Social Network Analysis and Mining (2022) 12:80
https://doi.org/10.1007/s13278-022-00908-6

ORIGINAL ARTICLE

How disinformation operations against Russian opposition leader Alexei Navalny influence the international audience on Twitter

Iuliia Alieva[1] · J. D. Moffitt[1] · Kathleen M. Carley[1]

Received: 1 March 2022 / Revised: 14 June 2022 / Accepted: 16 June 2022

Abstract
Previous research dedicated a lot of effort to investigation of the activities of the Internet Research Agency, a Russia-based troll factory, as well as other information operations. However, those studies are mostly focused on the 2016 U.S. presidential election, Brexit, and other major international political events. In this study, we have attempted to analyze how narratives about a domestic issue in Russia are used by malicious actors to promote harmful discourses globally and persuade an international audience on Twitter. We have identified bot and troll activities related to the Twitter discussions of a Russian opposition leader Alexei Navalny using social network analysis and bot detection. We have also implemented the BEND framework to find persuasion maneuvers that are used by bots in conversations about Navalny and found attempts to manipulate the opinion of the international audience on Twitter. Our findings have demonstrated that there is a significant presence of bot activities in information operations against Alexei Navalny as one of the leaders of the Russian opposition. We have observed how the Russian domestic issue is framed in the context of Russian confrontation with the West and how it is used to promote hostile narratives either against Navalny, an opposition movement, or democratic values. Many agents that we have identified pretend to be English speakers, who exhibit hostile attitudes towards Navalny and the Western democracies, express skepticism and distort the facts, promote a lack of trust in the democratic institutions as well as spread disinformation and conspiracy theories.

Keywords Disinformation · Social cybersecurity · Computational propaganda · Russia · Bots · Social network analysis

1 Introduction

With the development of technologies, journalism, and social media, the issue of investigating information disorder activities online became a new interdisciplinary area. Researchers, journalists, and NGOs attempt to find ways to detect coordinated inauthentic behavior, mis-/disinformation campaigns, and hate speech as well as to generate solutions to prevent harmful activities. The 2016 U.S. presidential election, Brexit, and other political issues globally have demonstrated that there are multiple attempts to manipulate public opinion online (Badawy et al. 2019; Lukito et al. 2020; Nyst and Monaco 2018). Widespread activities of bots and trolls attracted a lot of attention from scientists in various fields. After 2016, the global community revealed troll farms in various countries. To mention as an example, the Russian Internet Research Agency was identified as the main source of malicious activities that attempted to manipulate users' opinions online through divisive messages (Badawy et al. 2019; Bastos and Farkas 2019; Linvill and Warren 2020; Lukito et al. 2020). Numerous studies dedicated to the analysis of Internet information operations, including activities of bots and trolls, formed a new interdisciplinary area that investigates online disinformation and computational propaganda (Benkler et al. 2018; Woolley 2020).

To a greater extent, existing research is focused on the information disorder activities that are coordinated by the actors directly or informally affiliated with the current Russian political regime such as IRA and target international audience (Bodrunova 2021; Golovchenko et al. 2020; Lukito 2020). However, most of those studies focus on troll activities that implement domestic national discourse in the United States or other countries, as we see in the existing analysis of Internet operations and interference in the 2016 U.S. presidential election, Brexit campaign, or COVID-19

✉ Iuliia Alieva
ialieva@cmu.edu

[1] Carnegie Mellon University, Pittsburgh, PA, USA

Published online: 14 July 2022

vaccination coverage (Dawson and Innes 2019; Chen et al. 2021; Bastos and Mercea 2018). Previous studies also analyzed Russian disinformation campaigns, including activities of bots and trolls that target domestic audiences in the Russian segment of the Internet (Sanovich et al. 2018; Stukal et al 2019). This study expands previous research about activities of malicious actors online with an additional case study that shows the attempts to manipulate the attitudes of foreign audiences through the inauthentic negative narratives about Alexei Navalny and the Russian opposition movement.

The main goal of this analysis is to identify bot and troll activities related to the conversations about the opposition movement in Russia using the case of Alexei Navalny as well as to investigate promoted narratives using community detection analysis and network science metrics. This study identifies propaganda operations that demonstrate the domestic political confrontation between the Russian systemic political establishment and opposition movement through the case of Alexei Navalny and the Anti-Corruption Foundation. We present an analysis of how propaganda trolls and sock puppets frame the discussion around the opposition movement in Russia on Twitter as well as attempt to manipulate the opinion of the global audience. As the main tactics, trolls use accusations of collusion with foreign governments, spread disinformation, conspiracy theories, and negative discourse about Western democracies. In order to investigate the whole picture of Twitter communication about the issue, the study implements a bot detection algorithm, network analysis, and mixed-method approach for the analysis of disinformation and propaganda trolls.

2 Related works

2.1 Computational propaganda, social cybersecurity, and state-funded propaganda operations

The expansion of broad computational power, Internet resources, and big data is now being exploited by various actors and governments to serve their information control and manipulation goals. It opens up more opportunities to implement novel strategies that would penetrate the social discourse and guide it in a more convenient direction. Analyzing how governments use social media to control information environments adds a broader context to understanding the current situation around information operations online. Many states have switched their traditional strategies and realized that social media offers new opportunities for spreading disinformation, consolidating power and social control, and generating narratives to promote their agenda (Nyst and Monaco 2018; Weedon et al 2017). Countries like China, Turkey, Ecuador, Russia, Philippines, Bahrain, Azerbaijan, Venezuela, and the United States have implemented information operations for political goals as identified in the report by the Institute for the Future (Nyst and Monaco 2018). Very often these trolling attacks target journalists, activists, and other actors who criticize the government's actions. However, some countries such as China, Russia, and Turkey, have professionalized propaganda operations online and built systems with “troll farms” and youth groups to support government social media campaigns (Nyst and Monaco 2018; Sanovich et al. 2018). State-sponsored patriotic trolling uses political bots and trolls to exploit social media algorithms to amplify its propaganda and disinformation. In Russia, it is combined with the network of state-funded news websites, similar to so-called pink slime websites that are used to add visible credibility to the messages (Bengani 2020). Previous research has underlined that state-funded information operations include a set of coordinated channels such as bots, trolls, and websites that are used to create an alternative informational space and effectively disseminate state propaganda messages.

In this study, we decided to focus on analyzing the issue from Russia's agenda, such as the activities of the Russian opposition, and investigate related conversations in the English segment of Twitter. We assumed that the Russian government and affiliated actors would attempt to manipulate the existing discourse and build their own narratives for better persuasion of the foreign audiences. As previous research has identified, the Russian government dedicates substantial resources to developing information operations (Nyst and Monaco 2018; Sanovich et al. 2018; Stukal et al. 2019; Tsyrenzhapova and Woolley 2021). Russia has honed its propaganda tactics since the inception of the Soviet Union through the Cold War. Manipulating public opinion is one of the current government's priorities where the previous experience is used to control information flows. The main principles of propaganda communication remain the same: persuasion through symbols, emotions, stereotypes, and pre-existing frames with the purpose of shaping perceptions, manipulating cognition and behavior to fit the propagandist's goal (Hemanus 1974; Jowett and O'Donnell 2014). With social media and computational tools, this propaganda is designed to be more targeted. Within the current environment of echo chambers, it has become easier to detect communities of people who share the same range of ideals and target them by exploiting their beliefs, causing further polarization of their stances. Computational propaganda is similar to traditional propaganda but implements computational tools such as automation and algorithms to disseminate and amplify discourses and opinions on social media for ideological control and manipulation (Tsyrenzhapova and Woolley 2021).

