A Computational Analysis of the Coronavirus Pandemic Response of Tri-State Area Politicians on Twitter

Zachary Dau

July 1, 2021

A Computational Analysis of the Coronavirus Pandemic Response of Tri-State Area Politicians on Twitter

Zachary Dau
Department of Linguistics
Department of Computer Science
Montclair State University
Montclair, NJ, USA
dauz1@montclair.edu

Abstract

We present our findings on a computational analysis of coronavirus response-related tweets by prominent politicians that represent the Tri-state area of New York, New Jersey, and Connecticut. We extract Twitter data over the span of 20 months and collect tweets from Governors Andrew Cuomo, Phil Murphy, and Ned Lamont, Mayor Bill de Blasio, and President Donald Trump. With this data, we observe the drastic increase in Twitter activity during the pandemic and apply LDA and LSA models to discern changes in topic. Moreover, we analyze sentiment and lexical choices of these politicians.

1 Introduction

The 2020 novel coronavirus (COVID-19) pandemic has undoubtedly altered the normal way of life for people throughout the world. Currently, the virus has affected 218 countries and territories with the most cases occurring in the United States (World Health Organization, 2020). Of the approximately 26,595,540 cases in the country as of February, 4, 2021, 2,407,010 cases reside in the states New York, New Jersey, and Connecticut, commonly referred to in the region as the "Tri-state area" (Center for Disease Control, 2020). These states currently rank fourth, twelfth, and thirty-fourth, in total case count, respectively, but were first, second, and seventh in May 2020 at the peak of community spread of the virus in those areas. New York City, the most populous city in the United States, alone accounts for 621,218 cases, higher than thirty-seven other states (Center for Disease Control, 2020). As a result, residents have looked to state and federal politicians for guidance and leadership during this critical period. The highest ranking of these public officials are Governor of New York, Andrew Cuomo; Governor of New Jersey, Phil Murphy; Governor of Connecticut, Ned Lamont; Mayor of New York City, Bill de Blasio; and President of the United States, Donald Trump.

As there has thus been much governmental response to the COVID-19 outbreak presented on Twitter, the social media platform presents a rich amount of coronavirus-related textual data available for extraction. Analyzing this data allows for the observation of changes in the amount of communication on Twitter, topic, types of words used, and tweet sentiment as a result of the COVID-19 pandemic period.

2 Related Work

Twitter has long been a method of quickly spreading information, evidenced further by the vast amounts of governmental COVID-19 guidelines presented in such a short period of time. Hughes and Palen (2009) found that Twitter has historically been used in cases of both national security and environmental disasters. Moreover, prominent figures directly involved with a given situation generally see a rise in Twitter activity.

Yang, Chen, Maity, and Ferrara (2016) found Twitter to be an ideal platform for politicians to engage in conversation with constituents as well as set political agendas. Takikawa and Nagayoshi (2017) used Latent Dirichlet Allocation (LDA) to create topic models of politically-based tweets and generalize a conversation amongst a community.

Additionally, Sheikha (2020) compared the results of the LDA and Latent Semantic Analysis (LSA) (Deerwester, Dumais, Furnas, Landauer, and Harshman, 1990) to observe topic changes of media outlets during the COVID-19 pandemic on Twitter. Research found that NBC News and the New York Times exhibited topic changes that correlated well with major events during the pandemic.

Twitter has become a ubiquitous method of expressing opinion and emotion for individuals of all cultures and socio-economic backgrounds. Bae and Lee (2012) observed prominent figures on Twitter and measured sentiment and word usage via the lexicon-based analyzer Linguistic Inquiry and Word Count (LIWC) (Tausczik and Pennebaker, 2010). Research found that tweeting sentiment of prominent figures changed over time and affected the sentiment of their followers. Moreover, Tumasjan, Sprenger, Sandner, and Welpe (2010) used various LIWC categories to study the political sentiment and deliberation of German federal elections, finding that politicians often display an increase in usage of particular LIWC word categories depending on their role or situation.

More recently, research in sentiment analysis of Twitter has included work with the `TextBlob` library (Ahuja and Dubey, 2017), a tool that yields polarity and subjectivity scores to data. When analyzing Twitter data, `TextBlob` extracts polarity data and can return percentages of positive, negative and neutral tweets of a specified user.

3 Methods

As research has indicated Twitter to be the most practical social media source to observe political data (Mejova and Srinivasan, 2012), we gathered tweets from Governor Andrew Cuomo, Governor Phil Murphy, Governor Ned Lamont, Mayor Bill de Blasio, and President Donald Trump from two periods of time. The first was a pre-pandemic period of January 1, 2019 through February 29, 2020. The second was from the date of the first confirmed case of COVID-19 in the Tri-state on March 1, 2020, through September 1, 2020. We refer to this time as the pandemic period. From these data sets, we observed activity from each of the five politicians and ran an LDA, an LSA, and LIWC on all tweets. Lastly, we ran a sentiment analysis using the `TextBlob` library.

