Russian Influence? Let the Tweets Speak

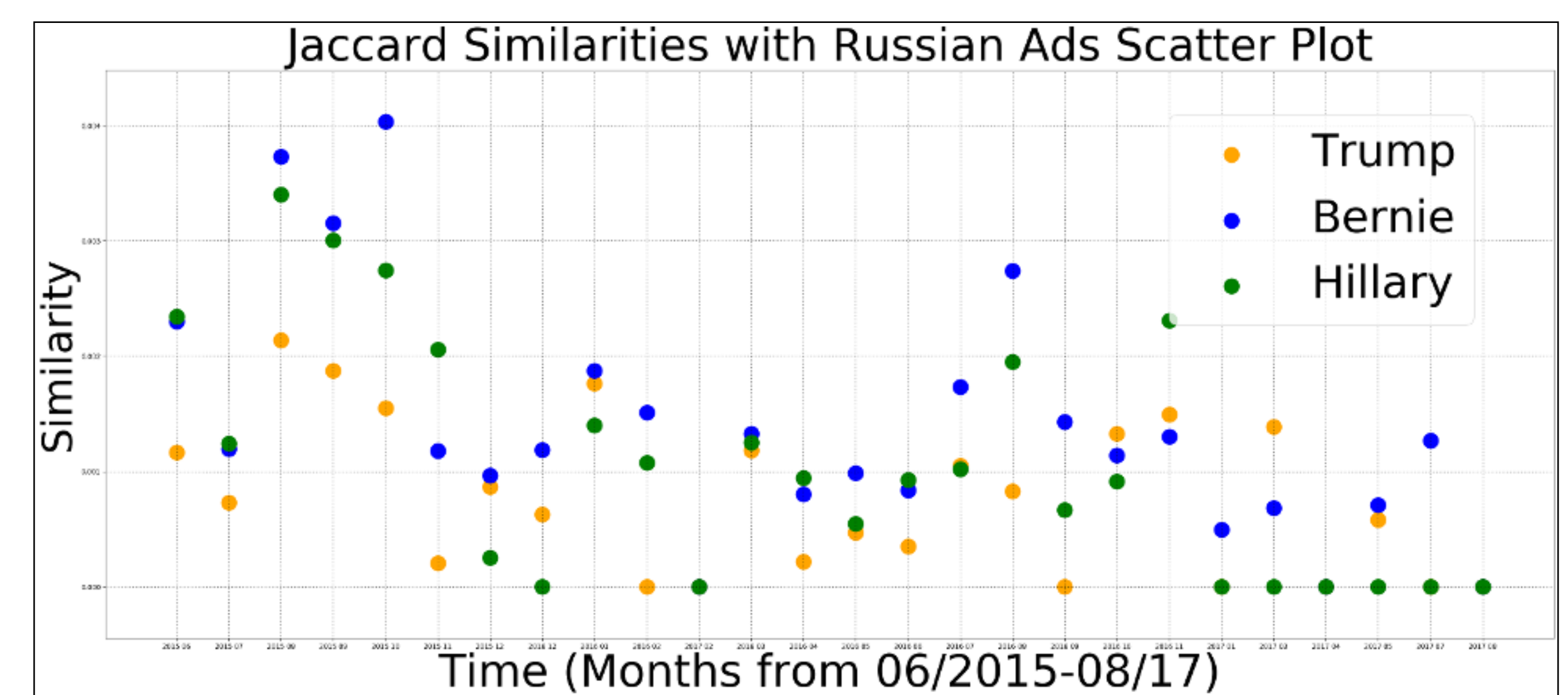