Xia, Yan; Gronow, Antti; Malkamäki, Arttu; Ylä-Anttila, Tuomas; Keller, Barbara; Kivelä, Mikko

The Russian invasion of Ukraine selectively depolarized the Finnish NATO discussion on Twitter

Published in:
EPJ Data Science

DOI:
10.1140/epjds/s13688-023-00441-2

Published: 03/01/2024

Document Version
Publisher's PDF, also known as Version of record

Xia et al. *EPJ Data Science* (2024) 13:1
https://doi.org/10.1140/epjds/s13688-023-00441-2

EPJ Data Science
a SpringerOpen Journal

EPJ.org

REGULAR ARTICLE Open Access

The Russian invasion of Ukraine selectively depolarized the Finnish NATO discussion on Twitter

Yan Xia[1*], Antti Gronow[2], Arttu Malkamäki[2], Tuomas Ylä-Anttila[2], Barbara Keller[1] and Mikko Kivelä[1]

*Correspondence: yan.xia@aalto.fi
[1]Department of Computer Science, Aalto University, Espoo, Finland
Full list of author information is available at the end of the article

Abstract

It is often thought that an external threat increases the internal cohesion of a nation, and thus decreases polarization. We examine this proposition by analyzing NATO discussion dynamics on Finnish social media following the Russian invasion of Ukraine in February 2022. In Finland, public opinion on joining the North Atlantic Treaty Organization (NATO) had long been polarized along the left-right partisan axis, but the invasion led to a rapid convergence of opinion toward joining NATO. We investigate whether and how this depolarization took place among polarized actors on Finnish Twitter. By analyzing retweet patterns, we find three separate user groups before the invasion: a pro-NATO, a left-wing anti-NATO, and a conspiracy-charged anti-NATO group. After the invasion, the left-wing anti-NATO group members broke out of their retweeting bubble and connected with the pro-NATO group despite their difference in partisanship, while the conspiracy-charged anti-NATO group mostly remained a separate cluster. Our content analysis reveals that the left-wing anti-NATO group and the pro-NATO group were bridged by a shared condemnation of Russia's actions and shared democratic norms, while the other anti-NATO group, mainly built around conspiracy theories and disinformation, consistently demonstrated a clear anti-NATO attitude. We show that an external threat can bridge partisan divides in issues linked to the threat, but bubbles upheld by conspiracy theories and disinformation may persist even under dramatic external threats.

Keywords: Political polarization; Social media; External threat; Conspiracy theory; Disinformation

1 Introduction

Despite a period of momentum building, the Russian invasion of Ukraine on Feb 24, 2022 came as a shock to most observers. The shock was most acute in Ukraine but was felt also in countries bordering Russia. Finland, the militarily non-aligned European country that shares a 1344-kilometer border with Russia, witnessed a sharp shift in its public opinion on NATO membership, based on a reappraisal of the external threat posed by Russia. Traditionally, around 20 percent of the Finnish population had been in favor of joining NATO [1]. Russia's invasion of Crimea in 2014 increased the number to 25–30 [1], but

Springer

after the invasion of Ukraine in 2022, support for joining NATO soared as high as 70–80 percent [2].

Behind this major change in opinion, the rising external threat seems to have had a depolarizing effect on the Finnish NATO discussion. For long, Finnish opinions on NATO embodied a polarization that was largely partisanship-based: voters of the main right-wing party (National Coalition) were largely in favor of joining, whereas voters of left-wing parties were the most vocal opponents of NATO [3]. After the invasion, however, many left-wing supporters changed their opinion, and eventually the Finnish parliament almost unanimously voted in favor of joining NATO (188 for, 8 against).

Social media opens an unobtrusive observation window [4] into whether and how this depolarization took place among the more politically active and partisan segment of the population [5, 6], including political elites who often play an important role in steering the discussion [7, 8], as well as fringe communities that subscribe to conspiracy theories and disinformation [9–11]. The digital traces of user interactions make it possible to measure structural polarization by constructing endorsement networks of individuals and observing cohesive groups in them [12–15], and provide insight into the information spreading and user interaction dynamics that drive opinion (de)polarization [16, 17]. While network analysis can reveal the structure of user interactions and how it changes over time [18], content analysis can uncover how the discussion climate evolves and what arguments connect or distinguish opposing sides [19, 20].

Previously, researchers have used social media data during the Russian invasion of Ukraine to study people's opinions and emotions toward the incident [21–24], how it changed people's sentiments toward green energy [25], and the spread of Russian propaganda and misinformation online [26, 27]. Closest to our work, Nisch [28] studied social media discussions on Finland's NATO membership. However, they were interested in the world's reaction to Finland's NATO application, and thus analyzed data from English Twitter after Finland announced its intention to join NATO in May 2022. By contrast, we are interested in the immediate depolarizing effect of the invasion on the NATO discussion among Finnish actors.

