Social Media vs. Democracy: Evidence from the January 6th Insurrection*

Karsten Müller† Carlo Schwarz‡ Zekai Shen§

February 11, 2026

Abstract

Social media platforms are often credited with empowering grassroots movements in the pursuit of political freedoms. In this paper, we show how social media can also be exploited by political elites to undermine democratic institutions, using the January 6th, 2021 Capitol insurrection as a case study. We present three main findings. First, by exploiting plausibly exogenous variation in Twitter usage, we document that social media exposure predicts participation in the Capitol attack, donations for anti-democratic causes, beliefs in election fraud, and support for the January 6th rioters. Second, Donald Trump's tweets questioning the election's integrity were followed by spikes in "Stop the Steal" activity on Twitter and pro-Trump donations originating from high Twitter usage counties. Third, the insurrection and Trump's account deletion were followed by a decrease in the public expression of toxic political and "Stop the Steal" messaging by pro-Trump users on Twitter, but had little effect on privately held beliefs about the election outcome and pro-Trump donations.

Keywords: Social Media, January 6th, Election Denial, Content Moderation
JEL Codes: L82, J15, O33.

*We are grateful to seminar participants at Bocconi, the SSRC, the CEPR Monash-Paris-Warwick-Zurich Text-as-Data workshop, the Hertie School of Governance, and the DAISI workshop for their valuable comments. We also thank Pablo Barbera for sharing the list of Twitter user ID data with us. Giovanni Moscariello provided outstanding research assistance. Müller acknowledges support from a Singapore MoE Start-up/PYP grant (number A-0003319-01-00). This paper incorporates and supersedes an earlier working paper by Karsten Müller and Carlo Schwarz entitled "The Effects of Online Content Moderation: Evidence from President Trump's Account Deletion."

†National University of Singapore, NUS Business School and Risk Management Institute, and CEPR, kmueller@nus.edu.sg.

‡Università Bocconi, Department of Economics and IGIER, and PERICLES, CEPR, CAGE, carlo.schwarz@unibocconi.it.

§Università Bocconi, Department of Economics, zekai.shen@phd.unibocconi.it

1 Introduction

Social media platforms have reshaped politics and collective action by lowering coordination costs, facilitating real-time information sharing, and creating new channels for political expression. Early commentary portrayed social media as a "liberation technology" that could enable dissent and expand political, social, and economic freedoms Diamond (2010), in part because its decentralized information flow limits the ability of authoritarian regimes to control it (Edmond, 2013). Consistent with this view, a growing body of empirical research finds that social media can catalyze collective action, including during the Arab Spring (Acemoglu, Hassan and Tahoun, 2017), in Russia (Enikolopov, Makarin and Petrova, 2020*a*), in China (Qin, Strömberg and Wu, 2017; Qin, Strömberg and Wu, 2024), and across several African countries (Manacorda and Tesei, 2016).

Yet a growing body of recent research challenges the intrinsically pro-democratic view of social media. Social media has been linked to political polarization (e.g., Allcott, Braghieri, Eichmeyer and Gentzkow, 2020; Levy, 2021), misinformation (e.g., González-Bailón et al., 2023; Allcott et al., 2024), and politically-motivated violence (e.g., Müller and Schwarz, 2021, 2023; Bursztyn, Egorov, Enikolopov and Petrova, 2019), and such concerns have led many governments to implement stronger oversight of digital platforms (e.g., The Wall Street Journal, 2022, 2023; The Guardian, 2023). In light of these findings, there is a need for empirical evidence on whether the decentralized nature of social media inherently makes it a force for democracy or whether it can also be co-opted by powerful leaders to undermine democratic institutions.

In this paper, we investigate this question by examining the case of the "Stop the Steal" campaign led by Donald Trump following the 2020 U.S. presidential election, which ultimately culminated in the January 6$^{\text{th}}$, 2021 attack on the U.S. Capitol. Leveraging county-level data on riot participation and political donations, survey evidence, and a unique dataset on Twitter activity, we provide causal evidence that Trump's social media messaging fueled anti-democratic sentiment online and spurred costly and persistent real-world political actions.

Our analysis proceeds in three parts. First, we examine the relationship between social media exposure, anti-democratic political mobilization, and distrust in the 2020 election results. To estimate causal effects, we use an instrumental variables strategy that generates plausibly exogenous variation in Twitter usage. Second, we study the impact of Trump's election-denial messaging on both online behavior and offline political actions. Third, we exploit Twitter's unprecedented suspension of Trump's account, along with the broader platform-wide crackdown on election-denial content following the January 6$^{\text{th}}$ insurrection, to assess whether such drastic measures were associated with reductions in both public and private support for the "Stop the Steal" movement.

In the first part of our analysis, we exploit plausibly exogenous variation in Twitter adoption generated by the 2007 South by Southwest (SXSW) festival (Müller and Schwarz, 2023; Fujiwara, Müller and Schwarz, 2024), a pivotal moment in Twitter's diffusion in the United States, to identify the causal effect of social media exposure on political mobilization. We find that a 10% increase in county-level Twitter usage increased participation in the January 6^{th} attack by 0.31 individuals per county (a 65% increase relative to the mean), even after controlling for a rich set of covariates, including the 2020 presidential vote. Greater Twitter exposure also boosted donations to the *Save America PAC*, which was established to support Trump's efforts to overturn the election. To assess the validity of our identification strategy, we implement placebo tests using other major U.S. festivals and find no comparable effects on political outcomes. We further document the importance of network linkages: January 6^{th} participants were disproportionately drawn from counties that were more connected to counties with other January 6^{th} participants via Twitter, based on an empirical strategy similar to Qin, Strömberg and Wu (2024). Finally, using individual-level survey data, we show that social media usage is associated with stronger beliefs in election fraud and greater approval of the January 6^{th} rioters, consistent with a mechanism through which online misinformation translates into offline mobilization.

In the second part of our analysis, we examine whether Trump's allegations of election fraud amplified similar narratives on Twitter and mobilized offline financial support. Using high-frequency event studies, we show that Trump's election-denial tweets were followed by a doubling in tweets containing the "Stop the Steal" hashtag. We further document that the county-level volume of "Stop the Steal" tweets strongly predicts both participation in the January 6^{th} attack and donations to the Save America PAC. Building on this evidence, we show that donations to the *Save America PAC* increased disproportionately in counties with higher Twitter-usage in direct response to Trump's election-denial tweets. This effect remains robust when instrumenting for county-level Twitter usage with the SXSW instrument. Taken together, these results suggest that political elites can leverage their social media following to disseminate misinformation and mobilize efforts to undermine democratic institutions.

In the third part of our analysis, we examine the effectiveness of countermeasures to limit the spread of anti-democratic rhetoric online and reduce private support for anti-democratic causes. We focus on the suspension of Trump's Twitter account on January 8^{th}, 2021, which was accompanied by a broader platform-wide crackdown on election-denial content. Using a difference-in-differences design that exploits variation in whether users followed Trump prior to his account suspension, we show that both the number of tweets containing the "Stop the Steal" hashtag and the toxicity of political tweets declined significantly. Consistent with network spillovers, we also observe declines among users who did not follow Trump directly but followed accounts that did. Importantly, however, these changes in public expression did not translate into corresponding shifts in private beliefs or financial support. Social media users continued to exhibit elevated levels of election denial, and counties with exogenously higher Twitter usage continued to donate more to the *Save America PAC*. We interpret these findings as evidence that the impact of Trump's anti-democratic campaign persisted even after such sentiments became less visible online, whether due to self-censorship or platform content removal.

Taken together, our findings indicate that although social media can amplify the voices of individuals who are typically marginalized in the political process, it can also be appropriated by illiberal actors within democratic systems to mobilize anti-democratic behavior. The long-lasting nature of these effects remained detectable in privately-held beliefs despite a substantive crackdown online. Put differently, a successful anti-democratic campaign on social media can undermine political norms and sow the seeds of longer-lasting political upheaval even when initially suppressed through extensive content moderation.

Related Literature Our findings contribute to three strands of research. First, we contribute to the literature on social media and collective action. A large body of evidence shows that social media platforms and mobile internet access can facilitate protest participation by lowering coordination costs and enabling rapid information diffusion (e.g., Steinert-Threlkeld, Mocanu, Vespignani and Fowler, 2015; Manacorda and Tesei, 2016; Acemoglu, Hassan and Tahoun, 2017; Christensen and Garfias, 2018; Enikolopov, Makarin and Petrova, 2020*a*; Fergusson and Molina, 2021; Auxier, Anderson et al., 2021; Qin, Strömberg and Wu, 2024; Gylfason, 2023). See Zhuravskaya, Petrova and Enikolopov (2020) for a review.

This existing work has largely emphasized pro-democratic mobilization, particularly in authoritarian contexts, where the idea is that digital platforms can weaken regime control and empower dissidents (Edmond, 2013; King, Pan and Roberts, 2013; Lorentzen, 2014; Chen and Yang, 2019; Boxell and Steinert-Threlkeld, 2022). In this view, social media is inherently a "liberation technology," because its decentralized structure constrains authoritarian repression. In contrast, we show that social media can also facilitate political violence and anti-democratic mobilization within liberal democracies, a theoretical possibility raised by Tucker, Theocharis, Roberts and Barbera (2017), but for which causal empirical evidence has been limited.

Related work documents that digital connectivity can increase political violence in some contexts (Pierskalla and Hollenbach, 2013). We go one step further in demonstrating how a powerful political actor—in this case, the U.S. president—can leverage social media to weaken democratic institutions. Our results suggest that social media may be far less decentralized in practice than previously assumed, given the extraordinary reach of (political) elites.

Second, we contribute to research on the effect of the media on political violence and anti-democratic attitudes. Existing work shows that radio propaganda contributed to the Rwandan genocide (Yanagizawa-Drott, 2014), fascist violence in Italy (Gagliarducci, Onorato, Sobbrio and Tabellini, 2020), lynchings and race riots in the United States (Ang, 2023), and political sympathies during the Nazi era (Adena, Enikolopov, Petrova, Santarosa and Zhuravskaya, 2015; Wang, 2021*a*). Related research also highlights radio's role in mobilizing Civil Rights activism (Wang, 2021*b*; Bernini, 2023). Our paper focuses on a fundamentally different technology, social media, that allows for unmediated, interactive communication between political elites and the public. We show how this direct channel enabled Donald Trump to amplify misinformation, shape online narratives, and mobilize costly offline actions.

Third, we contribute to the literature on elite influence and online behavior. Recent work shows that Trump's tweets increased hate crimes (Müller and Schwarz, 2023; Cao, Lindo and Zhong, 2023) and that tweets by Trump and Members of Congress affect economic expectations and financial markets (Bianchi, Gómez-Cram, Kind and Kung, 2023; Bianchi, Gómez-Cram and Kung, 2024). We extend this literature by linking the messaging of political elites to participation in anti-democratic offline actions.

We also complement studies on content moderation and digital governance. Prior work has examined platform-wide moderation policies (Jiménez-Durán, 2023; Horta Ribeiro, Cheng and West, 2023; Beknazar-Yuzbashev, Jiménez-Durán, McCrosky and Stalinski, 2025) and government regulation of harmful content (Andres and Slivko, 2021; Jiménez Durán, Müller and Schwarz, 2022). We provide evidence on the effect of removing a single influential account, Twitter's suspension of Trump, on various outcomes. The study of direct and indirect effects on Trump's followers is particularly unique because it requires a dataset on who followed Trump before the account deletion, which we collected at the time. Related works by Buntain, Innes, Mitts and Shapiro (2023) and McCabe, Ferrari, Green, Lazer and Esterling (2024) studies changes in online content around Trump's account deletion. In contrast to these works, our analysis also sheds light on the arguably more important offline effects of the account removal. In particular, we demonstrate that the online effects do not readily translate into changes in offline behaviour. Our findings thus inform ongoing policy debates about targeted moderation approaches (Thomas and Wahedi, 2023; Aridor, Jiménez-Durán, Levy and Song, 2024), suggesting that limiting the reach of high-influence accounts may have some of the desired effects on platform discourse, but limited effects on people's privately-held beliefs.

2 Data

Our analysis draws on three primary data sources. First, we assemble county-level information on (a) Twitter usage, (b) participation in the January 6$^{\text{th}}$ insurrection, (c) donations to the pro-Trump *Save America PAC*, and (d) a wide range of socioeconomic and political characteristics. Second, we use individual-level survey data from Nationscape, which provides information on self-reported social media usage and detailed political attitudes, including questions explicitly related to the Capitol attack. Third, we analyze Twitter data comprising the universe of Donald Trump's tweets, which we complement with approximately 400 million tweets from a representative sample of U.S. users, along with information on which of these users followed Trump's account prior to its suspension in early 2021. We describe each data source in detail below. Additional information and summary statistics are provided in Appendix A.

2.1 County-Level Data

Capitol Riot Participants We measure participation in the January 6, 2021 Capitol attack using NPR's interactive database of 1,575 individuals who were federally charged for their involvement.[1] While the precise number of insurrectionists remains unknown, these data likely cover the vast majority of individuals who violently entered the U.S. Capitol, given the considerable efforts from law enforcement in the prosecution of the participants.

For each defendant, the database records the legal name, arrest date, and place of residence (city and state), compiled from federal court records and Department of Justice announcements. We assign defendants to US counties based on their reported place of living.[2] We then use the number of court cases in each county to measure participation. Figure 1a visualizes the widely dispersed geographic distribution of January 6$^{\text{th}}$ cases.

Save America Donations As a measure of support for Trump's efforts to challenge the 2020 election results, we obtain individual contribution data to the *Save America Leadership PAC* (FEC Committee C00762591) from the Federal Election Commission. These data cover the period from November 2020 to September 2024. The *Save America PAC* was established on November 9, 2020 specifically to support Donald Trump's political activities following the 2020 presidential election, including efforts to challenge election results and promote *Stop the Steal* messaging. In total, the *Save America PAC* collected $114.8 million in donations from

[1] These data are available on NPR's website.

[2] For a small minority of cases (98 cases), we only know in which state the individual was living at the time of arrest, and we thus cannot assign a county.

175,768 individual donors. We assign contributions to the county level based on the donor's address as recorded in the FEC filings. As a leadership PAC, Save America faced a $5,000 cap on individual contributions, making it a natural fundraising channel for individual donors. We therefore interpret contributions to *Save America PAC* as a costly revealed expression of support for the *Stop the Steal* movement. Figure 1b shows there is a considerable regional variation in *Save America PAC* donations.

Regional Twitter Usage To measure exposure to social media, we use the county-level number of Twitter users from Müller and Schwarz (2023) and Fujiwara, Müller and Schwarz (2024). The data is based on geocoded data from approximately 475 million tweets. These data have been shown to provide a reliable measure of county-level Twitter usage in the United States and have been validated against several alternative measures of Twitter usage, including, among others, the Gfk Media Survey.

South by Southwest Festival To study the causal effect of social media exposure, we closely follow the identification strategy first developed by Müller and Schwarz (2023) that exploits plausibly exogenous variation in Twitter penetration stemming from early adoption of the platform at the South by Southwest (SXSW) festival.[3] The instrumental variable is the number of SXSW followers per county who joined Twitter in March 2007, capturing the festival's immediate effect on early adoption. To control for pre-existing interest in both Twitter and SXSW, we include the number of SXSW followers who joined Twitter before the March 2007 festival. This control also allows us to address concerns about other omitted county-level characteristics. We discuss the motivation and assumptions underlying this identification strategy in Section 4.1. Figure 1c plots the geographic distribution of SXSW followers who joined Twitter in/before March 2007.

