nature human behaviour

Article

https://doi.org/10.1038/s41562-026-02435-2

How deceptive online networks reached millions in the US 2020 elections

Received: 20 December 2024

Accepted: 27 February 2026

Published online: 06 April 2026

Check for updates

Ruth E. Appel[1], Young Mie Kim[2], Jennifer Pan[1], Yiqing Xu[3], Ben Nimmo[4], Daniel Robert Thomas[4], Hunt Allcott[5], Pablo Barberá[4], Taylor Brown[4], Adriana Crespo-Tenorio[4], Drew Dimmery[4,6], Deen Freelon[7], Matthew Gentzkow[8], Sandra González-Bailón[7], Andrew M. Guess[9], Shanto Iyengar[3], David Lazer[10,11], Neil Malhotra[12], Devra Moehler[4], Brendan Nyhan[13], Jaime Settle[14], Emily Thorson[15], Rebekah Tromble[16], Carlos Velasco Rivera[4], Arjun Wilkins[4], Magdalena Wojcieszak[17,18], Beixian Xiong[4], Chad Kiewiet de Jonge[4], Annie Franco[4], Winter Mason[4], Natalie Jomini Stroud[19] & Joshua A. Tucker[20]

Deceptive online networks are coordinated efforts that use identity deception to pursue strategic political or financial goals. During the US 2020 elections, these networks reached at least 37 million Facebook and 3 million Instagram users, representing 15% and 2% of the platforms' active US adult users, respectively. Only 3 networks out of 49—1 network with explicitly political aims and 2 that appeared to use politics as a lure for profit—were responsible for over 70% of users reached. Notably, accounts unaffiliated with the networks played an important role in facilitating this reach by resharing content the three networks produced. Deceptive networks, regardless of whether their goals were political or financial, reached users who were older, more conservative, more frequently exposed to content from untrustworthy sources, and spent more time on Facebook.

In the summer of 2020, Rally Forge, a US marketing firm acting on behalf of the nonprofit Turning Point USA, operated a network of Facebook and Instagram accounts with fake profiles that impersonated US citizens from a wide range of demographic and ideological backgrounds, making thousands of posts pretending to express authentic opinions[1,2] (see the report on CIB9 linked in Extended Data Table 1). Posts from Rally Forge were reshared millions of times by Facebook users who were unaffiliated with this deceptive online network.

From June to October 2020, a group from Kosovo used a compromised comedy Page and six fake Facebook user accounts to post memes and links to off-platform websites that copied content from Fox News. These posts were reshared by Facebook users unaffiliated with the Kosovo group hundreds of thousands of times (see the description of FMO27 in Extended Data Table 1).

Posting political opinions and sharing links are activities that people do every day on social media. But when these actions are taken by deceptive online networks—which use various types of identity deception such as using fake names and photos, deploying seemingly unrelated and independent shell accounts, and artificially boosting engagement—they violate democratic norms and can erode trust in democratic deliberation by contaminating public discourse and misleading audiences, thus potentially distorting collective preferences.

A large and growing body of research on influence operations and disinformation campaigns has focused primarily on networks pursuing political goals[3–6], particularly those originating from Russia active around the 2016 US elections[7–14]. Furthermore, these studies have focused on users' engagement with deceptive online networks rather than exposure. This limitation arises because data on when users are exposed to network content is available only to platforms, and not to external researchers. Except for one study that captured targeted ad exposure[14], previous research has inferred exposure through proxies, such as users following or mentioning network accounts, without directly measuring whether individuals viewed the content[4,10]. For these

A full list of affiliations appears at the end of the paper. e-mail: jp1@stanford.edu

two reasons, previous studies are fundamentally limited in their ability to explain whether and how deceptive networks reach audiences.

This pre-registered study uses platform-level aggregated data about deceptive online networks identified by Meta that targeted US-based users in the context of the 2020 US elections to analyse their characteristics, activities and reach. The data cover deceptive online networks that engage in political discourse around the elections, whether they pursue political or financial goals. The data that we analyse encompass 49 networks, including details on how many US adult Facebook and Instagram users saw content from deceptive online networks and how network content reached these users.

We find that deceptive networks are diverse in their origins, the number and type of accounts that they operate, and their level of posting activity. The main commonality of the networks that we study is the preponderance of political content that they produce, regardless of whether their primary motive is political or financial. These networks reached at least 37 million unique Facebook and 3 million unique Instagram users, representing 15% and 2% of Facebook and Instagram US adult active users, respectively. However, only a few networks accounted for most of the reach—over 70% of the users who viewed any content from deceptive online networks were exposed through 3 networks highly active on Facebook. In addition, we find that the way these three networks reached users on Facebook was through the participation of accounts that were unaffiliated with the networks resharing deceptive network content. Regardless of the network's primary motive or how content reached users (whether directly through network accounts or indirectly through accounts unaffiliated with networks), those exposed to deceptive network content were relatively older, more conservative, more frequently exposed to content from untrustworthy sources, and spent more time on Facebook.

Deceptive online networks

We introduce the concept of deceptive online networks and define it as coordinated efforts that take place online where audiences are misled about the identity of the people behind the network. Coordination means that members of a group organize among themselves to conduct a set of activities (for example, some members of the group administer multiple accounts or post content at the same time). Deception refers to the concealment of the coordination and/or misrepresentation of the identities of those involved in the effort—that is, who they really are or what they want to achieve. For example, in the Rally Forge case, workers coordinated on the content and timing of the content that they produced and engaged in identity deception when they operated multiple accounts with seemingly different identities, disguising their role behind these accounts and the payment they received for posting. This identity deception is not based on content; a network can engage in identity deception without posting deceptive content such as false information.

Our study focuses on deceptive online networks that engage in political discourse, regardless of their ultimate aims. Previous research has focused predominantly on deceptive online networks where the immediate and/or end goal of the effort is political influence. These efforts have been referred to as influence operations, influence campaigns, disinformation campaigns and propaganda campaigns[3,6,7,15,16]. Much less attention has been paid to deceptive online networks using political discourse whose only goal is financial gain. Networks pursuing financial gain talk about politics to monetize online activity, for example, by enticing users to click on ads, make payments or inadvertently share their private financial information[17–19]. Very little research has been conducted on these networks because they are often dismissed as mere spam and clickbait, deemed unworthy of public attention and data disclosure. What we know about them comes from journalistic investigations and one qualitative study of groups in North Macedonia that became well known during the 2016 US elections for taking advantage of the elections as a hot-button issue to garner public attention and ultimately generate profits from users' clicks[20–22]. To understand how deceptive online networks shape public discourse, there is a need to expand research and data disclosure beyond networks focused solely on political influence. Our concept of deceptive online networks addresses this need. Regardless of their motive, the networks that we study engage in identity deception and interact with users on politics and social issues.

This research is part of the US 2020 Facebook and Instagram Election Study (FIES), an effort between Meta and external researchers to study the political impact of Facebook and Instagram on US adults in the context of the 2020 elections (Supplementary Information section 1.1). The deceptive online networks examined in this study include two subsets of networks removed by Meta for inauthentic behaviour (IB), which covers efforts "to mislead people or Facebook about the popularity of content, the purpose of a community (that is, groups, Pages, Events), or the identity of the people behind it"[23]. The first subset are coordinated inauthentic behaviour (CIB) networks Meta removed for engaging in "coordinated efforts to manipulate public debate for a strategic goal where fake accounts are central to the operation"[24], which are "particularly sophisticated forms of Inauthentic Behavior"[23]. The second subset are financially motivated operations (FMO) networks Meta removed for IB that were flagged by an internal, informal Meta taskforce for producing some content related to politics in the United States leading up to the 2020 US elections. This operationalization captures the key elements of deceptive online networks: coordination, misrepresentation of identity and engagement in political discourse, regardless of a network's goal. We show CIB and FMO networks separately throughout the paper to highlight their similarities and differences, allowing for an empirical assessment of the importance of studying deceptive online networks as an overarching category. For more details on network identification, inclusion criteria and descriptions (including names of CIB networks in common usage), see Extended Data and Supplementary Information section 2.1.

We focus on networks operating on Facebook and Instagram that targeted the United States between 26 June 2020 and 15 February 2021 (the study period). Networks identified by Meta between 26 June 2020 and 3 November 2021 are included to cover networks operating during the elections but discovered afterwards. This may not capture all deceptive online networks engaged in political discourse on Facebook and Instagram. Some networks might never have been caught by Meta. Some networks and/or information related to the networks might not have been available to Meta at the time of data analysis because the data were not preserved by the company (see Supplementary Information section 2.1.3 for details on data availability). Takedowns under Meta's CIB subcategory are publicly disclosed, whereas other forms of IB are often not. All attributions of network characteristics and activity are provided by Meta, including the designation of accounts as either affiliated with a network (that is, part of it) or merely connected to network accounts (for example, through friend or followership ties) without being affiliated with the network. Any Facebook or Instagram account not identified by Meta as affiliated with a deceptive online network is referred to as a non-network account. Some non-network accounts are connected to network accounts, while others are not. These data encompass the broadest set of networks engaged in political discourse and their characteristics made available for academic research.

The dataset includes the characteristics, activity and reach of these deceptive networks on Facebook and Instagram (to access, see 'Data availability' section). The data on network reach that we analyse are based on active users, defined as unique adult US-based Facebook or Instagram users who were active for at least 1 day during the study period and excluding user accounts affiliated with networks. Network content can differ in how it reaches users—whether users see a post directly from the network or users see a post indirectly when a non-network account reshares a post from the network. We use the terms 'direct network content' and 'direct network post' to refer

https://doi.org/10.1038/s41562-026-02435-2

to original posts of network-affiliated accounts and reshares made by network-affiliated accounts. We use the terms 'indirect network content' and 'indirect network posts' to refer to reshares of posts of network-affiliated accounts by non-network accounts. Whenever the terms 'network content' and 'network post' are used in this paper, this indicates the inclusion of direct and indirect content (see Supplementary Information section 2.0.1 for all definitions).

Deceptive networks are diverse in origin and predominantly feature political content

Meta identified 13 CIB networks and 36 FMO networks from around the world targeting US users on Facebook and Instagram during the 2020 elections (Fig. 1a). CIB networks originated from Russia (5 networks), the United States (2 networks), Iran (2 networks), China and Romania, and included 2 networks whose organization spanned multiple countries (for links to detailed reports on CIB networks, see Extended Data Table 1). One transnational network, Truthmedia (CIB3—references to networks are standardized by network type (CIB or FMO) and a number, ordered from oldest to newest date of removal), was located in a dozen countries across Asia, Europe, Oceania and North America (Fig. 1a). As shown by the green areas in Fig. 1a, FMO networks were concentrated in the Balkans (16 networks from Albania, North Macedonia, Bosnia and Herzegovina, Kosovo and Romania) and South Asia (11 networks from Pakistan, Bangladesh and India), but also included networks from Canada, the United States, Italy, Russia, Algeria, Morocco and the Philippines (see Extended Data Table 1 for descriptions of FMO networks; see Extended Data Table 2 for more details on network characteristics).

Networks varied greatly in the type and number of accounts that they operated and the number of direct network posts that they made (Fig. 1b). On Facebook, all 49 CIB and FMO networks had activity but varied in the number of accounts (from 2 to 676) and number of direct network posts (from none in the study period to 15,273). On Instagram, 11 out of 13 CIB networks and 13 out of 36 FMO networks had accounts and the volume of direct network posts varied from none in the study period to 6,646 (see Extended Data Table 2 and Extended Data Fig. 1 for post activity over time; see Supplementary Information section 4.1 for more details). In total, networks operated 4,101 Facebook accounts, creating 81,729 direct network posts, and 613 Instagram accounts, creating 16,854 direct network posts. Networks varied in the number of ads that they purchased, with only 12 out of 49 networks running any ads. Those 12 networks posted 1 ad per 29 direct network posts on average, but varied from 1 ad per 0.6 direct network posts (FMO28) to 1 ad per 1,042 direct network posts (CIB6). Networks paid to make posts more visible only 0.6% of the time.

Politics and social issues are the most frequent topics for both CIB and FMO networks on Facebook (Fig. 1c based on the Meta Topic classifier for Facebook; Supplementary Information section 2.0.2). On Facebook, politics and social issues accounted for 65% of direct network content for CIB networks, and 32% of direct network content for FMO networks (based on Civic classifier; Supplementary Information section 2.0.2). On Instagram, politics and social issues are among the most frequent topics, making up 18% of direct CIB network content and 24% of direct FMO network content (based on Civic classifier; Supplementary Information section 2.0.2). The prevalence of political content is unsurprising given our aim to study deceptive online networks engaged in political discourse. However, it is notable that politics and social issues were the most frequently discussed topics not only for CIB networks but also for FMO networks. While FMO networks had to produce at least some politics content to be included in the study, there was nothing in the selection process that predetermined that political content would be as frequent. Put another way, FMO networks actively produced content related to politics and social issues targeting users in the United States. More generally, politics and social issues were widely debated, and networks appear to have utilized them, along with other contentious issues of the time such as COVID-19 and news, and niche content such as hunting, in their posts (see Supplementary Information section 4.2 and CIB reports linked in Extended Data Table 1 for additional details on network content).

A few networks dominate reach

Network content reached at least 36.79 million unique US adult active users on Facebook (14.63% of approximately 250 million unique US adult active users) and 2.98 million unique US adult active users on Instagram (1.85% of approximately 160 million unique US adult active users). An additional 545,299 unique active users on Facebook and 7,189 unique active users on Instagram were reached solely via network ads, for a total reach of 37.33 million (14.85%) unique active users on Facebook and 2.98 million (1.86%) on Instagram over our 8-month study period. This is roughly equivalent to 15% and 1% of the US voting-age population, respectively (based on US government census data; https://www.census.gov/content/dam/Census/library/publications/2022/demo/p20-585.pdf). For context, in the 2016 US elections, the potential exposure of the Russian Internet Research Agency is estimated to have been 126 million Facebook users over a 2-year period[25]. Distinguishing between exposure to CIB and FMO networks, and excluding users who were reached only via ads, 15.3 million (6.09%) Facebook users saw content from CIB networks and 27.7 million (11.01%) from FMO networks; 2.94 million (1.83%) Instagram users saw content from CIB networks and 34,570 (0.0215%) from FMO networks (the sum of those exposed to CIB and FMO networks is greater than the total number of unique active users exposed because some users saw content from both types of network).

Those reached by the networks were often repeatedly exposed. On Facebook, in total, there were 175.24 million views of network content (excluding ads), 55.87 million for CIB networks and 119.37 million for FMO networks (see Extended Data Table 3 for details). On average, viewers exposed to either CIB or FMO networks saw 4.76 pieces of network content, but the distribution was highly skewed such that the median is lower. On Instagram, in total, there were 69.65 million views of network content (excluding ads), 69.30 million for CIB networks and 356,399 for FMO networks (see Extended Data Table 4 for details). This means that viewers exposed to either or both CIB and FMO networks on Instagram had an average of 23.4 views of network content (see Supplementary Information section 4.3 for more details), but the distribution is also highly skewed with few people accounting for most exposures.