Recently exposed Internet Research Agency, or so-called Russian troll factory, actively uses social media platforms and algorithms to promote strategic narratives to create destabilization, polarization, information chaos, and distrust (Bastos and Farkas 2019; Freelon and Lokot 2020; Linvill and Warren 2020). Among the main features of IRA trolls, researchers found intent to deceive (Badawy et al. 2019), to sow political discord, doubt, lack of trust, disagreement (Golovchenko et al. 2020; Lucas and Nimmo 2015); implement "astroturfing" with online troll accounts to mimic grassroots activities (Golovchenko et al. 2020; Peng et al. 2017; Woolley 2020). It is worth noting that various types of accounts are used, such as automated bots, trolls, and sock puppets controlled directly by humans and positioned as real social media users (Golovchenko et al. 2020; Badawy et al. 2019). While it is hard to measure the effects of those activities, since those accounts sometimes may not have many followers, it is still possible that those opinions could be considered vox populi by the media agencies. Lukito et al. (2020) found that the IRA accounts were quoted by at least 71 news media organizations, such as The Washington Post and The Guardian.

To further study and analyze this area, a novel field of social cyber-security has emerged. It is particularly centered on the computational methods used to characterize, understand, and forecast cyber-mediated changes in human behavior, social, cultural, and political outcomes. It also focuses on building the cyber-infrastructure used by our current information society to tackle issues that arise in a cyber-mediated and constantly changing information environment with actual or imminent social cyber-threats (Carley et al. 2018a). This emerging area will help us understand the importance of analyzing the strategies used in propaganda information operations and investigating the methods for better identification and resistance (Carley 2020).

2.2 Russian political and media environment

Per its constitution, Russia is a democratic country. In reality, the current political establishment demonstrates significant democratic backsliding by not adhering to democratic principles such as freedom of the press, freedom of political expression, and legitimate access and participation in the political process. The growth of authoritarianism has been taking place in recent years (Weiss 2013), further undermining democratic institutions and putting pressure on journalists and political activists. The suppression of political expression resulted in new forms of alternative political participation for opposition members via social media since official mass media platforms are not accessible for them. Many traditional official media resources are funded by the government and follow the line of the current political regime, restricting access for the opposition and any critical voices.

It was also the case for opposition leader Alexei Navalny and his Anti-Corruption Foundation, who tried to use available means of political participation and activism and communicate their goals with their anti-corruption investigations via social media. However, all mentions of Navalny or his activities are banned on official media sources. The Russian media system consists of a mix of democratic norms and market principles (artificially implemented in the industry in the 1990s) and paternalistic institutions inherited from the Soviet era (Kiriya 2019). As Kiriya argues, the current Russian media system is a form of hybrid "transitional media," in which market-oriented trends and government control coexist. Still, the state uses the media to advance its political agenda, creating a conflict between liberal values and the informal rules imposed by the ruling groups, favored oligarchs, and groups of *siloviki (*security and military personnel) (Lipman et al. 2018).

Today, the Russian media system is characterized by significant "intrusion of the state in social life, which forms particular practices of commercial and state-dependent agents in the field of pressure on the media by means of control over content and news" (Kiriya 2019). As a result of the recent legal steps and state intervention, the authorities succeeded in the further segmentation and isolation of the liberal opponents who are presented by the state media as "foreign agents." The main party line is supported by the most accessible news organizations in the country, which remain owned by the state, oligarchs, or other elites, who are inclined to represent the point of view of a dominating party (Kiriya 2019).

Previous research demonstrated that the media systems significantly influence communication practices, which affect how stories and narratives are presented in public conversations (Hallin and Mancini 2011; Wells et al. 2020). By limiting media freedom and implementing more restrictions and barriers for the journalists and independent media, the government may gain control over the social discourses by manipulating and restricting the critical coverage of unsavory issues or forcing the media to present them in a more positive light. Currently, Russian state media have become mouthpieces of state propaganda that frame key issues in a way that would benefit the current political regime and its beneficiaries.

Another way for a non-democratic country to control the media narrative is by eliminating opposition leaders' participation in the public discourse and reducing the amount of attention they receive from mass media (Kazun 2019). However, after recent political protests and governmental intervention in traditional mass media, a large portion of the Russian audience has been attracted by the freedom of speech and lack of censorship on the Internet and social

media. After the 2011–2012 protests for a fair election in Russia, the share of monthly Internet users was 46%, while at the end of 2020, it increased to 78% (Kiriya 2021). From one side, there was an optimistic view of the new media that brought hopes and positive dynamics to the development of alternative discourse and liberal freedoms. Although, from another side, some researchers have expressed a rather pessimistic view that social media contributes to further isolation and marginalization of the critical voices. Russian disruptive Internet operations demonstrated how computational propaganda exploits vulnerabilities and partisan divide in the Western democracies. However, the same process is taking place in the Russian segment of the Internet, where the differences between pro-government conservative groups and more liberal oppositional voices are used to create informational disorder and distract citizens from discussing more substantial political problems and issues (Sanovich et al. 2018). This research illustrates how Russian domestic social and political divides and discourses can be used to further advance the political disagreements not only in Russia but also globally.

While state-funded Russian media cover political issues from one side that would be favorable to the current political regime, independent Russian outlets and Western media actively demonstrate a wider variety of issues presenting political critique and inconvenient coverage that the Russian government cannot control. As a result, the Russian political regime implemented multiple Internet operations with bots, trolls, and sock puppets to manipulate the narrative on social media for the foreign audience: “From creating troll factories and bots to distort communication in social media, the state is progressively moving towards a strategy of creating a huge state-oriented information flood to “litter” online space” (Kiriya 2019). The main goal of the state is to control the informational environment through fragmentation and isolation of opposition groups and to promote the main-party agenda via state-funded media as well as information operations online with cyber tools. State-funded media such as RT and Sputnik are also used as the most well-known sources of propaganda news. The purpose of those activities is to exercise online agenda control as well as demobilize and undermine opposition supporters (Sanovich et al. 2018; Stukal et al. 2019).

2.3 Alexei Navalny and Anti-Corruption Foundation

Alexey Navalny presented himself as a prominent figure in the Russian political environment back in 2011–2012 during the mass protests after presidential and parliamentary elections in Russia. However, he is still considered to be an "outsider of systemic politics" and a "non-elected politician operating outside of the political system" (Glazunova 2020). As one of the opposition leaders, he encounters numerous barriers trying to get access to a political arena. Social media serves as the primary platform for his political expression and for maintaining communication between his organization and his followers. Due to the limited freedom of the press in Russia, YouTube, Twitter, Facebook, Instagram, and other platforms help Navalny participate in an information battle with the current political regime. Navalny started the Anti-Corruption Foundation (ACF) to conduct anti-corruption investigations. In their activities, ACF members combine the practices of investigative journalism and civic activism with the primary goal of revealing wrongdoings and corruption. Through detailed investigations of corruption activities in the government, they stand for political rights such as fair elections and open political participation.

In 2013, Navalny received 27.24% of the vote in the Moscow mayor elections and got the second place after the current mayor of Moscow Sergey Sobyanin, while in 2016 he announced his intention to run in the Russian presidential election in 2018: "It can be concluded that Alexei Navalny is already a sufficiently famous federal-level politician" (Kazun 2019). However, the Central Election Committee announced that Navalny was not eligible to run in the 2018 election due to a conviction preventing him from participating until after 2028. In 2018, Anti-Corruption Foundation created an app called Smart Voting, an instrumental algorithm and a tactical voting strategy that selected the most popular candidate from systemic opposition and intended to consolidate the votes of those who opposed the main party *Yedinaya Rossiya* ("United Russia"). The goal was not to prioritize a particular opposition party but rather to decrease the number of seats in the parliament for United Russia.

