



ISBN: 978-9934-564-97-0

Project manager: Rolf Fredheim

Authors: Kristina Van Sant, Rolf Fredheim, and Gundars Bergmanis-Korāts

Copy-editing: Anna Reynolds

Design: Kārlis Ulmanis

Riga, February 2021

NATO STRATCOM COE

11b Kalnciema iela

Riga LV1048, Latvia

[www.stratcomcoe.org](http://www.stratcomcoe.org)

[Facebook/stratcomcoe](https://www.facebook.com/stratcomcoe)

[Twitter: @stratcomcoe](https://twitter.com/stratcomcoe)

This report was completed in November 2020, and based on data collected from March to July 2020

This publication does not represent the opinions or policies of NATO or NATO StratCom COE.

© All rights reserved by the NATO StratCom COE. Reports may not be copied, reproduced, distributed or publicly displayed without reference to the NATO StratCom COE. The views expressed here do not represent the views of NATO.

PAGE LEFT  
BLANK  
INTENTIONALLY

” The main topics triggering abusive messages were the COVID-19 pandemic, immigration, Finnish-EU relations, and socially liberal politics.

## Executive Summary

This report is an explorative analysis of abusive messages targeting Finnish ministers on the social media platform Twitter. The purpose of this study is to understand the scope of politically motivated abusive language on Finnish Twitter, and to determine if, and to what extent, it is perpetrated by inauthentic accounts. To this end, we developed a mixed methodology, combining AI-driven quantitative visualisations of the networks delivering messages of abuse with a qualitative analysis of the messages in order to understand the themes and triggers of abusive activity. We collected Twitter data between 12 March and 27 July 2020,

a period spanning the state of emergency declared in response to the COVID-19 pandemic.<sup>1</sup>

This report is informed by the findings of three recent Finnish studies, one of which investigated the extent and effects of online hate speech against politicians while the other two studied the use of bots to influence political discourse during the 2019 Finnish parliamentary elections. The first study, released by the research branch of the Finnish government in November 2019, found that a third of municipal decision-makers and nearly half of all members of Finnish Parliament have been subjected to hate speech online.

---

<sup>1</sup> Finland declared a state of emergency on 16 March 2020 that was in force for three months until 16 June 2020, although the Emergency Powers Act remained in force through the end of June.



The two studies tracking inauthentic activity during the 2019 parliamentary elections identified bot interference but concluded that the impact of these bots on Finland's political environment appeared limited. Based on these findings, and on our comprehensive literature review, we developed two hypotheses:

1. We expect to observe abusive language targeting Finnish politicians, with female politicians receiving gendered abuse;
2. We expect to observe low levels of coordinated inauthentic activity in the Finnish information space, with increased levels of inauthentic activity during periods of political significance.

Our quantitative and qualitative analyses confirmed both hypotheses and yielded multiple findings. Our investigation demonstrated that the messaging directed at Finnish government officials is largely free from automated activity. When it comes to abusive messaging, we find a number of users singularly focused on harassing the government. While both left- and right-leaning communities engaged in abusive activity, the bulk of abusive messaging originated from clusters of right-wing accounts.

Overall, we observed very low levels of both bot and coordinated activity. The majority of bots we identified were operating in foreign

languages and either not generally focused on Finland or used to push certain causes in multiple languages. We repeatedly came across a cluster of accounts throughout our monitoring period that posted the same messages about animal cruelty and climate change. These accounts predominantly post in English and appear in some cases to be automated or semi-automated. However, they represent a very small part of the conversation. Likewise, a small cluster of automated accounts amplified messaging by a number of right-wing voices. Again, there was a degree of coordination here, but these amplifications looked more like attempts at self-promotion rather than systematic manipulation of the information space. If large-scale inauthentic coordination exists in the Finnish information environment, we are either looking in the wrong place, or it is so sophisticated or so small in scale that it evades our detection methods.

We found that the main topics triggering abusive messages were the COVID-19 pandemic, issues of immigration, Finnish-EU relations, and socially liberal politics. We observed that female Finnish ministers received a disproportionate number of abusive messages throughout our monitoring period. A startling portion of this abuse contained both latent and overtly sexist language, as well as sexually explicit language. Although we found large volumes of offensive and abusive messaging, we did not observe threats of physical violence.



” Social media has become an essential platform for political engagement, granting citizens unprecedented access to their government representatives.

## Introduction

Lipstick brigade. Lipstick girls. Feminist quintet. Tampax team. These are all phrases used on Twitter to refer to the current coalition in Finland, in which all five party leaders are women, led by Prime Minister Sanna Marin of the Social Democratic Party. When the remarkably young and female leadership came into power in December 2019, they made international headlines as pioneers of gender equality in governance. Their election also provoked online resistance in the form of abusive messages. Many assumptions about their political inexperience were accompanied by sexist and misogynistic language.

Social media platforms provide individuals with virtually limitless opportunities for communication and self-expression. This potential, though transformative, has raised

challenges for states navigating the complex relationship between freedom of speech and protection from harmful discourse, as online hate speech and abusive messaging have become increasingly recognised as socio-political issues. Social media has become an essential platform for political engagement, granting citizens unprecedented access to their government representatives. Twitter in particular has provided candidates and constituents with an informal channel of communication, through which citizens can share feedback and politicians have the ability to engage with these concerns directly.

However, this unfettered access to politicians online, combined with the anonymous nature of social media platforms, has led to government officials being targeted with



abusive messages. This virtual vitriol can take many forms: it can be threatening, misogynistic, racist, vulgar, and so on. For governments, online harassment is a growing concern, as it can have the effect of discouraging participation in public service, particularly among women. Simultaneously, the rise in fake account activity in online political discourse is equally concerning, as recent examples highlight the impact inauthentic activity can have on public opinion and political participation.

This study is an analysis of how abusive messaging intersects with the activity of fake Twitter accounts in the political sphere. In this explorative analysis, we will be focusing on the state of politically motivated online abuse in Finland. Specifically, we will be analysing messages directed at Finnish ministers between 12 March and 27 July 2020, encompassing the three months Finland maintained a state of emergency

in response to the COVID-19 pandemic and several weeks after it was lifted.

The report is structured as follows. The literature review engages with the scholarly literature discussing definitions and methods of detecting abusive language on social media platforms, abuse of politicians online, misogyny online, and the use of bots for political purposes on Twitter. Having established this framework, we will describe our methodological approach for analysing the data, which combines social network analysis, bot detection, hate speech detection, and narrative analysis. This combination of quantitative and qualitative approaches is designed to identify instances when accounts coordinate to send abusive messages to politicians. The study continues with a social network analysis that informs the basis of the qualitative analysis. We conclude our study with a discussion of our findings, conclusions, and policy recommendations.



# The Proliferation of Online Abuse

## Defining hate speech and abusive language

The concept of **hate speech**, considered an umbrella term for abusive user-created content, does not have a single formal definition. Rather, hate speech is interpreted as a collection of overlapping terms, including designations such as cyberbullying, abusive language, and hostile language

(Waseem et al, 2017) and is the most frequently used phrase for describing the phenomenon of insulting user-generated content (Schmidt and Wiegand, 2017). Broadly, hate speech is defined as any communication that disparages a person or group on the basis of a particular trait, such as race, color, ethnicity, gender, sexual orientation, nationality, or religion, among other characteristics (Nockleby, 2000).

	Explicit	Implicit
Directed	<p>"Go kill yourself", "You're a sad little f*ck" (<i>Van</i>)</p> <p>"@User shut yo beaner ass up sp*c and hop your f*ggot ass back across the border little n*gga" (<i>Davidson et al., 2017</i>),</p> <p>'Youre one of the ugliest b*tches Ive ever fucking seen" (<i>Kontostathis et al., 2013</i>).</p>	<p>"Hey Brendan, you look gorgeous today. What beauty salon did you visit?" (<i>Dinakar et al., 2012</i>),</p> <p>"(((@User))) and what is your job? Writing cuck articles and slurping Google balls? #Dumbgoogles" (<i>Hine et al., 2017</i>),</p> <p>"you're intelligence is so breathtaking!!!!!!" (<i>Dinkar et al., 2011</i>).</p>
Generalised	<p>"I am surprised they reported on this crap who cares about another dead n*gger?", "300 missiles are cool! Love to see um launched into Tel Aviv! Kill all the g*ys there!" (<i>Nobata et al., 2016</i>),</p> <p>"So an 11 year old n*gger girl killed herself over my tweets? ^_^ that's another n*gger off the streets!!" (<i>Kwok and Wang, 2013</i>).</p>	<p>"Totally fed up with the way this country has turned into a haven for terrorists. Send them all back home." (<i>Burnap and Williams, 2015</i>),</p> <p>"most of them come north and are good at just mowing lawns" (<i>Dinakar et al., 2011</i>),</p> <p>"Gas the skypes" (<i>Magu et al., 2017</i>).</p>

Table 1: Typology of abusive language (Waseem et al, 2017)



Davidson et al (2017) differentiate hate speech from offensive language, defining it as, “language that is used to express hatred towards a targeted group or is intended to be derogatory, to humiliate, or to insult the members of the group.” In an attempt to synthesise varying definitions of hate speech, Waseem et al (2017) proposed a typology establishing a distinction between explicit and implicit abusive language. Explicit abuse is clearly derogatory, including language that contains racist or sexist slurs, while implicit abuse is less overt, often obscured by the use of sarcasm and other ambiguous language, making it more difficult to detect qualitatively and with machine learning approaches. When discussing **abusive language** in this report, we will refer to this typology (Table 1).

## Targets of online abuse

Existing studies on hate speech have typically focused on specific forms of abuse, such as racism and homophobia, as well as on rhetoric shared by hate groups and content circulated on radical forums (Silva et al, 2016). For the purposes of this study, we will discuss literature that examines hate speech directed at politicians, misogynistic hate speech, and abuse targeting female politicians.

### Political abuse

The scholarly debate surrounding hate speech and politicians is dominated by studies investigating the employment

of hate speech as a political tool. Accusations and prosecution of politicians for hate speech is an established reality in modern democracies, especially with the rise of anti-immigration parties in Western Europe in recent decades (van Spenje and de Vreese, 2013). Extreme-right political parties in Spain were found to imply discrimination on their Facebook pages, which was then exacerbated by their followers who used hate speech in the comment sections (Ben-David and Matamoros-Fernandez, 2016). But what is the state of abuse directed at politicians? Little academic work exists regarding the extent of abusive messages addressed to politicians online (Gorrell et al, 2018).

Social media has become an essential platform for political engagement, voter mobilisation, electoral campaigning, and intimate communication between political candidates and the public. Despite the opportunity for individual interaction, Theocharis et al (2016) found that politicians prefer to engage in broadcasting-style communication. The authors hypothesised that this is the case because politicians are reluctant to invite the vitriol of citizens empowered by the anonymity of Twitter. Their results lend support to this theory, as they found that candidates with more engaging messages are also more exposed to criticism and harassment online (Theocharis et al, 2016). Ward and McLoughlin (2020) identify four explanatory themes for abuse of British MPs: mental illness among members of the public;



the nature of social media; the increase in political polarisation and extremism; and social problems around identity issues, such as race, religion, and gender.

A study released by the research branch of the Finnish government in November 2019 explored the nature and extent of hate speech targeting Finnish politicians. The study, which constituted the first in Finland on how societal decision-making may be influenced by hate speech, found that a third of municipal decision-makers and nearly half of all members of Finnish Parliament have been subjected to hate speech online. Additionally, two-thirds of policymakers surveyed believe that hate speech has increased in recent years. Hate speech directed at public servants may have a negative impact on political participation, as 28% of municipal officials who were targeted with hate speech expressed that the experience decreased their willingness to participate in decision-making (Knuutila et al, 2019). The findings of the Finnish government provide the foundational basis for this project.

#### Online misogyny

Research has repeatedly found that women are subjected to more online abuse, bullying, hateful language, and threats than men (Bartlett et al, 2014). A 2017 survey by the US-based Pew Research Center found that women are much more likely to experience severe types of gender-based or sexual harassment than men. In fact, 21% of women aged 18

to 29 reported being sexually harassed online, a figure that is more than twice the percentage of men in the same age bracket (Pew Research Center, 2017). Between December 2016 and March 2018, Amnesty International conducted qualitative and quantitative research into women's experiences on social media platforms. During their investigation, the authors found that Twitter fosters a toxic and unregulated underbelly of violence and abuse against women (Amnesty International, 2018).

Mantilla (2013) distinguishes "gendertrolling", the targeting of women with abuse online, from generic trolling. Gendertrolling is defined by the following features: gender-based insults; vicious language; credible threats; the participation, often coordinated, of numerous people; unusual intensity, scope, and longevity of attacks; and reaction to women speaking out (Mantilla, 2013: 564–65). Mantilla argues that gendertrolling systematically targets women to prevent them from fully occupying public spaces, particularly traditionally male-dominated arenas (ibid., 569).

A 2016 Inter-Parliamentary Union (IPU) report discusses the three characteristics that distinguish violence against women in politics: (1) Women are targeted because of their gender; (2) the abuse itself can be highly gendered, as exemplified by sexist abuse and threats of sexual violence; (3) its impact is to discourage women from



” The targeting of women with gender-based abuse online, particularly women in positions of power, has become a global phenomenon.

becoming or continuing to be active in politics. The IPU identified a number of factors that exacerbate the vulnerability of certain women parliamentarians to gender-based abuse. These aggravating factors include belonging to the political opposition, being under 40 years old, and belonging to a minority group, where sexism is often compounded by racism. Alarming, the IPU found that this phenomenon exists to varying degrees in every country (Inter-Parliamentary Union, 2016).

In 2017, Amnesty International carried out a small-scale investigation into sexist and racist abuse faced by women in UK politics on Twitter. In the run-up to the 2017 election, Amnesty International researchers found that Labour MP Diane Abbott received almost half of all abusive tweets and that black and Asian women MPs received 35% more abusive tweets than white women. Notably, online abuse did not adhere to party lines: women from all UK political parties were targeted by sexist hate speech (Dhrodia, 2017). Southern and Harmer (2019) conducted a comparative study of the experiences of less prominent male and female MPs online in which they

found that female MPs were more likely to receive tweets that stereotyped them or questioned their position as government representatives. While they identified clearly gendered patterns in the data, the authors concluded that there are fewer differences in how female and male politicians are addressed than previously expected. In their explorative analysis of instances of sexist hate speech and abusive language against female politicians on Twitter in Japan, Fuchs and Schäfer (2020) found that negative attitudes expressed towards female politicians have become a common trend on the platform, especially towards controversial or more prominent female politicians.

The targeting of women with gender-based abuse online, particularly women in positions of power, has become a global phenomenon. The anonymous nature of social media platforms, such as Twitter, has empowered individuals to engage in abusive discourse online. Recent studies have identified patterns in gendered abuse on social media and have attempted to explain their ubiquitous prevalence. The possible impacts of sexist abuse against



women politicians are vast, ranging from psychological distress to discouragement from participating in public service. Given that 11 of the 19 ministers appointed to the current Finnish government are women, we expect to observe gender-based abuse targeting these ministers and the functioning of the government as a whole.

**H1** We expect to observe abusive language targeting Finnish politicians, with female

## Automated hate speech detection

As the volume of user-generated content on social media platforms has surged, so has the amount of hate speech circulating online and, consequently, the demand for hate speech detection tools (Schmidt and Wiegand, 2017). However, both manual and automated detection methods are hindered by the lack of a clear definition of hate speech and the often-ambiguous nature of verbal abuse. Previous studies have identified hateful and antagonistic content through various qualitative and quantitative methodological approaches. Among them is the bag-of-words (BoW) approach, a technique of natural language processing that uses words within a corpus to classify hate speech but is prone to misclassification (Greevy and Smeaton, 2004). In addition to keywords, Silva et al (2016) leveraged the

sentence structure of derogatory social media posts to identify instances of hate speech with a high rate of precision. In their article, Burnap and Williams (2015) developed a supervised machine learning classifier of hateful content on Twitter for the purpose of monitoring the public reaction to highly emotive events, such as a terror attack.

Fuchs and Schäfer (2020) adopted corpus-based discourse analysis (CDA), a mixed-methods approach that combines critical discourse analysis with corpus linguistics, particularly keyword analysis based on frequency and occurrence patterns. In order to investigate misogynistic hate speech, Hewitt et al (2016) gathered thousands of tweets using a range of sexist terms, disregarding irrelevant commercial messages, messages in foreign languages, or completely unintelligible tweets, and manually coded the remaining sample using a simple binary model. Badjatiya et al (2017) applied deep learning architectures to the problem of identifying hate speech on Twitter, which they define as being able to classify a tweet as racist, sexist, or neither. Despite their variance, most techniques for detecting hate speech on social media platforms—namely Twitter—incorporate machine-learning-driven data collection and identification algorithms with in-depth qualitative analysis to address the shortcomings of automated identification and enhance understanding of the content of hateful messages.





## Social Media Companies Struggle to Keep up with Hate Speech

Companies such as Microsoft, Twitter, YouTube, and Facebook have signed the EU's *Code of Conduct on Countering Illegal Hate Speech Online* (European Commission, 2020), which compels the companies to remove any post containing hate speech within 24 hours. Despite the huge resources and data availability of these tech giants, social media companies to this day find it hard to automate hate speech detection. One indication of this is their decision to publicly release datasets to allow the general public to contribute to the challenge (Vidgen and Derczynski, 2020; Davidson et al, 2017; de Gibert et al, 2018; Ahlgren et al, 2020).

Recently Facebook has invested heavily in measures to control toxic content and inauthentic behaviour of platform users (Ahlgren et al, 2020). As an example, Facebook previously mainly outsourced its content moderation to a relatively small group of reviewers around the world. However, the volume of potentially abusive messaging was such that Facebook is currently investing heavily in Artificial Intelligence solutions to automate the process. Despite recent advancements in AI, the current level of development is far behind necessary efficiency levels as it must be able to understand content holistically (the way we perceive).

