Efficient Detection of Online Communities and Social Bot Activity During Electoral Campaigns

Ludovic Rheault[†] and Andreea Musulan[†]

[†]Department of Political Science, University of Toronto

Abstract

Threats of social media manipulation during elections have become a central concern for modern democracies. This study tackles the problem of identifying the purpose and origins of social bots during electoral campaigns. We propose a methodology—uniform manifold approximation and projection combined with user-level document embeddings—that efficiently reveals the community structure of social media users. We show that this method can be used to predict the partisan affiliation of social media users with high accuracy, detect anomalous concentrations of social bots, and infer their geographical origin. We illustrate the methodology using Twitter data from the 2019 Canadian electoral campaign. Our evidence supports the thesis that social bots have become an integral component of campaign strategy for national actors. We also demonstrate how the methodology can be used to identify clusters of foreign bots, and we show that such accounts were used to share far-right and environment-related content during the campaign.

Keywords: Social bots; foreign interference; elections; social media user embeddings; fake news; Twitter

Author accepted version of forthcoming paper. Please cite as:

Rheault, Ludovic and Andreea Musulan. 2021. "Efficient Detection of Online Communities and Social Bot Activity During Electoral Campaigns." *Journal of Information Technology & Politics*. Forthcoming.

1 Introduction

The 2016 US election raised significant concerns regarding the potential of foreign interference on social media in democratic elections. Released in redacted form, the Mueller Report revealed some of the activities of the Russian-based Internet Research Agency (IRA) during that campaign, which targeted public opinion on the presidential candidates (Mueller 2019, 28). A key part of these operations were conducted on social media, and Twitter identified over 50 thousand bot accounts with Russian ties having shared content on the website during the weeks preceding the election (Twitter 2018). Moreover, there is widespread evidence that misinformation circulated on social media during the campaign, including a much-discussed false news story claiming that the Pope endorsed Donald Trump's candidacy (Bessi and Ferrara 2016; Silverman 2016; Allcott and Gentzkow 2017; Guess, Nyhan, and Reifler 2020; Lazer et al. 2018). While there is no consensus about the effectiveness of such persuasion efforts, the experience of 2016 highlighted in spectacular fashion the role played by digital platforms in modern elections.

This paper's objective is to investigate the nature and origins of social bot activity during electoral campaigns—that is, automated accounts posting content on social media. We introduce an efficient methodology that outperforms existing approaches at revealing the community structure of Twitter users along partisan lines. Specifically, we identify partisan clusters by applying the uniform manifold approximation and projection algorithm (McInnes, Healy, and Melville 2018; Becht et al. 2019) to a custom model of document embeddings (Le and Mikolov 2014). We provide extensive validation of this methodology and show that it can be used to predict partisan affiliations of social media users with a high level of accuracy. Moreover, we present a novel set of empirical tests to investigate signs of foreign interference during elections, by measuring anomalies in the distribution of geolocation tags across user clusters, and discrepancies in the distribution of social bots. To our knowledge, no existing work in the literature has proposed such a straightforward solution to the detection of electoral interference on social media. Our approach does not require prior information about the network structure, the access to which tends to be rate-limited by social media companies. Finally, the method we propose to generate

visualizations of user clusters is completed in a fraction of the time required by force-directed algorithms for network visualizations.

We implement our approach using an original collection of 18.9 million messages posted on the Twitter platform during the 2019 electoral campaign in Canada. We identify social bots using three different techniques, including machine learning models. We estimate that approximately 8 percent of accounts involved in political discussions during the electoral campaign were social bots, combining for 2.4 million messages (about 13% of the total volume). Although we find evidence of foreign interference, we show that many bots were used to spread official party materials during the campaign, and conclude that these bots were largely used by national partisans. Given the pressing concerns for democracy raised by social bots, our methods and findings have concrete implications for scholars and social actors involved in the monitoring of elections.

The Canadian 2019 election represents a fertile ground to assess the threat posed by automated content on social media. The election took place in a context of high alert, and followed the creation of entities such as Rapid Response Mechanism (RRM) Canada and the Canadian Centre for Cyber Security in 2018. The campaign featured a prominent scandal affecting the public image of the incumbent prime minister Justin Trudeau—the release of photographs featuring the Liberal leader costumed as a racial minority—which generated a flurry of reactions on Twitter, opening up an easy opportunity to launch targeted attacks toward the candidate. A series of articles posted by the Buffalo Chronicle, categorized as false news by the Agence France-Presse, eventually made the headlines with a fake sex scandal involving Trudeau (Chown, Lytvynenko, and Silverman 2019). Testing whether campaigns of political interference took place during this heated context can provide a useful ground to judge whether the events from the 2016 US election are likely to reemerge in future elections around the world.

2 The Role of Social Bots in Political Communication

The impact of social bots has received considerable attention in recent years with the publication of studies documenting their role in disseminating political information, in particular during elections (see Ratkiewicz et al. 2011; Bessi and Ferrara 2016; Ferrara 2017; Shao et al. 2018; Vosoughi, Roy, and Aral 2018; Bastos and Mercea 2019; Bradshaw et al. 2019). *Social bots* refer to user accounts whose posting behavior is achieved through the use of automated programs (Bessi and Ferrara 2016; Davis et al. 2016). On Twitter, these bot accounts can be set up relatively easily and bulk-managed to share website links, follow targeted users, retweet, and post original content (for an extended review, see Howard and Kollanyi 2016; Gorwa and Guilbeault 2020). Automated accounts may serve to amplify the visibility of candidates or to disparage political opponents, for instance by spreading promotional materials or inimical news about the opposition (see e.g. McKelvey and Dubois 2017; Howard, Woolley, and Calo 2018). In spite of a rapidly growing literature on the topic, retracing the origin of social bots is an ongoing challenge.

In this section, we examine assumptions about the motivations of actors behind the proliferation of social bots. Are bots used by national actors as part of their campaign strategy, or are they operated by foreign actors seeking to influence electoral outcomes? Both alternatives are plausible. The first one stems naturally from modern political communication theory, while the second has roots in a long history of electoral interventions by foreign actors. We review both lines of argument in turn below.

The incentives for political parties to deploy social bots arise naturally from the nature of campaigns. Electoral campaigns are competitions for visibility, and automated accounts can help to expose voters to campaign materials. In Benoit (2007)'s *functional theory* of electoral campaigns, candidates vie for distinction from their opponents. The theory also predicts that attack messages are risky, in that the strategy can inadvertently increase the visibility of opponents. As a result, self-promotion should be the most frequent strategy overall, in particular for incumbents (Benoit 2007, 55). Recent studies on social media campaigning confirmed that challengers and smaller parties are the ones relying more aggressively on negative campaigning (Auter and Fine

2016; Borah 2016). Parallel to this development, *mere exposure theory* has been invoked to explain the role of campaign messages in boosting name recognition (Grimmer, Messing, and Westwood 2012; Kam and Zechmeister 2013; Bright et al. 2019). Simply put, the theory posits that being exposed to a stimulus induces likeability (Zajonc 2001), thus providing a rationale for candidates and parties to send repeated messages during campaigns. Kam and Zechmeister (2013) find this type of effect to benefit challengers, who typically do not enjoy the same visibility as incumbents.

The logic of functional theory and mere exposure theory arguably extends to social bots, which are ideally suited to circulate campaign content broadly. In the presence of a national-based bot strategy, we should observe tangible efforts at relaying official campaign messages. Emerging parties and challengers, in particular, have a strong incentive to take advantage of social media automation to achieve name recognition. The cost effectiveness of digital media makes this option appealing: social bots can be operated at a trivial fraction of the cost of traditional advertisement. There is ample evidence that such national-based strategies have been deployed during campaigns. Early examples from Western democracies suggest that bots often served to increase politicians' number of followers (Woolley 2016), but practices are diversified. McKelvey and Dubois (2017) surveyed cases of parties having relied on bots during Canadian election campaigns between 2012 and 2015. They mention the @CAQbot account, created by a supporter of the Coalition Avenir Quebec in 2012, which helped to propel the nascent provincial party among the trending topics on the Twitter platform, at a time when it was most in need of visibility. We find concrete examples of that nature in our data.

We should distinguish between two types of national-based strategies involving social bots: transparent and covert. Bots that openly reveal their nature are common. For instance, they are used by news and weather channels to automatically post updates of interest (see Ferrara et al. 2016; Stieglitz et al. 2017). Similarly, several political bots in our dataset openly state their political affiliation and do not seek to disguise their purpose. The CAQbot example cited above fits that category; the username made the nature of the account transparent to the public. Used with moderation, we would argue that these accounts are even desirable for modern campaigns,

as a cost-effective way to circulate information to voters. The second type of strategy—covert—is a more delicate subject matter. There is considerable risk for an official party to be associated with the bulk-management of bot accounts. While the behavior is typically not subjected to legal restrictions, covert operations raise ethical issues. Many instances of social bots involving national actors are of that nature.[1] A growing body of literature provides evidence that political parties have been involved in the covert deployment of social bots (Woolley 2016; Schäfer, Evert, and Heinrich 2017; Howard, Woolley, and Calo 2018). A recent report from the Oxford Internet Institute, for instance, suggests that political parties or national governments from 48 countries have been involved in social media manipulation in 2018, typically involving account automation (Bradshaw and Howard 2018). In practice, the task of operating the bots will be delegated to social media outreach firms (Bradshaw and Howard 2018).