In August 2020, Navalny was hospitalized after he was poisoned with a Novichok nerve agent. He was medically evacuated to Berlin and discharged a month later. Navalny accused Russian president Vladimir Putin of being responsible for the incident. He also conducted his own investigation of the poisoning incident in collaboration with journalists from Bellingcat and Insider media outlets. The investigation showed evidence of FSB's (Russian Federal Security Service) involvement in the poisoning case. Russian authorities did not accept responsibility for Navalny's poisoning and have not initiated an official investigation of the case. The European Union and the U.S. responded by imposing sanctions on senior Russian officials. After his recovery, Navalny returned to Russia, where he was immediately detained on accusations of violating parole conditions from his previous case and was eventually arrested with 2.5 years of a prison sentence. After his arrest, ACF published the documentary investigation of *Putin's Palace,* which accused Vladimir Putin of corruption. These events have caused mass protests across the country in January 2021. A resolution by the ECHR (European Court of Human Rights) called for Navalny's release. Recently, Navalny’s political activism was

declared to be extremism by the Russian court. As a consequence of this decision, his organization and his supporters are currently under state prosecution.

2.4 Research questions and goals

Twitter has an enormous potential to form people's opinions. Social media affordances benefit politicians and propaganda spreaders who want to control the information discourse. As a result, Twitter often becomes a platform for bots, cyborgs, trolls, and other actors interested in spreading disinformation and propaganda that affects views and attitudes and shifts the narrative to fit the goal of the spreader. This study aims to analyze how one of the prominent opposition leaders in Russia is presented in the discourse on Twitter. The main goal is to identify user communities and message frames to detect information operations against Navalny. Our purpose is to identify and analyze attempts of discourse manipulation during the previous year after Navalny's poisoning. As we have mentioned, journalistic investigations found a connection between the poisoning by Novichok and the Russian intelligence services involved in the incident; therefore, we may expect to find information operations that would attempt to discredit Alexei Navalny and shift the discourse to undermining his personality.

Thereby, in order to investigate disinformation structures and strategies, the following research questions can be analyzed:

RQ1: What influential online communities and disinformation actors can be identified in a discussion about Alexei Navalny?

RQ2: What frames and goals can be detected in the disinformation messages spread by the malicious actors in order to distort the discourse about Alexei Navalny?

3 Data and methodology

The data collection and methodologies used in this study closely follow a pipeline from social cyber-security studies (Blane et al. 2022; Uyheng et al. 2020) to identify key influencers and their strategies for persuasion and information control (see Fig. 1).

3.1 Twitter data

To investigate our research questions, a sample of tweets was collected using the Python package twarc (Summers 2022) via an archive search with updated Twitter API version 2. For this study, we focus on analyzing tweets in the English-language segment of Twitter to find possible information operations and conversation manipulations targeting the international community. We collected tweets about Alexei Navalny from August 2020 until August 2021. This period covers Navalny's initial poisoning, imprisonment, and the subsequent time when journalistic investigations took place. The keyword search with the term 'navalny' was used to identify the tweets about Navalny for our dataset. As a result of data collection, our dataset contains 3,824,357 tweets, 3,081,054 retweets and 717,209 users (agents). We converted the raw Twitter data into a meta-network consisting of user-to-user communication networks, user-to-tweet, and user to various tweet artifacts (hashtags, URLs, location, etc.) networks with ORA software to conduct network analysis (Carley 2014).

3.2 Bot identification

We have used the BotHunter tool (Beskow and Carley 2018) to identify bot activities, a tiered supervised machine learning approach for bot detection and characterization. BotHunter is a random forest regressor trained on previously collected labeled tweets from information operation attacks

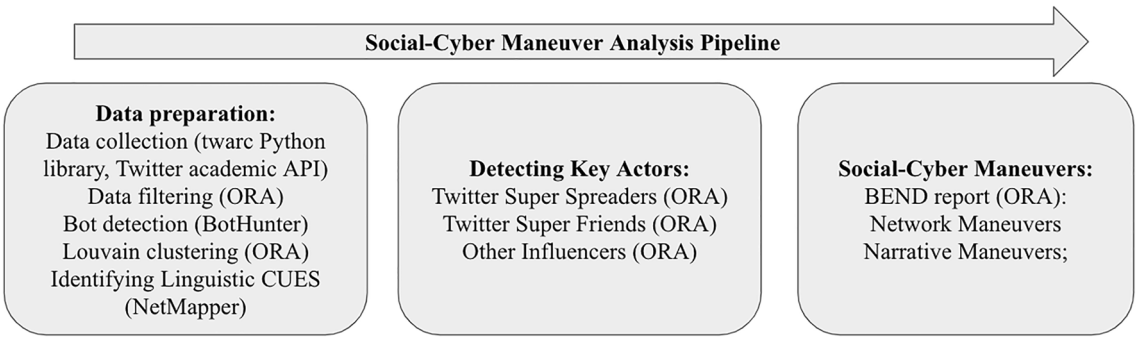

Fig. 1 Social-Cyber Maneuver Pipeline with the software tools used

on the Atlantic Council's Digital Forensic Lab and NATO as well as on the suspended Russian bot dataset released from Twitter in October 2018. The model uses several variables characterizing tweet content and user metadata. The model then produces a probability that the account is a bot. A general recommendation for a threshold to use is between 0.6 and 0.8, where a score closer to 0.6 would include more false positives, while a higher threshold would have more false negatives (Beskow & Carley, 2020). For this study, we use a threshold of 0.7 to label an account as a bot or non-bot as the most stable threshold for bot classification (Ng et al. 2022).

3.3 Community and key actor detection

We implement the Louvain method to identify network communities participating in the Navalny Twitter discussion (Blondel et al. 2008). The Louvain algorithm is a widely adopted method for community detection that allows more granular rendering of the network (Hagen et al. 2020; Uyheng and Carley 2019). Once identified, we use a mix of qualitative and quantitative methods to compare content and user characteristics between groups.

Influential (Key) actors are essential to successful information operations; thus, identifying them is an important step to finding and understanding potential operations (Larson et al. 2009). Leveraging the ORA network analysis tool and our Twitter-derived meta-networks, we focus on finding four types of key actors: super spreaders, super friends, influencers, and want-to-be-influencers. Super spreaders are users that generate often shared content and hence spread information effectively. In contrast, super friends are users that exhibit frequent two-way communication, facilitating large or strong communication networks. Measures that help identify these types of key actors in a derived Twitter all-communication network include out-degree centrality, page rank centrality, and metrics showing whether a user is a member of a large k-core.

Influencers send messages that reach many other users both directly and indirectly through their followership network and the followership networks of their followers (mentions, retweets, replies, etc.). Want-to-be-influencers share similar characteristics to influencers; they retweet, reply and quote frequently, but their follower network is often much smaller, thus decreasing reach.

3.4 Describing information and network maneuvers

We have compiled lists of 500 bot accounts with the highest scores in each measure and grouped them by their Louvain cluster. As a result, we have identified three main groups with the most influential users and qualitatively investigated their profiles. While the first and the second groups were represented by the politicians, journalists, and newsrooms, the third group included highly influential coordinated bot users. We have investigated the messages from the third group of bots using the BEND framework to identify the main goals of their communication. BEND maneuvers include 16 categories of maneuvers for online persuasion and manipulation (Beskow and Carley 2019). BEND framework serves as a tool for deciphering strategic engagement and information maneuvers (Carley 2020). The framework divides maneuvers in information space regarding positive and negative actions related to actions affecting narrative or network structure. Narrative maneuvers focus on the content of the message, while network maneuvers show network communities and structures.

To identify BEND maneuvers, we have extracted linguistic cues using NetMapper software that computes text features such as positive and negative sentiments and emotional attitudes of the source (Carley et al. 2018b). We have used ORA software to compute BEND analysis (see Table 1 for an overview of BEND maneuvers).

In addition, we conducted a qualitative analysis of the bot messages from the most influential community and analyzed the information maneuvers they implement. This methodology will allow us to see how the key influencers use Twitter infrastructure to communicate their narratives effectively.

4 Results

4.1 Bot activity

We found that 357,098 users from our full dataset with 717,209 users (~50%) are predicted bots. The users predicted as bots produced 2.8 million of the 3.8 million tweets (~73%) in our data. Figure 2 provides a temporal view of the volume of predicted bots and non-bots and the volume of tweets they produced from August 2020 to August 2021. Predicted bots dominate the Navalny conversation in presence and message volume across all months. Additionally, we find that the most active periods for the conversation occurred during January and April of 2021.

