Noname manuscript No.
(will be inserted by the editor)

Lying as a New Defensive Misinformation Strategy*

**Daniele Bellutta · Catherine King ·
Kathleen M. Carley**

Received: date / Accepted: date

Abstract This paper presents a new social media phenomenon that sees users lying about their deceptive motivations by either dishonestly claiming that they are not bots or by asserting that real news is actually fake news. We analyze these strategies by examining the use of the #FakeNews and #NotABot groups of hashtags in Twitter data collected on the 2019 Canadian federal elections. Our findings show that the #FakeNews hashtag was most likely to be connected to an established news source rather than an actual fake news site and that users of the #NotABot hashtag were no more likely to be human than other users in our data set. This phenomenon of lying about lying has therefore been used to discredit well-known news organizations and amplify political misinformation. This new defensive strategy used by online influence campaigns shows how they continue to evolve to manipulate social media users even as people have become more aware of the dangers of online misinformation. As these campaigns learn to adapt to avoid detection, it remains critical to characterize these changes in their campaigns and strategies in the hope of countering their influence on democratic societies.

Keywords Social media analytics · social networks · disinformation · elections

1 Introduction

The 2016 U.S. presidential election led to increased interest and concern over the safety and security of democratic election systems around the world. Because Russia notably interfered in that election to increase division and sup-

[*] This paper is an extension of the conference paper, “Lying about Lying: A Case Study of the 2019 Canadian Elections” [13].

Daniele Bellutta, Catherine King, and Kathleen M. Carley
E-mail: [dbellutt, cking2, kathleen.carley]@cs.cmu.edu
Address: Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh PA 15213

port then-candidate Donald Trump, other democratic nations have also grown concerned about potential foreign interference [2]. More broadly, the spread of misinformation and disinformation on social media can have many negative societal impacts, including political polarization, the undermining of trust in government institutions and the mainstream media, and growing contempt for members of the opposite political party [2].

Because of the many negative consequences of misinformation and influence campaigns, researchers have conducted several studies over the last few years to analyze their impacts [2][10]. Other research has focused on improving bot detection [3] or improving the automatic detection of potentially fake or misleading news stories [9][14]. As detection has continued to improve over time, we have seen malicious actors adapt their strategies to evade detection.

During the 2019 Canadian federal elections, several journalists covered various misinformation campaigns, including one directly targeting Prime Minister Justin Trudeau that was likely driven by bot accounts [15]. Caroline Orr, a reporter and research analyst at the National Observer, observed that the #NotABot hashtag was likely being used in an inauthentic manner to amplify the #TrudeauMustGo hashtag. There was a large spike in the usage of both the #NotABot and the #TrudeauMustGo hashtags in July of 2019 [15]. Recently, we have also seen the term "fake news" being used as a way to discredit real news stories and disparage political adversaries rather than being used for its original purpose as a way to expose false or deceptive news stories. Both regular users and malicious users have taken over the "fake news" term in this way [18].

As both researchers and the public become more aware of bots and information maneuvers, claiming to not be a bot appears to be a new defensive strategy that bots are using to try to convince others that they are authentic. In addition, claiming something is fake when it is accurate is also a way to try to convince others of your misinformation or disinformation. Therefore, these observations have raised concerns among journalists and government officials about a new type of disingenuous misinformation that could potentially undermine their elections in the future. Because there is now an awareness among the general public that bots exist and can seek to manipulate, bots are attempting to circumvent those concerns by lying.

This paper focuses on lying as a new type of disinformation. Both the #NotABot and #FakeNews hashtags have been observed to be used in false manners. In this extension of our conference paper, "Lying about Lying: A Case Study of the 2019 Canadian Elections" [13], we describe the types of users and targets of these hashtags in the context of the 2019 Canadian federal election.

2 Related Work

The emerging and interdisciplinary field of social cybersecurity focuses on describing and understanding how the online information environment can impact society, culture, and politics. Specifically, it looks at how information and network maneuvers can directly impact human behavior and opinions. Current researchers in this field have focused on how to best protect an open and free Internet given the existence of misinformation, hate speech, cyber-bullying, and other types of information attacks. These researchers have analyzed the potential impact of various misinformation campaigns that attacked several democratic nations in the last few years [2] [10].

It is challenging to quantify the impact that these misinformation campaigns may have had. Bail *et al.* found that the individuals who were most likely to interact with Russian bot accounts were already polarized before their interactions [2]. Similarly, other researchers found that while older and more conservative individuals were more likely to engage with false news and bot accounts than others, the total engagement with these stories was relatively low and highly concentrated [10]. However, it is still unclear how these misinformation campaigns may be impacting other types of offline behaviors, like voting, protesting, or even violence. In addition to analyzing the impact of these campaigns, it is important to be able to accurately detect misinformation campaigns as they are happening. Specifically, the line between fake news and satire has become more difficult to navigate; with some fake news sites claiming that they are satire, several researchers have been working on how to better differentiate between the two [9][14].

This new field of social cybersecurity affects the national security and democratic well-being of our country as well as other democratic nations around the world [5][7][8]. Influence campaigns have been shown to use many types of actions to manipulate the structure of social networks by either connecting or breaking up groups. In addition to manipulating the network structure, these campaigns also manipulate the information environment by spreading falsehoods, polarizing content, and amplifying certain individuals or groups. Bots are a crucial element to these information campaigns, as they are often used as a way to amplify these false messages to as large of an audience as possible [8]. Many researchers in this field have worked on improving bot detection algorithms [3] and detection of "deep fakes" (doctored videos or images) [11][21]. Because adversaries are continuing to adapt, it is essential to keep up.

As a result, after the 2016 election in the United States, Canada developed one of the most detailed plans among democratic nations to combat foreign interference in their 2019 elections [16]. Canadian journalists and researchers were actively monitoring for any evidence of foreign meddling. One of the most obvious influence campaigns came in July before the October 2019 elections. Many Twitter users started falsely claiming that they were #NotABot when they likely were bots, and they used this hashtag when amplifying potential misinformation or a specific political agenda. There was a large peak in the usage of both the #NotABot and the #TrudeauMustGo hashtags that summer [15]. Other researchers found that bots were involved in attempting to prolong the Trudeau blackface controversy, likely to hurt his re-election chances [17]. Ultimately, researchers found that while the bots were successful at spreading certain messages, they did not appear to have a long-term meaningful impact on public sentiment towards the political figures from the major parties [17]. Additionally, government researchers stated after the election that while they detected misinformation and disinformation and recognize they need to do better in the future, they do not believe it compromised the election [22].

