Australian and New Zealand Journal of European Studies
https://cesaa.org.au/anzjes/
Vol11 (3)
ISSN 1837-2147 (Print)
ISSN 1836-1803 (On-line)

Sharp power in social media: Patterns from datasets across electoral campaigns

Simo Hanouna
Ben-Gurion University of the Negev
hanouns@post.bgu.ac.il

Omer Neu
Ben-Gurion University of the Negev
neu@post.bgu.ac.il

Sharon Pardo
Ben-Gurion University of the Negev
pardos@bgu.ac.il

Oren Tsur
Ben-Gurion University of the Negev
orentsur@bgu.ac.il

Hila Zahavi
Ben-Gurion University of the Negev
hilape@bgu.ac.il[1]

Abstract
Using Christopher Walker's and Jessica Ludwig's 'sharp power' theoretical framework, and based on some preliminary findings from the May 2019 European Parliament election and the two 2019 rounds of elections in Israel, this article describes a novel method for the automatic detection of political trolls and bots active in Twitter in the October 2019 federal election in Canada. The research identified thousands of accounts invested in Canadian politics that presented a unique activity pattern, significantly different from accounts in a control group. The large-scale cross-cross-sectional approach enabled a distinctive perspective on foreign political meddling in Twitter during the recent federal election campaign. This foreign political meddling, we argue, aims at manipulating and poisoning the democratic process and can challenge democracies and their values, as well as their societal resilience.

Key words: Canada, sharp power, machine learning, Twitter, social networks, political trolls, bots, foreign political meddling

[1] The three principal investigators, marked with °, appear in alphabetical order and contributed equally to this article.

Introduction

Protecting democracies from fake news, propaganda, manipulation, confusion, division and foreign electoral interference is becoming a major challenge for Western liberal societies. Tackling these threats, both Canada and the European Union (EU) have put new laws and codes of practice in place.

In the EU, in view of the May 2019 election to the European Parliament (EP), the European Commission urged the leading social networks, as well as advertisers, to step up their efforts in the joint fight against disinformation and to adopt a self-regulatory Code of Practice on Disinformation (European Commission, 2019, p. 4). In the October 2018 Code of Practice, the signatories recognize and agree with the EU that "the exposure of citizens to large scale disinformation, including misleading or outright false information, is a major challenge for Europe" (European Commission, 2018, p.1). The signatories committed themselves to improve the transparency, accountability and trustworthiness of their online services (European Commission, 2018, pp. 3–4). Based on their commitment, from January to May 2019, just before the European elections, the European Commission together with the European Regulators Group for Audiovisual Media Services closely monitored Facebook, Google and Twitter on a monthly basis (European Commission, 2019, p. 4).

Following the cyber threat activity against the democratic processes in the EU and in the United States (US), in October 2017, the Communications Security Establishment (CSE) – Canada's centre of excellence for cyber operations – warned the Canadian parliament of major cyber threats to Canada's democratic process at the federal, provincial/territorial, and municipal levels of government. According to a CSE June 2017 report, "almost certainly, multiple hacktivist groups will deploy cyber capabilities in an attempt to influence the democratic process in 2019". CSE further warned that "nation-states are constantly deploying cyber capabilities" against Canada (Communications Security Establishment, 2017, p. 33).

Following the 2017 report and despite the fact that it is often almost impossible to identify the attackers, the recent federal election campaign was fought under a new amendment to the Canada Elections Act. Section 91 of the Act, which came into effect only 41 days before the October 2019 federal election, prohibits the publication of a false statement to affect election results during the election period.[2]

Both the EU and Canada have thus identified cyber capabilities, and mainly those that are deployed by nation-states, as a major threat to their societies and to their democracies. Cyber capabilities are now used "not to destroy an adversary, but rather to frustrate it, slow it, undermine its democratic institutions, and leave its citizens angry and confused" (Sanger, 2018, p. XVII). And these capabilities, David Sanger argues, "are almost always employed just below the threshold that would lead to retaliation" (Sanger, 2018, p. XVII).

In the following pages we focus on foreign deployment of cyber capabilities in the October 2019 federal election campaign in Canada. We do so by analyzing four Twitter datasets collected in different time spans, based on different filters, and covering the May 2019 European Parliament election and the two 2019 rounds of elections in Israel.

[2] S.C. 2000, c. 9. The full text is available at https://laws-lois.justice.gc.ca/eng/acts/E-2.01/FullText.html [accessed: 16 November 2019].

In our cross campaign datasets, we identify thousands of accounts invested in Canadian politics that present a unique activity pattern, significantly different from accounts in a control group. Drawing on Christopher Walker and Jessica Ludwig's 'sharp power' theoretical framework (National Endowment for Democracy, 2017), we argue that the pattern that emerges suggests a case of foreign political meddling in the recent federal election campaign. This meddling, we further argue, aims at manipulating and poisoning (Walker and Ludwig, 2017a) the democratic processes in both Canada and the EU.

From soft power to sharp power

Joseph Nye first developed the concept of 'soft power' in his 1990 seminal work *Bound to Lead: The Changing Nature of American Power*, in which he argued that the US is still the dominant world power, not only in military and economic power, but also in what he called 'soft power' (Nye, 1990, p. 31; Nye, 2001). For Nye, this aspect of power is:

> the ability to get what you want through attraction rather than coercion or payments. It arises from the attractiveness of a country's culture, political ideals, and policies. When our policies are seen as legitimate in the eyes of others, our soft power is enhanced (Nye, 2004, p. x).

Over the years the soft power concept was well received by the research community and the global mass media and was widely used by political leaders all over the world. However, as Nye himself argued, this "wide usage has sometimes meant misuse of the concept as a synonym for anything other than military force" (Nye, 2011, p. 81). The soft power concept was embraced also by non-ethical authoritarian governments that are "investing heavily in their own instruments of 'soft power' in order to compete with democracy in the realm of ideas" (Walker, 2016, p. 50). These illiberal regimes hijacked "the concept of soft power as part of a broad assault on democracy and its values" (Walker, 2016, p. 51). Soft power, however, is "a rather uncomfortable fit for these efforts" (Walker, 2016, p. 50), as the techniques used by authoritarian regimes were not really soft. International Relations (IR) were thereby in a need of a new term to describe this new concept of power.

The term 'sharp power' is relatively new and was coined in 2017 by Walker and Ludwig in a report by the National Endowment for Democracy (2017). Walker and Ludwig hold that authoritarian regimes' influence efforts cannot be seen through Nye's theoretical framework of soft power. Importantly, "some of the most visible authoritarian influence techniques used by countries such as China and Russia, while not 'hard' in the openly coercive sense, are not really 'soft' either" (National Endowment for Democracy, 2017, p. 6). Walker and Ludwig argue that Beijing's and Moscow's initiatives in the spheres of academia, culture, media, publishing and think-tanks – "sectors that are crucial in determining how citizens of democracies understand the world around them" (Walker, 2018, p. 12) – should not be seen as soft power efforts as they are neither "charm offensive" nor an effort to "win the hearts and minds" (National Endowment for Democracy, 2017, p. 6). For Walker and Ludwig "this authoritarian influence is not principally about attraction or even persuasion; instead, it centers on distraction and manipulation" (National Endowment for Democracy, 2017, p. 6; Walker and Ludwig, 2017a, p. 10).

