An investigation of differences in sentiment from tweets related to COVID-19 between Canada and US residents

Keshava Pallavi Gone
Dalhousie University, Master of Information Student, KeshavaPallavi.Gone@dal.ca

Colin Conrad
Dalhousie University, Assistant Professor, Colin.Conrad@dal.ca, 0000-0003-3891-1670

SUMMARY

Goal: This study aims to understand how individuals communicated and acknowledged to COVID-19 pandemic on Twitter. It mainly focused on identifying and demonstrating the differences in the perspective of United States and Canadian residents.

Design / Methodology / Approach: We performed sentiment analysis on a sample of 1 005 358 tweets from the states and provinces most affected by COVID-19 in USA and Canada between 23 March 2020 to 24 April 2020. To accomplish this, we used the Valence Aware Dictionary and sEntiment Reasoner (VADER), which is based on a dictionary of a set of words with positive or negative sentiment scores (Hutto & Gilbert, 2014). We also compared differences in word frequencies between the two countries and compared the sentiments trends over 1 month and analyzed frequent words from the tweets of both the countries to give a clear picture on the gap between the sentiment and how differently people of two countries expressed their emotions on Twitter.

Results: The project revealed the differences between the attitudes of people in USA and Canada on COVID-19 outbreak. The study highlighted two major differences. First, it revealed differences in positive and negative sentiment, as well as how it changed and each day of the given period. Second, it provided differences in how general public in two countries responded and reacted to the outbreak by analyzing the frequent words used on the social media platform.

Limitations of the research: The study is limited in its scope to focus only on the twitter data generated by US and Canadian residents. It also uses the VADER approach, which is classifies tweets into three discrete sentiments (positive, negative, neutral), which may not capture all possible sentiments.

Practical implications: We demonstrate a technique which could be implemented by policy makers to determine public sentiment to a particular emerging topic on social media. Specifically, policy makers could use this approach to complement traditional methods to determine public opinion about COVID-19.

Originality / value: While other research has explored sentiment on Twitter related to COVID-19, to the best of our knowledge, no research has specifically focused on the differences between Canada and the United States during the formative period of policy responses specifically.

1. INTRODUCTION

The outbreak of the COVID-19 virus in China in late 2019 initially triggered international concern, which was later followed by an unprecedented economic crisis. In the early months, the pandemic received an enormous amount of popular attention, which was amplified by the expression of opinions on social media platforms such as Twitter. It is now widely accepted that initial public opinion towards the crisis and governments' responses played a major role in shaping health policies, which has been particularly noted in the differences in COVID-19 outcomes between the United States and Canada (Gadarian, Goodman & Pepinsky, 2020; Merkley et al, 2020). By using sentiment analysis—a natural language processing technique— it is possible for researchers to extract the subjective information underlying the opinion or statement of the individuals related to the COVID-19 disease. Such analysis can give insights into differences in attitudes expressed by members of the public, potentially giving policy makers a tool for anticipating public reaction to the response to the disease, and consequently anticipate the impact of popular behaviour on health outcomes.

In this paper, we describe the results of a study of 1 005 358 tweets containing reference to COVID-19 generated by United States and Canadian residents between the March 23, 2020 and April 24, 2020, during the early stages of the global response to the outbreak. Our objective is to discern differences in the viewpoints of Canadian and American residents in an effort to understand how public opinion may have influenced health policy outcomes. In the knowledge discovery context, there are two fundamental classification tasks that can be considered in conjunction with the Twitter data: (a) We implement term frequency, that provides knowledge into the relative popularity of terms or phrases, and polarity, which measures the sentiment (i.e. positive, negative) of the phrases. (b) Word frequency of each sentiment, that gives most frequently used words with word cloud visualizations to discover people's attitudes and reaction towards pandemic on social media. By analyzing these two features, we can discover differences which may reflect intangible changes in public opinion which ultimately shaped the differences in outcomes in the two countries. The acquired information could be codified in a way that helps policy makers decide on expected reactions from the public. Our research question can be articulated as follows:

> **RQ:** Are there differences in word frequencies and polarity between tweets related to COVID-19 between Canada and US residents on Twitter?

Following this introduction, the remainder of the paper is structured as follows. Section 2 provides background literature on the COVID-19 virus in the United States and Canada during this period and summarizes the current state of relevant literature on Twitter sentiment analysis. Section 3 is devoted to explaining the study's methodological approach. Section 4 presents the results of the research, which is followed by discussion in Section 5 and conclusion in Section 6.

2. BACKGROUND

2.1 COVID-19 outbreak in the United States and Canada

In December 2019 and January 2020, a cluster of patients with infected pneumonia was reported in Wuhan, China (Li, 2020). It was discovered that the pneumonia cases were caused by a novel coronavirus, named 2019-nCoV (Zhu et al., 2020), and the disease caused by the virus was later named COVID-19 by the World Health Organization (WHO) (Li & Feng, 2020). Evidence was found that the infection is transmitted from human-to-human, it had a widespread to other provinces of China and many other countries (Munster et al., 2020). Thus, the outbreak caused panic across the globe. The social spread surrounding the outbreak had a significant impact on the public anxiety and social media became one of the concerning sources in spreading the information

(Nawrat, 2020). Since then, most governments have enacted public health responses to limit the spread of the disease.

Though literature on COVID-19 is developing rapidly, as of July 2020, there has been only preliminary research published on public attitudes towards COVID-19 on social media. Much of the related literature produced to date focus on identifying misinformation related to the COVID-19 outbreak (Cinelli et al. 2020; Ahmed et al, 2020), or about public attitudes expressed on Chinese social media (Li et al, 2020; Zhao & Xu, 2020). However, Lwin et al. (2020) conducted analysis on the emotional response to nearly 20 million tweets collected between January 28th and April 9th, 2020 and found that attitudes among Twitter users transitioned from fear to anger as the outbreak progressed. The study conducted by Lwin et al. (2020) did not observe differences among jurisdictions, and largely focused on the very early stages of the outbreak. There is therefore a gap in the research which we seek to address in this paper.

Canada and the United States share many similarities in history, culture, and living standards (Prus, 2011), though their health care systems are very different. The expenditures on health care in the US is relatively more than Canada, though is supported largely from private funds (Fuchs & Hahn, 1990). “Canada’s universal, publicly funded health-care system—known as Medicare—is a source of national pride and a model of universal health coverage. It provides relatively equitable access to a physician and hospital services through 13 provincial and territorial tax-funded public insurance plans” (Martin et al., 2018). In the US health care system, patients must acquire insurance coverage which is provided from either private and public sectors (Gohmann & Stephan, 2010). Uninsured Americans are much more likely to report serious access barriers than insured Americans in the United States (Kennedy et al., 2006). A survey of patients’ experiences with health care services is conducted in eleven different countries. The study revealed that “The United States trailed other countries in making health care affordable and ranked poorly on providing timely access to medical care (except specialist care)” (Osborn et al., 2016).