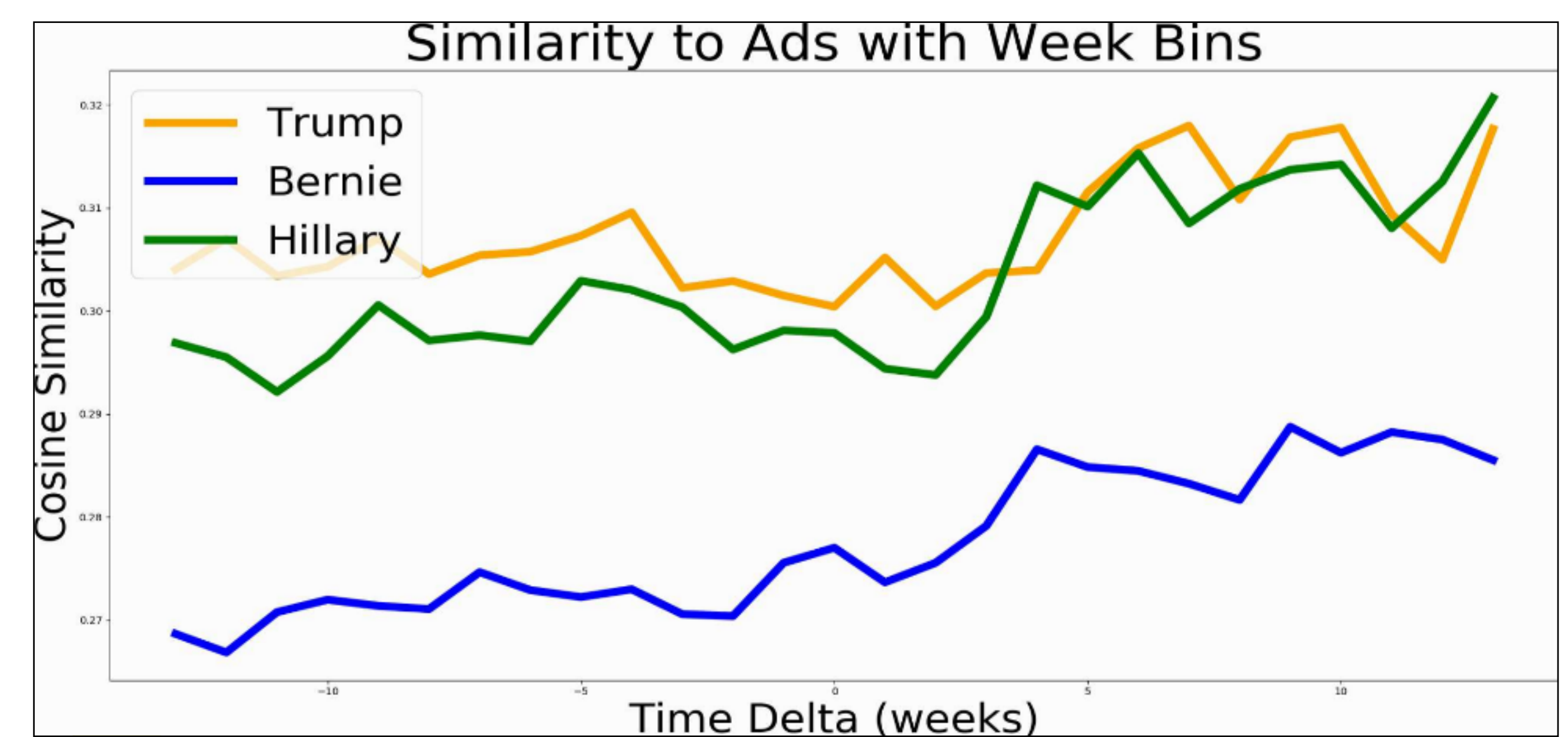*Brandon Mouser   Kyle Price   Jaden Simon*