Exposure to deceptive network content accounts for a very small share of users' overall political content consumption. Previous research shows that news and politics make up less than 15% of total digital content consumption[26,27]. During the 2020 election period, less than 10% of content that users saw on Facebook was political[28]. Analysing individual-level participant data, we find that for participants who encountered content from any deceptive online network in the 41 days before the election, such content represented, on average, 0.3% of their political content views on Facebook. The distribution was highly skewed: for users in the 99th percentile of the proportion of network content views to political content views, deceptive network content accounted for 3% of political content views.

On Instagram, 1% of active users accounted for over 96% of views of network content (Fig. 2a). On Facebook, 1% of active users accounted for 55% of views of network content (Fig. 2b). The greater concentration of exposure on Instagram aligns with the fact that viewers exposed to network content on Instagram are exposed to more content on average than viewers exposed to network content on Facebook. For context, previous research on presumed exposure on Twitter to content from the Russian Internet Research Agency in 2016 found that 1% of users accounted for 70% of exposures[10], and research on fake news on Twitter in 2016 found that 1% of users accounted for 80% of such views[29].

Notably, a small number of networks are responsible for reaching almost all viewers. On Instagram, nearly all viewers (95%) were reached

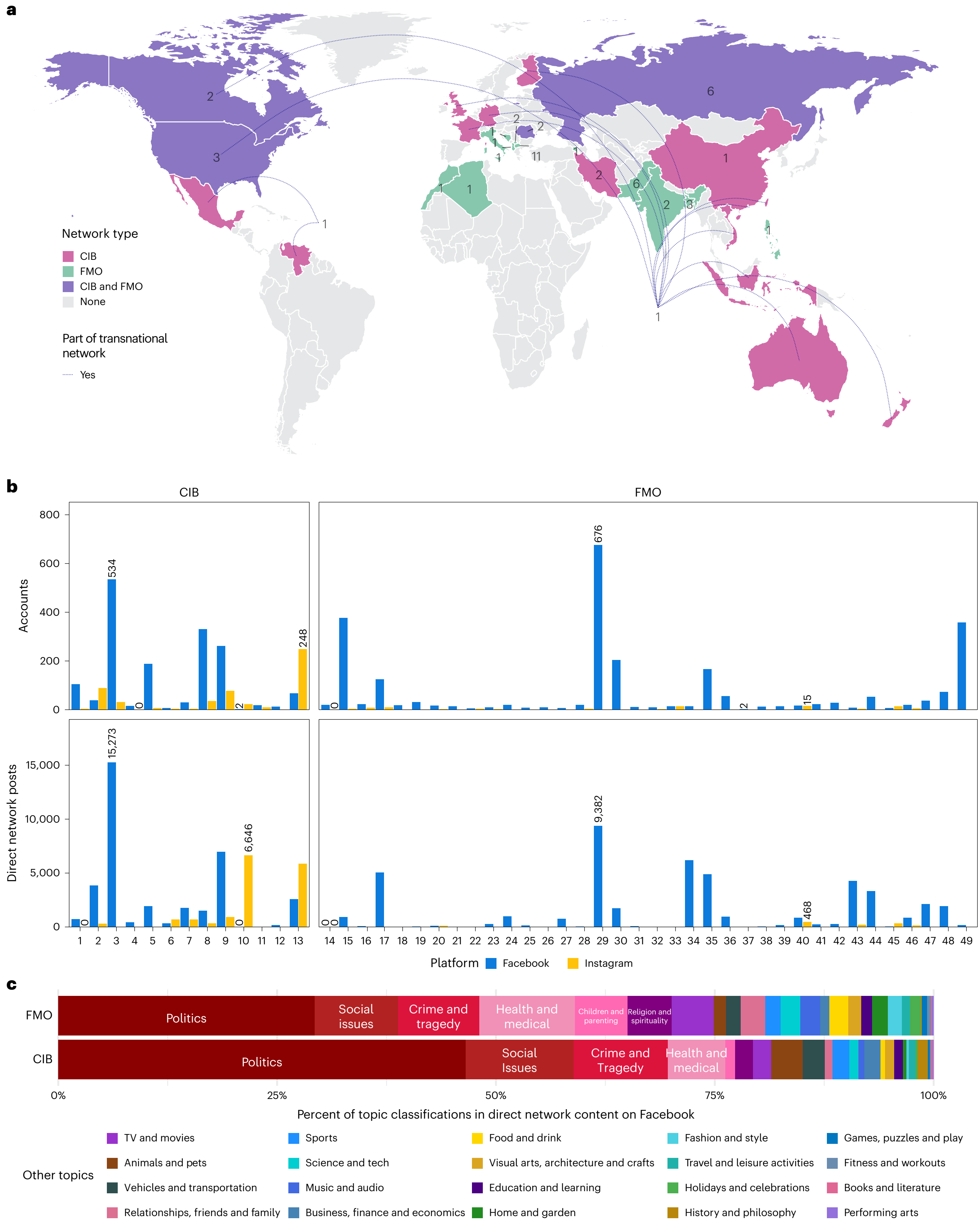


Fig. 1 | Characteristics of deceptive online networks. a, Deceptive online networks' attributed countries of origin. Numbers reflect the number of networks identified by Meta as originating in each country; transnational networks are shown as individual numbers at the approximate geographic midpoint of their attributed countries of origin and are connected to these countries via dotted lines. **b**, Number of accounts and direct network posts for each network on Facebook and Instagram. **c**, Topics for CIB and FMO networks on Facebook based on the Meta Topic Classifier.

 https://doi.org/10.1038/s41562-026-02435-2

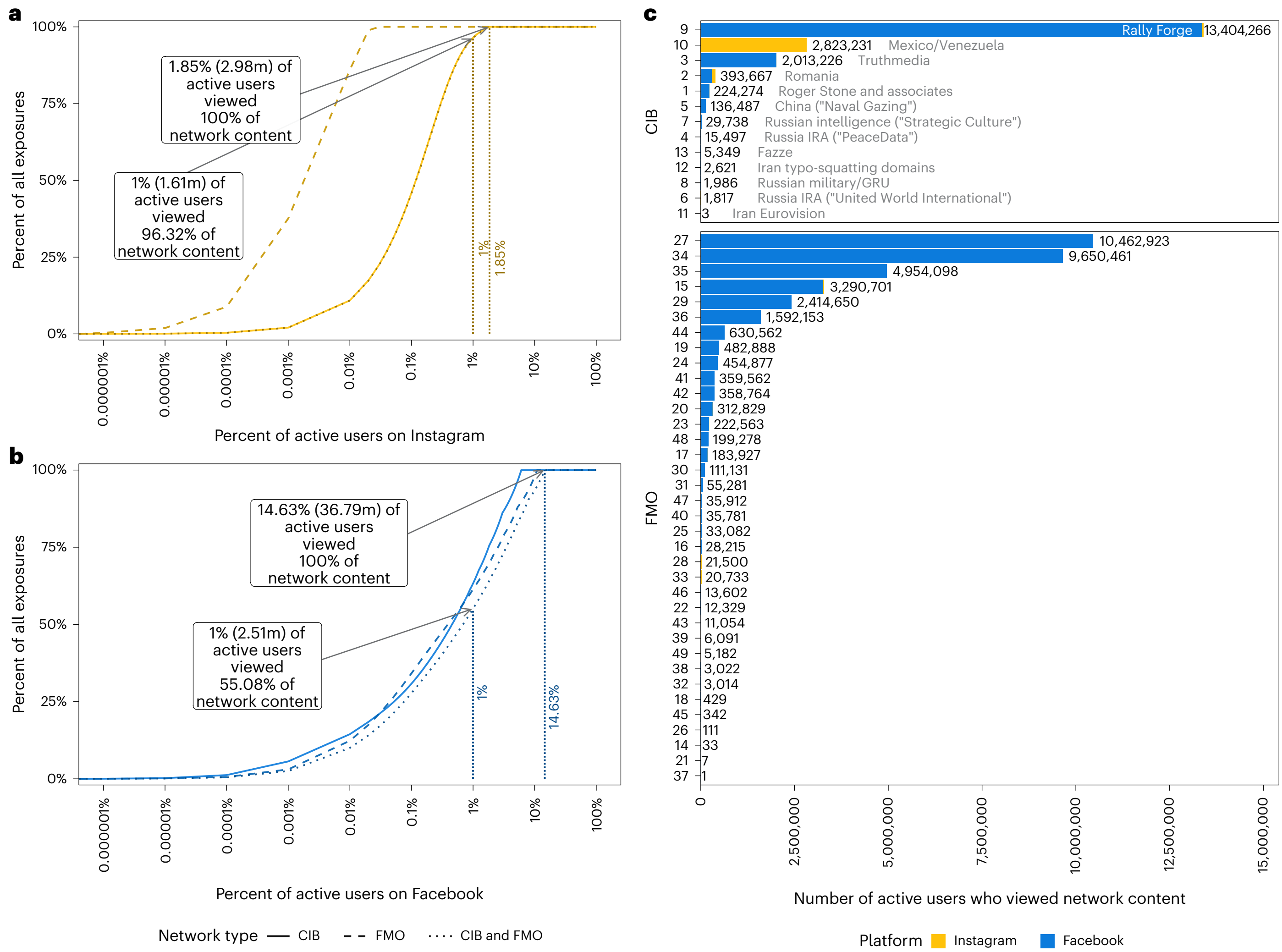


Fig. 2 | Reach of deceptive online networks. a, Cumulative percent of all exposures to network content for active users on Instagram for CIB networks, FMO networks, and their combination. **b**, Cumulative percent of all exposures to network content for active users on Facebook for CIB networks, FMO network, and their combination. **c**, Number of active users who viewed content by network on Facebook and Instagram. **a**,**b**, *x*-axes are log-transformed owing to high concentration. **c**, The sum of the viewers for all networks exceeds the total number of viewers reached on each platform because some viewers were exposed to multiple networks; white or grey text labels next to the viewer counts for CIB networks represent the network names in common usage.

by one network (CIB10; Instagram reach shown in yellow in Fig. 2c). On Facebook, nearly 80% of all unique viewers of deceptive networks were reached by three networks (CIB9, FMO27 and FMO34; Facebook reach shown in blue in Fig. 2c). Because more Facebook users were exposed to network content than Instagram users, the three dominant networks on Facebook account for over 70% of all users who saw deceptive online network content across both platforms. It is also worth noting that many networks have limited reach—8 networks on Facebook and 16 networks on Instagram reached fewer than 1,000 viewers (see Supplementary Table 10 in Supplementary Information section 4.3).

Accounts unaffiliated with networks play a critical role in reach

Focusing on Facebook, where networks reached far more viewers, a clear pattern differentiating the networks with the highest reach from other networks is the participation of non-network accounts, that is, accounts unaffiliated with the deceptive online network (green circles in Fig. 3a). A Facebook user can see content directly posted by network accounts—for example, if the user is friends with a network user account, follows a network Page account, is recommended content from a network-affiliated account or sees an ad from the network (green circles connected to blue circle in Fig. 3a). However, if a Facebook user, Page or Group that is not part of a network reshares an original network post, other Facebook users who see the reshare are exposed to network content indirectly through a non-network account (green circles connected to green circles in Fig. 3a). Figure 3b, where each column is a network and column width corresponds to the total number of viewers exposed to each network, shows that exposure through non-network accounts (green) vastly outpaces exposure through network accounts (blue), especially for networks that reached larger numbers of viewers (for details by network, see Extended Data Table 5). The CIB network with the highest overall reach, CIB9, only reached 1.3 million viewers directly, but through the participation of non-network accounts resharing its content, CIB9 reached 13 million viewers indirectly, for a total of 13.4 million viewers reached (the sum of directly reached and indirectly reached users is larger than the total because some viewers were reached both directly and indirectly). In addition, we find positive correlations between the share of viewers reached through indirect exposure and average cascade size, average cascade depth, average cascade maximum breadth and average structural virality of network original posts (see Supplementary Fig. 27 in Supplementary Information section 4.5). It is important to note that only a small

https://doi.org/10.1038/s41562-026-02435-2

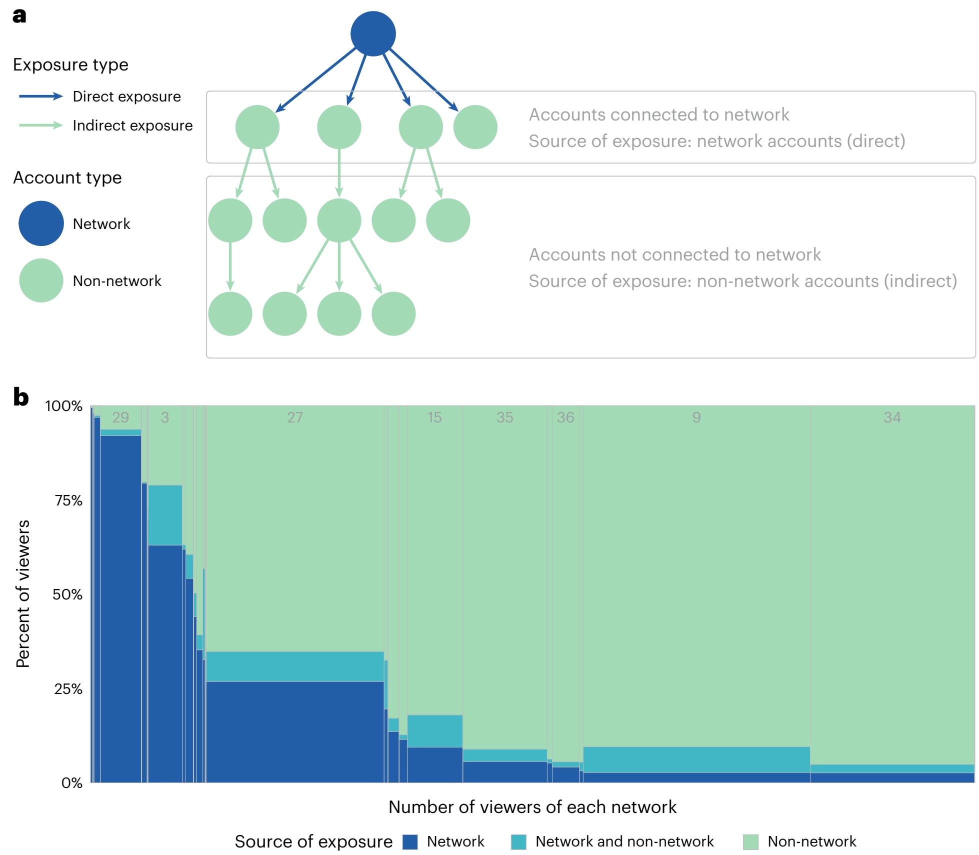


Fig. 3 | How users are exposed to deceptive network content. **a**, Diagram illustrating the distinction between direct exposure and indirect exposure to deceptive network content. Direct exposure is through network accounts and indirect exposure is through non-network accounts. **b**, Source of exposure for each network, ordered by proportion of exposure from non-network accounts. Numbers at the top of the plot refer to the network number, for example, 34 is FMO34, 9 is CIB9.

proportion of users who see network content (5.67% on Facebook and 0.34% on Instagram) reshared network content. This means that even though only a small proportion of those exposed to network content are involved in its propagation, their involvement disproportionately amplifies overall reach. In sum, during the study period, the number of Facebook users reached by network content was massively larger because non-network accounts reshared network content.