To address this problem and to boost AI development in this direction, Facebook launched a hate speech challenge to detect meme-based political hate speech using a data set of 10,000+ new multimodal examples (Kiela et al, 2020). Facebook recently claimed that its current AI solution is able

to proactively detect 88.8% of the total hate speech content they remove, up from 80.2% the previous quarter thanks to progress in two key areas: deeper semantic understanding of language (the detection of more subtle and complex meanings) and broadening the capacity of AI tools to understand content (holistic understanding of content) (Dansby et al, 2020). Even though almost 90% sounds impressive, this is relatively low compared to simpler classification tasks (for instance automatically detecting pornography), and certainly much too low to hand the process over to the machines entirely. In the case of content moderation, the ~10% of data missed may reach and harm a significant number of platform users, especially considering that Facebook currently has 2.7 billion active users worldwide (Clement, 2020). In this case, the emphasis should not be on the number of harmful posts removed, but on the total number of missed posts that may potentially cause harm.

# Automation and Political Contestation Online

## The role of bots and trolls in political discourse

**Bots**—short for software robots—are computer programs that perform tasks automatically and have been a staple of online activity since the advent of computers. While much bot activity is benign, they can also be utilised by economically or politically motivated actors to carry out malicious activities such as launching spam, distributed denial-of-service (DDoS) attacks, click fraud, cyberwarfare (Kollanyi et al, 2016), and the manipulation of public opinion (Woolley and Howard, 2016). The label **troll** describes persons who start arguments online to elicit an emotional response, typically outrage. This practice is widely known as **trolling**. Political trolling and astroturfing are terms that refer to the artificial promotion of messages online in order to manufacture a false sense of popularity or support for these messages (Bradshaw and Howard, 2018).

As contemporary social media ecosystems have increased in scope and complexity, the demand for bots that convincingly mimic human behaviour has risen in parallel. **Social bots**, scripts designed to produce content and to interact with

humans over a diverse range of social media platforms, have been used to infiltrate political discourse, manipulate the stock market, steal personal information, and spread disinformation (Ferrara et al, 2016). **Political bots** are social bots used as tools for politics and propaganda, such as posting carefully staged photos and well-crafted responses in pursuit of political objectives (Howard et al, 2018). The use of bots can be classified as inauthentic behaviour. Facebook defines **inauthentic behaviour** as the use of accounts, pages, groups, or events to mislead social media platform users and the platforms themselves about:

- the identity, purpose, or origin of the entity that they represent;
- the popularity of content or assets;
- the purpose of an audience or community;
- the source or origin of content;

or to evade enforcement of a platform's Community Standards.

Furthermore, **coordinated inauthentic activity** is defined by Facebook as the use of multiple assets, working in concert to engage in inauthentic behaviour, where



the use of fake accounts is central to the operation (Facebook, 2020).

Social media companies themselves estimate that about 5–8% of accounts are fake or bot accounts, but scholars tend to see these estimates as conservative. For instance, Varol et al (2017) estimate that bot accounts make up 9–15% of all Twitter users. However, studies of non-English data sets have pointed to even higher bot concentrations. Filer and Fredheim (2015) found that Russian-language discussions are at times conducted virtually exclusively by bots. In their 2016 Bot Traffic Report, the security company Imperva estimated that over half of online traffic was attributed to bots, nearly 30% of which were categorised as ‘bad bots,’ such as impersonators and spammers.

In recent years, Facebook, Twitter, and other social media platforms have received intense criticism for their treatment of users and their personal data, their insufficient responses to pervasive issues such as hate speech, the dissemination of mis- and disinformation by bots, and their systematic evasion of critical and independent oversight (Bruns, 2019). Ünver (2019) argues that the debate at the heart of the bot problem is whether technology companies are deliberately, or at least passively, facilitating negative political messaging. Despite strengthening their efforts against inauthentic activity, social media platforms have not been able to get malicious bot activity under control (Bay and Fredheim, 2019).

## Political use of bots on Twitter

Recent academic studies are replete with instances where social bots have been used as instruments to carry out political campaigns or to shape the political conversation on Twitter. Wooley (2016) identified three main ways political bots have been employed: to demobilise opposition, to disseminate pro-government messages, and to inflate follower counts. Political actors and governments use bots to manipulate public opinion, choke off debate, and muddy political issues (Howard and Kollanyi, 2017). Political bots can engage in weaponised communication—the strategic use of communication as an instrumental tool to gain compliance and avoid accountability (Mercieca, 2019). Such campaigns have been documented in Argentina, Australia, Azerbaijan, Bahrain, China, Iran, Italy, Mexico, Morocco, Russia, South Korea, Saudi Arabia, Turkey, the United Kingdom, the United States, and Venezuela (Wooley, 2016).

The two most widely studied cases of political bot interference both occurred in 2016, during the US presidential election and the UK-EU Referendum. In a study of the 2016 US presidential election debates, Kollanyi et al (2016) found one third of pro-Trump tweets were driven by bots and highly automated accounts, suggesting bots were used to amplify the popularity of pro-Trump messages on Twitter. The authors later found that political bot activity reached an



all-time high during the 2016 presidential campaign. By election day, the gap between highly automated pro-Trump and pro-Clinton messaging was 5:1, indicating a deliberate and strategic attempt to sway the outcome of the election (Kollanyi et al, 2016). Howard and Kollanyi (2016) found that bots generated a noticeable portion of all traffic surrounding the UK Brexit referendum. These messages predominantly supported the Vote Leave campaign.

Since 2016, the deployment of bots during important and divisive national and international political moments has become a staple of the 21st century information space. In 2019, during the Hong Kong protests, Twitter and Facebook took action against China for using hundreds of fake accounts to sow political discord. Twitter announced that it was suspending nearly a thousand Chinese accounts, citing a "significant state-backed information operation" (Baca and Romm, 2019). Similarly, during mass demonstrations seeking political change in Lebanon, it was discovered that accounts tweeting a pro-government hashtag had a higher likelihood of displaying automated behaviour (Skinner, 2019).

In Finland, several Aalto University studies have investigated the use of bots to influence political discourse during the 2019 Finnish parliamentary elections. Rossi (2019) developed a machine-learning tool for detecting bots on Twitter to analyse whether inauthentic accounts were used to influence voters. According to this model, over 30% of followers of Finland's top politicians were classified as bots, indicating that they may have been used to increase the Twitter following of certain politicians. However, apart from artificial follower inflation, the impact of these bots on Finland's political environment appears limited. Researchers working on the ELEBOT project funded by the Finnish Ministry of Justice concluded that the volume of Twitter-bots and their influence on the Finnish elections were minimal. The researchers visualised communities whose members were linked and identified the three largest communities. The smallest of the three communities, which included around 7% of all users and 8% of all bots, primarily used language relating to immigration and to the right-wing populist Finns Party. Overall, authentic Twitter users rarely engaged with tweets produced by bots (Salloum et al, 2019).



**H2** We expect to observe low levels of coordinated inauthentic activity in the Finnish information space, with increased levels of inauthentic activity during periods of political significance.

## Methodological Approach

For the purposes of this report, we used data collected from Twitter. From a technical perspective, this decision is easily justifiable: other social media platforms are increasingly closed to research of this type. Additionally, Twitter is a platform on which it is easy to engage directly and publicly with people outside of mutual friend relationships. However, one drawback of using Twitter data is the platform's relative unpopularity in Finland. According to the We Are Social #Digital2020 report for Finland, 95% of the total population is active on the internet. Based on data from SimilarWeb, the three most-visited websites are Google, YouTube, and Facebook. There are approximately 3.3 million active social media users in Finland, 773,000 of whom can reportedly be reached with adverts on Twitter (Kemp, 2020). According to Media Landscapes, the most popular social networks used weekly for news are Facebook (34%), YouTube (9%), Twitter (6%), and Suomi24 (5%) (Jyrkiainen, 2017).

We began monitoring Finnish activity on Twitter on 12 March 2020, a few days

before Prime Minister Marin declared a state of emergency in Finland due to the COVID-19 pandemic. Monitoring continued until 27 July 2020, encompassing the entirety of the state of emergency and the period immediately after the declaration was lifted.

For this research we used a range of methods to attempt to infer coordinated inauthentic behaviour from observational data:

1. **Bot detection**—useful for finding possible inauthentic behaviour.
2. **Community detection using social network analysis**—useful for identifying clusters of users with similar behavioural patterns.
3. **Narrative estimation**—identifying the subjects of conversation.
4. **Abusive language detection**—identifying possibly harmful material.
5. **Timeline comparison**—identifying users with similar posting patterns.



While none of these methods directly identify coordinated behaviour, our assumption was that taken together, they would reveal pockets of similar activity, be that in terms of account type, community position, subject of messaging, or level of abusiveness. Such pockets of activity could through qualitative analysis be verified as examples of coordinated inauthentic behaviour. Should the analysis fail to identify clusters of inauthenticity, it might be because they are uncommon in the Finnish-language Twitter conversation, that the clusters of activity are very small (only two or three accounts), or because our choice of methods was inappropriate.

## Data collection

We collected messages on the social media platform Twitter mentioning one or more members of the Marin cabinet. Data collection was conducted using the Twitter Search API, which Twitter describes as focused on relevance rather than completeness. Consequently, one limitation of this study is that some messages were not captured during collection. The data collection script was run periodically during the observation period. As data was collected, we automatically anonymised all personal identifiable information.

After data collection we calculated the bot-likelihood for each account in the dataset. The algorithm used for this process was trained on data collected from 2017 to 2020

for our quarterly *Robotrolling* report, which monitors English- and Russian-language messaging about the NATO presence in the Baltics and Poland. The algorithm works by calculating how similar behaviour of new (unseen) users is to observed examples of automated activity. To make this calculation, it draws on the metadata provided by Twitter, and a range of calculated metrics. These include, for instance, proportion of retweets, frequency of posting, average number of posts per day, standard deviation of message posting, and so on.

## Abusive language detection

When analysing abusive messaging in the Finnish information space, the language barrier poses a challenge. One possible solution to this problem is to translate the dataset to English via machine translation, and then apply mainstream models optimised for the English language to the translated dataset. However, this approach would result in the loss of vital information, as translation services may distort the original text and its meaning. Therefore, in order to preserve the original text of abusive messages, we tried to avoid translation.

We decided to conduct hate speech analysis in the original Finnish language. To do so, we relied on the prepared dataset used in Knuutila et al (2019). This dataset contained approximately 2,000 tweets identified by the researchers as one of two classifications, either abusive or neutral. Using this data,



we trained a neural network to recognise unseen examples of both categories.

Bidirectional Encoder Representations from Transformers (BERT), a concept pioneered by Google AI, is a Transformer-based machine learning technique for natural language processing (NLP). It learns contextual relations between words in order to create a model of a language. Unlike so-called bag-of-words models where language is stripped of context and reduced to counts of signifiers, the BERT model includes contextual information to separate the meaning of otherwise identical words (e.g. the word 'model' could refer to a person in a photoshoot, it could be a verb, it could be a noun referring to an abstract object, it could denote the make of a car, etc.). In each of these cases, the word 'model' would be positioned differently within a vector space. This contextualised positioning in vector space allows the comparison and even generation of text. In the context of abusive language, this allows the researcher to determine that the phrase 'he is a total failure' is closer in meaning to 'what a moron!' than to 'a failure of leadership'. We use the model to assess whether novel sentences are in some way contextually similar to the examples of abusive language in the training data, even in cases when a different vocabulary is chosen.

The Finnish language BERT model we used was provided by *TurkuNLP*, a group of researchers at the University of Turku (Virtanen et al, 2019). The training and classification tasks were carried out using the Python library *ktrain* (Maiya, 2020) which supports model training and deployment from the Transformer framework—a Natural Language Processing (NLP) for *Pytorch* and *TensorFlow 2.0*. The model achieved excellent precision on the given training corpus even after only a few training epochs. We were concerned that it might be prone to overfitting, and unable to generalise to unseen data. However, after performance analysis and training we assessed that the model was able to capture the tweets that contain abusive language with swear words allowing us to conduct the quantitative analysis. Messages predominantly containing expletives yielded very high abuse probabilities (0.97–0.99). Less blunt examples of abuse were classified within the 0.6–1.0 probability range. In order to capture the abusive portion of the dataset we decided to implement a classification threshold of 0.6 to act as a filter. Nonetheless, it is worth noting that the model has a higher accuracy for classifying explicit abuse than for implicit abuse (see Table 1).





## NexPlore or NView: A multilingual narrative explorer tool

For our analysis we wanted to understand whether abusive language was more common for some topics than for others. To explore possible correlations between subject matter and hate speech, we developed a bespoke exploratory tool that implements state-of-the-art NLP text similarity search methods and algorithms. The narrative estimation tool incorporates the similarity search project, *FAISS* developed by Facebook AI research lab (Johnson et al, 2017), and *ScaNN* developed by Google scientists (Guo et al, 2020). The Facebook algorithm was implemented in order to find similarities of each tweet to predefined topics. The algorithm iterates through all individual tweets and computes the *FAISS* distance to all topics of interest. As a result, one might use post-processing for clustering or simply to visualize the results. Tweets found to be too distant can be extracted separately and used for additional NLP analysis to estimate new narratives not considered before. We found this approach particularly useful when graph analysis methods are used. Estimated distances of predefined narratives serve as weights and features when creating complex graphs and analysing communities. Another useful application is to monitor selected topics over time to follow the activity of users engaging in those topics. When superimposed with a news timeline one might follow and estimate the reactions of communities.

The Narrative estimation tool also offers search-engine type functionality for the complete dataset where a fast tweet search using *ScaNN* finds the top 100 best matching tweets corresponding to rather abstract search queries instead of searching for exact text matches. Speed is achieved using efficient search algorithms by incorporating item indexing similarly to the way it is done in databases. This rapidly gives analysts insight into large collections and allows them to quickly estimate whether provisional narratives or keywords of interest are within the dataset.

Similarity search requires text translation to vector space. Currently various text encoders are available. One might use BERT sentence transformers (Reimers and Gurevych, 2019) for English-language texts or even a language-agnostic model in the case of a multi-language dataset (Yang and Feng, 2020). The rapid pace of development of NLP models and methods will create increasingly powerful text analysis tools for analysts to use regardless of which languages are involved. Future implementations will incorporate additional functionality such as synthetic (generated) text detection probabilities.

” The leadership in Finland is notably young, female, and left leaning.

## Case Study Background

Since December 2019, Finland has been governed by a centre-left coalition led by Prime Minister Sanna Marin of the Social Democrat Party. Marin's election drew international attention, as she simultaneously became the world's youngest prime minister and the head of a unique coalition in which all five party leaders are women. The four additional coalition party leaders, three of whom are in their 30s, are Li Andersson of the Left Alliance, Maria Ohisalo of the Green League, Annika Saariko of the Centre Party, and Anna-Maja Henriksson of the Swedish People's Party of Finland. Throughout the months that we gathered data, Katri Kulmuni was initially serving as leader of the Centre Party but was replaced by Annika Saariko on 5 September 2020.

The new leadership in Finland is notably young, female, and left-leaning. During our

monitoring period, eleven/twelve of the nineteen Finnish ministers were women. The leaders of the coalition also hold ministerial posts: Andersson serves as Minister of Education, Ohisalo is Minister of the Interior, Kulmuni served as Minister of Finance until her resignation on 5 June 2020, and Henriksson is Minister of Justice. Marin's political agenda prioritises climate change, equality, and social welfare (Special, 2020).

Marin took over the premiership from Prime Minister Antti Rinne, who resigned in December 2019 after the Centre Party expressed that it had lost confidence in him over his controversial handling of a postal workers' strike. Although Rinne tendered his government's resignation, President Sauli Niinisto requested Rinne's cabinet continue on as a caretaker government, allowing the coalition to remain intact (YLE, 2019). The



coalition was originally formed following parliamentary elections in April 2019, when the Social Democrats narrowly defeated the right-wing populist Finns Party with just 17.7% of the vote. The close victory resulted in a centre-left coalition and paved the way for Rinne to become the first leftist Prime Minister in nearly 20 years (Reuters, 2019).

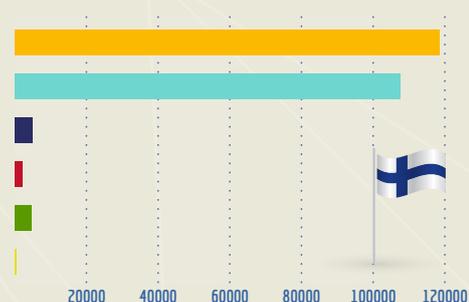
The 2019 elections also highlighted the increasingly fragmented nature of Finnish politics. The right-wing populist Finns party, which campaigned on a eurosceptic and anti-immigration platform, fell a mere 6,800 votes short of winning first place. The party is led by Jussi Halla-aho, who holds controversial political views concerning climate change and immigration policy. Halla-aho's election in 2017 caused tension within the ruling three-party coalition, which

at the time consisted of the Finns Party, the Centre Party, and the National Coalition Party. Finnish state media YLE reported that both Prime Minister Juha Sipilä and head of the National Coalition Party Petteri Orpo announced that their respective parties would no longer cooperate with a Finns Party led by Halla-aho due to his convictions of hate speech for comments made about Islam and Somalis (YLE, 2017). Meanwhile, Finland's traditional political parties struggled, as the Centre Party, conservative National Coalition Party, and left-leaning Social Democrats combined received just 49% of support from voters. This was one of the poorest election outcomes for the Social Democrats, while the Centre Party polled its lowest general election result ever (YLE, 2019). As a result, the current coalition is navigating in a polarised political climate.

