Accepted manuscript version, cite: Hodson, J., Veletsianos, G., & Houlden, S. (2021). Public responses to COVID-19 information from the public health office on Twitter and YouTube: Implications for research practice. Journal of Information Technology & Politics, 1–9. https://doi.org/10.1080/19331681.2021.1945987

Title: Public responses to COVID-19 information from the public health office on Twitter and YouTube: Implications for research practice

Abstract: We collected tweets directed at the official Twitter account of the Canadian Public Health Office as well as comments on a Canadian Public Health Office press conference posted to YouTube. We used a mixed method corpus-assisted discourse analysis approach to categorize and analyze these data. We found key differences between comments on each platform, namely differences in tone and sarcasm in YouTube comments, and more balance in Twitter mentions. Findings suggest that studying public responses to health information on one platform in isolation does not provide an accurate picture. To generate a fuller picture of misinformation, researchers should conduct studies across digital platforms using diverse methods. This research could influence how studies of health communication and public opinion are approached in the future.

Keywords: social media; COVID-19; public opinion; YouTube, Twitter; methods; interdisciplinary approaches

Introduction

As the COVID-19 pandemic unfolds, misinformation about the the disease has proliferated in what the World Health Organization (WHO) has called an “infodemic,” a term originating out of concern for a similar information crisis tied to SARS outbreak in 2003 (Rothkopf, 2003). COVID-19 misinformation has medical, cultural, social and political facets. Given this context, successful mitigation of this pandemic requires action linking public health, science and politics both internationally and in each country around the world (McCloskey and Heymann, 2020). Such action requires complex analyses of public opinion in response to health messaging, and thus necessitates innovative methods of doing so. While algorithmic methods can demonstrate how health misinformation travels online (Song and Gruzd, 2017), and how some information "goes viral" (Nahon and Hemsley, 2013), there is widespread agreement that algorithmic approaches to understanding much of the data on social media, including COVID-19 misinformation, are inadequate to fully account for the spectrum of public opinion (Bechmann et al., 2018; Vraga and Bode, 2018), particularly considering the unique challenges of the current COVID-19 misinformation environment. Algorithmic data analysis, such as the ones popularized recently by Big Data approaches to analysis, refer to the development of computer-based rules and processes to examine, unpack and interpret numerical or textual data. One example would be sentiment analysis software that identify sentiment or emotion in a sentence.

In this exploratory research, we examine social media comments directed at Canada's Chief Public Health Officer (CPHO) during the first wave of the COVID-19 pandemic on both YouTube and Twitter as a way to understand how different online publics respond to government efforts to provide pandemic-related information, and the qualitative nature of the misinformation occurring in these responses. Our research reveals key differences between the public responses to CPHO on the different platforms.

Review of Relevant Literature

Health communication researchers argue that "the production and circulation of biomedical knowledge is increasingly a complex public process" (Hallin and Briggs, 2015, p. 98). This complicated landscape is problematic at a time when many people receive a large amount of their news from social media, and is exacerbated during a public health crisis, like the COVID-19 pandemic..

To manage a pandemic, politicians must both ask the public to engage in voluntary social behaviors (e.g., vaccinations, wearing a mask, social distancing for example), and also must try to get public buy-in for policy measures that may initially seem to limit personal freedoms, such as travel restrictions and quarantines (Bennett, 2020; Freimuth et al., 2014; Fridman et al., 2020; Lovari, 2020; Siegrist & Zingg, 2014; Taha et al., 2013a). Thus, pandemic control is an issue that requires politicians and public health professionals to effectively communicate and engage the public. However, social media, and the misinformation that travels on it, complicate this task (Bennett, 2020).

One social media platform closely intertwined with politics such as election campaigns (Jungherr, 2016) and social issues such as abortion rights (Hunt, 2019), is Twitter. Twitter is a common platform that researchers use to study online information flows in large part because the popular Twitter API allows for relatively easy scraping of large datasets for research purposes (Giglietto, Rossi, and Bennato, 2012). For these reasons, information posted on Twitter has become a proxy for public opinion, wherein researchers often use algorithmic text mining and automated content analysis to assess public views on policy or political issues (see Rogstad, 2016; Cody et. al., 2016; Karami, Bennett, and He, 2018). Twitter activity also seems to be an important driver with respect to agenda setting on traditional news sources like print media (Su and Borah, 2019). Nevertheless, while the number of Twitter users is growing in Canada and globally (Gruzd et al., 2020), the platform may not represent public opinion on an issue, since Twitter publics appear to represent a non-uniform sample of the overall population, with some studies for example showing that Twitter users are younger and over-represent men (e.g., Mellon and Prosser 2017; Mislove et al, 2011).