The most challenging question for political science research, we argue, is how to distinguish between national and foreign bot activity. The idea that foreign powers seek to interfere in democratic elections should not be a surprise to political science scholars. The history of electoral interference is well documented (Levin 2016; Bubeck and Marinov 2019), and predates the advent of social media by many decades. While objectives may have evolved, the principal one—affecting the electoral outcome in the target country—has arguably remained. As emphasized previously, there is evidence of foreign-controlled social bots spreading content during the 2016 US election. The Mueller Report concluded that the Russian government "carried out a social media campaign that favored presidential candidate Donald J. Trump and disparaged presidential candidate Hillary Clinton" (Mueller 2019, 1). Aside from the attempt to influence the outcome, two features of the 2016 interference activities stand out. The first is the circulation of fake news, which received significant attention from scholars. According to the US Senate's Select Committee on Intelligence, "[a] free and open press—a defining attribute of democratic society—is a principal strategic target for Russian disinformation" (SSCI 2019, 20). The second strategy mentioned by

1. In this study, we refer to these bots as "national" in origin, even though formally associating these accounts to a party's central organization is an endeavor that falls beyond the scope of this research—bots may be operated by volunteers or party militants.

the committee consists of exacerbating "social fissures" (SSCI 2019, 21), for instance by fueling the debate on divisive issues, with the goal of accentuating polarization and social unrest (see also Stella, Ferrara, and De Domenico 2018).

More concretely, we identify three strategies by which foreign-controlled bots could aim to affect democratic debates during electoral campaigns. First, bots can be deployed to introduce *new* content in the campaign, with the objective of influencing the outcome. The typical case is the spread of false news, whereby bots are designed to disseminate links pointing to disreputable sources of information to either help or besmirch a targeted candidate (see Lazer et al. 2018). The extant literature has focused extensively on this type of interference, with the 2016 US election as the epitome. More generally, we may include any attempt to introduce claims that would not have appeared organically in the online debates involving national actors, whether they are supported by a link to a false news story or not. Thankfully, detecting this type of interference is more straightforward. Computer algorithms are particularly efficient at finding unusual patterns—new links, new expressions—and clustering them apart. We illustrate this with concrete examples in our empirical section. Even if the general public adopts a fake news story that was originally spread by foreign bots, those are usually noticed and quickly exposed by mainstream media, and retracing the initial account that shared a false news story is feasible.[2]

In our view, two other foreign strategies pose a tougher challenge for researchers. In the second type, foreign actors could hide their tracks by designing social bots that replicate the behavior of existing national partisans, thereby inflating the apparent size of a group of supporters. The objective, once again, would be to favor the electoral fortunes of a targeted party or candidate. This type of interference is more difficult to distinguish from strategies deployed by national actors. However, as we illustrate in this study, it is possible to test for anomalous concentrations of social bots among groups of partisans, and detect signs of this strategy empirically. While the existence of discrepancies does not rule out the possibility that specific groups of national partisans are the ones overflowing social media with automated accounts, finding disproportions in

2. In fact, the Twitter API always reports the first user to have posted a story retweeted on the site.

the distribution of social bots helps to narrow the scope of the investigation.

A related, third strategy is alluded to in the above-mentioned Senate study: the amplification of messages on multiple sides of the political spectrum. The objective of foreign actors may be to inflame existing social divisions, thereby fostering polarization. For empirical research, the difficulty is to separate this type of intervention from competing national interests, especially given the fact that extreme ideological positions tend to be overrepresented on social media platforms such as Twitter (Barberá and Rivero 2015). Nonetheless, the clustering approach presented in this study allows to reveal affinities between social bots and human users, identify the content they share and the language they use. Combined with a qualitative assessment, the method can help researchers to identify sophisticated strategies that aim to exploit divisive issues.

3 Data Collection

Our data collection comes from a real-time stream of the Twitter platform between September 3 and October 23, capturing the week preceding the launch of the campaign up to the day following the election. Messages posted on the site are called tweets or statuses. The stream was filtered with an extensive list of track terms comprising the principal hashtags and keywords used to refer to the election (e.g. #cdnpoli and #elxn43), as well as the names of parties and of party leaders. The free version of the streaming API returns up to 1% of the global volume posted on Twitter at any given time, and also reports to developers the cumulative amount of tweets not returned when exceeding this limit. Only 297 tweets could not be retrieved due to rate limiting, compared to 19 millions collected. As a result, our dataset contains the virtual totality of tweets meeting our search criteria.

We identify social bots with three different methods. Our primary method is a random forest classifier with the same features and specification used by Yang et al. (2020) for constructing the new Botometer Lite model. This model relies on user metadata to detect social bots—including, for instance, the lexical characteristics of usernames; the growth rates of tweets, followers, and

friends; and the followers to friend ratio. Because of its reliance on metadata, as opposed to tweet content, this model can be scaled easily to examine a large quantity of users. We use the same training datasets as in the original study, for which we fetched extended metadata from the Twitter API (see Online Appendix for an extended discussion of this model). Our training sample differs slightly due to missing data points,[3] but achieves a similarly high level of accuracy with an average F1 score of 0.97 computed using five-fold cross-validation. For all intents and purposes, our model is essentially the same as the scalable new version of Botometer. The benefit of our custom implementation is that it is not restricted by rate limits, such that we can predict user types for the entire collection of 1.7 million users. Secondly, we double-checked accounts that posted at least 50 tweets during the campaign (roughly one per day) using the full version of the Botometer API (Davis et al. 2016), which has rate limit restrictions.[4] There is a strong overlap between the two methods (89.9% of the users tagged using the live API have the same predicted label using the scalable model). Finally, we performed an exhaustive search of the 1.7 million users in our dataset to verify whether they had been suspended by Twitter. The company frequently suspends accounts for violations of terms of service, and such accounts are often social bots engaged in spamming or other nefarious behaviors (Twitter 2018).

Table 1 reports the breakdown of users as humans or bots. An account is categorized as a bot if it was flagged by either one of the three methods described above. According to this methodology, social bots represented approximately 8% of users engaged in discussions related to the electoral campaign, and 13% of the total volume of tweets. Note that the user counts are based on primary keys (i.e. identification numbers unique to a given user), as opposed to screen names: Twitter users can change their user name at any time on the site, whereas the primary id number remains constant. In total, 68 percent of the messages are "retweets" (shared messages with no original content).

3. The extended metadata for some of the users in the original training datasets could not be retrieved from the Twitter API due to account deletions or suspensions. We also included an additional political dataset due to its domain relevance.

4. The Botometer API returns probability scores. We mark a user as a bot when the aggregated score is greater than 0.5 (Davis et al. 2016).

Table 1: Summary of Data Collection

	Tweets		Users	
User Type	Count	%	Count	%
Social Bot	2,436,032	12.9	132,611	7.9
Human	16,455,410	87.1	1,539,803	92.1
Total	18,891,442	100.0	1,672,414	100.0

Notes: Total number of tweets in our data collection, broken down by user type, after removing non-political and spam content (the total sample size was 19.3 million). Social bots include accounts flagged by any one of the three methods for bot detection discussed earlier in this section. The Online Appendix provides a detailed breakdown of the number of accounts detected as bots by each method, and the number of overlaps.

4 Methodology

Our analysis relies on a set of methods straightforward to replicate in other contexts. We start by comparing the aggregate proportions of internet domains shared by social bots against those shared by human users. This comparison helps to assess for the presence of discrepancies in the origin and quality of the websites referred to by each type of user. As mentioned previously, a recurring strategy used by foreign actors operating social bots is to introduce disruptive contents during a campaign. Therefore, we recommend this preliminary step as a means to investigate signs of foreign interference.

The central piece of our methodology involves the computation of *social media user embeddings* to detect the presence of communities of interest, combined with a recent advance in clustering techniques. Several approaches have been proposed to extract numerical features associated with social media users (Pan and Ding 2019) or estimate their ideology (Barberá 2015; Temporão et al. 2018; Eady et al. 2019). We rely on an approach that is completely unsupervised and scalable, using document embeddings (Le and Mikolov 2014) computed with social media users as an index variable. Wu et al. (2019) provided extensive evidence of the reliability of a similar approach for measuring partisan associations. Like word embeddings, the model relies on a fully connected neural network predicting the occurrence of words in the tweet corpus by the preceding and following words. The methodology has become ubiquitous in recent years. Our implementation considers the users writing a tweet as an additional predictor in the model,

and we also rely on indicator variables for the calendar date to account for temporal changes in the content of online discussions. We fit the embeddings with a hidden layer of size 200, 5 epochs and a context window of 10 words. These embeddings allow us to map both users and words in a common vector space. Simply put, the methodology builds on the property that two users who rely on a similar language—and as a result, help predict the occurrence (or absence) of the same words—will have similar embeddings.

The next key step relies on uniform manifold approximation and projection (UMAP) (McInnes, Healy, and Melville 2018; Becht et al. 2019) to reveal and visualize neighboring communities of users. UMAP is a non-linear dimensionality reduction technique in the family of manifold learning, and can be construed as an unsupervised learning method. As with other dimensionality reduction techniques, the goal is to find the latent structure of a high-dimensional dataset. In our case, we have large vectors that represent social media users, and we want to reveal connections between subgroups of users. Manifold learning methods proceed by mapping points in a nearest neighbor graph while attempting to preserve the original distance between observations, which helps to represent local patterns of data points with higher fidelity (Maaten and Hinton 2008; McInnes, Healy, and Melville 2018). This usually comes at the cost of the global structure—that is, the dimensions of the reduced space do not reflect the variance between data points and lose their interpretability, in contrast to matrix factorization approaches such as principal component analysis. The UMAP algorithm improves upon previous manifold learning methods by using an objective function accounting for both the local and global structure of the data (McInnes, Healy, and Melville 2018). The balance between local and global structure is controlled by the size of local neighborhoods considered when fitting the algorithm. Moreover, the UMAP algorithm can be estimated in a fraction of the time needed for earlier methods.

Specifically, we start by extracting 50 principal components from the user embeddings (a common intermediary step to manifold learning that further enhances the clarity of visualizations, see Maaten and Hinton 2008), and we reduce the resulting user matrix to two dimensions using the UMAP model fitted with 200 neighbors. As illustrated in the next section, this method is par-

ticularly promising to retrieve the communities of users on social media platforms. Conveniently, our procedure also affords us with the opportunity to interpret the semantic attributes associated with the clusters, by retrieving the words closest to each user in the original embedding space. We fit the manifold learning model on the subset of the most active users who tweeted at least once a day on average during the campaign.