During the election cycle, several controversial and ultimately inaccurate stories about Canada were spread widely on social media, likely by right-wing actors in an attempt to damage the reputation of Prime Minister Trudeau [19] [20]. The Daily Star, a British tabloid, cited unnamed sources and claimed that a child murderer would be sent from Britain to Canada. While the British government issued a statement claiming that this story was false, the news and social media coverage of this tabloid article had already caused this inaccurate story to spread widely. The story even led to a response on Twitter by the Canadian Conservative Party leader, Andrew Scheer, who claimed he found the news disturbing and would never allow it to happen if he became Prime Minister [19]. In another case, Canadian officials also had to deny an article published in a Lebanon-based newspaper that claimed that Canada would be taking in 100,000 Palestinian refugees [20]. These events triggered a heightened awareness of influence campaigns among researchers and government officials, who were trying to safeguard the election and the conversation surrounding the democratic process.

As is the case with many algorithmic improvements in other fields of computer science, we have seen our information adversaries adapt and improve their strategies as a way to continue avoiding detection [21]. Many malicious actors have been using the #NotABot and #FakeNews hashtags to lie about their identity or the accuracy of a story. In this paper, we investigate these new defensive strategies employed by actors wishing to spread misinformation.

3 Methods

3.1 Data Collection

The Canadian federal election was held on 21 October 2019. We collected Twitter data from 20 July 2019, around when the campaigns began, to 6 November 2019, a couple of weeks after the election. A data set consisting of 16,784,400 tweets, 1,303,761 users, and 137,419 distinct hashtags was collected by streaming tweets matching a list of search terms that was augmented over time. Table 1 shows the final list of terms. Though these data are not necessarily representative of all Twitter activity surrounding the 2019 Canadian elections, the collection terms were chosen to cover a wide variety of political topics.

We identified two groups of interesting hashtags to study. The first group comprises hashtags used to call out what a user sees as misinformation. More specifically, these hashtags contain both the words “fake” and “news”. The second group comprises hashtags used to claim that the posting account is

2019 Canadian Election Twitter Collection	
Category	Terms
General Election	#Election2019, #elxn43, #cdnpoli, #ItsOurVote, #NotAbot, cccr2019
Liberal Party/Justin Trudeau	#lpc, TeamTrudeau, trudeau, #chooseforward
Conservative Party/Andrew Scheer	#cpc, scheer, #TrudeauMustGo, #LiberalsMustGo, #ButtsMustGo, #LavScam
Regional Politics/Other Parties	#ndp, #gpc, dougford, fordcutshurt, fordisfailing, #onpoli, BlocQuebecois, #blocqc, #TransMountain, #NoTMX, #TMX

Table 1: The final list of terms used to collect the Twitter data on the 2019 Canadian election.

not run by a bot, which are hashtags that contain both “not” and “bot”. The most popular hashtags in both of these groups are listed in Table 2.

Fake-News Hashtags		Not-A-Bot Hashtags	
Hashtag	Tweets	Hashtag	Tweets
#fakenews	9,741	#notabot	45,605
#fakenewsmedia	3,287	#iamnotabot	921
#fakenewscbc	70	#imnotabot	142
#fakenewsandy	62	#teamnotabot	62
#cbcisfakenews	59	#stillnotabot	53

Table 2: The most-used hashtags in the fake-news and not-a-bot groups.

3.2 Bot Detection

After collecting the data, our first goal was to identify the malicious bots present in the data set. Using the Tier-1 BotHunter algorithm developed by Beskow and Carley [3], we attained probability scores that indicated the likelihood that each user in our data was a bot. BotHunter is a random forest regressor that was trained on labeled data assembled from forensic analyses of events with bot activity that was widely reported, with a specific focus on the 2017 attack on the Atlantic Council Digital Forensic Labs. The machine learning model makes use of features stemming from both the account information and the specific tweets it is given. These include user attributes such as account age and screen name length, network attributes such as the number of followers or friends, and tweet attributes such as tweet content and timing. Beskow and Carley also developed Tier-2 and Tier-3 versions of BotHunter [3] that added user timeline data and the timelines of a user’s friends, respectively, but those algorithms take more time to run and require massive data collection efforts for a data set as large as ours. We therefore decided to use the Tier-1 version of the algorithm, which has been shown to not lose a significant amount of accuracy when compared to its Tier-2 and Tier-3 counterparts [4].

Given that the output of BotHunter is a probability value and not a classification, any threshold can be applied to separate bot accounts from non-bot accounts. For this study, we used multiple thresholds ranging from 0.6 to 0.8. Lower thresholds would have included more accounts as bots, but they might also have included some accounts that were actually operated by humans. Higher thresholds would have yielded a more conservative set of bots, but they would likely not have captured all of the bots present in the data.

An important concern that arises when detecting bots in Twitter data is that many accounts have bot-like behavior but are not malicious. For example, accounts run by news agencies often simply tweet links to their latest news articles. Since our interests were focused on malicious bots, we used the account identity classification system developed by Huang and Carley [12] to filter our sets of bot accounts so that they only included accounts that did not belong to legitimate organizations. This identity classifier uses a hierarchical self-attention neural network to determine whether a Twitter account falls into one of seven classes: government official, company, celebrity, sports, news media, news reporter, and normal. When using the various BotHunter thresholds to identify bots, we discarded any detected bots that the Huang and Carley classifier did not deem to be normal accounts.

3.3 #FakeNews Targets

Since the fake-news group of hashtags was used to call out supposed misinformation, it is important to understand which entities were being targeted using these hashtags. For each tweet mentioning a fake-news hashtag, we developed a set of possible targets for that tweet. This set was composed of (a) any users mentioned in the tweet, (b) the Web sites linked to in the tweet, (c) the author of the tweet being replied to if the given tweet was a reply, and (d) the targets of any fake-news hashtags that were geared towards specific entities. The links included in tweets were un-shortened and then manually tied to specific targets by inspecting their domains. Certain hashtags also stood out as targeting specific entities; for example, the hashtag "#fakenewscbc" was clearly targeting the Canadian Broadcasting Corporation. These hashtag-specific targets were identified manually and included in the sets of targets for the tweets mentioning those hashtags.

From these sets of targets for each tweet, we discarded entities that were not likely to be targets of fake-news accusations. More specifically, we only considered an entity to be a potential target if it belonged to one of the following categories: politicians, political organizations, entities claiming to be news agencies, and individuals claiming to be reporters. Importantly, we did not consider only mainstream news organizations and their reporters; any account or Web site claiming to be a journalist or news organization was included as a potential target. Also, we manually categorized the potential targets into four groups: news agencies, journalists, politicians, and other entities. For these categories, we again did not discriminate between well-known news organizations or reporters and other entities claiming to be journalistic.