Using a combination of network analysis and content analysis methods, we inspect how the Russian invasion of Ukraine changed the polarization dynamics of the Finnish NATO discussion on Twitter. We mainly analyze Finnish language tweets from Feb 10, 2022 to Mar 30, 2022 that contain any NATO-related keyword. By clustering the user retweet network before the invasion and reading through a sample of tweets in each user cluster, we identify three separate user groups: one pro-NATO group, one anti-NATO group of left-wing partisans, and another anti-NATO group of conspiracy theory and disinformation consumers. We find that depolarization took place among Twitter users after the invasion, but in a selective manner. In contrast to the expectation that partisanship-based polarization is considerably resistant to external threats [29], the invasion rapidly depolarized partisan actors: the left-wing anti-NATO group started retweeting much more from the pro-NATO group right after the invasion, while they also started posting less anti-NATO content. Instead, the conspiracy-charged anti-NATO group retained strong in-group retweeting and a consistent anti-NATO attitude. As opposed to previously reported evidence of bots (i.e., automated accounts) playing an important role in spreading disinformation [30] and Russian propaganda [26] on the English Twitter, the conspiracy-charged anti-NATO group in our data did not seem to be populated by bots, but more

likely consisted of spontaneous consumers of conspiracy theories and disinformation who exhibited persistent opinions and communication patterns.

Our study adds new empirical evidence to the long-held theory that external conflicts increase internal cohesion [31] by showing that this process may happen selectively. While limited partisan depolarization has been observed in the face of external threats in the United States context [29], we show that partisan actors can be depolarized overnight by a dramatic external threat in a way similar to the rally-around-the-flag reaction to terrorist and other direct attacks [32]. Meanwhile, in line with previous findings that consumers of conspiracy theories tend to form echo chambers [33, 34] and concentrate on within-community content [35], our results further underscore the resilience and danger of conspiracy-/disinformation-based polarization by showing how it can survive an external threat that even bridges partisan divides.

2 Data and methods

We collected all tweets in the Finnish language from Dec 30, 2021 to Mar 30, 2022 that contain any NATO-related keyword (see Appendix for the list of keywords) using the Twitter Application Programming Interface (API) v2 [36]. This gave us 320,407 tweet records produced by 28,887 users in total. We divided the timeline into two-week periods before Feb 24 and one-week periods after Feb 24, in consideration of the asymmetric activity level before and after the Russian invasion of Ukraine. For each period, we constructed a retweet network of users, where a directed link with weight w connects user A to user B if A retweeted B w times within the period. We used only retweet records (not including quote retweets) for constructing the user networks, as retweet is a relatively certain indicator of endorsement-based connection [37]. Following prior work on quantifying polarization in Twitter data [14, 18], we used only the largest connected component of each network for subsequent analysis.

Based on an observation of the retweet networks, we decided to focus our analysis mainly on four periods that are representative of the evolving retweeting dynamics: *before* (Feb 10 to Feb 23, 31,399 total tweets, 12,891 retweets), *right-after* (Feb 24 to Mar 2, 81,433 total tweets, 39,936 retweets), *1-week-after* (Mar 3 to Mar 9, 49,585 total tweets, 23,365 retweets), and *4-weeks-after* (Mar 24 to Mar 30, 20,792 total tweets, 9103 retweets). The retweet network (largest connected component) contains 3836 users and 10,774 links in the *before* period, 8986 users and 32,454 links in the *right-after* period, 6173 users and 19,309 links in the *1-week-after* period, and 3383 users and 7598 links in the *4-weeks-after* period.

In order to track stance changes induced by the invasion, we performed our analysis on users who were active in the *before* network. Using the Leiden graph partitioning algorithm [38], we find clusters of users in the *before* network who retweet mainly within their cluster, and inspect how these users change their behavior and stances in the subsequent periods. We ran the Leiden algorithm 50 times, each time with a different resolution parameter; for each run, the algorithm returned a partition of maximized modularity. In order to select the most appropriate resolution and number of communities, we chose the partition that gives the shortest description length of the data under the weighted stochastic block model [39], where link weights (i.e., count of retweets between a pair of users) are treated as covariates sampled from a binomial distribution. For each of the four time periods, we then calculated for each user cluster the number of external retweets, the

number of internal retweets, and the external-internal (E/I) ratio (i.e., the ratio between the number of external retweets and the number of internal retweets), in order to examine the change in intra-cluster and inter-cluster communication dynamics.

To get a sense of changes in the content of the discussion, we sampled a number of tweets from the data for manual content analysis. For each cluster and each of the four time periods, we randomly sampled 42 tweets from those that got retweeted at least once in the cluster in the period, which resulted in 504 sampled tweets (see Appendix for more sampling statistics). We preferentially sampled tweets that were popular within each cluster by setting the sampling probability of each tweet proportional to its number of in-cluster retweets in the period. A group of four coders (all co-authors) then labeled the stance of each tweet to be pro-NATO, anti-NATO, unclear, or unrelated to NATO. The coders were split into two teams, with two coders on each team. From the 504 tweets in total, 24 tweets were randomly sampled for both teams to code; for the remaining 480, one team coded half and the other team coded the remaining half. Within each team, each coder first labeled the 264 tweets independently, then the two coders discussed cases of disagreement and reached a consensus as a team. The inter-team agreement for the 24 double-coded tweets, as evaluated by Krippendorff's alpha [40], is 0.80.