As we illustrate in the empirical section, this methodology can be used to predict the partisan affiliation of specific users with high accuracy. After identifying the accounts that tweeted the most for each cluster, we proceed by assigning every other user to the cluster with which they share the highest cosine similarity, using the original embedding space to perform the vector arithmetic. The online appendix provides additional details. By comparing the proportions of social bots across clusters, this approach allows to test for unusual levels of bot activity across partisan groups. Likewise, the approach can be used to detect suspiciously low levels of Canadian geotags across clusters. To be sure, actors engaged in foreign interference are probably less likely to activate the Twitter geotagging functionality. However, we argue that the absence of Canadian geotags, in and of itself, can be informative. In other words, if the odds of observing geotagged users from Canada are significantly lower for some clusters, this provides indirect evidence of foreign activity.

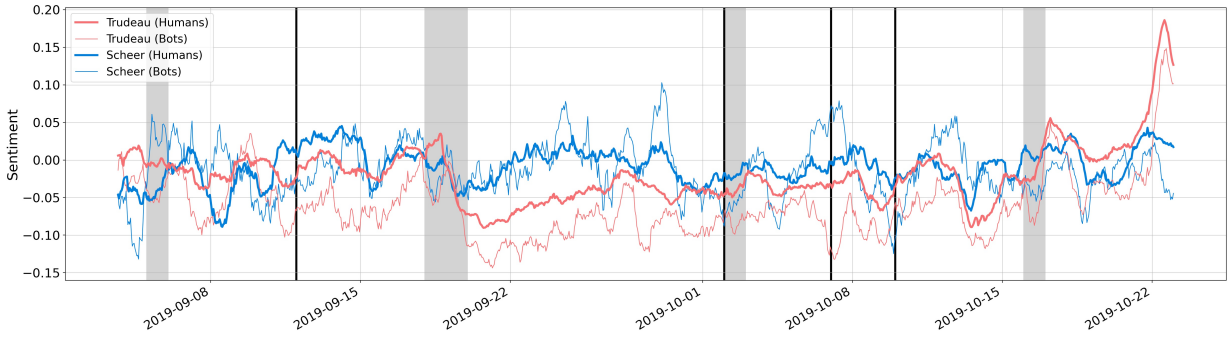
5 Empirical Evidence on the 2019 Campaign

We begin this section with a brief descriptive account of campaign events, before turning to our main analysis. In an oversimplified fashion, the unfolding of the 2019 electoral campaign can be summarized with two key events. The first was the scandal following the release of photographs presenting incumbent prime minister Justin Trudeau wearing "brownface" makeup (and later a blackface). The initial photo was published by Time Magazine on September 18, 2019, and was abundantly discussed in the media. The second event was Barack Obama's open endorsement of Trudeau on October 16, five days before the vote, which happened on the Twitter website. Both

these events mark turning points in the tone of discussions toward Trudeau during the campaign, as we illustrate below.

Figure 1 plots a moving average of the hourly sentiment toward the two main party leaders, Justin Trudeau (Liberal) and Andrew Scheer (Conservative), after breaking down social bots and human users. We compute tweet sentiment scores using the Vader library (Gilbert and Hutto 2014), using the subset of tweets that mention each leader.[5] While the average sentiment toward each leader is roughly similar in magnitude before the first incident, the series for Justin Trudeau takes a decided plunge with the release of the blackface photographs. Trudeau's sentiment remains lagging to that of his main rival for a large part of the campaign, and only surpasses Andrew Scheer's a few days before the election. In fact, this late upsurge coincides with the former US President's social media intervention. The final boost in sentiment in Figure 1 is the day following the election, during which many users congratulated the winner.

Figure 1: Hourly Leader Sentiment on Twitter, by User Type



The figure shows an 18-hour moving average of the mean sentiment toward each of the two principal party leaders during the campaign, for human users and suspected bots. Vertical lines indicate leaders' debates. Shaded areas represent, respectively from left to right, 1) the announcement that Justin Trudeau would not participate in some leaders' debates (September 5-6, 2019), 2) the blackface controversy (September 18-20, 2019), 3) the revelation that Trudeau uses two planes during the leaders' tour (October 2-3, 2019), and 4) Barack Obama's endorsement of Trudeau (October 16-17, 2019).

A second noticeable trend is that the average sentiment expressed by social bots was more negative, on the whole, than the average sentiment expressed by human users. The difference between user types, calculated over the full period, is statistically significant based on two-sample

5. The online appendix provides an extended analysis of leader sentiment.

t-tests and bootstrapped t-tests, for both leaders ($p < 0.001$). Overall, bots were even more negative toward Justin Trudeau, with a gap of −0.04 points on the sentiment scale relative to human users, compared to a gap of −0.02 for Andrew Scheer. This finding aligns with expectations from functional theory, which posits that challengers are more likely to engage in negative campaigning. Following the blackface scandal in particular, we observe a pronounced dip in sentiment toward Trudeau among social bots, to an even greater extent than among the general public. This downtrend suggests that some accounts have been used purposefully as a means to amplify reactions to the incident. Our online appendix reports additional results showing that social bots had no significant influence on leader sentiment during the campaign.

5.1 Investigating the Origins of Social Bots

We now turn our attention to inferring the origin of social bots. We start by comparing the substantive focus of social bots with that of regular users. Table 2 reports the top 20 domains most frequently linked to by each group of users. We interpret the overarching pattern as an indication that social bots were very similar to human users. Rather strikingly, social bots referred heavily to the same national news sources consulted by regular users. In fact, the top five domains shared by each group are essentially the same. Even a specific example like the Spencer Fernando website—a popular political blogger who has expressed critical views on the incumbent government—is referred to by both groups of users in similar proportions. Among the few anomalies are the Buffalo Chronicle, which ranks higher among social bots' favorite domains, as well as paper.li, a resource that can be used to create news-looking web pages.

We interpret the similarities in the distributions from Table 2 as an indication that social bots are used in large part by national partisans. It is possible—albeit far-fetched—that foreign operators of automated accounts had such a strong understanding of the Canadian political environment that they chose to share national news media links in proportions that align very closely with the preferences of the public. In our view, a more plausible interpretation is that social bots are largely operated by supporters of partisan groups. A piece of evidence support-

Table 2: Comparing the Internet Sources of Humans and Social Bots

Humans			Social Bots		
Domain	Count	Percent	Domain	Count	Percent
cbc.ca	511428	10.5	cbc.ca	67599	8.0
globalnews.ca	247281	5.1	youtube.com	56778	6.7
youtube.com	229705	4.7	globalnews.ca	36708	4.3
theglobeandmail.com	206064	4.2	torontosun.com	33802	4.0
torontosun.com	181506	3.7	theglobeandmail.com	28353	3.4
ctvnews.ca	174657	3.6	nationalpost.com	25894	3.1
thestar.com	172578	3.6	ctvnews.ca	23890	2.8
nationalpost.com	155359	3.2	thepostmillennial.com	23463	2.8
thepostmillennial.com	132284	2.7	thestar.com	20515	2.4
liberal.ca	66506	1.4	facebook.com	11150	1.3
huffingtonpost.ca	64064	1.3	spencerfernando.com	10368	1.2
facebook.com	60386	1.2	peoplespartyofcanada.ca	9440	1.1
time.com	56424	1.2	liberal.ca	8979	1.1
spencerfernando.com	52586	1.1	huffingtonpost.ca	8163	1.0
washingtonpost.com	42114	0.9	buffalochronicle.com	7426	0.9
conservative.ca	41049	0.8	paper.li	7191	0.9
macleans.ca	40436	0.8	tnc.news	6757	0.8
nytimes.com	39601	0.8	time.com	6649	0.8
pressprogress.ca	37064	0.8	conservative.ca	6513	0.8
theguardian.com	34623	0.7	washingtonpost.com	6195	0.7

Notes: The table reports the most frequently shared domains for each user type, along with counts and percentages. These calculations are performed after an exhaustive domain conversion for the links posted using url shorteners (by retrieving the original, expanded urls), including those from services external to the Twitter platform.

ing this interpretation is the large proportion of content shared by social bots from the official websites of political parties (liberal.ca, conservative.ca, and peoplespartyofcanada.ca in Table 2). This behavior serves the objective of parties, and is consistent with the type of national campaign strategy expected from political communication theory. In other words, many bots were ostensibly deployed to get the campaign message out.