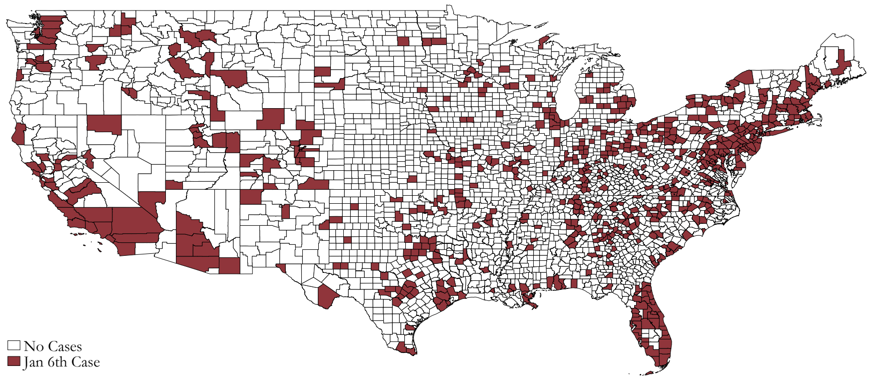
Additional County Characteristics We incorporate a comprehensive set of county-level control variables following the specification used in Fujiwara, Müller and Schwarz (2024), sourced from the U.S. Census, American Community Survey, and Simply Analytics.[4] These controls account for demographic, economic, geographic, and media consumption patterns that may correlate with both Twitter usage and political behavior.

[3]The same identification strategy has been used by, among others Fujiwara, Müller and Schwarz (2024), Müller, Pan and Schwarz (2023), Boken, Draca, Mastrorocco and Ornaghi (2023), and Gylfason (2023) to study the effects of social media on other outcomes.

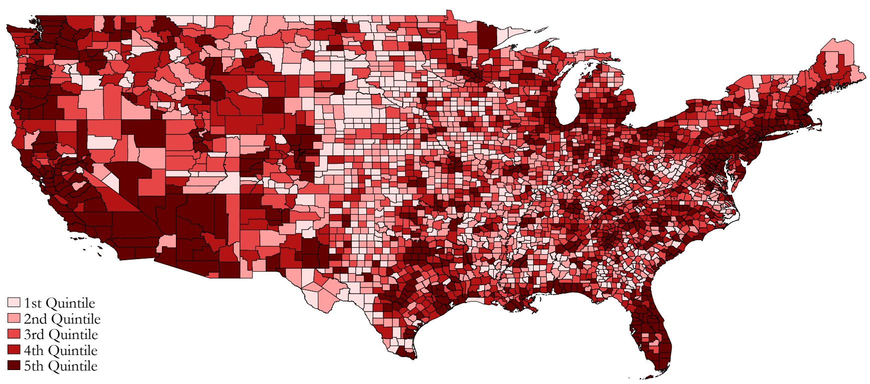
[4]The data are available as part of their replication file.

Figure 1: Geographic Variation of Main Variables

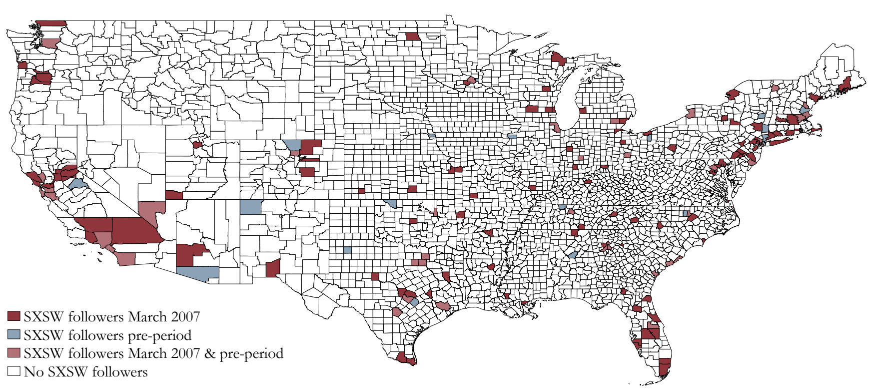
(a) Jan 6$^{\text{th}}$ Participation



(b) Save America Donations



(c) Identifying Variation



Notes: These figures show the geographic distribution of key variables across counties. Panel A shows participation in the January 6$^{\text{th}}$ Capitol insurrection, as measured by court cases. Panel B plots quintiles of donations to the pro-Trump *Save America PAC*, where the donation amounts represent county-level aggregates from November 2020 to January 6$^{\text{th}}$, 2021. Panel C illustrates the identifying variation for the SXSW instrumental variable, where we categorize counties by the timing of when SXSW followers are estimated to have joined Twitter. Note that we exclude Alaska, Hawaii, and U.S. territories.

2.2 Nationscape Survey Data

To examine the link between social media usage, political beliefs, and attitudes towards Trump, we use data from the Democracy Fund + UCLA Nationscape Project (Holliday et al., 2021), a major data collection effort on public opinion conducted during the 2020 U.S. election cycle. Fielded between July 2020 and February 2021, Nationscape surveyed approximately 6,250 respondents per week online.

In our analysis, we focus on Phase 3 of the survey, which overlaps with the post-election period. In addition to a question about respondents' self-reported social media usage for obtaining political information, we utilize a set of questions that capture three key dimensions of political attitudes. The first set of questions concerns beliefs about the integrity of the 2020 election (e.g., confidence in fair and accurate elections, perceived legitimacy of the results). The second captures attitudes towards Donald Trump (e.g., approval of Trump's handling of the Capitol events, support for his actions). The third set measures support for democratic norms (e.g., the importance of acknowledging legitimate winners and approval of Capitol riot participants). We provide a complete list of these questions in Appendix Table A.3 and provide summary statistics in Appendix Table A.2.

2.3 Twitter Data

Trump Tweets Data We compile a comprehensive dataset of Donald Trump's Twitter activity between October 10, 2020, and January 8, 2021 (the date of Trump's permanent account suspension) from the *Trump Twitter Archive*. The archive systematically captured all content from Trump's Twitter account (@realDonaldTrump), including each tweet's text and its engagement metrics (i.e., retweets and likes). Based on these tweets, we create a binary indicator for messages that undermine election integrity and incite riots. We conduct this annotation entirely manually, rather than relying on automated methods, to capture the nuanced and context-dependent nature of political rhetoric. Specifically, we code a tweet as 1 if it explicitly or implicitly promotes claims of election fraud, references the "Stop the Steal" movement, and encourages supporters to mobilize for the January 6^{th} Capitol events. Otherwise, we coded the tweet as 0.

Appendix Figure A.1 shows the daily time series of tweets sent by Trump that actively seek to undermine the election's integrity. Trump's *Stop the Steal* messaging occurs in bursts related to critical moments during the election: the immediate post-election period in early November, with peak activity in late November and early December, and then again leading up to January 6^{th}. While Trump's tweeting intensifies around the election period, the largest spike in his messaging occurs in late November and early December, with some days exceeding 30 *Stop the Steal* tweets.

Representative Twitter User Data We quantify the public's reaction to Trump's Twitter activity on social media based on a representative sample of American Twitter users, building on the sampling frame of Siegel et al. (2021). Their sample was created by querying the Twitter API with random numbers to determine whether each number was associated with an active U.S. Twitter account. We collected the profiles and tweets of 432,882 users, resulting in a dataset of approximately 399 million tweets. We assign Twitter users to counties based on the information provided in their profiles and identify followers of Trump's Twitter account using a follower list downloaded in November 2020. This follower information is unique because it was collected before January 6th, allowing us to precisely identify which users followed Trump *before* the insurrection.

The tweet-level data contains the full text content of posts, timestamps, engagement metrics, and metadata, including hashtags and mentions. For our analysis, we classify political tweets based on the transformer model developed by Habibi, Hovy and Schwarz (2024). We further identify election-denial tweets based on hashtags associated with the *Stop the Steal* movement online: #stopthesteal, #voter_fraud, #riggedelection, #fightfortrump, or #capitolriots. Additionally, we classify the toxicity of the Twitter content using Google's Perspective API (Wulczyn, Thain and Dixon, 2017; Dixon, Li, Sorensen, Thain and Vasserman, 2018). The Perspective API has been widely used as a measure of hate speech and is known to achieve state-of-the-art performance in classifying hateful text based on its toxicity score, which ranges from 0 (non-toxic) to 1 (highly toxic).

In Appendix Figure A.2, we separately show the average number of political and election-denial tweets that are sent by Trump's followers and non-followers, as well as the toxicity of these tweets. Trump's followers produce a significantly larger amount of both types of content. We provide examples of tweets classified as toxic in Appendix Table A.4.

3 Background: Social Media and the 2020 Election

Donald Trump's use of Twitter reshaped political communication in the United States. Twitter provided an unmediated channel that bypassed traditional news outlets, and Trump's use of it established new norms for presidential discourse. From 2009 until his suspension in January 2021, Trump posted more than 57,000 tweets, including over 25,000 during his presidency (Wikipedia, 2025). By the 2020 presidential election, his posting frequency had risen to an average of 35 tweets per day, with peaks of nearly 200 messages in a single day (Statista, 2019).

Trump's tweets were (and continue to be) characterized by a distinct, emotionally charged style that frequently included attacks on opponents. This rhetorical approach increased engagement and diffusion on the platform (Brady, Wills, Jost, Tucker and Van Bavel, 2017; Clarke and Grieve, 2019; Anderson, 2017) and influenced the broader media agenda (Lewandowsky, Jetter and Ecker, 2020). Research also shows that Trump's inflammatory messages fueled hate crimes against minorities (Müller and Schwarz, 2023; Cao, Lindo and Zhong, 2023) and shaped electoral participation, trust in institutions, and perceptions of democratic legitimacy (Fujiwara, Müller and Schwarz, 2024; Green, Hobbs, McCabe and Lazer, 2022; Clayton et al., 2021).

During and after the 2020 election cycle, Trump used his extensive online reach to cast doubt on the integrity of the electoral process. We summarize the timeline of key events in Table 1 in which we also provide examples of Trump's election-related rhetoric. The first explicit challenge to the 2020 election emerged on May 26, 2020, when Trump claimed that mail-in ballots would produce a "Rigged Election," prompting Twitter's first-ever fact-check of a presidential tweet. By September, far-right activists had begun organizing under the *Stop the Steal* slogan, laying the groundwork for post-election mobilization.

Following Election Day, Trump escalated his claims. On November 4th, he falsely asserted victory; when major news networks called the election for Joe Biden on November 7th, Trump denounced the results and encouraged protests nationwide. Offline mobilization followed, including the November 14th "Million MAGA March" in Washington, D.C., which drew thousands of supporters and participants from extremist groups. Trump also formed the *Save America PAC* on the 9th of November, only two days after the 2020 presidential election was declared. Donations to the PAC were explicitly marketed as supporting Trump's "Official Election Defense Fund" (ABC News, 2020).

Another pivotal moment came on December 19th, when Trump tweeted: "Big protest in D.C. on January 6th. Be there, will be wild!" This message circulated widely among extremist communities and was later cited in legal proceedings against Capitol rioters. On January 6th, 2021, Trump continued to repeat false claims on Twitter and encouraged protesters on the National Mall. Shortly after his remarks, thousands of protestors marched to the U.S. Capitol, where at approximately 12:53 PM the outer security perimeter was breached, followed by entry into the building. Roughly 140 law enforcement officers were injured, and seven individuals, including three officers, ultimately lost their lives. The attack delayed the certification of Electoral College votes by more than 13 hours, with Congress affirming the 2020 election results at 3:42 AM on January 7 (Senate, 2021).

Table 1: Timeline of Key Events and Trump's Election-Denial Tweets

Date	Event	Representative Tweet
May 26, 2020	First public attack on mail-in voting	*"There is NO WAY (ZERO!) that Mail-In Ballots will be anything less than substantially fraudulent. Mail boxes will be robbed, ballots will be forged & even illegally printed out & fraudulently signed..."*
Sep 7, 2020	*Stop the Steal* movement emergence	
Nov 4, 2020	False victory declaration	*"Last night I was leading, often solidly, in many key States... Then, one by one, they started to magically disappear as surprise ballot dumps were counted. VERY STRANGE..."*
Nov 7, 2020	Major networks call election for Biden	*"THE OBSERVERS WERE NOT ALLOWED INTO THE COUNTING ROOMS. I WON THE ELECTION, GOT 71,000,000 LEGAL VOTES. BAD THINGS HAPPENED WHICH OUR OBSERVERS WERE NOT ALLOWED TO SEE..."*
Nov 14, 2020	Million MAGA March in Washington D.C.	*"Hundreds of thousands of people showing their support in D.C. They will not stand for a Rigged and Corrupt Election!"*
Dec 12, 2020	Second mass protest in Washington D.C.	*"Wow! Thousands of people forming in Washington (D.C.) for Stop the Steal. Didn't know about this, but I'll be seeing them! #MAGA"*
Dec 19, 2020	Critical mobilization call for January 6^{th}	*"Big protest in D.C. on January 6^{th}. Be there, will be wild!"*
Jan 6, 2021	U.S. Capitol riot	*"The States want to redo their votes. They found out they voted on a FRAUD... BE STRONG!"*
Jan 8, 2021	Twitter permanently suspends Trump's account	*"These are the things and events that happen when a sacred landslide election victory is so unceremoniously & viciously stripped away from great patriots... Go home with love & in peace."*

Notes: This table presents a timeline of key election-related events and corresponding representative tweets from Donald Trump's Twitter account (@realDonaldTrump) during the period from May 2020 to January 2021. The timeline focuses on events related to mail-in voting controversies, the *Stop the Steal* movement, post-election protests, and the January 6^{th} Capitol riot. Event descriptions are primarily sourced from contemporary news reports from major media outlets, at times supplemented with information from Wikipedia. Trump's tweets are obtained from the Trump Twitter Archive, which systematically captured all content from his Twitter account, including tweet text and engagement metrics (retweets and likes). The tweets are representative examples that illustrate Trump's messaging around each event.

Trump's continued defense of the rioters prompted Twitter and other platforms to permanently suspend his account on January 8th, marking the first time a sitting president was deplatformed. His account remained inactive until Elon Musk's acquisition of Twitter in 2023. Following his victory in the 2024 election, Trump issued pardons to all individuals convicted for participating in the Capitol attack on January 20th, 2025.

Building on this sequence of events, our empirical analysis proceeds in three steps. First, we examine how local exposure to Twitter shaped offline mobilization and individual beliefs. Second, we analyze the effect of Trump's tweets on online election denial and on revealed political preferences, as measured by donations. Third, we investigate whether Trump's removal from Twitter altered patterns of political expression, beliefs, and financial support.

4 Social Media and the January 6th Riots

We begin by examining the impact of Twitter exposure on political support for the January 6th riots and the *Stop the Steal* movement. Section 4.1 introduces the empirical strategy, outlines the identifying assumptions, and presents the first-stage results. Section 4.2 presents evidence on the effect of Twitter exposure on participation in the Capitol riots. Section 4.3 investigates the effect of Twitter exposure on donations as another costly and thus credible measure of political preferences. Section 4.4 provides extensive robustness checks. Section 4.5 provides evidence of network spillovers. Finally, Section 4.6 presents complementary evidence on individual beliefs and attitudes from survey data.

4.1 Identification Strategy and First Stage Estimates

Our identification strategy leverages the 2007 South by Southwest (SXSW) festival as an exogenous shock to early Twitter adoption, building on the empirical framework developed in Müller and Schwarz (2023). The identification strategy leverages the fact that Twitter's "breakthrough moment" occurred during SXSW 2007, when the platform hosted a launch event that allowed SXSW participants to sign up by sending a text message to a designated phone number. This led to a surge in activity and registrations, with Twitter's growth rate quadrupling following the festival (Müller and Schwarz, 2023). Since SXSW attracts participants from across the United States, this surge translated into geographically dispersed increases in early adoption, as attendees brought the technology back to their home counties, where their usage had local network effects. Due to these local network effects, counties with residents who joined Twitter during the festival subsequently experienced higher long-run Twitter adoption rates, which have persisted to this day.