It is possible that network exposure could be affected by platform features such as the Facebook feed ranking algorithm. We conducted a post hoc exploratory analysis of the data from FIES experiments[28,30,31] where consenting Facebook users were randomly assigned to (1) a feed without reshares, (2) a reverse chronological feed or (3) a feed where like-minded content was reduced to examine how these treatments affected exposure to and engagement with deceptive online networks compared with the default feed ranking algorithm (control group). We found that the no reshares and chronological feed conditions decreased exposure to and engagement with deceptive networks. However, the treatments also decreased overall user engagement on Facebook, and owing to the small number of experiment participants who were exposed to network content, we were not powered to detect the effect of these treatments after overall views and engagements are taken into account (Supplementary Information section 4.6 figures and tables).

Network content viewers have similar characteristics

When we examine who is reached by deceptive online networks, we find that the characteristics of viewers are similar between CIB and FMO networks. In addition, we find that the characteristics of viewers are similar between viewers who are exposed directly by network accounts and viewers exposed indirectly through non-network accounts. We focus on Facebook here because some characteristics are not available for Instagram.

There are four panels in Fig. 4, each with two plots, that show for each network with more than 100 viewers the characteristics of users exposed to network content directly through network accounts (top plot in each panel) and indirectly through non-network accounts (bottom plot in each panel). The user characteristics are political ideology (panel a), age (panel b), quartiles of time spent on Facebook (panel c) and the share of users in the top 20% of Facebook users in terms of exposure to content from untrustworthy sources (panel d). The horizontal dashed lines in each plot correspond to that characteristic in the population of US adult active users in the study period (for example, the two dashed lines in the first pair of plots separate the share of liberal, moderate and conservative adult active Facebook users).

For both CIB and FMO networks, those exposed are more likely to be conservative—the red portions of Fig. 4a rise higher for the majority of networks than the lower dashed line, which shows that the proportion of conservatives exposed is higher than the proportion of conservatives in the US adult active Facebook user population. Those exposed are older than US adult active Facebook users in general—the dark and medium blue in panel b is higher than the second lowest dashed line, which shows the proportion of US adult active Facebook users aged 45 to 64 and 65 and above, for almost all networks. Both results align with previous research showing that conservatives and older adults are disproportionately reached by disinformation and misinformation[4,10,29], as well as research on echo chambers and selective exposure[32,33]. Those exposed to CIB and FMO networks were also highly active on Facebook

 https://doi.org/10.1038/s41562-026-02435-2

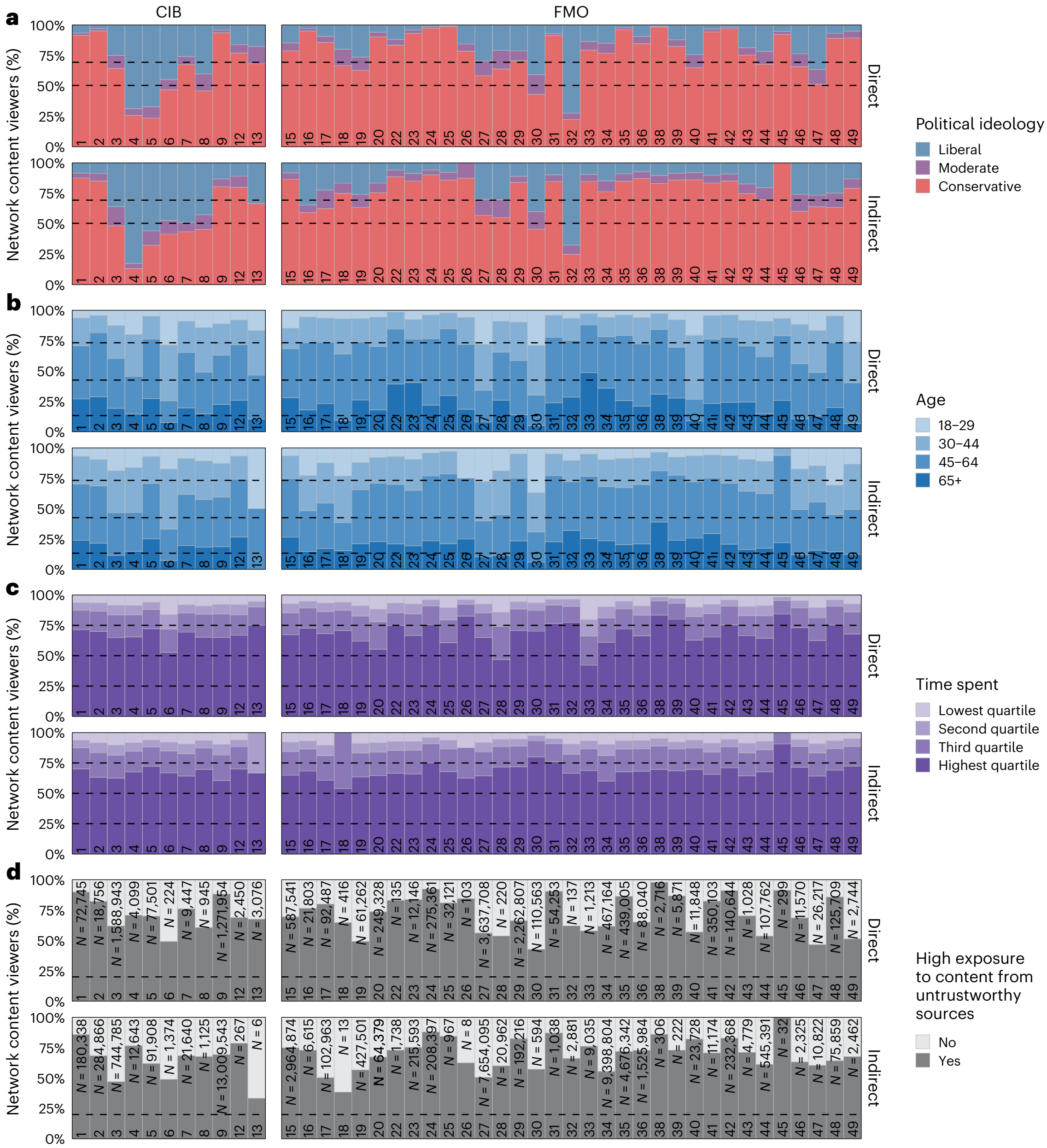


Fig. 4 | Characteristics of viewers of network content by how users are exposed to network content on Facebook. a–d, Characteristics of viewers of network content by how users are exposed to network content on Facebook; top plot in each panel is direct exposure and bottom plot is indirect exposure. Horizontal dashed lines correspond to how characteristics are distributed in the population of US adult active users in the study period. **a**, Political ideology of viewers. **b**, Age of viewers. **c**, Quartiles of time spent on Facebook. **d**, Whether viewers are in the top 20% of Facebook users in terms of exposure to content from untrustworthy sources; **d** shows the number of viewers by source of exposure at the top of each bar. Networks CIB10, CIB11, FMO14, FMO21 and FMO37 are excluded because they had less than 100 viewers in total and no observations in some of the panels.

in terms of time spent, and disproportionately among the top 20% of Facebook users who are exposed to content from untrustworthy sources (for other characteristics, see Supplementary Figs. 24–26 in Supplementary Information section 4.4). The characteristics of viewers who are exposed directly through network accounts (top plot in each panel in Fig. 4) and those exposed indirectly through non-network accounts (bottom plot in each panel in Fig. 4) are also generally similar. Several characteristics of those viewing and those resharing content are similar and deviate from the characteristics of the general user population (see Supplementary Figs. 18–22 in Supplementary Information section 4.3).

The importance of user characteristics is further highlighted by an analysis based on individual-level participant data of the relationship between exposure to and engagement with network content and downstream consequences. On the basis of a pre-registered analysis of 72,962 consenting survey participants, we examine the relationship between exposure to and engagement with deceptive online networks and individual-level outcomes controlling for a set of user characteristics, which were measured pre-exposure. The outcomes that we examine—the ability to discern true from false information, belief in the legitimacy of the 2020 elections and behavioural polarization (partisan news clicks on Facebook)—capture different aspects of politics (knowledge, attitudes and behaviour) identified by previous research as areas potentially affected by deceptive online networks[4,10,29]. We set a threshold to only analyse networks whose content was viewed by 500 or more consenting participants after Wave 2 and before Wave

 https://doi.org/10.1038/s41562-026-02435-2

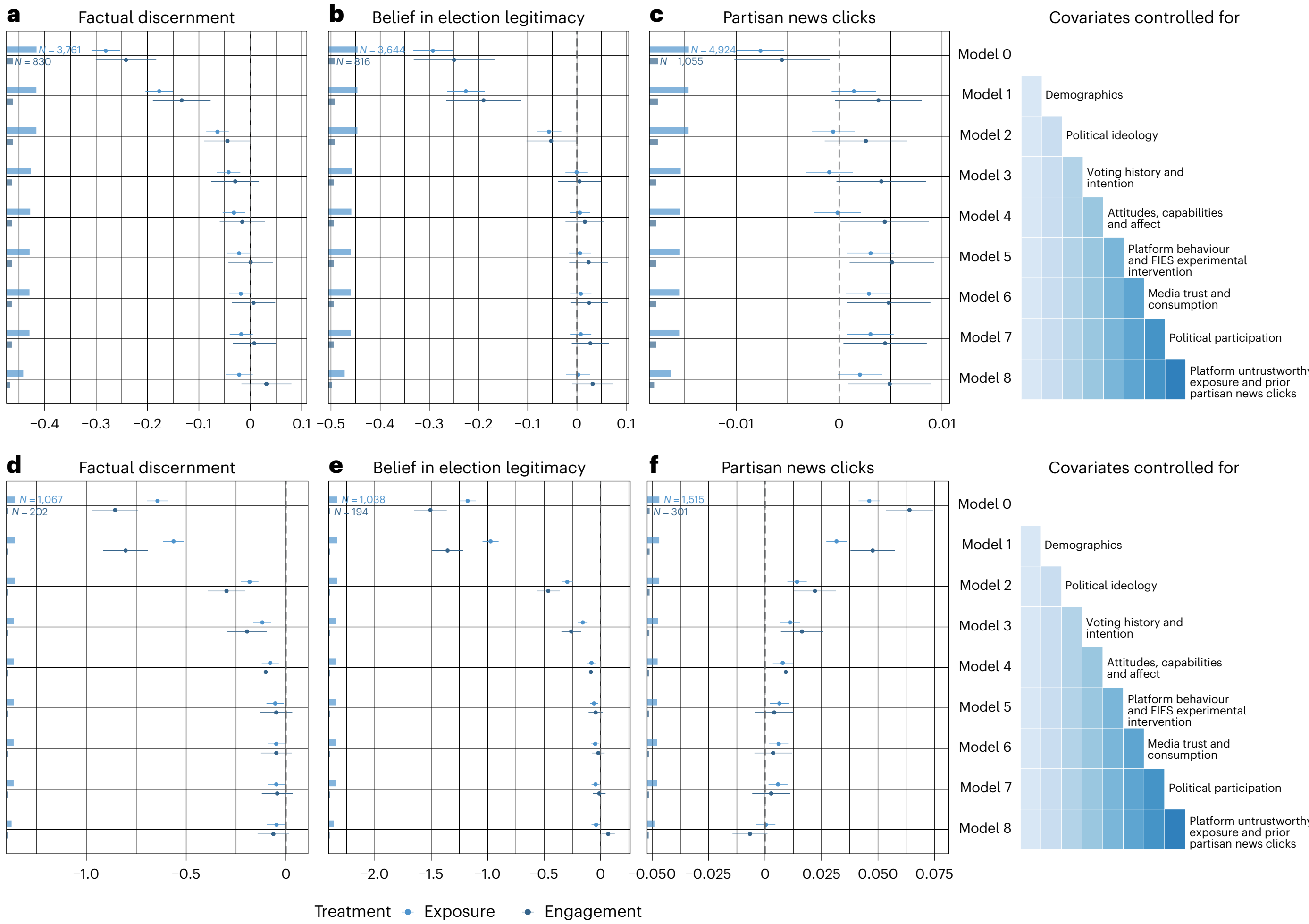


Fig. 5 | User characteristics correlate with exposure/engagement and outcomes: adding more control variables reduces the correlation between exposure/engagement and outcomes. a–c, Correlation between exposure to and engagement with FMO27 and outcomes. **d–f**, Correlation between exposure to and engagement with FMO35 and outcomes. **a** and **d** show the factual discernment outcome. **b** and **c** show the belief in the legitimacy of the election outcome. **c** and **f** show the partisan news clicks outcome. **a–f** show coefficient estimates and the 95% confidence intervals based on Huber–White robust standard errors (HC2) obtained by sequentially adding more pre-exposure covariates to the model and using entropy balancing to adjust for them. The number of participants who were exposed to or engaged with deceptive network content are shown for each model in the form of bars on the left of the coefficient plots; the model at the very top also shows exact numbers for orientation. See Supplementary Table 3 for full result details, including *P* values and confidence intervals.

4 of survey data collection (that is, between 24 September 2020 and 3 November 2020; see Supplementary Information section 3.1 for details). Only two FMO networks operating on Facebook (FMO27 from Kosovo and FMO35 from North Macedonia) met these criteria (for more details on these networks, see Supplementary Information section 3.2, Supplementary Table 5 in Supplementary Information section 4.2 for network content topics, Supplementary Fig. 9 in Supplementary Information section 4.2 for candidate mentions and Extended Data Table 1 for network descriptions).

For both FMO networks, without controlling for any individual-level characteristics, we observe that survey participants who were exposed to or engaged with the deceptive online networks tended to have less ability to discern true from false information, and were more likely to exhibit partisan news clicks (see Model 0 in the panels of Fig. 5 and Supplementary Table 3 for coefficient estimates, *P* value and confidence intervals). After adjusting for pre-exposure individual-level characteristics using entropy balancing[34]—which reweights non-exposed/non-engaged participants to achieve mean balance on these characteristics with the exposed/engaged group (see Supplementary Information section 3.1.5 for more details)—the previously strong associations between exposure to or engagement with deceptive online networks and users' factual knowledge largely disappear (FMO27: Treatment = Exposure, $t(30,003) = -1.603$, $P = 0.109$, coef = −0.022, 95% CI = (−0.048, 0.005); Treatment = Engagement, $t(30,003) = 1.267$, $P = 0.205$, coef = 0.031, 95% CI = (−0.017, 0.080); FMO35: Treatment = Exposure, $t(30,003) = -1.908$, $P = 0.056$, coef = −0.048, 95% CI = (−0.096, 0.001); Treatment = Engagement, $t(30,003) = -1.580$, $P = 0.114$, coef = −0.064, 95% CI = (−0.142, 0.015)), perceived election legitimacy (FMO27: Treatment = Exposure, $t(28,952) = 0.201$, $P = 0.840$, coef = 0.003, 95% CI = (−0.022, 0.027); Treatment = Engagement, $t(28,952) = 1.501$, $P = 0.134$, coef = 0.032, 95% CI = (−0.010, 0.074); FMO35: Treatment = Exposure, $t(28,952) = -1.931$, $P = 0.054$, coef = −0.040, 95% CI = (−0.081, 0.001); Treatment = Engagement, $t(28,952) = 2.230$, $P = 0.026$, coef = 0.067, 95% CI = (0.008, 0.126)) and subsequent online behaviour (FMO27: Treatment = Exposure, $t(36,211) = 1.834$, $P = 0.067$, coef = 0.002, 95% CI = (0, 0.004); Treatment = Engagement, $t(36,211) = 2.391$, $P = 0.017$, coef = 0.005, 95% CI = (0.001, 0.009); FMO35: Treatment = Exposure, $t(36,211) = 0.160$, $P = 0.873$, coef = 0, 95% CI = (−0.004, 0.005); Treatment = Engagement, $t(36,211) = -1.689$, $P = 0.091$, coef = −0.007, 95% CI = (−0.014, 0.001)).

