



Studying Toxicity on Twitter During the 2019 Federal Election in Canada

Anatoliy Gruzd, Philip Mai, Raquel Recuero and
Felipe Bonow Soares

EasyChair preprints are intended for rapid
dissemination of research results and are
integrated with the rest of EasyChair.

May 22, 2022

Title: Studying Toxicity on Twitter during the 2019 Federal Election in Canada

ABSTRACT

Objective

The goal of this study was to better understand the online dynamics of violence on Twitter against candidates running for political offices. Violence on online platforms is a pressing problem. It can be broadly characterized as: threats and expressions of intent to perpetrate violence or abuse online, trolling, cyberbullying, hate speech directed at members of an identifiable group, actions aimed at damaging reputations, or non-consensual distribution and doxing of intimate images (also known as revenge-porn). For the purpose of this exploratory research, we focused on two common characteristics of online violence: the presence of *toxic* and *insulting* messages. A message is toxic when it is rude, disrespectful, or unreasonable. A message is insulting when it is inflammatory or negative toward a particular person or a group of people. Our research was guided by the following research questions:

RQ1: What is the prevalence of toxic and/or insulting messages targeting political candidates?

RQ2: Is there a difference in frequency of toxic and/or insulting messages directed at women versus men candidates on Twitter?

Methods

We started by compiling a list of 2,144 candidates running for office during the 2019 federal election in Canada and recorded their Twitter handles, when available. In total, 1,344 candidates had a public profile on Twitter at the time of data collection during the fall of 2019.

We then used an open source program called Social Feed Manager to collect public tweets directed at each of the 1,344 candidates, from September 29 to October 28, 2019. We focused on 363,706 public tweets posted in English and directed at 1,116 candidates¹.

Once data was collected, three independent coders (one undergraduate racialized, female student, one graduate female student, and one graduate male student) hand coded a random sample of 3,637 tweets (1% sample) to categorize them as toxic or insulting (as defined above). To ensure the trustworthiness of the manual coding procedure, only tweets that were flagged by all three coders were considered as either toxic or insulting for the subsequent analysis.

Finally, we used a chi-square test to determine if there is an association between a candidate's gender and the likelihood of them receiving either a toxic or insulting tweet.

Results

RQ1: What is the prevalence of toxic and/or insulting messages targeting political candidates?

In total, there were 307 (8.4%) toxic and 101 (2.8%) insulting tweets. While most toxic tweets expressed partisan opinions using strong language, it is the insulting tweets directed at candidates that had an especially negative undertone directed at a specific individual(s); tweets such as "*@[username] Was praying the little f@cker would tip over lol*" and "*@[username] American citizen can not be a Canadian PM Resign you MF*".

This result is in line with related studies that examined toxic-type messages targeting the members of parliament in the UK and showed that the overall percentage of toxic messages on Twitter was under

¹ 1,344 minus 228 candidates who have not received any tweets directed at them during the studied time period.

10%. For example, one study found that 9.8% of tweets targeting British MPs were uncivil (Southern and Harmer, 2019), while another found that less than 4% of tweets directed at British MPs were abusive (Gorrel et al., 2019). Similarly, other studies found swearing, dismissive insults, and abusive words to make up around 3% of online communications more broadly (Mead, 2014; Subrahmanyam et al., 2006).

Even though the overall percentage of toxic and insulting tweets was relatively low (under 10%), it is not necessarily their quantity, but also their severity which may negatively impact one's well-being, potentially leading to reduced online participation and withdrawal. This is particularly important considering the growing expectation for politicians to maintain an active online presence and engage the public on sites like Twitter, even if faced with toxic and insulting messages daily.

RQ2: Is there a difference in frequency of toxic messages directed at women versus men candidates on Twitter?

The chi-square test confirmed that there was no significant association between gender and receiving a toxic or insulting tweet. The test was performed on the sample of 3,635 tweets (3637-2²), containing 868 tweets targeting women candidates and 2,767 targeting men candidates.

On the one hand, this is in line with a related study by Gorrel et al. (2019) who found that the "burstiness" of abuse on Twitter does not depend on gender in the context of Twitter use by the UK's MPs. On the other hand, in a different study, Southern & Harmer (2019) found that women MPs in the UK were more likely to receive uncivil, stereotypical or tweets questioning their position as politicians than their male counterparts. Yet, in an earlier study about the UK's MPs, Ward & Mcloughlin (2017) found that men were more likely to receive "abusive" messages on Twitter than women MPs. These somewhat contradictory results are likely due to the differences in the study context (different countries), study population (candidates vs members of parliament), different time frames, and different approaches to defining toxic, uncivil and abusive messages.

Even as more and more academic studies are starting to examine this area of research, there is still a need for standardization of definitions and approaches. For instance, because different researchers measure different things, it hampers our ability to conduct comparative studies and develop a theoretical foundation.