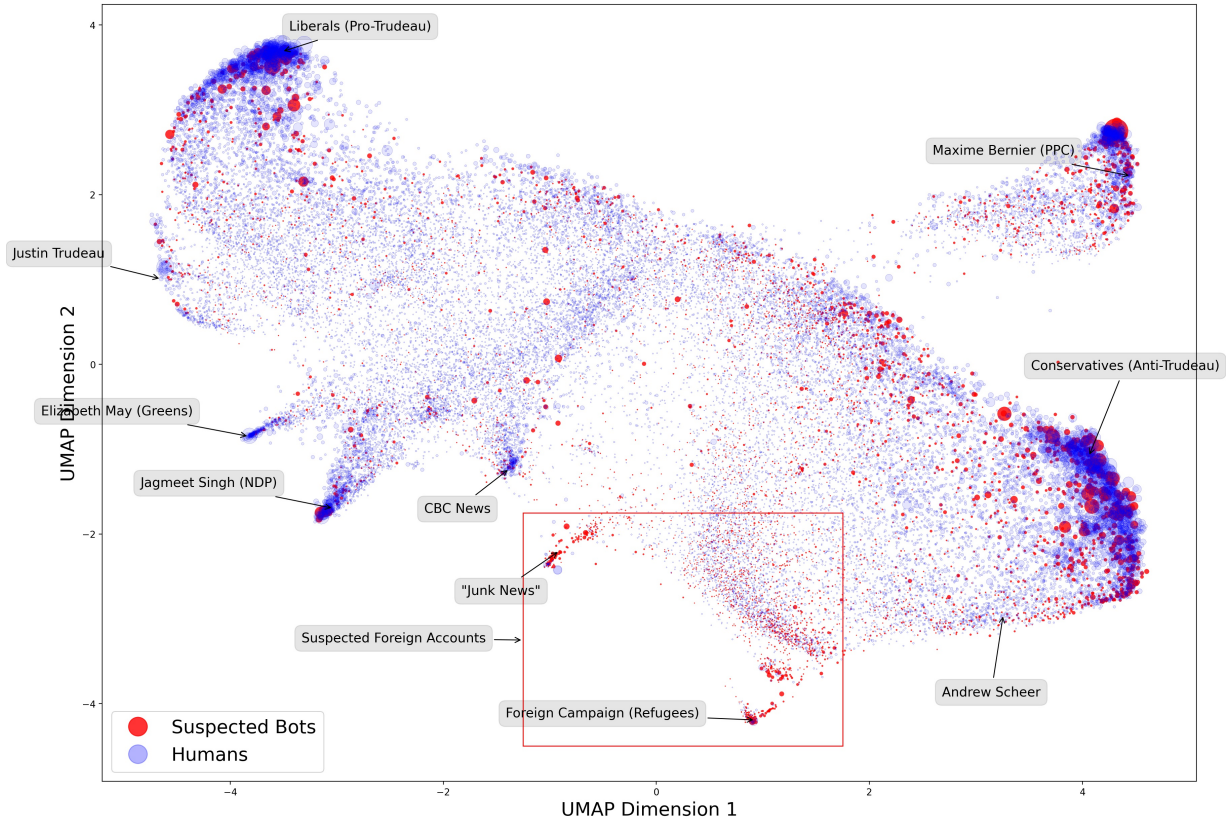
To examine this claim more thoroughly, we rely on social media user embeddings to generate a mapping of the Twitter ecosystem during the campaign, using the method described in the previous section. For this analysis, we exclude French speakers to improve the accuracy of the language model. Figure 2 depicts the location of users who tweeted at least 50 times during the campaign (roughly once a day) in a two-dimensional space estimated using the UMAP algorithm (n = 38,027). The size of the scatter points reflects the number of tweets posted by each user, whereas the color codes indicate whether users are bots or humans.

Figure 2a includes superimposed labels to facilitate interpretation. We indicate the location of relevant political actors, and we rely on the words closest to users in the labeled areas of the figure to describe clusters in substantive terms (see Table 3 for details). The largest clusters emphasize the primary axis of division (on the horizontal dimension), which separates Liberal supporters—the cluster on the top left—from Conservative ones—on the right. This is consistent with expectations, given the distribution of the vote in the 2019 election, primarily between these two major parties. We can further distinguish between Conservative partisans and supporters of Maxime Bernier's People's Party of Canada (PPC), a fringe political party embracing libertarian views that enjoyed a surprising amount of attention on Twitter, relative to its vote share. PPC supporters fall in a different cluster on the top right area of the plot.

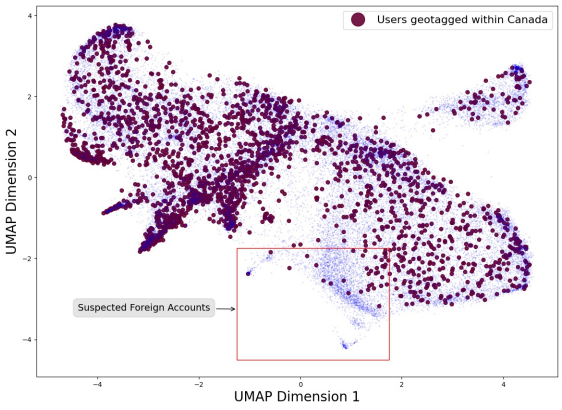
We may assess the face validity of this mapping in substantive terms by examining the top words associated with the principal clusters (Table 3). These top words were retrieved by identifying the word embeddings most similar to the users located near the center of each cluster, using cosine similarities. For example, the aforementioned Liberal (pro-Trudeau) cluster is associated with hashtags used to attack the leader opposite, Andrew Scheer, as well as endorsements

Figure 2: Mapping of Twitter Users during the 2019 Canadian Election

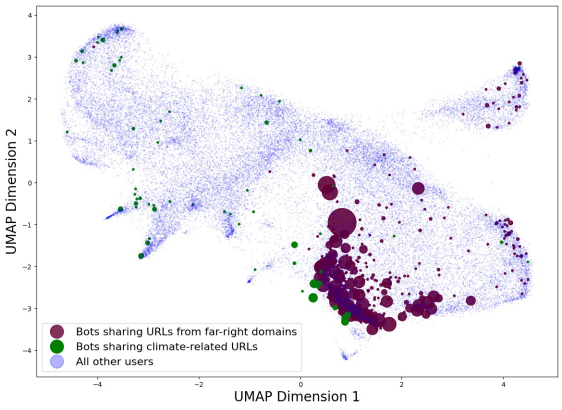
(a) UMAP projections for most active users



(b) Users geotagged within Canada



(c) Content shared by suspected foreign bots



of the Liberal Party. Conversely, the Conservative cluster contains hashtags used to attack the incumbent prime minister. Combined with the fact that party leaders also fall near the appropriate partisan clusters, these keywords leave little doubt regarding the ability of the proposed methodology to reveal substantively meaningful groups of users.[6]

Table 3: Top Words by UMAP Cluster

Cluster	Top Words
Liberal	#scheerdisaster, #neverscheer, #scheerlies, #yankeedoodleandy, #voteliberal
Greens	#votegreen, greenparty.ca, #gpc, @canadiangreens, #greenparty
NDP	#initforyou, #uprisingh, ndp.ca, layton, #ndp
PPC	#peoplesparty, #ppc, max, ppc, #peoplespartyofcanada
Conservative	#liberalsmustgo, #trudeaumustgo, #trudeauworstpm, #cpcmajority, #trudeauisdone
Foreign (Refugees)	@refugees, #helprefugeesinindonesia, @sbsnews, @jacindaardern, #resettlement4singlerefugees
Junk News	amid, #news, emerges, #breaking, surfaces

Notes: The table shows the five tokens most similar to the average user embeddings for accounts located within each of the principal clusters identified on Figure 2. NDP stands for New Democratic Party and PPC for People's Party of Canada.

The visualization in Figure 2 gives some indication of the concentration of social bots in each cluster. On the whole, we observe many bots among the two principal clusters (Liberal and Conservative). Noticeable on this mapping are the high concentrations of social bots forming isolated neighborhoods at the bottom of the figure. One of these peninsulas comprises accounts that bear the characteristics of what Bradshaw et al. (2019) refer to as "junk news", more specifically a combination of automated news aggregators and accounts with news outlet-sounding names, some of which were suspended by Twitter during the election. This group branches out from regular news media organizations (the plot shows the Canadian Broadcasting Corporation, CBC News, for orientation), since these accounts borrow from the same language. We note that these accounts did not engage with other users, and were likely not central actors during the campaign. The second noticeable bot cluster, labeled "Foreign Campaign (Refugees)" reveals an organized campaign with a specific agenda focusing on refugees and foreign affairs. These accounts posted repetitive content addressed to various world leaders (including Justin Trudeau), and referring to

6. In the online appendix, we further validate the interpretation of these UMAP clusters in terms of political ideology.

countries in the Middle East and South-East Asia.

More generally, a large area of Figure 2 contains what we infer to be foreign accounts. The area is outlined using a rectangle, and suggests a high concentration of social bots. We deduce that these accounts are foreign in origin because of the distinction in the choice of language they make—at the basis of our methodology—and because of the suspiciously low occurrence of geotagged users located inside Canada within that cluster, as illustrated in Figure 2b. Users from that cluster also differ from regular party supporters in terms of the type of content they shared during the campaign. In particular, we find that social bots belonging to that cluster were more likely to share URLs related to environmental activism and from specific domains associated with the far-right. To illustrate, Figure 2c displays the number of articles shared by social bots from News Punch and the QAnon merchandise website, as well as the share count for the top two environment-related domains in the data collection (the Fridays for Future and Green New Deal websites).

5.2 Discrepancies in the Distribution of Bots and Canadian Geotags

We proceed by examining our empirical conclusions more systematically. After assigning social media users to their most likely cluster, we compute the proportion of social bots and Canadian geotags by cluster. Table 4 reports the results. While the five main partisan groups account for the majority of social bots, there is no clear evidence that one of them enjoyed a disproportionate level of bot support, apart from the PPC. To be sure, the odds of observing a bot among Conservative supporters, relative to all other users, are higher than for Liberal supporters. However, the percentage of bots in the Conservative cluster (about 27.3%) is actually proportional to the estimated size of that digital constituency (also 27.3% of all users). Liberal supporters may be lagging behind in their adoption of this technology, which parallels previous findings that right-wing partisans are more likely to use social bots (Schäfer, Evert, and Heinrich 2017; Keller and Klinger 2019). Meanwhile, the PPC cluster was slightly over-represented in the population of social bots. This is an anomaly, although we should point out that new parties tend to rely more aggressively

on social media to drum up new adherents and get their message across (see e.g. Auter and Fine 2016).