Several works of literature present Canada-US differences in attitudes towards health care. For example, a study by Prus (2011) directly compared the social factors associated with health in two countries. Nakayama (2011) examined pediatric surgeons in the US and Canada regarding their attitudes and experience in healthcare insurance. Thus, researchers have often compared health care concerns between both the countries. However, differences in pandemic responses and outbreak trajectories between the two countries have generated considerable additional popular interest, in part due to the stark differences in cases between the two countries (Markusoff, 2020). A comparison between social media attitudes in the United States and Canada could thus establish a relationship between public support for action and policy choices that politicians eventually made.

2.2 Twitter, word frequency and sentiment analysis

Twitter is a microblogging platform which allows individuals to communicate through tweets. It is a powerful social networking tool for both business and personal use, it uses “tweets” – short texts of 280- character maximum to post messages on Twitter (Southern, 2019). It is considered as one of the fastest methods to spread information to a large number of users. As of the fourth quarter of 2019, Twitter had global usage of 152 million monetizable daily active users of which 59.35 million lived in the United States (Clement, 2020).

The tweeted information is often multidimensional, time varying and mutable, so it is difficult to visualize the features of popular topics/words. A study by Changbo et al., (2013) analyzed public sentiments and predicted the short-term trend of the sentiment about the event using cellular automata. The results highlighted the dominant viewpoint and short-term trend of public sentiments on the web. To further understand and discover public feelings based on social media conversation, word frequencies provides in-depth insights of the sentiment. Several studies adapted statistical and mathematical approaches for identifying word frequencies. Word frequency distributions were identified as a useful tool for learning about texts as far back as the early 1990s (Baayen, 1992) and continue to be used to this day. For example, a recent study by Raghupathi et al., (2020) reviewed attitudes towards vaccination to the rise of medical misinformation via social media. They used word frequencies as a major analysis to discover different vaccines related knowledge of public on social media.

Sentiment analysis is a form of opinion-mining and is mainly focused on the classification of sentiment polarity (i.e. positive or negative), typically detected by a machine-learning approach (Kim, 2016). “Sentiment Analysis is a term that includes many tasks such as sentiment extraction, sentiment classification, subjectivity classification, summarization of opinions or opinion spam detection, among others. It aims to analyze people's sentiments, attitudes, opinions emotions, etc. towards elements such as products, individuals, topics, organizations, and services” (Kharde & Sonawane, 2016). In recent years, considerable research work has been conducted using sentiment analysis on Twitter. It has been handled as a natural language processing task at many levels of granularity (Bagheri et al., 2017). Initially, it was intended for binary classification which provided only bipolar classes such as positive and negative (Kharde & Sonawane, 2016). Later, multiple machine learning algorithms were introduced such as naïve Bayes, maximum entropy classification, support vector machines, artificial neural networks, decision trees, k-nearest neighbors, and ensemble learning, since the sentiment polarity detection can be seen as a classification task (Han et al., 2018). Some studies used external resource-based lexicons to improve performance of sentiment analysis. The study by (Han et al., 2018) performed sentiment lexicon called SentiWordNet (SWN) and review data, the results showed bias processing strategy and reduced polarity bias rate (PBR).

Twitter has been used by the public to express their emotions on infectious disease outbreaks. A research study by Kim et al., (2016) investigated topic coverage and sentiment dynamics of the hot health issue of the Ebola virus on Twitter. The study used sentiment analysis to explore public response by analyzing articles, publications and tweets with Twitter stream API. The results revealed Twitter has narrow topic coverage than news media, it further provided with reports on distinct news outlets on Twitter and news media. Besides, Guidry et al. (2017) argued that major health organizations such as the World Health Organization (WHO) addressed the Ebola crisis on Twitter. Another study by García-Díaz et al., (2020) built an ontology-based sentiment analysis model to measure public's opinions as regards infectious diseases expressed in Spanish using tweets concerning Zika, Dengue, and Chikungunya viruses in Latin America. They measured the relationship between disease domains such as risks, drugs, symptoms and other concepts based on statistical and linguistic sentiment analysis features, applying deep learning models. All in all, the literature demonstrated Twitter use in providing multiple insights of public's sentiment on disease outbreak.

3. METHODS

3.1 Data collection

Effective data collection and cultivation was one of the goals of this project. As such, we used Netlytic, a web-based system to extract tweets from Twitter. Netlytic is a cloud-based text and social network analysis tool that allows users to capture and import online conversational data from several social media platforms such as Twitter, Instagram and so on (Gruzd et al., 2016).

In order to control for differences between states and provinces that had not yet been directly impacted by COVID-19 we first identified states and provinces most affected by COVID-19 in the USA and Canada and collected data only from those jurisdictions. Based on live tracking and statistics on the Johns Hopkins COVID-19 Map (n.d.), we chose 5 jurisdictions from each country: New York, New Jersey, California, Washington and Florida from USA; Ontario, British Columbia, Quebec, Alberta and Nova Scotia from Canada. Then, by using the Netlytic text analysis tool, we extracted tweets by developing a search query with these provinces as user location.

Initially, we collected a sample of 955,317 tweets (Twitter messages) about the Covid-19 from the users in the previously identified affected US states and and 695,985 tweets from Canada between March 23, 2020 to April 24,2020. Tweets were retrieved using Netlytic's NEAR: query, which collects tweets from a given radius around a specified jurisdiction and the keywords used to search these tweets are #covid, #coronavirus, and #covid-19. The time period was chosen because it was the time where the number of COVID-19 cases had suddenly raised in both the countries (COVID-19 Map, n.d.) and all the services in the two countries began to lock down.

However, there were limitations to the Netlytic tool. The datasets retrieved through Netlytic's analysis tool consisted data other than the specified jurisdictions and included Tweets made outside of the USA and Canada. There were also tweets with no detailed location or with only country data. To manage this, we combined the tweets into two major datasets representing the data of USA and Canada. We then filtered the datasets using by removing the tweets other than USA and Canada location. The final dataset of USA consisted 735,666 tweets and Canada dataset contained 269,692 tweets. Table 1 summarizes the number of tweets extracted from the users of different jurisdictions and the number of tweets remaining after our filtering process.

Table 1 - Total number of tweets extracted from between March 23 and April 24, 2020.

Jurisdiction	No. of Tweets	Jurisdiction	No. of Tweets
New York	194 871	Ontario	183 117
New Jersey	186 731	Quebec	156 010
California	188 419	Alberta	140 488
Washington	198 661	British Columbia	120 689
Florida	186 635	Nova Scotia	95 681
Total (USA)	955 317	Total (CAN)	695 985
Total (USA, filtered)	735 666	Total (CAN, filtered)	269 692

3.2 Sentiment analysis

Sentiment analysis is a type of opinion mining and was used in this study. It is a sub-field of natural language processing (NLP) which identifies and extract opinions to gauge the attitude, sentiments, and emotions of a writer based on the computational treatment of subjectivity in a text (Pandey, 2019). Sentiment analysis often concerns polarity detection, (i.e., only positive, negative and neutral sentiment) rather than discrete emotions (Bae & Lee, 2012).