The key advantage of this design is that it isolates plausibly exogenous variation in local Twitter usage. Müller and Schwarz (2023) and Fujiwara, Müller and Schwarz (2024) provide extensive evidence for the short and long-run impact of SXSW on the adoption of Twitter. They also show that, conditional on pre-existing interest in Twitter and SXSW, the SXSW adoption shock is uncorrelated with observable county characteristics related to their demographic composition, interest in information technology, or political preferences. Müller and Schwarz (2023) and Fujiwara, Müller and Schwarz (2024) also show that various political outcomes evolved similarly in *treated* and *control* counties before Twitter became politically relevant.

Empirical Strategy Our identification strategy exploits both the location and the timing of SXSW followers' adoption of Twitter. This dual source of variation allows us to isolate the effect of the March 2007 SXSW festival from the underlying interest in the SXSW brand or the self-selection of individuals into festival participation. Specifically, we compare counties whose SXSW followers first joined Twitter during the March 2007 festival with counties whose followers joined before the festival. The latter provides a natural control group, capturing pre-existing interest in SXSW and the baseline propensity to adopt new technologies. Intuitively, counties with followers who were induced to join Twitter as a result of SXSW 2007 experienced an inflow of additional early adopters that drove subsequent Twitter adoption patterns. In contrast, counties where users joined earlier did not receive additional early adopters and do not exhibit persistent differences in Twitter usage, and can therefore be regarded as a placebo check.

This intuition gives rise to the following system of 2SLS equations:

$$\text{Twitter Users}_c = \alpha + \beta \cdot \text{SXSW}_c^{\text{March2007}} + \gamma \cdot \text{SXSW}_c^{\text{Pre}} + X_c'\delta + \xi_c \tag{1}$$

$$y_c = \phi + \theta \cdot \widehat{\text{Twitter Users}}_c + \pi \cdot \text{SXSW}_c^{\text{Pre}} + X_c'\rho + \varepsilon_c \tag{2}$$

where Twitter Users_c is the log number of Twitter users in county c. $\text{SXSW}_c^{\text{March2007}}$ represents the log number of SXSW followers who joined Twitter during March 2007, and $\text{SXSW}_c^{\text{Pre}}$ captures the log number of followers who joined before (with one added inside the logarithm). In the second stage, we study the effect of Twitter penetration on y_c (e.g., participation in the January 6^{th} riots) by regressing it on the predicted component of Twitter usage $\widehat{\text{Twitter Users}}_c$ from the first stage. The vector X_c contains a comprehensive set of controls, including population size, demographics, media consumption patterns, and geographic proximity to Austin, Texas (where SXSW takes place). All specifications further include state fixed effects, which account for any state-level factors that correlate with Twitter adoption or January 6^{th}

participation. To account for the within-state correlation of outcomes, we cluster standard errors at the state level.

Identifying Assumptions Our identification strategy relies on the assumption that $SXSW_c^{March2007}$ affects political outcomes only through its impact on local Twitter adoption. Conditional on $SXSW_c^{Pre}$ and a comprehensive set of controls, the identifying variation comes from comparing counties with similar baseline interest in SXSW but differing exposure to the platform's breakthrough moment.

Several pieces of evidence support this exclusion restriction. First, counties with SXSW followers in March 2007 are observationally equivalent to counties with followers who joined earlier (see Appendix Table B.1). Second, as we will show, the first- and second-stage coefficients for $SXSW_c^{Pre}$ (π and γ) are insignificant and close to 0, suggesting that pre-period interest in the SXSW festival is unrelated to both Twitter usage and political outcomes. This indicates that our estimates for SXSW 2007 exposure capture the effect of the festival-induced adoption shock rather than underlying interest in new technologies or the SXSW brand. Third, we provide placebo checks using followers of other festivals in 2007, none of which have predictive power for Twitter adoption or support for Trump's election denial. Taken together, and consistent with prior work employing this strategy, we interpret estimates of θ as capturing the causal effect of plausibly exogenous variation in local Twitter penetration. For transparency, we report the corresponding 2SLS estimates alongside OLS and reduced-form results throughout.

First Stage Estimates Table 2 reports the first-stage estimates linking SXSW exposure to the subsequent level of Twitter usage across counties. The coefficient on $\text{SXSW}_c^{\text{March2007}}$ is consistently positive and highly significant across specifications, indicating that counties with more SXSW followers who joined during the March 2007 festival experienced greater Twitter penetration in later years. The estimates remain stable with the inclusion of an increasingly rich set of controls in columns 2-5, suggesting that the observed predictive power of SXSW followers cannot be explained by observable county characteristics. The coefficient in column 5 implies that a 10% increase in the number of SXSW followers who joined in March 2007 is associated with a 4.5% higher number of Twitter users.

In support of the exclusion restriction, the coefficient π for $\text{SXSW}_c^{\text{Pre}}$ becomes small and statistically insignificant once geographic and demographic controls are introduced. The coefficient estimate for π in column 5 is five times smaller than our main coefficient of interest θ and far away from statistical significance at conventional levels.[5] These results suggest that

[5]The estimates for θ and π are also statistically significantly different from each other.

Table 2: SXSW and Twitter Usage

	Dep. Var.: Log(Twitter Users)				
	(1)	(2)	(3)	(4)	(5)
Log(SXSW followers, March 2007)	0.614***	0.575***	0.479***	0.470***	0.450***
	(0.062)	(0.064)	(0.060)	(0.058)	(0.064)
Log(SXSW followers, Pre)	0.213***	0.166**	0.119	0.116	0.089
	(0.072)	(0.066)	(0.074)	(0.070)	(0.072)
Population deciles	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes
Geographical controls		Yes	Yes	Yes	Yes
Demographic controls			Yes	Yes	Yes
Socioeconomic controls			Yes	Yes	Yes
China shock controls				Yes	Yes
2020 election control					Yes
Observations	3,107	3,107	3,106	3,105	3,064
Mean of DV	5.29	5.29	5.29	5.29	5.28

Notes: This table presents first stage estimates from Equation (1). *Log(SXSW followers, March 2007)* is the number of Twitter users who joined in March 2007 and follow South by Southwest (SXSW) (in logs, with 1 added inside). *SXSW followers, Pre* is the number of SXSW followers who registered at some point in 2006, defined similarly. *Twitter users* is the number of Twitter users observed in 2014-2015 (in logs). The regressions include the indicated groups of control variables; see the Online Appendix for additional details. Standard errors in parentheses are clustered by state. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

variation in SXSW-induced Twitter adoption can isolate plausibly exogenous variation in social media exposure across counties.

4.2 January 6th Participation Results

We next examine the relationship between Twitter exposure and participation in the January 6th insurrection. Table 3 presents the estimates of OLS, reduced form, and IV estimates from Equation (2). The OLS results in Panel A, in line with the binscatter plot, suggest a statistically significant positive relationship between Twitter exposure and January 6th court cases, which persists when we flexibly control for a plethora of county-level characteristics, including demographics and the 2020 election results. If anything, the inclusion of these controls increases the estimated coefficients.

The reduced-form and IV estimates in Panels B and C confirm a strong positive relationship between Twitter exposure and participation in the January 6th attack. The IV estimates are consistently positive and statistically significant at the 1% level. In the fully saturated specification in column 5, the coefficient of 3.097 implies that a 10% increase in Twitter usage raises the number of January 6th participants by approximately 0.31, an increase

Table 3: Twitter and the January 6^{th}

	Dep. Var.: Nr. January 6^{th} Court Cases				
	(1)	(2)	(3)	(4)	(5)
Panel A: OLS					
Log(Twitter users)	0.629***	0.604***	0.678***	0.669***	0.685***
	(0.183)	(0.172)	(0.178)	(0.174)	(0.187)
Panel B: Reduced form					
Log(SXSW followers, March 2007)	1.299***	1.351***	1.395***	1.386***	1.392***
	(0.309)	(0.314)	(0.326)	(0.329)	(0.332)
Log(SXSW followers, Pre)	0.479	0.601	0.589	0.589	0.602
	(0.413)	(0.423)	(0.495)	(0.498)	(0.509)
Panel C: 2SLS					
Log(Twitter users)	2.117***	2.347***	2.915***	2.951***	3.097***
	(0.504)	(0.554)	(0.665)	(0.687)	(0.743)
Log(SXSW followers, Pre)	0.028	0.212	0.241	0.246	0.328
	(0.492)	(0.477)	(0.522)	(0.518)	(0.492)
Population deciles	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes
Geographical controls		Yes	Yes	Yes	Yes
Demographic controls			Yes	Yes	Yes
Socioeconomic controls			Yes	Yes	Yes
China shock controls				Yes	Yes
2020 election control					Yes
Observations	3,107	3,107	3,106	3,105	3,064
Mean of DV	0.47	0.47	0.47	0.47	0.47
Robust F-stat.	98.45	79.69	63.86	64.78	50.14

Notes: This table presents OLS, reduced form, and 2SLS estimates from Equation (2) examining the relationship between Twitter usage and January 6^{th} participation. The dependent variable is the number of January 6^{th} court cases by county. Panel A shows OLS results with Log(Twitter users) as the main explanatory variable. Panel B presents reduced form results, with Log(SXSW followers, March 2007) and Log(SXSW followers, Pre) as instruments. Panel C shows 2SLS results using Equation (2), where Log(Twitter users) is instrumented using SXSW followers who joined in March 2007. Log(SXSW followers, March 2007) measures the number of Twitter users (in logs, with 1 added inside) who joined in March 2007 and follow South by Southwest. Log(SXSW followers, Pre) measures SXSW followers who registered before March 2007, defined similarly. All regressions include population decile and state fixed effects, with successive columns adding control variables as indicated. The robust F-statistic tests instrument strength in the 2SLS specifications. Standard errors in parentheses are clustered by state. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

of about 66% relative to the mean. In contrast, the estimates for pre-2007 SXSW followers are always statistically insignificant and markedly smaller than the IV coefficients. We interpret this pattern as evidence that our results are not driven by unobservable characteristics such as a county's predisposition toward early technology adoption or January 6^{th} participation, but instead reflect the effect of the exogenous variation in Twitter usage generated by the SXSW 2007 festival.

As the first-stage F-statistics consistently exceed 50 across all specifications, well above conventional thresholds and those recommended by Stock and Yogo (2005), biases arising from weak instruments are unlikely in our setting. The fact that the IV estimates are roughly three times larger than the OLS estimates is consistent with two factors. First, prior to Elon Musk's takeover, Twitter's user base skewed strongly liberal, which would bias the OLS coefficients downward (see Fujiwara, Müller and Schwarz, 2021). Second, the IV strategy mitigates measurement error in county-level Twitter usage, thereby attenuating OLS estimates.

4.3 Effects on Political Donations

The armed storming of a country's legislative assembly represents one of the most extreme manifestations of anti-democratic mobilization that social media can facilitate. Such actions are typically enabled and legitimized by broader patterns of democratic backsliding and polarization (e.g., Adena, Enikolopov, Petrova, Santarosa and Zhuravskaya, 2015; Satyanath, Voigtländer and Voth, 2017; Levitsky and Ziblatt, 2018). Motivated by this insight, we examine political donations as a costly, and therefore credible, measure of anti-democratic preferences. Specifically, we study the relationship between Twitter exposure and donations to the *Save America PAC* between November 9$^{\text{th}}$, 2020, and January 6$^{\text{th}}$, 2021.

Table 4 reports the OLS, reduced-form, and IV estimates of Equation (2), where the dependent variable is the log number of donations to the *Save America PAC*. Mirroring our findings for participation in the January 6$^{\text{th}}$ attack, we obtain positive and statistically significant coefficients across all specifications, regardless of the included control sets. The IV estimates in our preferred specification in column (5) imply that a 10% increase in Twitter usage leads to a 7.2% increase in donations to the *Save America PAC*. The estimates of π, capturing the association between pre-existing SXSW followers and donations, are small in magnitude and insignificant, further supporting our identifying assumption that our findings are not driven by unobserved county characteristics.

4.4 Robustness

Placebo Tests We conduct additional falsification tests using Twitter followers of other major U.S. music festivals as placebo instruments. If our results were driven by omitted county-level factors that make counties more likely to adopt Twitter or participate in the *Stop the Steal* movement, then we should observe similar effects for followers of these alternative festivals. Table 5 reports the results using Twitter followers who joined during the festival months of five prominent U.S. festivals: Burning Man, Coachella, Pitchfork, Electric Daisy Carnival (EDC), and Austin City Limits (ACL). Panel A presents the first-stage estimates,

Table 4: Twitter and Save America Donations

	Dep. Var.: Log(Save America Donations)				
	(1)	(2)	(3)	(4)	(5)
Panel A: OLS					
Log(Twitter users)	0.569***	0.593***	0.656***	0.638***	0.702***
	(0.069)	(0.078)	(0.096)	(0.096)	(0.098)
Panel B: Reduced form					
Log(SXSW followers, March 2007)	0.338***	0.378***	0.375***	0.332***	0.321***
	(0.086)	(0.091)	(0.106)	(0.109)	(0.098)
Log(SXSW followers, Pre)	0.084	0.013	-0.131	-0.160	-0.196
	(0.146)	(0.132)	(0.162)	(0.170)	(0.163)
Panel C: 2SLS					
Log(Twitter users)	0.551***	0.657***	0.784***	0.706***	0.715***
	(0.130)	(0.151)	(0.204)	(0.213)	(0.208)
Log(SXSW followers, Pre)	-0.034	-0.096	-0.225	-0.242	-0.260
	(0.158)	(0.146)	(0.163)	(0.170)	(0.159)
Population deciles	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes
Geographical controls		Yes	Yes	Yes	Yes
Demographic controls			Yes	Yes	Yes
Socioeconomic controls			Yes	Yes	Yes
China shock controls				Yes	Yes
2020 election control					Yes
Observations	3,107	3,107	3,106	3,105	3,064
Mean of DV	6.67	6.67	6.67	6.67	6.67
Robust F-stat.	98.45	79.69	63.86	64.78	50.14

Notes: This table presents OLS, reduced form, and 2SLS estimates from Equation (2) examining the relationship between Twitter usage and donations for the pro-Trump *Save America PAC*. The dependent variable is the total amount of Save America donations (in logs, with 1 added) by county over the period from November 9$^{\text{th}}$, 2020 to January 6$^{\text{th}}$, 2021. Panel A shows the OLS results with Log(Twitter users) as the main explanatory variable. Panel B presents reduced form results, with Log(SXSW followers, March 2007) as an instrument. Panel C shows 2SLS results using Equation (2), where Log(Twitter users) is instrumented using SXSW followers who joined in March 2007. Log(SXSW followers, March 2007) measures the number of Twitter users (in logs, with 1 added inside) who joined in March 2007 and follow South by Southwest. Log(SXSW followers, Pre) measures SXSW followers who registered before March 2007, defined similarly. The regressions include population decile and state fixed effects, with successive columns adding control variables as indicated. The robust F-statistic tests instrument strength in the 2SLS specifications. Standard errors in parentheses are clustered by state. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

where followers of these festivals serve as potential "instruments" for Twitter usage. We find that only SXSW followers exhibit a significant relationship with subsequent Twitter adoption, while the estimates for all other festivals are small and statistically indistinguishable from zero.