When using its sharp power capabilities, the authoritarian "pierces, penetrates, or perforates the political and information environments in the targeted countries. […] the repressive regimes' 'sharp power' techniques should be seen as the tip of their dagger – or indeed as their syringe" (National Endowment for Democracy, 2017, p. 6). The authoritarian regimes have come to confuse, divide, repress and poison the information that reaches their target audiences (Walker and Ludwig, 2017a, pp. 12–13), with the aim "of polluting [their] audiences' understanding of the world" (Orttung and Walker, 2014). Sharp power enables authoritarian regimes "to cut, razor-like, into the fabric of a society, stoking and amplifying existing divisions" (Walker and Ludwig, 2017a, p. 13). Taking advantage of the democratic societies, "the authoritarians' 'sharp power' efforts are typically difficult to detect, meaning they benefit from a lag time before the targeted democracies realize there is a problem" (Walker and Ludwig, 2017a, p. 13). Walker and Ludwig conclude that "above all, the term 'sharp power' captures the malign and aggressive nature of the authoritarian projects" (Walker and Ludwig, 2017a, p. 13). By using sharp power techniques, "the generally unattractive values" (Walker and Ludwig, 2017b, p. 3) of countries such as China and Russia "are projected outward, and those affected are not so much audiences as victims" (Walker and Ludwig, 2017b, p. 3).

Walker returned to his sharp power concept in mid-2018 and clarified that through manipulation, sharp power may be used to degrade the integrity of independent institutions. By employing the arts of destruction, sharp power exploits the open electoral and free media sectors, and by working via modern forms of censorship, sharp power sharpens tensions and rifts within and between democracies (Walker, 2018, pp. 12–13).

Comparing between soft and sharp power, Walker argues that "sharp power takes advantage of the symmetry between free and unfree systems. […] It is within this context that sharp power, neither really soft nor hard, is able to flourish" (Walker, 2018, p. 17). For him, as well as for Nye, authoritarians simply cannot "do" soft power well (Walker, 2018, p. 18; Nye, 2013), simply because, unlike hard power, soft power does not really belong to governments (Nye, 2001). Hence, Walker concludes by advising us "to avoid conceiving of sharp power as soft power's polar opposite. It is *not* the case that countries can wield either 'sharp' or 'soft' power but not both" (Walker, 2018, p. 18).

In order to understand how foreign entities are confusing, dividing, repressing and poisoning information, we now turn to an analysis of foreign manipulation of political trolls and bots active in Twitter in the October 2019 federal election in Canada. Using the sharp power theoretical framework and based on a trans-continental approach, with some preliminary findings from the May 2019 European Parliament election and the two 2019 rounds of elections in Israel, we seek to understand how foreign entities are trying to influence the Western liberal democracies.

Sharp power and political campaigns in social media

Social network platforms, especially Facebook and Twitter have become the *de-facto* arena for public debate, political campaigns, agenda setting efforts, the diffusion of real and fake news, as well as out-right propaganda (Jamieson, 2018; Grinberg et al., 2019; Lazer et al., 2018).

Both legitimate and illegitimate campaigns are inevitable in the era of social networks. However, it is clear that some deliberate efforts are subversive, employing questionable

tactics, at best. The debate about the legitimacy and inevitability of dynamics of political campaigns in social media is reflected in the very different policy decisions, recently announced by the major social platforms – the total ban on political advertising (Twitter),[3] the complete exemption of political ads from fact checking (Facebook),[4] and some restrictions on micro-targeting (Google).[5] While these policies address official ads and the promotion of paid content, official accounts are only a small part of the social network eco-system, as most content is posted and shared by random users. It is known that some of these users are automated bots and paid trolls – working in coordination to disrupt the political discourse in social networks.[6] Twitter reported removing 3,814 accounts that were linked to the Russian Internet Research Agency (IRA) – an organization with ties to the Russian intelligence[7]– as well as other accounts originating in various countries, though not necessarily linked to official branches[8].

Vast efforts are dedicated to the detection of political trolls and bots, however, success is moderate and depends on the platform and the specific campaign, often requiring access to classified intelligence that is not readily available for researchers. Moreover, the service terms enforced by social platforms in the End User License Agreement (EULA) harshly restrict the amount of data that can be collected, severing the coverage of the data to be analyzed.

In the following section we describe a novel method for the automatic detection of political trolls and bots (*trolls*, for brevity) active in Twitter in the October 2019 federal election in Canada. Specifically, we show how a careful contrastive analysis of different datasets facilitates an efficient detection, even with limited resources.

Methodology and data

In this section we briefly review the computational tools we use and the data collection process.

Descriptive statistics and significance testing

Different variables could be used to characterize users in social media, for example, the number of posts a user posts a day, the number of followers and friends she has, the number of languages she uses, etc. Once we define groups of users, hypothesizing they are drawn from distinct populations, we can test the significance of the difference between the distribution of each of the relevant variables. Standard significant testing is used and the p-value is reported.

Data representation

[3] https://business.twitter.com/en/help/ads-policies/introduction-to-twitter-ads/twitter-ads-policies.html [accessed: 10 December 2019].
[4] https://about.fb.com/news/2019/09/elections-and-political-speech/[accessed: 10 December 2019].
[5] https://blog.google/technology/ads/update-our-political-ads-policy [accessed: 10 December 2019].
[6] Bots are accounts that automate content promotion. Some bots are harmless or even useful, while others might promote negative, inflammatory and low quality content (Stella et al., 2018; Shao et al., 2018). Unlike bots, troll-accounts are operated by human users who misrepresent their identities with the intention of promoting discord (Phillips, 2015; Broniatowski et al., 2018, p. 1378).
[7] https://blog.twitter.com/en_us/topics/company/2018/2016-election-update.html [accessed: 10 December 2019].
[8] https://blog.twitter.com/en_us/topics/company/2019/information-ops-on-twitter.html [accessed: 10 December 2019].