We caution against interpreting these results as causal—that exposure to and engagement with network content influences (or does not influence) downstream political knowledge, attitudes and behaviour. This analysis relies on individual-level participant data, and the relatively small number of exposed participants may not represent the broader population of users who encounter network content. More generally, owing to limited knowledge about how the treatments (exposure to or engagement with the networks) are assigned, causal inference using observational data based on the unconfoundedness assumption is inherently challenging[35,36]. Our pre-registered sensitivity analyses show that these estimates are highly vulnerable to potential unobserved confounding and therefore cannot be credibly interpreted as causal (see Supplementary Information section 3.2.3 for full sensitivity analysis results). For example, individuals who have lower trust in media, higher prior approval of Donald Trump as president, are aged 45 or older, and identify as Republican are both more likely to click on content from FMO35 and more likely to doubt the legitimacy of the 2020 elections. We control for these factors, but they are illustrative rather than exhaustive and other confounders likely remain.

The observed associations between exposure to, or engagement with, two FMO networks and political outcomes are likely confounded by unobserved individual-level characteristics and should not be interpreted as evidence for causal effects—or the absence of such effects. Altogether, these results underscore the importance of understanding the characteristics of individuals who are exposed to and engage with network content.

Discussion

Deceptive online networks use various tactics of identity deception, violating community standards and norms on social media. Deceptive online networks raise normative concerns as they may mislead the public, contaminate political discourse and corrode public confidence in democracy.

In the 2020 US elections, a diverse array of deceptive online networks targeted users in the United States. While previous research has focused primarily on foreign influence campaigns from Russia, we find that deceptive online networks targeting the US electorate originate domestically and from around the world. In fact, the network with the largest reach in this study, Rally Forge, was operated by US domestic actors.

The results show that all deceptive online networks producing and disseminating political content should be scrutinized, regardless of their goals. Previous work on coordinated networks, influence operations and disinformation campaigns has largely neglected financially motivated actors, dismissing them as mere spam or clickbait. This has resulted in little research or oversight[20,21]. In our study period, Meta identified three times as many financially motivated as politically motivated deceptive networks that engaged in political discourse. These financially motivated networks reached more users and shared a substantial amount of political content. The characteristics of users reached by deceptive online networks are similar, regardless of networks' goals. Furthermore, in practice, it can be difficult to make clear attributions of intent (for example, was a network pursuing political or financial goals?). It is also conceivable that some networks are 'repurposed' over time—for example, accounts originally used for financial gain may later be sold or leveraged for explicitly political purposes. All deceptive online networks engaged in political discourse—regardless of how they are classified—may shape election contexts and deserve scrutiny, and future research could benefit from delving more deeply into the function, structural organization and dissemination tactics of networks. To enable this, it is essential that platforms not only reinvest in trust and safety efforts—reversing recent trends[37,38]—but also take proactive steps to document, retain and publicly release data on all deceptive online networks engaged in political discourse.

Leveraging unique data on actual exposure, we find that 37 million active users on Facebook and 3 million active users on Instagram were exposed to some deceptive online network content in the 2020 US elections. These numbers are likely a lower bound on exposure for several reasons: not all networks may have been detected by Meta, we only analyse an 8-month study period, only adult users are included, and not all network data were preserved (see Supplementary Information section 2.1.1 for details on network identification and Supplementary Information section 2.1.3 on missing data). While reach was highly concentrated on Instagram, with 1% of users accounting for 96% of all views of network content, it was more diffuse on Facebook, where 1% of users accounted for 55% of all views. Notably, only a very small number of deceptive networks achieved broad reach, and they accounted for the vast majority of exposures to deceptive network content. As with all of our results, care should be taken when generalizing these findings beyond the context of this study.

Reshares by non-network accounts, that is, accounts that are unaffiliated with deceptive online networks, are crucial for exposing more users to deceptive online network content. Dovetailing with previous research, this suggests that social media users are often active—albeit likely unwitting—participants in amplifying deceptive network content, not just passive information consumers[10,39–45].

The importance of non-network accounts for network reach has implications for future research and the development of solutions and countermeasures. First and foremost, it highlights the need for future research to focus on understanding the motivations and behaviour of the relatively small number of users who choose to reshare deceptive network content. For example, these users may be influenced by social reinforcement or ideological affinities[46], or they may rely on heuristic processing and intuitive decision-making[47,48] when they share network content. These users, particularly those directly exposed to networks, play a key role in the transmission chain, connecting the networks to other users who would not have been directly exposed to deceptive networks but are nevertheless reached indirectly through reshares of non-network accounts. Since humans judge reliability based in part on others' interactions, when deceptive network content is reshared by non-network users, it may make identity deception more difficult to detect[49,50]. This suggests that countermeasures aimed solely at the networks themselves may be insufficient unless paired with behavioural interventions or preventative strategies that address the reasons that users choose to reshare deceptive net work content.

Although we illustrate the role of non-network accounts in users' exposure to deceptive networks, we do not know how exactly non-network accounts facilitate spread. Our analyses of engagement with network content, such as reshares, by different user characteristics such as age and posting activity give insight into who reshares network content (Supplementary Information section 5.1). Our data on network original post diffusion show averages of cascade size, depth, maximum breadth and structural virality across all diffusion trees per network (see Supplementary Information section 5.2, Supplementary Tables 39–42 for select diffusion results; results for all pre-registered analyses can be found in the Supplementary Repository (SR); see Supplementary Information section 7 and SR on OSF (pdf passcode aZXGdq), including all diffusion results in SR2.4). However, we cannot identify whether deceptive networks intentionally targeted accounts with many followers for amplification, whether networks targeted non-network accounts with specific characteristics, or whether characteristics or behaviours of non-network accounts, beyond what is measured in this study, promote content spread[51]. This suggests a potential role that platform-specific features—for example, reshare functionality on Facebook—may play in enabling broader reach of deceptive networks, which should be explored in future research (see Supplementary Information section 4.6 for an exploratory analysis). The effectiveness of countermeasures would be strengthened with a deeper understanding of the mechanisms under which non-network reshares occur.

The emphasis on non-network accounts does not necessarily mean that future research can ignore deceptive networks' direct actions and reach. On Instagram, which has different reshare features from Facebook, CIB10, the most successful network in terms of reach on Instagram, reached 2.8 million active users directly. Even on Facebook, FMO29 reached 2.2 million active users directly, without the amplification by non-network accounts. Networks can directly reach millions of users and do so in a number of different ways. Networks can compromise popular accounts—for instance, FMO27 compromised a comedy Page to connect directly with Facebook users. Networks can use targeted ads to reach desired audiences. Rally Forge (CIB9), the most successful network in terms of reach on Facebook, utilized Facebook ads (513 in our 8-month study period; Extended Data Table 2), spending about US$1.15 million according to Meta's public reporting[52] (albeit over a potentially longer period).

Finally, it remains important to understand the causal effects of deceptive online networks, including not only whether viewers' beliefs are changed but also whether beliefs are hardened, and how networks affect general public perceptions. In our observational data, we were unable to reach firm conclusions about the causal effects of networks. Future research could focus on sampling people likely to be exposed to networks to improve power or run longer-term studies.

The data underlying this study, including all figures and tables in the SI and SR, is available to create opportunities to analyse deceptive online networks further (for details on access, see 'Data availability'). Platform transparency and data sharing with researchers is essential in further investigating the role and impact of deceptive online networks.

We aim to promote clarity in public debates and to foster evidence-based policymaking in three main ways: (1) providing a conceptualization of deceptive online networks that moves beyond networks' goals, (2) directly measuring exposure to deceptive networks and (3) using a variety of approaches to show how platform users are exposed to deceptive networks. The nature of deceptive online networks is complex, and the mechanisms of their reach are multifaceted. However, this study highlights the key role of social sharing. Accounts unaffiliated with deceptive online networks played a pivotal role in spreading network content to broader audiences. Understanding the importance of resharing by non-network accounts and the characteristics of the users who inadvertently become accomplices in deception is crucial for effectively addressing the issue of deceptive online networks on social media.

Methods

The methods used in this research are part of the broader US 2020 FIES project[28,30,31,53].

Ethical considerations

The research design prioritized ethical considerations, with attention to minimizing risks both for individual respondents and for society more broadly. Meta applied for and received approval from the NORC Institutional Review Board (IRB) to conduct the experimental studies (protocol number 20.08.10, project number 8870). Each academic collaborator also coordinated with their home institution's IRB to meet Human Subjects Research requirements for analysing the aggregated, de-identified data that Meta and NORC collected. The same protocol was submitted to 13 academic institutions, 9 of them determined that the research was not human subjects, not research or irrelevant for the university IRB (that is, the university was not engaged in the research). Of the remaining four, three universities determined that the protocol was exempt (University of Texas Austin IRB number 2020-07-0062, William and Mary IRB number PHSC-2020-09-21-14523-jsettle and George Washington University IRB number NCR213250), and one university conducted expedited review (University of Pennsylvania IRB number 844651). In addition, Meta engaged Ethical Resolve, a data ethics consultancy, to advise both company and academic team members before the research was implemented. This external review assessed the project against established principles of research ethics and evolving standards for digital research.

Definitions

We define terms in Supplementary Information section 2.0.1, such as what constitutes a 'view' (for example, on Facebook, a view is counted whenever the post renders in the visible portion of a user's web browser or mobile device for more than 250 milliseconds).

We describe and show the performance of classifiers as well as other categorization methods that are applied to the data, such as those used to identify political content, in Supplementary Information sections 2.0.2 and 2.0.3.

Study design

This study was initially pre-registered on 16 October 2020 at https://doi.org/10.17605/OSF.IO/D7QES (all deviations and clarification can be found in Supplementary Information section 8). The study utilizes multiple types of data, including aggregated platform data, individual-level participant data and network data. We provide an overview of each data type here and refer to relevant Supplementary Information sections for more details.

Data

Aggregated platform data. Aggregated platform data refers to data on all adult US-based Facebook and Instagram users who were active for at least 1 day between 26 June 2020 and 15 February 2021 (the study period). The size of this sample is about 250 million adults for Facebook and about 160 million adults for Instagram. For Facebook, about 46.8% of users are male and 53.2% are female; 26.7% are aged 18–29, 30.7% 30–44, 29.2% 45–64 and 13.3% 65 plus; 28.6% are politically liberal, 19% moderate and 50.4% conservative. For Instagram, about 42.4% of users are male and 57.6% are female. This sample is representative of the adult active Facebook and Instagram US user population, but not the general US population. For more details, see Supplementary Information section 2.2.

Individual-level participant data. Individual-level participant data refers to data from US Facebook and Instagram users aged 18 and over who consented to participate in a study of social media and politics and completed 2 baseline survey waves. The data include individual-level survey responses as well as participants' on-platform behavioural data. Our sampling frames were drawn from adult (18+) monthly active users in the United States on Facebook and Instagram as of 17 August 2020. Eligible respondents were those who could receive general platform surveys, representing a random subset of each platform's total user population. Participants were asked to confirm that they were over 18 years of age and lived in the United States as part of the recruitment process. The total sample size is about 73,000, although we use different subsamples for different analyses. Of these users, 43.3% are male and 56.7% are female or another gender (after excluding missing values); 17.9% are aged 18–29, 41.8% 30–44, 33.3% 45–64 and 7% 65 plus (after excluding missing values). In terms of political identification, 52.4% are Democrat or Democrat-leaning, 34.9% Republican or Republican-leaning, and 12.8% Independent. This sample is neither representative of the adult active Facebook and Instagram US user population nor the general US population. For more details, see Supplementary Information section 2.3. For individual-level data, we excluded participants who deleted their account, withdrew from the study or requested data deletion from the analysis for ethical reasons since we could not access their data.

Network data. Deceptive online networks are operationalized as networks disabled under the CIB protocol, which consisted of influence operations, whether domestic or foreign in origin, that attempted to

https://doi.org/10.1038/s41562-026-02435-2

manipulate or corrupt public debate for a strategic goal. Networks also include those disabled under the IB protocol. These consisted of FMOs that were identified by an informal, internal Meta taskforce as networks using inauthentic accounts to post content on American political themes—usually to drive users towards off-platform websites where the operators could monetize users' clicks. For more details on data related to deceptive online networks, including how networks were identified, how these data were prepared, missing data and more detailed descriptions of networks, see Supplementary Information section 2.1. Note that our operationalization of network content excludes network comments, which some networks such as Rally Forge commonly used[52], implying that our numbers on content production and reach are underestimates. Further, owing to lower performance for other languages, we focus on posts in English and Spanish for topic classifications, meaning that we do not take posts in other languages into account when assessing the share of political content or topics about which a network posted.

Sampling strategy

Aggregated platform data. This represents the population of US-based adult active users on Facebook and Instagram. For more details, see Supplementary Information section 2.2.

Individual-level survey data. Below, we summarize our sampling strategy. Further details are available in Supplementary Information section 2.3.2. Participants were sampled to achieve specific sample targets across different stages of the study, which were, in turn, chosen to achieve desired minimum detectable effect sizes for experimental interventions related to other papers part of the US 2020 FIES across different subgroups among the set of respondents participating in the Wave 1 and Wave 2 surveys and Wave 4 or 5. The sampling frame consisted of all US Facebook and Instagram users aged 18 and above who were active monthly users and qualified to participate in standard platform surveys as of 17 August 2020. This participant pool constituted a randomly selected subset of the broader user bases across both social media platforms. The Facebook sampling frame was trimmed by removing predicted fake accounts, employees and advertisers. For Instagram, creators and business accounts also were removed. Finally, because the use of multiple accounts is common among Instagram users, the sampling frame was narrowed to include only a user's primary account (for users with multiple accounts, this is the oldest account). Having defined the sampling frames, sampling probabilities were computed to achieve specific sample distributions for the set of demographics encoded in the stratification step across each of the samples of interest. The sampling probabilities took into account (a) differential non-response across different demographics and (b) the desired sample size across the different studies. Recruitment occurred via a message displayed at the top of Facebook and Instagram feeds for a random subset of users, inviting them to share their opinion. Users who clicked 'Start Survey' were taken to an IRB-approved consent form describing the study procedures, potential benefits and risks, and compensation. As part of this process, participants confirmed that they were at least 18 years old and resided in the United States. During recruitment, individuals were required to verify their age (18+) and US residency status. The sample size is $N = 72{,}962$, with $N = 19{,}510$ participants randomized into the control group for the FIES platform interventions. Additional information regarding sampling methodology, recruitment and response rates can be found in Supplementary Information section 2.3. Data collection and analysis were not performed blind to the conditions of the experiments.