3.1 Tweet Frequency

Hughes and Palen (2009) found that during national security and environmental disasters, closely involved public officials will increase Twitter activity as a means of communicating with their constituents. As is the case, we sought to observe any Twitter activity increases from the five politicians during the pandemic compared to a baseline of the pre-pandemic period.

3.2 LDA

As Takikawa and Nagayoshi (2017) used an LDA to create topic models of political tweets, we sought to create topic models for each politician for both the pre-pandemic and pandemic periods. To do so, we cleaned the extracted tweets for punctuation and stop words, normalized words for lemmatization, and used the `gensim` library (Řehůřek and Sojka, 2010) to extract topics.

3.3 LSA

Similarly to the LDA, we also performed an LSA for the pre-pandemic and pandemic periods with the intention of observing changes in topic as was found in Sheikha (2020). We likewise cleaned tweets for punctuation and stop words, and normalized words for lemmatization. For LSA, however, we used the `sklearn` library (Pedregosa, Varoquaux, Gramfort, Michel, Thirion, Grisel, Blondel, Prettenhofer, Weiss, and Dubourg, 2011) to extract feature words and topics.

3.4 LIWC

Tumasjan et al. (2010) previously documented the word usage of politicians on Twitter through LIWC, observing various categories relevant to politics and psychological analysis. Our work also focused on these categories, adding the categories of *death*, *family*, and *health*, given the nature of the COVID-19 pandemic. We separated these into two subsets which we refer to as *tangibles*, those LIWC groupings that reflect more concrete concepts, and *emotions*, those groupings that are emotive. The *tangibles* subset consisted of *achievement*, *death*, *family*, *future-oriented*, *health*, *money*, *past-oriented*, and *work*, whereas the *emotions* subset included *anger*, *anxiety*, *certainty*, *negative emotions*, *positive emotions*, *sadness*, and *tentativeness*. We computed the average usage of these categories per tweet for each politician during the pandemic period and compared them against the pre-pandemic baseline.

3.5 Sentiment Analysis

Since `TextBlob` has become increasingly popular for analyzing sentiment in tweets (Ahuja and Dubey, 2017), we computed the percentages of positive, neutral, and negative tweets for each of the five politicians during the pandemic period to observe any increases or decreases from the pre-pandemic period.

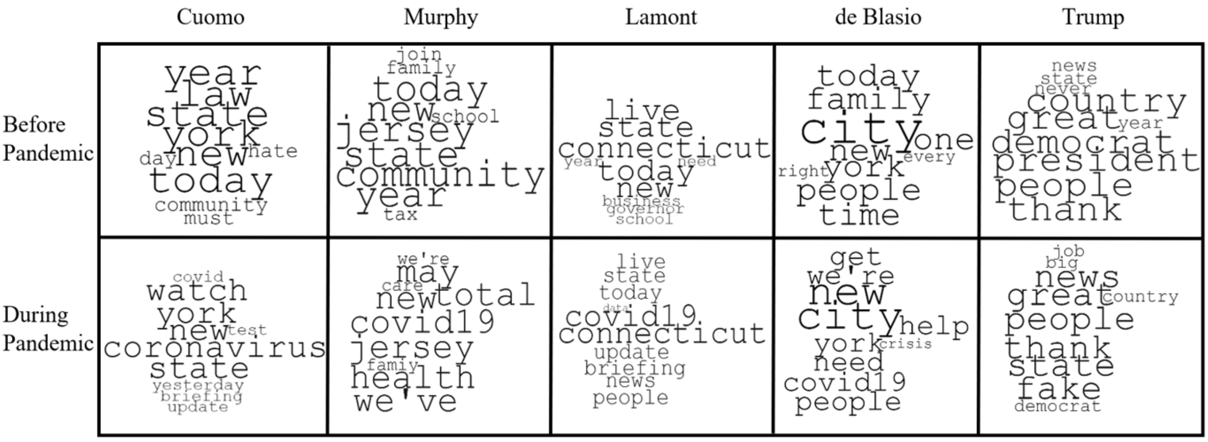

Figure 1: Twitter LDA results per politician as word clouds of the ten highest weighted topic words. The top row displays pre-pandemic LDA results, whereas the bottom row exhibits pandemic LDA results. Larger words designate a higher topic weight.