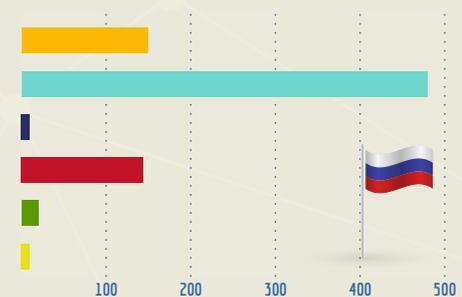
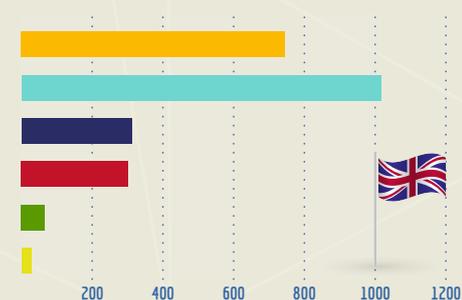


# Descriptive Statistics: Comparison with Robotrolling

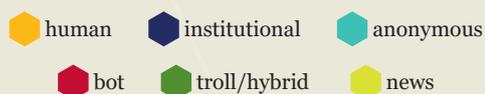
The algorithm we used to assess bot-likelihood for each account in the Finnish dataset was trained on data that tracks English- and Russian-language messaging about the NATO presence in the Baltic countries and Poland. Before we share the results of our analysis, we will provide a comparative overview of these two datasets. During our observation period of the Finnish Twitterspace, the majority of tweets were shared by human accounts (50%) and anonymous accounts (45%), with just 3% of messages originating from bot-like accounts (bot, troll, hybrid). When we apply the abuse filter, the amount of bot messaging remains the same, while anonymous activity jumps to 59% and human activity decreases to 35%. Compared to *Robotrolling* figures for the same time period, we can see that the Finnish information space fosters roughly 1/5 of the bot activity observed in both the Russian- and English-language networks discussing NATO in the Baltics and Poland. Our initial assessment is that the Finnish online space is a 'cleaner' space in which a greater number of legitimate actors are operating online.



## Robotrolling



Number of accounts by type

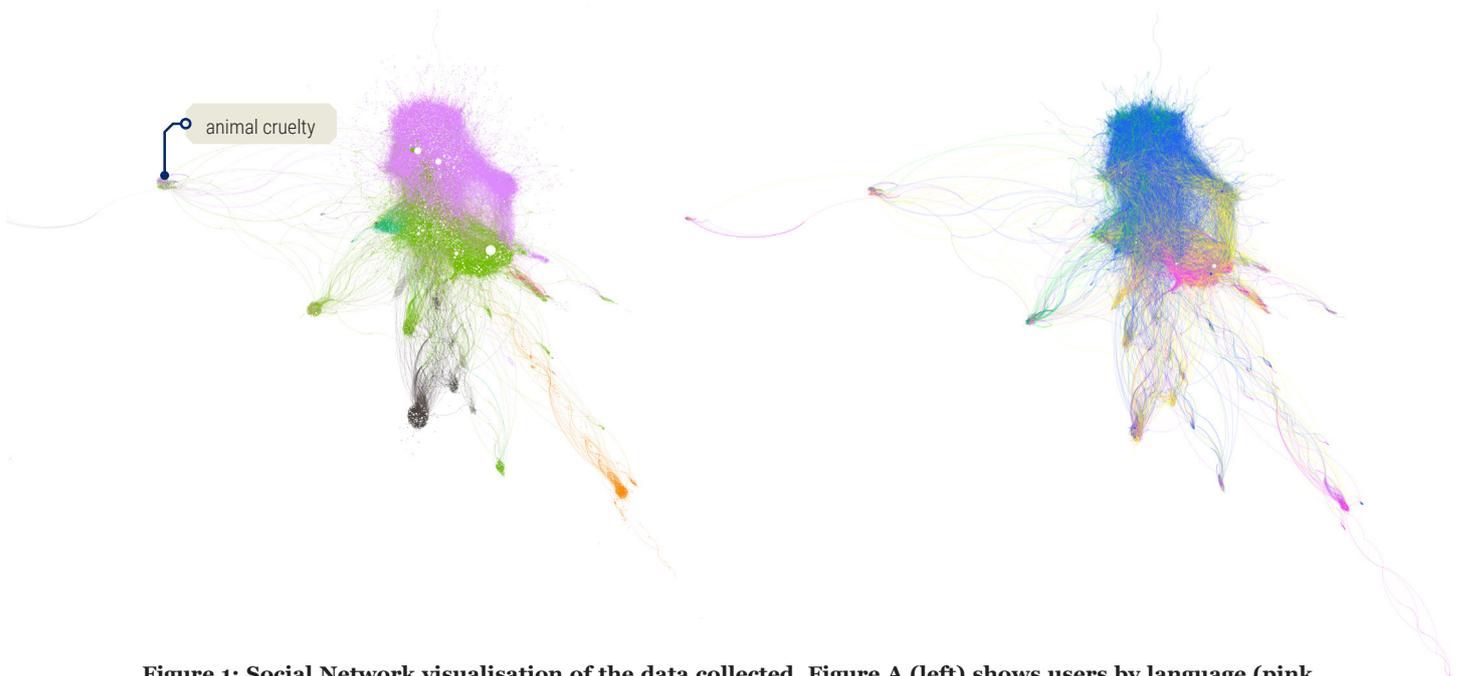


# Social Network Analysis

In order to understand communication patterns on Finnish Twitter, we mapped the connections between users to form a network visualisation. The positioning of each user within the following figures is relative to all the other users in our dataset. As a result, users who mention the same domains, use the same hashtags, and retweet the same users tend to be grouped closer together. Mapping the data in this way creates an approximation of the interests represented within the dataset. For example, users who frequently discuss issues related to climate change or education policy will tend to cluster.

The graphs are coloured to illustrate trends in the data. This allows us to see whether accounts identified as sources of abuse are randomly spread across the network, or whether particular communities are

disproportionately responsible for such messaging. Similarly, we map the calculated likelihoods that each account is automated or anonymous. While there are many legitimate uses for automation on social media, and



**Figure 1: Social Network visualisation of the data collected. Figure A (left) shows users by language (pink is Finnish), whereas figure B (right) shows the users by account type. Anonymous accounts are in yellow; automated accounts in pink. Users are positioned close to accounts that share the same hashtags, retweets, and links.**

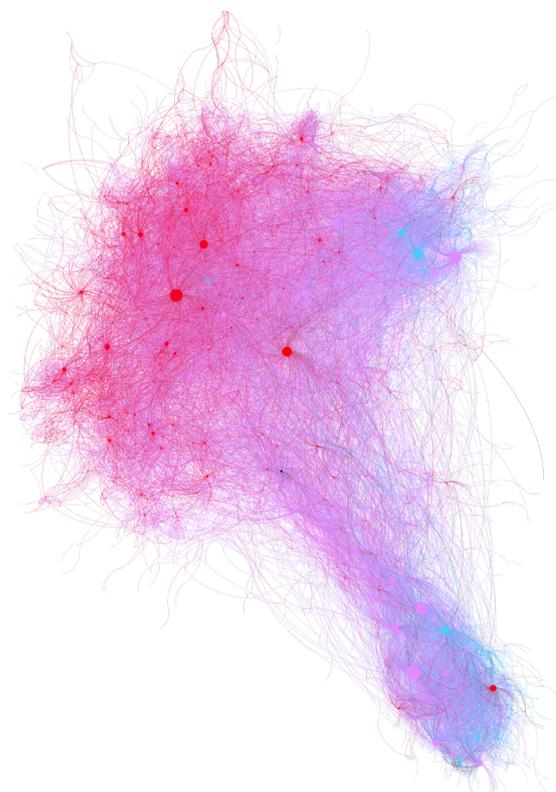


there are reasons why users may prefer to remain anonymous, areas of the graph where such users cluster may indicate coordinated inauthentic activity. We then project other connections onto this configuration, including characteristics such as which users mention each other, whether these mentions are categorised as abusive, and what topics are discussed.

In Figure 1a, users depicted in the constellation to the left have been assigned a colour to differentiate them by language. We note that the majority of users discussing Finnish politicians engage in Finnish (pink, 48%) and English (green, 29%). Figure 1b shows the same network pattern but is coloured according to account type—human (blue), anonymous (yellow), and automated (pink). This reveals that most of the potentially suspicious activity is primarily conducted in languages other than Finnish. Although examining foreign-language bot activity is helpful for comparative purposes, it is of little relevance for this analysis. For the remaining graphs, we are only including Finnish-language users.

## Finnish-language Twitter

Figure 2 depicts the Finnish-language conversation during our monitoring period. The graph is coloured according to account type, with pink representing human users, blue—anonymous users, red—institutional accounts, and black—automated users. We identified two principal communities in this

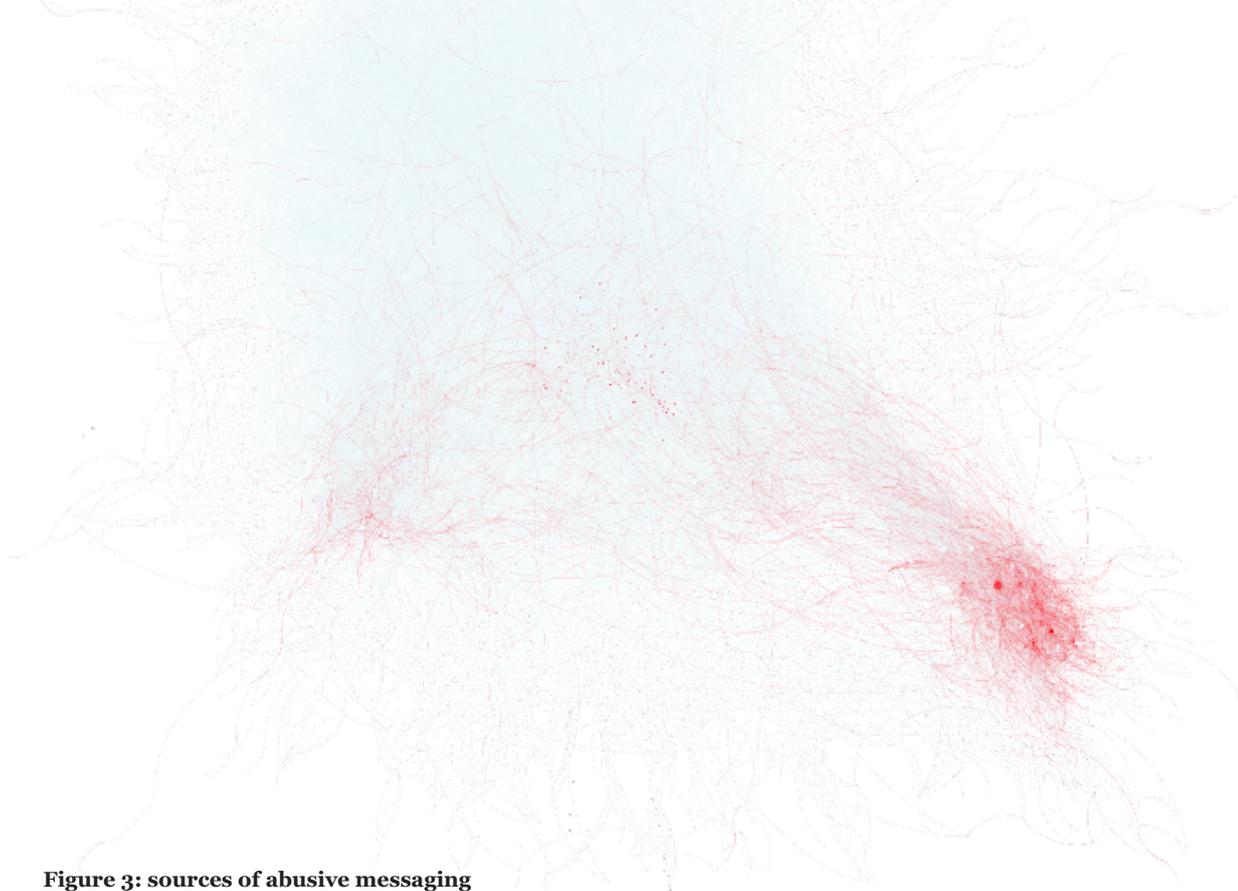


**Figure 2: Finnish language accounts by type**

cluster with a high proportion of anonymous accounts: one on the left and one on the right. Both communities feature patches of blue, reflecting the high proportion of anonymous account activity. The accounts on the right-hand side are centred around Jussi Halla-aho, the leader of the right-wing populist Finns Party. The community situated on the left side of the graph, featuring a smaller blue area, is populated by left-leaning Twitter users. Overall, the data do not suggest widespread use of automated accounts within the Finnish-language conversation.

One minor exception to this assertion is a handful of users from the area on the right whose messaging is regularly retweeted by automated accounts.

To visualise the proportion of abusive messaging, we overlay the calculation of



**Figure 3: sources of abusive messaging**

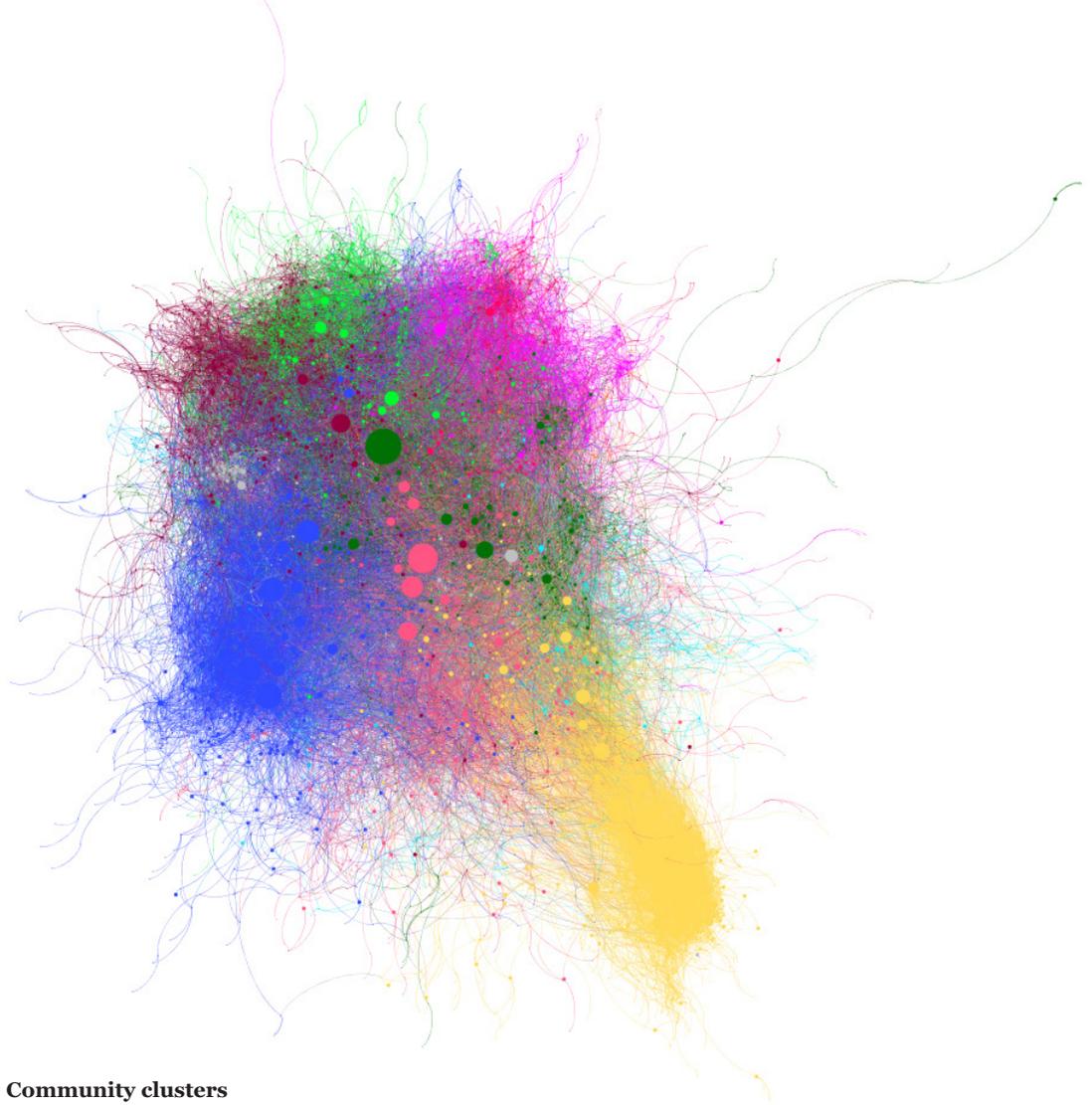
average abuse levels onto this network map. In the result, Figure 3, accounts that tend to engage in systematic abusive messaging are coloured red. As is demonstrated by the cloud of white nodes and edges, the vast majority of users do not use abusive language when engaging with current government officials. The bulk of abusive messaging is clustered in the lower right corner, among the online right-wing community. This finding is relatively unsurprising, as the Finns Party is currently part of the opposition to Marin's centre-left coalition government. A small collection of abusive messaging was produced by the left-wing community, in large part in response to anti-government rhetoric.