Machine learning techniques require the representation of data points as vectors in a vector-space defined by the relevant features. In this study we use three types of representation:

1. *Meta features*: Each user is represented by twenty-eight meta-features, e.g., number of friends and followers, publish rate, lingual context switches, retweet ratio, etc. Meta features are proven to be very useful for bot detection (Yang et al., 2019).
2. *Bag of Words* (BOW): The feature vector space could correspond to a predefined list of words (e.g. all words in the language, the thousand most frequent words, etc.). Each textual data point is represented by the (often weighted) occurrence of vocabulary words in the text (Jurafsky and Martin, 2019).
3. *Topic Vectors*: Topic Models can be used as a tool for dimensionality reduction and for the discovery of latent topics in large textual corpora. A topic is a distribution over a predefined vocabulary. Latent Dirichlet Allocation (Blei, 2003) and other methods are applied over the data in order to infer the word distribution in a number (k) of speculated topics. This is done in an unsupervised way and k – the optimal number of topics is decided heuristically. One of the benefits of the topic models framework is the potential interpretability of some of the topics in an ad-hoc manner.
4. *Word Embeddings* (word2vec): Word2vec is the name commonly used to refer to an array of methods for learning word representation and distributional semantics. We learn the embeddings over corpora in order to identify the words most associated with the different parties and groups of users.

Algorithmic tools

1. K-means: K-means is one of the common methods to cluster the data to k clusters (Lloyd, 1982). The k-means clustering algorithm takes the vector representations of the data point as inputs and finds the best partition of the data in the given vector space. We experimented with all the representations surveyed above.
2. Word embeddings: Embeddings were learnt using FastText module (Bojanowski et al., 2017).
3. T-SNEprojections: t-Distributed Stochastic Neighbor embedding is used for dimensionality reduction, suitable for clear visualization of high dimensional data (van der Maaten and Hinton, 2008).

Data collection

Twitter provides a number of ways to collect data. Broadly speaking, data could be collected using the streaming Application Program Interface (API)[9] or the search API.[10] The Streaming API allows real time capturing of up to 1% of the public tweets. The Search API allows querying historical data. Both API allow applying some criteria and filters on the data collected. For example, one can restrict the Streaming API to collect only data containing specific words. The number of filters is limited to 400. Similarly, the Search API limits the query rate. These restrictions force the researcher to narrow her data collection efforts, limiting both the amount of data and the contextual coverage.

[9] https://developer.twitter.com/en/docs/tutorials/consuming-streaming-data [accessed: 10 December 2019].
[10] https://developer.twitter.com/en/docs/tweets/search/overview [accessed: 10 December 2019].

In order to overcome these challenges, we have collected and distilled a number of complementary datasets, thereby allowing the detection of troll accounts and the study of political troll behavior. The datasets are described in the next subsection.

Dataset

1. Canada 2019 Terms (CT): A list of 104 terms, salient in the Canadian political discourse, were compiled by domain experts (e.g. *Canada, four more years, Trudeau, water crisis choose to include, progressive trade, here for YEG*). All tweets containing these terms were collected in real time, starting on September 4, 2019, using the Streaming API. In total, CT contains 150 million tweets, posted by 21 million unique users. While these data provide a full coverage of the use of the predefined terms, it may be noisy (e.g. concerns about *water crisis* are not unique to Canada), and some other, mostly emerging topics are absent from the predefined list (e.g. *blackface*), and may not be captured.
2. Canada 2019 Political Handles (CH): A list of 52 user handles of prominent political figures and institutions was compiled (e.g., *@JustinTrudeau, @CanadianPM, @VoteChris2019, @BlocQuebecois, @CanadianPolling*). Again, using the Streaming API, starting on September 4, 2019, we have captured all tweets by the specified accounts and all tweets replying, retweeting or mentioning them – a total of 1.7 million tweets by 258 thousand unique users.
3. EU Parliament Tweets (EP): A dataset of 646 million tweets, by 31 million unique users was collected from April 4, 2019 until August 4, 2019, in a similar way to CT, based on a list of terms related to EU politics and policies and the May 2019 European election, compiled by domain experts.
4. Hebrew Tweets (HT): Political developments in Israel resulted in two rounds of general elections in April and September 2019 (and a third round in March 2020). We started collecting Hebrew Tweets in November 2018, when the election was called. A dataset of 30 million Hebrew tweets, posted by 690,000 unique users between November 2018 and October 2019. This dataset is unique as it is not based on predefined lists of terms and users and constitutes an almost full coverage of public Hebrew tweets.[11]
5. Canada 2019 Flagged (8KO): Previous research suggests that exclusive retweeting of political accounts correlates with troll behavior. Out of the 258,000 accounts captured in the CH dataset, 100,000 were only retweeting political accounts, rather than replying to, mentioning and quoting them. We identified 8,602 accounts out of these 100,000 accounts that were active in HT as well. The odd behavior of exclusive retweeting (in the CH dataset), along with their unexpected presence in the HT dataset (different language, continent, time-span and campaign) promotes the flagging of these 8,602 as suspected trolls (hence, 8KO – O for odd). We used the Search API to download the most recent 3,200 tweets of each of these accounts.[12] This process resulted in a dataset containing about 25 million tweets. While restricted to these users, these data are unfiltered by topic, term or time span (depending on the temporal distribution of tweeting by each individual user).
6. Canada 2019 Control (8KR): In order to study the characteristics of the accounts in 8KO, we randomly sampled the same number of accounts from the same

[11] This was achieved due to the relatively small demographics of Hebrew speakers and a careful choice of stopwords as steaming filters.
[12] Twitter restricts the history that can be queried, as well as enforcing a strict rate limit on issuing Search queries.

100,000 retweeting accounts identified above. A dataset of about 25 million tweets was collected in a similar way to 8KO. These data serve as a control set (hence, 8KR – R for random).

Combining the top four datasets (CT, CH, EP, HT), collected in different time spans, based on different filters and covering different political campaigns in different states and different continents, yields a unique multi-faceted, cross-sectional political dataset. The cross campaign datasets (8KO, 8KR) distilled from the intersections between subsets, enables a distinctive perspective on foreign political meddling in social networks.