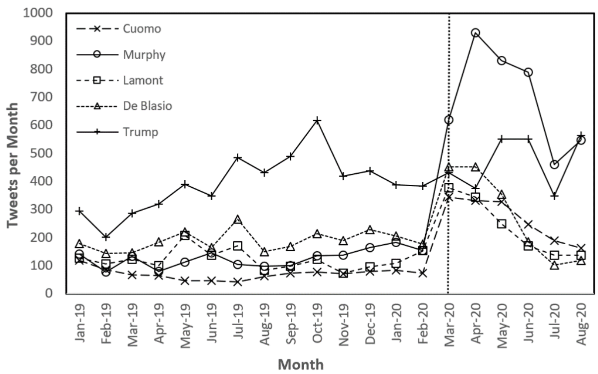

Figure 2: Monthly tweet frequency of the Tri-state Area politicians. The vertical dotted line represents the first confirmed case of COVID-19 in the Tri-state area on March 1, 2020.

4 Results and Discussion

4.1 Tweet Frequency

All five politicians saw a jump in tweets per month following the announcement of the first confirmed positive case of COVID-19 on March 1, 2020 (Figure 2). Governor Murphy had the highest increase in tweets per month, averaging 570 more tweets during the pandemic period. The average tweets per month of Governors Cuomo and Lamont rose by 196.1 and 114.26, respectively, and Mayor de Blasio's average climbed by 88.98 tweets per month. President Trump saw the smallest increase with an average of 78.38 more tweets per month. These increases demonstrate the reliance of politicians on Twitter to communicate with constituents during a time of crisis, consistent with the findings of Hughes and Palen (2009).

4.2 LDA

We retrieved the ten highest weighted words per politician via an LDA model for both the pre-pandemic and pandemic periods. Each individual instance is represented as a word cloud in Figure 1.

In the pre-pandemic period, each politician exhibits relatively normal tweeting topics, discussing their respective communities by name (e.g. "New Jersey," "city," "country," etc.) as well as political vernacular such as "law," "community," and "business." Topics changed systematically during the pandemic period, however, with Governors Cuomo, Murphy, and Lamont and Mayor de Blasio showing shifts in topics to COVID-19 related terms. The term "covid19" appeared for each of these four politicians, with "coronavirus" also appearing for Governor Cuomo. Other pandemic-related lemmas that emerged for various politicians included words relating to informing constituents about pandemic updates (e.g. "briefing," "update," "health," etc.). President Trump was the only exception to this shift, with much of his topic remaining similar to his pre-pandemic results and no COVID-19 related terms appeared.

4.3 LSA

Like our LDA analysis, we likewise retrieved the top weighted words per politician for the pre-pandemic and pandemic periods in an LSA model as well. For our LSA, however, we gathered the 15 highest weighted feature names for both periods

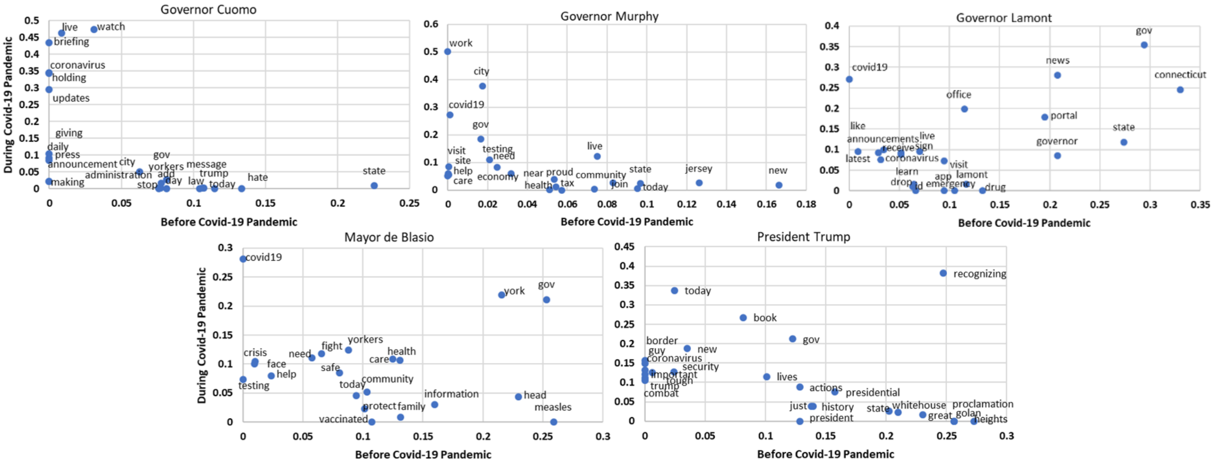

Figure 3: Top LSA word results per politician for the periods before and during the Covid-19 pandemic. Terms aligned to the x-axis are more closely related to topics before the pandemic, whereas terms closer to the y-axis are more closely related to topics during the pandemic. Terms not necessarily aligned to either axis are prevalent during both periods.

and plotted the weights of the two sets per feature name (Figure 3). This process allowed us to visualize the differences in topics between the two periods per politician. As there existed a degree of overlap between prominent feature names, each plot in Figure 3 contains less than 30 total feature names.