A qualitative analysis of the most influential nodes in the network classified as possible bots reveals critical insights into bot behavior and employment tactics. Bots in our data

Table 1 Categories for BEND Maneuvers

Narrative maneuvers		Community maneuvers	
Positive	Negative	Positive	Negative
Enhance	Dismay	Back	Neutralize
Excite	Distract	Build	Narrow
Explain	Distort	Bridge	Neglect
Engage	Dismiss	Boost	Nuke

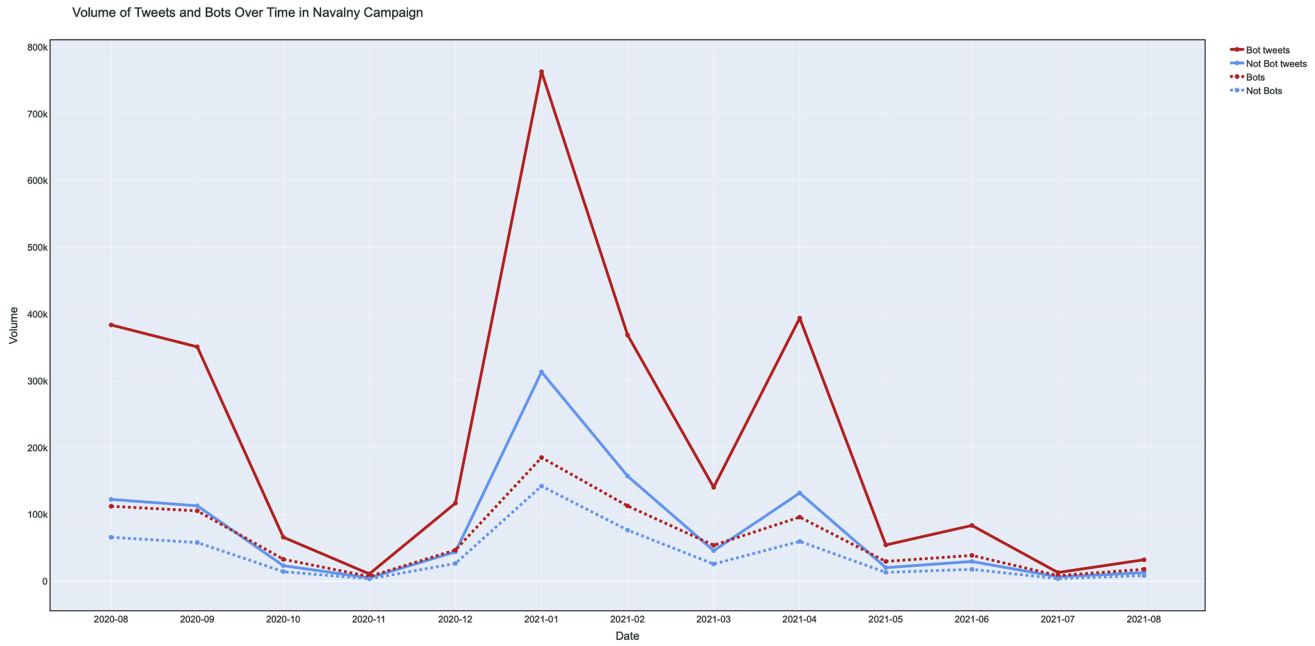

Fig. 2 This figure provides the timeline of tweets about Navalny and bot participation. The blue lines represent tweets produced by predicted non-bots and the number of non-bots in the Navalny conversation; in contrast, the red lines represent tweets produced by predicted bots and the presence of bots in the Navalny conversation.

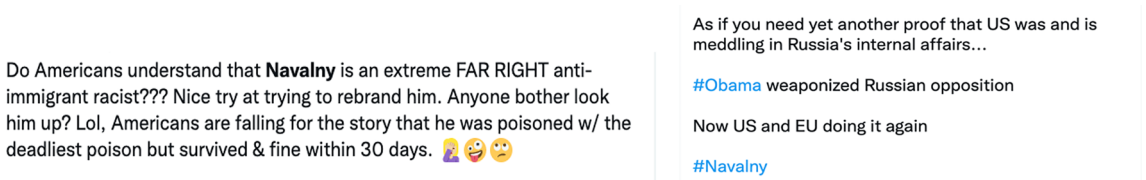

Fig. 3 Examples of anti-Navalny tweets

present their identities in short, standard bios, with random pictures as profile photos and self-report geolocations consistent with English-speaking countries. We identify that predicted bots produce tweets across four major themes: pro-opposition, anti-opposition, news-like bots, and conspiracy theory.

Though in the minority, pro-opposition tweets demonstrate support for Alexei Navalny in the messages. One of the most representative tweets of this theme follows, "Navalny would be president if the people of Russia could vote in a free and fair election."

Anti-opposition tweets focus on undermining Navalny's credibility, blaming the West for staging Navalny's poisoning, and for funding opposition with the purpose to harm Russia (one of the examples: "Germany reacted on Lukashenko statement about Navalny's poisoning being a fake (and that he has a record to prove it). Germany said, that "It goes without saying that Mr. Lukashenka's statement does not correspond to reality. As if they would admit that they faked it?"). Certain tweets attack independent journalists and media: "@BBCWorld As a verified @CIA and @USaid agent in the Russian Federation @navalny with 2% support is called a 'Russian' Opposition leader. LOL! Fake news promulgated by @BBCWorld. @nytimes @anneaplebaum @ASLuhn @ap @afp @npr @moscowtimes @meduza_en." Here is another example of the anti-Navalny tweet: "#Navalny is the opposition leader endorsed by the West (2% in polls, 6th party), he's the top opponent only in fairytales. Noone in #Russia pays attention to him as he has been caught taking money from abroad to weaken and dismember the country with destabilization campaigns. https://t.co/YF3vE4B9SZ." See Fig. 3 for more examples.

Conspiracy tweets focus on distorting the nature and obfuscating the actors behind Navalny's poisoning (example of the conspiracy tweet: "Moreover, false flag poisoning Navalny, with his 4% support, is profitable only to Navalny's people, who would want to make a martyr from him, thus eliminating the competitor for western funding & increasing

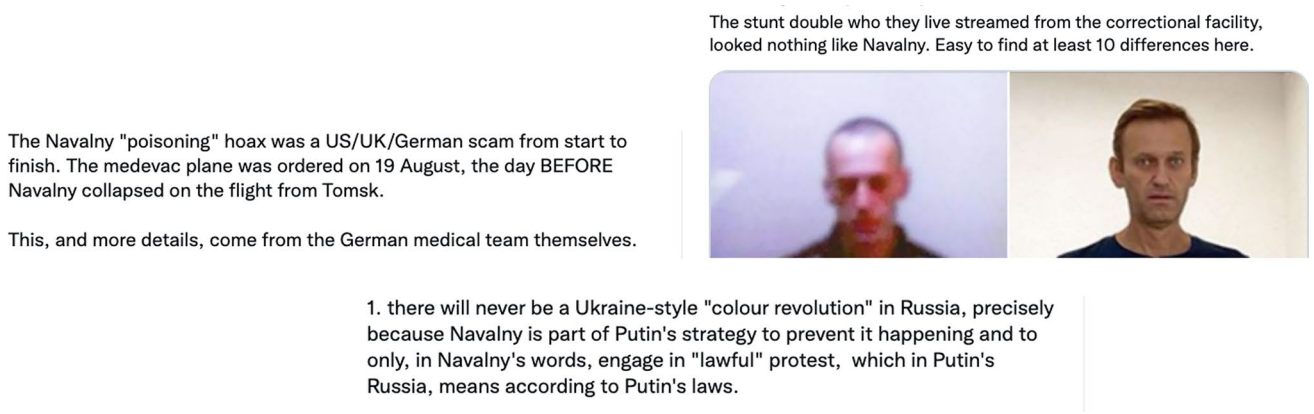

Fig. 4 Conspiracy tweets against Alexei Navalny.