Analysis

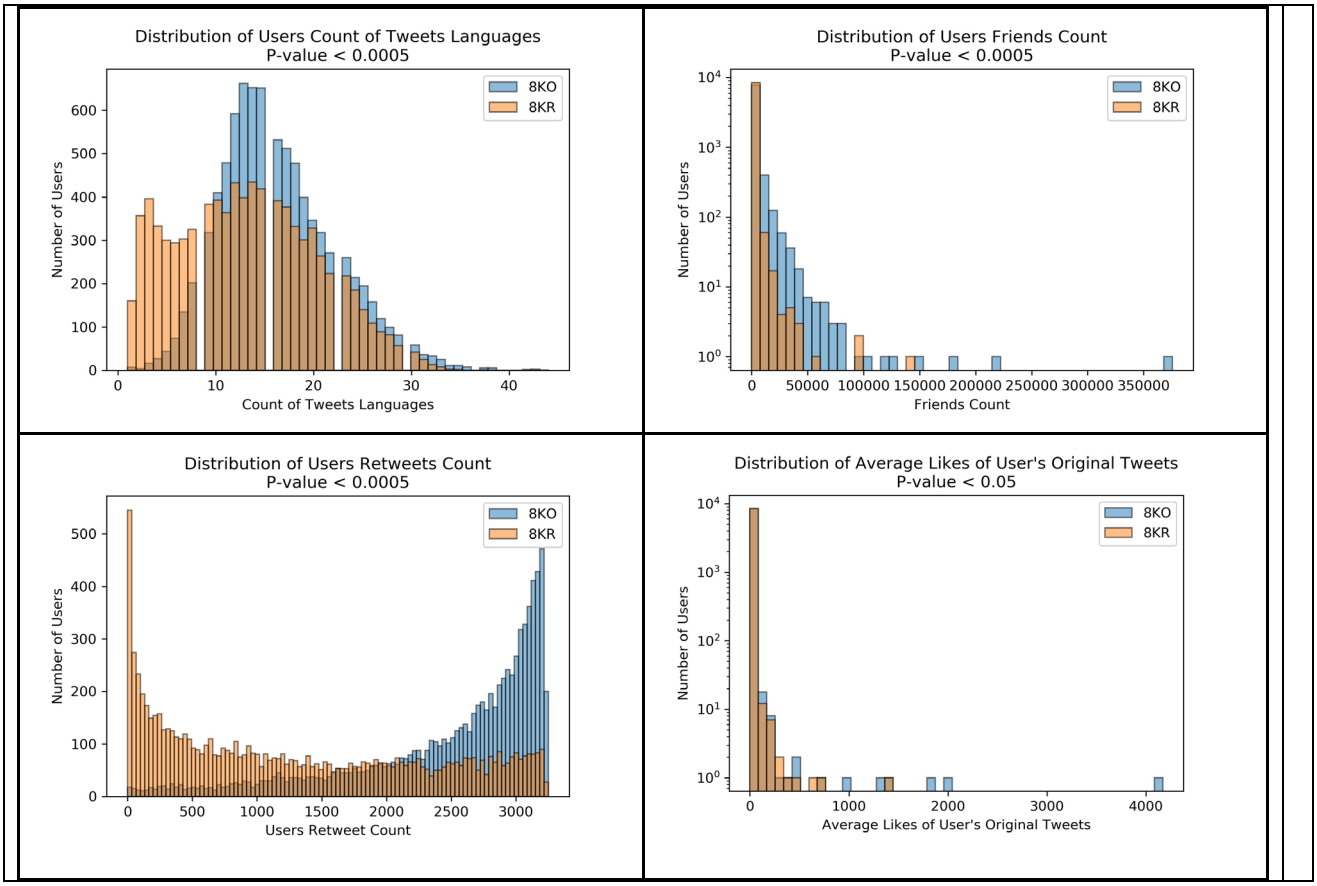
Twitter users can be characterized and modeled in different ways, ranging from the structural properties of the network, to user meta-data and content analysis. In this work we provide an initial analysis based on the meta-data relevant to the EP, Canada and Israel, showing a significant difference between clusters of users. We further expand this cluster analysis by coupling it with a content-based analysis, using Topic Models and User Embeddings.

Account meta-data includes 28 features such as the number of friends and followers, the total number of tweets, the age of the account, the number of tweets in the data, the number of languages the account uses, the tweet/retweet ratio, etc. We hypothesize that political trolls will exhibit different behavior than authentic accounts and that this behavior is reflected in the meta-data. For example, we expect trolls to retweet more, and post less original content compared to authentic accounts. We also expect them to post in more languages and have different social behavior, reflected in the number of friends and followers they have.

Indeed, we find statistically significant differences in the distribution of the values of the meta-features of 8KO/R users. Figure 1 presents the distribution of the number of languages used by an account, the number of followers, the number of retweets, and average number of likes (stars/favorite-votes). We find, for example, that flagged users (8KO) tend to have a significantly larger number of friends (users they 'follow'), while most authentic users have a manageable number of friends.

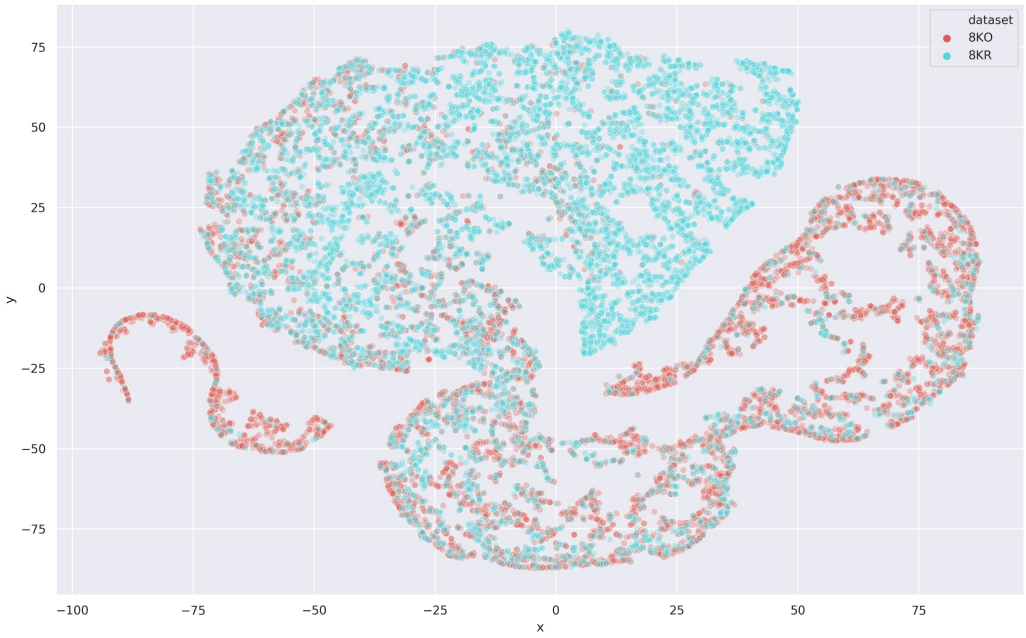
The 8KO users appear to be incredible polyglots, which correlates with their 'obsessive' retweeting activity. Examining the language distribution, while English and French are ranked 1 and 2 (respectively) in both 8KO and 8KR, the frequency of Russian is ranked #14 in 8KO while it is ranked #26 in the 8KR control set. Arabic is ranked #4 in the 8KO and 7th in 8KR, a difference of a whole order of magnitude in the raw number of Arabic tweets. Chinese, on the other hand, is ranked #22 in both 8KO/R, with roughly the same number of tweets.

Figure 1: Distributions of four meta-features in 8KO and 8KR datasets



The 2-dimensional T-SNE projection (van der Maaten & Hinton, 2008) of the users by their meta-data (Figure 2) presents a relatively clean separation between the datasets. We note, however, that some 8KO users are mixed in the 8KR dominated cluster. Indeed, a manual inspection revealed some authentic users, e.g. a liberal Canadian Rabbi and university professor, legitimately engaged in both Israeli and Canadian political and cultural exchanges. On the other hand, we also notice a non-negligible minority of the 8KR users appearing in the 8KO-dominance cluster. This suggests that while the troll detection precision we achieve using our cross dataset matching is high, we can still improve the recall, potentially by crossing with more datasets beyond HT.