To illustrate the interactions between communities, we split the graph into groupings calculated by a modularity algorithm. The number of communities is

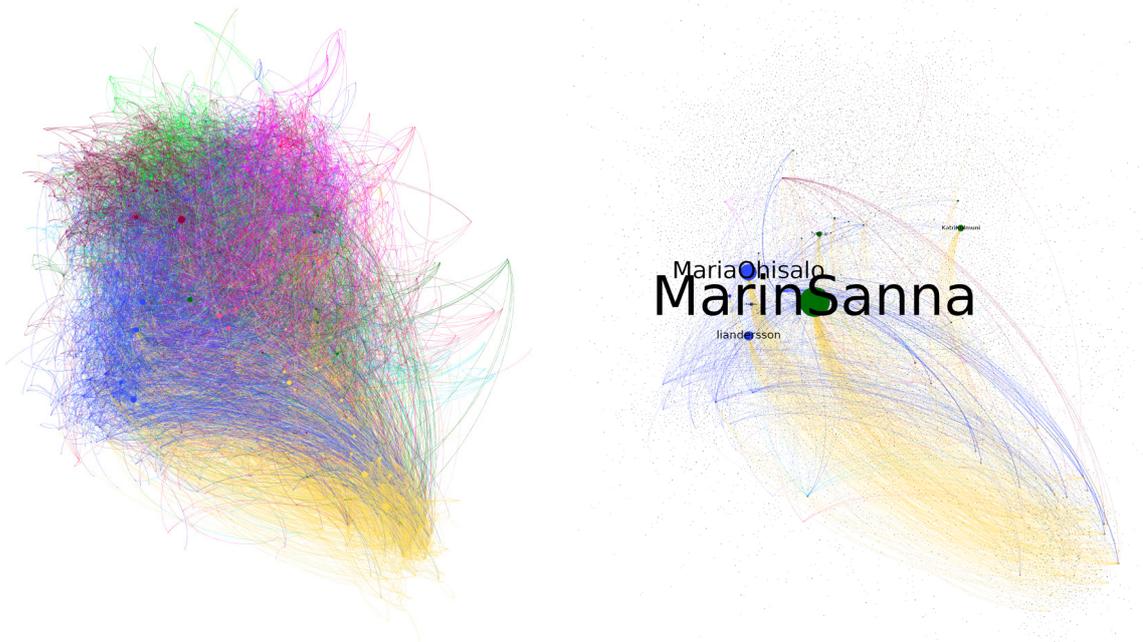
arbitrary and is shown only for illustrative purposes—the borders between the communities is to some degree random. The individual 'bubbles', which feature in several visualisations, vary in size depending on the number of messages sent. In Figure 4, we can see that the blue and yellow communities broadly coincide with the left- and right-wing activists who exhibit disproportionate levels of anonymous and abusive activity.

When we recreate the graph mapping by mentions, opposed to hashtags, domains, and retweets, we observe that the yellow and blue communities—ideologically opposed users—exchange a large volume of messages (Figure 5). Restricting the connections between users to those sharing abuse allows us to see the accounts that receive the highest volume of abusive messaging—the Twitter





**Figure 4: Community clusters**



**Figure 5: a) interactions between clusters, b) abusive messages between clusters**



accounts of Prime Minister Sanna Marin, Minister of the Interior Maria Ohisalo, and Minister of Education Li Andersson.

This reveals an additional dynamic: the targets of yellow community users do not respond directly to the abuse. Instead, it appears that other users positioned close to the targets in the map appear to respond on their behalf, presumably in their defence.

## Topics of abuse

We further analysed the data to understand which themes in Finnish politics attracted the highest levels of abuse from online users. We identified six prominent topic areas that accounts engaged with throughout our observation period. These topics are:

- **Government Corruption and Failure**
- **Sexism and Homophobia**
- **Racism and Islamophobia**
- **Government Handling of COVID-19**
- **Education (in the context of COVID-19)**

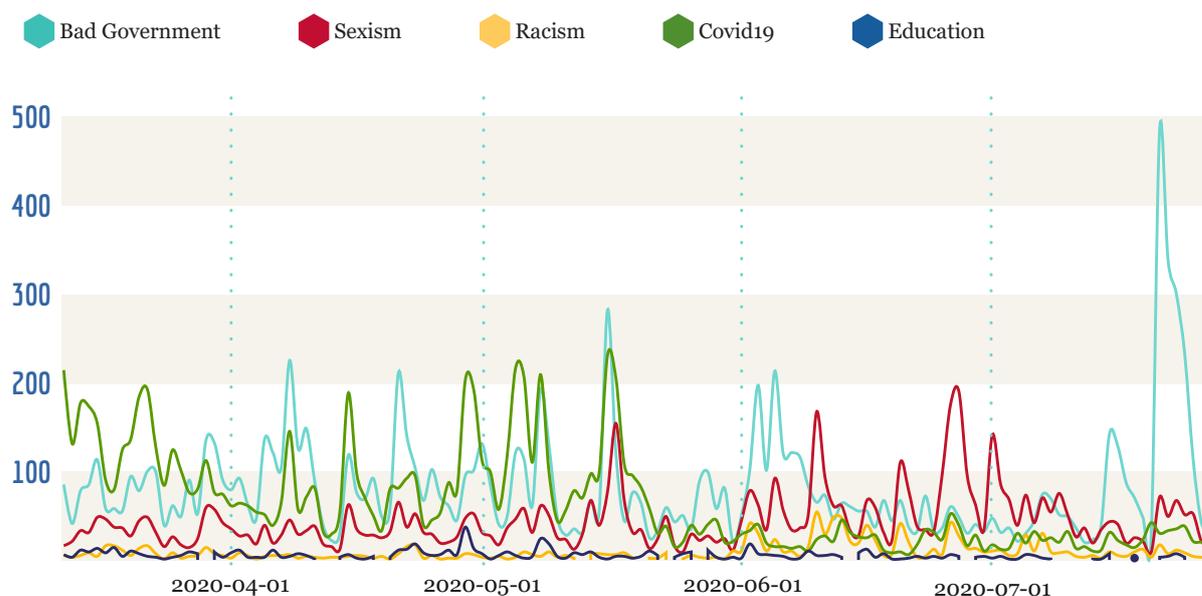
In the graphs on page 30, the same network graph is coloured to show the proportion of abusive messaging about each topic. The darker areas of the graph

show a higher concentration of inter-community engagement as well as a correlation between community interest and messaging about the topics. It is immediately clear that discussions about government incompetence are riddled with abusive language, particularly coming from the right-wing community, shown in yellow, and left-leaning community, shown in blue. Levels of abuse are similarly high among users engaging in sexist and homophobic discourse. Discussions of Topic 3, Racism and Islamophobia, garnered similar levels of activity from the yellow community and reduced activity from the blue community. Abusive messaging about COVID-19, Topic 5, appeared to largely originate from the left-leaning community directed at right-wing users. Topic 5 about education exhibited virtually no abusive messaging

## Was there coordination of inauthentic abusive messaging?

Around 7% of the messages shared on Finnish Twitter during our monitoring period were identified as abusive and over 5,000 users sent at least one abusive message. However, a handful of users shared high volumes of such messages. For instance, during the 138-day observation window, one user sent 520 messages, of which 199 were classified as abusive. Education Minister Li Andersson was the primary subject of the messaging, receiving 87 abusive messages from this one user. Additionally, this user directed 46, 44, 35, and 34 messages at Interior Minister





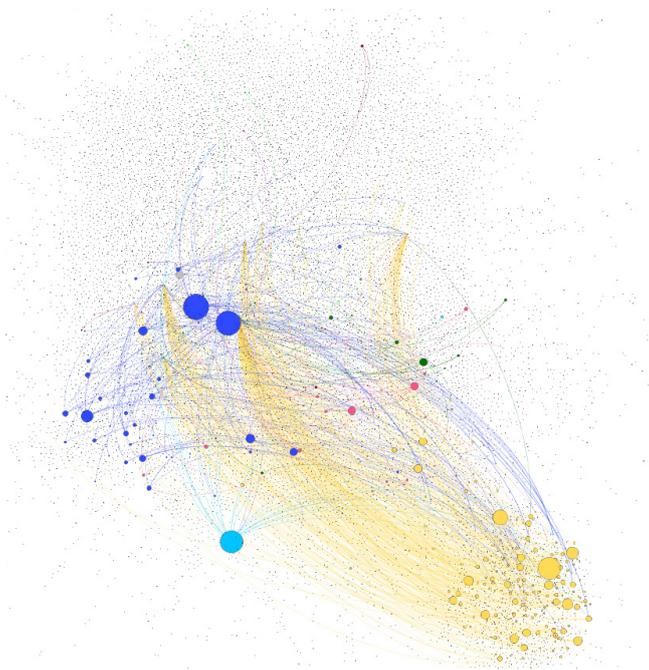
**Figure 6: total number of messages by theme**

Ohisalo, Prime Minister Marin, Minister of Family Affairs and Social Services Kiuru, and former Minister of Finance Kulmini, respectively.

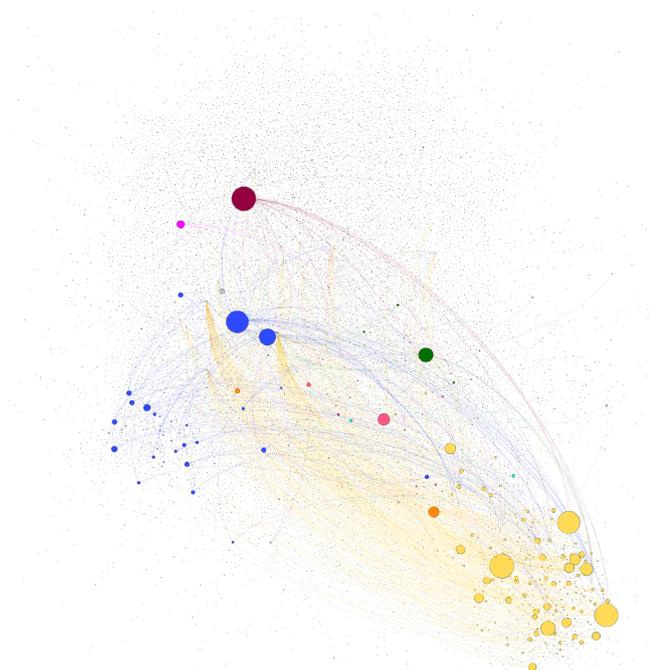
Like this user, the subsequent three most prolific posters of abusive content averaged more than one such message daily. None of these users were identified as part of a community in the social network analysis, because they never retweeted other users, shared links to news stories, or even commented on specific hashtags. Instead, more than 93% of their posts were directed at (@) other twitter users, in large part government ministers. That said, there is nothing about the posting patterns of these four hyperactive users that indicate they

are operated automatically, by the same person, or even in coordination with one another.

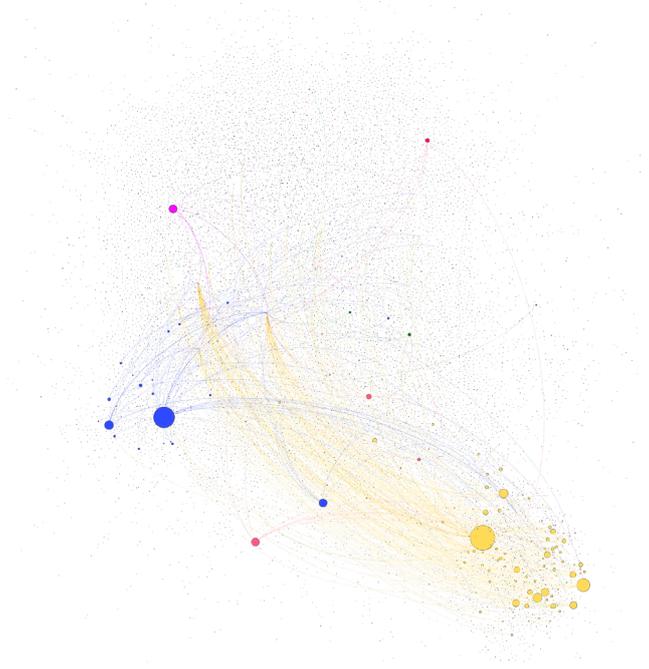
Beyond the example of these four users, we observed a tendency where the most abusive messages come from users who in their twitter activity are singularly focused on harassing the government. These accounts may be fake—certainly the owners of the accounts are not generally easily identified—but they do not appear to be tightly, if at all, coordinated. Thus, when it comes to abusive messaging of this kind, the story told in the data is less about messages being posted from coordinated accounts, but rather a stream of abusive messages coming from a few accounts.



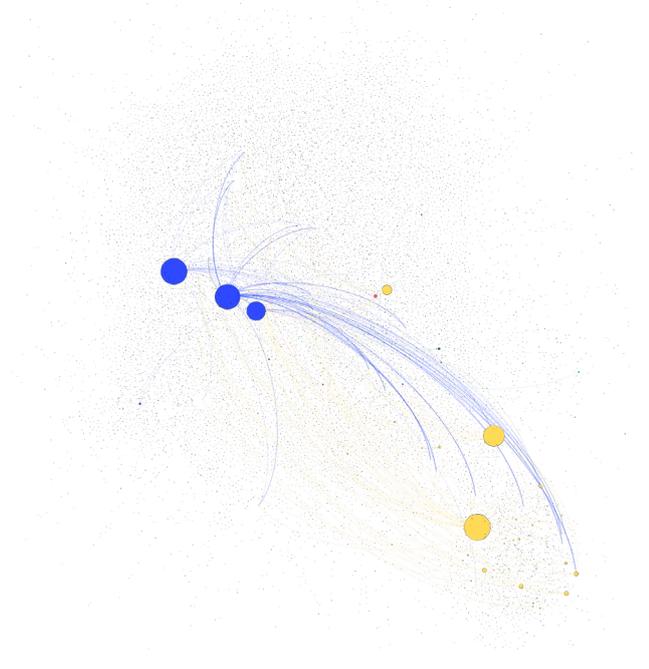
**Topic 1: Government Corruption and Failure**



**Topic 2: Sexism and Homophobia**



**Topic 3: Racism and Islamophobia**



**Topic 4: COVID-19**



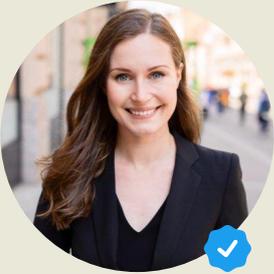
Users who mainly message directly at other users (@mention percentage of 95–100%) exhibit an abusive message rate of 14%. Two-thirds of these users are classified as anonymous. We observed more than 500 accounts sending abusive messages at politicians who do little else than engage in conversation with other Twitter users through @mentions. These users, though highly active, are disconnected from the network modelled in the Social Network Analysis section because they do not retweet other users, share articles, or use hashtags. Of these users, our algorithm identified 70% as anonymous, compared to 27% for the dataset as a whole, and 14% for users within the connected graph modelled previously. Many of these accounts, roughly 30%, were created in 2020. Our methods are not suited to detecting who is operating individual accounts, but together they do appear to be a category with a disproportionate number of suspicious users.

Overall, we observed very low levels of both bot and coordinated activity. The majority of bots we identified were operating in foreign languages and either not generally focused on Finland or used to push certain causes in multiple languages. We repeatedly came across a cluster of accounts throughout our monitoring period that posted the same messages about animal cruelty and climate change. These accounts predominantly message in English and appear in some cases to be automated or semi-automated. However, they represent a very small part of the conversation. Likewise, a small cluster of automated accounts amplified messaging by a number of right-wing voices. Again, there was coordination here, but this looks more like self-promotion than an attempt to systematically manipulate the information space. If large-scale inauthentic coordination exists in the Finnish information environment, we are either looking in the wrong place, or it is so sophisticated or so small in scale that it evades our detection methods.



# A Closer Look: Online Abuse of Finnish Politicians

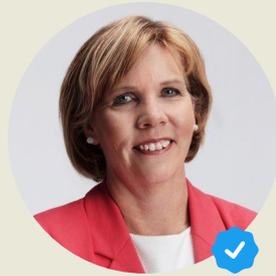
Informed by the insights from our social network analysis, we will now turn to our qualitative discussion of abuse aimed at Finnish politicians between 12 March and 27 July 2020. To this end, we identified the days, and clusters of days, during which levels of abusive language were significantly high. We investigated these individual spikes in abusive activity to determine what triggered intensified abuse, which cabinet members were targeted with abuse, and whether the abuse was driven by inauthentic accounts.

 <b>Sanna Marin</b> Prime Minister @Social Democrats	 <b>Pekka Haavisto</b> Minister for Foreign Affairs @Green League	 <b>Maria Ohisalo</b> Minister of the Interior @Green League
 <b>Katri Kulmuni</b> Minister of Finance (Dec 2019 - June 2020) @Centre	 <b>Matti Vanhanen</b> Minister of Finance @Centre	 <b>Li Andersson</b> Minister of Education @Left Alliance





**Ville Skinnari**  
Minister of Development  
and Trade  
[@Social Democrats](#)



**Anna-Maja  
Henriksson**  
Minister of Justice  
[@Swedish People's Party](#)



**Tuula  
Haatainen**  
Minister of Employment  
[@Social Democrats](#)



**Antti Kaikkonen**  
Minister of Defence  
[@Centre](#)



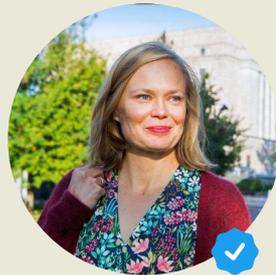
**Tytti  
Tuppurainen**  
Minister of European  
Affairs and Ownership  
Steering  
[@Social Democrats](#)



**Sirpa Paatero**  
Minister of Local  
Government  
[@Social Democrats](#)



**Timo Harakka**  
Minister of Transport and  
Communications  
[@Social Democrats](#)



**Hanna Kosonen**  
Minister of Science and  
Culture  
(Aug 2019 - Aug 2020)  
[@Centre](#)



**Annika Saarikko**  
Minister of Science and  
Culture  
[@Centre](#)





**Krista Mikkonen**

Minister of the Environment and Climate Change  
@Green League



**Jari Leppä**

Minister of Agriculture and Forestry  
@Centre



**Mika Lintilä**

Minister of Economic Affairs  
@Centre



**Aino-Kaisa Pekonen**

Minister of Social Affairs and Health  
@Left Alliance



**Krista Kiuru**

Minister of Family Affairs and Social Services  
@Social Democrats



**Thomas Blomqvist**

Minister of Nordic Cooperation and Equality  
@Swedish People's Party

**Abuse in figures**

Over the course of our four-month monitoring period, we collected over 350,000 tweets directed at or discussing the nineteen Finnish cabinet members. Our algorithm detected 24,885 tweets with an abuse probability above 0.6. Of these messages, only 2.5% were attributed to

automated users. Human user activity contributed 35% of abusive content. Anonymous users, on the other hand, were the most prolific abusive tweeters and responsible for 59% of abusive messages.

Approximately 5,426 unique users tweeted abusive messages at politicians. 170 of them, less than 3%, were classified as



bots by our algorithm. Compared to other datasets we have studied this is a low percentage. The counts of human and anonymous users were almost equal, suggesting that anonymous users produced, on average, more abusive tweets than users identified as human. This is hardly surprising, as previous studies have highlighted a correlation between online anonymity and increased trolling behaviour.