In Panels B, we present reduced-form estimates that examine the relationship between alternative festival followers and January 6$^{\text{th}}$ participation. We focus on the reduced-form estimates because participation in these other festivals does not predict Twitter usage. In line with the assumptions of our identification strategy, none of these other festivals exhibits a significant relationship with the number of participants in the January 6$^{\text{th}}$ insurrection. Comparing the estimated coefficients with our baseline estimates shows they are substantially

Table 5: Placebo Other Festivals

	SXSW (1)	Burning Man (2)	Coachella (3)	Pitchfork (4)	EDC (5)	ACL (6)
Panel A: First Stage						
	Dep. Var.: Twitter Usage					
Followers Festival Month	0.196***	-0.002	0.021	0.001	0.009	-0.015
	(0.020)	(0.022)	(0.016)	(0.014)	(0.009)	(0.022)
Population deciles	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,107	3,107	3,107	3,107	3,107	3,107
Panel B: Reduced Form: January 6$^{\text{th}}$ Court Cases						
	Dep. Var.: Nr. January 6th Court Cases					
Followers Festival Month	0.414***	0.034	0.128	0.069	0.093	0.204
	(0.099)	(0.140)	(0.216)	(0.115)	(0.192)	(0.229)
Population deciles	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,107	3,107	3,107	3,107	3,107	3,107
Panel C: Reduced Form: Save America Donations						
	Dep. Var.: log(Save America Donations)					
Followers Festival Month	0.108***	-0.023	0.020	0.013	-0.005	0.021
	(0.027)	(0.031)	(0.024)	(0.013)	(0.017)	(0.027)
Population deciles	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,107	3,107	3,107	3,107	3,107	3,107

Notes: This table presents falsification tests using other festivals as placebo instruments to validate the SXSW identification strategy. Panel A shows first-stage results with Log(Twitter Usage) as the dependent variable. Panel B presents reduced form results with the number of January 6$^{\text{th}}$ court cases as the dependent variable. Panel C shows reduced form results with Save America Donations(in logs, with 1 added inside), summed by county over the period of January 6$^{\text{th}}$, 2021, as the dependent variable. The key explanatory variable "Followers Festival Month" measures the log of users who started following each respective festival during their main month (with 1 added inside the log), and is thus constructed as an analogue to our SXSW instrument. Column 1 uses SXSW (March 2007), our instrument, to aid comparisons. Columns 2-6 test the alternative festivals Burning Man, Coachella, Pitchfork, Electric Daisy Carnival (EDC), and Austin City Limits (ACL). All regressions include population decile and state fixed effects. Standard errors in parentheses are clustered by state. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

smaller. In Panel C, we repeat the same analysis for donations to the *Save America PAC*. The coefficients for the Twitter followers of other festivals are again small in magnitude and statistically indistinguishable from zero. Notably, this includes the Austin City Limits festival, which is held at the same physical location as SXSW. This case is particularly informative, as it demonstrates that our results are not driven by unobserved county characteristics correlated with the festival venue.

Together, these findings support the conclusion that exposure to SXSW in March 2007 generated a unique shock to early Twitter adoption, rather than reflecting underlying demographic differences or pre-existing interest in new technologies.

Additional Robustness Appendix Table B.2 presents a series of robustness checks that examine the sensitivity of our main findings to alternative specifications and sample definitions. Column (1) weights the regression based on a county's population. Columns (2) and (3) introduce more flexible controls for the pre-period interest in the SXSW festival, controlling for either pre-period followers using polynomials or deciles. Column (4) restricts the sample to counties with non-zero January 6th participants, ensuring our findings are not exclusively driven by extensive margin effects, but even hold for the intensive margin of January 6th participants. In column (5), we exclude counties with no SXSW followers, focusing only on counties with direct festival exposure (either before or after). Finally, in column (6), we exclude counties with no donations to *Save American PAC*. Throughout these tests, the results remain statistically significant and of similar magnitude.

Appendix Table B.3 examines the sensitivity of our results to alternative functional form specifications of the dependent variables. For January 6th participation, we estimate effects using the number of cases in levels (baseline, column 1), a dummy indicator for any cases (column 2), and the asinh transformation (column 3). For donations to *Save America PAC*, we examine donations in levels (column 4), the number of individual donors (column 5), and the asinh of donations (column 6). The estimates are statistically significant and positive across the different functional forms, highlighting the effects of Twitter usage on political mobilization regardless of the specific transformation.

Finally, in Appendix Figure B.1 we show the estimates from a leave-one-state-out analysis, in which we estimate our IV estimates dropping one state at a time. We find that the estimates are stable, independent of the state we drop. In other words, no individual state appears to drive our findings.

4.5 The Role of Network Spillovers

The results above document a strong relationship between social media exposure and anti-democratic sentiments. Given the substantial evidence that social networks facilitate protest coordination (e.g., Enikolopov, Makarin and Petrova, 2020*b*; Enikolopov, Makarin, Petrova and Polishchuk, 2017) and can amplify extremist behaviors such as hate crimes (Müller and Schwarz, 2021; Bursztyn, Egorov, Enikolopov and Petrova, 2019), we next examine the role of network spillovers in propagating participation in the January 6th attack.

Specifically, we adopt a specification in the spirit of Qin, Strömberg and Wu (2024), who examine how social media facilitates the coordination and diffusion of protests in China. To isolate the role of online spillovers, Qin, Strömberg and Wu (2024) construct measures of a city's exposure to protests through social media networks and through physical proximity.

They then relate protest incidence in a given city to these network-based exposure measures, thereby distinguishing between digital and geographic channels of diffusion.

We adapt this methodology to our setting by constructing analogous exposure measures for the January 6$^{\text{th}}$ rioters. The first measure captures exposure through the Twitter network, defined by the number of Twitter connections between a focal county and other counties with January 6$^{\text{th}}$ participants. The second measure captures exposure through local, offline networks, proxied by an inverse-distance-weighted count of January 6$^{\text{th}}$ participants in nearby counties. Formally, these two measures are defined as:

$$\text{Twitter Network Exposure}_c = \sum_{j \neq c} \frac{\text{Twitter Connections}_{cj}}{\text{Twitter Connections}_c} \cdot y_j$$

$$\text{Local Network Exposure}_c = \sum_{j \neq c} \frac{1}{\text{Distance}_{cj}} \cdot y_j$$

where y_j denotes the number of January 6$^{\text{th}}$ rioters in county $j \neq c$ to which county c is connected. Twitter Connections$_{cj}$ is the number of Twitter connections from county c to county j and Twitter Connections$_c$ is the total number of Twitter connections from county c, which we construct based on data of the follower network of Twitter users from Siegel et al. (2021), collected separately. Accordingly, we restrict our analysis to the subset of counties for which we observe at least one Twitter user.[6] Distance$_{cj}$ denotes the geodetic distance between county centroids. To make the magnitudes comparable, we standardize both measures to have a mean of 0 and a standard deviation of 1. We then estimate network exposure regressions of the form:

$$\begin{aligned} y_c = \phi + \theta \cdot \text{Twitter Users}_c + \lambda_1 \cdot \text{Twitter Network Exposure}_c \\ + \lambda_2 \cdot \text{Local Network Exposure}_c + X_c'\rho + \varepsilon_c \end{aligned} \tag{3}$$

All remaining variable definitions follow Equation (2), and we additionally control for each county's total number of Twitter connections. The results of this analysis are presented in Table 6. Consistent with the presence of spillovers operating through the Twitter network, we find that a county's connections to other counties with many January 6$^{\text{th}}$ rioters predict its own number of participants. Using the estimates from column (5), a one standard deviation increase in Twitter network exposure predicts 0.7 additional participants in the January 6$^{\text{th}}$

[6] There are two differences between our measure and the social media exposure measure in Qin, Strömberg and Wu (2024). First, because following someone on Twitter is directional, TwitterConnections$_{cj}$ and TwitterConnections$_{jc}$ need not be equal. Second, since we do not observe temporal variation in January 6$^{\text{th}}$ participation, our analysis only exploits cross-sectional rather than panel variation.

attacks, which is large relative to the mean of 0.812. This suggests that ties to Twitter users in other counties can amplify local political mobilization.

In contrast, exposure through local offline networks (proxied by physical distance) does not appear to matter in our setting, as the corresponding coefficients are statistically insignificant and negative throughout. Importantly, including network exposure measures does not affect the coefficient on the direct effect of Twitter exposure, which remains positive and significant across all specifications. Network spillovers through Twitter thus independently predict higher involvement in the storming of the capitol, even when we hold the overall level of Twitter exposure constant.

Table 6: Network Effects of Social Media

	Dep. Var.: Nr. Jan 6 Cases				
	(1)	(2)	(3)	(4)	(5)
Twitter Network Exposure	0.611	0.776**	0.712**	0.715**	0.737**
	(0.395)	(0.326)	(0.313)	(0.314)	(0.314)
Local Network Exposure	-0.153	-0.234	-0.181	-0.189	-0.227
	(0.246)	(0.187)	(0.177)	(0.179)	(0.182)
Log(Twitter users)	0.701***	0.756***	0.993***	0.990***	1.049***
	(0.168)	(0.180)	(0.208)	(0.205)	(0.217)
Population deciles	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes
Geographical controls		Yes	Yes	Yes	Yes
Demographic controls			Yes	Yes	Yes
Socioeconomic controls			Yes	Yes	Yes
China shock controls				Yes	Yes
2020 election control					Yes
Observations	1,591	1,591	1,591	1,590	1,562
Mean of DV	0.819	0.819	0.819	0.819	0.812

Notes: This table presents OLS estimates from Equation (3) examining the relationship between Twitter connections and January 6^{th} participation. The dependent variable is the number of January 6^{th} court cases by county. The independent variables Twitter Network Exposure and Local Network Exposure are constructed as defined in the equation above, and capture a county's connections to January 6^{th} rioters via Twitter networks and linear distance (as a proxy for local networks), respectively. All regressions include population decile and state fixed effects, with successive columns adding control variables as indicated. Standard errors in parentheses are clustered by state. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1.p < 0.1$.

Note that, since we do instrument for the network exposure measures, we do not claim that these estimates are causal. That said, these patterns are consistent with a mechanism in which social media influences violent anti-democratic actions both through direct content exposure and by facilitating coordination within online networks.

4.6 Distrust in Election Results

As a final piece of evidence in the first part of our analysis, we examine whether Twitter exposure mattered for attitudes and perceptions about the 2020 election and the legitimacy of anti-democratic actions, which are plausible mechanisms that may underlie the documented offline effects. For this purpose, we turn to individual-level survey data from Nationscape. We focus on beliefs regarding the legitimacy of the 2020 election results, the credibility of Trump's claims of election fraud, and attitudes toward political violence, and estimate regressions of the following type:

$$y_i = \alpha + \beta \cdot \mathbb{1}[\text{Social Media User}]_i + \mathbf{X_i'}\beta + \epsilon_i \tag{4}$$

where y_i is a binary response of individual i to one of the questions from the Nationscape Survey (Nov 5, 2020 to Jan 21, 2021). We focus particularly on the false belief that Trump won the 2020 election. In additional exercises, we also look at approval of the January 6^{th} rioters, approval of Trump's actions on January 6^{th}, confidence in a fair election, and concerns about Twitter Bans. We recode all questions such that 1 indicates support for Trump, and 0 indicates no support. $\mathbb{1}[\text{Social Media User}]_i$ is an indicator equal to 1 if the respondent is using social media. $\mathbf{X_i}$ is a large set of control variables accounting for respondents' gender, age, race, income, education, and congressional district fixed effects. We cluster standard errors by congressional district.

We present the estimates from Equation (4) in Table 7, which focuses on the relationship between social media usage and the belief that Trump won the 2020 election, i.e. the denial of the official election results. Column (1) reports estimates for all respondents in the Nationscape survey. Social media users are 3.2 percentage points more likely to affirm this belief, an increase of roughly 15% relative to the mean of 22%. Columns (2)–(4) split the sample by party affiliation. We find that this pattern is especially pronounced among self-identified Republicans and Independents, for whom social media usage is associated with increases of 6.2 and 3.7 percentage points, respectively. In contrast, among Democrats, social media use is not associated with a greater likelihood of believing that Trump won the 2020 election.

We present additional evidence from other survey questions that were less frequently asked in the Nationscape survey in Appendix Table B.4, including approval of the January 6^{th} rioters, approval of Trump's actions on January 6^{th}, confidence in a fair election and concerns about Twitter Bans. For all of these questions, we observe that social media users are significantly more likely to affirm anti-democratic attitudes (columns 1–4) and are perhaps unsurprisingly more concerned about Twitter bans, a question the third part of our paper tackles more directly.

Table 7: Evidence from the Nationscape Survey

	Dep. Var.: Trump Won 2020 Election			
	All (1)	Republicans (2)	Democrats (3)	Independents/Others (4)
Social Media User	0.032*** (0.006)	0.062*** (0.012)	0.003 (0.004)	0.037*** (0.010)
Cong. District FE	Yes	Yes	Yes	Yes
Gender	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes
Hispanic	Yes	Yes	Yes	Yes
Race	Yes	Yes	Yes	Yes
Income	Yes	Yes	Yes	Yes
Education	Yes	Yes	Yes	Yes
Observations	56,519	17,170	21,115	18,153
Mean of DV	0.22	0.53	0.03	0.18

Notes: This table presents regression estimates using data from ten waves of the Nationscape survey conducted weekly between November 5, 2020 and January 21, 2021. The dependent variable is based on the survey question "Who do you think won the November 2020 Presidential election?" coded as 1 if the respondent answered "Trump" and 0 otherwise. The key explanatory variable, Social Media User, captures whether respondents have seen or heard news about politics on social media (e.g., Facebook, Twitter) in the past week. Exact question wording is provided in Appendix Table A.3. Column (1) includes all respondents, while columns (2), (3), and (4) restrict the sample to Republicans, Democrats, and Independents/Others, respectively, based on self-reported political ideology. All specifications include controls for gender, age, Hispanic ethnicity, race, income, education, and congressional district fixed effects. Standard errors in parentheses are clustered by state. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Taken together, these results indicate that social media users were more likely to endorse anti-democratic views in the aftermath of the 2020 election. This evidence complements existing work showing that younger and more partisan voters exhibit weaker commitments to democratic norms (e.g., Foa and Mounk, 2016; Svolik, 2019), which could contribute to an environment in which political violence can become legitimized.

5 Trump's Tweets and Anti-Democratic Sentiment

Having established that broader social media exposure is associated with anti-democratic beliefs and behaviors, we now examine whether direct elite messaging can serve as a trigger. In this second part of the paper, we turn to Donald Trump's role in promoting election denial online and mobilizing his supporters. We begin by studying the high-frequency dynamics between Trump's tweets and the behavior of other Twitter users, documenting how election-fraud narratives spread rapidly across the platform following his tweets. We then show that this online rhetoric translated into offline mobilization, with Trump's messages driving additional donations to the *Save America PAC*, particularly in counties with high Twitter usage.

5.1 High-Frequency Time Series Estimates

To examine whether and how political elites can use social media to mobilize their followers and stimulate anti-democratic action, we exploit the precise timing of Trump's tweets, casting doubt on the integrity of the 2020 election. We analyze the behavior of Twitter users in the 96 hours before and after each of Trump's posts, using 8-hour intervals to capture short-run dynamics.[7] Formally, we estimate the following regressions:

$$\text{User Stop the Steal Tweet}_t = \sum_{k=-12}^{12} \beta_k \cdot \text{Trump Stop the Steal Tweet}_{t+k} + \mathbf{X}_t' \gamma + \epsilon_t \qquad (5)$$

where User Stop the Steal Tweet_t denotes the number of *Stop the Steal* tweets posted by users in time interval t. Likewise, Trump Stop the Steal Tweet_t is the number of *Stop the Steal* tweets posted by Donald Trump at time t. The specification includes 12 leads and 12 lags of Trump's posting activity to capture dynamic changes in outcomes before and after each tweet. We also include time-varying controls, $\mathbf{X}_t$, including tweet characteristics (tweets, retweets, favorites) and date fixed effects to absorb temporal variation associated with major political events during this period.