Therefore, despite Twitter's popularity, the platform may not provide a full picture of political or social issues if studied in isolation (Bechmann et al., 2018). Other platforms are also important for public engagement. For example, the Pew Research center reports that YouTube is becoming increasingly popular as a news source for young people (van Kessel, 2019).Additionally, research on political debates via YouTube suggests that populations that are underrepresented via other media may be more engaged with political issues via YouTube (Ricke, 2010); however, YouTube users also report exposure to misinformation on the platform, which could be problematic during a public health emergency.

Current approaches to understanding social media publics generally approach the issue from a single discipline, and use either algorithmic approaches (see for example, Jin et al., (2014), Song and Gruzd, (2017), and Jin et al. (2016)) or surveys (see overview in Authors, 2020). Surveys rely on self-reported data, which can result in participants telling researchers what they think will earn them social approval (Gove and Geerken, 1977). For this reason, algorithmic approaches are thought to be more accurate. One solution is to use multiple methods (Vraga and Bode, 2018; Hunt and Gruszczynski 2019). For example, Conroy, Rubin & Chen (2016) recommended that algorithmic methods be used "to augment human judgement, not replace it" (p. 4), suggesting the need for both quantification and also context-rich information for understanding social media. Results from such efforts may help to understand complex issues more fully. In the context of the unfolding COVID-19 pandemic, text mining of social media has been used to understand different aspects of this crisis (see for example, Tommasel et. al, 2021; Glowacki, Wilcox & Glowacki, 2020). We thus examined public responses to official government messaging by using a mixed methods approach as a way to address the following research questions:

- In what ways, if any, do the posts directed at the CPHO differ between YouTube and Twitter?
- What do the differences tell us about the efficacy of hybrid methods like corpus assisted discourse analysis for understanding COVID-19-related public opinion on social media?

Methods

To explore the public response to COVID-19-related government messaging on different platforms, we collected data from two sources: a sample of tweets mentioning the Canadian Public Health Office (CPHO) on Twitter and comments on video released and posted on YouTube by the CPHO. The selection of the CPHO allows us to draw boundaries around the choice of communication to examine on each platform so that we can compare two different platforms that are contributing to the same national political and social conversation. We were intentional about choosing two different social media platforms for comparison because prior research highlights the biases that can arise when focusing on a single platform (Tufekci, 2014). By juxtaposing Twitter (a platform favored by researchers conducting big data analysis), and YouTube (a growing source of news and political information, but less-studied than Twitter), we hoped to bring to light potential differences that arise between the two platforms, as a way to understand the limitations of algorithmic public opinion mining. We anticipated observing differences between the two platforms not only because the affordances of the two platforms vary (e.g., YouTube is centered on video; Twitter has a strict character length; they employ different algorithms), but also because (a) prior studies identified that user demographics differ between platforms (e.g., Mellon and Prosser 2017), and (b) as we were keeping ourselves informed about the pandemic, we observed anecdotal differences between the two.

Starting on April 18, 2020 we launched a web script which ran every few minutes. It used the Twitter Search Application Programming Interface (API), and it searched for and retrieved tweets that included the keyword *@CPHO_Canada*. A more detailed description of this script is provided in earlier studies (e.g., Authors 2016, 2018, 2020). Data collection continued until April 29, 2020, and collected a total of 16,563 tweets posted by 7,392 unique accounts. The median user contributed one tweet (M = 2.1; S.D. = 3.55). Approximately 25% of tweets were posted by 10% of users, and about 7.5% of tweets were posted by 1% of users.

On May 6, the office of the CPHO posted a video of a press conference on YouTube[1]. We used the netvizz YouTube Data Tool (Rieder, 2015) three times over a 24-hour period on May 6 and 7 to scrape the comments posted in response to this video. A total of 1,875 comments posted by 802 users were retrieved and the median user contributed one comment (M = 2.34; S.D. = 4.02). Approximately 61% of comments were posted by 10% of users, and about 20% of comments were posted by 1% of users. The Twitter data set contained 232,984 words, and the YouTube data set 42,266 words.