3 Results

The graph partitioning algorithm reveals three clusters of users in the *before* network (Fig. 1A). Based on the coded stances of sampled tweets in each user group, we find one of the groups to be pro-NATO (hereinafter referred to as "the pro group") and the other two to be anti-NATO (Fig. 1F-H). A qualitative reading of the sampled tweets suggests that one of the anti-NATO groups based their arguments on traditional leftists' concerns, such as pacifism and feminism not being compatible with joining a military alliance, and NATO having been involved in the violation of human rights. We will hereinafter refer to this group as "the left-anti group". The other anti-NATO group showed a clear engagement in conspiracy theories and disinformation in framing their opposition to NATO. For example, they claimed NATO equals supporting globalism, the global elite, and the World Economic Forum, all of which are supposed co-conspirators that are set out to destroy the Finnish nation, and that those who want people to inject themselves with poisonous vaccines are the ones who want to join NATO. We will hereinafter refer to this group as "the conspiracy-anti group". It is worth noting that we did not observe a prevalence of bot-like accounts in our sample, which aligns with previous studies that found limited bot activity in political discussions on Finnish Twitter [41, 42].

Our user partisanship analysis further enriches the profile of each group. Specifically, we plot a list of Finnish politician accounts in the *before* network, colored by their publicly available party affiliation (Fig. 1B). In our analysis, we focus on the six main parties in Finland, each with over 10 Members in the current Finnish Parliament: the Left Alliance (Left), the Social Democratic Party (SDP), the Green League (Green), the Centre Party (Centre), the National Coalition Party (Coalition), and the Finns Party (Finns). Quite surprisingly, we find that politicians of most parties, including many that traditionally took a neutral or an anti-NATO stance, already fell on the pro-NATO side in the *before* network; presumably, this results from the buildup to the war since the end of 2021. However, politicians affiliated with the Left Alliance – the traditionally most anti-NATO party – still fell exclusively in the left-anti group. Meanwhile, the conspiracy-anti group seems

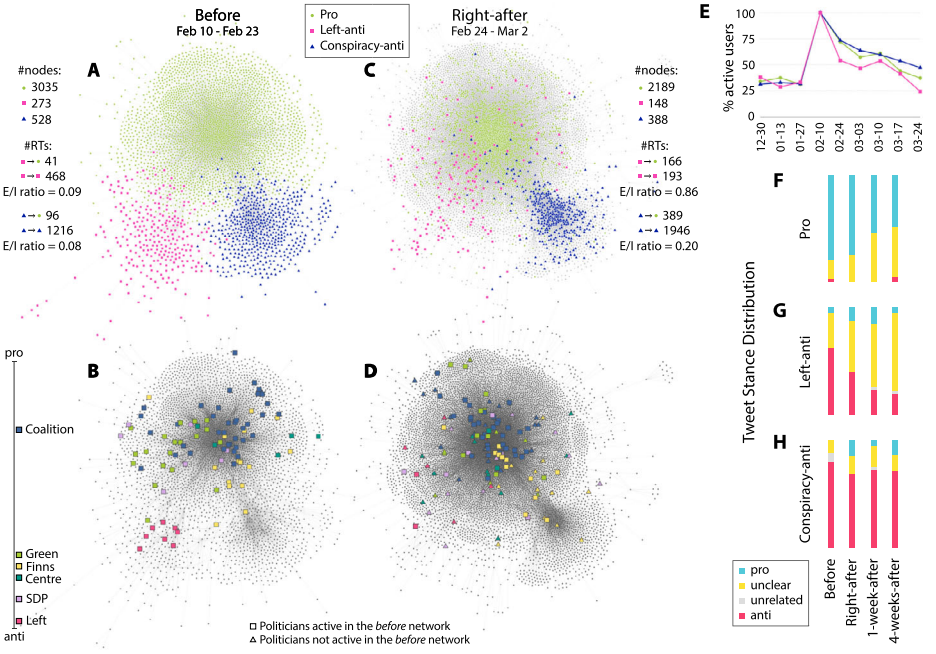

Figure 1 *Retweet networks and tweet stance distributions before and after the invasion.* Retweet networks (A)-(B) before and (C)-(D) right after the Russian invasion of Ukraine. Networks are drawn using the SFDP spring-block layout [43]. Node colors in (A) and (C) correspond to the three groups detected in the *before* network, and the statistics beside each network show the number of users, the number of external retweets of the pro group, the number of internal retweets, and the E/I ratio in each anti group. Node colors in (B) and (D) denote the party affiliation of politicians. Parties in the legend are positioned based on the mean of their candidates' attitude toward NATO in 2019, according to an election poll conducted by the Finnish Broadcasting Company (Yle) [44]. (E) Change in percentage of active users in each group. Change in distribution of stances on joining NATO among sampled tweets in the (F) pro, (G) left-anti, and (H) conspiracy-anti group

to accommodate few politicians, which indicates its relatively fringe position in political communication.

3.1 Change in network structure

Plotting the pro group, left-anti group, and conspiracy-anti group members in the retweet networks after the invasion, we observe a substantial change in the network structure. In the *right-after* network, members of the left-anti group became much less connected internally and more connected to the pro group, while most members of the conspiracy-anti group largely remained in their own internally connected bubble (Fig. 1C). This observation is confirmed by the number of external retweets of the pro group, the number of internal retweets, and the E/I ratio in each anti group (Table 1): although the E/I ratio of the conspiracy-anti group also more than doubled in the first week after the invasion (some of this change might be explained by overfitting [45]), the E/I ratio of the left-anti group had an almost tenfold increase in the same period. This structural change in the retweet network also remains in the *1-week-after* and *4-weeks-after* periods (see Appendix for a visualization of the retweet networks in these two periods). A statistical test further confirms that the larger change of E/I ratio in the left-anti group than in the conspiracy-anti group is not explained by statistical fluctuations (see Appendix).