Table 4: Odds of Observing Social Bots and Canadian Geotags, by Cluster

			Social Bot		Canadian Geotag	
Group	Percent (%)	Bot Density (%)	Odds Ratio	p-value	Odds Ratio	p-value
Conservative	27.29	27.26	0.998	0.958	1.107	0.426
Liberal	26.18	13.47	0.393	0.000	2.174	0.000
NDP	10.98	5.56	0.436	0.000	1.110	0.432
PPC	10.63	13.15	1.333	0.000	0.600	0.012
Greens	4.66	2.77	0.544	0.000	1.134	0.564
News	4.19	3.71	0.862	0.051	0.858	0.420
Foreign Accounts	16.06	34.07	3.463	0.000	0.205	0.000

Notes: The table reports the percentage of users and the percentage of social bots (bot density) by cluster, after attributing each user to their most probable cluster. Odds ratios for a cluster k are calculated as $\frac{B_k}{H_k} \Big/ \frac{B-B_k}{H-H_k}$ where B_k is the number of positive cases in cluster k, H_k the number of negative cases, B the total count of positives, and H the total count of negatives.

The results presented in Table 4 reinforce the claim that a non-trivial level of foreign interference occurred during the campaign. Social media accounts belonging to the larger cluster previously identified as foreign in origin—accounts outlined in the rectangular shape of Figure 2, encompassing the special cases of the refugee campaign and "junk news" accounts—are significantly less likely to be geolocalized in Canada, and feature an exceptionally large concentration of bot accounts. The odds ratio reported in the penultimate column is well below the reference value of 1 for the Foreign Account clusters, a statistically significant result. Foreign activity may also help explain the over-representation of PPC supporters, relative to that party's vote share. To account for the fact that bot density may be a confounder—bots may be less likely to have an active geolocation, hence reducing the chances of observing a Canadian geotag—we calculated the odds ratios on the subset of users who posted geotagged content during the campaign. That is, we compute the odds of a Canadian geotag relative to a foreign geotag. Absent systematic forms of foreign activity, we would expect this ratio to be statistically indistinguishable from 1.

Combined with the previously mentioned evidence, the cluster of foreign accounts aligns with the first type of foreign strategy identified in the theoretical section, whereby bots are deployed

to introduce external content into the campaign. However, these bots was not as common as those that we infer to be national in origin. Overall, social bots behaved in a manner largely consistent with expectations derived from political communication theory. They were largely used to share party materials and to spread contents from trusted national news media, they were used more heavily by challengers, and with the exception of the PPC, their distribution was not disproportionate given the size of each partisan group in the Canadian Twittersphere.

5.3 Validation of the Methodology

A central property of the proposed methodology is that we can assign social media users to clusters. We used this procedure to infer statistical conclusions about the online communities in the previous subsection. Each user is assigned to the cluster with which they exhibit the highest semantic similarity (see the methodology section and online appendix for additional details). In plain terms, this means that each user is assigned a party affiliation based on whether they rely on a language and URL sharing behavior similar to the most vocal partisans.

We can validate the accuracy of this particular step of the methodology. Included among the most frequent users depicted on Figure 2 are the official accounts of 505 candidates during the 2019 election, for whom we know the true party affiliation. If the UMAP and document embeddings approach described therein successfully assigns users to a correct cluster, we should achieve high accuracy at predicting the party affiliation for these candidates. Table 5 is a classification matrix reporting results from this validation test. Overall, the methodology correctly predicts the affiliation of close to 91% of the candidates. The improvement from the model is substantial: using the mode of the most frequent category would generate a 28% accuracy score.

As an additional form of validation, we compare the observed distribution of users across our predicted partisan clusters to the actual distribution of the vote on election day. Table 6 displays the results, based on the same sample of frequent users discussed in the previous subsections. We calculated values in the first column using the proportion of users observed in the five partisan groups only; that is, excluding news media and suspected foreign accounts.

Table 5: Predicting the Party Affiliation of Known Candidates

	Predicted Label						
True Label	Conservative	Foreign	Greens	Liberal	NDP	News	PPC
Conservative	56	11	0	2	6	2	1
Greens	0	0	42	0	0	0	0
Liberal	1	10	3	115	9	0	0
NDP	0	0	0	0	100	0	0
PPC	0	1	0	0	0	0	146

Notes: Classification matrix of party affiliations for the 505 official candidate accounts featured among the most frequent users. The percentage of party affiliations correctly predicted by the model is 90.9% and the weighted F1 score is 0.928.

Table 6: Estimated Size of Partisan Groups vs Popular Vote

Party Group	Estimated Proportion (%)	Popular Vote (%)
Conservative	34.2	34.4
Liberal	32.8	33.1
NDP	13.8	15.9
PPC	13.3	1.6
Greens	5.8	6.5

Notes: Percentages of users classified in each of the five main partisan groups (excluding News and Foreign Accounts categories), compared to the popular vote in the 2019 federal election. The numbers in the "Popular Vote" column do not sum to 100 due to the exclusion of the Bloc Québécois (7.6%) and other minor parties (0.9%) not included in this analysis.

Table 6 suggests a close correspondence between the estimated size of partisan groups and percentages of the popular vote in the 2019 election. One party is notably absent (the Bloc Québécois), since our analysis focused on English speakers. The other exception is the People's Party of Canada (PPC), clearly overrepresented on Twitter compared to its share of vote. However, the estimated proportions for Conservatives, Liberals, NDP and Greens are not far away from the popular vote. This result gives additional credence to the methodology, in that predicted user affiliations align with realistic expectations.

6 Discussion

The threat of social media interference during political campaigns will likely remain a concern for democracies in the years to come. Our results on the 2019 Canadian election are somewhat reassuring in that regard. To be sure, we do find that a sizeable number of suspected social bots were active during the campaign, sharing over 2 million messages on the Twitter platform and representing approximately 13% of the total volume posted. On the other hand, the spread of disruptive content associated with false news websites represented a small fraction of all bot activity. We used an original methodology to detect foreign accounts, which revealed the existence of a user cluster comprising a disproportionately high density of social bots. These foreign accounts circulated targeted content—in particular, links to far-right websites, and specific campaigns revolving around climate change and refugees. The majority of social bots, however, were tightly aligned with national partisan groups. These Twitter bots shared content from official party websites and from national news outlets. Although the bot density was higher among Conservative supporters, it remains consistent with the proportion expected from the distribution of human users. Only one fringe party—the People's Party of Canada—enjoyed an overproportion of bot support.

These findings raise practical implications for the future of research on social media and elections. First, the main approach deployed in this paper, UMAP clustering in conjunction with

document embeddings, appears particularly efficient for analyzing the structure of social media communities during an election. The method captures substantively meaningful clusters, and will easily reveal attempts to introduce contents that differ semantically from the flow of messages posted by regular users. In particular, the method does not require privacy-invading techniques to investigate the existence of anomalies such as unusually low proportions of bots and national geotags. The various components of this methodology can help researchers, policymakers and regulatory agencies to monitor upcoming electoral campaigns. Second, the findings emphasize the importance of social bots as a strategic campaigning tool. Our evidence suggests that social bots were used by party supporters, and that many of these accounts were likely national in origin. This means that researchers and stakeholders need to account for the fact that national actors are also making use of this technology. In our view, a promising agenda for future research is the role of social bots as a component of the modern campaign toolkit.

References

Allcott, Hunt, and Matthew Gentzkow. 2017. "Social Media and Fake News in the 2016 Election." *Journal of Economic Perspectives* 31 (2): 211–36.

Auter, Zachary J, and Jeffrey A Fine. 2016. "Negative Campaigning in the Social Media Age: Attack Advertising on Facebook." *Political Behavior* 38 (4): 999–1020.

Barberá, Pablo. 2015. "Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data." *Political Analysis* 23 (1): 76–91.

Barberá, Pablo, and Gonzalo Rivero. 2015. "Understanding the Political Representativeness of Twitter Users." *Social Science Computer Review* 33 (6): 712–729.

Bastos, Marco T, and Dan Mercea. 2019. "The Brexit Botnet and User-Generated Hyperpartisan News." *Social Science Computer Review* 37 (1): 38–54.

Becht, Etienne, Leland McInnes, John Healy, Charles-Antoine Dutertre, Immanuel WH Kwok, Lai Guan Ng, Florent Ginhoux, and Evan W Newell. 2019. "Dimensionality Reduction for Visualizing Single-Cell Data Using UMAP." *Nature Biotechnology* 37 (1): 38.

Benoit, William L. 2007. *Communication in Political Campaigns.* New York: Peter Lang.

Bessi, Alessandro, and Emilio Ferrara. 2016. "Social Bots Distort the 2016 US Presidential Election Online Discussion." *First Monday* 21 (11-7).

Borah, Porismita. 2016. "Political Facebook Use: Campaign Strategies Used in 2008 and 2012 Presidential Elections." *Journal of Information Technology & Politics* 13 (4): 326–338.

Bradshaw, Samantha, and Philip N Howard. 2018. *Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation.* Oxford: The Computational Propaganda Project.

Bradshaw, Samantha, Philip N Howard, Bence Kollanyi, and Lisa-Maria Neudert. 2019. "Sourcing and Automation of Political News and Information Over Social Media in the United States, 2016-2018." *Political Communication:* 1–21.

Bright, Jonathan, Scott Hale, Bharath Ganesh, Andrew Bulovsky, Helen Margetts, and Phil Howard. 2019. "Does Campaigning on Social Media Make a Difference? Evidence From Candidate Use of Twitter During the 2015 and 2017 UK Elections." *Communication Research.*

Bubeck, Johannes, and Nikolay Marinov. 2019. *Rules and Allies: Foreign Election Interventions.* Cambridge: Cambridge University Press.

Chown, Marco, Jane Lytvynenko, and Craig Silverman. 2019. "A Buffalo Website Is Publishing 'False' Viral Stories About Justin Trudeau—And There's Nothing Canada Can Do About It." *The Toronto Star* Oct. 18.

Davis, Clayton Allen, Onur Varol, Emilio Ferrara, Alessandro Flammini, and Filippo Menczer. 2016. "Botornot: A System to Evaluate Social Bots." In *Proceedings of the 25th International Conference Companion on World Wide Web (WWW 15)*, 273–274.