We analyzed the data using the mixed methods approach of computer assisted discourse analysis (CADA). While this method is relatively common in the digital humanities, it is still growing in use with respect to social media data, and to the best of our knowledge has not yet been used to understand COVID-19 discourses. To operationalize this method, we did the following: First, we used the WMatrix linguistic analysis platform (Rayson, 2009) to compare each sample against the AmE06 reference corpus containing one million words of American written and spoken English. The use of a reference corpus allowed us to determine whether word

[1] https://www.YouTube.com/watch?v=BwvY76-6gbQ

usage in our sample is significant in its difference from common word usage, and also allowed us to determine the relative scale of the differences - a standard and important step in corpus assisted discourse analysis (Gabrielatos, 2018). The scale of the differences is the important reason to compare the sample corpora to a reference corpus, because they will allow us to compare the YouTube comments to Twitter comments in a way that compensates for differences in the relative affordances of each platform with respect to post length. This takes the guesswork out of the resultant qualitative analysis of key parts of the text, and provides rigor by allowing us to focus on those parts of the text that are statistically significant (see also Aluthman, 2018; Altoaimy, 2018; Lewis, Zamith and Hermida, 2013). After identifying statistically significant keywords, we sorted by the LogRatio statistic to isolate the key words and phrases, and topics by effect size (Gabrielatos, 2018) and used the effect size to guide a qualitative reading of the sentences and paragraphs in which top keywords were located, in order to understand the context and nature of how the words were being used. We used a grounded theory approach (Charmaz, 2014) to understand the key themes in the sentence and paragraph level data, coding for patterns related to the posed research questions. The qualitative analysis enabled us to develop a more robust understanding of misinformation than the one that algorithmic methods alone generally afford, and allowed us to locate important differences between YouTube and Twitter responses. This approach to understanding the data through corpus linguistics is growing in popularity, but is still relatively rare with respect to the study of social media (Authors, in press). A good overview of how effect size helps to guide discourse analysis can be found in Gabrielatos (2018), which will help other interested researchers to see the details necessary for conducting their own CADA.

Results

Analysis of key @CPHO Tweet phrases

Tweets directed at the CPHO were mixed, with posts made both in support of the information coming out of the CPHO, and also posts made to insult or criticize the CPHO. While there were some references to conspiracy-type misinformation, notably accusations of collusion between the CPHO and the WHO in the provision of incorrect information to the public, there were also a number of posts encouraging the Government of Canada and supporting the Chief Public Health Officer, Dr. Tam. Table 1 shows the key phrases from the @CPHO_Canada tweets with a cut-off significance level of $p = 0.000014$ log-likelihood. This p-value is standard in (nonlinear) corpus analysis due to large corpora (Gabrielatos, 2018). In this table, the log ratio (LRatio) value shows the effect size, or in other words, the magnitude of the difference between the @CPHO_Canada tweets and the reference corpus for each phrase. Log ratio is a base 2 binary logarithmic measure, so a log ratio of 1 means the phrase is twice as common in this data set than in the reference corpus, 2 means the phrase is 4 times more common in our data set than the reference, 3 means the phrase is 8 times more common, and so on (Hardie, 2014).

Seven phrases are statistically significant, three of which relate to the name or title of the Chief Public Health Officer, Dr. Tam. A close qualitative reading of these words in context reveals that often, when Dr. Tam is mentioned, it is accompanied by the hashtag #IStandWithDrTam or expressions of appreciation. Similarly, the phrase “keep up” usually refers to tweets urging Tam to keep up the good work. On the other side of the coin, the phrase “step down” is used as a command for Dr. Tam to resign. Directives to step down accuse the CPHO of misinformation or incompetence, and of collusion with the WHO which is also accused of providing misinformation. The phrase "public health" has mixed use, with about half its uses being to complain about the incompetence or "failure" of Dr. Tam and the Government of Canada and the other half providing thanks.

Conspiracy and misinformation are not, upon a closer read, the main focus of key Twitter content directed at the CPHO account. Instead, unsupportive content seems to focus on personal attacks on the CPHO's intelligence and integrity. These tweets suggest a gendered and racial component to the results, which is an area that requires further research.This content was often countered by the tweets using the #IStandWithTam hashtag which provides support for the CPHO and policy, often in direct response to accusatory and negative content. These findings indicate that some users are supporting the official flow of COVID-19-related information provided by the government.