It should be noted that this method of identifying the targets of fake-news tweets has its disadvantages. For example, in situations in which a user replies to a person of the same political leaning and uses a fake-news hashtag to call out an entity not mentioned in the Twitter conversation [18], our method would incorrectly identify the target of the reply. Though this presents a clear limitation of this scheme for identifying the targets of fake-news accusations, our method is nevertheless advantageous in that it does not require further data collection and runs quickly on large data sets.

4 Results

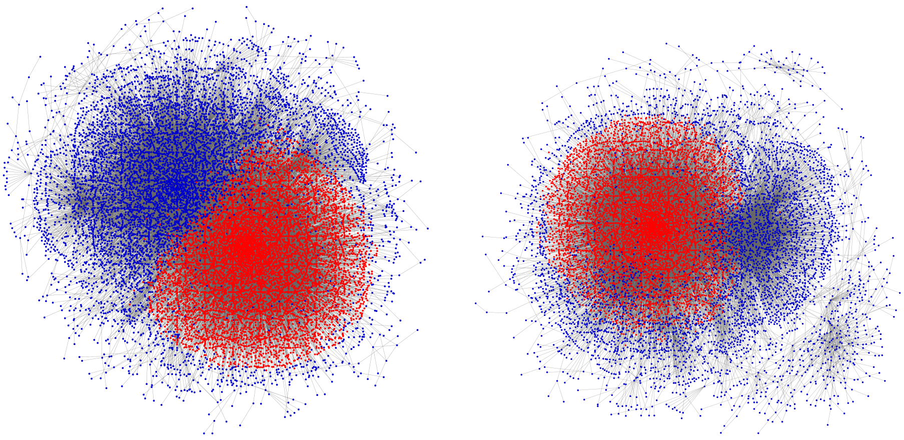
4.1 Reciprocal Communication Networks

We examined the reciprocal communication networks of Twitter users who used either fake-news or not-a-bot hashtags. The reciprocal communication networks for these two groups of hashtags show a connection between two hashtag users if both users have communicated with each other in some way. For example, the first user may have mentioned the second user, and the second user may have retweeted the first user. Visually, we noticed that both networks appeared to be divided into two large groups of users. This is shown in Figure 1a for the fake-news hashtag users and in Figure 1b for the not-a-bot hashtag users.

To better understand the reason for the existence of these two clusters, we separated the users into their respective groups using CONCOR [6], which groups nodes based on the similarity of their connections. We then counted the number of times any hashtag was used by the accounts in each group and normalized the count by the group's total number of hashtag uses. We were particularly interested in hashtags with a clear partisan stance (such as "#scheerlies" and "#trudeaumustgo"), which turned out to have different levels of usage across the two clusters. Table 3 shows the usage frequencies of an equal number of popular liberal-leaning and conservative-leaning hashtags in the two clusters present in each of the reciprocal communication networks. All of these hashtags show that in each network, one group consistently uses liberal-leaning hashtags more than the other group, and the other group uses conservative-leaning hashtags more than the first group. The reciprocal communication networks, therefore, appear to be divided on a partisan basis.

4.2 Co-Occurring Hashtags

To understand which discussion topics were associated with the use of fake-news or not-a-bot hashtags, we studied the hashtags that appeared alongside the fake-news or not-a-bot hashtags. Figure 2a shows the ten hashtags that



(a) Users of fake-news hashtags.

(b) Users of not-a-bot hashtags.

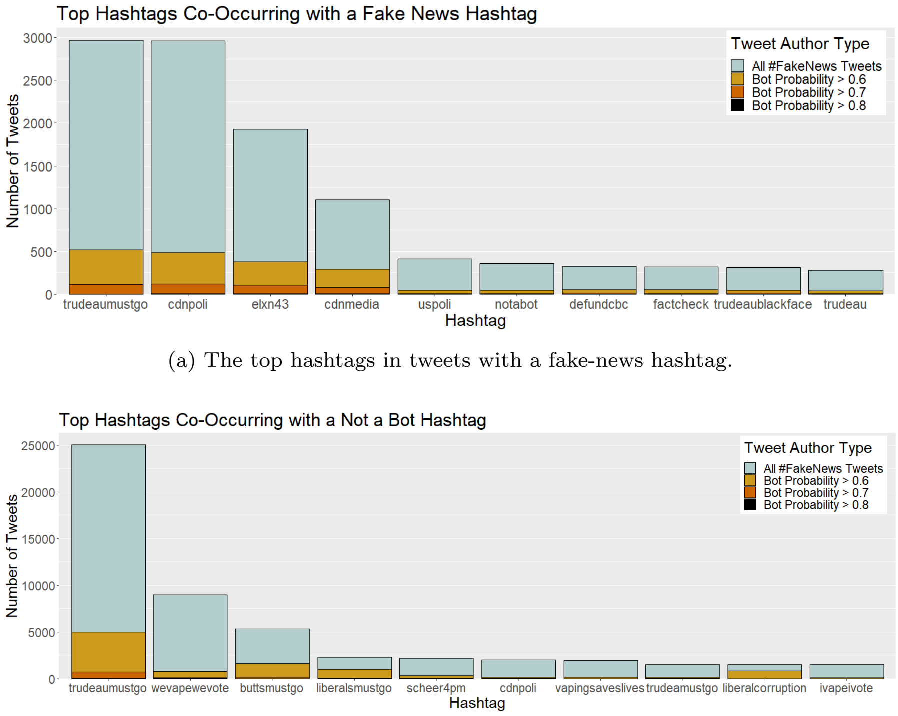
Fig. 1: The reciprocal communication networks for the users of fake-news hashtags and not-a-bot hashtags. Each network has been divided into two clusters using CONCOR, with the two clusters shown in red and blue.