We conducted sentiment analysis on the tweets of two datasets separately, using the Valence Aware Dictionary and sEntiment Reasoner (VADER), which is based on a dictionary of a set of words with positive or negative sentiment scores (Hutto & Gilbert, 2014). VADER calculates the sentiment score of a comment by summing up the sentiment scores of each word in the dictionary. The sentiment score of the tweet is counted as positive if the words in the tweet match with a positive emotion in the dictionary. It is counted as negative, when they match with negative emotion in the dictionary.

The metric used to compute the compound score is the sum of valence scores of each word in lexicon, adjusted according to VADER's rules, and then normalized to be between -1 (most extreme negative), +1 (most extreme positive) and 0 (neutral). Sentiment neutral, negative, and positive scores thus provide multi-dimensional measures of the sentiment (Hutto & Gilbert, 2014). Compound score and sentiment type specifies sentiment score and type of the tweet; in this paper we hereafter describe states below the -0.05 threshold as "negative" and above the 0.05 threshold as "positive," as specified by Hutto & Gilbert (2014).

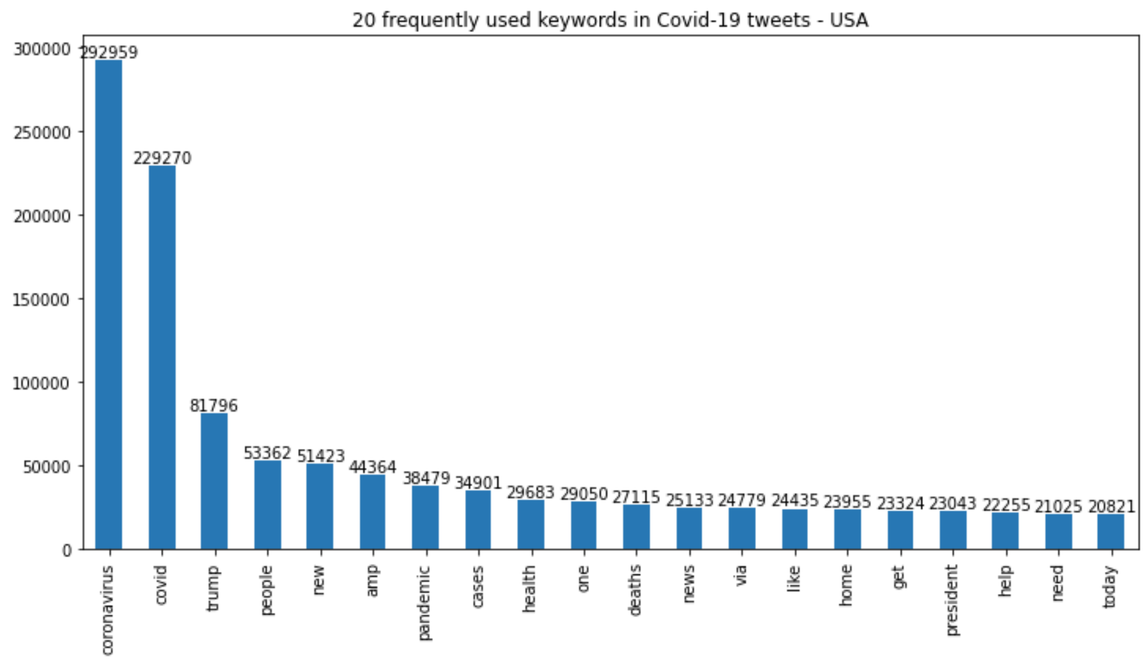
The tools used to complete this task were the Python Natural Language Toolkit (NLTK) (Bird, Klein & Loper, 2009) and the Scikit-learn machine learning library (Pedregosa et al., 2011). The SklearnClassifier class is a wrapper class around a scikit -learn model to make it conform to NLTK's classifier interface (Perkins et al., 2014). To provide more insights on the sentiment analysis, non-informative texts and stop words are filtered by via NLTK library called NLTK corpus, which contains a series of stop words. A variety of analyses were then conducted using analyzed data with the goal of generating descriptive characteristics of the sentiment within the country and comparisons between the countries. These analyses either compared the number of tweets classified as positive or negative for the two countries and the most frequently used terms for each class.

4. RESULTS

We began by analyzing data from the USA dataset. 37.6% of the tweets in this dataset were classified as negative time, while 36.6% were classified as positive and 25.8% were classified as neutral. Figure 1 illustrates the 20 frequent words used by Twitter users in this dataset. The words like 'new', 'cases', 'help, 'death' gives a glimpse at their panic and concern towards the pandemic. The political topics like 'Trump', 'president' also appear among the 20 most frequent words.

When applied to the Canadian dataset however, more tweets were classified as positive, as opposed to negative. Only 33.0% of tweets were classified as negative while 40.3% of tweets classified as positive, and 26.7% as neutral. Figure 2 illustrates the 20 most frequent words used by Twitter users in the Canadian dataset. The words like 'cases', 'care', 'help' and 'government', 'response', and 'update' shows concern towards the outbreak and provides favourable situations of the users. Interestingly, we observed different relative frequencies of the terms "coronavirus" and "covid" in this dataset. Figure 3 compares the frequencies of sentiment between the two datasets.

Figure 1 – Most frequent words used in the USA dataset



When comparing overall sentiment, tweets in the USA dataset were more likely to be classified as negative, whereas tweets in the Canada dataset had more positive. Figure 3 reveals that USA tweets has a minor difference in positive and negative sentiments, though the differences were pronounced among Canadian tweets. This is due to the low compound polarity score maintained by the tweets. Other interesting finding noticed in the analysis is the low number of neutral sentiments in both the countries. This reveals that people always conveyed an emotion (either positive or negative) about the situation. The above two graphs represent the differences in the sentiment from the tweets related to COVID-19 between Canada and USA. A two-sided independent t-test was conducted among the USA and Canadian tweets and revealed that the difference in positive sentiment was significant (t = -51.95; p < 0.0001). The trend also held throughout the duration of the dataset. Figure 4 illustrates variations in net sentiment score over the time period. Even though USA tweets had marked eight days which exceeded a threshold of 0, these days were, it maintained low sentiment score, always less than 0.005 polarity. On the other hand, Canada tweets always preserved high sentiment score between 0.018 and 0.074 polarity score over the duration.