Eady, Gregory, Jonathan Nagler, Andy Guess, Jan Zilinsky, and Joshua A Tucker. 2019. "How Many People Live in Political Bubbles on Social Media? Evidence From Linked Survey and Twitter Data." *Sage Open* 9 (1): 2158244019832705.

Ferrara, Emilio. 2017. "Disinformation and Social Bot Operations in the Run Up to the 2017 French Presidential Election." *First Monday* 22 (8).

Ferrara, Emilio, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. 2016. "The Rise of Social Bots." *Communications of the ACM* 59 (7): 96–104.

Gilbert, C.J., and Eric Hutto. 2014. "Vader: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text." In *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media (ICWSM-14),* 216–225.

Gorwa, Robert, and Douglas Guilbeault. 2020. "Unpacking the Social Media Bot: A Typology to Guide Research and Policy." *Policy & Internet* 12 (2): 225–248.

Grimmer, Justin, Solomon Messing, and Sean J Westwood. 2012. "How Words and Money Cultivate a Personal Vote: The Effect of Legislator Credit Claiming on Constituent Credit Allocation." *American Political Science Review* 106 (4): 703–719.

Guess, Andrew M, Brendan Nyhan, and Jason Reifler. 2020. "Exposure to Untrustworthy Websites in the 2016 US Election." *Nature Human Behaviour:* doi.org/10.1038/s41562-020-0833–x.

Howard, Philip N, and Bence Kollanyi. 2016. "Bots,# StrongerIn, And# Brexit: Computational Propaganda During the UK-EU Referendum." *Available at SSRN 2798311.*

Howard, Philip N, Samuel Woolley, and Ryan Calo. 2018. "Algorithms, Bots, and Political Communication in the US 2016 Election: The Challenge of Automated Political Communication for Election Law and Administration." *Journal of Information Technology & Politics* 15 (2): 81–93.

Kam, Cindy D, and Elizabeth J Zechmeister. 2013. "Name Recognition and Candidate Support." *American Journal of Political Science* 57 (4): 971–986.

Keller, Tobias R, and Ulrike Klinger. 2019. "Social Bots in Election Campaigns: Theoretical, Empirical, and Methodological Implications." *Political Communication* 36 (1): 171–189.

Lazer, David M. J., Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, et al. 2018. "The Science of Fake News." *Science* 359 (6380): 1094–1096.

Le, Quoc, and Tomas Mikolov. 2014. "Distributed Representations of Sentences and Documents." In *Proceedings of the 31st International Conference on Machine Learning.*

Levin, Dov H. 2016. "When the Great Power Gets a Vote: The Effects of Great Power Electoral Interventions on Election Results." *International Studies Quarterly* 60 (2): 189–202.

Maaten, Laurens van der, and Geoffrey Hinton. 2008. "Visualizing Data Using t-SNE." *Journal of Machine Learning Research* 9 (86): 2579–2605.

McInnes, Leland, John Healy, and James Melville. 2018. "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction." *arXiv Preprint arXiv:1802.03426.*

McKelvey, Fenwick, and Elizabeth Dubois. 2017. *Computational Propaganda in Canada: The Use of Political Bots.* Oxford: The Computational Propaganda Project.

Mueller, Robert S. 2019. *Report on the Investigation Into Russian Interference in the 2016 Presidential Election.* Washington, DC: US Department of Justice.

Pan, Shimei, and Tao Ding. 2019. "Social Media-Based User Embedding: A Literature Review." In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19),* 6318–6324.

Ratkiewicz, Jacob, Michael Conover, Mark R Meiss, Bruno Gonçalves, Alessandro Flammini, and Filippo Menczer. 2011. "Detecting and Tracking Political Abuse in Social Media," 11:297–304.

Schäfer, Fabian, Stefan Evert, and Philipp Heinrich. 2017. "Japan's 2014 General Election: Political Bots, Right-Wing Internet Activism, and Prime Minister Shinzō Abe's Hidden Nationalist Agenda." *Big Data* 5 (4): 294–309.

Shao, Chengcheng, Giovanni Luca Ciampaglia, Onur Varol, Kai-Cheng Yang, Alessandro Flammini, and Filippo Menczer. 2018. "The Spread of Low-Credibility Content by Social Bots." *Nature Communications* 9 (1): 1–9.

Silverman, Craig. 2016. "This Analysis Shows How Viral Fake Election News Stories Outperformed Real News on Facebook." [Online; accessed September 27, 2018], *BuzzFeed News* November 16.

SSCI. 2019. *Report of the Select Committee on Intelligence: Russian Active Measures Campaigns and Interference in the 2016 U.S. Election.* Vol. 2. Washington, DC: United States Senate.

Stella, Massimo, Emilio Ferrara, and Manlio De Domenico. 2018. "Bots Increase Exposure to Negative and Inflammatory Content in Online Social Systems." *Proceedings of the National Academy of Sciences* 115 (49): 12435–12440.

Stieglitz, Stefan, Florian Brachten, Björn Ross, and Anna-Katharina Jung. 2017. "Do Social Bots Dream of Electric Sheep? A Categorisation of Social Media Bot Accounts." In *ACIS 2017 Proceedings,* 89.

Temporão, Mickael, Corentin Vande Kerckhove, Clifton van der Linden, Yannick Dufresne, and Julien M Hendrickx. 2018. "Ideological Scaling of Social Media Users: A Dynamic Lexicon Approach." *Political Analysis* 26 (4): 457–473.

Twitter. 2018. *Update on Twitter's Review of the 2016 US Election* . https://blog.twitter.com/en_us/topics/company/2018/2016-election-update.html: Page Consulted on February 14, 2020.

Vosoughi, Soroush, Deb Roy, and Sinan Aral. 2018. "The Spread of True and False News Online." *Science* 359 (6380): 1146–1151.

Woolley, Samuel C. 2016. "Automating Power: Social Bot Interference in Global Politics." *First Monday* 21 (4).

Wu, Patrick Y, Walter R Mebane Jr, Logan Woods, Joseph Klaver, and Preston Due. 2019. "Partisan Associations of Twitter Users Based on Their Self-Descriptions and Word Embeddings." *2019 Annual Meeting of the American Political Science Association.*

Yang, Kai-Cheng, Onur Varol, Pik-Mai Hui, and Filippo Menczer. 2020. "Scalable and Generalizable Social Bot Detection Through Data Selection." In *Proceedings of the 14th International Conference on Web and Social Media (ICWSM-2020).*

Zajonc, Robert B. 2001. "Mere Exposure: A Gateway to the Subliminal." *Current Directions in Psychological Science* 10 (6): 224–228.

Supplementary Information

(Online Appendix)

Efficient Detection of Online Communities and Social Bot Activity During Electoral Campaigns

Additional Information on Data Collection

Our dataset originally contained 19,316,613 tweets. The final version used for this study was carefully pruned out to remove spam content unrelated to politics (particularly live streams of sports events) as well as false positives. The false positives are tweets that matched our stream filters but for the wrong reasons. This was caused largely by the acronym "CPC" as one of our filters (commonly used when referring to the Conservative Party of Canada), which has multiple meanings. We removed the off-topic instances of "CPC" by excluding tweets that did not contain at least one other token referring to Canada or the election. While these steps ultimately had little impact on our substantive conclusions, they further reduce the noise ratio in the dataset.

Table A1 provides detailed information about the three bot detection methods mentioned in the paper. The number of suspected social bot accounts detected by each method is indicated in the second column, and listed in decreasing order. We also indicate the number of accounts flagged by multiple methods. Note that the Botometer API was applied only to a subset of 52,374 accounts due to rate limits. As a result, the numbers in Table A1 should not be interpreted as a direct comparison of positive prediction rates. In general, the methods have an agreement approximating 90% using pairwise comparisons on equivalent samples. For example, the agreement between our implementation of the Botometer Lite model and the Twitter suspended accounts—that is, the percentage of accounts assigned consistently as either human or suspected bot—is 90.4%. As mentioned in the main text, the agreement between Botometer Lite and the Botometer API, using the common sub-sample of users, is 89.9%.

Our implementation of the Botometer Lite model relies on the same features mentioned in

Table A1: Breakdown of Bot Detection by Method

Detection Method	Social Bots
Custom Botometer Lite	74,541
Suspended	46,446
Custom Botometer Lite & Suspended	7,616
Botometer	3,331
Custom Botometer Lite & Botometer	652
Botometer & Suspended	21
All three methods	4
Total	132,611

Notes: The table reports the number or bots detected by each method mentioned in the main text. The Botometer API was only applied to the most frequent users, a subset of 52,374 accounts, due to rate limit restrictions.

the Yang et al. (2020) paper. The key predictive features are comprised of a binary indicator measuring the presence of a custom profile image, the length of the profile description, the tweet frequency, the followers growth rate, the favorites growth rate, the friends growth rate, the listed accounts growth rate (Twitter lists), the followers to friends ratio, and lexical attributes of the screen name, including the likelihood of character bigrams. There are two principal differences between our implementation and the original model discussed in Yang et al. (2020). First, we calculated the likelihood of character sequences in screen names using our own data collection, and we ignored casing since Twitter usernames are not case sensitive. Second, the number of training observations differs from the paper as we included only the account information which was retrievable from the datasets.[1] As mentioned in the main text, the Botometer Lite approach achieves a high rate of accuracy. We obtain an average F1 score of 0.97 using five-fold cross-validation. While we have not performed an exact replication of the authors' original model, the fact that we can easily achieve accuracy rates close to those reported in the original paper validates the usefulness of this approach to bot detection. Obviously, prospective users should remain wary that social bot practices evolve over time: a model that performs well on a dataset from 2019 may not achieve the same accuracy in the future if the actors operating bot accounts

1. The public training datasets for bot detection are available at https://botometer.osome.iu.edu/bot-repository/datasets.html. Our selection of training datasets is based on Yang et al. (2020), but we made sure to include both political datasets ('political-bots-2019' and 'midterm-2018') since they are domain relevant.

adapt their strategy to avoid detection.