Network data. Deceptive online networks are operationalized as networks Facebook (now Meta) disabled between 26 June 2020 and 3 November 2021 (1 year after election day) that targeted US users during the study period (26 June 2020 and 15 February 2021). This resulted in 49 networks. For more details, including on identification of networks, see Supplementary Information section 2.1. For one analysis on the downstream effects of networks, we subset to networks that reached a pre-registered reach threshold between different survey waves (see Supplementary Information section 3.1.1 for details).

Data collection

The study's data collection involved a partnership between Meta and NORC, a survey research organization at the University of Chicago. While Meta handled the majority of participant recruitment and gathered information from Facebook and Instagram platforms, NORC managed the survey administration process, including expanding the participant pool through additional recruitment, gathering supplementary information beyond the social media platform data and ensuring participant anonymity by stripping identifying information before data integration and distribution to researchers. Participants provided data through their personal devices, including smartphones and computers. NORC operated without knowledge of the research hypotheses or participants' experimental assignments.

Meta's internal logging systems captured on-platform behavioural and network data. NORC administered the surveys through their established survey infrastructure and collaborated with two vendors—MDI Global and RealityMine—to gather the passive measurement data.

Aggregated platform data. For Facebook, aggregated data are based on computing the total count for the study period (26 June 2020 until 15 February 2021). For Instagram, metrics are aggregated between 17 July 2020 and 15 February 2021 owing to an unanticipated data gap.

Individual-level survey data. Survey data collection started on 31 August 2020. Two surveys were fielded pre-treatment: Wave 1 (31 August–12 September) and Wave 2 (8–23 September). The treatments that we analyse in more detail (less like-minded content, no reshares and reverse chronological feed) ran from 24 September to 23 December; other treatments had different timelines (Supplementary Information section 3.1). During the treatment period, three more surveys were administered: Wave 3 (9–23 October), Wave 4 (4–18 November) and Wave 5 (9–23 December). See Supplementary Information section 2.3 for details on the survey and platform behavioural data for individual-level participants and data collection period details.

Network data. Network data are measured for the study period (26 June 2020 until 15 February 2021). For more details, see Supplementary Information section 2.1.

Data transparency

The study relies on data provided by Meta, which offers an important view of deceptive networks targeting US adult users on Facebook and Instagram between 26 June 2020 and 3 November 2021. At the same time, this reliance raises methodological constraints. First, we can only analyse what Meta made available. Not all data from the identified networks were retained (see Supplementary Information section 2.1 for details on missing data), and the informal taskforce engaged in flagging FMOs may not have captured all relevant networks. Importantly, we cannot independently verify the completeness of the dataset, and as a result, there may be biases in the data (for example, selection bias in identifying networks) that we cannot assess. Second, all attributions of network characteristics and activity come from Meta. Despite requests, Meta did not provide descriptions of all FMO networks because they did not have the resources in their investigative teams to conduct post hoc investigations of all FMO networks. Meta provided limited explanation of how they determine whether an account is affiliated with a network versus merely connected to it (non-network accounts). Despite repeated requests for details, all that was provided to us by Meta regarding network identification is what is presented in Supplementary Information section 2.1.1. These constraints should be kept in mind when interpreting the findings. Nonetheless, the dataset has advantages: it captures actual exposure to deceptive networks across the entire population of US adult users on Facebook and Instagram, in contrast to user surveys or qualitative studies that rely on self-reports, presumed exposure or limited samples. These data also provide a fine-grained view of networks and their activity over time. While the data lack the depth of ethnographic or interview-based approaches, they enable a population-level analysis of deceptive networks that has not previously been possible.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

De-identified data from this study are stored in the Social Media Archive (SOMAR) housed by ICPSR. This is accessible for university research approved by the Institutional Review Board (IRB) related to elections, or for the purpose of validating the results of this study. ICPSR will handle and vet all applications requesting access to these data. Data access is controlled to protect the privacy of the study participants and to be consistent with the consent form signed by study participants. Requests for data can be made via SOMAR. The data for this paper are based on the following datasets stored in SOMAR: https://socialmediaarchive.org/record/87, https://socialmediaarchive.org/record/88, https://socialmediaarchive.org/record/89, https://socialmediaarchive.org/record/90, https://socialmediaarchive.org/record/91, https://socialmediaarchive.org/record/92, https://socialmediaarchive.org/record/93, https://socialmediaarchive.org/record/94, https://socialmediaarchive.org/record/95, https://socialmediaarchive.org/record/96, https://socialmediaarchive.org/record/97, https://socialmediaarchive.org/record/98, https://socialmediaarchive.org/record/99, https://socialmediaarchive.org/record/100, https://socialmediaarchive.org/record/101, https://socialmediaarchive.org/record/102 and https://socialmediaarchive.org/record/103. In addition, we used public data sources for generating map plots and cleaning emojis in hashtags. In particular, for the world map in the main text, the ggplot2 package map_data function relies on world map data from https://www.naturalearthdata.com. We also downloaded centroid data from https://raw.githubusercontent.com/gavinr/world-countries-centroids/master/dist/countries.csv. For the Congressional district maps in the supplementary repository, we used a Congressional district shapefile from https://www2.census.gov/geo/tiger/GENZ2020/shp/cb_2020_us_cd116_20m.zipand Congressional district code file from https://www2.census.gov/geo/docs/reference/codes/files/national_cd113.txt. To clean emojis in hashtags, we used Unicode data to filter out emojis. The raw dictionary was downloaded from http://www.unicode.org/Public/emoji/1.0//emoji-data.txt. Emoji lists were downloaded from http://www.unicode.org/reports/tr51/index.html#emoji_data. For a comparison to the voting-age population, we gathered the voting-age population size from https://www.census.gov/content/dam/Census/library/publications/2022/demo/p20-585.pdf. Source data are provided with this paper.

Code availability

The code from this study is stored in the Social Media Archive (SOMAR) housed by ICPSR and is accessible under the same terms as the de-identified data. Access to the data can be requested in the following links: https://socialmediaarchive.org/record/87, https://socialmediaarchive.org/record/88, https://socialmediaarchive.org/record/89, https://socialmediaarchive.org/record/90, https://socialmediaarchive.org/record/91, https://socialmediaarchive.org/record/92, https://socialmediaarchive.org/record/93, https://socialmediaarchive.org/record/94, https://socialmediaarchive.org/record/95, https://socialmediaarchive.org/record/96, https://socialmediaarchive.org/record/97, https://socialmediaarchive.org/record/98, https://socialmediaarchive.org/record/99, https://socialmediaarchive.org/record/100, https://socialmediaarchive.org/record/101, https://socialmediaarchive.org/record/102 and https://socialmediaarchive.org/record/103. The data in this study were analysed using R (v.4.3.3), which was executed on via R notebooks on JupyterLab (v.4.2.4). The analysis code imports several R packages available on CRAN, including tidyverse (v.2.0.0), dplyr (v.1.1.4), ggplot2 (v.3.5.2), estimatr (v.1.0.4), Hmisc (v.5.1-0), sf (v.1.0-16), ggrepel (v.0.9.6), patchwork (v.1.2.0), scales (v.1.3.0), RColorBrewer (v.1.1-3), lubridate (v.1.9.4), stringr (v.1.5.1) and hbal (v.1.2.12).

References

1. *Reply-Guys Go Hunting: An Investigation into a U.S. Astroturfing Operation on Facebook, Twitter, and Instagram* (Stanford Internet Observatory, 2020); https://purl.stanford.edu/vh222ch4142
2. Stanley-Becker, I. Pro-Trump youth group enlists teens in secretive campaign likened to a 'troll farm,' prompting rebuke by Facebook and Twitter. *The Washington Post* (2020); https://www.washingtonpost.com/politics/turning-point-teens-disinformation-trump/2020/09/15/c84091ae-f20a-11ea-b796-2dd09962649c_story.html
3. Alizadeh, M., Shapiro, J. N., Buntain, C. & Tucker, J. A. Content-based features predict social media influence operations. *Sci. Adv.* **6**, eabb5824 (2020).
4. Bail, C. A. et al. Assessing the Russian Internet Research Agency's impact on the political attitudes and behaviors of American Twitter users in late 2017. *Proc. Natl Acad. Sci. USA* **117**, 243–250 (2020).
5. Ferrara, E., Chang, H., Chen, E., Muric, G. & Patel, J. Characterizing social media manipulation in the 2020 US presidential election. *First Monday* **25**, 11431 (2020).
6. Rid, T. *Active Measures: The Secret History of Disinformation and Political Warfare* (Farrar, Straus and Giroux, 2020).
7. Arif, A., Stewart, L. G. & Starbird, K. Acting the part: examining information operations within #BlackLivesMatter discourse. In *Proc. ACM on Human-Computer Interaction* Vol. 2 20 (ACM, 2018).
8. Badawy, A., Ferrara, E. & Lerman, K. Analyzing the digital traces of political manipulation: the 2016 Russian interference Twitter campaign. In *Proc. IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* 258–265 (IEEE, 2018).
9. DiResta, R. et al. *The Tactics & Tropes of the Internet Research Agency* (New Knowledge, 2019); https://digitalcommons.unl.edu/senatedocs/2/
10. Eady, G. et al. Exposure to the Russian Internet Research Agency foreign influence campaign on Twitter in the 2016 US election and its relationship to attitudes and voting behavior. *Nat. Commun.* **14**, 62 (2023).
11. Golovchenko, Y., Buntain, C., Eady, G., Brown, M. A. & Tucker, J. A. Cross-platform state propaganda: Russian trolls on Twitter and YouTube during the 2016 U.S. presidential election. *Int. J. Press Politics* **25**, 357–389 (2020).
12. Howard, P. N., Ganesh, B., Liotsiou, D., Kelly, J. & François, C. The IRA, social media and political polarization in the United States, 2012–2018. *Digital Commons* https://digitalcommons.unl.edu/senatedocs/1 (2019).
13. Jamieson, K. H. *Cyberwar: How Russian Hackers and Trolls Helped Elect a President: What We Don't, Can't, and Do Know* (Oxford Univ. Press, 2020).
14. Kim, Y. M. et al. The Stealth Media? Groups and targets behind divisive issue campaigns on Facebook. *Polit. Commun.* **35**, 515–541 (2018).
15. Nizzoli, L., Tardelli, S., Avvenuti, M., Cresci, S. & Tesconi, M. Coordinated behavior on social media in 2019 UK general election. In *Proc. International AAAI Conference on Web and Social Media* 443–454 (AAAI, 2021).

16. Woolley, S. & Howard, P. *Computational Propaganda Worldwide: Executive Summary* (Oxford Internet Institute, 2017).
17. Chakraborty, A., Paranjape, B., Kakarla, S. & Ganguly, N. Stop clickbait: detecting and preventing clickbaits in online news media. In *Proc. IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* 9–16 (IEEE, 2016).
18. Lu, Y. & Pan, J. Capturing clicks: how the Chinese government uses clickbait to compete for visibility. *Polit. Commun.* **38**, 23–54 (2021).
19. Munger, K. All the news that's fit to click: the economics of clickbait media. *Polit. Commun.* **37**, 376–397 (2020).
20. Cvetkovska, S., Belford, A., Silverman, C. & Feder, J. L. The secret players behind Macedonia's fake news sites. *OCCRP* https://www.occrp.org/en/project/spooks-and-spin-information-war-in-the-balkans/the-secret-players-behind-macedonias-fake-news-sites (2018).
21. Hughes, H. C. & Waismel-Manor, I. The Macedonian fake news industry and the 2016 US election. *PS Polit. Sci. Polit.* **54**, 19–23 (2021).
22. Subramanian, S. Inside the Macedonian fake-news complex. *WIRED* https://www.wired.com/2017/02/veles-macedonia-fake-news (2017).
23. *Inauthentic Behavior* (Meta, 2024); https://transparency.meta.com/policies/community-standards/inauthentic-behavior/
24. Gleicher, N. *Coordinated Inauthentic Behavior Explained* (Meta, 2018); https://about.fb.com/news/2018/12/inside-feed-coordinated-inauthentic-behavior/
25. *Report on Russian Active Measures Campaigns and Interference in the 2016 U.S. Election, Volume 2: Russia's Use of Social Media with Additional Views* (United States Senate Select Committee on Intelligence, 2019); https://www.intelligence.senate.gov/sites/default/files/documents/Report_Volume2.pdf
26. Allen, J., Howland, B., Mobius, M., Rothschild, D. & Watts, D. J. Evaluating the fake news problem at the scale of the information ecosystem. *Sci. Adv.* **6**, eaay3539 (2020).
27. Hosseinmardi, H. et al. Examining the consumption of radical content on YouTube. *Proc. Natl Acad. Sci. USA* **118**, e2101967118 (2021).
28. Nyhan, B. et al. Like-minded sources on Facebook are prevalent but not polarizing. *Nature* **620**, 137–144 (2023).
29. Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B. & Lazer, D. Fake news on Twitter during the 2016 US presidential election. *Science* **363**, 374–378 (2019).
30. Guess, A. M. et al. How do social media feed algorithms affect attitudes and behavior in an election campaign? *Science* **381**, 398–404 (2023).
31. Guess, A. M. et al. Reshares on social media amplify political news but do not detectably affect beliefs or opinions. *Science* **381**, 404–408 (2023).
32. Del Vicario, M. et al. The spreading of misinformation online. *Proc. Natl Acad. Sci. USA* **113**, 554–559 (2016).
33. Flaxman, S., Goel, S. & Rao, J. M. Filter bubbles, echo chambers, and online news consumption. *Public Opin. Q.* **80**, 298–320 (2016).
34. Hainmueller, J. Entropy balancing for causal effects: a multivariate reweighting method to produce balanced samples in observational studies. *Polit. Anal.* **20**, 25–46 (2012).
35. Imbens, G. W. & Rubin, D. B. *Causal Inference in Statistics, Social, and Biomedical Sciences* (Cambridge Univ. Press, 2015).
36. Imbens, G. & Xu, Y. Comparing experimental and nonexperimental methods: what lessons have we learned four decades after Lalonde (1986)? *J. Econ. Perspect.* **39**, 173–202 (2025).
37. Jackson, D. W., Adler, W. T., Dougall, D. & Jain, A. *Seismic Shifts: How Economic, Technological, and Political Trends are Challenging Independent Counter-Election-Disinformation Initiatives in the United States* (Center for Democracy & Technology, 2023); https://cdt.org/wp-content/uploads/2023/09/seismic-shifts-091823-1302.pdf
38. Menn, J., Schaffer, A., Nix, N. & Morse, J. Chinese propaganda accounts found by Meta still flourish on X. *The Washington Post* (2024); https://www.washingtonpost.com/technology/2024/02/16/x-meta-china-disinformation/
39. Asmolov, G. The effects of participatory propaganda: from socialization to internalization of conflicts. *J. Des. Sci.* **6**, (2019).
40. DiResta, R. *Invisible Rulers: The People Who Turn Lies into Reality* (PublicAffairs, 2024).
41. Nemer, D. Disentangling Brazil's disinformation insurgency: President Jair Bolsonaro is already casting doubt on the validity of the 2022 elections. As in 2018, WhatsApp is poised to be a key spreader of viral falsehoods. *NACLA Rep. Am.* **53**, 406–413 (2021).
42. Ong, J. C. & Cabañes, J. V. A. *Architects of Networked Disinformation: Behind the Scenes of Troll Accounts and Fake News Production in the Philippines* (Newton Tech4Dev Network, 2018).
43. Prochaska, S. et al. Mobilizing manufactured reality: how participatory disinformation shaped deep stories to catalyze action during the 2020 US presidential election. *Proc. ACM Hum.-Comput. Interact.* **7**, CSCW1 (2023).
44. Starbird, K., Arif, A. & Wilson, T. Disinformation as collaborative work: surfacing the participatory nature of strategic information operations. In *Proc. ACM on Human-Computer Interaction*, CSCW, Vol. 3 (ACM, 2019).
45. Wanless, A. & Berk, M. in *Social Media and Social Order* (eds Herbert, D. & Fisher-Høyrem, S.) 111–139 (De Gruyter Open Poland, 2021).
46. Messing, S. & Westwood, S. J. Selective exposure in the age of social media: endorsements Trump partisan source affiliation when selecting news online. *Commun. Res.* **41**, 1042–1063 (2014).
47. Chaiken, S. & Ledgerwood, A. in *Handbook of Theories of Social Psychology* Vol. 1 (eds Van Lange, P. A. M., Kruglanski, A. W. & Higgins, E. T.), 246–266 (Sage, 2012).
48. Petty, R. E., Cacioppo, J. T., Petty, R. E. & Cacioppo, J. T. *The Elaboration Likelihood Model of Persuasion* (Springer, 1986).
49. Bhadani, S. et al. Political audience diversity and news reliability in algorithmic ranking. *Nat. Hum. Behav.* **6**, 495–505 (2022).
50. Groh, M. et al. Human detection of political speech deepfakes across transcripts, audio, and video. *Nat. Commun.* **15**, 7629 (2024).
51. Lorenz-Spreen, P., Lewandowsky, S., Sunstein, C. R. & Hertwig, R. How behavioural sciences can promote truth, autonomy and democratic discourse online. *Nat. Hum. Behav.* **4**, 1102–1109 (2020).
52. *October 2020 Coordinated Inauthentic Behavior Report* (Facebook, 2020); https://about.fb.com/wp-content/uploads/2020/11/October-2020-CIB-Report.pdf
53. González-Bailón, S. et al. Asymmetric ideological segregation in exposure to political news on Facebook. *Science* **381**, 392–398 (2023).

