysis. We will try to address these facts in our analysis.

**Characters Used:**

Alexander Hamilton

Aaron Burr

George Washington

Elizabeth Schuyler

Angelica Schuyler

John Laurens

Marquis de Lafayette

Hercules Mulligan

Thomas Jefferson

**Vocabulary Size: 2370**

**Number of Tokens: 10534**

**Analysis Sections:**

Character Classifcation - More like Hamilton or Burr? (Adit)

Topic Modelling - Assigning a Topic to each Character. (Kevin)

Topics in LDA model:

Topic #0:

angelica sister father revolution eliza word change cuz want way york schuyler freedom leave city til story whoa life thomas

Topic #1:

president ry life gon eye city mr look wait break france guy stand word want na ask hold revolution lafayette

Topic #2:

happen father city angelica home york wait day story eliza win thomas meet turn sure people way jefferson yo mr

Topic #3:

wait happen sir look jefferson die story world run rise oh alive meet na gon way whoa ry day son

Topic #4:

president stand gon cuz ha day god inside word plan france mr ry revolution ya guy assume thing york change

Topic #5:

whoa yo rise ooh father ta ya john matter freedom win want hold god word oh tonight schuyler sister great

Topic #6:

vote day god revolution alive word watch want york tonight hold lafayette eliza freedom ta oh eye gon life plan

Topic #7:

win thomas president france stand change word guy home hold ask adams great john people plan meet tonight vote assume

Topic #8:

wait na gon sir tonight em rise freedom story jefferson son ta mr hand oh til lafayette run meet watch

Topic #9:

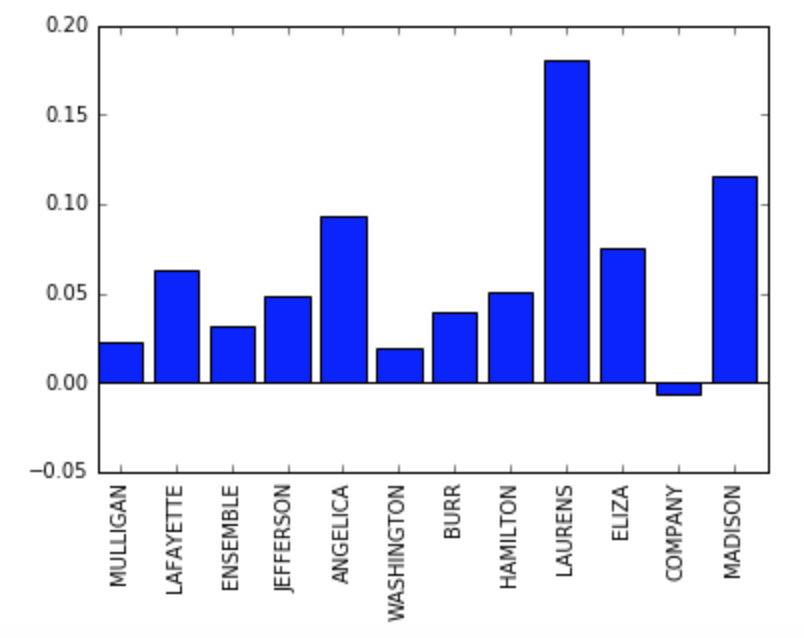
jefferson na em gon leave watch ya life alive eliza ry revolution freedom hand meet hear son sir oh city

Sentiment Analysis (Ik-Hwan)

Overall sentiment score by character:

Each lines of a character is given a score based on the words contained within those lines. Each line is passed through a pre-trained sentiment analysis model in *nltk.sentiment.vader* and the intensity scores as well as probability of the sentence being positive, negative, or neutral are returned by the model.

**Average sentiment score of lines by character**



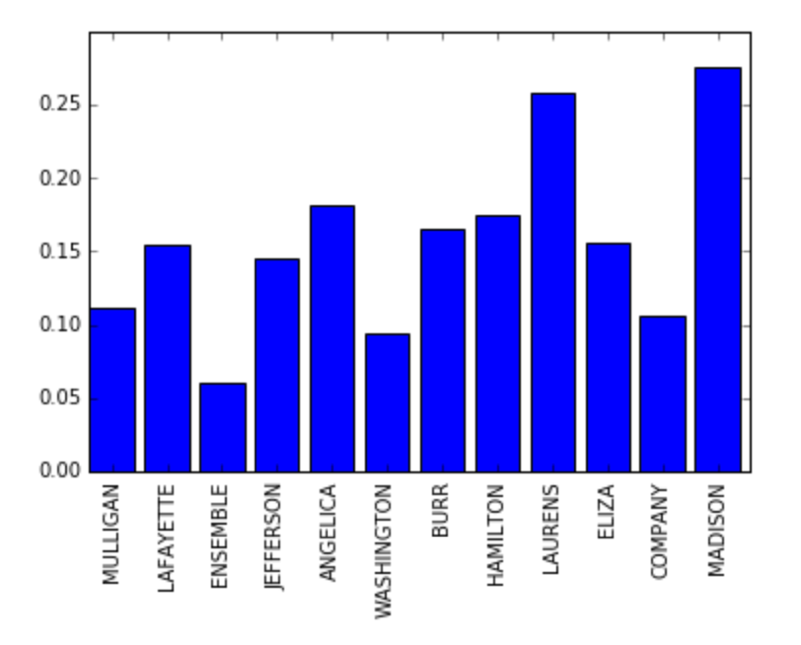
It can be seen that most character on average have more positive lines than negative lines, and when there are similar number of positive and negative lines, intensity of the positive lines are usually stronger, except for the *Company*. In particular, *Laurens’* lines are most positive, while the *Company* is the only one that has negative sentiment.

*Laurens’* positive sentiment is driven by the lines like “But I’ve seen wonders great and small,” which was given an intensity score of 0.6281, or “Then I’ll do it Alexander, you’re the closest friend I’ve got” with intensity score of 0.4939. Overall, the character had much more positive vocabularies than others, including words like *well, freedom, wonder, satisfaction, hope* etc.

*Company* seems to have the most negative sentiment, and these seems to be driven by especially negative lines describing some of the characters. Most of its lines were in fact neutral (126 out of 150 lines) but negative lines carried were much stronger in their intensity. For example, one line included words like *bastard, whore, obnoxious,* and *arrogant* which had a score of -0.891. Other negative lines included words such as *discriminate, sinner* which were also heavily weighted as negative.

Another interesting result was that *Madison* had the largest proportion of positive lines, even though his average sentiment score was much higher than other characters. This suggests that Madison’s lines include positive words more often, but they are not as strong in their intensity. These less intensive but positive words include the following: *well, significant, great, uniquely, virtue,* etc.

**Proportion of Positive Lines**



We can also look at the average intensity of positive and negative lines by the character, which confirms that *Company* has very intense negative lines, while *Mulligan*’s positive lines are much weeker in its intensity compared to its negative lines. *Lafayette* seems to be the character who has most intense lines, both in positive and negative direction, while the show’s protagonist *Hamilton* remains in the middle.