The five ministers targeted with the most abusive messaging during our monitoring period were Prime Minister Sanna Marin, Minister of the Interior Maria Ohisalo,

Minister of Education Li Andersson, now former Minister of Finance Katri Kulmuni, and Minister of European Affairs Tytti Tuppurainen. Notably, all five ministers are female and, at the time, four of them served as the leaders of their respective parties. Prime Minister Marin was targeted with the most abuse, as she received more than one third of abusive tweets (34%). There is a significant gap between the first- and second-most tweeted at ministers; Marin was trailed by Ohisalo, who was tagged in 18% of abusive messages. Following Ohisalo were Andersson (12%), Kulmuni (6%), and Tuppurainen (4%).

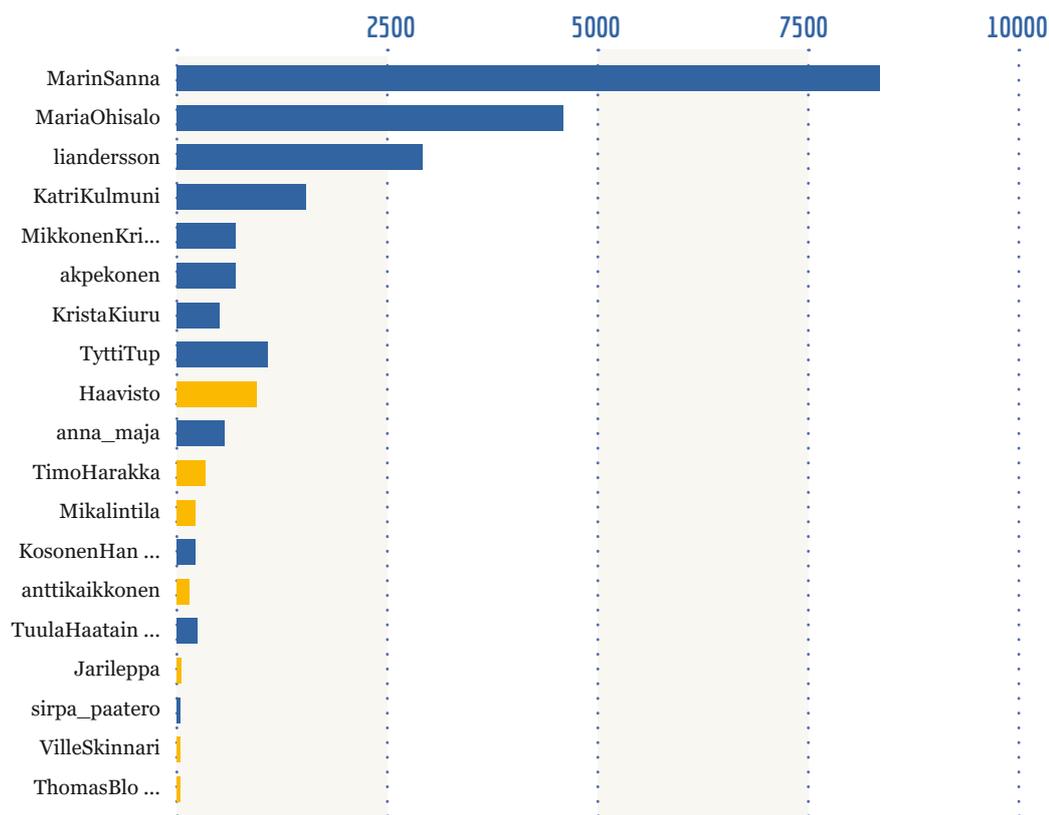
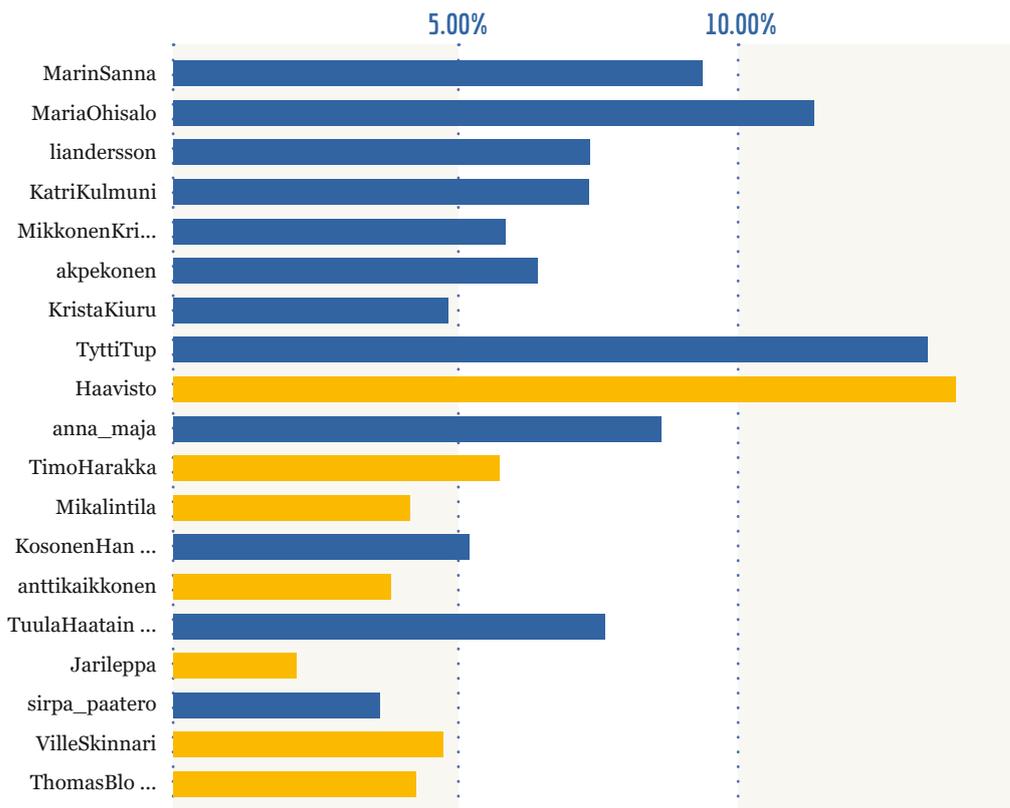


Figure 7: Number of abusive messages directed at ministers—women in blue, men in yellow.





**Figure 8: Percentage of messages directed at ministers which are abusive—women in blue, men in yellow.**

By contrasting the absolute number of abusive tweets with the total volume of tweets targeting each individual minister, we can compare the relative number of abusive tweets each minister received. Although Marin received the highest absolute number of abusive messages, they made up 9% of all tweets sent to her. Tuppurainen, who was sent the lowest count of abusive tweets, received the greatest share of abusive tweets (13.5%). 11% of the tweets sent to Ohisalo were abusive, while 7% of all tweets sent to Andersson and Kulmuni were identified as abusive.

### Spikes in abuse: when, who, and why

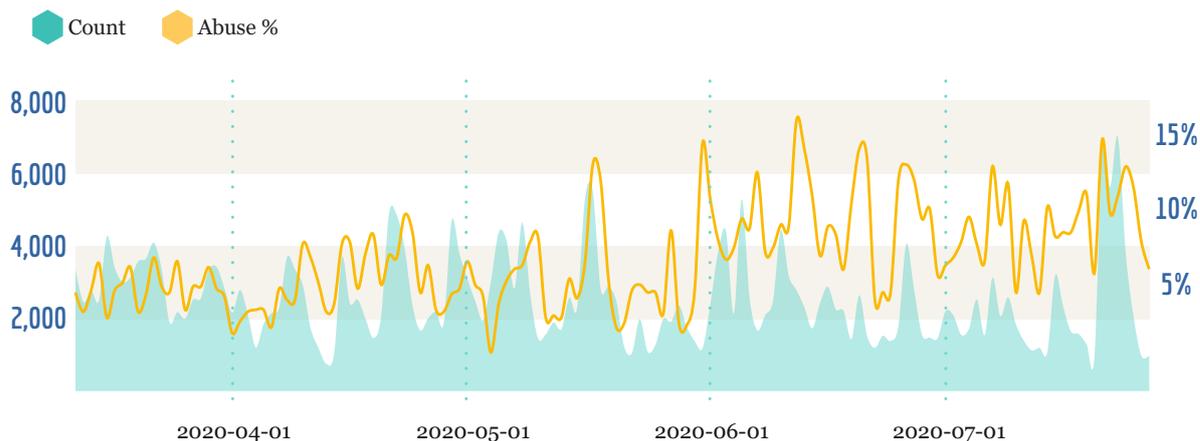
The timeline in Figure 9 shows abusive activity on Twitter during our monitoring period, 12 March – 27 July 2020. The timeline features various peaks and troughs, which represent the changing volume of abusive messaging over time. From this visualisation, we can infer that levels of abusive activity were relatively stable throughout March and April, picking up in May and intensifying throughout June and July. In this section, we will take a deeper look at this timeline



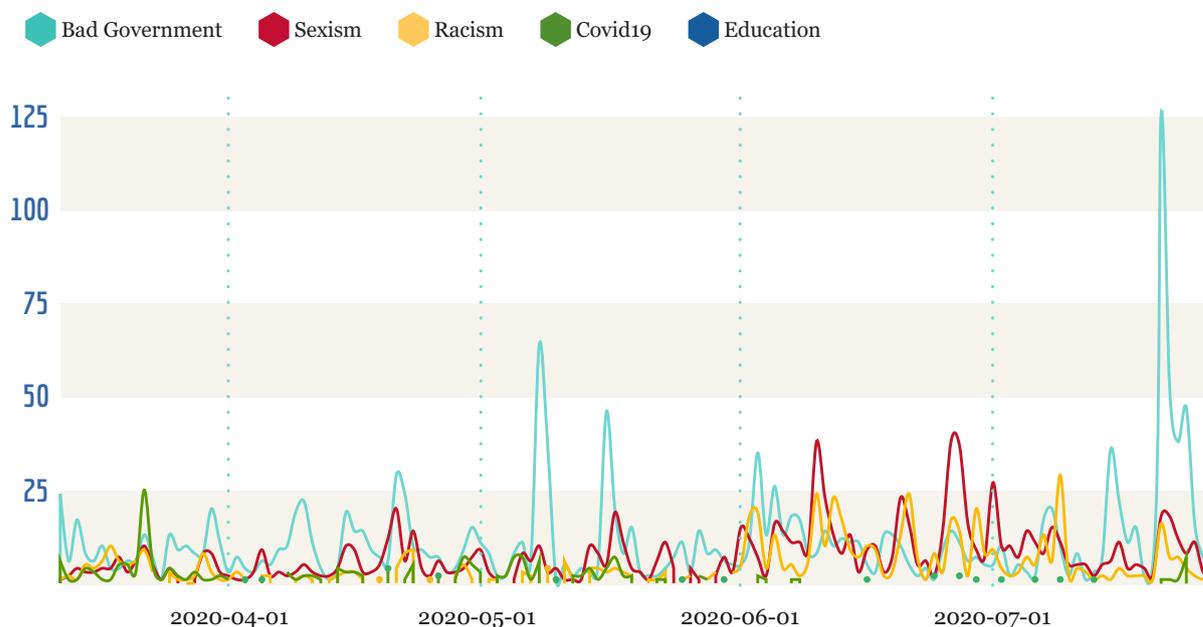
by investigating particularly significant spikes in abusive activity. Our aim is to develop a greater understanding of the nature, method, and triggers of politically charged abusive activity targeting Finnish politicians online.

## March

Because we began tracking abusive language in the Finnish information space in March, we were able to capture individuals' reactions to the initial



**Figure 9: Message count and percentage of abusive messages over time**



**Figure 10: Count of abusive messages by theme and over time**



escalation of the COVID-19 pandemic in Europe. The latter half of March featured a cluster of modest increases in abusive language. Over half of these messages were shared by anonymous accounts, while less than 1% were attributed to bots. The volume of abusive messages peaked twice, first on 22 March and again on 29 March. On 22 March, abusive messaging primarily targeted Prime Minister Marin with criticism of her government's handling of COVID-19. Education Minister Li Andersson received significant attention on 29 March for an opinion piece she published in which she analysed how the pandemic has highlighted class divisions in Finland (Andersson, 2020). Among the 30 most unusual hashtags detected in this abusive cluster were *vihreät* (green), *koronavirusfi* (coronavirus Finland), *vasemmistoliitto* (Left Alliance), *kakarahallitus* (cockatoo government), and *papukaijahallitus* (parrot government).

The COVID-19 pandemic, and government policies intended to curb its spread, were the central focus of users sending abusive messages throughout March. On 16 March, the Finnish government declared a national state of emergency due to the acceleration of COVID-19 cases throughout Europe. Shortly afterwards, the Ministry of Foreign Affairs announced that Finland would close its borders to non-essential travel on 19 March (Helsinki Times, 2020).

In a series of tweets, Interior Minister Maria Ohisalo underscored that the travel restrictions would not impede filings for asylum in Finland.<sup>2</sup> This assertion was met with fierce anti-immigrant and anti-Muslim rhetoric, with many users accusing Ohisalo of prioritising the safety of refugees over that of native Finns.

As the COVID-19 pandemic intensified in March, Prime Minister Marin was criticised on Twitter for her government's early response to the public health crisis. Abusive messages targeting Marin and her fellow young female ministers questioned their leadership abilities and decision-making skills due to their age and gender. Many abusive messages referred to the current Finnish administration with derogatory phrases intended to undermine their authority on the basis of gender, such as: "lipstick government", "lipstick girls", "feminist quintet", and "tampax team". These phrases were often paired with "red-green", "communist", and "left-wing", creating an association between presumed feminine incompetence and the parties in the current coalition. Both Marin and Andersson were accused of personally being responsible for the deaths of Finns due to COVID-19 and criticised for not adopting restrictions, particularly school closures, earlier.

---

<sup>2</sup> <https://twitter.com/untamokuikka/status/1240367910910070786>



### Examples: trolling Maria Ohisalo

13 March:

luuletko että korona ja karanteenit lopettaa hutisalon kiiman rahdata tänne miljoona tummaa miestä omien 2,6 miljoonan työttömän jatkoksi?

Translation: *Do you really think the coronavirus and quarantine will stop Ohisalo, who's just ravenous to bring a million black men here, in addition to our own 2.6 million unemployed??*

22 March: @MariaOhisalo:

on Suomen julmin tunteettomin fanaatikko joka on saanut ministerisalkun. Eroa, et ole suomalaisten puolella pätkääkään. Unelmasi on tehdä Suomesta muslimimaa.

Translation: *@MariaOhisalo is the cruellest and coldest fanatic in Finland ever to be appointed a minister. Resign: not one bit of you is on Finland's side. Your dream is to*

### Examples: sexist attacks of Marin and government

15 March:

@MarinSanna Ei ole mitään Suomelle vaarallisempaa kuin Suomea kohtaava turvallisuusuhka ja että samaan aikaan Suomen hallituksessa istuu punavihreitä feministityttöjä, joilla ei ole elämäkokemusta, ei selkärankaa eikä ymmärrystä turvallisuudesta.  
@valtioneuvosto @STM\_Uutiset @THLorg

Translation: *@MarinSanna There's nothing more dangerous for the security threat facing Finland than the members of the government being socialist-green feminist girls with no life experience, no backbone and no understanding of security. @valtioneuvosto @STM\_Uutiset @THLorg*

29 March:

painu helvettiin demari pasaka, sulla ei ole pätevyyttä hoitaa maan asioita ja vie mukana muutkin huulipuna ämmät

Translation: *F\*ck off to hell you Social Democrat piece of sh\*t, you're not qualified to run the country, and take the rest of the lipstick b\*tches with you too.*

29 March:

Olette koko huulipuna-hallitus mokanneet perusteellisesti alusta alkaen.

Lentoasema sai rauhassa toimia Suomen pääviruslinkona 2-3 viikkoa ennenkuin edes siihen heräsitte. Olette kokoajan 2-3 viikkoa jäljessä ja jälki on sen mukaista. Antakaa kokeneemmille tilaa, kiitos

Translation: *You lipstick government have messed everything up royally right from the start. The airport was allowed to be Finland's biggest virus spreader for 2-3 weeks before you even faced the fact. You're always 2-3 weeks behind the facts and it shows. Move over and let the more experienced take over, please*



## April

Throughout April, the COVID-19 pandemic and the Finnish government's response to the crisis continued to attract abusive language. Marin and female members of her cabinet continued to receive gendered abuse, particularly from anonymous accounts. Bot activity increased slightly, as our algorithm found 2% of abusive messages were sent by automated accounts. Anonymous activity remained high, with 60% of messages attributed to anonymous accounts. The first major spike this month occurred on 10 April, coinciding with a face mask procurement scandal. The government of Finland announced the resignation of Tomi Lounema, the managing

director of Finland's National Emergency Supply Agency, after he admitted to a multi-million euro purchase of face masks from China that were either unsuitable for use or not delivered (YLE, 2020). As this agency is overseen by the Ministry of Economic Affairs and Employment, several users blamed Minister Tuula Haatainen for the scandal and called for her resignation, labelling her inexperienced and incompetent.

Abusive activity rose again on 15 April in response to Foreign Minister Pekka Haavisto's announcement that Finland would be granting additional funding to the WHO, pledging that the Finnish government would commit the same funding to the organisation that was previously provided in 2015, around

### Examples: more gendered abuse

*8 April:*

Syy tähän hallituksen mielisairauteen on aate "punavihreä feminismi".

Translation: *The reason for this government's mental illness is the ideology of "red-green feminism".*

*21 April:*

Sanna voi työntää ton rokotteen limaiseen pilluunsa! Meille ei noita tule.

Translation: *Sanna, you can shove that vaccine up your sticky c\*nt! We're not taking it.*

*21 April:*

@MarinSanna Miten sinä vasemmistolaisläähkö jaksat suoltaa tuota joutavaa paskaa. Mene leipomaan tai täyttämään pyykkikone. Tyhjöpäinen vasemmistolainen. Sinulla ei selvästikään ole hajuakaan miten tästä selvittää.