As was found in the LDA, all politicians maintained normal topics on Twitter during the pre-pandemic period in the LSA model, and the respective communities for each politician (e.g. “New York,” “state,” “city,” etc.) were highly weighted in the LSA as well. Since these terms were so highly weighted, they were at times omitted from Figure 3 as to generate better visualization for key terms. General political vernacular (e.g. “administration,” “community,” “tax,” etc.) were again present in pre-pandemic topics. Once again, the topics of the pandemic period changed to varying degrees. Topics changed through the introduction of new vernacular like “covid19” and “coronavirus” (at least one form appearing for every politician) as well as the prevalence of semantically related pandemic terms such as “testing,” “briefing,” and “crisis.” Unlike the LDA model, the LSA indicated the occurrence of the term “coronavirus” for President Trump.

4.4 LIWC

We observed two subsets of LIWC word categories that we referred to as *tangibles* and *emotions*.

Within the *tangibles* subset (Figure 4), we found that each politician saw a decrease in *achievement* words during the pandemic period. The use of *future-oriented* words either stayed the same or showed a slight increase, whereas *past-oriented* words were used far less frequently by all politicians aside from President Trump whose levels stayed the same during both periods. The use of *health* words increased for all politicians during the pandemic, excluding President Trump, who again maintained pre-pandemic levels. Lastly, the use of *money* words decreased for all five politicians, whereas *work* words drastically increased during the pandemic period, aside from President Trump, where *work* related words decreased.

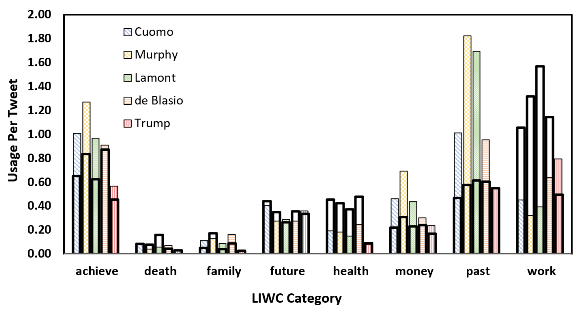

Figure 4: LIWC *tangibles* word usage per tweet. The colored bars represent baseline data whereas the overlaying, bolded bars show data from after March 1, 2020.

In the *emotions* subset (Figure 5), all five politicians saw a decrease in *anger* words per tweet during the pandemic, however also saw a decrease in *positive emotion* words as well. *Negative emotion* words decreased for all politicians during the pandemic, except for Governor Murphy who again had a slight increase. Lastly, Governors Murphy and Lamont both saw an increase in *certainty* words, whereas the remaining three politicians saw a decrease or stayed the during both periods.

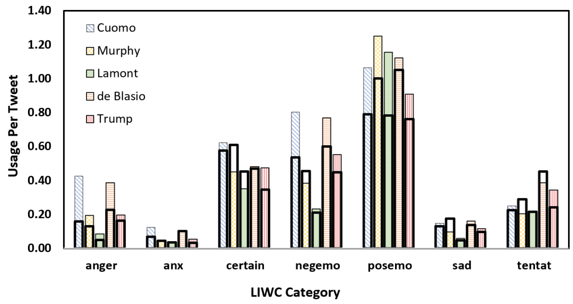

Figure 5: LIWC *emotions* word usage per tweet. The colored bars represent baseline data whereas the overlaying, bolded bars show data from after March 1, 2020. Note the difference in scale for usage per tweet

4.5 Sentiment Analysis

We found systematic differences between tweet sentiment in both the pre- and pandemic periods (Table 1). Every politician tweeted a higher percentage of neutral tweets during the pandemic and all politicians aside from Governor Murphy exhibited a higher percentage of negative tweets in the pre-pandemic period. Governor Lamont and Mayor de Blasio showed an increase in percentage of positive tweets during the pandemic while the rest decreased.

