ISS/PHIL/CS/PUBPOL 110 Final Project Proposal
Investigation into Social Media in Political Campaign

Nathan Shaul
Jacob Zeitlin
Webster Bei Yijie

October 30, 2016

Social media became important to American politics in 2008, when Barack Obama effectively used Facebook to lead him to victory. Since then, the use of social media has been prominent in numerous political strategies. Twitter and Facebook have been especially salient to Hillary Clinton's and Donald Trump's political strategies in the 2016 election so far. It is likely that social media's role in the two candidates'campaigns will only grow as we approach election day on November 8, 2016. Even after the election, social media is likely to play a large role as the candidates express their opinions about the result.

For our final project, we intend to quantitatively analyze the ways in which Hillary Clinton and Donald Trump use social media to enhance their campaigns. We will use web mining techniques to address the research question, How do Hillary Clinton and Donald Trump use social media to enhance their campaign strategies, both before and after Election Day? We will analyze data from Facebook and Twitter to identify similarities and differences between the two candidates'social media strategies, to observe how their strategies change once the next President of the United States is announced on November 8, 2016, and to find aspects of social media posts that voters are most receptive to.

Data for this project will be obtained from Hillary Clinton's Facebook page, Donald Trump's Facebook page, Hillary Clinton's Twitter feed, and Donald Trump's Twitter feed. Data mined from these sources will include the text content of each posted message, the date and time of each posted message, and the number of people who like each posted message. The date and time of each message will be used to determine the number of posts per day by each candidate. The data will be collected using a python crawler. Data will be collected from messages posted from November 1, 2016 through November 15, 2016, and messages will be divided into three categories based on the date they were posted: (1) Before Election: November 1, 2016 to November 7, 2016; (2) Election Day: November 8, 2016; and (3) After Election: November 9, 2016 to November 15, 2016.

Once we have collected these data, we will analyze trends in the number of posts on Twitter and Facebook by Donald Trump and Hillary Clinton before, on, and after Election Day (November 8, 2016). This analysis is intended to shed light on each candidate's temporal activity on social media, and will be useful in our comparison of Trumps and Clintons social media activity. It will also be interesting to observe shifts in the candidates social media use that are induced by the announcement of the winner of the election on November 8. For example, perhaps a large shift (increase or decrease) will occur in each candidate's social media use following the announcement of the winner of the election.

Additionally, we will calculate word frequency to observe the terminology, excluding filler words like and, a, and the, that both candidates use most frequently in their social media posts. The words used most frequently by Trump and Clinton on social media should be indicative of their rhetorical strategies. Therefore, we will be able to draw conclusions about how the candidates react to the events that occur in the days leading up to the election, the events on Election Day, and the immediate results of the election.

Furthermore, we will weight word frequency by likes on a posted message to determine voters receptiveness to particular rhetorical strategies. To do this, we will assign the number of likes each message receives to the words it contains, and sum the like values for each word across all messages to obtain a numerical representation of word popularity. These data are intended to capture trends in the words used by the two candidates that voters express the most support of.

Overall, we hope the results and analysis that this study produces can provide insight into the strategic use of social media by candidates in the 2016 election.

Although each of us has a specialty that can enhance the work of the group, we are all able to help in each of the tasks that we will complete. Therefore, we will each take command of one specific task, but we will all work collaboratively on everything we do.

- Webster is a skilled coder of python, and thus will lead our groups collaboration in developing a program that can collect these data from the candidates Twitter and Facebook posts. He will create a program that compiles the data from the posts into a database and sorts it based on usage frequency of each word or phrase, and assigns it a score based on the popularity of the post (i.e. how many likes it receives)

- Nathan will lead the groups data analysis. He took a statistics course in high school, and as a result is knowledgeable about various methods of quantitative data analysis. Furthermore, he has worked on the analysis of data from a healthcare consulting project, and therefore is comfortable working with large quantities of data. For this project, Nathan will lead the analysis of the data that is compiled by Websters program. This analysis will include cleaning the data, identifying trends, testing for statistical significance, and creating graphical representations that display the identified trends.

- Jacob will primarily focus on the organization of the results and conclusions that we conjecture into the final paper and presentation. He will be the main writer for the paper, as he has experience in the past writing formal research papers. In addition, he will be in charge of creating the PowerPoint (or some other type of presentation) that will be shown to the class, illustrating our question, purpose, methodology, results, and conclusions. The PowerPoint presentation will essentially be an overview of the entire process that we will carry out, as well as a summary of our results and answers to our research questions.