Table 1. @CPHO_Canada Tweets key phrases

Item	**LRatio**
Dr_Tam	11.55
Theresa_Tam	9.29
step_down	7.97
keep_up	6.29
thank_you	5.94
public_health	5.8

hard_work	4.48

Analysis of key YouTube phrases

In contrast to Twitter, YouTube comments directed at the CPHO contained more aggressive language, and were more focused on conspiracy, a strong anti-government theme, and a racist anti-China sentiment that was more amplified than the one observed on Twitter. These are shown in Table 2. The YouTube dataset was smaller than the @CPHO_Canada dataset, but since we analyzed each corpus independently in relation to a reference corpus, statistical analysis of effect size using log ratio measures should neutralize any differences related to corpus size (Gabrielatos, 2018). For that reason, the differences are even more striking. Significantly, the sheer number of statistically significant key phrases, which were 51 for YouTube in contrast to the 7 found in the Twitter corpus, is a major difference. Likely, there are more significant key phrases because YouTube comments are generally longer than tweets, producing a much richer dataset, even with fewer data points at the level of each individual comment. This suggests that commenters are more expressive on YouTube than on Twitter, a finding supported in work by Hajar et. al. that found YouTube comments to be similar to instant messages in terms of emotion classification (yasmina et. al., 2016).

Since there are such a large number of key phrases for YouTube, we won't discuss every single one here, so we focus on the key phrases that are notable due to their surface similarity to, or difference from, the significant Twitter phrases in an effort to answer our second RQ. First, "Bill Gates," rather than "Dr. Tam," is the most mentioned phrase, even though the comments were taken from a Canadian government press conference video. “Dr. Tam” is the third most mentioned phrase, despite being the main speaker in the video.

In the YouTube comments, “Bill Gates” is one of many phrases associated with conspiracy theories about coronavirus. Other phrases include: “chinese communist party,” “death certificates,” “wake up” or “waking up,” “see through,” “Hong Kong,” and “go out.” In these comments, the government is seen and portrayed as working against the will of the people or is traitorous, and some of its efforts are positioned as distraction for the “sheep.” A close read of the text surrounding these key phrases shows links between the Canadian government’s new gun control laws, and coronavirus-related stay-at-home recommendations, with many commenters suggesting that Canadians should defy both of those things. The context around the phrase “Immune system” is another flag for misinformation. It contains comments about false ways to prevent COVID-19, such as for example, that eating garlic can help prevent the virus, or that having a strong immune system is preferable to a vaccine for preventing COVID-19.

Other notable phrases are the ones that are overtly aggressive in nature. Aggressive language with the key statistically significant topics include “shut up”, “go to hell” and also calls for Dr. Tam to “go back” (usually followed by “to China”). Phrases like “shut up”, “step down” (as in a command), “go-to-hell,” “fight back,” and “wake up,” are part of a general tonality that upon a close read of contextual discourse reflects some alt-right hallmarks, such as the use of the term “cuck” to describe the Prime Minister, or the conflation of COVID-19 vaccinations with recent Canadian government policy limiting automatic weapons. Additionally, several key phrases reflect racist positions, such as, “Dr Tam is corrupt and chinese spy in Canada,” or “Dr Tam looks like she has the COVID19 ... Keep her away from the bat soup!”.

This racism is more overt than the racism we observed on Twitter, and whether it's due to the affordances of the medium (longer YouTube comments) or the public using the platform, represents a key difference between the two platforms.

Another key difference between YouTube and Twitter comments is tonal. Unlike in the Twitter dataset, the phrase "thank you" is not directed at the CPHO, but is typically used in a sarcastic way (e.g., "Thank you ! Oh yes , please enlighten us with your wisdom mr smarty pants") or used within to indicate agreement with conspiracies and misinformation (e.g., "Thank you canadians for waking up to our psycho leader and Public health officer who both should be locked up"). While "Thank you" statistically significant on both platforms, the context is very different on each when close reading is performed. For example, while the example phrase earlier in this paragraph "Thank you ! Oh yes , please enlighten us with your wisdom mr smarty pants") would be classified by algorithmic sentiment analysis as positive, a human qualitative coder can read it for they way it is intended: as sarcasm. So while "Thank you for your work Dr. Tam" is a phrase that occurred in some form on both platforms, the shorter nature of Tweets made"Thank you Dr. Tam" suggest they are genuine, whereas in the YouTube comments, this phrase was part of longer posts that were not expressions of actual appreciation (in a phrase such as "Thank you Dr. Tam.. for selling us out to a foreign government"). While at first glance, there appears to be similar phrases used on YouTube and Twitter, this analysis reveals the context to be very different.