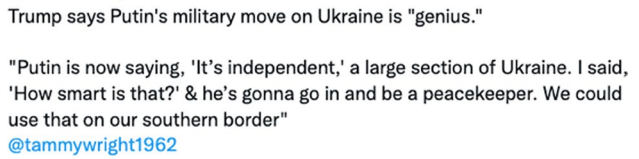

Fig. 5 A recent tweet from an anti-Navalny bot account.

the support base https://t.co/8NvXqo7rv8"). Conspiracy tweets call Navalny's poisoning "a false flag operation" by NATO to stop utilization of the Nord Stream 2 gas pipeline from Russia to Europe (example of the tweet text: "Do you need more proof that #Navalny "poisoning" is a NATO project to stop #NordStream2?"). Other tweets say that in fact Navalny was not poisoned: "Putin's assassination attempt Medical tests on Alexey Navalny have shown he has no toxic substances in his system. UT Oh! Well that blows the theory of assassination attempt out of the water. But the mainstream media will continue on. The anti-russian rhetoric is hilarious." See Fig. 4 for other examples of tweets promoting conspiracy theories about Alexey Navalny.

Another interesting observation of those bot accounts shows that they exhibit features of users in the United States, adding ideological context from the U.S. political discourse to their tweets ("Trump is Putin's puppet Trump tries to convince us that Putin & KGB are fine. The same Putin who meddled in the last election, which got Trump elected, who took over part of Ukraine and Crimea, and just poisoned his opponent Navalny Trump is a TRAITOR. https://t.co/78PMOOGj9p"). After checking recent posts from the bot accounts from our dataset, we could also identify attempts to influence U.S. users using the case of recent Russia's invasion of Ukraine (see Fig. 5).

4.2 Identifying groups

We identified three major groups participating in the conversation about Navalny through Louvain clustering and the identification of key actors. Group 1 and Group 2 consist of journalists and political figures, most of whom can be easily identified. Group 1 presents mostly pro-Navalny messages with highly influential actors such as the U.S. President Joe Biden, former U.S. ambassador to Russia Michael McFaul, the U.S. Secretary of State Antony Blinken, and other politicians and political activists. The list of influencers for Group 2 consists of political activists and journalists, including Navalny himself, the Bellingcat investigative journalists, RFERL, BBC, Washington Post correspondents, and others. In contrast, Group 3 has RT (a state-affiliated Russian newsroom) and its journalists, as well as highly influential bots and trolls, among the top influencers in the list. Many bots from this group spread anti-Navalny messages, and other hostile narratives focused on causing polarization and information chaos. Figure 6 provides a circular representation of a reciprocal communication network for three groups; nodes are colored by group membership and shaped by bot prediction. In the next section, we will present our BEND maneuver reports focusing on Group 3. This group was selected for detailed analysis since, in this group, we have identified multiple negative frames and narratives that were used to promote harmful discourse.

4.3 BEND maneuvers

After computing BEND analysis, we have identified specific differences in BEND maneuvers between communities. For positive narrative maneuvers, actors in Group 3 use less Explain, Engage and Excite but more Enhance maneuvers. In contrast, for negative narrative maneuvers, they would be more likely to use Distort, Dismay, Distract maneuvers and slightly less likely to use Dismiss maneuvers in comparison

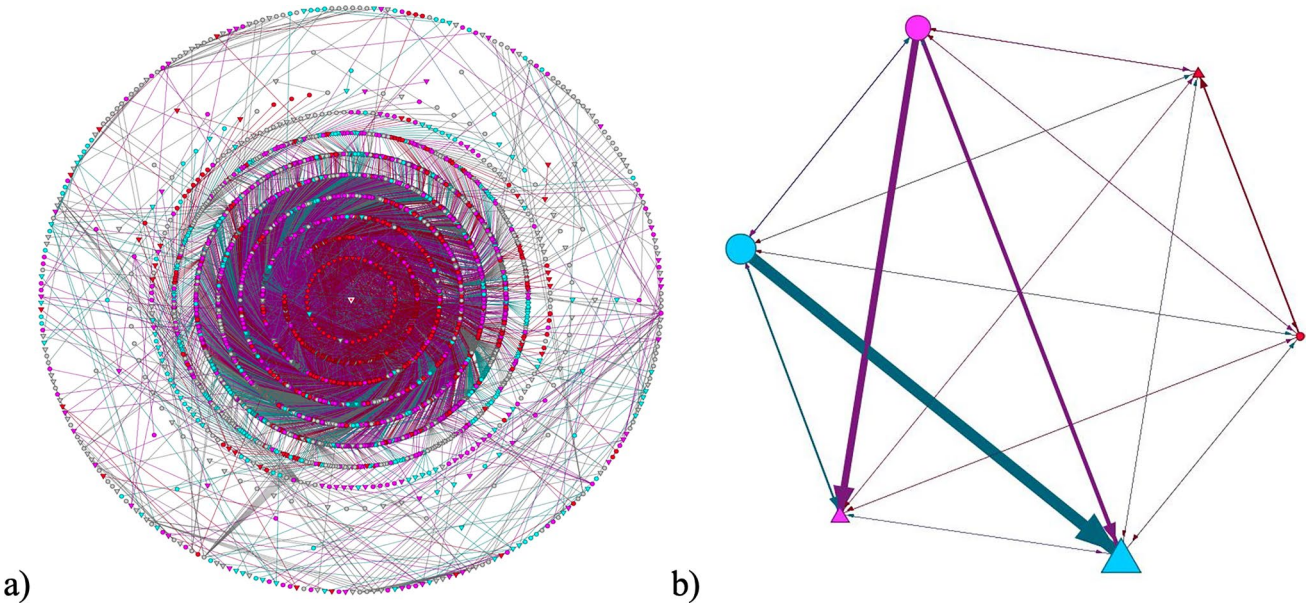

Fig. 6 This figure presents an overview of the three communities: **a** on the left, the graph presents the reciprocal communication network in a circle layout with the highest betweenness nodes towards the center; **b** on the right, the graph shows the all-communication network where nodes are sized by the number of agents, and the number of connections scales link weight. Nodes are colored by group membership and shaped by bot prediction for both graphs. Group 1 is light blue, Group 2 is pink, and Group 3 is red. Not bots are shaped as circles and bots as triangles.

with the other two groups. For positive network maneuvers, Group 3 would be more likely to use Bridge and Boost maneuvers and less likely to use Build and Back maneuvers. For negative network maneuvers, this group would be more likely to use all four maneuvers as Narrow, Neglect, Neutralize and Nuke (see Fig. 7). To understand how anti-Navalny bots build their communication strategies for spreading disinformation, we need to conduct a thorough analysis of each of the BEND maneuvers that are demonstrated by bots in Community #3.

Negative narrative maneuvers such as Dismiss, Distract, Distort, and Dismay are often used to amplify disinformation. Dismiss is used to express denial or downplaying facts as inconsequential, not worthy of one's attention, irrelevant, or wrong. Distort maneuver is implemented with the purpose to change or reinterpret information, neglect the context, and promote disinformation. Distract is used to misdirect the audience offering a new distracting topic or adding noise and confusion. Dismay causes the attitudes of sadness, fear, anxiety, or anger. See examples of how these maneuvers are used against Russian opposition in Fig. 8.

Positive narrative maneuvers such as Explain, Enhance, Excite, and Engage are used by the bots to affect the narrative about Alexei Navalny. According to the BEND framework theory, a positive approach can advance disinformation diffusion. Explain is used to provide more details and context, while Enhance covers the views of others and provides additional information about the narrative. Excite is used to attract the audience through positivity, happiness, and joy, while Engage provides more arguments for better associations with the narrative. See examples of these maneuvers implemented by bots against Alexei Navalny in Fig. 9.