		#FakeNews Users		#NotABot Users	
		Blue (%)	Red (%)	Blue (%)	Red (%)
Conservative	#trudeaumustgo	0.81	20.93	1.52	21.67
	#scheer4pm	0.03	1.86	0.05	1.93
	#trudeauworstpm	0.05	1.34	0.08	1.34
	#liberalsmustgo	0.02	1.19	0.03	1.25
	#trudeaumustresign	0.03	1.17	0.07	1.19
Liberal	#istandwithtrudeau	0.64	0.08	0.65	0.11
	#teamtrudeau	0.61	0.27	0.66	0.29
	#scheerlies	0.45	0.02	0.48	0.02
	#scheerdisaster	0.42	0.02	0.45	0.02
	#neverscheer	0.40	0.02	0.35	0.03

Table 3: The usage frequency of popular conservative-leaning and liberal-leaning hashtags in the CONCOR clusters for the reciprocal communication networks of fake-news hashtag users and not-a-bot hashtag users. Usage frequency was calculated as the number of tweets using a hashtag divided by the total number of hashtag uses in that CONCOR group.

most commonly co-occur with any of the fake news hashtags, whereas Figure 2b shows the hashtags co-occurring with the not-a-bot hashtags.

The hashtag that co-occurred most often with a fake-news hashtag was #cdnpoli, which is used to discuss Canadian politics in general. Interestingly, the hashtag #NotABot appears in the ten most co-occurring hashtags, indicating that the communities of people who use fake-news and not-a-bot hashtags were linked. Figure 2a also shows that fake-news hashtags were used in conjunction with the hashtags #defundcbc and #TrudeauBlackFace. Using a bot score threshold of 0.6, the hashtag with the highest fraction of tweets



(a) The top hashtags in tweets with a fake-news hashtag.

(b) The top hashtags in tweets with a not-a-bot hashtag.

Fig. 2: The hashtags used most often in tweets containing fake-news or not-a-bot hashtags. Different colors show the portions of tweets coming from bots detected at various BotHunter thresholds.

coming from bots was #cdnmedia, with 26% of its usages coming from bots. The next highest hashtags in terms of bot usage at the 0.6 level are #elxn43 (20% bot tweets) and #TrudeauMustGo (18% bot tweets). All ten of these top co-occuring hashtags exceed at least 10% bot usage at the 0.6 threshold.

For the not-a-bot hashtags, #TrudeauMustGo was by far the most commonly co-occurring hashtag, which is in line with the previous reporting conducted by the Canadian National Observer that suspected that the #NotABot hashtag was used by bots to promote the #TrudeauMustGo hashtag [15]. This hashtag was used almost three times as frequently as the next most popular co-occurring hashtag, providing more evidence of a coordinated campaign. The hashtag with the largest fraction of usages from bots exceeding the 0.6 threshold is #liberalcorruption, with 54% of usages coming from bots. At the same threshold, the hashtag with the next highest proportion of bot tweets were #liberalsmustgo (45% bot tweets), #ButtsMustGo (31% bot tweets) and #TrudeauMustGo (20% bot tweets). All of these hashtags are associated with the Liberal party, as Justin Trudeau was the Prime Minister of Canada and Gerald Butts was the Principal Secretary to Prime Minister Trudeau.

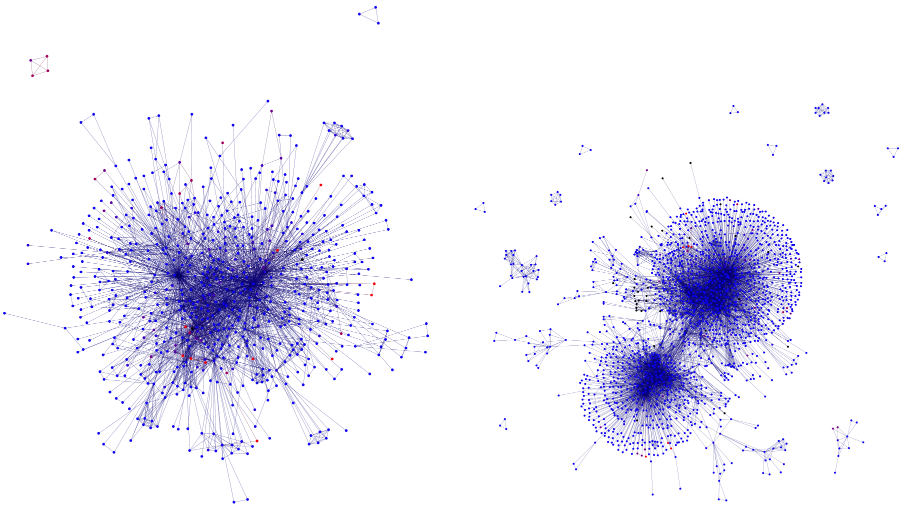
Interestingly, the hashtags #WeVapeWeVote, #VapingSavesLives, #IVape IVote, and #VapeBan were often used in conjunction with not-a-bot hashtags, meaning that not-a-bot hashtags were being used to promote voting within the vaping community. Out of the top fifty hashtags co-occurring with the not-a-bot hashtags, nine of them were related to vaping: #wevapewevote, #vapingsaveslives, #ivapeivote, #vapeban, #flavorsaveslives, #vapefam, #vaping, #vapingsavedmylife, #vape, #flavorban, and #vapersunitedworldwide. However, these hashtags had substantially fewer usages coming from bots, perhaps indicating a more genuine movement. Table 4 contains summary statistics for the top fifty hashtags co-occurring with the not-a-bot hashtags broken up by vaping-related versus not vaping-related. This table shows the average number of usages coming from bots at the three bot thresholds. Notice how the average bot usage is over twice as high in the non-vaping group for all thresholds except 0.8, in which they are close.

	Bot Score Threshold	Vaping Hashtags	Non-Vaping Hashtags
Average Proportion of Tweets from Bots	≥ 0.6	9.95%	21.3%
	≥ 0.7	0.73%	1.59%
	≥ 0.8	0.17%	0.10%
Average Usage		1,649 tweets	1,463 tweets

Table 4: The fifty hashtags most commonly co-occurring with the not-a-bot hashtags. There were eleven hashtags about vaping and thirty-nine not about vaping.

We additionally looked at the structure of the hashtag co-occurrence networks for both the fake-news and not-a-bot hashtags. These two networks show connections between hashtags that were used together within the same tweet. Figure 3a shows the hashtag co-occurrence network for tweets that contain fake-news hashtags, but the fake-news hashtags themselves have been removed from the network. Figure 3b shows the corresponding network for tweets containing not-a-bot hashtags (with the not-a-bot hashtags removed). In both figures, the hashtag nodes have been colored according to the proportion of their uses coming from bots (using a BotHunter threshold of 0.7), with red indicating more bot usage.

