Assignment 1

Digital Strategies for the Social Sciences

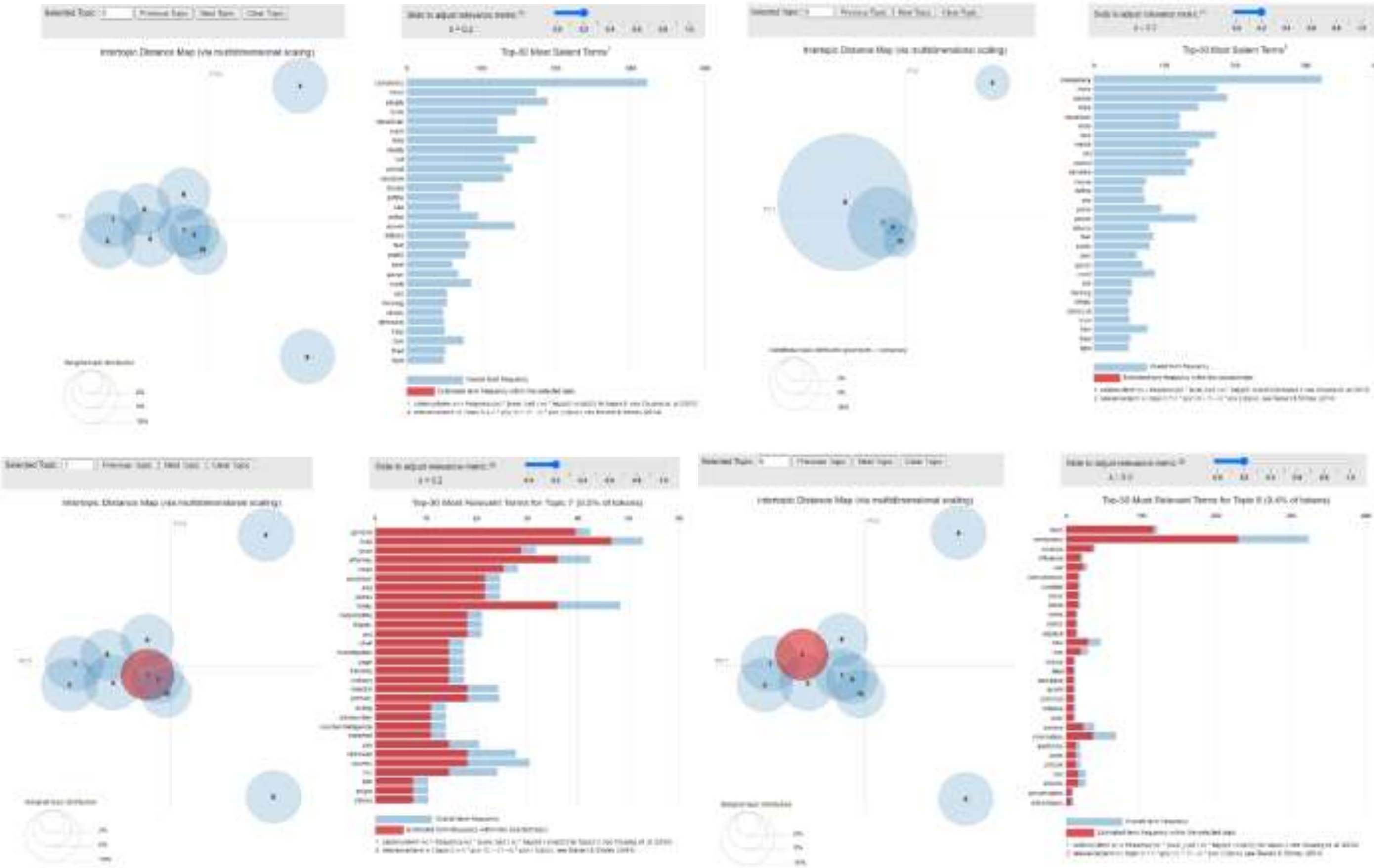
For this assignment I am working with data collected from the site www.qalerts.app. This site provides easy access to all posts from the anonymous 8chan user Q. The posts made by Q can be thought of as the bedrock of the alt-right conspiracy theory QAnon in which former U.S president Donald Trump and his allies is waging a secret war against a powerful cabal of pedophile cannibal Democrats. (Britannica, 2021)

Crazy as it may seem this conspiracy theory has taken root in America, with outspoken QAnon-believers elected to the House of Representatives. I believe history shows us the importance of disentangling conspiracy theories and not sweep them under the rug. We should investigate them, understand them and confront them.

A study from Switzerland utilized unsupervised machine learning to investigate stylometrics in "Q-drops". Q posts is referred to as a "Q-drops" within the QAnon community and I will utilize this terminology in the sections to come. The results from the Swiss study show how the anonymous account of Q bears two distinct individual signatures, indicating that the account has been controlled by at least two individuals. (OrphAnalytics, 2020) The study's approach highlights how computational methods can be utilized to investigate the origin and character of conspiracy theories.