Negative community maneuvers such as Neutralize, Nuke, Narrow, and Neglect target certain actors in order to reduce their impact on the conversation. The effect of Neutralize maneuver diminishes the impact of a particular opinion leader. Nuke is a maneuver that is used to split the community, targeting the key members and disabling their activities. Narrow maneuver is implemented with the purpose to polarize communities, isolate groups and break connections between them while Neglect is the maneuver that is used to undermine and downplay the topic. See examples of how these maneuvers are used by bots against Alexey Navalny in Fig. 10.

Positive community maneuvers such as Build, Back, Boost, and Bridge are used to strengthen connections between actors in a community network. Bots use Build maneuver to create communities and link influential actors while Back maneuver is used for support and better promotion of the opinions. The Boost maneuver is used to increase and enhance the connections between network actors and Bridge maneuver serves for adding linkages between various groups and communities. See how these maneuvers are used by bots against Navalny and Russian opposition in Fig. 11.

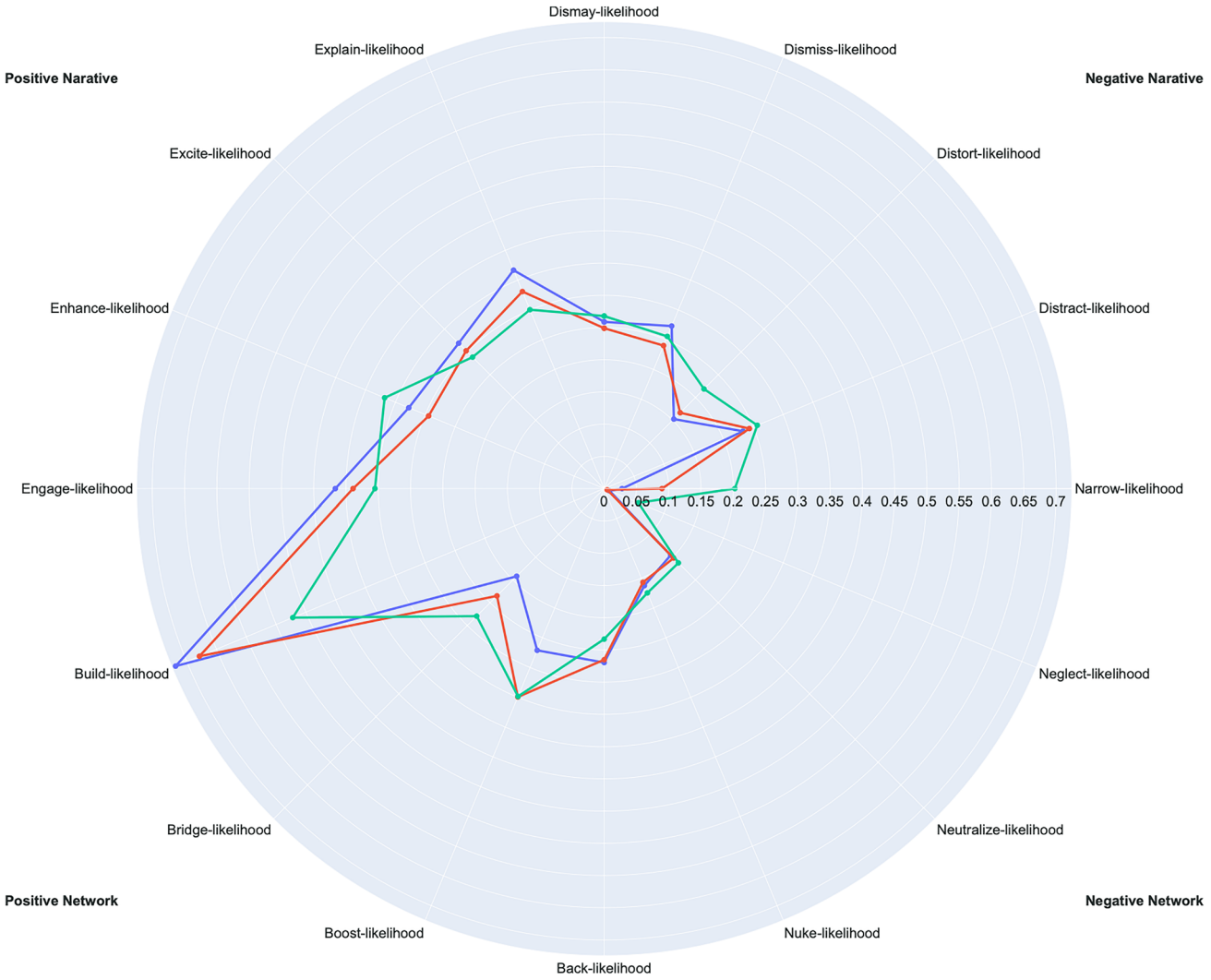

Fig. 7 This figure provides a comparison of the BEND maneuver average likelihood for each of the three highly influential communities found in our dataset

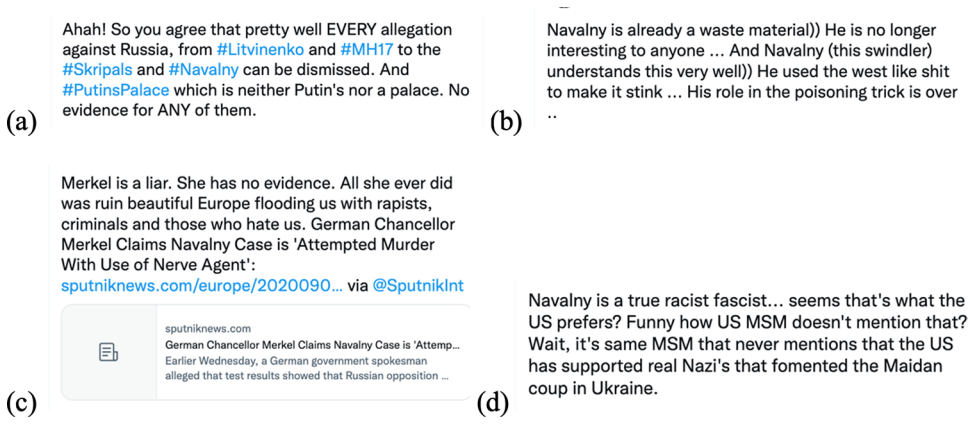

Fig. 8 Dismiss **a**, Distract **b**, Distort **c**, and Dismay **d** bot maneuvers in a Twitter discourse about Navalny

As a result, to answer our research questions we found influential online communities and disinformation actors in a Twitter discussion about Alexei Navalny. In addition, we have also detected a group of highly influential bots communicating with each other, retweeting messages from each other, trying to amplify the same disinformation

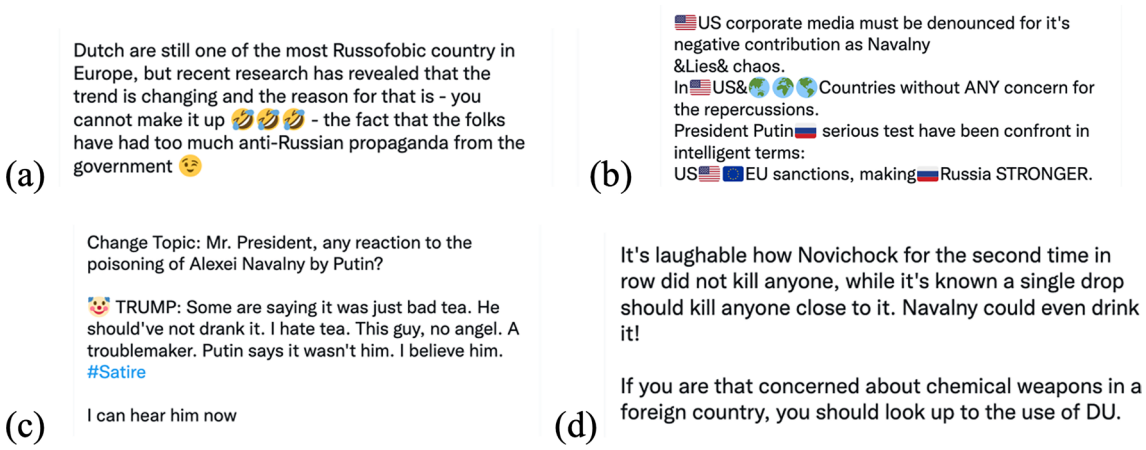

Fig. 9 Explain **a**, Enhance **b**, Excite **c**, and Engage **d** bot maneuvers in a Twitter conversation about Navalny