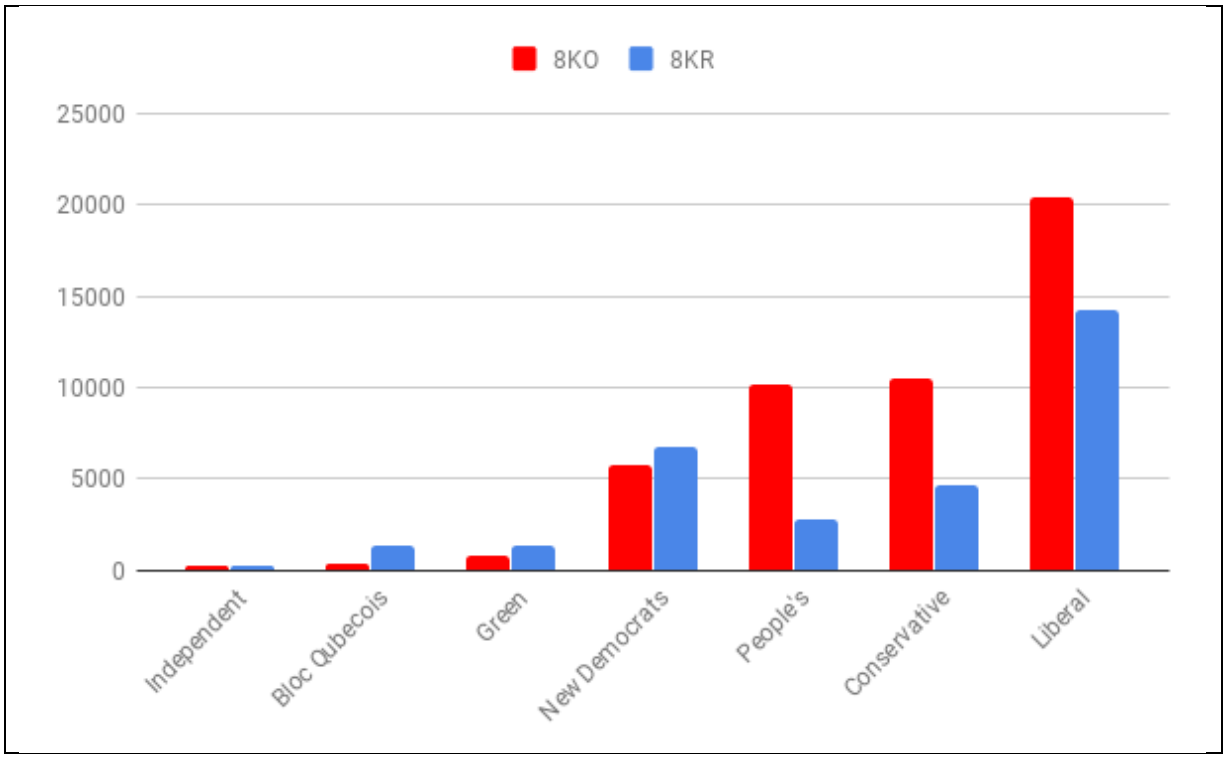
Figure 2: T-SNE projection of the flagged users and the control users based on their meta-data.



The distribution of meta-data values and the meta-data based projection indicates that the two populations – the flagged users (8KO) and the random users (8KR) – are indeed very different. The flagged users exhibit automated behavior: they tweet, mostly retweet, at a much higher frequency, are involved in multiple campaigns and tend to promote content in multiple languages. 2,605 of these accounts have more than five tweets in at least three of the four initial datasets (CT, CH, EP & HT), compared to only 890 of the 8KR users.

However, while we have managed to identify trolls, we are yet to understand the way they meddled in the Canadian elections and what type of information was promoted in support of which party. Figure 3 presents the number of retweets each party (official accounts of the party and its candidates as defined in the CH list) received from the 8KO/R users. It is notable that the 8KO users promote content from both sides of the aisle.

Figure 3: Retweets of party affiliated accounts by the 8KO and 8KR users



In order to evaluate the content promoted by different groups of users we used Gensim's FastText module (Bojanowski et al., 2017) to train an embedding model over the CT dataset. The embedding vectors are used to identify the concepts that are semantically related to each official political account. We aggregated these concepts on the party level,[13] effectively capturing the agenda and spin promoted by each party.[14]

We found that 8KO users use almost twice the number of unique partisan terms than those used by the 8KR users (Figure 4, top). It seems that this is not a result of the much higher frequency of tweeting but a result of a difference in the modus operandi of the groups. A statistical analysis and verification of this hypothesis is left for future work. The distribution of partisan tweets by the 8KO is skewed toward the two big parties (Liberals and Conservatives) with an increase of almost 15% in tweets echoing their agenda, compared to the partisan distribution of the tweets of the 8KR (Figure 3 and Figure 4, middle and bottom). This pattern suggests that trolls are using sharp power techniques and are focused in sawing the divide between the main actors, practically ignoring the smaller parties. It is interesting to note that troll accounts were also heavily invested in retweeting the People's party of Canada, which split from the Conservative party in 2018 (Figure 3). Looking at the retweet breakdown on the partisan level, it seems that the 8KO users retweet accounts linked to the Liberal party twice as much as