Overall Governor Lamont had the highest percentage of positive tweets during the pandemic at 72.71%, while Mayor de Blasio had the highest increase in percentage of positive tweets with 3.19%. Likewise President Trump had the highest percentage of neutral tweets during the pandemic with 30.54% and Governor Cuomo showed the highest increase in percentage of neutral tweets with 9.79%. Lastly President Trump had the highest percentage of negative tweets during the pandemic with 20.06%, although his overall percentage of negative tweets had decreased from pre-pandemic levels. As Governor Murphy was the only politician to display a rise in percentage of negative tweets during the pandemic, he accordingly had the highest increase of 2.65%.

Politician	Positive	Neutral	Negative
Before Pan.			
Cuomo	**66.63%**	12.36%	**21.01%**
Murphy	**75.61%**	13.66%	10.73%
Lamont	69.69%	20.66%	**9.65%**
de Blasio	66.82%	13.92%	**19.26%**
Trump	**53.96%**	23.36%	**22.68%**
During Pan.			
Cuomo	60.51%	**22.15%**	17.34%
Murphy	69.63%	**16.99%**	**13.38%**
Lamont	**72.71%**	**19.46%**	7.83%
de Blasio	**70.01%**	**14.24%**	15.75%
Trump	49.40%	**30.54%**	20.06%

Table 1: Positive, neutral, and negative tweet sentiments before and during the pandemic. The higher values between the two periods are in **bold.**

5 Conclusion

In this paper, we presented our findings of a computational analysis of the Twitter activity of Tri-state area politicians during the COVID-19 pandemic period and compared these results to pre-pandemic tweets. We found there was a drastic increase in Twitter activity during the pandemic, consistent with prior research. Moreover, our LDA, LSA, and LIWC analyses detected differences in topic and word usage by these politicians during the pandemic as well. These findings indicate that the onset of the COVID-19 pandemic has altered the vocabulary, approach, and style of communication of these political leaders. Lastly, a sentiment analysis of the rate of positive, negative, and neutral tweets during the pandemic found that the percentage of neutral tweets increased for all politicians, and negative tweets decreased for all politicians during the pandemic except Governor Murphy.

Ultimately, our work shows that there has been an increase in Twitter usage by these politicians to increasingly engage with their constituents during the pandemic and create a narrative of COVID-19 mobility and awareness. Additionally, our work finds this adjustment has created a comparatively more neutral sentiment as a result of this message. Future research could compare our results with that of other politicians across the country as well as continue this work as vaccines continue to be distributed in the Tri-state area.

References

Shreya Ahuja and Gaurav Dubey. 2017. Clustering and sentiment analysis on twitter data. In *2017 2nd International Conference on Telecommunication and Networks (TEL-NET)*, pages 1–5. IEEE.

Younggue Bae and Hongchul Lee. 2012. Sentiment analysis of twitter audiences: Measuring the positive or negative influence of popular twitterers. *Journal of the American Society for Information Science and Technology*, 63(12):2521–2535.

Center for Disease Control. 2020. CDC COVID Data Tracker.

Scott Deerwester, Susan T Dumais, George W Furnas, Thomas K Landauer, and Richard Harshman. 1990. Indexing by latent semantic analysis. *Journal of the American society for information science*, 41(6):391–407.

Amanda Lee Hughes and Leysia Palen. 2009. Twitter adoption and use in mass convergence and emergency events. *International journal of emergency management*, 6(3-4):248–260.

Yelena Mejova and Padmini Srinivasan. 2012. Political speech in social media streams: Youtube comments and twitter posts. In *Proceedings of the 4th Annual ACM Web Science Conference*, pages 205–208.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, and Vincent Dubourg. 2011. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830.

Radim Řehůřek and Petr Sojka. 2010. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50, Valletta, Malta. ELRA. http://is.muni.cz/publication/884893/en.

Hassan Sheikha. 2020. Text mining twitter social media for covid-19: Comparing latent semantic analysis and latent dirichlet allocation.

Hiroki Takikawa and Kikuko Nagayoshi. 2017. Political polarization in social media: Analysis of the “twitter political field” in japan. In *2017 IEEE International Conference on Big Data (Big Data)*, pages 3143–3150. IEEE.

Yla R. Tausczik and James W. Pennebaker. 2010. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1):24–54.

Andranik Tumasjan, Timm O Sprenger, Philipp G Sandner, and Isabell M Welpe. 2010. Predicting elections with twitter: What 140 characters reveal about political sentiment. In *Fourth international AAAI conference on weblogs and social media*. Citeseer.

World Health Organization. 2020. WHO Coronavirus Disease (COVID-19) Dashboard.

Xinxin Yang, Bo-Chiuan Chen, Mrinmoy Maity, and Emilio Ferrara. 2016. Social politics: Agenda setting and political communication on social media. In *International Conference on Social Informatics*, pages 330–344. Springer.