The hashtag co-occurrence network for tweets containing fake-news hashtags does not show a remarkably interesting structure, though there is a separate group with a noticeable bot presence consisting of the hashtags #Melanie LovesTrudeau, #WheresSophie, #JustinLovesButts, and #MelaniaLovesTrudeau. The network for tweets containing not-a-bot hashtags is much more interesting in that it seems to contain two main groups. Similarly to the analysis we carried out for the reciprocal communication networks, we used CONCOR [6] to separate the not-a-bot hashtag co-occurrence network into two clusters. One



(a) Hashtags used with fake-news hashtags. (b) Hashtags used with not-a-bot hashtags.

Fig. 3: The hashtag co-occurrence networks for tweets containing either fake-news or not-a-bot hashtags. Note that the fake-news or not-a-bot hashtags themselves have been removed from their respective networks.

of the two clusters is dominated by the vaping-related hashtags, which rank highly in terms of usage count, degree centrality, and betweenness centrality. The most important hashtags in the other cluster are predominantly political hashtags, such as #TrudeauMustGo, #ButtsMustGo, #cdnpoli, and #LiberalsMustGo. The not-a-bot hashtag co-occurrence network also shows several separate components of hashtags, including a group of three hashtags on the Honk Kong protests of late 2019, a set of several hashtags on computer gaming, and a group of hashtags about Tulsi Gabbard, the U.S. Representative from Hawaii. This all makes it exceedingly clear that the not-a-bot set of hashtags is not exclusively used in Canadian politics or political contexts in general. These hashtags were used to promote several different topics of discussion on Twitter.

4.3 #FakeNews Target Analysis

In examining the use of the fake-news hashtags, we wanted to find which entities were being targeted the most with accusations of spreading misinformation. We were also interested in which entities were being targeted by Twitter bots. Figure 4 shows the number of tweets targeting the ten most-targeted entities, with the last bar showing the average number of times all other entities were targeted. The bars in the figure also show the proportion of tweets coming from bots identified using three different BotHunter probability thresholds: 0.6, 0.7, and 0.8. As mentioned previously, we did not include bots that were deemed organizational accounts by Huang and Carley's identity classifier [12].

Figure 4 reveals that the Canadian Broadcasting Corporation (CBC) was the entity targeted the most with accusations of spreading misinformation. Amy McPherson, a freelance journalist for *HuffPost* based in Ontario [1], was the second-most targeted entity. Overall, the most commonly targeted entities were mainly important Canadian news sources like CTV News and the Toronto Star as well as prominent politicians like Justin Trudeau and Andrew Scheer.

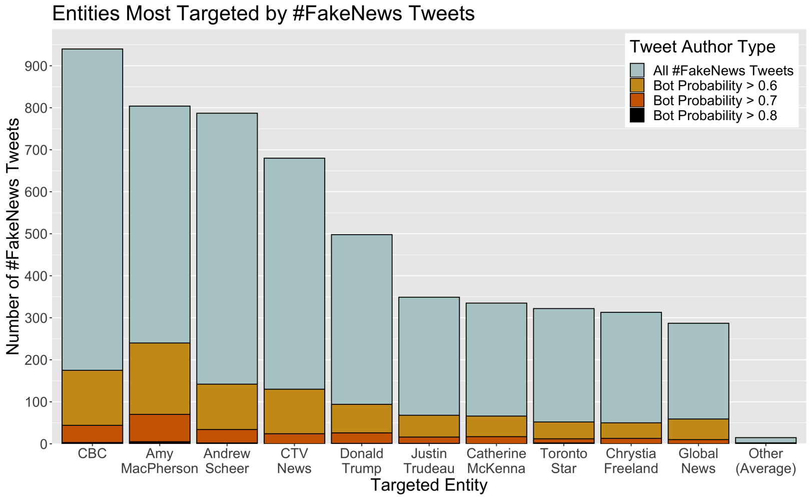


Fig. 4: A bar plot showing the number of times the most-targeted entities were accused of spreading misinformation. Different colors show the portions of tweets coming from bots detected at various BotHunter thresholds.

Using the four categories we manually assigned to each target, we also checked which category of entities was targeted the most with fake-news accusations. Figure 5 shows the number of times each category was targeted, along with the proportion of accusations coming from bots. The plot shows that despite the most-targeted individual entities being a news agency and a journalist, the most targeted category of entities was politicians. This makes sense because our data was collected to cover discussions on Canadian politics.

4.4 Analysis of Bots Using #NotABot

We wanted to investigate whether users of the not-a-bot hashtags were more or less likely to be bots than regular Twitter users in the Canadian data set. We found that there was not a substantial difference in the proportion of bot users in the not-a-bot hashtag users when compared to the rest of the population. As

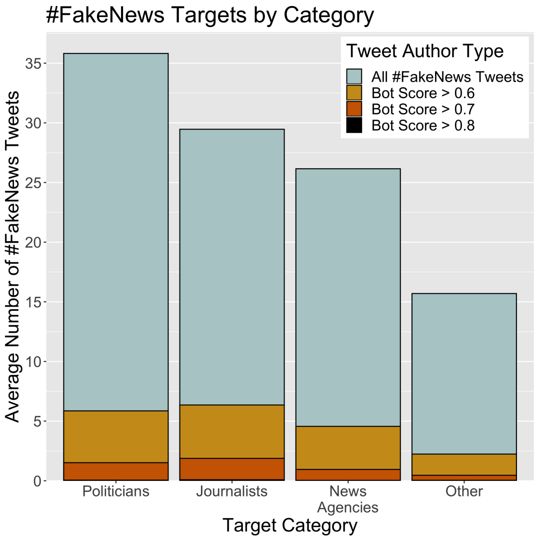


Fig. 5: A bar plot showing the number of times each category of entities was accused of spreading misinformation. Different colors show the portions of tweets coming from bots detected at various BotHunter thresholds.

seen by the blue bars in Figure 6 and the blue lines in Figure 7, the fraction of users exceeding a specific BotHunter probability threshold is similar between the two groups.

We chose to focus on the bot probability thresholds of 0.6, 0.7, and 0.8. Lower thresholds may mis-classify some cyborgs or humans as bots, while higher thresholds may miss some of the bots. As displayed in Table 5, we ran two-sample proportion tests of equality on the percentage of users that are bots between the two groups. For all three thresholds, the p-values were not statistically significant, indicating that there is no evidence to suggest that the proportion of users that are bots is different between the two groups of users. Considering this range of probability thresholds makes our results more robust.