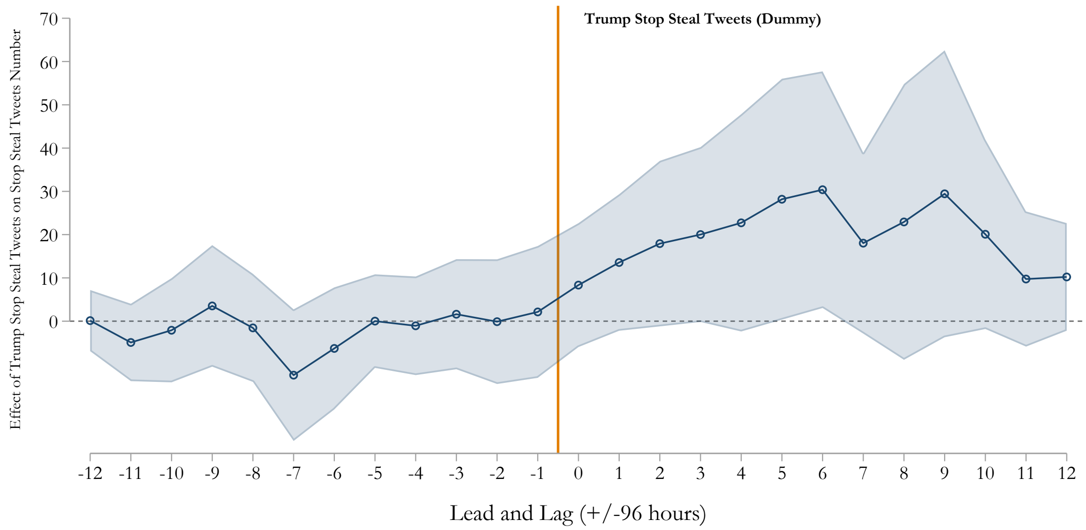
We present the regression estimates in Figure 2. The frequency of *Stop the Steal* tweets posted by users is essentially flat during the 96 hours preceding a *Stop the Steal* tweet by Donald Trump. In contrast, immediately after Trump's tweet, we observe a sharp and statistically significant increase in users' *Stop the Steal* activity, an effect that persists for 80 hours. These results are consistent with the view that Trump's messaging amplified the anti-democratic sentiments that ultimately contributed to the storming of the Capitol. The estimates indicate a doubling in *Stop the Steal* tweets relative to the mean.

5.2 Anti-Democratic Online Views and Offline Political Action

Our previous analyses reveal two key results: (1) increased Twitter usage fosters anti-democratic political mobilization offline, and (2) Trump's rhetoric generates cascades of similar content on the platform. We now bring these findings together to show how the online mobilization of followers can enable political elites to pursue anti-democratic objectives offline.

[7] The 8-hour frequency balances the precision of the timing with the declining number of *Stop the Steal* tweets observed in narrower time windows.

Figure 2: Trump's Tweets and Follower Reactions



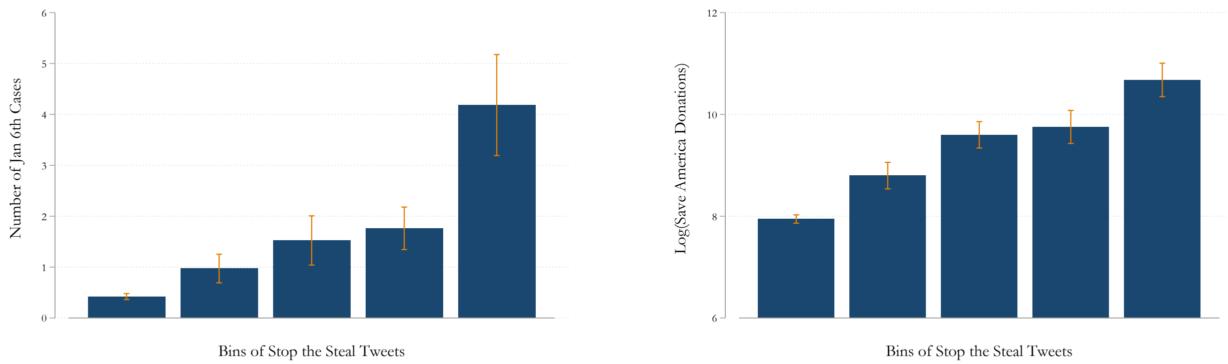
Notes: This figure plots the event study coefficients β_k from Equation (5). The dependent variable is the number of user tweets containing Election Denial topics in each 8-hour interval. The x-axis shows leads and lags in 8-hour intervals (±96 hours). Point estimates are displayed with 95% confidence intervals. Controls include Trump tweet characteristics (tweets, retweets, favorites) and date fixed effects.

We begin by presenting a suggestive relationship in Figure 3, which plots the county-level relationship between bins of *Stop the Steal* tweets and subsequent offline political participation, measured by January 6$^{\text{th}}$ attendance (Panel A) and donations to the *Save America PAC* (Panel B). Both panels display a significant positive association: counties that expressed more online election denial also engaged more heavily in costly offline political actions. These patterns indicate that online expressions of anti-democratic support are predictive of real-world political behavior. Building on this evidence, we next exploit the temporal variation in Trump's tweets within a panel regression framework to more systematically investigate the link between online mobilization and offline outcomes.

Our empirical strategy uses an instrumental variables approach that follows Bartik-style logic (Goldsmith-Pinkham, Sorkin and Swift, 2020), combining cross-sectional variation in Twitter exposure with time-varying shocks generated by Trump's messaging intensity. The key insight is that counties with historically higher Twitter adoption should exhibit stronger responses to Trump's tweets, undermining the integrity of the 2020 election. Formally, we

Figure 3: *Stop the Steal* Tweets and Offline Participation at County Level

(a) January 6 Participation (b) Save America Donations



Notes: These bar charts show the relationship between county-level *Stop the Steal* tweets over the period of January 6$^{\text{th}}$, 2021 (in five bins) and offline participation in anti-democratic activities. Panel (a) plots the number of January 6$^{\text{th}}$ riot participants; Panel (b) plots log donations to the *Save America PAC* (with 1 added inside).

estimate the following system of equations:

$$\begin{aligned} \text{Twitter Users}_c \times \text{Trump Tweets}_t = \pi \cdot \text{SXSW}_c^{\text{March2007}} \times \text{Trump Tweets}_t \\ + \gamma \cdot \text{SXSW}_c^{\text{Pre}} \times \text{Trump Tweets}_t \\ + \delta_c + \lambda_t + X'_{ct}\theta + \nu_{ct} \end{aligned} \tag{6}$$

$$\begin{aligned} y_{ct} = \beta \cdot \widehat{\text{Twitter Users}_c \times \text{TrumpTweets}_t} \\ + \kappa \cdot \text{SXSW}_c^{\text{Pre}} \times \text{TrumpTweets}_t \\ + \delta_c + \lambda_t + X'_{ct}\rho + \varepsilon_{ct} \end{aligned} \tag{7}$$

In the first-stage equation, the dependent variable Twitter $\text{Users}_c \times \text{Trump Tweets}_t$ is the interaction between the number of Twitter users in county c and the number of *Stop the Steal* tweets posted by Donald Trump on day t. As before, $\text{SXSW}_i^{\text{March2007}}$ denotes the log number of SXSW followers in county i who joined Twitter during March 2007, while $\text{SXSW}_i^{\text{Pre}}$ captures the log number of followers who joined prior to March 2007 (with one added inside the logarithm). The vector X_{ct} includes the same county-level controls used in Equation (2), each interacted with Trump Tweets_t. We include county fixed effects δ_i and date fixed effects λ_t, and cluster standard errors at the state level.

In the second-stage equation, the dependent variable y_{ct} is either the log number of *Stop the Steal* tweets originating from county c or the log number of Save America donations (with one added inside the logarithm). We regress these outcomes on $\widehat{\text{TwitterUsers}_c \times \text{TrumpTweets}_t}$, the fitted values from the first stage. The coefficient β therefore captures the differential increase in online or offline anti-democratic activity in counties with exogenously higher Twitter penetration due to the SXSW adoption shock, motivated by the intuition that these counties should respond more strongly to Trump's election-denial tweets.

We present the regression results in Table 8. As before, we report the OLS (Panel A), reduced-form (Panel B), and 2SLS estimates (Panel C). In Columns 1–3, the dependent variable is the log number of *Stop the Steal* tweets sent in a county on a given day (with one added inside the logarithm). The estimates in Column 1 indicate a positive and statistically significant increase in *Stop the Steal* activity on days when Trump posts more election-denial tweets. Columns 2 and 3 add interactions with county-level characteristics. The estimated coefficients remain positive and significant at the 1% level and increase in magnitude, suggesting that the relationship is not driven by observable county attributes. The 2SLS estimates imply that a doubling of Trump's tweets is associated with a 10% increase in *Stop the Steal* activity in counties at the 75^{th} percentile of Twitter usage relative to counties at the 25^{th} percentile.[8] Consistent with our identification assumptions, the interaction between $\text{SXSW}_i^{\text{Pre}}$ and Trump's tweets is small and statistically insignificant throughout.

Columns 4–6 use the log amount of donations to the *Save America PAC* as the dependent variable. We again find positive and statistically significant coefficients throughout. In the fully saturated specification in Column 6, the coefficient of 0.195 implies that a doubling of Trump's tweets is associated with a 36% increase in donations in counties at the 75^{th} percentile of Twitter usage relative to counties at the 25^{th} percentile.[9]

Taken together, these results indicate not only that broader social media exposure shapes the geography of anti-democratic sentiment, but also that targeted elite messaging, delivered directly and unfiltered to large online audiences, can mobilize costly political actions aimed at undermining democratic institutions. In this case, messages sent by the U.S. president had disproportionately strong effects in counties with greater social media penetration.

Robustness We conduct a battery of robustness checks for the panel estimation results. First, we replicate the placebo tests using followers of other 2007 festivals in Appendix Table C.1. Consistent with our identification strategy, none of the counties with followers from other festivals exhibit a disproportionate response to Trump's tweets. Second, the results are robust to alternative functional forms of the dependent variables (Appendix Table C.2). Third, a leave-one-state-out analysis in Appendix Figure C.1 shows that the estimates are highly stable and not driven by any single state.

[8] To see this, note that $0.065 \times \ln(2) \times 2.267 \approx 0.102$, and $e^{0.102} - 1 \approx 0.107$.
[9] To see this, note that $0.195 \times \ln(2) \times 2.267 \approx 0.306$, and $e^{0.306} - 1 \approx 0.358$.

Table 8: Trump's Tweets and Local Anti-Democratic Political Mobilization

	Log(Number of Stop the Steal Tweets)			*Log(Save America Donations)*		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: OLS						
Log(Twitter users) × Log(Trump StopSteal Tweets)	0.004***	0.008***	0.008***	0.177***	0.046***	0.066***
	(0.000)	(0.002)	(0.001)	(0.003)	(0.008)	(0.007)
Panel B: Reduced form						
Log(SXSW followers, March 2007) × Log(Trump StopSteal Tweets)	0.032***	0.026***	0.025***	0.556***	0.059***	0.076***
	(0.004)	(0.004)	(0.004)	(0.047)	(0.022)	(0.019)
Log(SXSW followers, Pre) × Log(Trump StopSteal Tweets)	0.011	0.012	0.014	-0.067	0.013	-0.013
	(0.009)	(0.009)	(0.009)	(0.099)	(0.036)	(0.033)
Panel C: 2SLS						
Log(Twitter users) × Log(Trump StopSteal Tweets)	0.013***	0.053***	0.065***	0.224***	0.121***	0.195***
	(0.002)	(0.008)	(0.010)	(0.008)	(0.041)	(0.041)
Log(SXSW followers, Pre) × Log(Trump StopSteal Tweets)	0.014	0.005	0.010	-0.006	-0.003	-0.026
	(0.010)	(0.008)	(0.009)	(0.032)	(0.034)	(0.028)
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes
Population deciles		Yes	Yes		Yes	Yes
Log(Total donations)		Yes	Yes		Yes	Yes
Demographic controls			Yes			Yes
Socioeconomic controls			Yes			Yes
China shock controls			Yes			Yes
2020 election control			Yes			Yes
Observations	261,072	261,072	257,460	261,072	261,072	257,460
Mean of DV	0.01	0.01	0.01	0.45	0.45	0.46
Robust F-stat.	190.95	66.34	43.25	190.95	66.34	43.25

Notes: This table presents the estimates of the county-day panel regressions in Equation (7). In columns 1-3, the dependent variable is the number of *Stop the Steal* tweets (in logs with one added inside). In columns 4-6, the dependent variable is the total dollar amount of donations to the *Save America PAC* between November and January 6$^{\text{th}}$, 2021. *Log(Num Trump Tweets on StopSteal* is the number of tweets by Trump undermining the integrity of the 2020 election results. *Log(SXSW followers, March 2007)* is the number of Twitter users (in logs, with 1 added inside) who joined in March 2007 and follow South by Southwest (SXSW). *SXSW followers, Pre* is the number of SXSW followers who registered at some point in 2006, defined similarly. *Twitter users* are the number of users, collected in 2014-2015. Standard errors in parentheses are clustered by state. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

6 Trump's Account Suspension and the Crackdown on Election Denial

In the third and final part of the paper, we examine the online and offline aftermath of the January 6$^{\text{th}}$ insurrection. After the attack on the Capitol, all major platforms, including Twitter, Facebook, Instagram, and YouTube suspended Donald Trump's accounts, citing the risk of further incitement to violence. In addition to removing Trump, platforms took action against a broad network of users associated with the *Stop the Steal* movement. Twitter alone suspended approximately 70,000 accounts spreading election-denial rhetoric (McCabe, Ferrari, Green, Lazer and Esterling, 2024). At the same time, federal law enforcement initiated large-scale prosecutions of January 6$^{\text{th}}$ participants, many of whom were charged and sentenced in the months that followed (Wikipedia, 2026). For brevity, we refer to this period as the time after Trump's account removal, with the understanding that it encompasses both the broader platform-wide crackdown and the federal prosecution effort following the insurrection, as well as potential shifts in social norms after the event that might affect individual behavior.

Our analysis proceeds in two steps. First, we study the online behavior of users who were not themselves deplatformed but may have been exposed to online election denial. Second, we investigate whether enforcement efforts were sufficient to weaken or break the relationship between social media exposure and anti-democratic beliefs and donations.

6.1 Crackdown on Election Denial and Online Rhetoric

We investigate the effects on online behavior by studying the activity of Trump's followers relative to other Twitter users around the timing of the January 6$^{\text{th}}$ insurrection and the January 8$^{\text{th}}$ account deletion. Specifically, we use a difference-in-differences design to estimate changes in Twitter activity depending on whether users followed Trump *before* the account deletion, who were plausibly more exposed to his rhetoric. The estimating equation is:

$$y_{it} = \alpha_i + \gamma_t + \beta \cdot \text{Trump follower}_i \times \text{Post Jan } 6_t + \theta \mathbf{X}_{it} + \varepsilon_{it}, \tag{8}$$

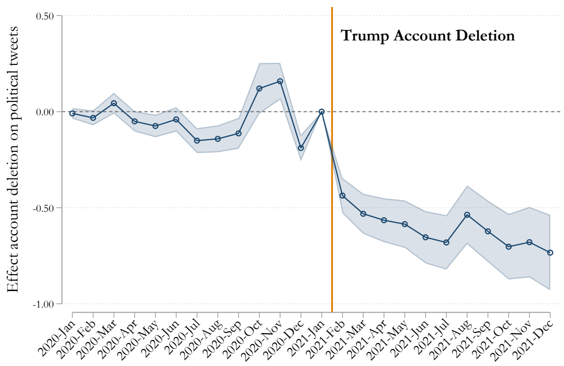
where Y_{it} denotes a measure of user behavior for account i in month t. In our baseline regressions, Y_{it} is either the number of political tweets or the number of *Stop the Steal* tweets posted by a user in a given month. We also examine the toxicity of users' tweets by counting how many exceed a Perspective API toxicity score of 0.8 on any toxicity dimension, a commonly used threshold. This allows us to isolate the subset of tweets that are most emotionally charged and plausibly more inflammatory. Trump follower$_i$ is an indicator variable equal to 1 for users who followed Trump before his account was deleted, and 0 otherwise. Post Jan 6_t is an indicator equal to 1 after January 2021. α_i and γ_t are a full set of user and month fixed effects, which account for average differences in toxicity across users and time. $\mathbf{X}_{it}$ is a set of additional control variables, like user-specific linear or quadratic time trends, or additional fixed effects. We cluster standard errors at the user level.