The Impact of Bots on Leader Sentiment

The main text provides a descriptive account of the trends in leader sentiment during the campaign. The focus on leaders is justified by a large body of political science literature suggesting that party leaders represent the dominant actors of Canadian elections, a trend also observed across many parliamentary democracies (Johnston 2002; Bakvis and Wolinetz 2005; Aarts, Blais, and Schmitt 2011; Pruysers and Cross 2016; Small 2016). We proceeded by flagging tweets that mention each of the two major candidates, Justin Trudeau and Andrew Scheer (excluding ambiguous cases of tweets that mention both of them). Next, we computed sentiment scores using the Vader library for Python, which has a lexicon developed specifically for social media analysis, and includes conveniences such as valence shifting (accounting for negative statements) and adjustments in presence of amplifiers (Gilbert and Hutto 2014). The sentiment scores range from −1 to 1 and were computed on tweets in English language only. This part of the analysis focuses on the 32% of messages containing original content, excluding retweets but including the original text posted using the "quote" functionality.

In this section, we supplement the main text by examining whether the content posted by social bots had a significant influence on public sentiment toward leaders. Table A2 reports the outcome of Granger causality tests for each leader, considering both possible directions. A statistically significant result represents evidence of temporal causality; in other words, that changes in the sentiment series for one type of users have a significant influence on future values in sentiment for the other type of users. We report results with the number of lags selected automatically using the Akaike information criterion (AIC), as well as results based on 12 and 24 lags (respectively a half and a full daily cycle). Regardless of the lag length selected, the tests indicate that bots respond to human users, rather than the reverse. Combined with our previous observations, this result suggests that bots may be used strategically to amplify the patterns already existing among social media users, as opposed to setting new trends.

Table A2: Granger Causality Tests

Series	Direction	Lags	χ^2	p-value
Trudeau	Human -> Bot	9 (Auto)	94.668	0.000
Trudeau	Human -> Bot	12	97.844	0.000
Trudeau	Human -> Bot	24	93.262	0.000
Trudeau	Bot -> Human	9 (Auto)	13.660	0.135
Trudeau	Bot -> Human	12	14.902	0.247
Trudeau	Bot -> Human	24	30.225	0.177
Scheer	Human -> Bot	5 (Auto)	21.720	0.001
Scheer	Human -> Bot	12	31.184	0.002
Scheer	Human -> Bot	24	47.212	0.003
Scheer	Bot -> Human	5 (Auto)	5.645	0.342
Scheer	Bot -> Human	12	13.542	0.331
Scheer	Bot -> Human	24	29.376	0.206

Notes: Chi-square test of the null of Granger non-causality. Automatic lag selection (Auto) is based on AIC.

We tested the robustness of this conclusion in two different ways. First, we considered bots detected using our machine learning algorithms only (that is, excluding suspended accounts). For simplicity, we do not report the full results, but this alternative definition of the bot variable leads to a similar conclusion. Second, we replicated the analysis after including the content of retweeted posts. Once again, we find that the conclusions reported in Table A2 remain unchanged. Only when considering daily time series in conjunction with lag lengths covering week-long periods do we observe a rejection of the null that bots do not Granger-cause public sentiment. Simply put, we do not find robust evidence indicating that bots have inflected online sentiment toward party leaders during the campaign.

Humans Sharing Social Bot Content

To further support the analysis presented in the main text, we examined whether human users retweeted the content posted by social bots, in line with the findings from Shao et al. (2018). Table A3 shows a cross-tabulation of "tweeters" (the type of the user posting a comment) by "retweeters" (the type of users retweeting), calculated on ∼11.8 million retweets in our data collection.[2]

2. We performed this analysis on a dataset restricted to original tweeters who also appear at least once in our dataset as retweeters, so that we can assign social bot labels on both tweeters and retweeters. This subset represents

About six percent of all retweets were human users sharing messages originally posted by suspected bots, representing approximately 735,000 cases in total.[3] While a fraction of the content introduced by social bots definitely reached members of the general public, the cross-tabulation indicates that human users were in fact reluctant to retweet social bots. The Table reports percentages tabulated within the categories of tweeters. If human users did not distinguish between users posting content (that is, under the null hypothesis), these percentages should be equal across rows. Instead, human users were more likely to share messages posted by human users than those posted by social bots (a difference of 7.2 percentage points).

Table A3: Did Humans Retweet Bots?

		Retweeter		
		Bot	Human	Total
Tweeter	Bot	170,974 (18.9%)	734,652 (81.1%)	905,626 (100%)
	Human	1,276,514 (11.7%)	9,652,358 (88.3%)	10,928,872 (100%)
				$\chi^2 = 40373.2; p < 0.001$

The previous finding, however, may be affected by the low-quality content shared by subgroups of social bots, which humans can more easily filter and ignore. Consequently, we look at specific URL links that were most frequently shared by social bots during the campaign and retain the top 30 political links (Figure A1). For each URL, we report the percentage of instances in which a human user retweeted a link originally shared by a social bot, over the total number of times that URL was shared in the data collection, along with 95% confidence intervals (compare against the 6% benchmark evoked earlier for the rate at which humans generally retweet messages posted by bots). Note that these proportions are also affected by how many human users decide to post the same URL: popular links introduced by human users on the site will decrease the probability that another human user retweets the same link from a bot. The numbers printed on the figure indicate, for each URL, the total number of occurrences where a human retweeted

91.4% of the retweets from the full data collection.

3. Note that "tweeters" indicate the user originally posting a message. In the language of the platform, a user can retweet a retweeted message, which happens frequently. In that case, the tweeter is still the user at the origin of the chain.

a bot who originally shared the link. Both measures give an indication of how much the external content introduced by bots circulated on the platform during the campaign.

Figure A1: Top URL Links Shared by Social Bots and Retweeted by Humans

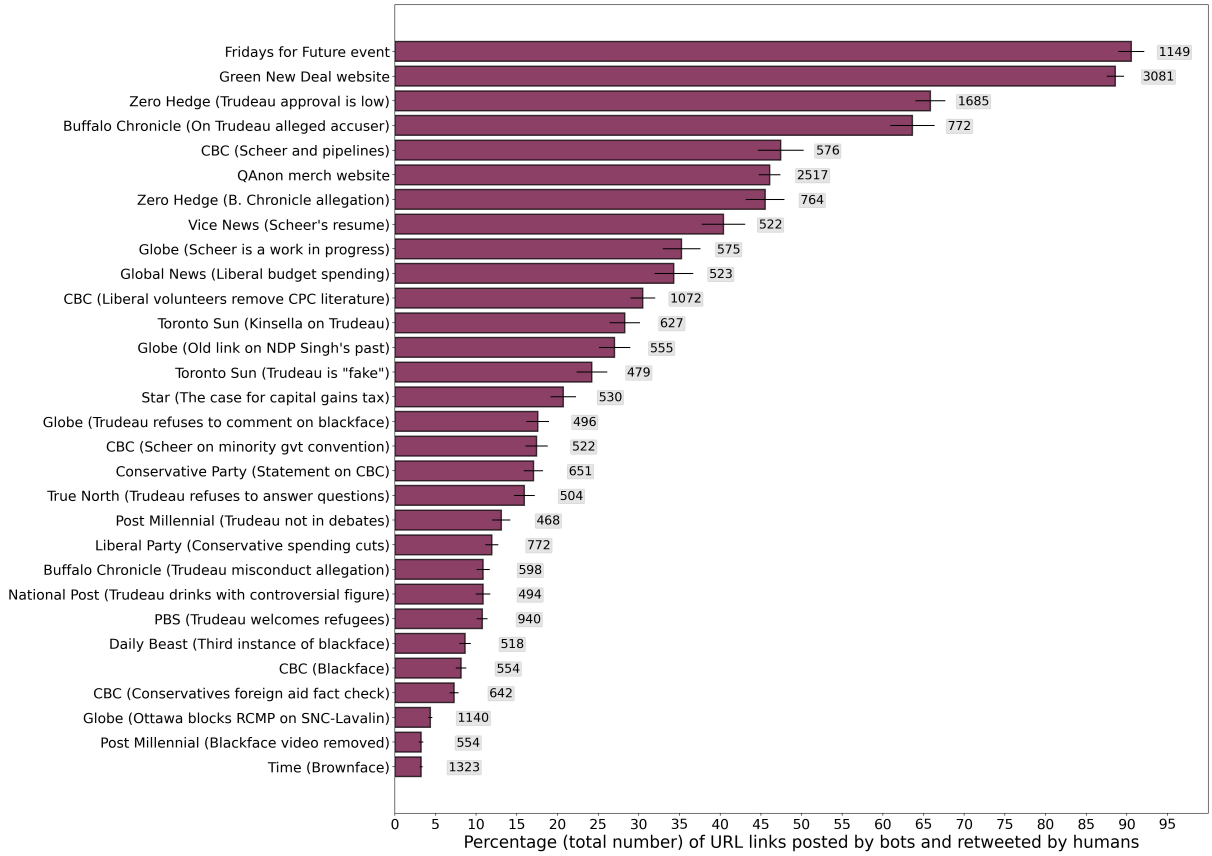


Figure A1 represents evidence that some social bots have been successful at directing the attention of regular users to specific news items during the 2019 electoral campaign. The most successful instances of that nature, however, were not directly related to the election. They were aimed instead at mobilizing users toward environment-related causes. The top two links in Figure A1 concern protest activities related to the Fridays for Future movement, and news associated with the Green New Deal website.

