Kevin Zhang

Final Paper and Project Proposal Difference

After finalizing my draft for "Ad Captandum: The Efficacy of Information Operations on Electoral Interference," I realized there are marked differences between the completed project and my initial proposal. Namely, the focus of my paper shifted away from examining the 2016 US Presidential Election, but the impact of information operations on elections in general. I used the 2016 Election and subsequent Special Counsel investigation to better frame my analysis of election interference as it was the catalyst for many social media companies to begin curating their platforms. Ultimately, there were three major ways in which my final research project differed from the initial proposal:

1) Problem Statements: My initial problem statement, as mentioned above, was focusing specifically on determining whether Twitters' ability to remove accounts increased over time. However, I realized after receiving the data that I would not be able to resolve this issue. For one, the date which I would be relying on (the exact date in which accounts were removed from Twitter) was not included within the dataset. I had hoped that a more detailed and un-hashed dataset provided by request from the Trust and Safety team would include this information; however, I found out relatively late that the data set was very similar to the hashed version (no extra data columns added). In my final project, I chose to examine a more malleable in the form of strategies used by Information Operation accounts to evade detection and increase spread.

2) Country Selection: Initially, I had sought to use banned account data pulled from the Internet Research Agency (IRA) accounts and Russia within the Twitter Transparency Report. However, I realized that in order to examine the Russia data better, I would be best comparing it to countries with similar amounts of data. I chose the Iranian and Chinese data sets to pair with the Russian data as the sets contain a similarly large amount of tweets, while also remaining geographically distinct (I did not want to pick Russia and Ghana/Nigeria, as many of the Ghana/Nigeria accounts included were actually proxy Russian accounts).

3) Data & Statistical Analytics: My initial research proposal was very vague about the method I would process my data. I had mentioned my interest in identifying correlation values but did not go deeper into the datasets themselves to pick variables. Additionally, I did not recognize at the time that I would need to deal with the memory debugging and hardware issues that come with large datasets. My previous statistical analysis was done in Stata and with markedly smaller datasets.

I realize that a large portion of this assignment was a learning experience for myself. I had never used Python to individually complete a project from start to finish. Different techniques such as correlation, variance, and attempting regressions helped formulate the ultimate shape of the project. The statistical analysis and methodology implemented in this project is by far the most advanced Python I have used to this date.

The direction of my project indeed shifted in many ways. However, I am happy the way it has turned out as I believe it makes a much stronger paper than initially. Examining strategies of information operation accounts and determining which strategies worked to evade the Twitter censor while propagating additional spread lays a good framework to examine further electoral manipulation. I am certainly interested in potentially taking this paper forward and examining further country data to establish a better understanding of the data set.