Figure 2 – Most frequent words used in Canada dataset

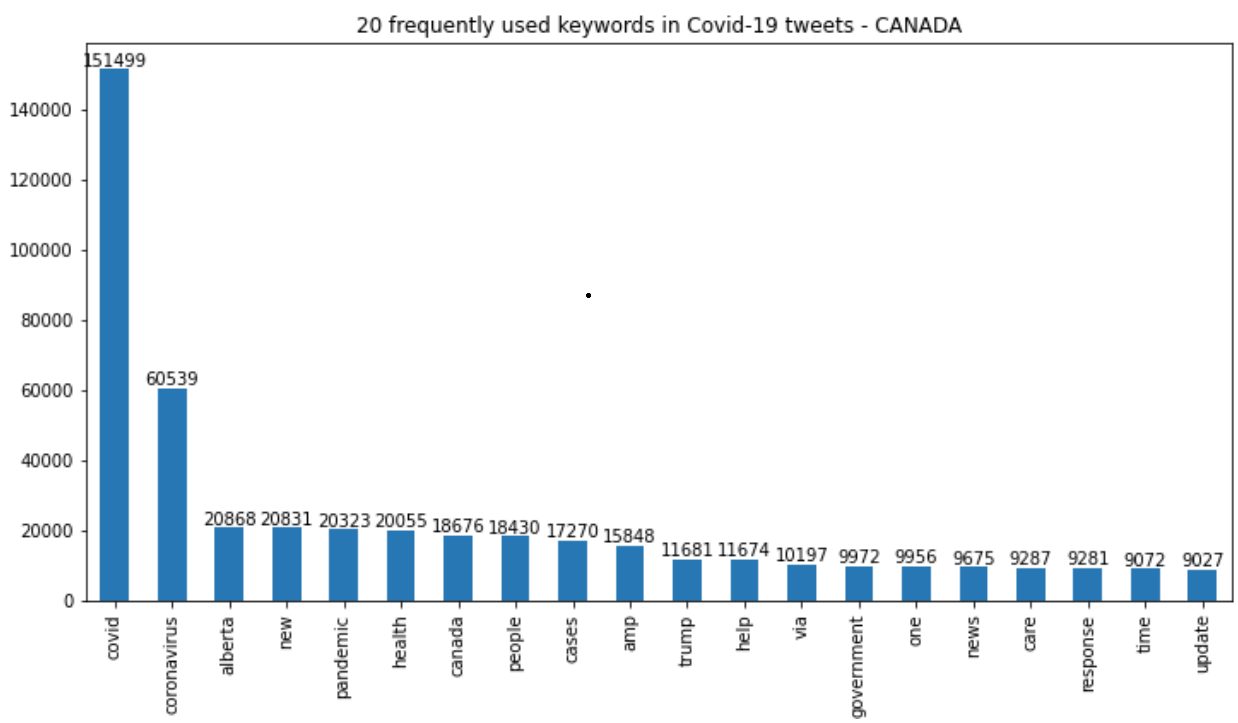


Figure 3 – Comparison of tweets in the USA and Canada datasets by percent of sentiment

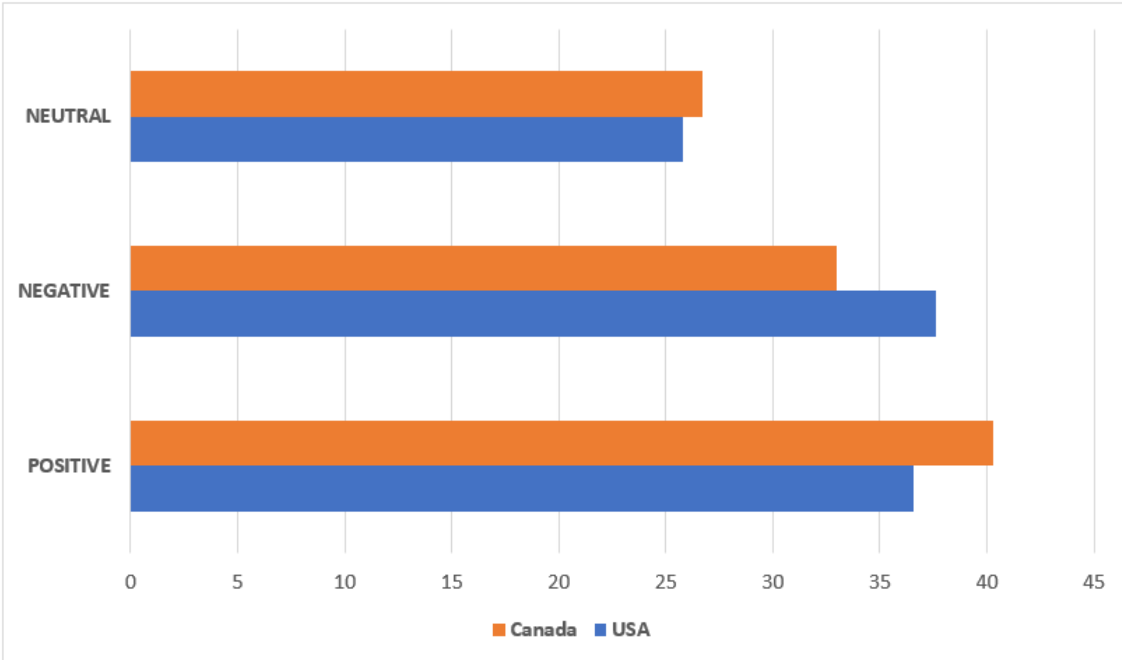
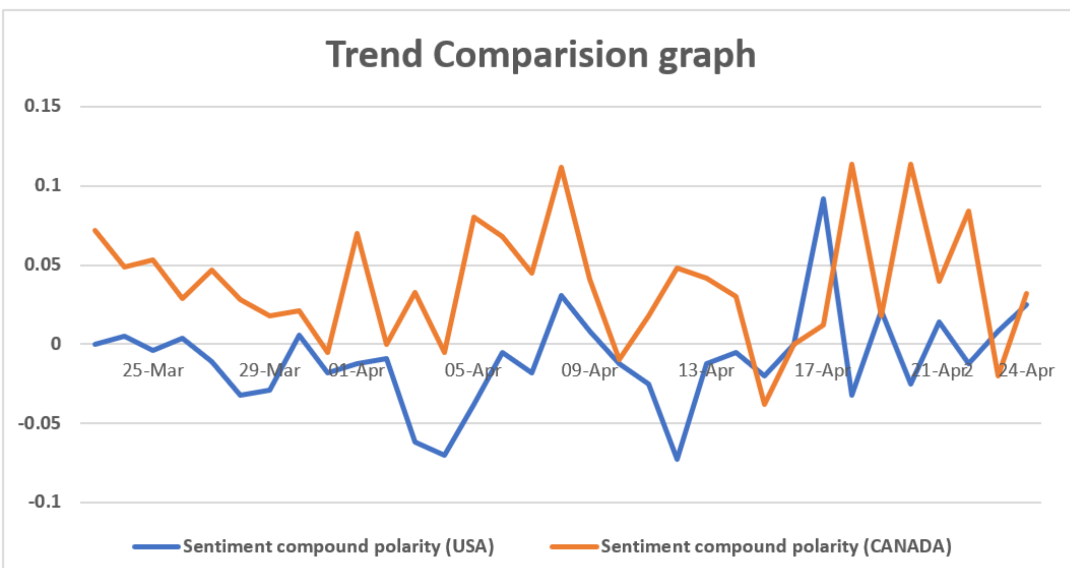


Figure 4 – Comparison of sentiment trends in USA and Canada during 22 March – 24th April



Finally, we also visually investigated the word frequencies within the subset of positive and negative tweets for each country. To effectively communicate this, we generated word clouds which depict term frequencies using size; term size is illustrated proportionately to their relative frequency in the dataset. Figures 5a and 5b depict the word frequencies for the positively classified and negatively classified tweets from the United States, while Figures 6a and 6b similarly depict the Canadian tweets. Note that terms investigated may have been processed to account for plurals and alternative spellings of the words (e.g. "United State").