Translation: @MarinSanna How do you manage to spew out such dreadful sh\*t all the time, you left-wing hag? Go and bake something or load the washing machine. Empty-headed left-winger. You clearly have no clue how to manage this situation.



## Example: WHO reactions

15 April:

@MariaOhisalo Tällä ilmoituksella hyväksytte Kiinan ja WHO:n välisen korruption jonka seurauksena miljoonat kuolevat ja maailmantalous ajautuu pahimpaan kriisiin ikinä. Onko sinun ja puolueesi typeryydellä mitään limittiä? Eroa.

Translation: @MariaOhisalo With this announcement, you accept the corruption between China and the WHO, as a result of which millions will die and the world economy will plunge into the worst crisis ever.

5.5 million euros (YLE, 2020). Prime Minister Marin released a statement of support for the WHO in the wake of US president Donald Trump's order to suspend funding to the health organisation. Marin received abuse from several users engaging with conspiracy theories, labelling the WHO as a criminal organisation controlled by China. These messages accused Marin of being a "Chinese puppet", supporting a Marxist-terrorist, and driving Finland into bankruptcy by sending funding abroad.

Mika Lintilä, the Minister for Economic Affairs, was significantly targeted with abuse on 21 April amid allegations of corruption in the distribution of COVID-19 relief funding from the government (YLE, 2020). The government announced an audit into a 200-million-euro programme that granted funding to software companies, consultancies, and television production companies, among others, as a facet of COVID-19 related emergency funding. This

drew considerable criticism from users, who condemned government spending and labelled the initiative as communist.

On the following day, 22 April, the volume of abusive tweets remained high as the Finnish criminal police announced that they would launch an investigation into suspected fraud and forgery in connection with a 2019 European Court of Human Rights (ECHR) ruling on Finland (Mäki, 2020). In November 2019, the ECHR ruled that Finland violated sections of the European Convention on Human Rights by denying asylum to a middle-aged Iraqi man who was reportedly killed shortly after returning to Iraq (Valtioneuvosto Statsrådet, 2019). In their announcement, Finnish authorities claim to have had reason to suspect that the documents that influenced the ruling were forged, and that the man is, in fact, still alive. Maria Ohisalo, who had condemned Finland's decision to deport the asylum seeker, was targeted with abuse for her



### Example: Ohisalo and terrorism

28 April:

Meidän oma mielenvikainen jihadististi

@MariaOhisalo

on täysi ääliö. Jos kiima ääriliikkeisiin ja niiden edustajiin on näin kova...

Translation: *Our own mentally retarded jihadist*

@MariaOhisalo

*is a total idiot. If she's got such hots for extreme movements and their representatives...*

previous statements and support for resettling refugees in Finland.

Levels of abusive language rose again towards the end of the month, on 29 and 30 April. The primary targets of abuse were Minister of the Department of Transportation and Communication Timo Harakka and Minister of Education Li Andersson. Twitter users attacked Harakka for his support for and investigation into urgent COVID-19 relief for journalism to support news agencies and secure five-day newspaper distribution across the country. Hostile users accused Harakka of funding state propaganda, sending taxpayer funds to his 'left-wing friends' in the media, and enabling the continuation of a "corrupt", "communist" government. Simultaneously, Andersson was criticised for the Education Ministry's decision to resume in-person teaching on 14 May as health authorities determined opening schools under controlled conditions would be safe for both students and staff.

Andersson was accused of endangering children's lives, not caring about the welfare of Finnish children, and intentionally spreading the virus to achieve herd immunity.

### May

Compared to April, we observed a slight decrease in overall abusive activity conducted by anonymous accounts throughout May. This was complemented by a minimal increase in human-driven activity, while the level of bot engagement remained negligible. Prime Minister Marin, Interior Minister Ohisalo, and Education Minister Andersson received the most significant portions of online vitriol. During her May Day speech on 1 May, Andersson introduced a COVID-19 economic relief initiative, suggesting a €100 "revitalisation voucher" be distributed to all Finns in order to support the service sector (YLE, 2020). This was met with accusations that



### Example: COVID-19 conspiracy

9 May:

feikkipandemia toteutettiin juuri sen takia, että loiset pääsette rakentamaan teidän ihanan vihreän valvontayhteiskunnan joka toisin sanoen on valvontakapitalismia. Minä en ainakaan teidän projektiinne osallistu.

Translation: *The COVID-19 fake pandemic was invented precisely so you parasites could build your lovely green surveillance society, in other words, surveillance capitalism. I for one am not going to take part in your project.*

Andersson is attempting to buy the approval of constituents.

Levels of abuse on Twitter rose again on 8 May, coinciding with a statement by Pasi Pohjola, the Director of COVID-19 Operations at the Ministry of Social Affairs and Health, addressing the government's public health strategy. Pohjola explained that the COVID-19 virus must spread throughout the Finnish population in order to maintain the legal justification for continuing emergency measures (Helsingin Sanomat, 2020). Prime Minister Marin responded to this statement on Twitter, asserting that Pohjola's statement does not represent the administration and that, "the aim of the government is to prevent the spread of the virus in society and therefore restrictive measures have been taken and are still in force".<sup>3</sup> Marin was subsequently targeted with anger and accusations of

using the Emergency Preparedness Act to both expand the power of the central government and to safeguard her position as its leader. Some abusive messages fell into the category of conspiracy, with claims ranging from COVID-19 being a hoax to Marin personally overseeing "Finnish genocide".

The volume of politically motivated abusive messaging skyrocketed on 17 May, resulting in the third-highest spike in abuse we observed during our monitoring period. We observed that half of the abusive messaging on this day was conducted by anonymous accounts and 2.6% of messages were circulated by bots. This slew of abusive messaging was triggered by ministers tweeting to observe the UN-recognised International Day against Homophobia, Transphobia, and Biphobia (IDHTB), which happened to coincide with

<sup>3</sup> <https://twitter.com/MarinSanna/status/1258708466228760578>



## Examples: homophobic abuse

17 May:

@vihreat and @MariaOhisalo

Vihreät muuttuu päivä päivältä kuvottavampaan suuntaan. Suomalaisuuden päivänä hehkutitti islamia ja Kaatuneiden muistopäivänä homoja sekä transuja. Aivan kertakaikkisen vastenmielinen puolue

Translation: *The Greens are moving in an increasingly sickening direction every day. On the Day of Finnishness [J.V. Snellman's birthday, 12 May – IME] they sang the praises of Islam and on Commemoration Day of Fallen Soldiers they praised gays and trans people. All round, a simply disgusting party.*

17 May:

@MikkonenKrista Homot sitä lesbot tätä transut tota. Onko mitään yhtä polttavaa puheenaihetta ja kriittisemmin huomiota kaipaavaa asiaa enää kuin joku vitun sateenkaarilippupropaganda? Alkaa tulla jo korvista ulos tämä hevonpaska.

Translation: *Gays this, lesbians that, trans people the other. Is there no other burning discussion topic or issue that needs our more urgent attention than some f\*\*king rainbow flag propaganda? This horses\*\*t is starting to come out of my ears already.*

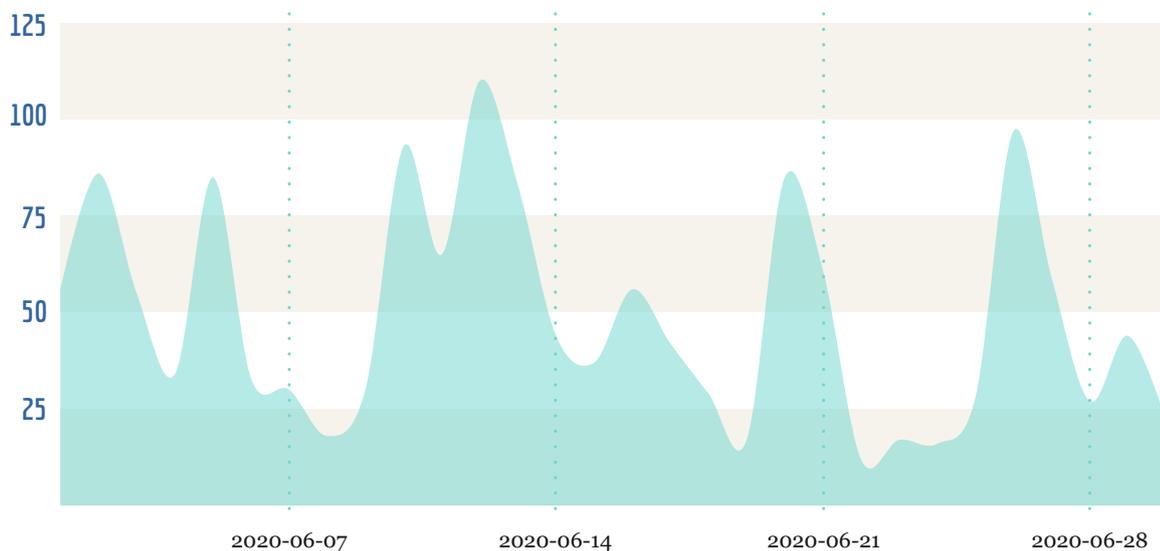


Figure 11: Volume of racism-themed abusive messages in June



Finland's Memorial Day for the War Dead, a day of remembrance commemorated on the third Sunday in May. Maria Ohisalo was the recipient of multiple homophobic tweets after the official Green party Twitter account shared a blog post she authored entitled, "Let's build a Finland where no one is afraid to be themselves".<sup>4</sup> Minister of the Environment and Climate Change Krista Mikkonen, who otherwise does not feature prominently in our data as a target of abuse, received 31 abusive messages on 17 May for tweeting in support of IDHTB. Several messages drew a connection between supporting the LGBTQ+ community and Finnish immigration policy, referring to both as contributing to the current coalition's goal to erode traditional Finnish culture.

## June

We observed a notable increase in the volume of abuse directed at Finnish ministers throughout June; accounts published nearly 2,000 more abusive tweets in June than in May. Abusive language that month was particularly racially charged, reflecting global affairs at the time. In early June, several Finnish ministers voiced their support for the demonstrations in the US over the killing of George Floyd by police and the global Black Lives Matter (BLM) movement (Shubber, 2020). On 2 June, Interior Minister

Ohisalo and Education Minister Andersson shared tweets about the BLM movement and expressed solidarity with the protesters.<sup>5</sup> In response, both ministers received a flurry of abusive and racist tweets condemning their statements. Ohisalo was accused of hypocrisy for welcoming potentially dangerous refugees into Finland, being racist against Finns, and supporting domestic terrorism. Ohisalo received 119 abusive tweets, some of which included the hashtags #AllLivesMatter, #AntifaTerrorists, and #QueenofSpades. Our algorithm found that less than 1% of these messages were sent by bots, indicating that human and anonymous users were responsible for this trolling.

On 5 June, Finance Minister Katri Kulmuni announced her resignation following revelations that she spent over 50,000 euros of public funds on media training (Bateman, 2020). Consequently, we observed a spike in abusive messaging. Kulmuni, who also served as leader of the Centre Party until 5 September, was the subject of online debate over whether her actions warrant charges of corruption. On the same day, Anna-Maja Henriksson, leader of the Swedish People's Party of Finland and Minister of Finance, shared her support for the BLM movement on Twitter. Abusive messages accused her of misusing her ministerial post to promote "looting and violence" and a left-wing conspiracy to overthrow US president Donald Trump.

---

4 <https://twitter.com/vihreat/status/1261908248866762752>

5 <https://twitter.com/MariaOhisalo/status/1267759582438359042>



## Examples: anti-BLM

*2 June:*

Mene sinä terroristi helvettiin siitä. Haluaisit varmaan samat saatanan negridi-laumat riehumaan suomen kaduille.

*Translation: Go to hell, you terrorist. You'd probably want the same f\*cking Negroid herds running riot on the streets of Finland.*

*2 June:*

Etkö vois vaan aloittaa valkosen rodun tuhoamista itestäs jos niin tuo ihonväri hävettää. Älä saarnaa siitä muille. Tuo on vitun sairasta ja sääliittävä.

*Translation: Couldn't you start destroying the white race with yourself, if you're so ashamed of that skin colour? Don't preach to me about it. It's f\*cking sick and pathetic.*

## Examples: responses to ministers speaking out against racism

*11 June:*

Tyypillistä kaltaisellesi joka vinkuu mustan miehen perään. Suomalaiset miehet ulos kotimaastaan jos teidän suvakkaatit ja meetoot ei miellytä ja nuorta ja tummaa miestä tilalle vanhoja akkoja ja läskisuvakki-feministejä tydyttämään. Itku teille tästä vielä seuraa.

*Translation: Typical for someone like you who loves black men. Finnish men should leave their homeland if they don't like their tolerant ideologies and #MeToo movements, and they should be replaced by a young black man to satisfy the bitches and fat tolerant feminists. It will all end in tears for you.*

*12 June:*

Kun olet aivopesty eliitin ja juutalaisten lobbaamaan Euroopan afrikkalaistamiseen ja pyrkyrytesi ylittää järjen äänen ja faktat, sinusta tulee uhka omalle kansallesi. Nämä aivopestyt viherkommunistit eivät kykene enään kuulemaan lastemme ääntä. Tämä on sairaiden ihmisten unelma

*Translation: When you're brainwashed into the Africanization of Europe lobbied by the elite and Jews, and your careerism outweighs all sense and facts, you have become a threat to your own people. These brainwashed green Communists are not able to hear our children's voices anymore. This is a dream of sick people.*



We observed a collection of spikes in abusive activity between 10 and 12 June. On 10 June, Hanna Kosonen, Minister of Science and Culture, was targeted with abusive tweets for denouncing a book published by Suomen Perusta, the Finns Party think tank, describing the book as having “cruel and disturbing” views on women. Suomen Perusta subsequently pulled the book *Totuus kiihottaa* [*The Truth Provokes*], described as a “philosophical study of the information and truth crisis of the left-wing populist mainstream media”, from circulation. The book discussed topics such as women’s sexual independence, the role of men in society, equality, and immigration (YLE, 2020). Prime Minister Marin addressed the controversy, tweeting, “racism, hatred and discrimination are not part of a civilized society”.<sup>6</sup> Criticism of the book by high-ranking officials drew allegations of corruption and censorship, with users claiming that the administration was attempting to weaken the Finns Party and censor dissenting views. More extreme abuse labelled the controversy a totalitarian seizure of power and part of Marin’s “feminist agenda”.

Issues surrounding the BLM movement, equality, and feminism continued to attract heightened levels of abusive activity throughout the end of June. On 20 June, a confrontation between Finnish police and a group of young individuals became violent. Ohisalo thanked the police for calming the situation, but condemned the officials’ use

of excessive force against minors. She was targeted with abusive messages blaming the incident on the Green’s open support of the BLM movement. Abuse levels rose again on 26 June, coinciding with the passage of a government equality programme intended to promote gender equality in Finland and in the country’s foreign policy. Users tweeted abuse at Ohisalo, criticising the initiative as extreme, communist, and depriving Finnish men of equality.

## July

Between 1 and 27 July, the final period of our data gathering, we observed an output of abusive messages similar to that of June. Nearly 60% of abuse was shared by anonymous accounts, while human-driven accounts sent 33% of abusive content. Bot activity hovered at around 3%. The first spike in abusive messaging occurred on 7 July, with Interior Minister Ohisalo’s announcement on Twitter that the Finnish Ministry of Justice would be reforming legislation on sexual offences to include consent.<sup>7</sup> She credited the #MeToo movement with making the reform possible. Abusive messaging ensued, with users criticising #MeToo as a movement against men. Several messages contained a racist element as well, as users claimed Ohisalo’s asylum policy welcomes rapists and pedophiles into the country. On the same day, the government decided to drop a plan to

---

<sup>6</sup> <https://twitter.com/MarinSanna/status/1270691667654443008>

<sup>7</sup> <https://twitter.com/MariaOhisalo/status/1280426332053295105>



## Examples: anger over EU

21 July:

@iltasanomat @MarinSanna Petit koko Suomen kansan. Hävettääkö huora, edes vähän? #maanpetturi #valtiopetos

Translation: @iltasanomat @MarinSanna You've betrayed the entire Finnish people. Are you ashamed, you whore, even a little? #traitor #hightreason

21 July:

@MarinSanna Häpeä "isätön" sateenkaari huora, HÄPEÄ. Toivottavasti saat vielä tuomiosi. #maanpetturi #valtiopetos #maanpetos #demarit

Translation: @MarinSanna Shame, you "fatherless" rainbow whore, SHAME. I hope you'll face judgement yet. #traitor #hightreason #treason #SocialDemocrats

21 July:

"Luulisi että köyhästä ja sairaasta sateenkaariperheestä tuleva maansa myynyt huora ja toverimme @MarinSanna 'n ymmärtävän sen, että jos tulot laskevat, niin silloin karsitaan myös menoja?!? #demarit #sateenkaarifriikit"

Translation: I would have thought that the whore from a sick and poor rainbow family, our comrade who has sold her country, @MarinSanna, would understand that if your income drops, you have to cut your expenses too?!? #SocialDemocrats #rainbowfreaks

adopt ankle monitors for unsuccessful asylum seekers, an announcement that came just as Finland was expected to receive 25 underage refugees from Greece (Helsinki Times, 2020). Users spouting abuse pointed to this as an example of how the current government invites dangerous refugees to Finland.

Foreign Minister Pekka Haavisto, who had thus far not featured prominently in our data, became the centre of abusive attention on 9 July. The spike in abusive activity coincided with reports that Haavisto was suspected of official misconduct the previous autumn,

when he decided to re-assign consular chief Pasi Tuominen following a disagreement over a plan to repatriate Finnish citizens from the al-Hol refugee camp in Syria (YLE, 2020). Accounts responding to these allegations with abusive language accused Haavisto of corruption and only remaining in his post because he is politically aligned with Prime Minister Marin's "green communist" ideology.