Discussion and Future Work

Some candidates tend to receive more extreme cases of online violence, such as the case of a hate campaign targeting a former environment minister and 2019 candidate Catherine McKenna, who received numerous online (and offline) threats (Mia Rabson, 2019); however, that variation is not necessarily explained by the candidate's gender. This study found that women candidates did not necessarily receive more tweets that are toxic or insulting. Since we only tested a sample of tweets (1%), future work will examine the full dataset to see whether our findings apply to the broad set of tweets that were aimed at candidates in the 2019 Federal Election.

And while platforms like Twitter are starting to offer new features to report, block or hide hostile individuals or tweets, it may be viewed as undemocratic by the public if politicians or elected officials rely on such features to block their critics. In this context, social media platforms ought to take a more proactive role in preventing online harassment campaigns against their users. A case in point is the fact that out of 307 (8.4%) toxic and 101 (2.8%) insulting tweets flagged by our coders, the majority of these posts (255 toxic and 85 insulting tweets) are still publicly available as of January 2, 2020³. Our future work will examine the trade-offs between individual's vs social media platforms' responsibilities to deal

² Prior to running the test, we excluded two candidates who self-identified as non-binary due to the insufficient number of observations for statistical testing (a qualitative approach will be used instead to examine these two cases in our future work).

³ 24 toxic and 8 insulting tweets are no longer available because Twitter suspended the original posters; and 25 toxic and 8 insulting tweets have been deleted by the original posters.

with online violence, because if left unchecked, the proliferation of toxicity and insults would undermine the capacity of platforms like Twitter to support deliberative and networked democracy (Loader & Mercea, 2011).

Going forward, we plan to apply a machine learning approach to analyze the full dataset of tweets in a systematic and automated way. We will also apply topic modelling techniques to examine differences and similarities between the types of toxic and insulting messages directed at women versus men candidates on Twitter. Finally, we will expand our work to other popular social media platforms like Facebook and reddit.

References

- Costello, M., Hawdon, J., & Ratliff, T. N. (2017). Confronting Online Extremism: The Effect of Self-Help, Collective Efficacy, and Guardianship on Being a Target for Hate Speech. *Social Science Computer Review*, 35(5), 587–605. <https://doi.org/10.1177/0894439316666272>
- Gorrell, G., Bakir, M. E., Greenwood, M. A., Roberts, I., & Bontcheva, K. (2019). Race and Religion in Online Abuse towards UK Politicians: Working Paper. *ArXiv:1910.00920 [Cs]*. <http://arxiv.org/abs/1910.00920>
- Kowalski, R. M., Giumetti, G. W., Schroeder, A. N., & Lattanner, M. R. (2014). Bullying in the digital age: A critical review and meta-analysis of cyberbullying research among youth. *Psychological Bulletin*, 140(4), 1073–1137. <https://doi.org/10.1037/a0035618>
- Loader, B. D., & Mercea, D. (2011). NETWORKING DEMOCRACY? *Information, Communication & Society*, 14(6), 757–769. <https://doi.org/10.1080/1369118X.2011.592648>
- Lohmann, R. C. (2014, January 28). *Trolling or Cyberbullying? Or Both?* Psychology Today. <http://www.psychologytoday.com/blog/teen-angst/201401/trolling-or-cyberbullying-or-both>
- Mead, D. (2014, February 19). *People Sure Tweet “Fuck” a Lot, Finds Science*. Motherboard. https://motherboard.vice.com/en_us/article/8qxn8a/people-sure-tweet-fuck-a-lot-says-science
- Mia Rabson. (2019, September 7). Environment Minister Now Needs Security As Threats Against Her Spike. *HuffPost Canada*. https://www.huffingtonpost.ca/entry/catherine-mckenna-security-threats_ca_5d73d831e4b06451356ee2fe
- Phillips, W. (2015). *This Is Why We Can't Have Nice Things: Mapping the Relationship between Online Trolling and Mainstream Culture*. MIT Press. <http://www.jstor.org/stable/j.ctt17kk8k7>
- Rosalynd Southern, & Emily Harmer. (2019). Twitter, Incivility and “Everyday” Gendered Othering: An Analysis of Tweets Sent to UK Members of Parliament. *Social Science Computer Review*, 1–17. <https://doi.org/10.1177/0894439319865519>
- Subrahmanyam, K., Smahel, D., & Greenfield, P. (2006). Connecting developmental constructions to the internet: Identity presentation and sexual exploration in online teen chat rooms. *Developmental Psychology*, 42(3), 395–406. <http://dx.doi.org.ezproxy.lib.ryerson.ca/10.1037/0012-1649.42.3.395>
- Ward, S. J., & Mcloughlin, L. (2017, April 26). Turds, traitors and tossers: The abuse of UK MPs via Twitter. *ECPR Joint Sessions 2017*. ECPR Joint Sessions of Workshops, University of Nottingham. <https://ecpr.eu/Events/PaperDetails.aspx?PaperID=34203&EventID=104>