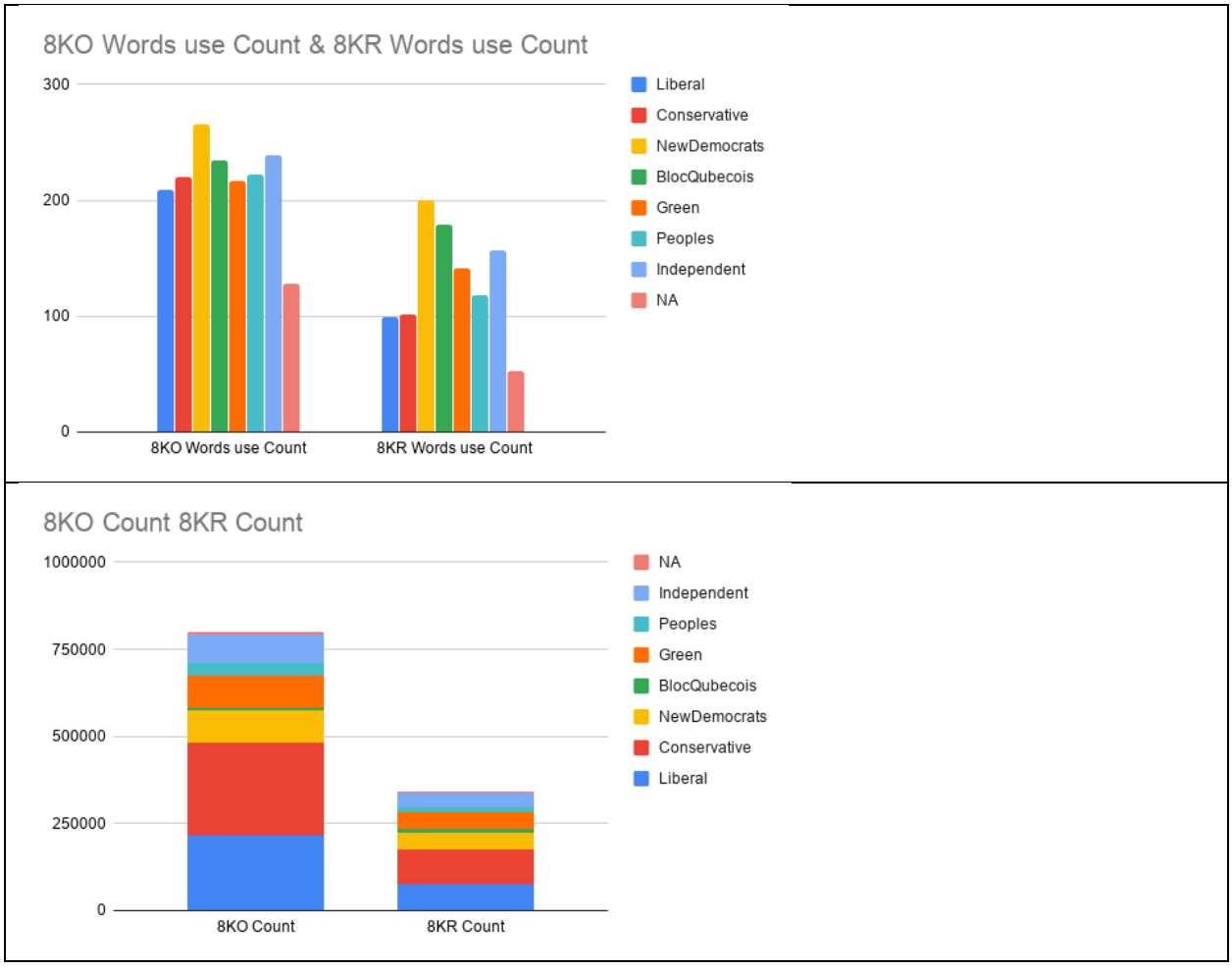
[13] For example, combine the list of hashtags most associated with the accounts of Justin Trudeau, the official account of the prime minister and the official account of the Liberal party.

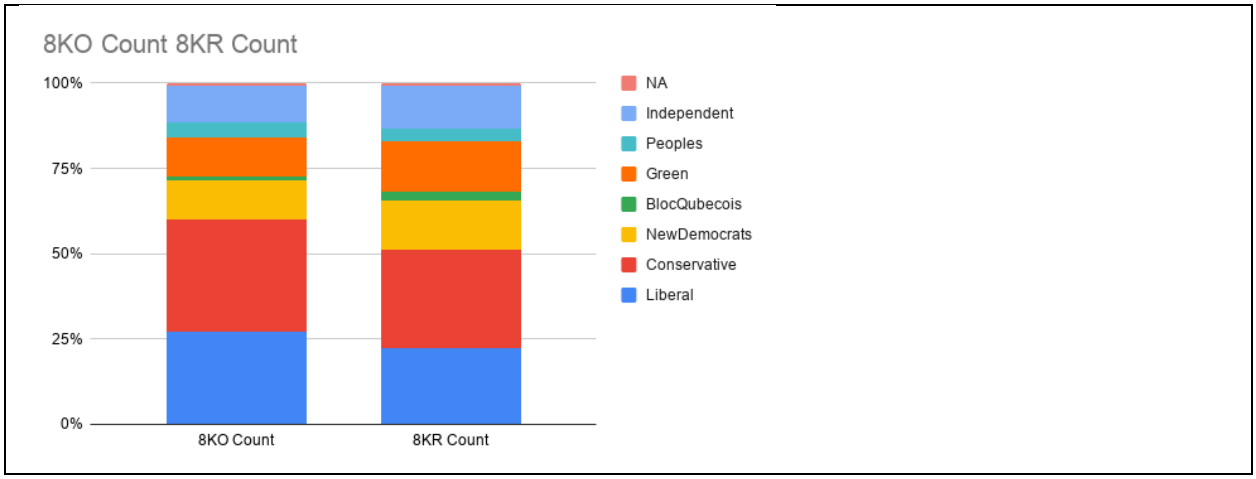
[14] We note that not all of the terms associated with a party reflect the party's agenda. Embeddings capture semantic similarity. Hence, party leaders, for example, will have high similarity. Indeed, we find that Justin Trudeau, the prime minister and head of the Liberal party, is very similar to Chrystia Freeland, his former foreign minister as well as to Andrew Scheer, the leader of the Conservative party. In fact, many of the words most associated with a party are party members, their rivals and their nicknames. The agenda associated with each party, as reflected by distributional semantics trained on social media data, is highly personal, rather than substantial. This trend is observed in other datasets as well, including the HT, which was collected clear of the bias induced by the filters used in the other datasets.

they retweet accounts linked to the conservative party. However, comparing it to the 8KR baseline, the 8KO users retweet the Conservatives twice as much as the baseline (~10,000 vs. ~5000) while they retweet the Liberals only 1.4 times than the baseline (~20,000 vs. ~14,500). We attribute the fact that the baseline of the Liberal party is significantly higher than that of the Conservative party to the fact that the Liberal party was in power, hence had a higher visibility to begin with. It is also worth noting that looking at the Conservative party and the People's party as one conservative block, we observe an even split between conservatives and liberals. It is worth noting however, that the engagement of the 8KO users with the People's party is disproportional to its limited political support. Reiterating the sharp power paradigm, the numbers in Figure 3 are aligned with the Russian interests with regards to Canadian politics – supporting Trudeau directly (promoting content of the Liberal party), and at the same time sowing discord on the right, promoting the more populist Brenier at Scheer's expense, in order to thwart the anti-Russian politics promoted by Shceer.

Surprisingly, there is not much promotion of the separatist agenda of Bloc Qubecois, although separatist tendencies are expected to be encouraged under the sharp power paradigm.

Figure 4: Distribution of partisan terms learnt by similarity of word embedding (Top). Distribution and raw counts of tweets containing partisan terms (Middle). Normalized distributions of partisan terms (Bottom)





Finally, we find that the 8KO users are engaged in the promotion of ideas borrowed from the American right, pushing issues such as immigration and border control and referring to American conservative politicians, pundits and Fox News fixtures.