The change in retweet network structure reflects a breakage of the cohesive cluster formed by the left-anti group members, as they instantly developed connection and align-

Table 1 Retweet network statistics. Number of active users in each group and the E/I ratio (i.e., the ratio between the number of external retweets of the pro group and the number of internal retweets) of the two anti groups in each time period

	Number of users			E/I ratio	
	Pro	Left-anti	Conspiracy-anti	Left-anti	Conspiracy-anti
Before	3035	273	528	41/468 = 0.09	96/1216 = 0.08
Right-after	2189	148	388	166/193 = 0.86	389/1946 = 0.20
1-week-after	1743	128	337	136/93 = 1.46	262/1443 = 0.18
4-weeks-after	1136	67	250	25/23 = 1.09	166/792 = 0.21

ment with the pro group after the invasion. The partisanship plot after the invasion (Fig. 1D) confirms that the invasion bridged the communication divide between politicians of the Left Alliance and those of the other parties. By contrast, the sustained bubble structure of the conspiracy-anti group suggests that the invasion did not change its communication dynamics as much.

3.2 Change in tweet content

Our reading of the sampled tweets suggests that the left-anti group shared with the pro group a critical attitude toward Russia's invasion of Ukraine, which potentially connected them in the retweet network. After the invasion, many people in the left-anti group also moved away from explicitly voicing anti-NATO stances to asking for more discussion on NATO, in addition to arguing that NATO opponents should not be ostracized. Although this might imply that they did not shift their opinion completely toward the other end, the change in their expression opened up a possibility for their interaction with the pro group, as some NATO supporters also argued that an open discussion involving both sides should be acceptable. Thus, the left-anti group and the pro group were also connected by a shared understanding of the discussion norm that embraces diverse opinions and open debate.

Meanwhile, members of the conspiracy-anti group consistently built explicitly anti-NATO arguments upon conspiracy theories and disinformation. Many were also repeating messages of the Russian state propaganda [46], and some of them, as well-known figures in the Finnish disinformation and conspiracy theory scene, have been interviewed on Russian state television as supposed experts. Thus, this conspiracy-charged and pro-Russia group presumably did not find much common ground with the pro group, and was not changed much by the invasion.

The stance distribution of the sampled tweets confirms that the conspiracy-anti group held a consistently strong anti-NATO attitude even after the invasion (Fig. 1H). Meanwhile, the left-anti group saw a notable decrease in the expression of anti-NATO attitude after the invasion (Fig. 1G), yet it also did not turn clearly pro-NATO. While in general tweets labeled "unclear" do not necessarily reflect an attitude toward NATO (see Appendix for a discussion), many users in the left-anti group posted a specific type of "unclear" stance tweets that discussed the pros and (especially) cons of joining NATO, or asked for more discussions on that. This potentially reflects some extent of self-censorship in this group: some users who retained an anti-NATO leaning might have avoided stating anti-NATO stances explicitly after becoming a minority in the discussion.

Our user activity analysis reveals another possible form of self-censorship in the left-anti group. Before the invasion, the two anti-NATO groups had a comparable percentage

of active users in each retweet network (Fig. 1E); yet after the invasion, the percentage was consistently lower in the left-anti group. The partisanship plot (Fig. 1D) also shows that only two of the eight Left Alliance politicians in the *before* network were still present in the *right-after* network. Coupled with the change in retweeting dynamics, the decreased user activity in the left-anti group hints at a spiral of silence [47] among a part of the left-anti group members, who may have chosen not to share their opinions in response to the shifted discussion climate.

4 Discussion

Our analyses provide an overview of how the Russian invasion of Ukraine selectively depolarized the Finnish NATO discussion on Twitter: the left-wing anti-NATO group members broke out of their retweeting bubble and connected with the traditionally right-wing pro-NATO group based on established common ground, but the conspiracy-charged anti-NATO group mostly remained a densely connected cluster of its own and persisted in holding an anti-NATO attitude.

Our results demonstrate how a dramatic external threat can change the discussion dynamics between partisan actors. While previous empirical research has found that terrorist and other direct attacks can lead to a rally-around-the-flag phenomenon where the political elite presents a united front [32], there is less evidence that an indirect external threat posed by an adversarial state would decrease partisan polarization [29]. Our results show that polarization in partisanship-divided issues can be weakened overnight by a dramatic external threat, as actors of opposite leanings build connections on the basis of a shared target of criticism (Russia) and a shared understanding of democratic norms (discussion about, and even opposition to joining NATO are part of democracy).

While the observed depolarization of the endorsement network and the change in expressed stances are conclusive, with observational data it is difficult to gauge the amount of actual opinion change or the level of ideological depolarization at large, especially given the possibility of a spiral of silence. Nevertheless, even when depolarization takes the form of self-censored opposition [48], it can still create opportunities for information exposure and conversation between different bubbles, which can serve as a first step toward actual ideological depolarization [49].