However, just using the BotHunter probability threshold does not differentiate between bot types. Some bots are official accounts for celebrities, news agencies, or reporters; these are not typically malicious and are allowed by the platform. Therefore, as described previously, we used the algorithm developed by Huang and Carley [12] to classify the user types. After removing all "official" bots from consideration in both the group of users that use the not-a-bot hashtag and the group of users that do not, we found that the proportion of bots was higher in the not-a-bot groups no matter the threshold. Additionally, we ran the two-sample proportion tests for equality at the 0.6, 0.7, and 0.8 bot probability thresholds. The results were statistically significant, indicating that there is evidence that the proportion of bots between the two groups

	All Bots			All Non-Official Bots		
Bot Threshold	#NotABot	Canada	P-Value	#NotABot	Canada	P-Value
≥ 0.6	16.21%	15.97%	0.55	14.38%	9.59%	2.2e-16
≥ 0.7	5.10%	5.25%	0.54	4.38%	3.00%	1.9e-14
≥ 0.8	1.47%	1.70%	0.10	1.22%	0.87%	4.3e-04

Table 5: The proportion of users classified as a bot in both the group of not-a-bot hashtag users and the rest of the Canadian users, over three bot thresholds. The p-values are the results from the two-sample proportion tests for equality.

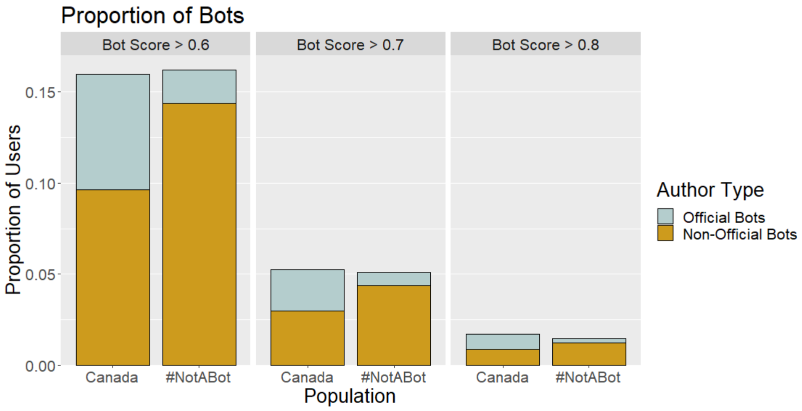


Fig. 6: A bar plot showing the percentage of users in the #NotABot dataset and the rest of the Canadian dataset that were detected as bots at three thresholds. Official accounts were detected using Huang and Carley's identity classifier [12].

is different. In this case, the proportion of bots in the group of users of the not-a-bot hashtags was higher than the users that did not use those hashtags, indicating that many accounts are lying about their bot status. Table 5 contains more detailed results.

We also ran a Mann-Whitney U test, a non-parametric statistical test evaluating the null hypothesis that the distribution of the two populations is the same. The p-value for this test was virtually zero, indicating that the distribution of bot scores was likely different in the two groups of users. As shown in Table 6, which contains the summary statistics for both groups of users, there are noticeable differences. For example, the median bot score for users of not-a-bot hashtags is 0.427, and the median bot score for the rest of the data is 0.410. While this difference may seem small (less than two percent) it shows that these differences exist across all users. Both the median and mean are higher in the not-a-bot group.

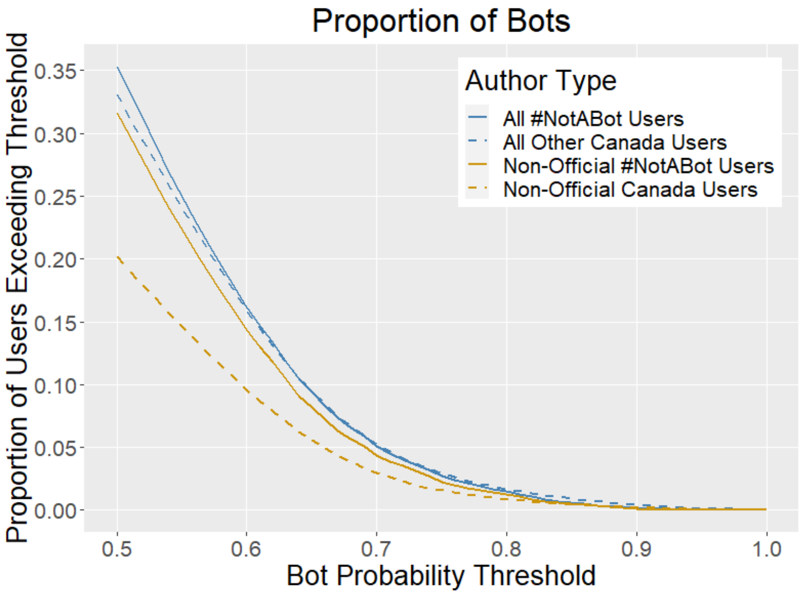


Fig. 7: A plot showing the percentage of users that were detected as bots at all thresholds above 50%. First shown in [13].

	Min	First Quartile	Median	Mean	Third Quartile	Max
#NotABot Users	2.7%	29.2%	42.7%	42.6%	55.0%	99.8%
Canadian Users	1.0%	27.0%	41.0%	41.2%	54.5%	100%

Table 6: The summary statistics for the bot scores in the #NotABot group and the rest of the Canadian users.

4.5 Analysis of Bot Behavior

After discovering the noticeable bot presence in the users of fake-news and not-a-bot hashtags, we became interested in understanding more about the behavior of these bots. Of primary concern was the extent to which these bots were being used to amplify existing messages as opposed to publicizing new content. In order to measure this, we calculated the proportion of bot tweets that were retweets of pre-existing tweets. We were also interested in seeing whether bots were being used to bridge or build communities between people. Bots that attempt to build or bridge communities may be doing so by mentioning multiple people within the same tweet, since that would introduce the same type of content into the feeds of the mentioned users. We therefore tallied how many bot tweets mentioned two or more users, excluding retweets. Table 7 shows the results of evaluating the extent to which bots were amplifying existing messages or were bridging or building communities.

			Retweets		Bridging Tweets	
	Bot Score	Bot Tweets	Retweets	Proportion Retweets	Bridging Non-Retweets	Proportion Bridging Non-Retweets
Fake-News Hashtag Tweets	0.6	2,433	1,910	78.50%	172	32.89%
	0.7	590	503	85.25%	24	27.59%
	0.8	25	19	76.00%	0	0.00%
Not-A-Bot Hashtag Tweets	0.6	7,386	3,588	48.58%	850	22.38%
	0.7	1,132	642	56.71%	154	31.43%
	0.8	140	49	35.00%	21	23.08%

Table 7: The proportion of bot tweets consisting of retweets for both the fake-news and not-a-bot hashtags, along with the proportion of non-retweets from bots that mention multiple people.