The blackface scandal: a case study

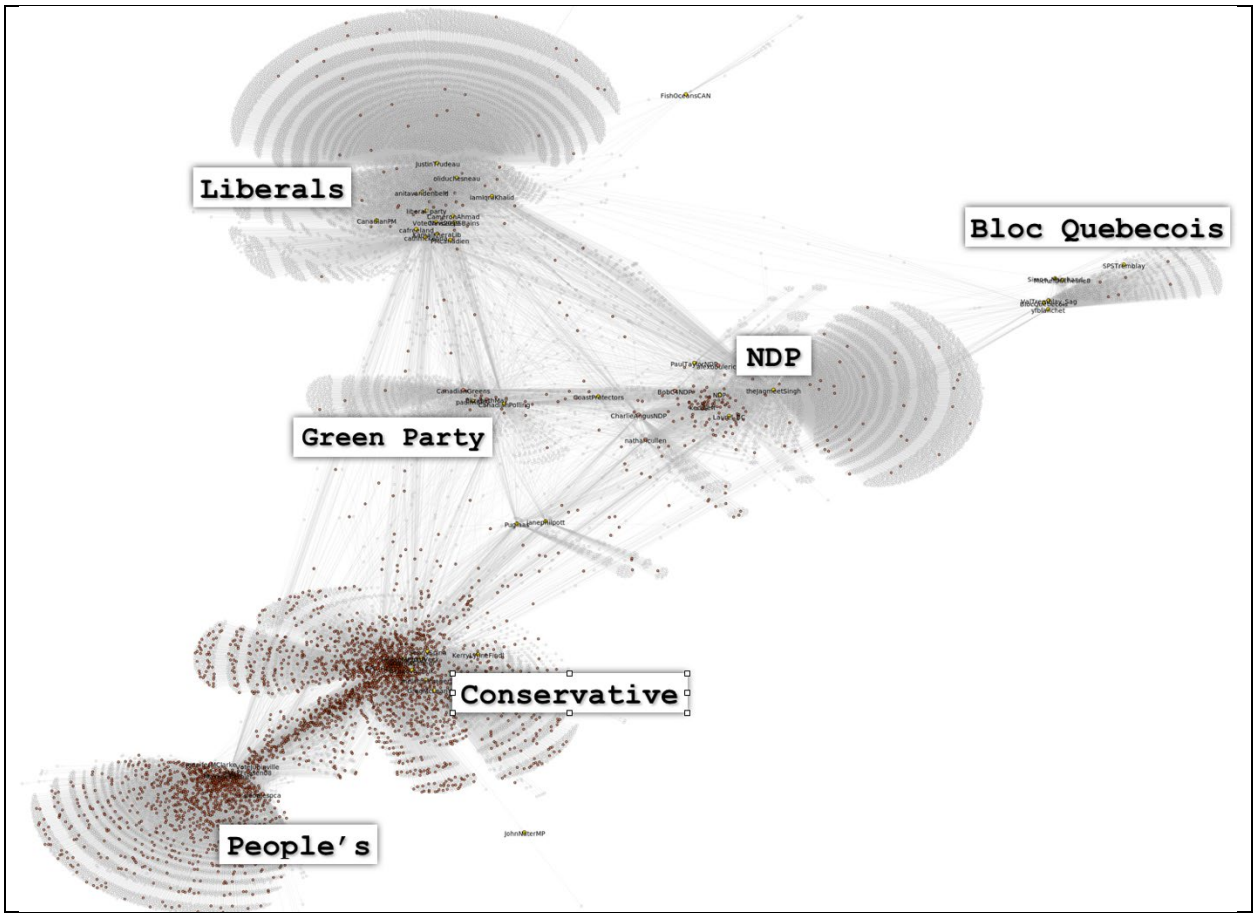
In this section, we focus on one specific story that dominated much of the news cycle during the 2019 Canadian campaign – Justin Trudeau's blackface scandal. Scandals involving party leaders often change the course of a campaign and have the potential to tip an election. It is therefore clear that adversaries, local and foreign alike, have the incentive to uncover scandals and fan the flames using social media in an attempt to capitalize on their foe's missteps.

The first photo of Trudeau featuring blackface[15] was published by *Time Magazine* on September 18, 2019. Trudeau quickly apologized, trying to contain the damage, while leaders of the competing parties issued (justified) condemnations, possibly trying to capitalize on the revelations. Users on social media raved. We note that the 'blackface' scandal was an emerging topic; therefore, it was not included in the predefined list of political terms that was used when collecting the data. Finding a large number of related tweets in both the CH and CT datasets serve as a validation to our data collection method.

Figure 5 presents a network of the Twitter users, reconstructed from the CH dataset. A node indicates a user and an edge indicates a retweet relation between two users (a threshold was applied). Official political nodes are yellow. Users referring to the blackface scandal are colored red. While some blackface references are sporadically scattered across the network, the vast majority of users promoting the story are related to The People's Party of Canada and the Conservative party (bottom-left). This pattern suggests a coordinated effort to keep the story in the political discourse. It is interesting to note that the 8KO trolls tweeted about the blackface scandal seven times as much as the users in the 8KR control group, even though a significant number of the 8KR users regularly promote the agenda of the Conservative party and the People's Party of Canada (see Figures 3 and 4). This disparity suggests that the blackface discourse was heavily promoted by trolls, rather than by authentic supporters.

[15] 'Blackface' is an artform in which a non-black actor paints his face black as an impersonation of a black character. Blackface performances were often used to ridicule the black community and promote racial stereotypes and are considered taboo.

Figure 5: A network based on the CH dataset. Nodes correspond to Twitter users. (Edges denote a retweeting relation (with a threshold of 3). Network rendering was done using force-directed layout. Clusters correspond to parties, as indicated by the label).



Conclusion

Already in 2017, the CSE acknowledged that "Cyber threat activity affecting the democratic process in Canada is a small fraction of the much larger global experience. During the 2015 federal election, Canada's democratic process was targeted by low-sophistication cyber threat activity" (Communications Security Establishment, 2017, p. 33). The CSE concluded that "hacktivist groups will deploy cyber capabilities in an attempt to influence the democratic process in 2019" (Communications Security Establishment, 2017, p. 33).

Drawing on the sharp power theoretical framework for our analysis, we analyzed four datasets collected in different time spans, based on different filters and covering the May 2019 European election and the two 2019 rounds of elections in Israel, to study foreign deployment of cyber capabilities in the 2019 federal election campaign in Canada. While finding a 'smoking gun"' explicitly attributing social network activity to a nation-state effort to interfere with and influence the outcome of political campaigns is almost impossible (Jaimieson, 2018), we did identify thousands of accounts invested in Canadian politics that present a unique activity pattern, significantly different from the activity of the accounts in a control group. We found that troll accounts were highly active in divisive discourses and latched onto controversies such as the blackface scandal. Many of these trolls were found to be active in other political campaigns,

occurring in different times and continents. Our large-scale cross-sectional approach (using four datasets CT, CH, EP, HT) enabled a distinctive perspective on foreign political meddling in Twitter during the recent federal election campaign, while mitigating some of the problems caused by the constraints on data collection. Our analysis of the blackface scandal reveals irregular activity. Following Walker and Ludwig, we are of the opinion that this foreign political meddling aimed at piercing and penetrating the political and information environments in Canada (Walker and Ludwig, 2017a; National Endowment for Democracy, 2017).