For the scope of this assignment, I choose a smaller subset of Q-drops. I choose drops that related to the term "conspiracy". I accessed these by simply typing "conspiracy" in the search bar at the website and hit enter. This returned 156 drops. However, drops can be nested inside each other so the actual sample consisted of about 200 observations. I extracted the textual elements from the drops as well as their unique tripcodes. I cleaned the observations using regular expressions per instructions and proceeded to build a topic model. I want to give KBLab and the topic modeling workshops they provided 21st-22nd April credit for a lot of the code that was utilized. The topic model I present is quite crude. I only removed stop-words, decapitalized all letters and removed extra spaces. I also did not qualitatively investigate the drops which I think would have been the correct way to go about it in order to pre-determine how many possible topics there might be. In other words, I lack a "ground-truth" for the number of topics I specify to the model. I set number of topics in the model to 10. However, I think the model still has some interesting results.

Using this model, we can visualize the frequency of the term "conspiracy" given a topic. Unsurprisingly this term has a high frequency in our model overall but we also see that words/abbreviations such as "news", "media", "msm" (mainstream media), "people" and "fake" is common. In the figures displayed on the next page, I have decreased lambda to 0.2 which puts more weight to the ratio of the frequency given the topic to the overall frequency of the word. Doing this simplifies classification of the topics. From the figures we see that the term is most common in topic 8. Albeit not present among the 30 most frequented words in topic 7 when lambda is 0.2, this topic is second-place carrier of the term "conspiracy" as indicated by figure 2.



However, when looking at these figures we see that the topics is not evenly distributed over the scatterplot and that all except 2 topics (4, 5) tend to cluster together. This is not reassuring since the distance between for instance topic 8 and 7 is an approximation of the difference between these two topics. In other words, an approximation of their semantic relationships. So, what I essentially have is a large cluster of very similar topics and two distinct topics. Looking at topic 7 for instance, one would initially be inclined to label this topic as “Public jobs” or maybe “Individual actors”. Common words include: “general”, “attorney”, “assistant”, “deputy”, “witness”, “director”. In regards to topic 8, its most common words are “msm”, “conspiracy” and “controls” so one might label this as “Media-Conspiracy symbiosis”. But as mentioned previously, the difference between the topics is very small hence an interpretation would be that these topics are semantically conjoined which makes labeling problematic.

What is interesting is that topic 4 and 5, which I do not display here but that can be accessed through R script, have a huge semantic difference. Topic 5 includes words such as: “people”, “think”, “trust”, “right”, “good”, “united”. Topic 5 includes: “news”, “media”, “fake”, “attacks”, “disinformation”, “Russia”. My interpretation of this is that Q clearly pits a negative image of media against a good image of people. So, the bad media is the enemy of the good people and of course this was a message we heard a lot during Trump’s term of office.

Furthermore, I utilized RSelenium to automize collection of Q-drops. This time I wanted to collect text data surrounding two concrete political conspiracies in the U.S. One being the Watergate-scandal and the other being Biden-Ukraine conspiracy. There were more than four times as many drops that related to the Biden-Ukraine conspiracy (~30) than to the Watergate-scandal (7). An interpretation of this is that Q purposefully choose to propagate conspiracies that are contemporary but inaccurate rather than historical but proven accurate. Further on, it demonstrates how the account propagate conspiracies in which the conspiring actors are linked to DNC rather than the RNC.

I created corpuses of the text data and extracted word clouds over the most frequented words. The minimum frequency for a word to be included was 10 in the drops that related to Ukraine and 5 for Watergate. I choose to keep the threshold lower for Watergate since there were so few drops.

Watergate

Ukraine

Looking at the above figures, we see that drops relating to Ukraine predominantly targets Biden, Obama (Hussein) and the DNC which makes sense since the conspiracy was spread in an effort to thwart Joe Bidens presidential campaign. I find the results for Watergate drops most interesting since they demonstrate efforts to target the media. This fits in with the overall anti-media semantics derived when looking at the overarching term “conspiracy”. One interpretation of this is that Q actively replaces historically validated conspiring actors belonging to one side of the political aisle, actors tied to the RNC, with the media and that Q thus purposefully tries to distort the legacy of the Watergate scandal.

References:

Reid, S. A. (2021, January 27). Conspiracy theory. Encyclopedia Britannica. https://www.britannica.com/topic/conspiracy-theory

OrphAnalytics. (2020, December 15). Style analysis by machine learning reveals that two authors likely shared the writing of QAnon's messages at two different periods in time. https://www.orphanalytics.com/en/news/whitepaper202012