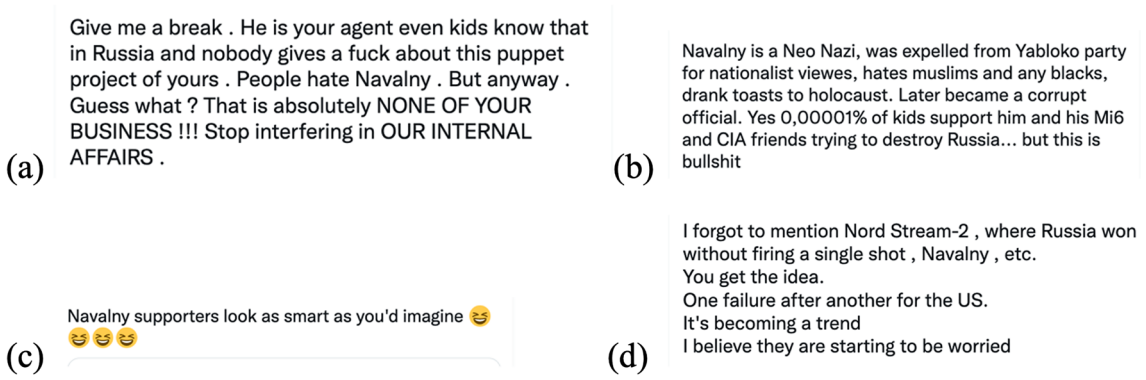

Fig. 10 Neutralize **a**, Nuke **b**, Narrow **c**, and Neglect **d** bot maneuvers in a Twitter conversation about Navalny

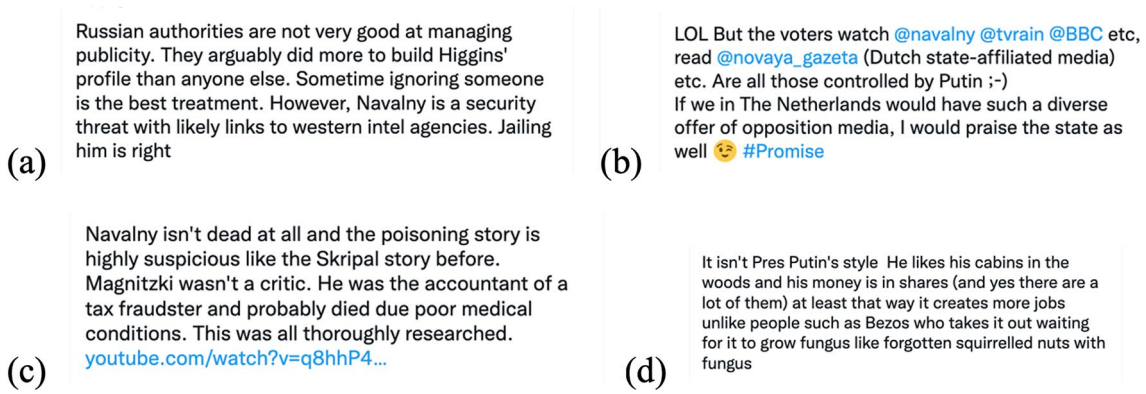

Fig. 11 Build **a**, Back **b**, Boost **c**, and Bridge **d** bot maneuvers in a Twitter conversation about Navalny

messages, and engaging a broad international audience with their content. We have identified frames and goals with the 16 BEND maneuvers in the disinformation messages spread by the malicious actors in order to distort the discourse about Alexei Navalny.

5 Conclusion

We have identified bot and troll activities related to the conversations about Alexei Navalny and the opposition movement in Russia as well as investigated promoted

narratives using community detection analysis and network science scores. We have also implemented BEND framework analysis to provide an overview of various persuasion techniques used by bots in this information operation. Our findings have demonstrated that there is a significant presence of bot activities in information operations against Alexei Navalny as one of the leaders of the Russian opposition. We have observed how the Russian domestic issue is framed in the context of Russian confrontation with the West and how it is used to promote hostile narratives against Navalny, an opposition movement, or democratic values. Many agents that we have identified pretend to be English speakers, who exhibit hostile attitudes towards Navalny and the Western democracies, express skepticism, distort the facts, promote a lack of trust in the democratic institutions as well as spread disinformation and conspiracy theories. The appearance of those various bot accounts is characterized by short, standard bios, random pictures as profile photos, and geolocation abroad to mimic a foreign citizen. Often, they exploit the democratic values and institutions or focus on the issues from the discourse of Western democracies, further intensifying political polarization, promoting a lack of trust in politicians, government, and the media by amplifying negative sentiments among the English-speaking users. Certain bot messages demonstrate a possible trace of the Russian state-sponsored campaign since they intended to improve the image of the Russian government and undermine Navalny as the primary opponent. We see that the strategy of choosing the fake personality of a foreigner is selected to pretend to be unbiased about the political situation in Russia and make the propagated messages to be seen as more truthful and accurate.

This study expands previous research about the activities of malicious actors online with an additional case study that shows the attempts to manipulate the attitudes of foreign audiences using the discourse about Alexei Navalny. Our analysis demonstrates that bots posted three times more messages. It illustrates the significant domination of bot information campaigns in the overall discourse on Twitter. We have identified several types of bots, such as bots with pro-opposition attitudes, bots that tweet in a news agency style, and anti-Navalny bots that spark conspiracy discussions and distorted information and facts. These tactics are used to create more noise and distract people's attention. While those messages may not significantly affect people with strong pre-existing attitudes about Navalny and politics, they could be influential for foreigners whoare unfamiliar with Russia's political situation and could be easily persuaded by distorted arguments. The bot disinformation messages have the potential to strengthen anti-government attitudes, polarizing the community on various political issues such as trust in the media, government, and democratic institutions. We have also observed bot accounts with the liberal and conservative focus, which used Navalny's case to manipulate the political agenda in the United States. Some of those accounts actively used polarizing language with the political context from various democratic countries such as the United States, Germany, the United Kingdom, the Netherlands, and others. We could see that certain bots were explicitly designed for a particular country to target the audience with pre-designed messages focused on the political discourse of that country.

We have identified three influential communities and analyzed the key influencers in each of them. As a result, we have selected one group with the highest bot activity and analyzed bot messages with the BEND framework. We found that a community with the highest number of anti-Navalny bots and trolls used more negative narrative and negative network maneuvers. As the main tactics, trolls spread disinformation, conspiracy theories, and strategic negative narratives about Western democracies, as well as use accusations of collusion with foreign governments. Certain messages were spread not only to diminish opposition but also to strengthen support for the Russian government, belittle Western countries and democratic institutions, and induce polarization.

According to our results and scores, anti-Navalny bots were mainly present in the top influential users for several types of scores such as super friends, total degree centrality, and members of large k-core for reciprocal communication. Bots are also characterized by higher scores for high weighted degree in hashtag network and high in-degree centrality for mentions and quotes. For further research, more case studies related to various countries and contexts should be analyzed to improve the detection process for harmful information operations.