Our study also sheds new light on the critical role that conspiracy theories and disinformation play in shaping and sustaining polarization on social media. Prior to our work, researchers have found that consumers of conspiracy news tend to form echo chambers with homophilic ties in friendship networks [34] and diffusion networks [33]. Bessi et al. [35] showed that conspiracy news consumers are more focused on diffusing within-group content and interacting with within-group actors, which points to the potential stability of conspiracy-based polarization. Zollo et al. [50] further added that users within the conspiracy echo chamber rarely interact with debunking posts, and when they do, their interest in conspiracy content actually increases after the interaction. Echoing these findings, our results more concretely demonstrate the resilience of conspiracy-/disinformation-based polarization. We show that consumers of conspiracy theories and disinformation formed a separate retweeting bubble, and further, that they were reluctant to change their opinions or communication patterns even in the face of a dramatic external threat and otherwise bridged partisan divides. This alerts us to the fact that conspiracy theories and

disinformation are consumed by polarized actors that are even more entrenched than partisan actors, and that it can be extremely difficult to pave a way toward conversation and consensus with them.

Appendix

A.1 Data collection keywords

Two of the authors who are experts on Finnish politics developed a list of keywords related to the Finnish NATO discussion. We here present the original keywords in Finnish along with their rough translations into English (many of the words are context specific): liittoutua (to ally), liittoutumaton (non-aligned), liittoutumattomana (as non-aligned), liittoutumattomuuden (non-alignment), liittoutumattomuus (non-alignment), liittoutuminen (allying), liittoutumisen (allying), nato (NATO), nato-kumppani (NATO partner), nato-kumppanien (NATO partners'), nato-kumppanit (NATO partners), nato-kumppanuus (NATO partnership), nato-yhteistyö (NATO cooperation), nato-yhteistyön (of NATO cooperation), nato-yhteistyössä (in NATO cooperation), nato-yhteistyötä (NATO cooperation), naton (of NATO), natoon (into NATO), natossa (in NATO), natosta (from NATO), puolustusliiton (defense alliance's), puolustusliitosta (from the defense alliance), puolustusliitto (defense alliance), puolustusliittoon (into the defense alliance), sotilasliiton (military alliance's), sotilasliitosta (from the military alliance), sotilasliitto (military alliance), sotilasliittoon (into the military alliance), suominatoon (Finland into NATO), natojäsenyyttä (NATO membership), natojäsenyyden (NATO membership's), nato-trolli (NATO troll), nato-trollit (NATO trolls), nato-trollien (NATO trolls'), nato-trollaajat (NATO troll users), nato-trollaajien (NATO troll users'), nato-kiima (NATO heat), nato-kiiman (NATO heat's), nato-kiimailijat (NATO enthusiasts), nato-kiimailijoiden (NATO enthusiasts'), natoteatteri (NATO theater), natoteatteria (NATO theater), and natoteatterista (from the NATO theater).

A.2 Tweet sampling statistics

For each group and each period, we sampled 42 tweets from those that got retweeted at least once in the group in the period. In total, 1800/221/416 tweets in the *before* period, 4188/343/1118 tweets in the *right-after* period, 2698/257/779 tweets in the *1-week-after* period, and 1022/88/481 tweets in the *4-weeks-after* period got retweeted at least once in respectively the pro, left-anti, and conspiracy-anti group.

A.3 Extra retweet network plots

Retweet networks in the *1-week-after* and *4-weeks-after* periods are plotted in Fig. 2.

A.4 Statistical test of network structure change

Due to the varying user group size in the retweet networks across different time periods, the observed change in E/I ratio can potentially be explained by statistical fluctuations. Here, we conduct a statistical test to see if the observed E/I ratio change after the invasion is higher in the left-anti group than in the conspiracy-anti group despite statistical fluctuations.

We suppose the retweets by each anti group are generated by a hypothetical model where each retweet is an external retweet of the pro group with probability p_E, or an

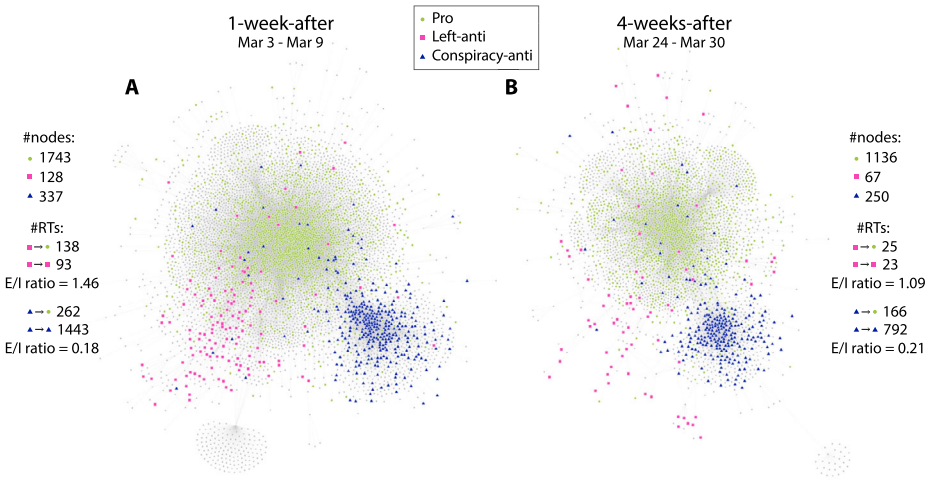

Figure 2 *Retweet networks 1 week after and 4 weeks after the invasion.* Retweet network (A) 1 week after and (B) 4 weeks after the Russian invasion of Ukraine. Networks are drawn using the SFDP spring-block layout [43]. Node colors correspond to the three groups detected in the *before* network, and the statistics beside each network show the number of users, the number of external retweets of the pro group, the number of internal retweets, and the E/I ratio in each anti group