Previous analyses have found evidence for foreign meddling in political campaigns around the world, especially in the US Presidential election of 2016. These analyses build on qualitative analyses of limited datasets and on classified intelligence reports or other data not available to the research community (e.g. National Endowment for Democracy, 2017; Jamieson, 2018; Mueller, 2019). These works clearly substantiate the sharp power theoretical framework. Therefore, while failing short of finding a smoking gun, this article suggests some evidence for irregular activity that could be further interpreted by the sharp power paradigm. Foreign political meddling in Twitter in election campaigns aims at manipulating and poisoning the democratic process, posing a serious challenge to Western democracies and their values, as well as their societal resilience. In the 2019 federal election, however, such meddling posed less of a threat to Canada's democratic institutions and society.

References

Blei, David M., Ng, Andrew Y., & Jordan, Michael I. (2003) "Latent dirichlet allocation", *Journal of Machine Learning Research*, 3(March): 993–1022.

Bojanowski, Piotr, Grave, Edouard, Joulin, Armand, & Mikolov, Tomas (2017) "Enriching word vectors with subword information", *Transactions of the Association for Computational Linguistics*, 5: 135–146.

Broniatowski, David A., Jamison, Amelia M., Qi, SiHua, AlKulaib, Lulwah, Chen, Tao, Benton, Adrian, Quinn, Sandra C., & Dredze, Mark (2018) "Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate", *American Journal of Public Health*, 108(10): 1378–1384.

Communications Security Establishment (2017) Cyber Threats to Canada's Democratic Process, 16 June, https://cyber.gc.ca/sites/default/files/publications/cse-cyber-threat-assessment-e.pdf [accessed: 10 December 2019].

European Commission (2018) EU Code of Practice on Disinformation. Brussels 16 October, https://ec.europa.eu/commission/news/code-practice-fight-online-disinformation-2018-oct-16_en [accessed: 10 December 2019].

European Commission (2019) Joint Communication to the European parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions: Report on the Implementation of the Action Plane Against Disinformation. Brussels 14 June JOIN(2019) 12 final, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52019JC0012&from=EN [accessed: 10 December 2019].

Grinberg, Nir, Joseph, Kenneth, Friedland, Lisa, Swire-Thompson, Briony, and Lazer, David (2019) "Fake news on Twitter during the 2016 US presidential election", *Science*, 363(6425): 374–378.

Jamieson, Kathleen Hall (2018) *Cyberwar: How Russian Hackers and Trolls Helped Elect a President What We Don't, Can't, and Do Know*. New York: Oxford University Press.

Jurafsky, Dan, and Martin, James H. (2019) *Speech and language processing*, 3rd edition. London: Prentice-Hall Inc.

Lazer, David M. J., Baum, Matthew A., Benkler, Yochai, Berinsky, Aadam J., Greenhill, Kelly M., Menczer, Fillipo, Metzger, Miriam J., Nyhan, Brendan, Pennycook, Gordon, Rothschild, David and Schudson, Michael (2018) "The science of fake news", *Science*, 359(6380): 1094–1096.

Lloyd, Stuart P. (1982) "Least squares quantization in PCM", *IEEE Transactions on Information Theory*, 28(2): 129–137.

Mueller, Robert S. (2019) *Report on the investigation into Russian interference in the 2016 presidential election*. Washington, DC: US Department of Justice.

National Endowment for Democracy (2017) *Sharp power: Rising authoritarian influence*. Washington, DC: National Endowment for Democracy.

Nye, Joseph S. Jr. (1990) *Bound to lead: The changing nature of American power*, New York, NY: Basic Books Harper Collins Publishers.

Nye, Joseph S. Jr. (2001) *The paradox of American power: Why the world's only superpower can't go it alone*, Oxford: Oxford University Press.

Nye, Joseph S. Jr. (2004) *Soft power: The means to success in world politics*, New York, NY: Public Affairs.

Nye, Joseph S. Jr. (2011) *The future of power*, New York: Public Affairs.

Nye, Joseph S. Jr. (2013) "What China and Russia don't get about soft power". *Foreign Policy*, https://foreignpolicy.com/2013/04/29/what-china-and-russia-dont-get-about-soft-power/ [accessed: November 16, 2019].

Orttung, Robert, and Walker, Christopher (2014) "Russia's international media poisons minds", *The Moscow Times*, https://themoscowtimes.com/articles/russias-international-media-poisons-minds-40194 [accessed: November 16, 2019].

Phillips, Whitney (2015) *This is why we can't have nice things: Mapping the relationship between online trolling and mainstream culture*, Cambridge: MIT Press.

Sanger, David E. (2018) *The perfect weapon: War, sabotage, and fear in the cyber age*, New York, NY: Crown.

Shao, Chengcheng, Ciampaglia, Giovanni Luca, Varol, Onur Yang, Kai-Cheng , Flammini, Alessandro , and Menczer, Filippo (2018) "The spread of lowcredibility content by social bots", *Nature Communications*, 9(1): 4787.

Stella, Massimo, Ferrara, Emilio, and De Domenico, Manlio (2018) "Bots increase exposure to negative and inflammatory content in online social systems", *Proceedings of the National Academy of Sciences*, 115(49):12435–12440.

van der Maaten, Laurens, and Hinton, Geoffrey (2008) "Visualizing data using t-SNE", *Journal of Machine Learning Research*, 9(November): 2579–2605.

Walker, Christopher (2016) "The authoritarian threat: The hijacking of 'soft power'", *Journal of Democracy*, 27(1): 49–63.

Walker, Christopher (2018) "What is 'sharp power'?", *Journal of Democracy*, 29(3): 9–22.

Walker, Christopher, and Ludwig, Jessica (2017a) "From 'soft power' to sharp power: Rising authoritarian influence in the democratic world", in *Sharp power: Rising authoritarian influence*, National Endowment for Democracy. (Ed.). Washington DC: National Endowment for Democracy: 8–25.

Walker, Christopher, and Ludwig, Jessica (2017b) "The Meaning of sharp power", *Foreign Affairs*, https://www.foreignaffairs.com/articles/china/2017-11-16/meaning-sharp-power [accessed: November 16, 2019].

Yang, Kai-Cheng, Varol, Onur, Hui, Pik-Mai, and Menczer, Filippo (2019) "Scalable and generalizable social bot detection through data selection", *arXiv preprint arXiv:1911.09179*.