The results show that when using fake-news hashtags, the vast majority of bot activity (76% to 85%, depending on the bot score threshold) actually consists of retweeting. This means that bots primarily seek to amplify pre-existing tweets containing fake-news hashtags rather than generating new messages with fake-news hashtags. This is less true for the not-a-bot hashtags, for which only 35% to 57% of the activity consists of retweeting, depending on the bot score threshold. This makes sense, since not-a-bot hashtags should be most useful when used by bots to conceal their identities when tweeting new content.

When considering bot tweets that mention two or more other users and are not retweets, we find more similarity between the fake-news hashtags and the not-a-bot hashtags. For the fake-news hashtags, 28% to 33% of non-retweets coming from bots (at the 0.6 and 0.7 thresholds) appear to be attempting to bridge or build communities of users by mentioning multiple people. Meanwhile, for the not-a-bot hashtags, 22% to 31% of non-retweets coming from bots mention multiple people. It therefore seems that a noticeable portion of original tweets sent by bots may be intended for building up groups of Twitter users or bridging different communities within the platform.

We also wanted to see which communities bot users may have been attempting to bridge. Table 8a shows the five users mentioned most often in bridging tweets sent by bots containing fake-news hashtags, whereas Table 8b shows the corresponding information for the not-a-bot hashtags. The fake-news hashtag tweets mention the kinds of users that might be expected: politicians and news organizations. The results for the not-a-bot hashtag tweets are somewhat more interesting, in that they reveal a noticeable overlap with U.S. politics. Accounts based in the U.S., such as those of Donald Trump, a U.S. campaign for reducing tobacco use by children, and two *Wall Street Journal* reporters, all rank highly in the number of times they are mentioned in bots' bridging tweets. This is likely a result of the data we were served by Twitter, but it may also be an indication that the communities of bots used in Canadian political discussions have some overlap with those used in U.S. political discussions.

Users Mentioned in Bridging Tweets by Bots Using Fake-News Hashtags			
Bot Score > 0.6		Bot Score > 0.7	
User	#	User	#
JustinTrudeau	26	CBCNews	6
CBCNews	25	CTVNews	3
AndrewScheer	24	MaximeBernier	2
CTVNews	16	CPC_HQ	2
liberal_party	13	globalnews	2

(a) The most mentioned users in bot bridging tweets containing fake-news hashtags.

Users Mentioned in Bridging Tweets by Bots Using Not-A-Bot Hashtags					
Bot Score > 0.6		Bot Score > 0.7		Bot Score > 0.8	
User	#	User	#	User	#
JustinTrudeau	92	JustinTrudeau	17	TobaccoFreeKids	3
gmbutts	73	gmbutts	14	johndmckinnon	3
realDonaldTrump	52	realDonaldTrump	13	TomBurtonWSJ	3
AndrewScheer	50	AndrewScheer	9	realDonaldTrump	2
CTVNews	45	cathmckenna	6	parscale	2

(b) The most mentioned users in bot bridging tweets containing not-a-bot hashtags.

Table 8: The five users mentioned most often in non-retweets sent by bots that mention two or more users.

5 Discussion

Our results show that the structures of the reciprocal communication networks for both the fake-news hashtag users and the not-a-bot hashtag users reflect the partisan divide between the hashtags' users. The structure of reciprocal all-communication network of these hashtags cleanly break into two opposing political groups. Additionally, we found that anti-liberal hashtags were among the most popular to be used in conjunction with either of these two groups of hashtags, with #TrudeauMustGo being the most likely hashtag to co-occur with these hashtags. We found that the users of the #TrudeauMustGo hashtag were more likely to be bots than the users of the other most popular co-occurring hashtags. This confirms previous reporting done by the National Observer that speculated that the #TrudeauMustGo movement was in part fueled by bots [15]. Additionally users of vaping-related hashtags, most of which were against vaping or flavor bans, were on average less likely to be bots. This specific political issue and movement seemed to have garnered an authentic following.

Our analysis of the usage of the #FakeNews hashtags show that well-known, important news agencies were the entities most associated with fake-news hashtags, meaning that they were often targeted with accusations of propagating misinformation. We further found that the use of a not-a-bot hashtag was not a good indicator for an account not being a bot. After disregarding organizational bots, such as those from government agencies and news reporters, the proportion of bots present in the population using not-a-bot hashtags was higher than among accounts that did not use those hashtags. This noticeable difference in the proportion of users that were classified as bots was statistically significant at the 0.6, 0.7, and 0.8 bot probability threshold levels.

Because the structure of the reciprocal communication networks show a clear partisan divide between users of both the #FakeNews and #NotABot hashtags, this noticeable divide indicates that both liberal and conservative users level accusations of propagating misinformation against each other. The fake-news accusations aimed at established news organizations also show that accusations of spreading misinformation were being used to deceive people about what information is actually true. This also extends to claims of not being a bot, which are used by both liberal and conservative groups of users to conceal the true natures of their identities. We also showed that both these groups of hashtags were used in non-political contexts, such as for promoting the vaping community.

6 Conclusions

This work investigates some of the new defensive strategies that malicious actors use to help facilitate the spread of misinformation on social media, specifically in the context of a democratic election. The label "fake news" was used more commonly against mainstream news sources and reporters rather than actual, lower-quality reporting that may be satirical, intentionally misleading, or in general not of good quality. The usage of the term "fake news" in this disingenuous way may prevent users of these social media platforms from being able to accurately differentiate between what is accurate and what is not. Similarly, the usage of the term "not a bot" was also found to be more commonly used by likely bot accounts than other users discussing the Canadian elections in our data set. Therefore, this label is also not helpful for users to be able to distinguish between a bot campaign and genuine support for something. This work demonstrates a type of textual defense that many of these bot campaign or malicious actors have been employing. However, it is unclear whether other users on the platform believe these misleading users when they use these hashtags in inauthentic ways.

Future work could build on this research and investigate whether these defensive mechanisms are effective. It is important to determine if these hashtags and misinformation campaign strategies are working as intended and potentially altering the opinions or behaviors of voters. Determining the impact of these influence campaigns is essential in the development and deployment of effective counter-measures, such as third-party fact-checking, accuracy nudges, or others. While the general public has become more aware of bots and misinformation campaigns since media coverage of these issues has increased since 2016, they may not be aware of specific adjustments in misinformation strategies including the increased chance of lying when using these popular hashtags.