As noted above, the estimates capture the combined change in user behavior following January 6$^{\text{th}}$; we do not seek to isolate the causal effect of Trump's deplatforming alone. Accordingly, the coefficient β reflects how public expression on Twitter shifted differentially for users with greater versus lesser exposure to election denial, capturing the aggregate impact of both the online platform crackdown and the broader offline enforcement actions taken against election denial.

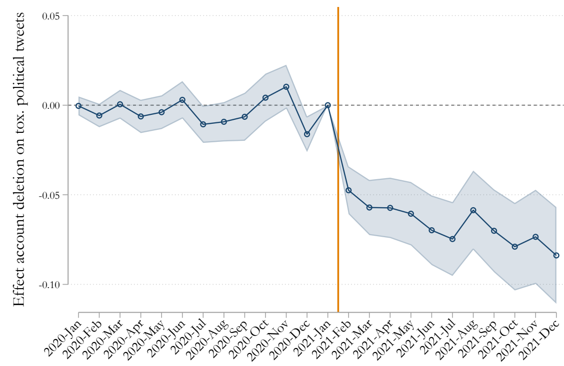
We present the event-study estimates from Equation (8) in Figure 4. The dependent variables are the number of political tweets (Panel a), toxic political tweets (Panel b), *Stop the Steal* tweets (Panel c). Two main patterns emerge. First, Trump supporters increasingly tweeted about politics in the run-up to the election, and consistent with our earlier findings, we observe a pronounced spike in *Stop the Steal* content immediately after the 2020 election. Second, all types of content exhibit a sharp relative decline in the aftermath of January 6$^{\text{th}}$. In other words, users with greater baseline exposure to Trump, and thus to election denial, show a large post-insurrection reduction in political content, toxic political content, and *Stop the Steal* messaging.

Figure 4: The Effect of Trump's Account Deletion on Online Content

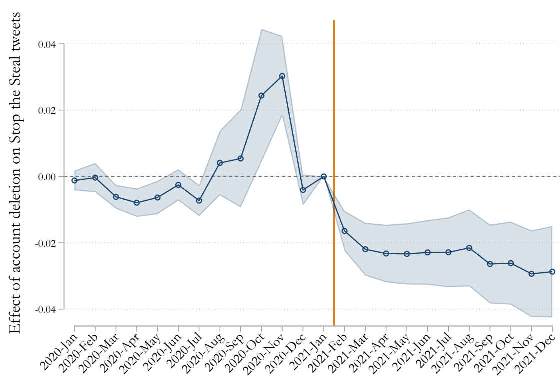
(a) Political Tweets



(b) Toxic Political Tweets



(c) *Stop the Steal* Tweets



Notes: These figures plot the estimates of an event study version of Equation (8). The dependent variable is the number of tweets as listed in the subcaption sent by user i in month t. The coefficient estimates measure the difference in tweets between Trump followers and a random sample of Twitter users around the time of Trump's account deletion on January 8$^{\text{th}}$ (two days after the January 6$^{\text{th}}$ insurrection). We plot 99% confidence intervals based on standard errors clustered by user.

We report the corresponding difference-in-differences estimates in Table 9. We present two specifications: one that includes user and month fixed effects along with a linear user-specific time trend, and a second that additionally incorporates user $\times$ month-of-year fixed effects to capture seasonality and a user-specific quadratic time trend. Across both specifications, the results closely mirror the event-study findings. The estimated effects are large in magnitude. The coefficient of -0.480 in Column 1 indicates that Trump followers sent 0.480 fewer political tweets per month after the account deletion, more than 2.5 times the pre-deplatforming average. Column 3 shows a similarly pronounced decline for toxic political tweets. For *Stop the Steal* tweets, the implied magnitudes are even larger, approximately six times the pre-period average. Note that the pre-period averages are calculated over all users, and Trump's Twitter followers are significantly more likely to produce such content (see Appendix Figure A.2).

Our takeaway from these estimates is that the January 6$^{\text{th}}$ insurrection marked a major inflection point in political expression on Twitter: both the frequency and toxicity of political tweets declined sharply, especially for *Stop the Steal* content. Given the scope of the enforcement actions against election denial, these declines likely reflect two forces. First, the mass removal of election-denial accounts reduced users' exposure to such content and thereby limited its propagation. Second, many users may have self-censored in response to heightened scrutiny and fear of repercussions (social or legal). For these reasons, the estimates should not be interpreted as evidence of changing support for the *Stop the Steal* movement, a distinction we return to in our analysis of offline outcomes.

Table 9: Trump's Account Deletion and Twitter Behavior

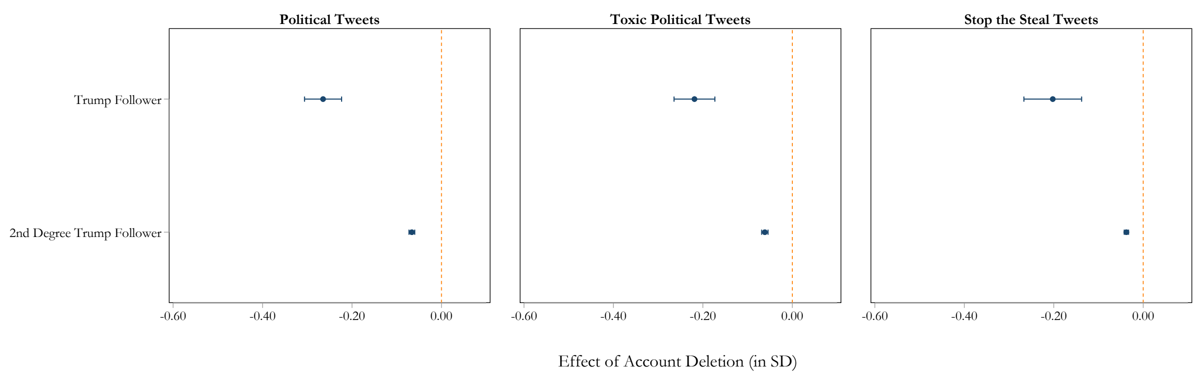
	Political Tweets		*Toxic Political Tweets*		*Stop the Steal Tweets*	
	(1)	(2)	(3)	(4)	(5)	(6)
Trump Follower × Account Deletion	-0.480***	-0.479***	-0.051***	-0.053***	-0.032***	-0.025***
	(0.032)	(0.036)	(0.005)	(0.006)	(0.004)	(0.003)
User FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
User Linear Time Trend	Yes	Yes	Yes	Yes	Yes	Yes
User × Month of Year FE		Yes		Yes		Yes
User Quadratic Time Trend		Yes		Yes		Yes
Observations	10,822,050	10,822,050	10,822,050	10,822,050	10,822,050	10,822,050
Pre-Period Mean of DV	0.186	0.186	0.019	0.019	0.005	0.005
R^2	0.42	0.68	0.38	0.65	0.19	0.60

Notes: This table presents the estimates from Equation (8), where the dependent variable is the number of tweets sent by a user in a given month. Standard errors are clustered by user. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

We next examine how political expression propagates through the network structure of social media platforms. In Figure 5, we report estimates from a modified version of Equation (8) that includes an interaction between the account-removal indicator and an indicator for users who do *not* follow Trump directly but do follow at least one account that follows him (i.e., second-degree Trump followers). The point estimates indicate substantial spillover effects. Across all tweet categories, Trump's deplatforming is associated with a significant reduction in activity even among these second-degree users. As expected, the magnitude of these effects is consistently smaller than for users directly exposed to Trump, underscoring the attenuation of influence as one moves further out in the network. Nonetheless, given the far larger number of second-degree followers, the overall role of these second-round effects on the Twitter network is substantial.

We also examine how the Twitter behavior of Trump followers varied by their estimated party affiliation. Although we do not directly observe users' partisanship, we can proxy for it using the set of politicians they follow on Twitter, a strategy also used in Mosleh and Rand (2022) and Renault and Stantcheva (2025). For this classification, we use follower lists of all U.S. Congress members of the 110th to 115th U.S. Congress, which enable us to quantify how many Democrats and Republicans each user follows.[10] The resulting classifications align with external evidence on the partisan composition of Twitter's user base at the time (Barberá, 2015). Consistent with prior findings, the vast majority of users for whom we can infer a political leaning are Democrats, by a ratio of roughly 8:1. Users who cannot be classified as either Republican or Democrat, because they follow no politicians or follow equal numbers from both parties, are labeled as "independents."

Figure 5: Network Effects of Trump's Account Deletion



Notes: These figures plot the point estimates from a regression specification akin to Equation (8) for different outcome variables Y. We include interactions of *Trump Follower* and *2nd Degree Trump Follower* with *Account Deletion* as independent variables in the same regression. Standard errors are clustered by user. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Using these estimates of user ideology, we report separate estimates by party affiliation in Table D.1. We observe by far the largest estimates for Republicans and somewhat smaller estimates for Independents. For Democrats, the coefficients are smaller and in many cases indistinguishable from zero. We interpret these results as evidence that changes in online behavior following the insurrection and deplatforming were mainly driven by users who were ex-ante more likely to support anti-democratic actions and were hence more heavily affected by the online crackdown.

[10] We collected this list in the run-up to the 2020 election.

Robustness Tables D.2 and D.3 present robustness checks using alternative variable transformations and different thresholds for defining toxicity. Across all specifications, the results consistently confirm that the insurrection and Trump's subsequent deplatforming were followed by substantial declines in toxic, *Stop the Steal* messaging among users with greater prior exposure to such content.

6.2 Crackdown on Election Denial and Offline Outcomes

The preceding results show that the January 6$^{\text{th}}$ insurrection was followed by a sharp contraction in online political expression. As discussed, these changes likely reflect both shifts in individuals' willingness to express such sentiments publicly and Twitter's direct removal of election-denial content. As such, the results are not directly informative about possible changes in beliefs. Thus, we now turn to whether the crackdown on election denial was accompanied by corresponding shifts in offline attitudes.

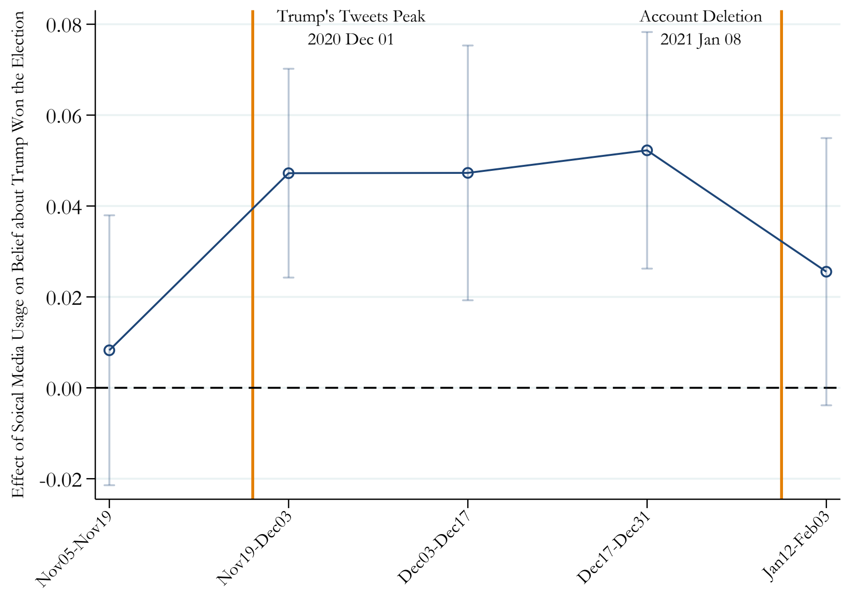
We begin by examining how the relationship between social media use and the belief that Trump won the 2020 election evolved over time. As shown in Section 4.6, this relationship is positive on average. Because the Nationscape survey provides repeated cross-sections, we can track this correlation across different periods. To do so, we estimate the relationship in two-week intervals and plot the resulting coefficients in Figure 6.

We find that in the first two weeks after the election, there is no detectable association between social media use and election denial. A positive relationship emerges only toward the end of November, coinciding with Trump's escalation of claims about election fraud and the growing traction of the *Stop the Steal* movement on Twitter. This correlation persists through early January, including the period of the January 6$^{\text{th}}$ attack. Following the insurrection, the association declines somewhat but remains positive and statistically significant at the 10% level.

Our interpretation is that the extreme events of January 6$^{\text{th}}$, together with Trump's subsequent deplatforming, may have modestly weakened anti-democratic sentiments. Crucially, however, any such shift is small relative to the dramatic collapse in toxic, anti-democratic content observed on Twitter. In other words, the sharp reduction in *public* expression, driven both by user self-censorship and by the platform's removal of election-denial posts, was not matched by a comparable change in *private* beliefs.

Because eliciting truthful beliefs is inherently difficult, and individuals may have become less willing to express their views even in anonymous surveys, we turn to donations to the *Save America PAC*. Donations are a costly and relatively private form of political support and therefore provide a credible measure of revealed preferences at any given point in time. By examining how the predictive power of (instrumented) social media usage for Save America

Figure 6: Twitter Usage Impact on Election Denial after Trump Account Deletion



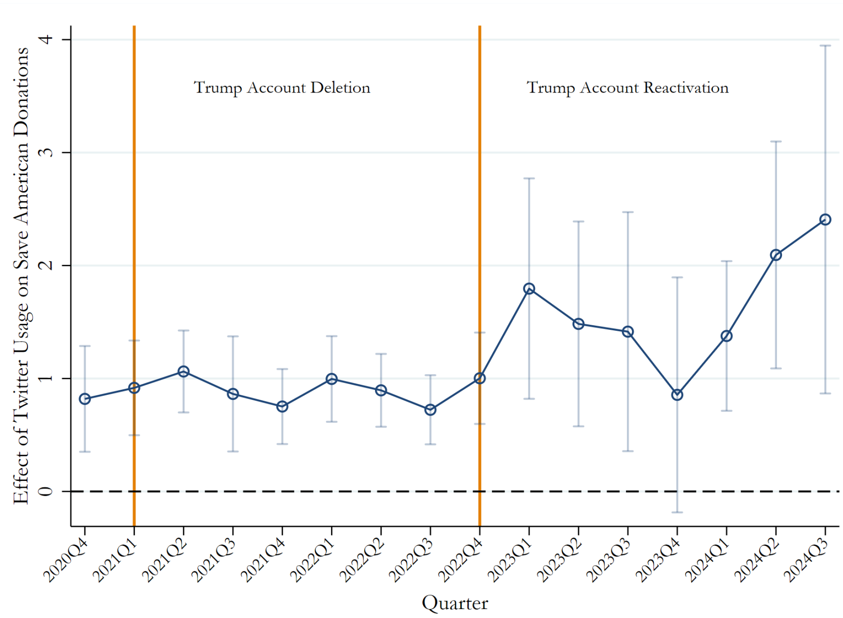
Notes: This figure plots the coefficients and 95% confidence intervals from regressions of election-denial beliefs on social media usage across five biweekly subsamples, labeled by their corresponding two-week periods. Each regression follows the same identification and control specification as column (1) in Table 7. The dependent variable is whether respondents believe Trump won the 2020 election, and the key independent variable indicates whether the respondent reported consuming political news via social media (e.g., Facebook, Twitter) in the past week. We use ten waves from the Nationscape Survey conducted between November 5, 2020, and February 3, 2021, aggregated into five two-week periods for the analysis. The vertical lines mark key events: Trump's election denial tweet peak on December 1, 2020, and his Twitter account deletion on January 8, 2021.

donations evolves over time, we can assess whether the insurrection, Trump's deplatforming, and the eventual reactivation of his account coincided with changes in support for anti-democratic causes.[11]

We plot the coefficients from separate instrumental-variables regressions of the form in Equation (2), where the dependent variable is the amount of donations to *Save America PAC* in a given quarter. The results are in Figure 7. In stark contrast to the collapse in *Stop the Steal* activity on Twitter, we observe no weakening of the relationship between Twitter usage and Save America donations following January 6th. The coefficient estimates remain remarkably stable from 2020 through the end of 2022 and are consistently highly statistically significant. This pattern indicates that, despite the dramatic decline in public anti-democratic expression online, the underlying private attitudes of individuals willing to contribute to the *Save America PAC* remained essentially unchanged.