Motivation:

A foreign hostile government knowingly influenced the US 2016 presidential election. We wanted to determine which (if any) of the leading political figure's rhetoric was emulated in order to cause discord, confusion, and prejudice. We also explored the possibility of political candidates emulating Russian ads.

Datasets:

Russian ads were collected from the House Intelligence Committee public database. We originally tried to use Twitter's REST API for tweets, but ended up just scraping the website for tweets. There were 3474 Russian ads and 22432 Tweets (9087 for Trump, 6222 for Bernie, and 7123 for Hillary). All datasets were in JSON format.

Ad/Tweet Similarity:

We began by calculate similarities between tweets and ads and averaging over month-long bins.

*Bernie measured the highest similarity overall*

Realizing that this wasn't helpful, we decided to enhance the model by moving from term matching to word-embedding and plotting similarities as a function of time differences.

Tweets appear more similar after ads are posted

Word-embedding and Topic Extraction:

We extracted topics from each dataset using LDA, and then measured the cosine similarity between them in order to evaluate how well our word-embedding model was capturing the contextual subtleties in the text.

Topics do not differentiate candidates
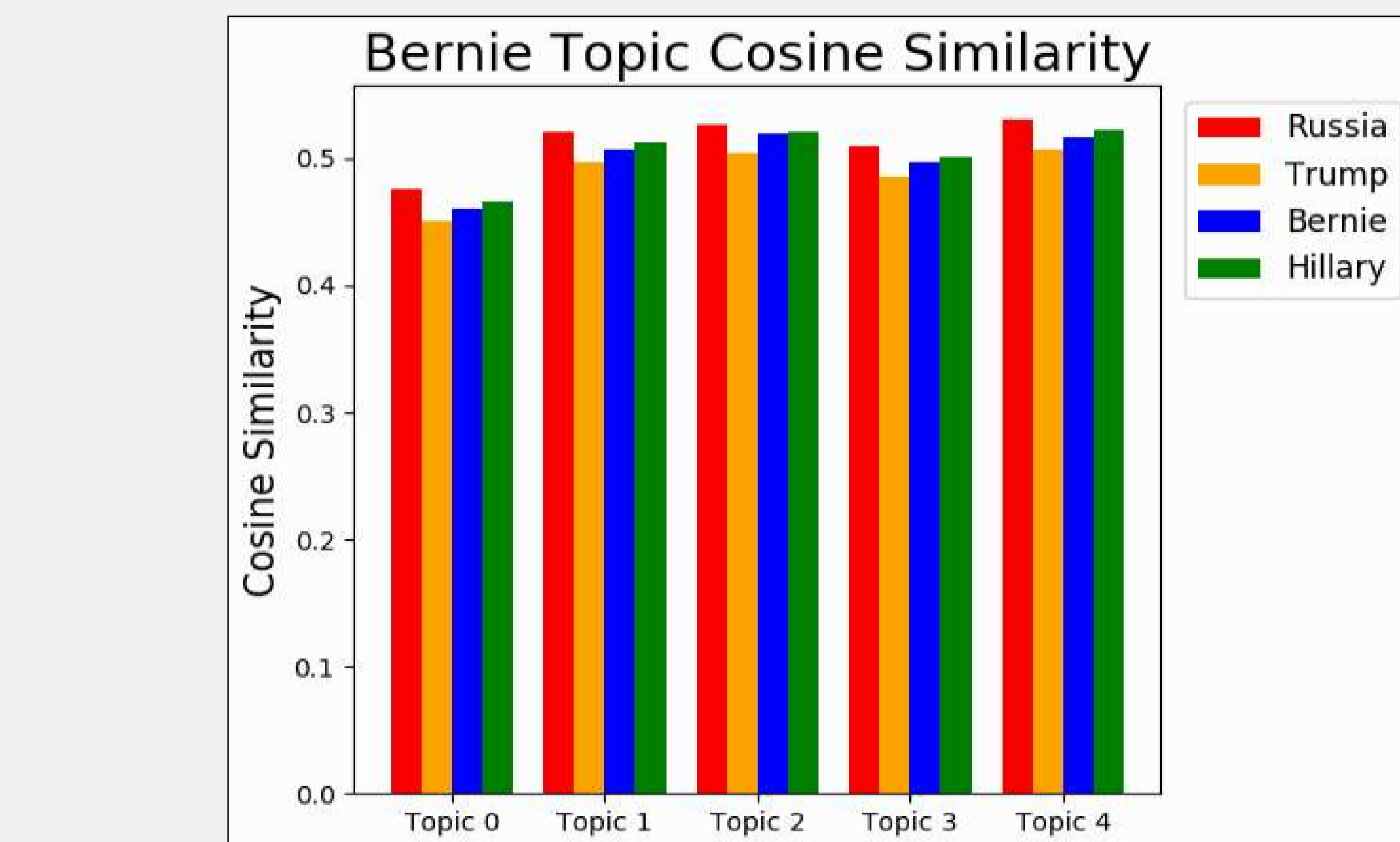
Unfortunately the results were inconsistent in that the similarities between the topics and datasets weren't consistently higher than the similarities between that topic and the other datasets.

EXAMPLE

Consistent Similarity Result   *Inconsistent Similarity Result*

Other Forms of Analysis:

Term Frequency and Sentiment Analysis are used to visualize and compare candidates to the ads. Who would have thought one of Trump's favorite words is 'crooked'?

Top Positive: enjoy, wonderful, proud, online, fair, great, facemusic, beautiful, hampshire, happy, amazing, important, sure, player
Top Negative: horrible, sick, terrible, crazy, dangerous, disastrous, false, badly, bad, weak, tired, fake, pathetic, wrong

Sentiment Analysis of All Candidates and Ads

Top Trump: thank, poll, tonight, debate, crooked, am, cruz, big, donald, cnn, great, tomorrow, win, iowa
Top Russia: black, police, matters, community, officer, cops, free, racism, free, brutality, student, shoot, follow

Corpus Characteristics Between Trump and Russia

Sentiment Analysis:

Seems like something affected the positivity of each candidate and the ads around November of 2016...?

Word-cloud:

Word-clouds quickly and effectively convey the general focus of each candidate and the Russian ads.

Conclusion:

Multiple methods of similarity analysis did not turn up any conclusive results, and term frequency analysis showed unique results for each dataset. While it may be possible that the Russian ads did indeed influence the election, our results do not conclusively show support for any single candidate.

Sources:

Tweets - https://twitter.com/explore
Twitter Scraper - https://pypi.org/project/GetOldTweets3/
Russian Ads - https://github.com/russian-ad-explorer/russian-ad-datasets
Spacy - https://spacy.io/
Scattertext - https://github.com/JasonKessler/scattertext
LDA - https://pypi.org/project/lda/
Wordcloud - https://pypi.org/project/wordcloud/
Project Github - https://github.com/JadenSimon/ForeignInfluence/