Additional work in this area could focus on how misinformation campaigns continue to evolve. These actors could replace their usage of these hashtags with other hashtags, or they could continue using these hashtags in other ways. As certain techniques become noticed by the social media platforms and the general public, these accounts may co-opt other less obvious hashtags to appear less deceptive. Also, misinformation strategies may differ in different countries, with some influence techniques being more effective in certain areas than others. It would be useful to compare across countries and regions to monitor patterns and changes. Understanding how these hashtags and influence operations continuously evolve and change over time is a tough problem requiring future work in multiple areas.

Acknowledgements This work was supported in part by the Office of Naval Research (ONR) Award 00014182106 and the Center for Computational Analysis of Social and Organization Systems (CASOS). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the ONR or the U.S. government. The authors would also like to thank David Beskow for collecting the data used in this study and for running his BotHunter algorithm on the data. We would also like to thank Binxuan Huang for access to his Twitter user classification system.

References

1. Amy MacPherson. URL https://www.huffingtonpost.ca/author/amy-macpherson/
2. Bail, C.A., Guay, B., Maloney, E., Combs, A., Hillygus, D.S., Merhout, F., Freelon, D., Volfovsky, A.: Assessing the Russian Internet Research Agency's impact on the political attitudes and behaviors of American Twitter users in late 2017 (2019)
3. Beskow, D., Carley, K.: Bot-hunter: A tiered approach to detecting characterizing automated activity on twitter. In: International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation (2018)
4. Beskow, D.M., Carley, K.M.: Bot Conversations are Different: Leveraging Network Metrics for Bot Detection in Twitter. In: 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 825–832. IEEE, Barcelona (2018). DOI 10.1109/ASONAM.2018.8508322. URL https://ieeexplore.ieee.org/document/8508322/
5. Beskow, D.M., Carley, K.M.: Social cybersecurity: An emerging national security requirement (2019). URL https://www.armyupress.army.mil/Journals/Military-Review/English-Edition-Archives/Mar-Apr-2019/117-Cybersecurity/
6. Breiger, R.L., Boorman, S.A., Arabie, P.: An algorithm for clustering relational data with applications to social network analysis and comparison with multidimensional scaling. Journal of Mathematical Psychology **12**(3), 328 – 383 (1975)
7. Carley, K.M., Cervone, G., Agarwal, N., Liu, H.: Social cyber-security. In: R. Thomson, C. Dancy, A. Hyder, H. Bisgin (eds.) Social, Cultural, and Behavioral Modeling, vol. 10899, pp. 389–394. Springer International Publishing (2018)
8. Committee on a Decadal Survey of Social and Behavioral Sciences for Applications to National Security, Board on Behavioral, Cognitive, and Sensory Sciences, Division of Behavioral and Social Sciences and Education, National Academies of Sciences, Engineering, and Medicine: A Decadal Survey of the Social and Behavioral Sciences: A Research Agenda for Advancing Intelligence Analysis. National Academies Press (2019)
9. Golbeck, J., Mauriello, M.L., Auxier, B., Bhanushali, K.H., Bonk, C., Bouzaghrane, M.A., Buntain, C., Chanduka, R., Cheakalos, P., Everett, J.B., Falak, W., Gieringer, C., Graney, J., Hoffman, K.M., Huth, L., Ma, Z., Jha, M., Khan, M., Kori, V., Lewis,

E., Mirano, G., WilliamT.Mohn, I.V., Mussenden, S., Nelson, T.M., Mcwillie, S., Pant, A., Shetye, P., Shrestha, R., Steinheimer, A., Subramanian, A., Visnansky, G.: Fake news vs satire: A dataset and analysis. In: WebSci '18 (2019)

10. Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., Lazer, D.: Fake news on Twitter during the 2016 U.S. presidential election **363**(6425), 374–378 (2019)
11. Gupta, A., Lamba, H., Kumaraguru, P., Joshi, A.: Faking sandy: Characterizing and identifying fake images on twitter during hurricane sandy. In: Proceedings of the 22nd International Conference on World Wide Web, pp. 729–736. Association for Computing Machinery (2013)
12. Huang, B., Carley, K.M.: Discover your social identity from what you tweet: a content based approach. Disinformation, Misinformation, and Fake News in Social Media-Emerging Research Challenges and Opportunities (2020)
13. King, C., Bellutta, D., Carley, K.M.: Lying about lying on social media: A case study of the 2019 canadian elections. In: B.H.D.C.H.A. Thomson R., H. Muhammad (eds.) Social, Cultural, and Behavioral Modeling, *Lecture Notes in Computer Science*, vol. 12268, pp. 75–85. Springer (2020)
14. Levi, O., Hosseini, P., Diab, M., Broniatowski, D.: Identifying nuances in fake news vs. satire: Using semantic and linguistic cues. In: Proceedings of the Second Workshop on Natural Language Processing for Internet Freedom: Censorship, Disinformation, and Propaganda, pp. 31–35. Association for Computational Linguistics (2019)
15. Orr, C.: A new wave of disinformation emerges with anti-Trudeau hashtag. URL https://www.nationalobserver.com/2019/07/25/analysis/new-wave-disinformation-emerges-trudeaumustgo
16. Panetta, A., Scott, M.: Unlike U.S., Canada plans coordinated attack on foreign election interference (2019). URL https://politi.co/30WXcIa
17. Rheault, L., Musulan, A.: Investigating the role of social bots during the 2019 canadian election (2019). DOI http://dx.doi.org/10.2139/ssrn.3547763
18. Ribeiro, M.H., Guerra, P.H.C., Jr., W.M., Almeida, V.: "Everything I disagree with is #FakeNews": Correlating political polarization and spread of misinformation. In: Data Science + Journalism Workshop @ KDD 2017. Halifax, Canada (2017)
19. Rizza, A.: British government says reports that child murderer will be sent to canada are false (2019). URL https://globalnews.ca/news/5850357/uk-child-murderer-canada-reports/
20. Staff, J.P.: Canada denies arab media report it will take in 100,000 palestinians (2019). URL https://www.jpost.com/Middle-East/US-has-agreement-with-Canada-to-accept-100000-Palestinians-Arab-report-600602
21. Tariq, S., Lee, S., Kim, H., Shin, Y., Woo, S.S.: Detecting both machine and human created fake face images in the wild. In: Proceedings of the 2nd International Workshop on Multimedia Privacy and Security, pp. 81–87. Association for Computing Machinery (2018)
22. Thompson, E.: Gould says after some online election meddling detected (2019). URL https://www.cbc.ca/news/politics/election-misinformation-disinformation-interference-1.5336662