Acknowledgements

The Facebook Open Research and Transparency (FORT) team provided substantial support in executing the overall project. We are grateful for support on various aspects of project management from C. Nayak, S. Zahedi, I. Rosenn, L. Ahmad, A. Bhalla, C. Chan, A. Gruen, B. Hillenbrand, D. Li, P. McLeod, D. Rice and N. Shah; engineering from Y. Chen, S. Chen, T. Huang, T. Lohman, R. Moodithaya, R. Pyke, L. Saloumi, Y. Wan and F. Yan; data engineering from S. Chintha, J. Cronin, D. Desai, Y. Kiraly, T. Li, X. Liu, S. Pellakuru and C. Xie; data science and research from H. Connolly-Sporing, S. Tan and T. Wynter; academic partnerships from R. Mersey, M. Zoorob, L. Harrison, S. Aisiks, Y. Rubinstein and C. Qiao; privacy and legal assessment from K. Benzina, F. Fatigato, J. Hassett, S. Iyengar, P. Mohassel, A. Muzaffar, A. Raghunathan and A. Sun; and content design from C. Bernard, J. Breneman, D. Leto and S. Raj. NORC at the University of Chicago partnered with Meta on this project to conduct the fieldwork with the survey participants and pair the survey data with web tracking data for consented participants in predetermined aggregated forms. We are particularly grateful for the partnership of NORC principal investigator J. M. Dennis and NORC project director M. Montgomery. We thank participants of the Digital Repression Workshop at the Hertie School, participants at the Trust and Safety Research Conference at Stanford, Stanford Communication Department Colloquium participants, Graphika, DFRLab, Jigsaw, the Social Technologies Lab at Cornell, the Stanford HCI group, the Stanford Social Media Lab and the Stanford Internet Observatory for helpful feedback on earlier versions of this paper. Meta funded all aspects of project implementation and data collection. Ancillary support (for example, research assistants and course buyouts), as applicable, was sourced by academics from (authors' initials in parenthesis): the Democracy Fund (N.J.S.), the European Research Council Starting Grant (EXPO-756301) (M.W.), the Guggenheim Foundation (B. Nyhan), the Hewlett Foundation (N.J.S. and J.A.T.), the Hopewell Fund (J.A.T.), the John S. and James L. Knight Foundation (D.F., D.L., N.J.S., J.A.T. and R.T.), the Charles Koch Foundation (J.A.T.), New York University (J.A.T.), the Alfred P. Sloan Foundation (M.G.), Stanford University (H.A., M.G., N.M., J.P. and Y.X.), the Stanford Institute for Economic Policy Research (M.G.), the Stanford Freeman Spogli Institute for International Studies (J.P.), the Hoover Institution at Stanford University (Y.X.), the University of Texas at Austin (N.J.S.) and the University of Wisconsin-Madison (Y.M.K.). R.E.A. was supported by an SAP Stanford Graduate Fellowship in Science and Engineering, a Stanford Center on Philanthropy and Civil Society PhD Research Fellowship and a Stanford Impact Labs Summer Collaborative Research Fellowship. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

R.E.A., Y.M.K., J.P., Y.X., B. Nimmo and D.R.T. supervised all analyses and wrote the paper. R.E.A., Y.M.K., J.P. and Y.X. were academic lead authors, and Y.M.K., J.P. and Y.X. had final control rights. B. Nimmo and D.R.T. were the lead authors at Meta. R.E.A., J.P. and Y.X. analysed the data. R.E.A., Y.M.K., J.P., Y.X., B. Nimmo and D.R.T. designed the study. R.E.A., Y.M.K., J.P., Y.X., D.R.T., P.B., D.D., D.F., N.M., D.M., B. Nyhan, E.T., R.T., C.V.R., A.W. and M.W. contributed study materials (for example, survey questionnaires, classifiers and software). R.E.A., Y.M.K., J.P., Y.X., B. Nimmo, D.R.T., P.B., H.A., D.F., M.G., S.G.-B., A.M.G., S.I., D.L., N.M., B. Nyhan, J.S., E.T., R.T., M.W., A.C.-T., D.M., C.V.R., C.K.d.J., A.F., W.M., N.J.S. and J.A.T. contributed to the design of the project. D.R.T., P.B., T.B., A.C.-T., D.D., D.M., C.V.R., B.X., A.W., A.F. and W.M. coordinated the implementation of the experimental intervention and collected and curated all platform data. B. Nimmo coordinated the collection of qualitative deceptive online network descriptions. R.E.A., J.P. and Y.X. contributed the figures and tables. D.L., R.T., M.W., A.C.-T., T.B., B. Nyhan, P.B., M.G., S.G.-B., A.F., W.M., N.J.S. and J.A.T. provided feedback on the paper. N.J.S. and J.A.T. were joint principal investigators for the academic involvement on this project, responsible for management and coordination. C.K.d.J., A.F. and W.M. led Meta's involvement on this project and were responsible for management and coordination.

Competing interests

The costs associated with the research (such as participant fees, recruitment and data collection) were paid by Meta, and some authors are employed by Meta. To ensure transparency and integrity in the research process, we adopted the following conventions. First, none of the academic researchers nor their institutions received financial or any other compensation from Meta for their participation in the project. Second, all of the papers resulting from the US 2020 FIES, including this one, were pre-registered at the Open Science Foundation. Third, for every paper, a set of authors with control rights over the final content of the paper were specified. The authors with control rights for this paper are Y.M.K., J.P. and Y.X. Fourth, Meta publicly agreed that there would be no pre-publication approval of papers for publication on the basis of their findings. Finally, we appointed a rapporteur for the project—M. Wagner of the University of Wisconsin, Madison—who was neither a paid employee of Meta nor a member of the independent academic research team. For more information, see Supplementary Information section 1. The following authors are employed by Meta: B. Nimmo, D.R.T., P.B., T.B., A.C.-T., D.D., A.F., C.K.d.J., W.M., D.M., C.V.R., A.W. and B.X. Below, we list additional declarations from the academic author team: owns Meta Stock (J.S. and Y.X.), conducted paid consulting work for Meta (N.J.S.), received direct research funding from Meta (A.M.G., S.I., B. Nyhan, J.P., J.S., N.J.S., R.T., J.A.T. and M.W.), received an honorarium/fee from Meta for attending and/or hosting an event/serving as outside expert (M.G., J.P., J.A.T. and N.M.), attended a Meta event where food, travel or lodging was paid for by the company (R.E.A., D.F., M.G., S.G.-B., A.M.G., S.I., Y.M.K., D.L., N.M., B. Nyhan, J.P., J.S., N.J.S., E.T., R.T., J.A.T. and M.W.), own individual stocks at a related company (J.S.), received direct research funding from a related company (N.J.S. and R.T.), received an honorarium/fee from a related company for attending and/or hosting an event/serving as outside expert (M.G.), and attended an event at a related company where food, travel or lodging was paid for by the company (M.G., D.L., N.M., B. Nyhan, N.J.S., J.A.T. and M.W.). The other authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41562-026-02435-2.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-026-02435-2.

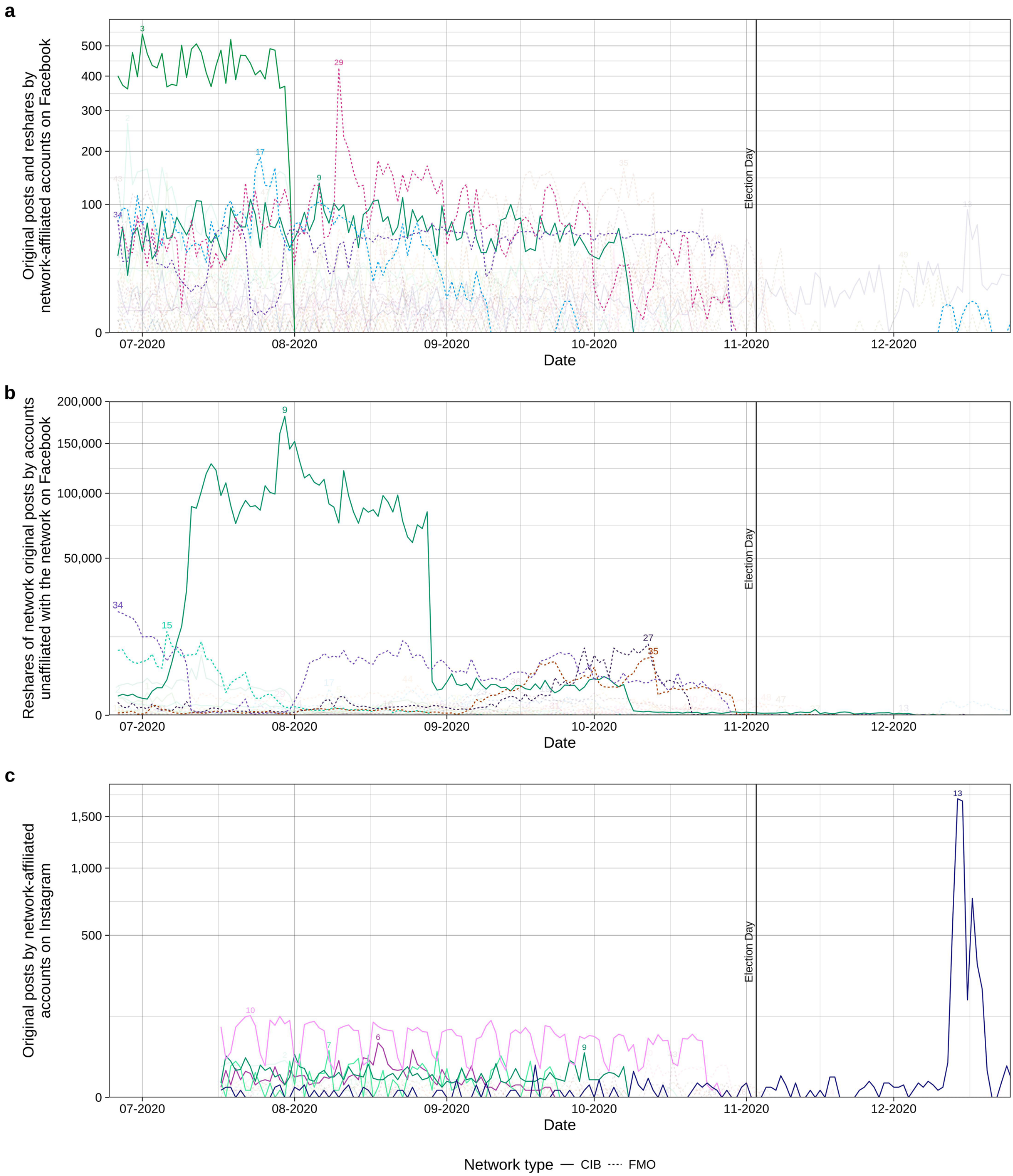
Correspondence and requests for materials should be addressed to Jennifer Pan.