Spikes in abusive messaging escalated towards the end of the month, around 21–24 July. On 21 July, EU leaders agreed on a 750-billion-euro COVID-19 recovery



fund (Helsinki Times, 2020). The agreement generated an intensification of abusive activity, with nearly 4% of abusive tweets shared by bot accounts. Tytti Tuppurainen, Minister of European Affairs, announced the achievement on Twitter and was subsequently targeted with 192 abusive tweets. Users angry over the European relief agreement ridiculed Tuppurainen and Marin with accusations of treason, referring to them as “EU whores” and “euro-prostitutes”.

Finnish ministers faced EU-related criticism laced with abusive language throughout the remainder of our monitoring period.

Prime Minister Marin received the bulk of these messages. Users sharing abusive messages engaged with eurosceptic narratives, accusing Marin of funding the “Italian mafia” and paying for the economic mistakes of Southern Europe. Messages characterised contributions to the EU as punishment for Finland’s economic wealth and Marin as overseeing the bankruptcy of Finland. Additionally, we observed a number of messages containing the hashtag #FIXIT, referring to a Finnish exit from the European Union. Following 24 July, abusive activity appeared to taper off until the end of our collection period on 27 July.

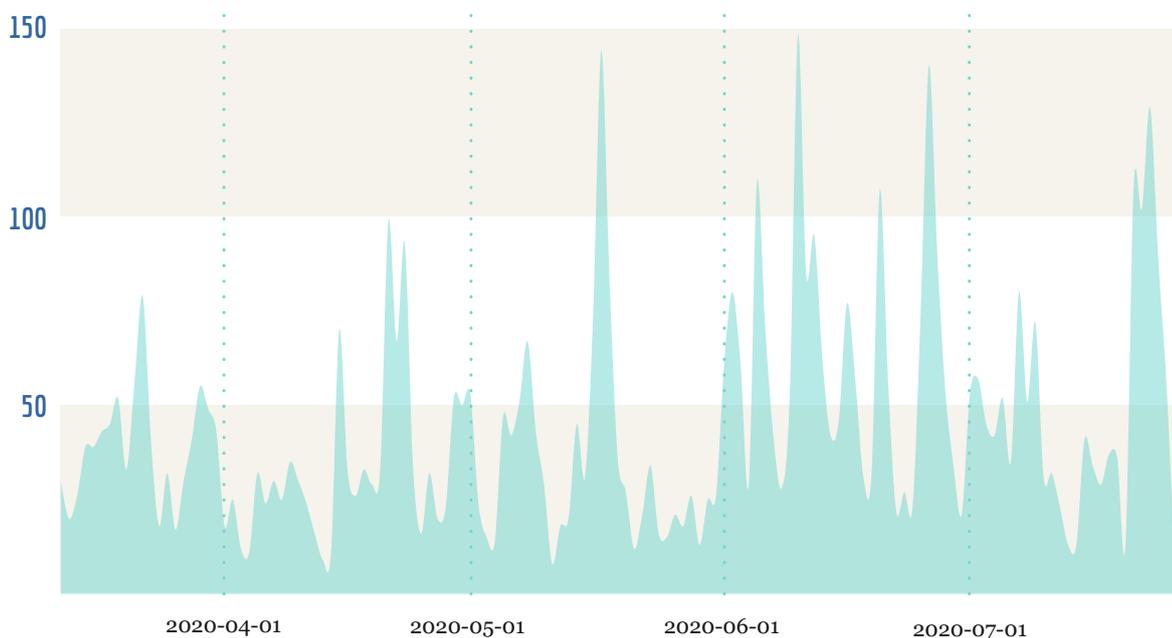


# Findings and Analysis

In this report, we analysed a collection of abusive tweets sent to Finnish politicians, either in reply to their own tweets or published independently. As a result of our in-depth analysis of abusive messages, we uncovered several findings about the nature of online abuse perpetrated against Finnish parliamentarians. To recap, we hypothesised that we would observe Finnish politicians being targeted with abusive language on Twitter. Additionally, we expected that female representatives would receive higher proportions of gendered abuse. Through our

quantitative and qualitative analyses, we gathered evidence to support both aspects of our hypothesis.

We observed that female Finnish ministers received a disproportionate number of abusive messages throughout our monitoring period. A startling portion of this abuse contained both latent and overtly sexist language. The five-most targeted ministers, all female, were overwhelmingly victimised by misogynistic abuse attacking their values,



**Figure 12: Timeline of abuse labelled sexist or homophobic**



## Examples of misogynistic abuse:

*16 March:*

Mitäs Huoru Li, onko tullut neekerin kyrpää imettyä paljon? Uskoisin että on kun oot tollanen juoppo huora. 😂

Translation: *What about the Whore Li, have you sucked a lot of n\*gger cock? I'd believe it, given you're an alcoholic whore.*

*25 June:*

@MariaOhisalo Haista vittu äläkä uhkaile idiootteja liberaaleja perheenjäseniäni saatanan huora

Translation: *@MariaOhisalo Go f\*ck yourself and don't threaten my family members, you liberal idiots, you f\*cking whore*

demeaning their decision-making skills, and questioning their leadership abilities. Many abusive tweets included gendered expletives, such as “b\*tch”, “whore”, and “slut”—among others—to degrade female ministers solely based on their identities as women. We found that female ministers were targeted with sexist messages regardless of the political event or announcement that prompted the increase in abusive activity; gendered language was used to criticise female ministers’ performance as government officials no matter what the topic.

Additionally, we noted multiple instances of female ministers receiving abusive tweets with sexually explicit language. Interior Minister Ohisalo in particular received sexually explicit abuse from users who disagree with her immigration policy,

conflating her sympathy for refugees as sexually motivated. In these instances, sexist and racist abuse was often combined. The misogynistic and objectifying language we observed may be intended to humiliate and intimidate female politicians, likely for the purpose of discouraging their participation in government, as has been found in previous studies.

The concept of feminism was repeatedly disparaged throughout our monitoring period. On several occasions, the terms “feminism” and “feminist” were appropriated by users attempting to insult female ministers. The term was used in phrases such as “feminist quintet” to undermine the authority of the female-led government or “feminist agenda” to portray female ministers as politically radical and, in some cases, totalitarian.



In their study of the treatment of female politicians on social media, Rheault, Rayment and Musulan (2019) found that, “the association between gender and the likelihood of being targeted is conditional on visibility: women who achieve a higher status in politics are more likely to receive uncivil messages than their male counterparts” (6). Given that in Finland’s traditionally male-dominated political sphere multiple leadership positions are currently occupied by women, visibility may play a role in the volume and type of abusive messages that female ministers received during the monitoring period. Although we could not determine whether the female ministers who received significant volumes of abuse were targeted due to their positions as high-ranking members of the Finnish government or due to their gender identity, we can definitively conclude that (1) female ministers received high levels of abuse online, and (2) this abuse was often permeated with sexist and misogynistic language.

### Topics triggering abusive activity

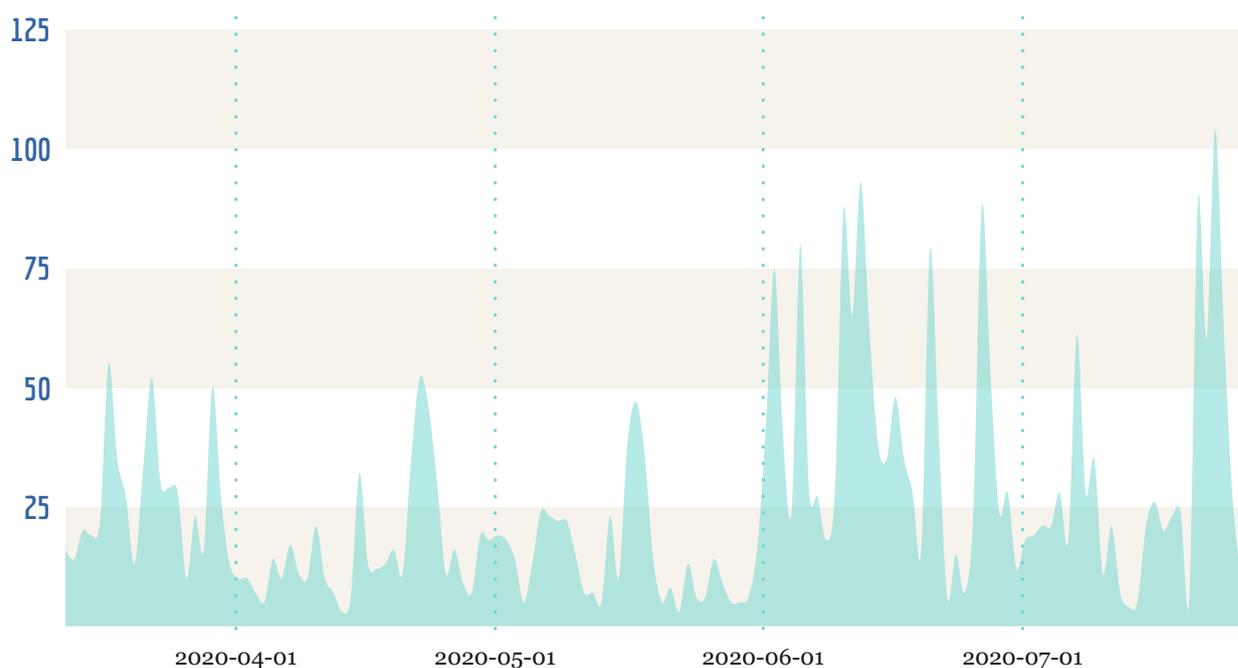
Our study further sought to explore the factors that influence the volume and type of abuse directed at Finnish politicians. To this end, we investigated the most significant spikes in abusive activity and identified the main topics of concern among users that tweet abuse, which include: the COVID-19 pandemic, immigration, the EU, and left-wing politics.

## COVID-19

When we began monitoring the Finnish information space, the escalation of COVID-19 infection rates in Europe drove several countries to take drastic public health measures, including Finland. Prime Minister Marin announced a state of emergency on 16 March. Assuming that a state of emergency, restrictions on public life, and an unprecedented global health crisis would lead to an increase in emotional speech of any form, we were surprised to find that more spikes in abusive activity occurred after the state of emergency was lifted on 1 June. COVID-19-related messaging did not appear organised as the topics of abusive messages ranged from criticising the government for not acting swiftly or strictly enough to accusing the current administration of inflating the crisis to secure their position in power.

The pandemic-triggered abusive messaging we observed consistently accused the government of incompetence and corruption. This was especially apparent in April amid the face mask procurement scandal, with many users calling for the resignation of multiple officials associated with the deal. We also noted examples of COVID-19 conspiracy theories. While some messages labelled the pandemic as a hoax, others spread the narrative that Prime Minister Marin and the “global left” are determined to carry out ethnic cleansing of native Finns.





**Figure 13: Timeline of abuse labelled racist or Islamophobic**

## Immigration

Immigration is a divisive topic in Finnish politics. Prime Minister Marin's left-leaning government supports a liberal asylum policy and several members of the coalition are outspoken about their pro-refugee stance. Immigration issues, particularly the topic of refugees, regularly triggered abusive activity throughout our monitoring period. This finding on its own is not exceptionally surprising, as the current political opposition in Finland supports tightening Finland's refugee policy. However, the degree of abusive language is striking. We observed multiple examples of extreme anti-immigrant

rhetoric and racial slurs in abusive tweets targeting politicians. Female politicians, notably Interior Minister Ohisalo, were targeted with sexually aggressive tweets due to their support for asylum seekers. These messages discriminate against and attempt to dehumanise people of colour. Additionally, by dehumanising people of colour and suggesting that left-wing female ministers' policy decisions are dictated by sexual desire, these users seek to undermine their legitimacy as government officials. Notably, we did not observe Foreign Minister Haavisto, a man, being targeted with the same type of sexually explicit messaging when he was at the centre of a refugee-related controversy.

## Examples of racism, anti-immigration:

16 May:

@TyttiTup Minä taas peräänkuulutan niitä miljoonia mitä vasemmistolaisporsaat syydätte ympäri maailmaa neekereille ja vastenmielisille muslimeille. Emme halua jakaa rahaa heille. Tarvitsemme ne itse. Ymmärtääkö höpön löppö?

Translation: @TyttiTup I, on the other hand, would like to know what happened to the millions your left-wing pigs send round the world to n\*ggers and horrible Muslims. We don't want to send them money. We need that money ourselves. Do you get that, idiot?

2 June:

Miksi Sanna tuet pikkutyttöjen sukupuolielinten silpomista. Miksi jaatte rahaa mielenvikaisille muslimeille jotka silpovat lapsia. Miksi vasemmistolaiset vihaa lapsia? Antaisitko rahaa oman lapsesi silpojalle?

Translation: Why, Sanna, do you support mutilating little girls' genitals? Why do you give money to retarded Muslims who mutilate children? Why does the left hate children? Would you give money to someone who mutilated your own child?

10 June:

Teidän muslimimaahantuontilemmikit jo raiskaavat 10-20-kertaisella innolla, että oot sä yksi saatanan idiootti.

Translation: Your Muslim immigrant pets are already raping with ten or twenty times the enthusiasm – you are one f\*cking idiot.

Examples of such immigration- and race-related sexual harassment have featured throughout this study.

Anti-immigrant abusive activity generally engaged with two core narratives. First, left-wing parties are more concerned with the health and wellbeing of refugees than those of the native Finnish population. Second, these same parties are determined to “bring” Islam to Finland. These narratives were represented in abusive activity responding to virtually all immigration-related developments.

## Finnish-EU relations

Another significant trigger in politically driven abusive activity was Finland's relationship with the European Union. The two highest peaks in abusive messages occurred on 21 July (871 tweets) and 23 July (711 tweets), coinciding with the finalisation of an EU plan to provide 750 billion euros of COVID-19 recovery funding. In these instances, messages containing abusive language expressed clearly Eurosceptic views. While the traditional



### Example of EU-related abuse:

21 July:

[@MarinSanna](#) Miten voit sitoa kaikki veronmaksajat tähän? Et ilmeisesti ymmärrä että olet tehnyt semmoisen virheen mikä saattaa jopa johtaa sisällissotaan. Lupaani omalta osaltani muistaa kirotta nimeäsi joka päivä, kunnes tuo raha on saatu takaisin. Saatanan Eu-huora!

Translation: [@MarinSanna](#) How can you plough taxpayers' money into this? You clearly don't understand that you've made the kind of mistake that could lead to a civil war. For my part, I vow to curse your name every day until that money is paid back. F\*cking EU whore!

right- and left-wing debates regarding EU policies, particularly spending, are normal deliberative contestation, the offensive language used by many of those tweeting abuse went beyond standard criticism. The content of these messages often criticised Finland's financial contribution to the EU, accusing the current administration of prioritising the interests of the global European elite over those of Finnish citizens. Once again, we observed female politicians targeted with gendered name-calling and expletives. The language used to express anger over Finland's membership in the EU mirrors that used to admonish Finland's refugee policy: the left-wing government places foreign interests—those of the EU and asylum seekers—ahead of domestic Finnish interests.

### Socially liberal politics

Discussions of socially liberal politics persistently attracted high levels of abusive activity throughout our monitoring period, highlighting the ideological divide between Finnish Twitter users. Within this general category, three topics stood out: LGBTQ+ rights, the international Black Lives Matter movement, and gender equality. The third-highest peak in abusive messaging occurred on 17 May, with ministers receiving 643 abusive tweets. This sudden and significant spike in abuse was driven by outrage at the ministers' public recognition of the International Day against Homophobia, Transphobia, and Biphobia. We observed abusive messages using vulgar and homophobic language to troll ministers. Abusive messages condemned left-wing ministers for spreading "rainbow flag propaganda", accused them of "teaching"



Finns to be LGBTQ+, and, in reference to transgender individuals, disseminated the outlandish claim that they want to give sexual predators access to girls' locker rooms.

Throughout June, Finnish politicians' support of the Black Lives Matter (BLM) movement and public condemnations of racism triggered torrents of abusive messages. During this time, several Finnish ministers tweeted about racial injustice and police violence in the US, advocating for equality. The abusive messages rejected the BLM protests as violent riots, with some claiming they were orchestrated by radical left-wing entities or intended to overthrow US president Trump. When analysing the tweets criticising ministers individually, we observed the emergence of a central theme: left-wing

ministers condemn racism in the US but are racist against Finns by bringing potentially violent refugees to Finland and prioritising their needs. Interior Minister Ohisalo was targeted in particular with abusive messages of this nature, including sexually explicit abuse.

As mentioned previously, issues involving gender equality and feminism were often discussed with abusive language. Two prime examples of this were the reactions to the proposed reform of legislation on sexual offences and to criticism of a book recently published by the Finns Party think tank containing "cruel and disturbing" views on women. Abusive tweets accused the left of pushing an extreme, anti-man agenda in which any dissenting opinions, especially those of the political opposition, are censored.



# Conclusions

Our investigation has demonstrated that the messaging directed at Finnish government ministers is largely free from automated activity. When it comes to abusive messaging, we find a number of users singularly focused on harassing the government. These accounts may be fake—certainly the owners of the accounts are not generally easily identified—but they do not appear to be tightly coordinated. The story told in the data is less about messages being sent from coordinated accounts, but rather a stream of abusive messages sent from a few accounts.

Our analysis has found that the targeting of Finnish government ministers with abusive language on Twitter is a common occurrence. This study further revealed the troubling extent to which female ministers receive gendered, sexist, and misogynistic abuse online. On Finnish Twitter, not only did female politicians receive more abuse than their male counterparts, but the abuse displayed a gendered pattern. Gendered abuse was used to criticise and delegitimise women in ministerial positions no matter the political topic of the moment, be it the Finnish government's COVID-19 response, its immigration policy, or its involvement in EU affairs. Our analysis suggests that female ministers are likely to be targeted with abusive messages by users aligning themselves with the political opposition,

which in this case was the right. These findings reinforce those of previous studies; the phenomenon of online harassment against female parliamentarians knows no boundaries and is prevalent even in Finland, a country that ranks among the best in the world in terms of gender parity (World Economic Forum, 2020).