Table 2. YouTube comments by key phrase

Item	**LRatio**	**Item (continued from previous column)**	**LRatio (continued from previous column)**
Bill_Gates	9.69	go_back	3.93
immune_system	9.42	shut_down	3.88
Dr_Tam	8.84	check_out	3.61
Tax_payer, Theresa_Tam	8.32	come_out	3.52
Chinese_communist_party, care_homes, Dr. Tam	8.1	coming_from	3.35
Flu_vaccine, Joe_Smith, Debra_Smith	7.52	thank_you	3.34
mental_health	7.32	due_to	2.65
Taking_away, life_guards, lande_Dukel, death_certificates, coronavirus_vaccine, Teresa_Tam, Stu_Harris, Nick_Smith, Justin_Trudeau	7.1	so_many	2.45

go_to_hell	6.52	look_at	2.24
death_rate	6.1	going_to	1.86
taken_down	5.84	have_to	1.63
public_health	5.35	a_lot	1.56
side_effects	5.32	Immune_systems, front_line, at_heart	5.52
go_out	5.05	stand_up	4.25
Hong_Kong	4.78	have_no_idea	3.52
wake_up	4.64	off_of	4.03
look_up	4.63	if_anything	4.52
In_charge, i_agree, all-over_the_word	4.52	Step_down, i_bet, health_and_safety, for_ever, fight_back	6.1
shut_up	4.35	is_over	3.67
let_know	4.15	your_own	3.1
long_term	4.1	Waking_up, see_through	4.19

back_up	4.07		

Discussion

Our mixed methods analysis of content directed at the CPHO on Twitter and YouTube combined algorithmic analysis with qualitative discourse analysis. Findings suggest that statistically significant phrases from YouTube comments have a different tone, and show evidence of harsh, racist, and conspiracy theory language, whereas statistically significant phrases in Twitter mentions contain more supportive language, and sometimes serve to counter misinformation. These findings reveal the importance of understanding how platforms foster different public engagement. They also suggest a need for both platform-diversity in misinformation research, and a need for more human coding of small data to augment big data approaches to misinformation research.

Researchers are making use of the opportunities afforded to them by large-scale social media data to examine online public opinion pertaining to COVID-19. Such studies may offer many important insights, such as for example revealing the kinds of misinformation claims that users make (Mai and Gruzd, 2020). Our findings, however, reveal that a more complex picture of public opinion arises if a cross-platform and multi-methods approaches are used. If researchers and policymakers are to fully understand public opinion in an environment characterized by misinformation, like the COVID-19 information environment, they must seek to gather and analyze user-generated content on multiple platforms. Earlier research has shown that computational social science studies may face a number of limitations and biases (e.g., boyd & Crawford, 2012; Tufecki, 2014). One significant limitation is researchers' tendency to focus such research on a single platform—most often Twitter—resulting in findings that are bound by the context and affordances of that platform. Nonetheless, when researchers focus on a single platform, as our research shows, the results they generate may not fully capture the full spectrum of public responses to political decisions. As a result, policymakers may not gain a robust account of what public opinion looks like, what it focuses upon, and how to counter the misinformation that is spreading through various online communities. Our research reveals important differences between platforms, and future research could help illuminate the reasons why, beyond platform affordances and demographics, different platform publics respond to health information in the ways they do..