Peer review information *Nature Human Behaviour* thanks Jiyoung Lee, Tamar Mitts and Manoel Horta Ribeiro for their contribution to the peer review of this work.

https://doi.org/10.1038/s41562-026-02435-2

[1]Department of Communication, Stanford University, Stanford, CA, USA. [2]School of Journalism and Mass Communication, University of Wisconsin-Madison, Madison, WI, USA. [3]Department of Political Science, Stanford University, Stanford, CA, USA. [4]Meta, Menlo Park, CA, USA. [5]Stanford Doerr School of Sustainability, Stanford University, Stanford, CA, USA. [6]Data Science Lab, Hertie School, Berlin, Germany. [7]Annenberg School for Communication, University of Pennsylvania, Philadelphia, PA, USA. [8]Department of Economics, Stanford University, Stanford, CA, USA. [9]Department of Politics and School of Public and International Affairs, Princeton University, Princeton, NJ, USA. [10]Network Science Institute and Department of Political Science and Khoury College of Computer Sciences, Northeastern University, Boston, MA, USA. [11]Institute of Quantitative Social Science, Harvard University, Cambridge, MA, USA. [12]Graduate School of Business, Stanford University, Stanford, CA, USA. [13]Department of Government, Dartmouth College, Hanover, NH, USA. [14]Department of Government, William and Mary, Williamsburg, VA, USA. [15]Department of Political Science, Syracuse University, Syracuse, NY, USA. [16]Department of Political Science and Khoury College of Computer Sciences, Northeastern University, Boston, MA, USA. [17]Department of Communication, University of California, Davis, Davis, CA, USA. [18]Center for Excellence in Social Science, University of Warsaw, Warsaw, Poland. [19]Moody College of Communication and Center for Media Engagement, University of Texas at Austin, Austin, TX, USA. [20]Wilf Family Department of Politics and Center for Social Media, AI and Politics, New York University, New York, NY, USA. ✉e-mail: jp1@stanford.edu

https://doi.org/10.1038/s41562-026-02435-2



Extended Data Fig. 1 | Network activity over time by post type on Facebook and Instagram. **a**, Number of original posts and reshares by network-affiliated accounts on Facebook by network, over time. **b**, Number of reshares of network original posts by accounts unaffiliated with the network on Facebook by network, over time. **c**, Number of original posts by network-affiliated accounts on Instagram. In each panel, the top 5 networks in terms of total activity are shown in full colors, while the remaining networks are shown in opaque colors. See SI Section Definitions of Terms and Metrics for definitions of original posts and reshares.

 https://doi.org/10.1038/s41562-026-02435-2

Extended Data Table 1 | Descriptions of CIB and selected FMO networks

ID	Network type	Name in common usage	Description
1	CIB	Roger Stone and associates	https://about.fb.com/news/2020/07/removing-political-coordinated-inauthentic-behavior/ (#4)
2	CIB	Romania	https://about.fb.com/news/2020/08/july-2020-cib-report/ (#1)
3	CIB	Truthmedia	https://about.fb.com/news/2020/08/july-2020-cib-report/ (#2)
4	CIB	Russia IRA ("PeaceData")	https://about.fb.com/news/2020/09/august-2020-cib-report/ (#1)
5	CIB	China ("Naval Gazing")	https://about.fb.com/news/2020/09/removing-coordinated-inauthentic-behavior-china-philippines/ (#1)
6	CIB	Russia IRA ("United World International")	https://about.fb.com/news/2020/09/removing-coordinated-inauthentic-behavior-russia/ (#2)
7	CIB	Russian intelligence ("Strategic Culture")	https://about.fb.com/news/2020/09/removing-coordinated-inauthentic-behavior-russia/ (#3)
8	CIB	Russian military / GRU	https://about.fb.com/news/2020/09/removing-coordinated-inauthentic-behavior-russia/ (#1)
9	CIB	Rally Forge	https://about.fb.com/wp-content/uploads/2020/11/October-2020-CIB-Report.pdf (#11)
10	CIB	Mexico / Venezuela	https://about.fb.com/news/2020/10/removing-coordinated-inauthentic-behavior-mexico-iran-myanmar/ (#1)
11	CIB	Iran Eurovision	https://about.fb.com/news/2020/10/removing-coordinated-inauthentic-behavior-mexico-iran-myanmar/ (#2)
12	CIB	Iran typo-squatting domains	https://about.fb.com/wp-content/uploads/2021/01/December-2020-CIB-Report-.pdf (#1)
13	CIB	Fazze	https://about.fb.com/news/2021/08/july-2021-coordinated-inauthentic-behavior-report/ (#2)
15	FMO		Taken down on July 28, 2020 and originated in Italy, consisted of 366 Facebook users, 10 groups and 4 Instagram users. The network posted links to ad-laden websites that featured articles on political topics copied from American outlets. The sources of the copied articles included outlets such as Patriot Journal, Fox News, Conservative Daily Brief and MAGA conservative.
17	FMO		Taken down on September 9, 2020 and originated in Armenia, consisted of 66 Facebook users, 57 Pages, 1 group and 9 Instagram users. The network primarily focused on Ukraine, but also targeted audiences in the US. The network posted links to ad-laden websites with English-language and non-English-language content, including civic content. The English-language content included articles copied from outlets such as Business Insider.
19	FMO		Taken down on September 30, 2020 and originated in North Macedonia, consisted of 7 Facebook users, 13 Pages and 11 groups. This network created fake accounts posing as Americans, as well as Pages and groups focused on US civic issues. They used these to post links to off-platform websites that featured articles on political topics copied from American outlets. The sources of the copied articles included outlets such as VOA News, Forbes, and Fox News.
27	FMO		Taken down on October 20, 2020 and originated in Kosovo, consisted of 6 Facebook users and 1 Page. The Page was a compromised comedy Page which, after being compromised, was used to post non-civic memes and links to off-platform websites that featured content copied from Fox News. User comments on these posts pointed out the Page had been hacked.
29	FMO		Taken down on October 20, 2020 and originated in Pakistan, consisted of 676 Facebook users. The network posted links to an off-platform website, which included copied articles from a number of other news sources.
34	FMO		Taken down on October 28 2020 and originated in Kosovo, consisted of 5 Facebook users, 6 Pages and 2 groups. This network used a fake account posing as an American to admin Pages and post links towards an ad-laden website that featured articles copied from American outlets. The sources of the copied articles included outlets such as America Now.
35	FMO		Taken down on October 29 2020 and originated in North Macedonia, consisted of 154 Facebook users and 12 groups. The network was a cluster of inauthentic accounts posting links to ad-laden websites that featured articles on political topics copied from American outlets. The sources of the copied articles included outlets such as the LA Post, The Washington Times, and The Federalist, among others.
36	FMO		Taken down on October 29 2020 and originated in North Macedonia, consisted of 45 Facebook users and 10 Pages. The network posted links to ad-laden websites which included articles copied from other news sources. The sources of the copied articles included outlets such as Fox News and the Daily Mail.

https://doi.org/10.1038/s41562-026-02435-2

Extended Data Table 2 | Network characteristics overview

ID	Network type	Modal takedown date	Attributed countries of origin	Top language used		Total posts		Total ads
				FB	IG	FB (original and network reshare)	IG (original)	FB & IG
1	CIB	2020-07-08	United States	English		731	0	
2	CIB	2020-07-31	Romania	English	English	3,843	317	
3	CIB	2020-07-31	United States, Canada, Australia, New Zealand, Vietnam, Taiwan, Hong Kong, Indonesia, Germany, United Kingdom, Finland, France	Chinese		15,273	0	251
4	CIB	2020-08-31	Russia	English		434	0	2
5	CIB	2020-09-18	China	English		1,940	0	
6	CIB	2020-09-24	Russia	English	English	346	696	1
7	CIB	2020-09-24	Russia	Russian	Russian	1,759	697	
8	CIB	2020-09-24	Russia	English	English	1,500	354	
9	CIB	2020-10-08	United States	English	English	6,977	911	513
10	CIB	2020-10-27	Mexico, Venezuela		English	0	6,646	
11	CIB	2020-10-27	Iran	English	Farsi, Persian	3	11	
12	CIB	2020-11-16	Iran	English		183	0	
13	CIB	2021-07-30	Russia	Spanish	Hindi	2,597	5,853	
14	FMO	2020-07-14	Philippines			0	0	
15	FMO	2020-07-28	Italy	English	English	928	27	
16	FMO	2020-08-18	Albania	English		92	0	
17	FMO	2020-09-09	Armenia	Russian	Armenian	5,042	9	
18	FMO	2020-09-18	Canada	English		12	0	
19	FMO	2020-09-30	North Macedonia	English		63	0	
20	FMO	2020-09-30	North Macedonia	English	English	119	105	
21	FMO	2020-10-02	North Macedonia	Macedonian		3	0	
22	FMO	2020-10-07	North Macedonia	English	English	2	21	6
23	FMO	2020-10-09	United States	English		272	0	5
24	FMO	2020-10-13	Canada	English		1,001	0	
25	FMO	2020-10-16	North Macedonia	English		144	0	
26	FMO	2020-10-19	Bosnia and Herzegovina	English		11	0	
27	FMO	2020-10-20	Kosovo	English		771	0	
28	FMO	2020-10-20	Russia	Russian	English	63	29	157
29	FMO	2020-10-20	Pakistan	English		9,382	0	
30	FMO	2020-10-21	Bangladesh	Bengali		1,744	0	
31	FMO	2020-10-21	Pakistan	English		64	0	
32	FMO	2020-10-23	Pakistan	English		18	0	
33	FMO	2020-10-27	North Macedonia		English	6	21	10
34	FMO	2020-10-28	Kosovo	English		6,199	0	189
35	FMO	2020-10-29	North Macedonia	English		4,883	0	
36	FMO	2020-10-29	North Macedonia	English		958	0	53
37	FMO	2020-10-29	Pakistan			2	0	
38	FMO	2020-11-03	Pakistan	English		52	0	
39	FMO	2020-11-03	North Macedonia	English		176	0	
40	FMO	2020-11-03	Romania	English	English	855	468	6
41	FMO	2020-11-03	Algeria	English		232	0	
42	FMO	2020-11-04	India	English		288	0	
43	FMO	2020-11-04	India	English	English	4,274	219	
44	FMO	2020-11-05	Pakistan	English		3,334	0	
45	FMO	2020-11-05	North Macedonia	Macedonian	English	49	338	
46	FMO	2020-11-05	North Macedonia	English	English	866	132	2
47	FMO	2020-11-09	Bangladesh	English		2,128	0	
48	FMO	2020-11-12	Morocco	French		1,938	0	
49	FMO	2021-01-06	Bangladesh	English		172	0	

FB refers to Facebook, IG to Instagram. Total posts in this table include original posts and reshares by network-affiliated accounts for Facebook, and only original posts by network-affiliated accounts for Instagram. Top language used refers to the most common language in direct network posts. See SI Section Definitions of Terms and Metrics for detailed definitions of original posts and reshares, see SI Section Network Country Attribution for more details on country attribution.

 https://doi.org/10.1038/s41562-026-02435-2

Extended Data Table 3 | Number of active users exposed to or engaged with, and active users' total exposure to and engagement with network content on Facebook

Network type	Viewers	Clickers	Resharers	Views	Clicks	Reshares
CIB	15,305,161	4,412,302	1,691,761	55,870,634	6,148,881	4,967,135
FMO	27,695,164	3,490,107	521,402	119,369,610	6,821,529	963,792
CIB & FMO	36,788,238	7,474,595	2,084,173	175,240,244	12,970,410	5,930,927

The total number of active users on Facebook was approximately 250 million. On Facebook, network content includes original posts and reshares by network-affiliated accounts, and reshares of network posts by accounts unaffiliated with the network. For reshares, only network original posts are considered. See SI Section Definitions of Terms and Metrics for definitions of active users, network content and engagement metrics.

https://doi.org/10.1038/s41562-026-02435-2

Extended Data Table 4 | Number of active users exposed to or engaged with, and active users' total exposure to and engagement with network content on Instagram

Network type	Viewers	Resharers	Views	Reshares
CIB	2,944,266	9,625	69,296,390	12,167
FMO	34,570	365	356,399	798
CIB & FMO	2,977,009	9,989	69,652,789	12,965

The total number of active users on Instagram was approximately 160 million. On Instagram, network content includes only original posts by network-affiliated accounts. See SI Section Definitions of Terms and Metrics for definitions of active users, network content and engagement metrics.

 https://doi.org/10.1038/s41562-026-02435-2

Extended Data Table 5 | Overview of active users' exposure to and engagement with network content on Facebook by network and source of exposure

				Average per viewer exposed directly			Average per viewer exposed indirectly		
ID	Network type	Viewers Exposed Directly	Viewers Exposed Indirectly	Views	Clicks	Reshares	Views	Clicks	Reshares
1	CIB	72,745	180,338	7.08787	0.32979	0.09855	1.68895	0.11705	0.00466
2	CIB	18,756	284,866	15.73525	1.03620	0.22116	1.34877	0.05933	0.01340
3	CIB	1,588,943	744,785	4.83241	0.27670	0.03425	2.45609	0.13574	0.01199
4	CIB	4,099	12,643	1.88851	0.04847	0.02254	1.48675	0.02950	0.01736
5	CIB	77,501	91,908	2.91894	0.18087	0.01577	2.51562	0.20265	0.00793
6	CIB	224	1,374	25.95111	0.60889	0.03111	1.68873	0.08000	0.00364
7	CIB	9,447	21,640	13.74383	0.71074	0.11144	1.80995	0.09056	0.01229
8	CIB	945	1,125	7.75052	0.13872	0.04451	2.50044	0.10168	0.01061
9	CIB	1,271,954	13,009,543	7.20644	0.18120	1.83283	2.30651	0.38822	0.09827
10	CIB								
11	CIB	2		1.50000	0.00000	0.00000			
12	CIB	2,450	267	1.84628	0.04514	0.00041	1.24719	0.01498	0.00000
13	CIB	3,076	6	3.98669	0.07530	0.00032	1.50000	0.16667	0.00000
14	FMO	0	33	0.00000	0.00000	0.33333	1.25000	0.13889	0.00000
15	FMO	587,541	2,964,874	12.03691	1.38323	0.21217	1.76519	0.13783	0.01367
16	FMO	21,803	6,615	4.17788	0.03361	0.00004	1.42089	0.03385	0.00000
17	FMO	92,487	102,963	4.15291	0.20608	0.02062	2.10093	0.04713	0.00711
18	FMO	416	13	5.31100	0.46411	0.00239	0.86667	0.00000	0.00000
19	FMO	61,262	427,501	2.72657	0.12944	0.05547	1.41164	0.07347	0.00804
20	FMO	249,328	64,379	2.08595	0.09587	0.00519	1.20600	0.05914	0.00367
21	FMO	7		1.57143	0.00000	0.00000			
22	FMO	135	1,738	1.49701	0.01198	0.20958	1.31764	0.06693	0.00794
23	FMO	12,146	215,593	24.20606	1.55281	0.68200	1.88314	0.05518	0.01914
24	FMO	275,361	208,397	4.17743	0.51007	0.03726	1.51061	0.09606	0.00560
25	FMO	32,121	967	28.61297	2.34971	0.00208	1.72934	0.01963	0.00310
26	FMO	103	8	2.89524	0.02857	0.00952	2.37500	0.12500	0.00000
27	FMO	3,637,708	7,654,095	7.38086	0.32910	0.02582	1.72231	0.11503	0.00570
28	FMO	220	20,962	0.36970	0.00833	0.84545	1.22826	0.02652	0.00095
29	FMO	2,262,807	192,216	2.93323	0.27135	0.00274	1.41893	0.10428	0.00447
30	FMO	110,563	594	1.50944	0.08274	0.00023	1.27227	0.03361	0.00168
31	FMO	54,253	1,038	1.62643	0.15493	0.00088	1.32757	0.08934	0.00384
32	FMO	137	2,881	1.34211	0.00000	0.05263	1.56940	0.06500	0.02268
33	FMO	1,213	9,035	1.27814	0.00000	0.17801	1.29600	0.05622	0.00562
34	FMO	467,164	9,398,804	36.03519	1.87329	0.43252	1.92434	0.04968	0.01767
35	FMO	439,005	4,676,342	5.97054	0.47190	0.08050	1.70595	0.09335	0.01830
36	FMO	88,040	1,525,984	7.51170	0.23262	0.08663	1.50238	0.06804	0.01343
37	FMO	1		2.00000	0.00000	0.00000			
38	FMO	2,716	306	1.57842	0.16115	0.02626	1.31046	0.04902	0.00000
39	FMO	5,871	222	1.61804	0.08617	0.00236	1.13216	0.02643	0.00000
40	FMO	11,848	23,728	4.19404	0.23215	0.16860	1.81211	0.07485	0.01359
41	FMO	350,103	11,174	2.79629	0.08266	0.00192	1.69848	0.06718	0.00544
42	FMO	140,644	232,368	3.69991	0.44543	0.05501	1.54891	0.11805	0.00845
43	FMO	1,028	4,779	8.02027	0.14274	0.08108	1.31686	0.04044	0.00313
44	FMO	107,762	545,391	9.48022	0.41697	0.17170	2.16713	0.21066	0.03135
45	FMO	299	32	1.71572	0.07023	0.00669	1.56250	0.03125	0.00000
46	FMO	11,570	2,325	1.81610	0.04546	0.00198	1.41116	0.14592	0.00773
47	FMO	26,217	10,822	5.25330	0.09454	0.01343	1.46230	0.04379	0.00277
48	FMO	125,709	75,859	3.13568	0.26370	0.01219	1.57539	0.07786	0.00403
49	FMO	2,744	2,462	5.79058	0.03924	0.02105	1.31671	0.02190	0.00041

Different rates of engagement by those exposed directly or indirectly may be due to different engagement on the same content, or due to different content that is spread directly or indirectly and therefore receives different engagement. See SI Section Definitions of Terms and Metrics for definitions of network content, direct and indirect exposure, and engagement metrics.

nature portfolio

Corresponding author(s): Jennifer Pan

Last updated by author(s): July 23, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

The study's data collection involved a partnership between Meta and NORC, a survey research organization at the University of Chicago. While Meta handled the majority of participant recruitment and gathered information from Facebook and Instagram platforms, NORC managed the survey administration process, including expanding the participant pool through additional recruitment, gathering supplementary information beyond the social media platform data, and ensuring participant anonymity by stripping identifying information before data integration and distribution to researchers. Participants provided data through their personal devices, including smartphones and computers. NORC operated without knowledge of the research hypotheses or participants' experimental assignments.