Figure 5a – Word cloud generated from positive sentiment tweets from the United States dataset.

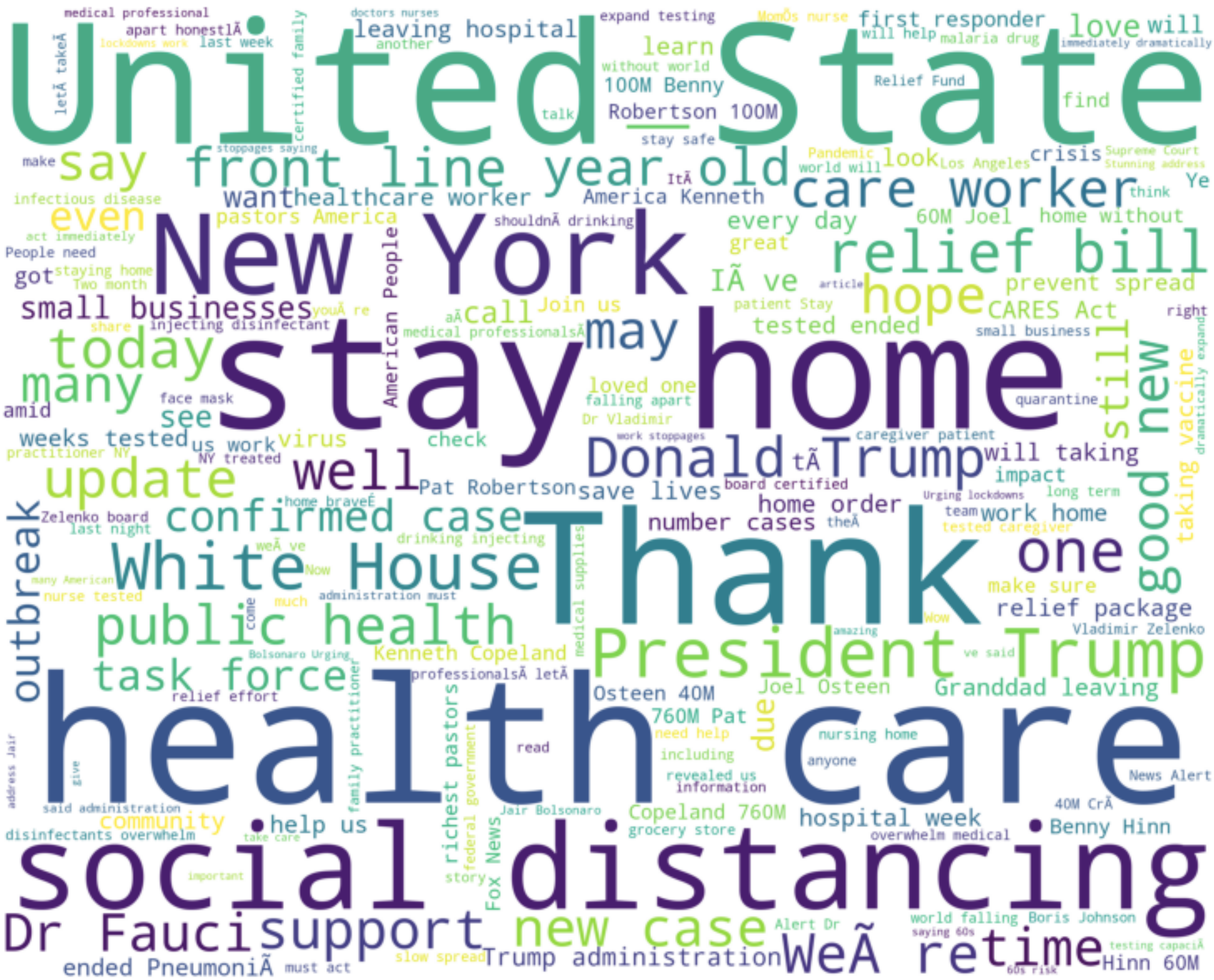


Figure 5b – Word cloud generated from negative sentiment tweets from the United States dataset.