In addition to adopting a cross- or multi-platform approach, researchers and policymakers need to consider studying misinformation *in context*. Studying misinformation in context means adopting both quantitative big data approaches to gathering and analyzing misinformation, as well as smaller data-set qualitative approaches, using newer and underused methods such as corpus assisted discourse analysis. While algorithmic topic and keyword analysis reveals some trends, certain nuances (such as for example sarcasm, racist undertones, and irony) are much more difficult to capture and categorize. The differences between, for example, a phrase coded algorithmically as “Thank you” in the YouTube data vs. the Twitter data are significant but require discourse analysis rather than algorithmic topic analysis to actually see. To understand the scope, tone, and nuance of policy debates on social media platforms, it is imperative to look beyond key words, word frequencies, and even key topics. While automated analysis helps researchers to consider a larger social media dataset, the hybrid method of corpus assisted discourse analysis researchers to be sensitive to nuance, such as the sarcasm we identified in our data (Lewis, Zamith and Hermida, 2013). Recent research has noted that while sarcasm detection has come a long way, there are still many challenges, particularly with respect to the use of emojis or emoticons in posts, and sarcasm in languages other than English (Elke et. al, 2020). One particular promise of corpus-assisted discourse analysis is that it helps to overcome these challenges, and thus may even be useful as a tool to help train deep learning text classification models in the future.

To summarize, our findings suggest a need to cast a wide net when studying public opinion on social media, particularly in the context of a misinformation heavy event like COVID-19. We must recognize the unique and context-specific nature of content, preferably across multiple platforms, by employing a rich methodological approach. By using both methods often associated with Big Data (i.e. the scraped posts and comments, analyzed through linguistic analysis software) *and* small data (i.e. the close reading and discourse analysis), this research highlights important differences between COVID-19 public opinion on Twitter and YouTube.

References

Authors. (2020).

Authors. (2021).

Altoaimy, L. (2018). Driving change on Twitter: A corpus-assisted discourse analysis of the Twitter debates on the Saudi ban on women driving. *Social Sciences*, *7*(5), 81.

Aluthman, E. S. (2018). A Corpus-assisted Critical Discourse Analysis of the Discursive Representation of Immigration in the EU Referendum Debate. *Arab World English Journal*, *9*(4).

Bechmann, A., Bruns, A., Gruzd, A., Quinn, A., and Rogers, R. (2018). Plenary Panel: Accessing social media data after Cambridge Analytica. *International Conference on Social Media and Society*. July 20, 2018: Copenhagen

Bennett, M. (2020). Should I do as I'm Told? Trust, experts, and COVID-19. *Kennedy Institute*

of Ethics Journal. *30* (3-4), 242-263.

boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*, *15*(5), 662-679.

Charmaz, K. (2014). *Constructing grounded theory*. New York: Sage.

Cody, E. M., Reagan, A. J., Dodds, P. S., & Danforth, C. M. (2016). Public opinion polling with Twitter. arXiv preprint arXiv:1608.02024.

Eke, C. I., Norman, A. A., Shuib, L., & Nweke, H. F. (2020). Sarcasm identification in textual data: systematic review, research challenges and open directions. Artificial Intelligence Review, 53(6), 4215-4258.

Gabrielatos, C. (2018). Keyness Analysis: Nature, Metrics and Techniques." In *Corpus Approaches To Discourse: A Critical Review*. Oxford: Routledge.

Giglietto, F., Rossi, L., & Bennato, D. (2012). The open laboratory: Limits and possibilities of using Facebook, Twitter, and YouTube as a research data source. *Journal of technology in human services, 30*(3-4), 145-159

Glowacki, E. M., Wilcox, G. B., & Glowacki, J. B. (2020). Identifying# addiction concerns on twitter during the COVID-19 pandemic: A text mining analysis. Substance abuse, 1-8.

Gove, W. R., & Geerken, M. R. (1977). Response bias in surveys of mental health: An empirical investigation. *American journal of Sociology*, *82*(6), 1289-1317.

Hallin, D.C., and Briggs, C.L. (2015). Transcending the medical/ media opposition in research on news coverage of health and medicine. *Media culture and society, 37*(1), 85-100.

Hunt, K. (2019). Twitter, social movements, and claiming allies in abortion debates. *Journal of Information Technology & Politics*, *16*(4), 394-410.

Hunt, K., & Gruszczynski, M. (2019). The influence of new and traditional media coverage on public attention to social movements: the case of the Dakota Access Pipeline protests. *Information, Communication & Society*, 1-17.

Jungherr, A. (2016). Twitter use in election campaigns: A systematic literature review. *Journal of information technology & politics*, *13*(1), 72-91.