[11]We cannot analyze the correlation between Trump's account reactivation and election-denial beliefs using Nationscape data because the survey ended in February 2021.

Figure 7: The Time-Varying Effect of Twitter Usage on Save America Donations



Notes: This figure plots coefficients and 95% confidence intervals from two-stage least squares (2SLS) regressions following Equation (2), estimated separately for each quarter. The dependent variable is the quarterly donation amount to the *Save America PAC* (in logs, with 1 added inside). The independent variable is Log(Twitter usage), instrumented with the number of SXSW followers who joined Twitter in March 2007. The vertical lines mark Trump's Twitter account deletion on January 8, 2021, and subsequent reactivation on November 19, 2022.

A notable inflection occurred with the reactivation of Trump's Twitter account in November 2022, following Elon Musk's acquisition of the platform. The estimated effect of Twitter usage on *Save America PAC* donations nearly *doubles* immediately afterward and remains elevated through 2024.[12]

Taken together, the temporal patterns in both election-denial beliefs and donations to anti-democratic causes carry important implications. Although the January 6th insurrection and the subsequent crackdown on election denial precipitated a dramatic collapse in toxic, partisan online content, these interventions had only limited effects on offline behavior. Support for the *Save America PAC* remained robust, and even increased in some periods, in areas with plausibly exogenous levels of social media penetration, persisting through 2024. These findings suggest that social media can entrench private anti-democratic beliefs and actions even when public expressions of these sentiments are curtailed.

[12] The dip in 2023Q4 likely reflects several headwinds faced by Trump during this period, including Save America's payment of approximately $50 million in legal fees, which may have dampened fundraising, and the indictment of six fake electors in Nevada for submitting false electoral certificates.

7 Conclusion

This paper provides comprehensive evidence that social media can be a powerful instrument for political elites seeking to mobilize followers toward anti-democratic action. Analyzing Donald Trump's use of Twitter during the aftermath of the 2020 U.S. presidential election, we show that online rhetoric translated into meaningful offline behavior, including political violence, culminating in the January 6th Capitol attack.

Our findings are consistent with three core mechanisms. First, social networks generate sizable spillovers: counties more connected on Twitter to areas that sent rioters were substantially more likely to contribute participants themselves. Second, social media reshaped political beliefs and revealed preferences, increasing adherence to false claims of election fraud and driving donations to the *Save America PAC*. Third, in a polarized environment, direct and unfiltered elite communication can trigger large online cascades that translate into offline mobilization: Trump's election-denial tweets produced immediate spikes in both online activity and costly political actions, and his deplatforming was followed by a collapse of anti-democratic online messaging.

Crucially, we also show that removing Trump and thousands of related accounts, while effective in reducing toxic online content, did not shift offline anti-democratic attitudes or donation behavior. The persistence of such beliefs despite deplatforming highlights the limits of content-removal strategies once misinformation and attitudes have already taken root.

Taken together, our results challenge the view that social media is inherently a "liberation technology" that empowers marginalized voices. Instead, social networks can magnify the influence of political elites, enabling them to bypass traditional gatekeepers, shape the information environment, and mobilize anti-democratic actions. The January 6th insurrection demonstrates that unregulated platforms can pose risks not only in authoritarian contexts but also within established democracies. While our case focuses on an extreme episode, the mechanisms we identify (i.e., elite amplification, network spillovers, and belief entrenchment) are likely to extend to other settings.

References

ABC News. 2020. "Trump's Fundraising for 'Save America' PAC Surges While Contesting Election Results." Accessed: 2025-10-04.

Acemoglu, Daron, Tarek A. Hassan, and Ahmed Tahoun. 2017. "The Power of the Street: Evidence from Egypt's Arab Spring." *The Review of Financial Studies*, 31(1): 1–42.

Adena, Maja, Ruben Enikolopov, Maria Petrova, Veronica Santarosa, and Ekaterina Zhuravskaya. 2015. "Radio and the Rise of The Nazis in Prewar Germany." *The Quarterly Journal of Economics*, 130(4): 1885–1939.

Allcott, Hunt, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow. 2020. "The Welfare Effects of Social Media." *American Economic Review*, 110(3): 629–76.

Allcott, Hunt, Matthew Gentzkow, Winter Mason, Arjun Wilkins, Pablo Barberá, Taylor Brown, Juan Carlos Cisneros, Adriana Crespo-Tenorio, Drew Dimmery, Deen Freelon, Sandra González-Bailón, Andrew M. Guess, Young Mie Kim, David Lazer, Neil Malhotra, Devra Moehler, Sameer Nair-Desai, Houda Nait El Barj, Brendan Nyhan, Ana Carolina Paixao de Queiroz, Jennifer Pan, Jaime Settle, Emily Thorson, Rebekah Tromble, Carlos Velasco Rivera, Benjamin Wittenbrink, Magdalena Wojcieszak, Saam Zahedian, Annie Franco, Chad Kiewiet de Jonge, Natalie Jomini Stroud, and Joshua A. Tucker. 2024. "The Effects of Facebook and Instagram on the 2020 Election: A Deactivation Experiment." *Proceedings of the National Academy of Sciences of the United States of America*, 121(21): e2321584121.

Anderson, Bryan. 2017. "Tweeter-in-Chief: A content analysis of President Trump's tweeting habits." *Elon Journal of Undergraduate Research in Communications*, 8(2): 36–47.

Andres, Raphaela, and Olga Slivko. 2021. "Combating online hate speech: The impact of legislation on Twitter." ZEW Discussion Papers.

Ang, Desmond. 2023. "The Birth of a Nation: Media and Racial Hate." *American Economic Review*, 113(6): 1424–1460.

Aridor, Guy, Rafael Jiménez-Durán, Ro'ee Levy, and Lena Song. 2024. "The economics of social media." *Journal of Economic Literature*, 62(4): 1422–1474.

Auxier, Brooke, Monica Anderson, et al. 2021. "Social media use in 2021." *Pew Research Center*, 1(1): 1–4.

Barberá, Pablo. 2015. "Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data." *Political Analysis*, 23(1): 76–91.

Beknazar-Yuzbashev, George, Rafael Jiménez-Durán, Jesse McCrosky, and Mateusz Stalinski. 2025. "Toxic content and user engagement on social media: Evidence from a field experiment."

Bernini, Andrea. 2023. "The Voice of Radio in the Battle for Equal Rights: Evidence from the U.S. South." *Economics & Politics*, 35(1): 163–226.

Bianchi, Francesco, Roberto Gómez-Cram, and Howard Kung. 2024. "Using social media to identify the effects of congressional viewpoints on asset prices." *The Review of Financial Studies*, 37(7): 2244–2272.

Bianchi, Francesco, Roberto Gómez-Cram, Thilo Kind, and Howard Kung. 2023. "Threats to central bank independence: High-frequency identification with twitter." *Journal of Monetary Economics*, 135: 37–54.

Boken, Johannes, Mirko Draca, Nicola Mastrorocco, and Arianna Ornaghi. 2023. "The Returns to Viral Media: The Case of US Campaign Contributions." *CEPR Discussion Papers.*

Boxell, Levi, and Zachary Steinert-Threlkeld. 2022. "Taxing dissent: The impact of a social media tax in Uganda." *World Development*, 158: 105950.

Brady, William J, Julian A Wills, John T Jost, Joshua A Tucker, and Jay J Van Bavel. 2017. "Emotion shapes the diffusion of moralized content in social networks." *Proceedings of the National Academy of Sciences*, 114(28): 7313–7318.

Buntain, Cody, Martin Innes, Tamar Mitts, and Jacob Shapiro. 2023. "Cross-platform reactions to the post-January 6 deplatforming." *Journal of Quantitative Description: Digital Media*, 3.

Bursztyn, Leonardo, Georgy Egorov, Ruben Enikolopov, and Maria Petrova. 2019. "Social Media and Xenophobia: Evidence from Russia." National Bureau of Economic Research Working Paper 26567.

Cao, Andy, Jason M Lindo, and Jiee Zhong. 2023. "Can social media rhetoric incite hate incidents? Evidence from Trump's "Chinese Virus" tweets." *Journal of Urban Economics*, 137: 103590.

Chen, Yuyu, and David Y. Yang. 2019. "The Impact of Media Censorship: 1984 or Brave New World?" *American Economic Review*, 109(6): 2294–2332.

Christensen, Darin, and Francisco Garfias. 2018. "Can You Hear Me Now? How Communication Technology Affects Protest and Repression." *Quarterly Journal of Political Science*, 13(1): 89–117.

Clarke, Isobelle, and Jack Grieve. 2019. "Stylistic variation on the Donald Trump Twitter account: A linguistic analysis of tweets posted between 2009 and 2018." *PloS one*, 14(9): e0222062.

Clayton, Katherine, Nicholas T Davis, Brendan Nyhan, Ethan Porter, Timothy J Ryan, and Thomas J Wood. 2021. "Elite rhetoric can undermine democratic norms." *Proceedings of the National Academy of Sciences*, 118(23): e2024125118.

Colella, Fabrizio, Rafael Lalive, Seyhun Orcan Sakalli, and Mathias Thoenig. 2019. "Inference with Arbitrary Clustering." IZA Discussion Papers 12584.

Diamond, Larry. 2010. "Liberation Technology." *Journal of Democracy*, 21(3): 69–83.

Dixon, Lucas, John Li, Jeffrey Sorensen, Nithum Thain, and Lucy Vasserman. 2018. "Measuring and Mitigating Unintended Bias in Text Classification." *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, 67–73.

Edmond, Chris. 2013. "Information Manipulation, Coordination, and Regime Change." *The Review of Economic Studies*, 80(4): 1422–1458.

Enikolopov, Ruben, Alexey Makarin, and Maria Petrova. 2020*a*. "Social media and protest participation: Evidence from Russia." *Econometrica*, 88(4): 1479–1514.

Enikolopov, Ruben, Alexey Makarin, and Maria Petrova. 2020*b*. "Social Media and Protest Participation: Evidence from Russia." *Econometrica*, 88(4): 1479–1514.

Enikolopov, Ruben, Alexey Makarin, Maria Petrova, and Leonid Polishchuk. 2017. "Social Image, Networks, and Protest Participation." *Universitat Pompeu Fabra*.

Fergusson, Leopoldo, and Carlos Molina. 2021. "Facebook Causes Protests." Universidad de los Andes - CEDE Documentos CEDE 018002.

Foa, Roberto Stefan, and Yascha Mounk. 2016. "The Danger of Deconsolidation: The Democratic Disconnect." *Journal of Democracy*, 27(3): 5–17.

Fujiwara, Thomas, Karsten Müller, and Carlo Schwarz. 2021. "The Effect of Social Media on Elections: Evidence From the United States." *NBER Working Papper*.

Fujiwara, Thomas, Karsten Müller, and Carlo Schwarz. 2024. "The Effect of Social Media on Elections: Evidence from The United States." *Journal of the European Economic Association*, 22(3): 1495–1539.

Gagliarducci, Stefano, Massimiliano Gaetano Onorato, Francesco Sobbrio, and Guido Tabellini. 2020. "War of the Waves: Radio and Resistance during World War II." *American Economic Journal: Applied Economics*, 12(4): 1–38.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift. 2020. "Bartik instruments: What, when, why, and how." *American Economic Review*, 110(8): 2586–2624.

González-Bailón, Sandra, David Lazer, Pablo Barberá, Meiqing Zhang, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Deen Freelon, Matthew Gentzkow, Andrew M. Guess, Shanto Iyengar, Young Mie Kim, Neil Malhotra, Devra Moehler, Brendan Nyhan, Jennifer Pan, Carlos Velasco Rivera, Jaime Settle, Emily Thorson, Rebekah Tromble, Arjun Wilkins, Magdalena Wojcieszak, Chad Kiewiet de Jonge, Annie Franco, Winter Mason, Natalie Jomini Stroud, and Joshua A. Tucker. 2023. "Asymmetric Ideological Segregation in Exposure to Political News on Facebook." *Science*, 381(6656): 392–398.

Green, Jon, William Hobbs, Stefan McCabe, and David Lazer. 2022. "Online engagement with 2020 election misinformation and turnout in the 2021 Georgia runoff election." *Proceedings of the National Academy of Sciences*, 119(34): e2115900119.

Gylfason, Gísli. 2023. "From Tweets to the Streets: Twitter and Extremist Protests in the United States." SSRN Manuscript 4551807. Available at `https://papers.ssrn.com/abstract=4551807`.

Habibi, Mahyar, Dirk Hovy, and Carlo Schwarz. 2024. "The Content Moderator's Dilemma: Removal of Toxic Content and Distortions to Online Discourse." *arXiv preprint arXiv:2412.16114*.

Holliday, Derek, Tyler Reny, Alex Rossell Hayes, Aaron Rudkin, Chris Tausanovitch, and Lynn Vavreck. 2021. "Democracy fund+ UCLA nationscape methodology and representativeness assessment." *University of California, Los Angeles. https://www. voterstudygroup. org/data/nationscape.*

Horta Ribeiro, Manoel, Justin Cheng, and Robert West. 2023. "Automated content moderation increases adherence to community guidelines." 2666–2676.

Jiménez-Durán, Rafael. 2023. "The economics of content moderation: Theory and experimental evidence from hate speech on Twitter." *George J. Stigler Center for the Study of the Economy & the State Working Paper*, , (324).

Jiménez Durán, Rafael, Karsten Müller, and Carlo Schwarz. 2022. "The Effect of Content Moderation on Online and Offline Hate: Evidence from Germany's NetzDG." *Available at SSRN.*

King, Gary, Jennifer Pan, and Margaret E. Roberts. 2013. "How Censorship in China Allows Government Criticism but Silences Collective Expression." *The American Political Science Review*, 107(2): 326–343.

Levitsky, Steven, and Daniel Ziblatt. 2018. *How Democracies Die.* New York:Crown Publishing Group.

Levy, Ro'ee. 2021. "Social Media, News Consumption, and Polarization: Evidence from a Field Experiment." *American Economic Review*, 111(3): 831–70.

Lewandowsky, Stephan, Michael Jetter, and Ullrich KH Ecker. 2020. "Using the president's tweets to understand political diversion in the age of social media." *Nature communications*, 11(1): 5764.

Lorentzen, Peter. 2014. "China's Strategic Censorship." *American Journal of Political Science*, 58(2): 402–414.

Manacorda, Marco, and Andrea Tesei. 2016. "Liberation Technology: Mobile Phones and Political Mobilization in Africa." C.E.P.R. Discussion Papers CEPR Discussion Papers 11278.

McCabe, Stefan D., Diogo Ferrari, Jon Green, David M. J. Lazer, and Kevin M. Esterling. 2024. "Post-January 6th Deplatforming Reduced the Reach of Misinformation on Twitter." *Nature*, 630(8015): 132–140.