As social media platforms continue to grow in political importance, so does their use as a means for engaging with and criticising individual government officials with little or no consequences. An additional aim of our study was to determine the role, if any, bot accounts play in disseminating abusive messages, and whether such bot activity displayed characteristics of coordination. Based on previous Finnish studies analysing the impact of bots during election periods, we hypothesised that we would observe low levels of automation and coordination. Our findings confirmed this theory; our algorithm attributed less than 3% of abusive messages to bot-like accounts. However, the more significant finding was that over half of abusive messages were sent by anonymous accounts. Anonymity erases accountability online. This can have the effect of emboldening users to voice their dissatisfaction with ministers through unfiltered, abusive messages. It is possible for people to operate many anonymous



accounts. However, our data do not show clear patterns indicating single users sending abusive messaging from multiple fake accounts. The unfortunate conclusion is that much of the offensive, sexually explicit, expletive-filled abuse targeting government officials is written and published by individuals.

What can states do to address or mitigate the problem of online abuse of public servants? Governments must manoeuvre a regulatory gray area in which it can be difficult to distinguish between freedom of speech and protection from harmful verbal abuse. Continuing to shed light on the problem, which lacks comprehensive study, would raise public awareness of the extent of politically motivated abuse online and perhaps lead to creative solutions. Government officials themselves should address the phenomenon, uniting to support one another and set a positive example of online conduct.

Content moderation is ultimately the responsibility of social media and big tech companies. Social media platforms, Twitter included, are far more adept at moderating content in mainstream languages, most notably English. We expect to witness the development of powerful tools drawing on advances in artificial intelligence to understand content across less-widely-spoken languages and allow for the analysis of content with a higher degree of language variation. As a result, such technology would ensure more equitable security measures across the linguistically diverse digital space, ultimately benefiting the smaller language branches of the Nordic and Baltic regions. Finnish-language Twitter appears to have been comparatively shielded from coordinated inauthentic manipulation, in part due to the complexity of the local language. It remains to be seen how long this relative protection will last; advances in artificial intelligence may remove this barrier to manipulation.



## ✦ Implications for NATO

Finland is not a NATO member, and many of these findings apply primarily to Finland. Nevertheless, this report does contain generalisable information of value also to the Alliance.

This report serves as a reminder that a large slice of hostile messaging is created by authentic domestic voices. Coordinated inauthentic activity is a real threat and a serious challenge, but it should not serve as the default label to describe unwelcome criticism on social media.

This report can be considered in conjunction with the quarterly *Robotrolling* published by the NATO Strategic Communications Centre of Excellence. The report tracks bot activity in social media conversations about the NATO presence in the Baltics and Poland. Using the same algorithm,

we find consistently high levels of bot activity, especially in Russian-language content, whereas we find negligible quantities when analysing Finnish-language content. This demonstrates that the online space, even within a single platform, is not monolithic. Some subjects are more likely to be the target of bot activity than others. Some languages are more vulnerable than others. And it appears, at least compared to the example of Finnish politicians, that Russian-language content about NATO acts as something of a bot magnet—activity from automated accounts is many times higher.



# Bibliography

Ahlgren, John, Maria Eugenie Berezin, Kinga Bojarczuk, Elena Dulskyte, Inna Dvortsova, Johann George, Natalija Gucevskaja, Mark Harman, Ralf Lämmel, Erik Meijer, Silvia Sapora, and Justin Spahr-Summers (2020) "WES: Agent-based User Interaction Simulation on Real Infrastructure."

Amnesty International (2018) "Toxic Twitter." Report in *Amnesty International Research*.

Andersson, Li (29 March 2020) "Li Andersson: Luokkaristiriitoja poikkeusoloissa" *Kansan Uutiset*.

Baca, Marie C. and Tony Romm (2019) "Twitter and Facebook take first actions against China for using fake accounts to sow discord in Hong Kong." *The Washington Post*.

Badjatiya, Pinkesh, Shashank Gupta, Manish Gupta, and Vasudeva Varma (2017) "Deep Learning for Hate Speech Detection in Tweets." In *International World Wide Web Conference Committee (IW3C2)*.

Bartlett, Jamie, Richard Norrie, Sofia Patel, Rebekka Rumpel, and Simon Wibberley (2014) "Misogyny on Twitter." Report published by DEMOS.

Bateman, Tom (5 June 2020) "Katri Kulmuni: Finnish minister quits over media training row" *BBC News*.

Bay, Sebastian and Rolf Fredheim (2019) "Falling Behind: How Social Media Companies are Failing to Combat Inauthentic Behaviour Online." *NATO StratCom COE*.

Ben-David, Anat and Ariadna Matamoros-Fernandez (2016) "Hate Speech and Covert Discrimination on Social Media: Monitoring the Facebook Pages of Extreme-Right Political Parties in Spain." *International Journal of Communication*, 10: 1167–93.

Boyd, Ryan L., Alexander Spangher, Adam Fourney, Besmira Nushi, Gireeja Ranade, James W. Pennebaker, and Eric Horvitz (2018) "Characterizing the Internet Research Agency's Social Media Operations During the 2016 US Presidential Election using Linguistic Analyses."

Bradshaw, Samantha, and Philip N. Howard (2018) "Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation." *Project on Computational Propaganda*. Working Paper 2018, 1.

Bruns, Axel (2019) "After the 'APocalypse': social media platforms

and their fight against critical scholarly research." *Information, Communication & Society*, 22: 1544–66.

Burnap, Pete and Matthew L. Williams (2015) "Cyber Hate Speech on Twitter: An Application of Machine Classification and Statistical Modeling for Policy and Decision Making." *Policy & Internet*, 7(2): 223–41.

Clement, J. (2020) "Global social networks ranked by number of users 2020." *Statista*.

Dansby, Ryan, Han Fang, Hao Ma, Chris Moghbel, Umot Ozertem, Xiaochang Peng, Ves Stoyanov, Sinong Wang, Fan Yang, and Kevin Zhang (2020) "AI advances to better detect hate speech." *Facebook AI*.

Davidson, Thomas, Dana Warmesley, Michael Macy, and Ingmar Weber (2017) "Automated Hate Speech Detection and the Problem of Offensive Language." In *Proceedings of the International AAAI Conference on Web and Social Media*.

de Gibert, Ona, Naiara Perez, Aitor Garcia-Pablos, and Montse Cuadros (2018) "Hate Speech Dataset from a White Supremacy Forum."

Dhrodia, Azmina (2017) "Unsocial Media: Tracking Twitter Abuse against Women MPs." *Amnesty Global Insights*.

European Commission (2020) "The code of conduct on countering illegal hate speech online."

Facebook (2020) "Community Standards."

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

Filer, Tanya, and Rolf Fredheim (2016) "Sparking debate? Political deaths and Twitter discourses in Argentina and Russia." *Information, Communication & Society* 19(11): 1539–55.

Fuchs, Tamara and Fabian Schäfer (2020) "Normalizing misogyny: hate speech and verbal abuse of female politicians on Japanese Twitter." *Japan Forum*, 0(0): 1–27.

Greevy, Edel and Alan F. Smeaton (2004) "Classifying Racist Texts Using a Support Vector Machine." In *Proceedings of the 27th Annual International Conference on Research and Development in Information Retrieval*: 468–69.



- Gorrell, Genevieve, Mark A. Greenwood, Ian Roberts, Diana Maynard, and Kalina Bontcheva (2018) "Twits, Twats and Twaddle: Trends in Online Abuse towards UK Politicians." In *Proceedings of the Twelfth International AAAI Conference on Web and Social Media*.
- Guo, Ruiqi, Philip Sun, Erik Lindgren, Quan Geng, David Simcha, Felix Chern, and Sanjiv Kumar (2020) "Accelerating Large-Scale Inference with Anisotropic Vector Quantization." *International Conference on Machine Learning*.
- Hardaker, Claire and Mark McGlashan (2015) "Real men don't hate women': Twitter rape threats and group identity." *Journal of Pragmatics*, 91: 80–93.
- Hegelich, S. and Janetzko, D. (2016) "Are social bots on Twitter political actors? Empirical evidence from a Ukrainian social botnet." In *Tenth International AAAI Conference on Web and Social Media*.
- Hellman, Maria and Charlotte Wagnsson (2017) "How can European states respond to Russian information warfare? An analytical framework." *European Security*, 26(2): 153–70.
- Helsingin Sanomat (8 May 2020) "[STM:n koronajohtaja: Epidemiaa ei voi tukahduttaa, koska silloin rajoitustoimien edellytykset poistuisivat – Marin: 'Ei edusta hallituksen linjaa.'](#)"
- Helsinki Times (18 March 2020) "[Finland to close borders to non-essential travel at 12am on Thursday.](#)"
- Helsinki Times (7 June 2020) "[Ohisalo: Finland drops plan to adopt ankle monitors for unsuccessful asylum seekers.](#)"
- Helsinki Times (21 July 2020) "[EU leaders strike agreement on €750bn recovery fund – €390bn to be disbursed in grants.](#)"
- Hewitt, Sarah, Dr. T. Tiropanis, and Dr. C. Bokhove (2016) "The Problem of Identifying Misogynist Language on Twitter (and other online social spaces)." In *Proceedings of the 2016 ACM Web Science Conference*.
- Howard, Philip N. and Bence Kollanyi (2016) "Bots, #Stronger in, and #Brexit: Computational Propaganda during the UK-EU Referendum." *SSRN Electronic Journal*.
- Howard, Philip N., Samuel Woolley, and Ryan Calo (2018) "Algorithms, bots, and political communication in the US 2016 election: The challenge of automated political communication for election law and administration." *Journal of Information Technology and Politics*.
- Imperva (2016) "Bot Traffic Report 2016."
- Inter-Parliamentary Union (2016) *Sexism, Harassment and Violence against Women Parliamentarians*, Issue Brief.
- Johnson, Jeff, Matthijs Douze, and Hervé Jégou (2017) "[Billion-scale similarity search with GPUs.](#)"
- Jyrkiäinen, Jyrki (2017) "Finland." *Media Landscapes*.
- Kemp, Simon (2020) "Digital 2020: Finland." *Datareportal*.
- Kiela, Douwe, Hamed Firooz, and Aravind Mohan (2020) "Hateful Memes Challenge and data set for research on harmful multimodal content." *Facebook AI*.
- Knuutila, Aleks, Heidi Kosonen, Tuija Saresma, Paula Haara and Reeta Pöyhtäri (2019) "Viha vallassa: Vihapuheen vaikutukset yhteiskunnalliseen päätöksentekoon." *Valtioneuvoston selvitys- ja tutkimustoiminnan julkaisusarja*, 2019: 57.
- Kollanyi, Bence, Philip N. Howard, and Samuel C. Woolley (2016) "Bots and Automation over Twitter during the US election." *Project on Computational Propaganda*. Data Memo, 4.
- Maiya, Arun S. (2020) "[A Low-Code Library for Augmented Machine Learning.](#)"
- Mantilla, Karla (2013) "Gendertrolling: Misogyny Adapts to New Media." *Feminist Studies*, 39(2): 563–70.
- Martin, Diego A., Jacob N. Shapiro, and Michelle Nedashkovskaya (2019) "Recent Trends in Online Foreign Influence Efforts." *Journal of Information Warfare*, 18(3): 15–48.
- Mäki, Markus (22 April 2020) "[Finland convicted of violating human rights treaty last year—KRP launches preliminary investigation: suspects Iraqi man still alive.](#)" *YLE*.
- Mercieca, Jennifer R. (2019) "Dangerous Demagogues and Weaponized Communication." *Rhetoric Society Quarterly*, 49: 264–79.
- Miro-Llinares, F and J.J. Rodriguez-Sala (2016) "Cyber Hate Speech on Titter: Analyzing Disruptive Events from Social Media to Build a Violent Communication and Hate Speech Taxonomy." *International Journal of Design & Nature Ecodynamics*, 11(3): 406–15.
- Nockleby, John T. (2000) "Hate Speech." In Leonard W. Levy, Kenneth L. Karst, and Dennis J. Mahoney, editors, *Encyclopedia of the American Constitution: 1277–79*.
- Pew Research Center (2017) "Online Harassment 2017."
- Pohjonen, Matti and Sahana Udupa (2017) "Extreme Speech Online: An Anthropological Critique of Hate Speech Debates." *International Journal of Communication*, 11.
- Reimers, Nils, and Iryna Gurevych (2019) "[Sentence-bert:](#)



[Sentence embeddings using siamese bert-networks.](#)

Reuters (2 June 2019) "Finland's center-left coalition concludes talks to form new government."

Rheault, Ludovic, Erica Rayment, and Andreea Musulan (2019) "Politicians in the line of fire: Incivility and the treatment of women on social media." *Research and Politics*: 1–7.

Rossi, Sippo (2019) "Detecting and Analyzing Bots on Finnish Political Twitter." *Aalto University School of Business Information and Service Management*. Helsinki, Finland.

Salloum, Ali, Tuomas Takko, Markus Peuhkuri, Raimo Kantola, and Mikko Kivelä (2019) "Botit ja informaatiovaikuttaminen Twitterissä Suomen eduskunta- ja EU-vaaleissa 2019." *Aalto University ELEBOT Project*. Helsinki, Finland.

Schmidt, Anna and Michael Wiegand (2017) "A Survey on Hate Speech Detection using Natural Language Processing." In *Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media*: 1–10.

Shubber, Kadhim (7 June 2020) "[US 'Black Lives Matter' protests go global.](#)" *Financial Times*.

Silva, Leandro, Mainack Mondal, Denzil Correa, Fabricio Benevenuto, and Ingmar Weber (2016) "Analyzing the Targets of Hate in Online Social Media." In *Proceedings of the Tenth International AAAI Conference on Web and Social Media*.

Skinner, Helena (2019) "Lebanon protests: are bots fuelling counter demonstrations?" *Euronews*.

Southern, Rosalynnd and Emily Harmer (2019) "Twitter, Incivility and 'Everyday' Gendered Othering: An Analysis of Tweets Sent to UK Members of Parliament." *Social Science Computer Review*, 20(10): 1-17.

Specia, Megan (10 December 2019) "Who is Sanna Marin, Finland's 34-Year-Old Prime Minister?" *The New York Times*.

Steward, Leo, Ahmer Arif, and Kate Starbird (2018) "Examining Trolls and Polarization with a Retweet Network." In *MIS2*.

Tanner, Jari (15 April 2019) "Riding populist wave, Finns Party nearly wins Finland vote" *Associated Press*.

Theocharis, Yannis, Pablo Barbera, Zoltan Fazekas, and Sebastian Adrian Popa (2016) "A Bad Workman Blames His Tweets: The Consequences of Citizens' Uncivil Twitter Use when Interacting with Party Candidates." *Journal of Communication*, 66(6): 1007–31.

Ünver, Hamid Akin (2018) "Digital Challenges to Democracy:

Politics of Automation, Attention and Engagement." *Journal of International Affairs*, 71: 127–46.

Valtioneuvosto Statsrådet (14 November 2019) "[European Court of Human Rights gives judgement concerning Finland in the case of the return of an asylum seeker to Iraq.](#)"

van Spenje, Joost and Claes de Vreese (2013) "The good, the bad and the voter: The impact of hate speech prosecution of a politician on electoral support for his party." *Party Politics*, 1–16.

Varol, Onur, Emilio Ferrara, Clayton A. Davis, Filippo Menczer, and Alessandro Flammini (2017) "[Online Human-Bot Interactions: Detection, Estimation, and Characterization.](#)"

Vidgen, Bertie and Leon Derczynski (2020) "[Directions in Abusive Language Training Data: Garbage In, Garbage Out.](#)"

Vilmer, Jean-Baptiste J., Alexandre Escorcía, Marine Guillaume and Janaina Herrera (2018) "Information Manipulation: A Challenge for Our Democracies." Technical report, *Policy Planning Staff (CAPS) of the Ministry for Europe and Foreign Affairs and the Institute for Strategic Research (IRSEM) of the Ministry for the Armed Forces*.

Virtanen, Antti, Jenna Kanerva, Rami Ilo, Jouni Luoma, Juhani Luotolahti, Tapio Salakoski, Filip Ginter, and Sampo Pyysalo (2019) "[Multilingual is not enough: BERT for Finnish.](#)"

Ward, Stephen and Liam McLoughlin (2020) "Turds, traitors and tossers: the abuse of UK MPs via Twitter." *The Journal of Legislative Studies*, 26(1): 47–73.

Waseem, Zeerak, Thomas Davidson, Dana Warmusley, and Ingmar Weber (2017) "Understanding Abuse: A Typology of Abusive Language Detection Subtasks." *Proceedings of the First Workshop on Abusive Language Online*, 78–84.

Woolley, Samuel C. (2016) "Automating power: Social bot interference in global politics." *First Monday*, 21(4).

Woolley, Samuel C. and Philip N. Howard (2016) "Political Communication, Computational Propaganda, and Autonomous Agents." *International Journal of Communication*, 10(2016): 4882–90.

World Economic Forum (2020) "[Global Gender Gap Report 2020.](#)" *Insight Report*.

Yang, Yinfei and Fangxiaoyu Feng (2020) "Language-Agnostic BERT Sentence Embedding." *Google AI Blog*.

YLE (12 June 2017) "[Finnish govt implodes: Centre, NCP say deal off with Finns Party.](#)"



YLE (15 April 2019) "SDP takes top spot, populists rally, heavy losses for PM's Centre in Finnish election."

YLE (4 December 2019) "Finnish PM Rinne resigns."

YLE (9 April 2020) "Agency boss admits paying out millions of euros without getting hospital-grade PPE."

YLE (15 April 2020) "Finland to increase WHO funding."

YLE (21 April 2020) "Gov't to audit Business Finland's 100k coronavirus grants."

YLE (1 May 2020) "Vasemmistoliiton Li Andersson: Sadan euron elvytysseteli kaikille – koronakriisintalouden vaikutukset huolestuttivat vappupuhujia."

YLE (11 June 2020) "Thursday's papers: Finns Party in sexist book storm, citizen's initiative on masks and STD spike."

YLE (8 July 2020) "Daily: Foreign Minister suspected of two offences over al-Hol flap."

