Acknowledgements This paper is the outgrowth of research in the Center for Computational Analysis of Social and Organizational Systems (CASOS) and the Center for Informed Democracy and Social-cybersecurity (IDeaS) at Carnegie Mellon University. This work was supported in part by both centers, the Knight Foundation, and the Office of Naval Research through the Minerva program N00014-17-1-2675. 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 Knight Foundation, the Office of Naval Research, or the U.S. government.

References

Badawy A, Addawood A, Lerman K, Ferrara E (2019) Characterizing the 2016 Russian IRA influence campaign. Soc Netw Anal Min 9:1–11

Bastos M, Farkas J (2019) "Donald Trump is my President!": the internet research agency propaganda machine. Soc Media Soc. https://doi.org/10.1177/2056305119865466

Bastos M, Mercea D (2018) The public accountability of social platforms: Lessons from a study on bots and trolls in the Brexit

campaign. Proc R Soc A: Math Phys Eng. https://doi.org/10.1098/rsta.2018.0003
Bengani P. (2020) As election looms, a network of mysterious "pink slime" local news outlets nearly triples in size. Columbia Journalism Review.
Benkler Y, Faris R, Roberts H (2018) Network propaganda: Manipulation, disinformation, and radicalization in American politics. Oxford University Press
Beskow DM, Carley KM (2018) Bot-hunter: a tiered approach to detecting & characterizing automated activity on twitter. Conference paper. SBP-BRiMS: International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation 3:3.
Beskow DM, Carley KM (2019) Social cybersecurity: an emerging national security requirement. Mil Rev 99(2):117–127
Blane J, Bellutta D, Carley KM (2022) Social-Cyber maneuvers analysis during the COVID-19 vaccine initial rollout. J Med Internet Res. https://doi.org/10.2196/34040
Blondel VD, Guillaume J, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. J Stat Mech: Theory Exp. https://doi.org/10.1088/1742-5468/2008/10/P10008
Bodrunova SS (2021) Information disorder practices in/by contemporary Russia. The Routledge Companion to Media Disinformation and Populism, Routledge, pp 279–289
Carley KM (2020) Social cybersecurity: an emerging science. Comput Math Organ Theory 26(4):365–381
Carley KM, Cervone G, Agarwal N, Liu H (2018a) Social cyber-security. International conference on social computing, behavioral-cultural modeling and prediction and behavior representation in modeling and simulation. Springer, Cham, pp 389–394
Carley LR, Reminga J, Carley KM (2018b) ORA & NetMapper. In: International conference on social computing, behavioral-cultural modeling and prediction and behavior representation in modeling and simulation. Springer (Vol. 3, No. 3.3, p. 7).
Carley KM (2014) ORA: A Toolkit for Dynamic Network Analysis and Visualization. In: Alhajj R and Rokne J (eds) Encyclopedia of Social Network Analysis and Mining, Springer.
Chen E, Chang H, Rao A, Lerman K, Cowan G, Ferrara E (2021) COVID-19 misinformation and the 2020 US presidential election. The Harvard Kennedy School Misinformation Review.
Dawson A, Innes M (2019) How Russia's internet research agency built its disinformation campaign. Political Q 90(2):245–256
Freelon D, Lokot T (2020) Russian Twitter disinformation campaigns reach across the American political spectrum. Harvard Kennedy School Misinformation Review, 1(1).
Glazunova S (2020) 'Four Populisms' of Alexey Navalny: an analysis of Russian non-systemic opposition discourse on Youtube. Media and Commun 8(4):121–132
Golovchenko Y, Buntain C, Eady G, Brown MA, Tucker JA (2020) Cross-platform state propaganda: Russian trolls on Twitter and YouTube during the 2016 US presidential election. Int J Press/politics 25(3):357–389
Hagen L, Neely S, Keller TE, Scharf R, Vasquez FE (2020) Rise of the machines? Examining the influence of social bots on a political discussion network. Soc Sci Comput Rev. https://doi.org/10.1177/0894439320908190
Hallin DC, Mancini P (2011) Comparing media systems beyond the Western world. Cambridge University Press
Hemánus P (1974) Propaganda and indoctrination; a tentative concept analysis. Gazette (leiden, Netherlands) 20(4):215–223
Jowett GS, O'Donnell V (2014) Propaganda & Persuasion. Sage, London
Kazun A (2019) To cover or not to cover: Alexei Navalny in Russian media. Int Area Stud Rev 22(4):312–326
Kiriya I (2019) New and old institutions within the Russian media system. Russ J Commun 11(1):6–21
Kiriya I (2021) From "troll factories" to "littering the information space": control strategies over the Russian internet. Media Commun 9(4):16–26
Larson EV, Darilek RE, Gibran D, Nichiporuk B, Richardson A, Schwartz LH, Thurston CQ (2009) Foundations of effective influence operations: a framework for enhancing army capabilities. RAND ARROYO CENTER Santa Monica, CA.
Linvill DL, Warren PL (2020) Troll factories: manufacturing specialized disinformation on Twitter. Pol Commun 37(4):447–467
Lipman M, Kachkaeva A, Poyker M (2018) Media in Russia: Between modernization and monopoly. In: Treisman D (ed) The new autocracy: information, politics, and policy in Putin's Russia. Brookings Institution Press, Washington, D.C, pp 159–190
Lucas E, Nimmo B (2015) Information warfare: What Is It and How to Win It. CEPA Infowar Paper.
Lukito J (2020) Coordinating a multi-platform disinformation campaign: internet research agency activity on three US social media platforms, 2015 to 2017. Pol Commun 37(2):238–255
Lukito J, Suk J, Zhang Y, Doroshenko L, Kim SJ, Su MH, Wells C (2020) The wolves in sheep's clothing: How Russia's Internet Research Agency tweets appeared in US news as vox populi. Int J Press/politics 25(2):196–216
Ng LHX, Robertson DC, Carley KM (2022) Stabilizing a supervised bot detection algorithm: How much data is needed for consistent predictions? Online Soc Net Media 28:100198
Nyst C, Monaco N (2018) How governments are deploying disinformation as part of broader digital harassment campaigns. Institute for the Future. ly/2Mi8DYm.
Peng J, Detchon S, Choo KKR, Ashman H (2017) Astroturfing detection in social media: a binary n-gram–based approach. Concurr Comput Pract Exp 29(17):1–14
Sanovich S, Stukal D, Tucker JA (2018) Turning the virtual tables: Government strategies for addressing online opposition with an application to Russia. Comp Polit 50(3):435–482
Stukal D, Sanovich S, Tucker JA, Bonneau R (2019) For whom the bot tolls: a neural networks approach to measuring political orientation of Twitter bots in Russia. SAGE Open 9(2):2158244019827715
Summers E (2022) Twarc 2.9.4. The python package index. https://pypi.org/project/twarc/
Tsyrenzhapova D, Woolley SC (2021) The evolution of computational propaganda: Theories, debates, and innovation of the Russian model. In: The Routledge Companion to Media Disinformation and Populism (pp. 121–130). Routledge.
Uyheng J, Magelinski T, Villa-Cox R, Sowa C, Carley KM (2020) Interoperable pipelines for social cyber-security: assessing Twitter information operations during NATO Trident Juncture 2018. Comput Math Organ Theory 26(4):465–483
Uyheng J, Carley KM (2019) Characterizing bot networks on Twitter: An empirical analysis of contentious issues in the Asia-Pacific. In International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation (pp 153–162). Springer, Cham.
Weedon J, Nuland W, Stamos A (2017) Information operations and facebook. Menlo Park, CA: Facebook.
Weiss M (2013) Rights in Russia: Navalny and the opposition. World Affs 176:72
Wells C, Shah D, Lukito J, Pelled A, Pevehouse JC, Yang J (2020) Trump, Twitter, and news media responsiveness: a media systems approach. New Media Soc 22(4):659–682
Woolley SC (2020) Bots and computational propaganda: automation for communication and control. Social media and democracy. In: Persily N and Tucker JA (eds) The state of the field, prospects for reform (pp 89–110).Cambridge University Press.