On-platform behavioral data and network data were collected via Meta's internal systems for logging user behavior. Survey data were collected by NORC using their existing survey infrastructure. To collect the passive measurement data, NORC partnered with two vendors: MDI Global and RealityMine.

Aggregated platform data: For Facebook, aggregated data are based on computing the total count for the study period (2020-06-26 until 2021-02-15). For Instagram, metrics are aggregated between 2020-07-17 and 2021-02-15 due to an unanticipated data gap.

Individual-level survey data: Survey data collection started on August 31, 2020. Two surveys were fielded pre-treatment: Wave 1 (August 31-September 12) and Wave 2 (September 8-23). The treatments that we analyze in more detail (less like-minded content, no reshares, reverse-chronological feed) ran from September 24-December 23, other treatments had different timelines (see SI 3.1). During the treatment period, three more surveys were administered: Wave 3 (October 9-23), Wave 4 (November 4-18), and Wave 5 (December 9-23). See SI S2.3 for details. See SI S2.3 for details on the survey and platform behavioral data for individual-level participants and data collection period details.

Network data: Network data are measured for the study period (2020-06-26 until 2021-02-15). For more details, see SI S2.1.

Data analysis	The data in this study were analyzed using R (version 4.3.3), which was executed on via R notebooks on JupyterLab (4.2.4). The analysis code imports several R packages available on CRAN, including tidyverse (2.0.0), dplyr (1.1.4), ggplot2 (3.5.2), estimatr (1.0.4), Hmisc (5.1-0), sf (1.0-16), ggrepel (0.9.6), patchwork (1.2.0), scales (1.3.0), RColorBrewer (1.1-3), lubridate (1.9.4), stringr (1.5.1) and hbal (1.2.12).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

De-identified data from this study is stored in the Social Media Archive (SOMAR) housed by ICPSR. This is accessible for university research approved by the Institutional Review Board (IRB) related to elections, or for the purpose of validating the results of this study. ICPSR will handle and vet all applications requesting access to this data. Data access is controlled to protect the privacy of the study participants and to be consistent with the consent form signed by study participants. Requests for data can be made via SOMAR (https://socialmediaarchive.org/). The data for this paper is based on the following datasets stored in SOMAR: https://socialmediaarchive.org/record/87, https://socialmediaarchive.org/record/88, https://socialmediaarchive.org/record/89, https://socialmediaarchive.org/record/90, https://socialmediaarchive.org/record/91, https://socialmediaarchive.org/record/92, https://socialmediaarchive.org/record/93, https://socialmediaarchive.org/record/94, https://socialmediaarchive.org/record/95, https://socialmediaarchive.org/record/96, https://socialmediaarchive.org/record/97, https://socialmediaarchive.org/record/98, https://socialmediaarchive.org/record/99, https://socialmediaarchive.org/record/100, https://socialmediaarchive.org/record/101, https://socialmediaarchive.org/record/102, https://socialmediaarchive.org/record/103..
In addition, we used public data sources for generating map plots and cleaning emojis in hashtags. In particular, for the world map in the main text, the ggplot2 package map_data function relies on world map data from \url{https://www.naturalearthdata.com}. We also downloaded centroid data from \url{https://raw.githubusercontent.com/gavinr/world-countries-centroids/master/dist/countries.csv}. For the Congressional district maps in the supplementary repository, we used a Congressional district shapefile from \url{https://www2.census.gov/geo/tiger/GENZ2020/shp/cb_2020_us_cd116_20m.zip} and Congressional district code file from \url{https://www2.census.gov/geo/docs/reference/codes/files/national_cd113.txt}. To clean emojis in hashtags, we used Unicode data to filer out emojis. The raw dictionary was downloaded from \url{http://www.unicode.org/Public/emoji/1.0//emoji-data.txt}. Emoji lists were downloaded from \url{http://www.unicode.org/reports/tr51/index.html#emoji_data}. For a comparison to the voting age population, we gathered the voting age population size from \url{https://www.census.gov/content/dam/Census/library/publications/2022/demo/p20-585.pdf}.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	We confirm that we do not use the terms gender or sex in the main text. Several analyses in the SI employ gender, which is measured via survey self-report (male, female, and other) or derived from the stated gender on the user profile. Gender was determined based on survey or use profile self-reports. Informed consent was provided prior to collecting survey data
Reporting on race, ethnicity, or other socially relevant groupings	We used survey self-reports to collect data on race and ethnicity (white non-hispanic, black non-hispanic, hispanic, asian-american and pacific islander, multi-racial. We collected the data for subgroup analyses, where we also considered a variety of other subgroups. We also looked at census data to classify congressional districts as majority non-hispanic/non-latino white or not for geographic subgroup analyses.
Population characteristics	Aggregated platform data refers to data on all adult U.S.-based Facebook and Instagram users who were active for at least one day between June 26, 2020 and February 15, 2021 (the study period). The size of this sample is about 250 million adults for Facebook, and about 160 million adults for Instagram. For Facebook, about 46.8% of users are male, 53.2% are female; 26.7% are aged 18-29, 30.7% 30-44, 29.2% 45-64, and 13.3% 65 plus. For Instagram, about 42.4% of users are male and 57.6% are female. This sample is representative of the adult active Facebook and Instagram US user population, but not the general US population. Individual-level participant data has a total sample size is about 73,000 consenting participants. Of these users, 43.3% are male, 56.7% are female or another gender (after excluding missing values); 17.9% are aged 18-29, 41.8% 30-44, 33.3% 45-65, and 7% 65 plus (after excluding missing values). This sample is neither representative of the adult active Facebook and Instagram US user population, nor the general US population. The aggregated platform data was selected to capture the aggregated characteristics of adult active Facebook and Instagram US user populations, which is the focus of this study. The individual-level participants data augments this analysis to enable individual-level estimates.
Recruitment	Aggregated platform data is based on adult U.S.-based Facebook and Instagram users who were active for at least one day between June 26, 2020 and February 15, 2021 (the study period). Individual-level data is based on survey participants who saw a recruitment message on Facebook or Instagram asking them if they would like to share their opinion. Those clicking "Start Survey" were directed to a consent form. Participants gave their consent to participate using an IRB-approved consent form that outlined the study procedure, benefits and risks, and compensation.
Ethics oversight	The research design prioritized ethical considerations, with attention to minimizing risks both for individual respondents and for society more broadly. Meta applied for and received approval from the NORC Institutional Review Board to conduct the experimental studies (Protocol number 20.08.10, Project number 8870). Each academic collaborator also coordinated with

their home institution's IRB to meet Human Subjects Research requirements for analyzing the aggregated, de-identified data that Meta and NORC collected. The same protocol was submitted to 13 academic institutions, nine of them determined that the research was not human subjects, not research, or irrelevant for the university IRB (i.e., the university was not engaged in the research). Of the remaining four, three universities determined that the protocol was exempt (University of Texas Austin FWA\# 00002030, William and Mary IRB\# PHSC-2020-09-21-14523-jsettle, George Washington University IRB\# NCR213250), and one university conducted expedited review (University of Pennsylvania IRB\# 844651). Additionally, Meta engaged \href{https://ethicalresolve.com/}{Ethical Resolve}, a data ethics consultancy, to advise both company and academic team members before the research was implemented. This external review assessed the project against established principles of research ethics and evolving standards for digital research.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description

The study combines multiple types of quanitative data, including aggregated platform data, individual-level survey data, and network data.

Research sample

Aggregated platform data:
Aggregated platform data includes data on adult U.S.-based Facebook and Instagram users who were active for at least one day between June 26, 2020 and February 15, 2021 (the study period). The size of this sample is about 250 million adults. For more details, see SI Section S2.2.

Individual-level survey data:
Individual-level participant data refers to data from U.S. Facebook and Instagram users age 18 and over who consented to participate in a study of social media and politics and completed two baseline survey waves. The data include individual-level survey responses as well as participants' on-platform behavioral data. The sampling frames included all Facebook and Instagram monthly active U.S.-based users 18 years of age or older eligible to receive general surveys on a given platform (these represent a random set of users from the overall Facebook and Instagram populations) as of August 17, 2020. Participants were asked to confirm they were over 18 years of age and lived in the United States as part of the recruitment process. For more details, see SI S2.3. We excluded participants who deleted their account or requested data deletion from the analysis for ethical reasons and since we could not access their data.

Network data:
Deceptive online networks are operationalized as networks disabled under the Coordinated Inauthentic Behavior (CIB) protocol, which consisted of influence operations, whether domestic or foreign in origin, that attempted to manipulate or corrupt public debate for a strategic goal. Network also include those disabled under the Inauthentic Behavior (IB) protocol. These consisted of financially motivated operations (FMO) that were identified by an informal, internal Meta taskforce as networks using inauthentic accounts to post content on American political themes—usually to drive users towards off-platform websites where the operators could monetize users' clicks. For more details on data related to deceptive online networks, including how networks were identified, how these data were prepared, missing data, and more detailed descriptions of networks, see SI S2.1. Note that our operationalization of network content excludes network comments, which some networks like Rally Forge commonly used, implying that our numbers on content production and reach are underestimates. Further, due to lower performance for other languages, we focus on posts in English and Spanish for topic classifications, meaning that we do not take posts in other languages into account when assessing the share of political content or topics about which a network posted.

Sampling strategy

Aggregated platform data:
This represents the population of US-based adult active users on Facebook and Instagram. For more details, see SI S2.2.

Individual-level survey data:
Below, we summarize our sampling strategy. Further details are available in SI S2.3.1. Participants were sampled to achieve specific sample targets across different stages of the study, which were in turn chosen to achieve desired minimum detectable effect sizes (MDEs) for experimental interventions related to other papers part of the US 2020 FIES across different subgroups among the set of respondents participating in the Wave 1 and Wave 2 surveys and Waves 4 or 5. The sampling frame consisted of all U.S. Facebook and Instagram users aged 18 and above who were active monthly users and qualified to participate in standard platform surveys as of August 17, 2020. This participant pool constituted a randomly selected subset of the broader user bases across both social media platforms. The Facebook sampling frame was trimmed by removing predicted fake accounts, employees, and advertisers. For Instagram, creators and business accounts also were removed. Finally, because the use of multiple accounts is common among Instagram users, the sampling frame was narrowed to include only a user's primary account (for users with multiple accounts, this is the oldest account). Having defined the sampling frames, sampling probabilities were computed to achieve specific sample distributions for the set of demographics encoded in the stratification step across each of the samples of interest. The sampling probabilities took into account (a) differential non-response across different demographics and (b) the desired sample size across the different studies. At the top of their Facebook feed, randomly selected participants saw a recruitment message asking them if they would like to share their opinion. Those clicking ``Start Survey'' were directed to a consent form. Participants gave their consent to participate using an IRB-approved consent form that outlined the study procedure, benefits and risks, and compensation.

	During recruitment, individuals were required to verify their age (18+) and U.S. residency status. The sample size is N = 72,962, with N = 19,510 participants randomized into the control group for the FIES platform interventions. Additional information regarding sampling methodology, recruitment, and response rates can be found in SI S2.3. Data collection and analysis were not performed blind to the conditions of the experiments. Network data: Deceptive online networks are operationalized as networks Facebook (now Meta) disabled between June 26, 2020 and November 3, 2021 (one year after election day) that targeted US users during the study period (June 26, 2020 and February 15, 2021). This resulted in 49 networks. For more details, including on identification of networks, see SI S2.1. For one analysis on the downstream effects of networks, we subset to networks that reached a pre-registered reach threshold between different survey waves, see SI S3.1.1 for details.
Data collection	The study's data collection involved a partnership between Meta and NORC, a survey research organization at the University of Chicago. While Meta handled the majority of participant recruitment and gathered information from Facebook and Instagram platforms, NORC managed the survey administration process, including expanding the participant pool through additional recruitment, gathering supplementary information beyond the social media platform data, and ensuring participant anonymity by stripping identifying information before data integration and distribution to researchers. Participants provided data through their personal devices, including smartphones and computers. NORC operated without knowledge of the research hypotheses or participants' experimental assignments. On-platform behavioral data and network data were collected via Meta's internal systems for logging user behavior. Survey data were collected by NORC using their existing survey infrastructure. To collect the passive measurement data, NORC partnered with two vendors: MDI Global and RealityMine.
Timing	Aggregated platform data: For Facebook, aggregated data are based on computing the total count for the study period (2020-06-26 until 2021-02-15). For Instagram, metrics are aggregated between 2020-07-17 and 2021-02-15 due to an unanticipated data gap. Individual-level survey data: Survey data collection started on August 31, 2020. Two surveys were fielded pre-treatment: Wave 1 (August 31-September 12) and Wave 2 (September 8-23). The treatments that we analyze in more detail (less like-minded content, no reshares, reverse-chronological feed) ran from September 24-December 23, other treatments had different timelines (see SI 3.1). During the treatment period, three more surveys were administered: Wave 3 (October 9-23), Wave 4 (November 4-18), and Wave 5 (December 9-23). See SI S2.3 for details. See SI S2.3 for details on the survey and platform behavioral data for individual-level participants and data collection period details. Network data: Network data are measured for the study period (2020-06-26 until 2021-02-15). For more details, see SI S2.1.
Data exclusions	For individual-level data, we excluded participants who deleted their account, withdrew from the study, or requested data deletion from the analysis for ethical reasons since we could not access their data.
Non-participation	We cannot distinguish numbers of participants whose data was excluded for different reasons because this data was not retained. For individual-level data, there were 180k people who consented to platform interventions on Facebook, 139k of whom completed they wave 1 survey. Of the 180k, there were a total of 439 people who withdrew. Of the 139k, 329 people withdrew. There were 123k people who consented to platform interventions on Instagram, 92k of whom completed wave 1. Of the 123k, there were a total of 275 people who withdrew. Of the 92k, 198 withdrew.
Randomization	For the analysis of experimental interventions using individual-level survey data, respondents were randomly assigned to treatment or control with probabilities that maximized statistical power using block randomization. A combination of survey-based pre-treatment outcomes and Facebook data were used to define the blocks in the sample of interest.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*