internal retweet with probability $1 - p_E$. We assume the uniform Beta$(1, 1)$ prior on p_E, which leads to the posterior distribution of $p_E \sim \text{Beta}(1 + n_E, 1 + n_I)$, where n_E is the observed number of external retweets, and n_I is the observed number of internal retweets in a certain period. For respectively the *before* period and the *right-after* period, we calculate the posterior distribution of p_E in respectively the left-anti group and the conspiracy-anti group. For example in the left-anti group, $p_E \sim \text{Beta}(1 + 41, 1 + 468)$ in the *before* period, and $p_E \sim \text{Beta}(1 + 166, 1 + 193)$ in the *right-after* period; in the conspiracy-anti group, $p_E \sim \text{Beta}(1 + 96, 1 + 1216)$ in the *before* period, and $p_E \sim \text{Beta}(1 + 389, 1 + 1946)$ in the *right-after* period.

We run 100,000 rounds of simulations. In each round, for respectively the *before* period and the *right-after* period, we sample $\hat{p}_E^L$ from the posterior distribution of p_E in the left-anti group, and $\hat{p}_E^C$ from the posterior distribution of p_E in the conspiracy-anti group. We then numerically calculate the expected E/I ratio in the left-anti group R^L (resp. in the conspiracy-anti group R^C) based on the sampled $\hat{p}_E^L$ (resp. $\hat{p}_E^C$). Then we obtain the E/I ratio change induced by the invasion in the left-anti group $Q_R^L = R_{\text{after}}^L / R_{\text{before}}^L$ (resp. in the conspiracy-anti group $Q_R^C = R_{\text{after}}^C / R_{\text{before}}^C$). Finally, we obtain distributions of Q_R^L and Q_R^C over 100,000 simulations. We conduct a similar analysis also for the E/I ratio change in the *1-week-after* period (as compared with the *before* period) and in the *4-weeks-after* period (as compared with the *before* period).

As shown in Fig. 3 and Table 2, there is a certain range of variance in E/I ratio change that can be explained by statistical fluctuations, and the variance increases in later periods as the group size decreases. However, despite statistical fluctuations, the E/I ratio change induced by the invasion is still consistently higher in the left-anti group than in the conspiracy-anti group.

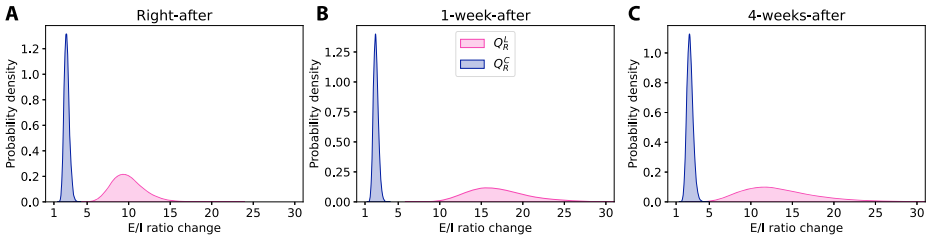

Figure 3 *Statistical test results: distributions.* Distribution of simulated E/I ratio change in respectively the left-anti group (Q_R^L) and the conspiracy-anti group (Q_R^C) in the (A) *right-after*, (B) *1-week-after*, and (C) *4-weeks-after* periods, over 100,000 simulations

Table 2 Statistical test results: distribution statistics. Mean and standard deviation of distributions of simulated E/I ratio changes induced by the invasion in respectively the left-anti group (Q_R^L) and the conspiracy-anti group (Q_R^C), over 100,000 simulations

	Q_R^L	Q_R^C
Right-after	9.93 ± 1.96	2.54 ± 0.31
1-week-after	16.99 ± 3.67	2.31 ± 0.29
4-weeks-after	13.56 ± 4.72	2.68 ± 0.37

A.5 Tweets with unclear stance

In our tweet stance coding, a tweet is labeled "unclear" if it does not explicitly express a positive or negative attitude toward NATO. Thus, in general, the label "unclear" does not necessarily imply an ambiguous attitude toward NATO, but rather that the tweet does not clearly indicate any attitude. For example, tweets labeled "unclear" can be reactions to what was currently taking place in the Ukraine war (while NATO was also mentioned) or in the Finnish NATO policy process.

More specifically in the pro-NATO group, many tweets were labeled "pro" in the earlier periods because they were advocating for two citizen initiatives that were pro-NATO; but later on, these initiatives became irrelevant because the needed signatures were collected, and the NATO policy process moved on. Thus in later periods, many clearly pro-NATO tweets disappeared from the pro-NATO group and, for example, many tweets condemning Russia's actions in Ukraine took their place. The latter are often labeled "unclear" as they are less clearly in favor of NATO, even though such a stance might be implicit. In general, the increase of tweets with unclear stance does not suggest that the group moved toward an ambiguous stance on NATO.

Acknowledgements
We want to thank Ted Hsuan Yun Chen, Risto Kunelius, and the anonymous reviewers for giving extremely insightful feedback on our study.