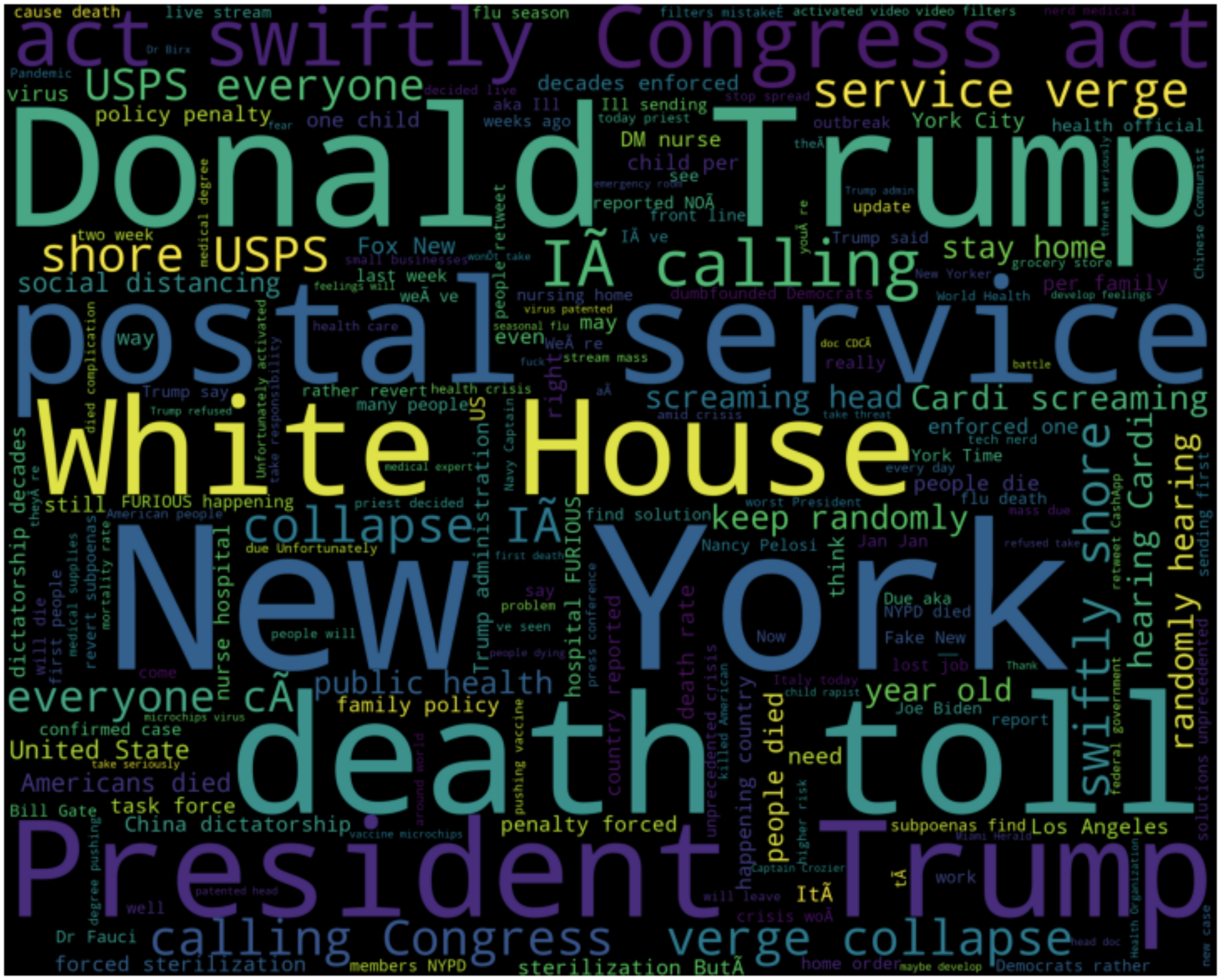


Fig. 5(a) and (b) shows the word cloud for positive and negative sentiments respectively. It was found that both the sentiments contained words like ‘President Trump’, ‘social distancing’, ‘White House’. These results were not investigated for statistical significance but were provided to effectively illustrate prospective differences in classified sentiment. Results were not compared using statistical significance because results that were classified differently by machine learning cannot be assumed to originate from similar population sources.

Figure 6a – Word cloud generated from positive sentiment tweets from the Canada dataset.

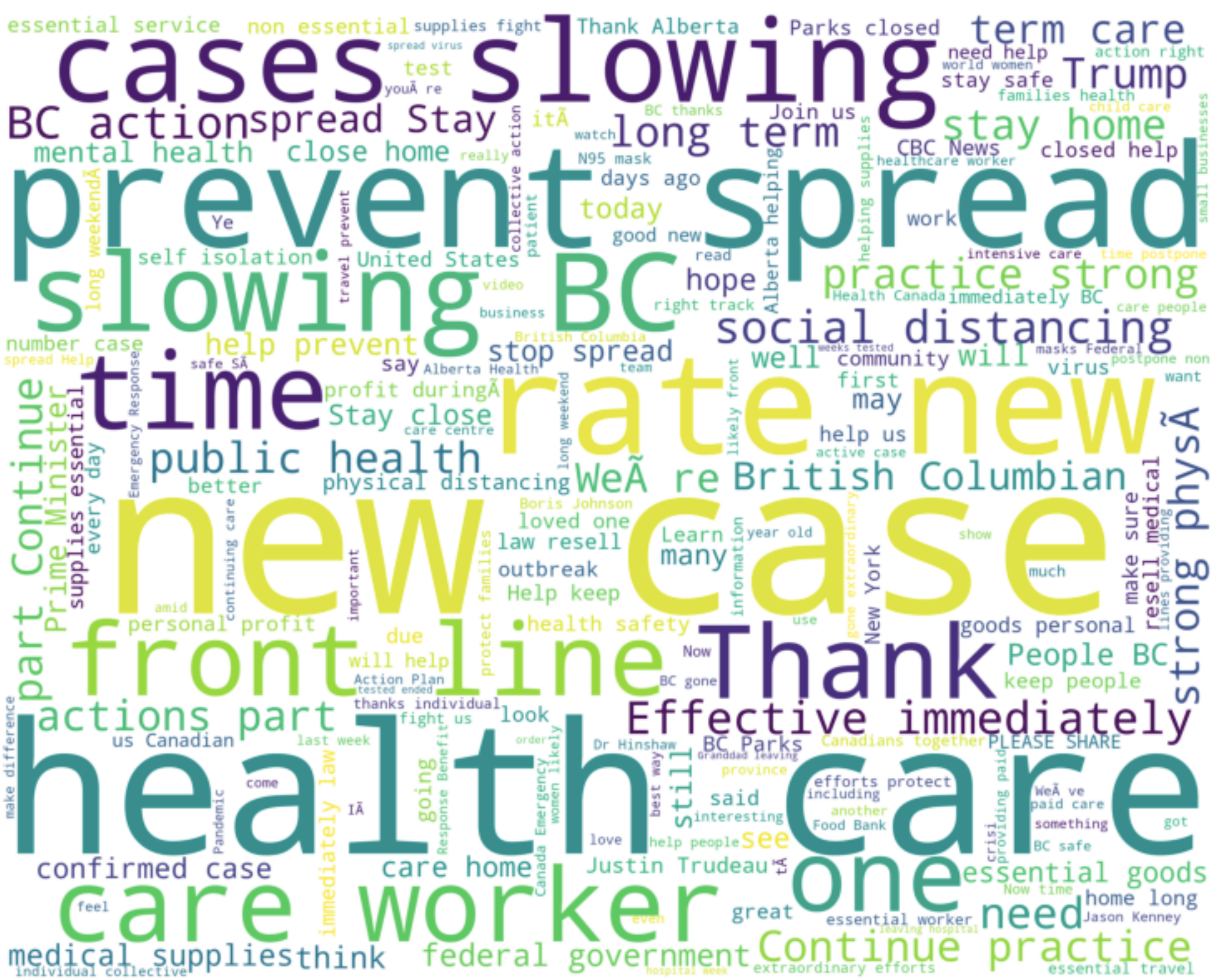
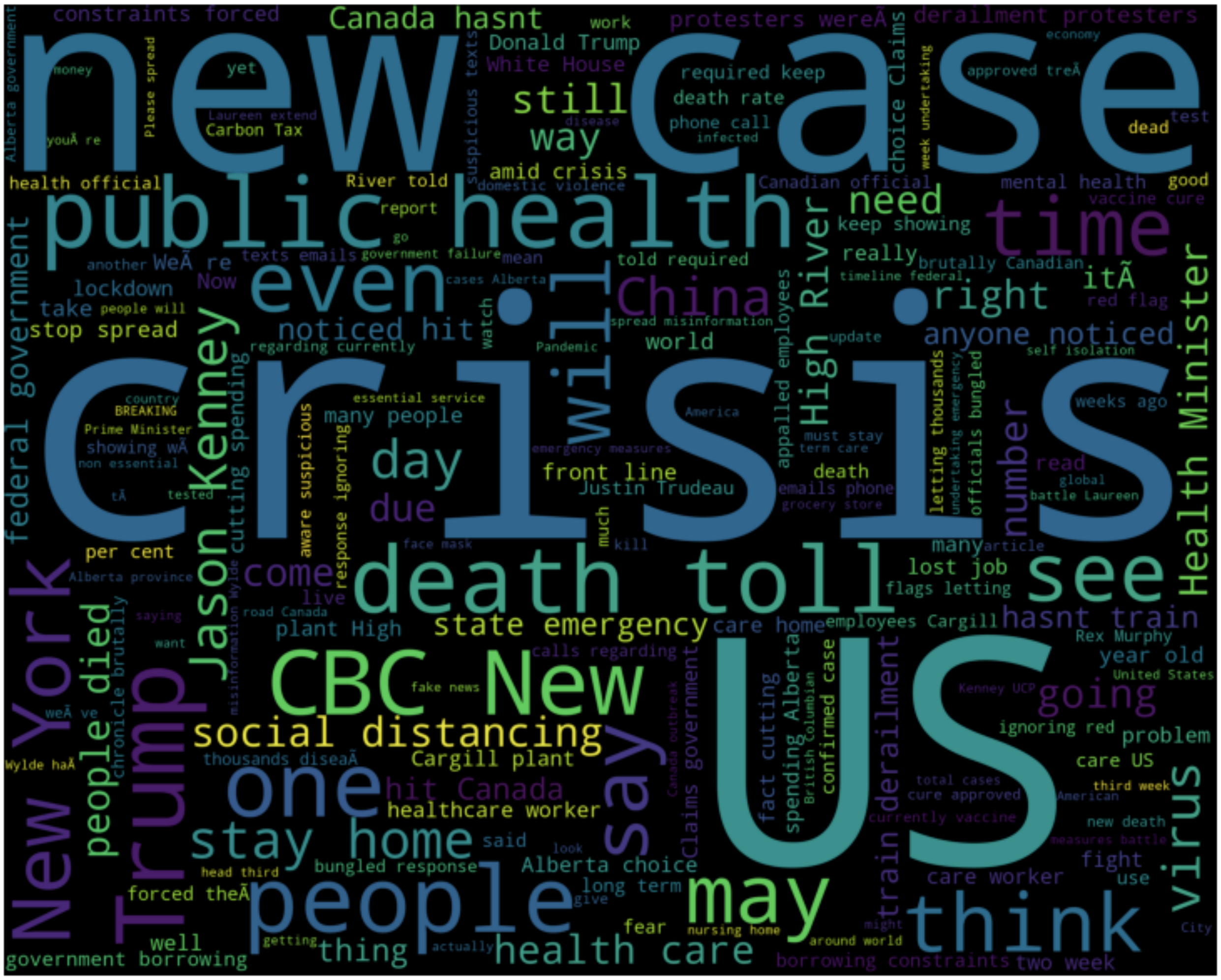