A few stories from false news websites (Buffalo Chronicle, News Punch) feature among the top influential URLs. None of the URLs preferred by social bots, however, reached a very large proportion of users. The most talked-about false news story published by the Buffalo Chronicle (an allegation of misconduct involving Justin Trudeau published on October 7) was retweeted by

human users from social bot accounts, although fewer than 600 times. To put things in perspective, the same link was posted by human users over 8,000 times. More importantly perhaps, this episode had no discernible effect on public sentiment toward Justin Trudeau (see Figure 1, from the main text). Moreover, the stories that arguably had a significant impact on the turn of events during the campaign (e.g. the release of Trudeau's controversial photographs by Time Magazine) rank among the least influential efforts made by social bots. These pivotal stories were shared in large part by general members of the public. Since they originate from established media organizations, they correspond to what Goel et al. (2016) refer to as broadcast diffusion, as opposed to viral diffusion that would be engineered from the ground up by social bots.

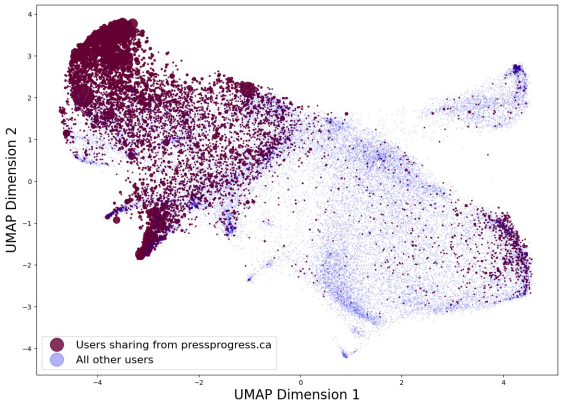
Additional Assessment of UMAP Visualizations

Figure A2 reprises the same mapping used in Figure 2 of the main text, this time after highlighting the users who shared content from substantively relevant web domains during the campaign. The size of the dark circles reflects the number of times each user referred to URLs from the associated domain. The top panels help to illustrate that the clustering achieved with our methodology can be interpreted in terms of political ideology. Users sharing from the openly left-leaning PressProgress news website are consistently located in the clusters associated with the NDP, Greens and Liberals. In contrast, users sharing from the openly right-leaning Post Millennial are more consistently found in the Conservative and PPC clusters, also right-leaning political parties. Both patterns are consistent with natural expectations.

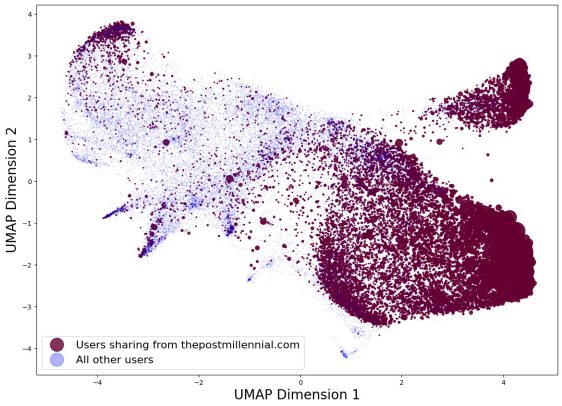
Finally, the bottom panel of Figure A2 suggests that articles from the Buffalo Chronicle were not only shared by social bots associated with the cluster of foreign accounts, but also—and largely—by social bots associated with Conservative and PPC supporters. We conclude that national actors contributed to sharing content from that outlet.

Figure A2: Domain Sharing Preferences by Cluster

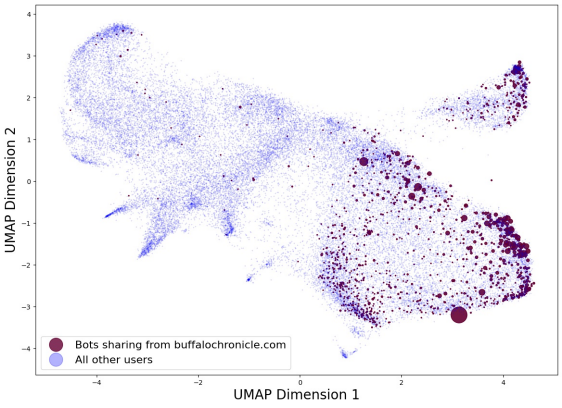
(a) PressProgress (Left-Leaning)



(b) The Post Millennial (Right-Leaning)



(c) Buffalo Chronicle (Bots Only)



Assigning Users to Substantive Clusters

Our approach to assigning users in Table 4 of the main text takes advantage of the user embedding model at the basis of our methodological approach. We relied on the UMAP clusters to identify meaningful groupings, and used this information to retrieve the ten users who tweeted the most among those located in the center of each cluster. For partisan clusters, we also include the party leaders. We treat the average embeddings of these representative supporters as anchor vectors. Next, we calculate the cosine similarity of every remaining user's embedding with each anchor vector, in the original embedding space. Users are assigned to the cluster with which they have the highest cosine similarity. The analysis is based on frequent users (those who tweeted 50 times or more during the campaign, the same subset used to produce Figure 2), which ensures that embeddings are fitted using a sufficient sample size for each user.

As emphasized in the methods section of the text, this approach builds on the idea that document and word embedding models have the property to map similar entities at proximity to each other in a vector space. Users tweeting similar content will help to predict the occurrence of the same hashtags and the same words. As a result, after identifying reference points of interest, we can classify users based on the textual attributes they have in common. Note that in our implementation, we also append the domain names of the URLs shared by each user to the textual content of their tweets, which we find to improve the substantive relevance of this methodology. The intuition is similar to models relying on the frequency of domain shares to infer user ideology (see e.g. Eady et al. 2019).

In addition to the validation tests presented in the main text, we manually examined the face validity of this user attribution procedure. In particular, we inspected the most active accounts classified in the foreign cluster to validate their geographical origin. While the foreign account cluster is a mixed bag in terms of substantive focus and interests, we did observe a high occurrence of accounts that are ostensibly foreign. In fact, many of these accounts explicitly mentioned a foreign location in their profile description.

In the main text, we illustrated two characteristic patterns among suspected foreign accounts:

the tendency to post environment-related and far-right content. The sub-cluster of users tweeting about the environment contains accounts seemingly part of an international network focusing on climate change activism and environment-related themes. Our manual inspection of these accounts reveals that many users probably rely on applications that automatically retweet environment-related stories originating from the same sources. These users were often flagged as bots by our predictive models. The origin of the retweeted environment stories can be traced back to a handful of users associated with Green New Deal Canada, Greenpeace, and Fridays for Future. The content of these environmental tweets typically has a global appeal, touching on stories from multiple countries, yet they included one of the principal Canadian political hashtags: #cdnpoli. While some users from that network can be traced back to Canada, we did identify many accounts ostensibly foreign in origin, in particular users mentioning locations in Europe and the United States.

A second sub-cluster of interest contains users who frequently posted content associated with far-right news websites such as News Punch, Zero Hedge and trump-train.com. A similar inspection of these accounts leaves little doubt that many were foreign in origin. The content tweeted often mixes Canadian and US politics, and the phrasing sometimes openly reveal that the opinions expressed come from outsiders. Table A4 gives three concrete examples of tweets from that category. A caveat to our approach is that we would need further investigation before determining whether these accounts are actually American based, or part of a campaign of interference targeting US politics that eventually spilled over to the Canadian Twittersphere.

Table A4: Example Tweets from Most Active Foreign Accounts

God, Trudeau is as bad as Barack Obama Sin Laden
@Patriots, tommorow is a YUUUUUGGGE day for our cousins to the north. Let's pray Canada still has enough people with common sense. Nationalism over Globalism. @AndrewScheer for Prime Minister
OOPS! Looks like someone found the video of soy boy Trudeau in Blackface! Typical Liberal! Do as I say, not as I do! Isn't the election coming soon? I hope he gets SLAUGHTERED! RT to remind liberals that their favorite Canadian SJW is a HUGE RACIST!

References

Aarts, Kees, André Blais, and Hermann Schmitt. 2011. *Political Leaders and Democratic Elections.* Oxford: Oxford University Press.

Bakvis, Herman, and Steven B Wolinetz. 2005. "Canada: Executive Dominance and Presidentialization." In *The Presidentialization of Politics: A Comparative Study of Modern Democracies,* edited by Thomas Poguntke and Paul Webb, 199–220. Oxford: Oxford University Press.

Eady, Gregory, Jonathan Nagler, Andy Guess, Jan Zilinsky, and Joshua A Tucker. 2019. "How Many People Live in Political Bubbles on Social Media? Evidence From Linked Survey and Twitter Data." *Sage Open* 9 (1): 2158244019832705.

Gilbert, C.J., and Eric Hutto. 2014. "Vader: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text." In *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media (ICWSM-14),* 216–225.

Goel, Sharad, Ashton Anderson, Jake Hofman, and Duncan J Watts. 2016. "The Structural Virality of Online Diffusion." *Management Science* 62 (1): 180–196.

Johnston, Richard. 2002. "Prime Ministerial Contenders in Canada." In *Leaders' personalities and the outcomes of democratic elections,* edited by Anthony King, 158–183. Oxford: Oxford University Press.

Pruysers, Scott, and William Cross. 2016. ""Negative" Personalization: Party Leaders and Party Strategy." *Canadian Journal of Political Science* 49 (3): 539–558.

Shao, Chengcheng, Giovanni Luca Ciampaglia, Onur Varol, Kai-Cheng Yang, Alessandro Flammini, and Filippo Menczer. 2018. "The Spread of Low-Credibility Content by Social Bots." *Nature Communications* 9 (1): 1–9.

Small, Tamara A. 2016. "Parties, Leaders, and Online Personalization: Twitter in Canadian Electoral Politics." In *Twitter and elections around the world: Campaigning in 140 characters or less,* edited by Richard Davis, Christina Holtz-Bacha, and Marion R Just. London: Taylor & Francis.

Yang, Kai-Cheng, Onur Varol, Pik-Mai Hui, and Filippo Menczer. 2020. "Scalable and Generalizable Social Bot Detection Through Data Selection." In *Proceedings of the 14th International Conference on Web and Social Media (ICWSM-2020).*