Funding
This work was supported by the Academy of Finland (320780, 320781, 332916, 349366, 352561), the Kone Foundation (201804137), and the Helsingin Sanomat Foundation (20210021).

Abbreviations
NATO, North Atlantic Treaty Organization; API, Application Programming Interface; E/I ratio, external-internal ratio.

Data availability
The code, tweet IDs, and anonymized retweet networks for generating the results described in the paper are available at https://github.com/ECANET-research/finnish-nato.

Declarations

Competing interests
The authors declare that they have no competing interests.

Author contributions
YX, AG, AM, TY, BK, and MK designed research, performed research, and analyzed data; YX, AG, AM, TY, and MK wrote the paper. All authors read and approved the final manuscript.

Author details
[1]Department of Computer Science, Aalto University, Espoo, Finland. [2]Faculty of Social Sciences, University of Helsinki, Helsinki, Finland.

Received: 23 August 2023 Accepted: 13 December 2023 Published online: 03 January 2024

References
1. Haavisto I (2022) At nato's door. EVA Analysis (104)
2. Yle (2022) Yle poll: support for NATO membership soars to 76%. https://yle.fi/a/3-12437506. Accessed 12 April 2023
3. Forsberg T (2018) Finland and NATO: strategic choices and identity conceptions. In: The European neutrals and NATO: non-alignment, partnership, membership? pp 97–127
4. Barberá P, Steinert-Threlkeld ZC (2020) How to use social media data for political science research. In: The SAGE handbook of research methods in political science and international relations, vol 2, pp 404–423
5. Ruoho I, Kuusipalo J (2019) The inner circle of power on Twitter? How politicians and journalists form a virtual network elite in Finland. Observatorio 13(1):70–85. https://doi.org/10.15847/obsobs13120191326
6. Bail C (2021) Breaking the social media prism: how to make our platforms less polarizing. Princeton University Press, Princeton
7. Matsubayashi T (2013) Do politicians shape public opinion? Br J Polit Sci 43(2):451–478
8. Barberá P, Zeitzoff T (2018) The new public address system: why do world leaders adopt social media? Int Stud Q 62(1):121–130
9. Suresh VP, Nogara G, Cardoso F, Cresci S, Giordano S, Luceri L (2023) Tracking fringe and coordinated activity on Twitter leading up to the US Capitol attack. arXiv preprint. arXiv:2302.04450
10. Sharma K, Ferrara E, Liu Y (2022) Characterizing online engagement with disinformation and conspiracies in the 2020 US presidential election. In: Proceedings of the international AAAI conference on web and social media, vol 16, pp 908–919
11. Erokhin D, Yosipof A, Komendantova N (2022) Covid-19 conspiracy theories discussion on Twitter. Soc Media Soc 8(4):20563051221126051
12. Conover M, Ratkiewicz J, Francisco M, Gonçalves B, Menczer F, Flammini A (2011) Political polarization on Twitter. In: Proceedings of the international AAAI conference on web and social media, vol 5, pp 89–96
13. Barberá P, Jost JT, Nagler J, Tucker JA, Bonneau R (2015) Tweeting from left to right: is online political communication more than an echo chamber? Psychol Sci 26(10):1531–1542
14. Garimella K, Morales GDF, Gionis A, Mathioudakis M (2018) Quantifying controversy on social media. ACM Trans Soc Comput 1(1):1–27
15. Cossard A, Morales GDF, Kalimeri K, Mejova Y, Paolotti D, Starnini M (2020) Falling into the echo chamber: the Italian vaccination debate on Twitter. In: Proceedings of the international AAAI conference on web and social media, vol 14, pp 130–140
16. Morales AJ, Borondo J, Losada JC, Benito RM (2015) Measuring political polarization: Twitter shows the two sides of Venezuela. Chaos, Interdiscip J Nonlinear Sci 25(3):033114
17. Esteve Del Valle M, Broersma M, Ponsioen A (2022) Political interaction beyond party lines: communication ties and party polarization in parliamentary Twitter networks. Soc Sci Comput Rev 40(3):736–755
18. Chen THY, Salloum A, Gronow A, Ylä-Anttila T, Kivelä M (2021) Polarization of climate politics results from partisan sorting: evidence from Finnish twittersphere. Glob Environ Change 71:102348
19. Weber I, Garimella VRK, Batayneh A (2013) Secular vs. Islamist polarization in Egypt on Twitter. In: Proceedings of the 2013 IEEE/ACM international conference on advances in social networks analysis and mining, pp 290–297
20. Borge-Holthoefer J, Magdy W, Darwish K, Weber I (2015) Content and network dynamics behind Egyptian political polarization on Twitter. In: Proceedings of the 18th ACM conference on computer supported cooperative work & social computing, pp 700–711
21. Garcia MB, Cunanan-Yabut A (2022) Public sentiment and emotion analyses of Twitter data on the 2022 Russian invasion of Ukraine. In: 2022 9th international conference on information technology, computer, and electrical engineering (ICITACEE). IEEE, New York, pp 242–247
22. Mir AA, Rathinam S, Gul S, Bhat SA (2023) Exploring the perceived opinion of social media users about the Ukraine–Russia conflict through the naturalistic observation of tweets. Soc Netw Anal Min 13(1):44
23. Evkoski B, Kralj Novak P, Ljubešić N (2023) Content-based comparison of communities in social networks: ex-Yugoslavian reactions to the Russian invasion of Ukraine. Appl Netw Sci 8(1):1–24
24. Caprolu M, Sadighian A, Di Pietro R (2023) Characterizing the 2022-Russo-Ukrainian conflict through the lenses of aspect-based sentiment analysis: dataset, methodology, and key findings. In: 2023 32nd international conference on computer communications and networks (ICCCN). IEEE, New York, pp 1–10
25. Ibar-Alonso R, Quiroga-García R, Arenas-Parra M (2022) Opinion mining of green energy sentiment: a Russia-Ukraine conflict analysis. Mathematics 10(14):2532
26. Geissler D, Bär D, Pröllochs N, Feuerriegel S (2023) Russian propaganda on social media during the 2022 invasion of Ukraine. EPJ Data Sci 12(1):35
27. Pierri F, Luceri L, Jindal N, Ferrara E (2023) Propaganda and misinformation on Facebook and Twitter during the Russian invasion of Ukraine. In: Proceedings of the 15th ACM web science conference 2023, pp 65–74