Figure 6b – Word cloud generated from negative sentiment tweets from the Canada dataset.



5. DISCUSSION

These findings demonstrate that the content and sentiment of tweets related to on COVID-19 are different between the two countries. We discovered that Canadian tweets were statistically more likely to be classified as positive by the VADER classifier than the USA tweets (t = -51.95; p < 0.0001). . In addition, the investigation on the frequency words used by the users of both the countries revealed differences that may reflect perceptions towards the outbreak. In the USA dataset, tweets about President Donald Trump were relatively frequent, while in Canada many tweets contained words related to the government's pandemic response, possibly reflecting differences in political dialogue between the countries. Among United States tweets, we discovered a greater frequency of negative tweets concerning the state of New York, which had then suffered the first widely publicized outbreak of COVID-19 in North America, as well as many negative tweets concerning Donald Trump. In Canada by contrast, discussion about political figures were infrequent in the overall dataset, and reference to new cases were found among both positive and negative tweets. References to the actions of the Government of British Columbia were found among positive sentiment tweets, while the frequency of the term "US" may reflect increased concern with the happenings of the United States. The illustrations in Figures 5a, 5b, 6a and 6b reinforce this observation; the size of political terms (e.g. Donald Trump, white house) were proportionately large in negatively classified tweets, while not in positively classified tweets.

When applied to public policy, it is important to note that the situations in the two countries were very different during this time period. Moreover, health care policies and medical response systems are very different among the two countries (Kant, 2014). The public response to COVID-19 were ultimately very different among the two countries, with the Canadian Federal government taking initiative to prevent the spread of the virus in March (Cohen, March 20, 2020). Consequently, people in Canada may have been more prone to support the government decision and involved more in sharing information during COVID-19 crisis. This offers an explanation for the differences among the two countries. We can infer from this that the differences in negative sentiment towards public offices reflected differences in public support for the government's actions on COVID-19 between the two countries.

Nonetheless, there are limitations to our findings. First, our investigation only observed three dimensions to sentiment: positive, negative and neutral, as provided by the VADER approach. It is possible that there are is further information which could be discovered using a different approach to sentiment, such as by analyzing the attitudes into more specific moods (e.g. happy, kind, calm, afraid, sad, angry). Second, due to technical challenges with our data collection, we were unable to observe differences in jurisdictions within the countries which might have been useful to policymakers. By overcoming this technical limitation, future research efforts could benefit by observing differences between jurisdictions to inform policies on the state or provincial level.

6. CONCLUSION

By employing appropriate data collection and sentiment analysis techniques, we revealed the differences in the attitudes of people in USA and Canada during COVID-19 outbreak. The study discovered two major differences. First, it revealed significant differences in positive and negative attitudes between users from the two countries. Second, it discovered differences in word frequencies, illustrating differences in reactions to the outbreak. Both observations may be the result of difference in public attitudes towards the coronavirus response or due to differences in the countries' healthcare systems. We believe that these results and especially the methods employed by this study may guide future research into public reaction to COVID-19 and ultimately inform better policies in the future.

7. REFERENCES

Ahmed, W., Vidal-Alaball, J., Downing, J., & Seguí, F. L. (2020). COVID-19 and the 5G conspiracy theory: social network analysis of Twitter data. *Journal of Medical Internet Research, 22*(5), e19458.

Alessa, A., & Faezipour, M. (2019). Flu outbreak prediction using Twitter posts classification and linear regression with historical centers for disease control and prevention reports: Prediction framework study. *JMIR Public Health and Surveillance, 5*(2), E12383.

Bae, Y., & Lee, H. (2012). Sentiment analysis of twitter audiences: Measuring the positive or negative influence of popular twitterers. *Journal of the American Society fo*r *Information Science and Technology, 63*(12), 2521-2535.

Bagheri, H., & Islam, M. (2017). Sentiment analysis of twitter data. *arXiv preprint*. arXiv: 1711.10377.

Baayen, H. (1992). Statistical models for word frequency distributions: A linguistic evaluation. *Computers and the Humanities, 26*(5), 347-363.

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: analyzing text with the natural language toolkit*. O'Reilly Media, Inc.

Changbo Wang, Zhao Xiao, Yuhua Liu, Yanru Xu, Aoying Zhou, & Kang Zhang. (2013). SentiView: Sentiment Analysis and Visualization for Internet Popular Topics. *IEEE Transactions on Human-Machine Systems, 43*(6), 620-630.

Cinelli, M., Quattrociocchi, W., Galeazzi, A., Valensise, C. M., Brugnoli, E., Schmidt, A. L., ... & Scala, A. (2020). The covid-19 social media infodemic. *arXiv preprint*. arXiv:2003.05004.

Clement, J. (2020, February 14). Countries with most Twitter users 2020. Retrieved from https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/

Cohen, A. (2020, March 26). Cohen: Why Canada's response to COVID-19 is so different from that of the U.S. Retrieved from https://ottawacitizen.com/opinion/cohen-why-canadas-response-to-covid-19-is-so-different-from-that-of-the-u-s/

COVID-19 Map. (n.d.). *Johns Hopkins University & Medicine*. Retrieved from https://coronavirus.jhu.edu/map.html

Fuchs, V., & Hahn, J. (1990). How does Canada do it? A comparison of expenditures for physicians' services in the United States and Canada. *The New England Journal of Medicine, 323*(13), 884-890.