Karami, A., Bennett, L. S., & He, X. (2018). Mining public opinion about economic issues: Twitter and the us presidential election. International Journal of Strategic Decision Sciences (IJSDS), 9(1), 18-28.

Lewis, S. C., Zamith, R., & Hermida, A. (2013). Content analysis in an era of big data: A hybrid approach to computational and manual methods. Journal of broadcasting & electronic media, 57(1), 34-52

Mai, P., and Gruzd, A. (2020). We can inoculate ourselves against COVID-19 related misinformation. *Policy Options*. April 14, 2020. Available at: https://policyoptions.irpp.org/magazines/april-2020/we-can-inoculate-ourselves-against-covid-19-misinformation/

McCloskey, B., & Heymann, D. L. (2020). SARS to novel coronavirus–old lessons and new lessons. *Epidemiology & Infection, 148*.

Mellon, J., & Prosser, C. (2017). Twitter and Facebook are not representative of the general population: Political attitudes and demographics of British social media users. *Research & Politics*, *4*(3), 2053168017720008.

Mislove, A., Lehmann, S., Ahn, Y. Y., Onnela, J. P., & Rosenquist, J. (2011, July). Understanding the demographics of Twitter users. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 5, No. 1).

Nahon, K., & Hemsley, J. (2013). *Going viral*. New York: Polity.

Rayson, P., Archer, D., Piao, S. and McEnery, T. (2020) "The UCREL Semantic Analysis System." Accessed May 13, 2020. https://eprints.lancs.ac.uk/id/eprint/1783/1/usas_lrec04ws.pdf.

Rayson, P. (2009) Wmatrix: a web-based corpus processing environment, Computing Department, Lancaster University. http://ucrel.lancs.ac.uk/wmatrix/

Ricke, L. (2010). A new opportunity for democratic engagement: The CNN-YouTube presidential candidate debates. Journal of Information Technology & Politics, 7(2-3), 202-215.

Rieder, Bernhard (2015). YouTube Data Tools (Version 1.11) [Software]. Available from https://tools.digitalmethods.net/netvizz/YouTube/

Rogstad, I. (2016). Is Twitter just rehashing? Intermedia agenda setting between Twitter and mainstream media. Journal of Information Technology & Politics, 13(2), 142-158.

Shin, J., Jian, L., Driscoll, K., & Bar, F. (2018). The diffusion of misinformation on social media: Temporal pattern, message, and source. *Computers in Human Behavior*, *83*, 278-287.

Song, M. Y. J., and Gruzd, A. (2017, July). Examining Sentiments and Popularity of Pro- and Anti-Vaccination Videos on YouTube. In *Proceedings of the 8th International Conference on Social Media and Society* (p. 17). ACM.

Su, Y., & Borah, P. (2019). Who is the agenda setter? Examining the intermedia agenda-setting effect between Twitter and newspapers. Journal of Information Technology & Politics, 16(3), 236-249.

Tommasel, A., Diaz-Pace, A., Rodriguez, J. M., & Godoy, D. (2021). Forecasting mental health and emotions based on social media expressions during the COVID-19 pandemic. Information Discovery and Delivery. [DOI 10.1108/IDD-01-2021-0003]

Tufekci, Z. (2014, May). Big questions for social media big data: Representativeness, validity and other methodological pitfalls. In *Eighth International AAAI Conference on Weblogs and Social Media*.

Van Kessel, P. (2019). 10 facts about Americans and YouTube. Pew Research Centre. Retrieved Jan 26, 2021 from: https://www.pewresearch.org/fact-tank/2019/12/04/10-facts-about-americans-and-youtube/

Vraga, E. K., & Bode, L. (2018). I do not believe you: How providing a source corrects health misperceptions across social media platforms. Information, *Communication & Society*, *21*(10), 1337-1353.

yasmina, D, Hajar, M., and Hassan, M. (2016). Using YouTube comments for text-based emotion recognition. Procedia Computer Science, 83, 292-299.

Data Availability

This paper uses social media data that was publicly available at the time of data collection. To honor Canadian ethics guidelines and user intent as per Twitter guidelines, the Twitter dataset is not made publicly available. To honor Canadian ethics guidelines, the YouTube dataset is not made publicly available, though the publicly available comments are available at: https://www.YouTube.com/watch?v=BwvY76-6gbQ