28. Nisch S (2023) Public opinion about Finland joining nato: analysing Twitter posts by performing natural language processing. J Contemp Eur Stud. https://doi.org/10.1080/14782804.2023.2235565
29. Myrick R (2021) Do external threats unite or divide? Security crises, rivalries, and polarization in American foreign policy. Int Organ 75(4):921–958
30. Bessi A, Ferrara E (2016) Social bots distort the 2016 US presidential election online discussion. First Monday 21(11-7)
31. Coser LA (1956) The functions of social conflict, vol 9. Routledge, Abingdon
32. Chowanietz C (2011) Rallying around the flag or railing against the government? Political parties' reactions to terrorist acts. Party Polit 17(5):673–698
33. Del Vicario M, Bessi A, Zollo F, Petroni F, Scala A, Caldarelli G, Stanley HE, Quattrociocchi W (2016) The spreading of misinformation online. Proc Natl Acad Sci 113(3):554–559
34. Bessi A, Petroni F, Vicario MD, Zollo F, Anagnostopoulos A, Scala A, Caldarelli G, Quattrociocchi W (2016) Homophily and polarization in the age of misinformation. Eur Phys J Spec Top 225:2047–2059
35. Bessi A, Coletto M, Davidescu GA, Scala A, Caldarelli G, Quattrociocchi W (2015) Science vs conspiracy: collective narratives in the age of misinformation. PLoS ONE 10(2):0118093
36. Twitter (2022) Twitter API documentation | Docs | Twitter Developer Platform. https://developer.twitter.com/en/docs/twitter-api. Accessed 12 April 2023
37. Metaxas P, Mustafaraj E, Wong K, Zeng L, O'Keefe M, Finn S (2015) What do retweets indicate? Results from user survey and meta-review of research. In: Proceedings of the international AAAI conference on web and social media, vol 9, pp 658–661
38. Traag VA, Waltman L, Van Eck NJ (2019) From Louvain to Leiden: guaranteeing well-connected communities. Sci Rep 9(1):1–12
39. Peixoto TP (2018) Nonparametric weighted stochastic block models. Phys Rev E 97(1):012306
40. Hayes AF, Krippendorff K (2007) Answering the call for a standard reliability measure for coding data. Commun Methods Meas 1(1):77–89
41. Xia Y, Gronow A, Kukkonen A, Chen THY, Kivelä M et al (2021) Botit ja informaatiovaikuttaminen twitterissä vuoden 2021 kuntavaaleissa-elebot-2021-hanke
42. Salloum A, Takko T, Peuhkuri M, Kantola R, Kivelä M et al (2019) Botit ja informaatiovaikuttaminen twitterissä suomen eduskunta-ja eu-vaaleissa 2019-elebot-hanke
43. Hu Y (2005) Efficient, high-quality force-directed graph drawing. Math J 10(1):37–71
44. Seuri V (2019) Nyt tutkimaan: Ylen eduskuntavaalien vaalikoneen aineisto julkaistu avoimena datana [Let's analyse: YLE election assistant tool data released as open access]. https://yle.fi/a/3-10725384. Accessed 26 October 2023
45. Salloum A, Chen THY, Kivelä M (2022) Separating polarization from noise: comparison and normalization of structural polarization measures. In: Proceedings of the ACM on human-computer interaction 6 (CSCW1), pp 1–33
46. Hanley HW, Kumar D, Durumeric Z (2023) Happenstance: utilizing semantic search to track Russian state media narratives about the Russo-Ukrainian war on Reddit. In: Proceedings of the international AAAI conference on web and social media, vol 17, pp 327–338
47. Noelle-Neumann E (1974) The spiral of silence a theory of public opinion. J Commun 24(2):43–51
48. Brody RA, Shapiro CR (1989) Policy failure and public support: the Iran-contra affair and public assessment of president Reagan. Polit Behav 11(4):353–369
49. Mutz DC (2002) Cross-cutting social networks: testing democratic theory in practice. Am Polit Sci Rev 96(1):111–126
50. Zollo F, Bessi A, Del Vicario M, Scala A, Caldarelli G, Shekhtman L, Havlin S, Quattrociocchi W (2017) Debunking in a world of tribes. PLoS ONE 12(7):0181821

Publisher's Note