Gadarian, S. K., Goodman, S. W., & Pepinsky, T. B. (2020). Partisanship, health behavior, and policy attitudes in the early stages of the COVID-19 pandemic. *SSRN preprint*. SRN: 3562796

Gohmann, Stephan F. A comparison of health care in Canada and the United States: The case of pap smears. *Medical Care 48*(11), 1036-040.

García-Díaz, J., Cánovas-García, M., & Valencia-García, R. (2020). Ontology-driven aspect-based sentiment analysis classification: An infodemiological case study regarding infectious diseases in Latin America. *Future Generation Computer Systems, 112*, 641-657.

Gruzd, *A. (2016).* Netlytic: Software for automated text and social network analysis. Retrieved from http://Netlytic.org

Gruzd, A., Paulin, D., & Haythornthwaite, C. (2016). Analyzing social media and learning through content and social network analysis: A faceted methodological approach. *Journal of Learning Analytics*, *3*(3), 46–71. DOI: 10.18608/jla.2016.33.4

Guidry, J., Jin, Y., Orr, C., Messner, M., & Meganck, S. (2017). Ebola on Instagram and Twitter: How health organizations address the health crisis in their social media engagement. *Public Relations Review, 43*(3), 477-486.

Han, Hongyu, Zhang, Yongshi, Zhang, Jianpei, Yang, Jing, & Zou, Xiaomei. (2018). Improving the performance of lexicon-based review sentiment analysis method by reducing additional introduced sentiment bias. *PLoS One, 13*(8), E0202523.

Hutto, C. J. & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the Eighth International Conference on Weblogs and Social Media (ICWSM-14), Ann Arbor, MI, June 2014*.

Kant P. & Mark E. Rushefsky (2014) Healthcare politics and policy in America. *Public Integrity, 17*(1), 94-96. doi: 10.2753/PIN1099-9922170107

Kennedy, J., & Morgan, S. (2006). Health care access in three nations: Canada, Insured America, and Uninsured America. *International Journal of Health Services, 36*(4), 697-717.

Kim, Erin Hea-Jin, Jeong, Yoo Kyung, Kim, Yuyoung, Kang, Keun Young, & Song, Min. (2016). Topic-based content and sentiment analysis of Ebola virus on Twitter and in the news. *Journal of Information Science, 42*(6), 763-781.

Kharde, V., & Sonawane, P. (2016). Sentiment Analysis of Twitter Data: A Survey of Techniques. *arXiv preprint*. arXiv: 1601.06971

Li, Q., & Feng, W. (2020). Trend and forecasting of the COVID-19 outbreak in China.

Li, J., Xu, Q., Cuomo, R., Purushothaman, V., & Mackey, T. (2020). Data mining and content analysis of the Chinese social media platform Weibo during the early COVID-19 outbreak: retrospective observational infoveillance study. *JMIR Public Health and Surveillance, 6*(2), e18700.

Li, Qun, Guan, Xuhua, Wu, Peng, Wang, Xiaoye, Zhou, Lei, Tong, Yeqing, . . . Feng, Zijian. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *The New England Journal of Medicine, 382*(13), 1199-1207. doi: 10.1056/NEJMoa2001316

Lwin, M. O., Lu, J., Sheldenkar, A., Schulz, P. J., Shin, W., Gupta, R., & Yang, Y. (2020). Global sentiments surrounding the COVID-19 pandemic on Twitter: analysis of Twitter trends. *JMIR public health and surveillance, 6*(2), e19447.

Markusoff, J. (15 July 2020). Every U. S. state is accumulating new virus cases at a faster clip than Canada. Yep, all 50. *Maclean's*. Retrieved from: https://www.macleans.ca/society/health/every-u-s-state-is-accumulating-new-virus-cases-at-a-faster-clip-than-canada-yep-all-50/

Martin, D., Miller, A., Quesnel-Vallée, A., Caron, N., Vissandjée, B., & Marchildon, G. (2018). Canada's universal health-care system: Achieving its potential. *The Lancet, 391*(10131), 1718-1735.

Merkley, E., Bridgman, A., Loewen, P., Owen, T., Ruths, D., & Zhilin, O. (2020). A Rare Moment of Cross-Partisan Consensus: Elite and Public Response to the COVID-19 Pandemic in Canada. *Canadian Journal of Political Science,* 1-8. doi:10.1017/S0008423920000311

Munster, V., Koopmans, M., Van Doremalen, N., Van Riel, D., & De Wit, E. (2020). A Novel coronavirus emerging in China - Key questions for impact assessment. *The New England Journal of Medicine, 382*(8), 692-694.

Nakayama , D., & Langer, K. (2011). Single payer health insurance in pediatric surgery: US impressions and Canadian experience. *Pediatric Surgery International, 27*(3), 329-334.

Nawrat, A. (2020, February 27). Covid-19 outbreak: How misinformation could fuel global panic. Retrieved from https://www.pharmaceutical-technology.com/features/covid-19-outbreak-how-misinformation-could-spark-global-panic/

Odlum, M., & Yoon, S. (2018). Health Information Needs and Health Seeking Behavior During the 2014-2016 Ebola Outbreak: A Twitter Content Analysis. *PLoS Currents, 10*, PLoS currents, March 23, 2018, Vol.10.

Osborn, R., Squires, D., Doty, M., Sarnak, D., & Schneider, E. (2016). In new survey of eleven countries, US adults still struggle with access to and affordability of health care. *Health Affairs (Project Hope), 35*(12), 2327-2336.

Pandey, P. (2019, November 8). Simplifying Sentiment Analysis using VADER in Python (on Social Media Text). *Analytics Vidhya*. Retrieved from https://medium.com/analytics-vidhya/simplifying-social-media-sentiment-analysis-using-vader-in-python-f9e6ec6fc52f

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*(Oct), 2825-2830.

Perkins, J., & Fattohi, Faiz. (2014). *Python 3 text processing with NLTK 3 cookbook: Over 80 practical recipes on natural language processing techniques using Python's NLTK 3.0* (Second ed.). Birmingham, England: Packt Publishing.

Prus, S. G. (2011). Comparing social determinants of self-rated health across the United States and Canada. *Social Science & Medicine*, *73*(1), 50–59. doi: 10.1016/j.socscimed.2011.04.010

Raghupathi, Viju, Ren, Jie, & Raghupathi, Wullianallur. (2020). Studying Public Perception about Vaccination: A Sentiment Analysis of Tweets. *International Journal of Environmental Research and Public Health, 17*(10), International journal of environmental research and public health, May 15, 2020, Vol.17(10).

Southern, S. (2019, February 1). The Purpose of Twitter. Retrieved from https://itstillworks.com/purpose-twitter-12292.html

Zhao, Y., & Xu, H. (2020). Chinese public attention to COVID-19 epidemic: Based on social media. *medRxiv preprint*. doi: https://doi.org/10.1101/2020.03.18.20038026

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., … Tan, W. (2020). A Novel Coronavirus from Patients with Pneumonia in China, 2019. *New England Journal of Medicine*, *382*(8), 727–733. doi: 10.1056/nejmoa2001017