Mosleh, Mohsen, and David G Rand. 2022. "Measuring exposure to misinformation from political elites on Twitter." *Nature Communications*, 13(1): 7144.

Müller, Karsten, and Carlo Schwarz. 2021. "Fanning the flames of hate: Social media and hate crime." *Journal of the European Economic Association*, 19(4): 2131–2167.

Müller, Karsten, and Carlo Schwarz. 2023. "From hashtag to hate crime: Twitter and antiminority sentiment." *American Economic Journal: Applied Economics*, 15(3): 270–312.

Müller, Karsten, Yuanyuan Pan, and Carlo Schwarz. 2023. "Social Media and Stock Market Participation." CEPR Discussion Papers.

Pierskalla, Jan H., and Florian M. Hollenbach. 2013. "Technology and Collective Action: The Effect of Cell Phone Coverage on Political Violence in Africa." *American Political Science Review*, 107(2): 207–224.

Qin, Bei, David Strömberg, and Yanhui Wu. 2024. "Social Media and Collective Action in China." *Econometrica*, 92(6): 1993–2026.

Qin, Bei, David Strömberg, and Yanhui Wu. 2017. "Why Does China Allow Freer Social Media? Protests versus Surveillance and Propaganda." *The Journal of Economic Perspectives*, 31(1): 117–140.

Renault, Y, and Stefanie Stantcheva. 2025. "Emotions and policy views."

Satyanath, Shanker, Nico Voigtländer, and Hans-Joachim Voth. 2017. "Bowling for Fascism: Social Capital and the Rise of the Nazi Party." *Journal of Political Economy*, 125(2): 478–526.

Senate, US. 2021. "Examining the US capitol attack: A review of the security planning, and response failure on January 6." *Washington, DC.*

Siegel, Alexandra A, Evgenii Nikitin, Pablo Barberá, Joanna Sterling, Bethany Pullen, Richard Bonneau, Jonathan Nagler, Joshua A Tucker, et al. 2021. "Trumping Hate on Twitter? Online Hate in the 2016 US Election Campaign and its Aftermath." *Quarterly Journal of Political Science*, 16(1): 71–104.

Statista. 2019. "End of the road for Trump's Twitter account." *https://www.statista.com/chart/19561/total-number-of-tweets-from-donald-trump/*, Accessed: 2025-06-20.

Steinert-Threlkeld, Zachary C, Delia Mocanu, Alessandro Vespignani, and James Fowler. 2015. "Online social networks and offline protest." *EPJ Data Science*, 4(1): 1–9.

Stock, James, and Motohiro Yogo. 2005. "Testing for Weak Instruments in Linear IV Regression." *Identification and Inference for Econometric Models*, , ed. Donald W.K. Andrews, 80–108. New York:Cambridge University Press.

Svolik, Milan W. 2019. "Polarization versus Democracy." *Journal of Democracy*, 30(3): 20–32.

The Guardian. 2023. "How the EU Digital Services Act Affects Facebook, Google, and Others." The Guardian, By Lisa O'Carroll. Accessed 2026-01-02.

The Wall Street Journal. 2022. "EU Lawmakers Approve Sweeping Digital Regulations." The Wall Street Journal, By Sam Schechner and Kim Mackrael. Accessed 2026-01-02.

The Wall Street Journal. 2023. "U.K. to Toughen New Social-Media Law, Threatening CEOs With Jail Time." The Wall Street Journal, Accessed 2026-01-02.

Thomas, Daniel Robert, and Laila A Wahedi. 2023. "Disrupting hate: The effect of deplatforming hate organizations on their online audience." *Proceedings of the National Academy of Sciences*, 120(24): e2214080120.

Tucker, Joshua A., Yannis Theocharis, Margaret E. Roberts, and Pablo Barbera. 2017. "From Liberation to Turmoil: Social Media And Democracy." *Journal of Democracy*, 28(4): 46–59.

Wang, Tianyi. 2021*a*. "Media, Pulpit, and Populist Persuasion: Evidence from Father Coughlin." *American Economic Review*, 111(9): 3064–92.

Wang, Tianyi. 2021*b*. "Waves of Empowerment: Black Radio and the Civil Rights Movement." University of Copenhagen Working Paper.

Wikipedia. 2025. "Social media use by Donald Trump." *https://en.wikipedia.org/wiki/Social_media_use_by_Donald_Trump#cite_note-BytheNumbers-2*, Retrieved June 20, 2025.

Wikipedia. 2026. "List of cases of the January 6 United States Capitol attack." *https://en.wikipedia.org/wiki/List_of_cases_of_the_January_6_United_States_Capitol_attack*, Accessed: 2026-01-06.

Wulczyn, Ellery, Nithum Thain, and Lucas Dixon. 2017. "Ex Machina: Personal Attacks Seen at Scale." *Proceedings of the 26th International Conference on World Wide Web*, 1391–1399.

Yanagizawa-Drott, David. 2014. "Propaganda and Conflict: Evidence from the Rwandan Genocide." *The Quarterly Journal of Economics*, 129(4): 1947–1994.

Zhuravskaya, Ekaterina, Maria Petrova, and Ruben Enikolopov. 2020. "Political Effects of the Internet and Social Media." *Annual Review of Economics*, 12.

Online Appendix

Social Media vs. Democracy: Evidence from the January 6$^{\text{th}}$ Insurrection

This Online Appendix consists of four parts.

1. Appendix A provides details on the data sources.
2. Appendix B provides additional cross-sectional results.
3. Appendix C presents additional results on the effects of Trump's tweets.
4. Appendix D presents additional results on temporal variation in the link between social media use and political actions around Trump's account deletion.

A Detail on Data Sources

A.1 Summary Statistics Cross-Section

Table A.1: Summary Statistics (County-Level)

	Mean	Std. Dev.	Min.	Median	Max.	N
Political outcomes and Twitter data						
Nr. Jan 6th Court Cases	0.47	1.47	0.00	0.00	24.00	3,108
Log(Save America Donations)	6.67	3.02	0.00	7.26	13.68	3,108
Republican two-party vote share (2016)	0.63	0.16	0.04	0.66	0.95	3,108
Change in Republican two-party vote share, 2000-2016	0.06	0.10	-0.33	0.07	0.45	3,065
Republican two-party vote share (2020)	0.65	0.16	0.05	0.69	0.96	3,067
Change in Republican two-party vote share, 2000-2020	0.08	0.11	-0.34	0.09	0.48	3,065
Log(Twitter users)	5.29	1.76	0.00	5.13	12.35	3,108
Log(SXSW followers, March 2007)	0.06	0.32	0.00	0.00	4.98	3,108
Log(SXSW followers, Pre)	0.02	0.18	0.00	0.00	3.61	3,108
Geographical controls						
Population density	261.27	1733.47	0.10	45.60	69468.40	3,108
Log(County area)	6.53	0.86	0.69	6.47	9.91	3,108
Distance from Austin, TX (in miles)	1450.64	612.61	5.04	1464.66	3098.88	3,108
Distance from Chicago (in miles)	1055.48	597.27	7.16	971.91	3040.38	3,108
Distance from NYC (in miles)	1643.93	908.30	6.48	1533.05	4191.67	3,108
Distance from San Francisco (in miles)	2766.26	921.38	41.11	2874.85	4565.01	3,108
Distance from Washington, DC (in miles)	1423.57	894.22	0.10	1280.27	3983.08	3,108
Demographic controls						
% aged 20-24	0.06	0.02	0.01	0.06	0.27	3,108
% aged 25-29	0.06	0.01	0.03	0.06	0.15	3,108
% aged 30-34	0.06	0.01	0.03	0.06	0.12	3,108
% aged 35-39	0.06	0.01	0.03	0.06	0.11	3,108
% aged 40-44	0.06	0.01	0.02	0.06	0.10	3,108
% aged 45-49	0.06	0.01	0.02	0.06	0.09	3,108
% aged 50+	0.39	0.07	0.11	0.39	0.75	3,108
Population growth, 2000-2016	0.06	0.18	-0.43	0.03	1.32	3,108
% white	0.77	0.20	0.03	0.84	0.98	3,108
% black	0.09	0.14	0.00	0.02	0.85	3,108
% native American	0.02	0.06	0.00	0.00	0.90	3,108
% Asian	0.01	0.02	0.00	0.01	0.37	3,108
% Hispanic	0.09	0.14	0.01	0.04	0.96	3,108
% unemployed	5.50	1.94	1.80	5.30	24.10	3,108
Socioeconomic controls						
% below poverty level	16.74	6.58	1.40	16.00	53.30	3,108
% employed in IT	0.01	0.01	0.00	0.01	0.21	3,108
% employed in construction/real estate	0.07	0.05	0.00	0.06	1.00	3,108
% employed in manufacturing	0.14	0.13	0.00	0.11	0.72	3,108
% adults with high school degree	34.77	7.07	7.50	35.20	54.80	3,108
% adults with college degree	21.89	3.81	8.40	21.80	35.60	3,108
% watching Fox News	0.26	0.01	0.23	0.26	0.30	3,107
% watching prime time TV	0.43	0.01	0.40	0.43	0.47	3,107
China shock controls						
Exposure to Chinese import competition	2.81	2.75	-0.63	2.18	43.08	3,107
Share of routine occupations	29.64	2.82	22.23	29.57	36.66	3,107
Average offshorability index	-0.50	0.47	-1.64	-0.56	1.24	3,107
Republican two-party vote share (2020)	0.65	0.16	0.05	0.69	0.96	3,067

Notes: This table presents descriptive statistics for the main estimation sample.

A.2 Nationscape Survey Data

Table A.2: Summary Statistics (Nationscape Survey)

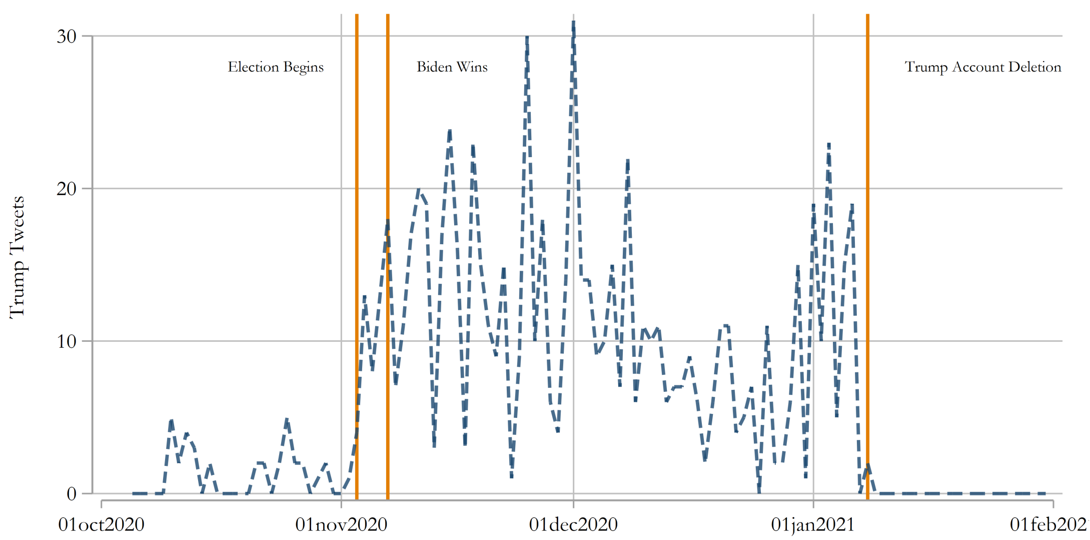
	Mean	Std. Dev.	Min.	Median	Max.	N
Nationscape Survey Variables						
Social Media User	0.70	0.46	0.00	1.00	1.00	178,336
Trump Won 2020 Election	0.24	0.43	0.00	0.00	1.00	62,151
Approve Rioters Jan 6th	0.10	0.30	0.00	0.00	1.00	11,648
Approve Trump Jan 6th	0.23	0.42	0.00	0.00	1.00	12,580
Non-Confident Fair Election	0.45	0.50	0.00	0.00	1.00	75,400
Concerned Twitter Ban	0.41	0.49	0.00	0.00	1.00	4,131

Notes: This table presents descriptive statistics for the estimation in Table 7 and Table B.4.

A.3 Twitter Data

Trump Tweets Data

Figure A.1: Trump *Stop the Steal* Tweets



Notes: This figure shows the daily frequency of Trump's *Stop the Steal* tweets over time from October 2020 to February 2021. Tweets are classified as related to the "Stop the Steal" or election fraud movement based on manual coding by research assistants. The tweet frequency begins increasing from October 10, 2020, peaks around the November 2020 election and January 6, 2021, and ends abruptly on January 8, 2021, when Trump's Twitter account was permanently suspended.

Table A.3: Description of Main Variables (Nationalscape)

	Variable	Time Coverage	Survey Prompt	Choices	Regression Meaning
1	Social Media User	phase3: 08/27/20–12/24/20; 01/12/21; parallel: 01/21/21	We're interested in where you might have heard news about politics in the past week. Have you seen or heard news about politics on any of the following outlets in the past week? — Social Media (e.g., Facebook, Twitter)	1=Yes; 0=No; .=Respondent Skipped	Dummy: 1=use; 0=not use
2	Trump won	phase3: 11/05/20–01/12/21; parallel: 01/21/21	Who do you think won the November 2020 Presidential election? — Selected Choice	1=Donald Trump; 2=Joe Biden; 3=Other; 999=Don't know; .=Respondent Skipped	Dummy: 1=Trump won; 0=others
3	Approve Capitol Rioters	phase3: 01/12/21; parallel: 01/21/21	Do you approve or disapprove of the actions of the people who stormed the U.S. Capitol?	1=Strongly approve; 2=Somewhat approve; 3=Somewhat disapprove; 4=Strongly disapprove; 999=Not sure; .=Respondent Skipped	Dummy: 1=approve; 0=others
4	Approve Trump Jan 6	phase3: 01/12/21; parallel: 01/21/21	Do you approve or disapprove of the actions of Donald Trump on January 6?	1=Approve; 2=Disapprove; 999=Not sure; .=Respondent Skipped	Dummy: 1=approve; 0=others
5	TrumpDidEnoughJan6	phase3: 01/12/21; parallel: 01/21/21	Do you think Donald Trump should have done more to end the violence at the Capitol?	1=Yes; 2=No; 999=Not sure; .=Respondent Skipped	Dummy: 1=No; 0=others
6	No Confidence Fair Election	phase3: 08/27/20–10/29/20	Thinking about the presidential election this November, how confident, if at all, are you that the election will be conducted fairly and accurately?	1=Very confident; 2=Somewhat confident; 3=Not too confident; 4=Not at all confident; .=Respondent Skipped	Dummy: 1=not confident ; 0=confident
7	No Confidence Able to Vote	phase3: 08/27/20–09/10/20; 10/29/20	Thinking about the presidential election this November, how confident, if at all, are you that all citizens who want to vote in the election will be able to?	1=Very confident; 2=Somewhat confident; 3=Not too confident; 4=Not at all confident; .=Respondent Skipped	Dummy: 1=not confident ; 0=confident
8	No Confidence Fair Election (Retro)	phase3: 11/05/20–01/12/21	Thinking about the presidential election in November 2020, how confident, if at all, are you that the election was conducted fairly and accurately?	1=Very confident; 2=Somewhat confident; 3=Not too confident; 4=Not at all confident; .=Respondent Skipped	Dummy: 1=not confident ; 0=confident
9	No Confidence Able to Vote (Retro)	phase3: 11/05/20–01/12/21	Thinking about the presidential election this November, how confident, if at all, are you that all citizens who want to vote in the election will be able to?	1=Very confident; 2=Somewhat confident; 3=Not too confident; 4=Not at all confident; .=Respondent Skipped	Dummy: 1=not confident ; 0=confident
10	Not Important Loser Concceds	phase3: 10/29/20	Thinking about the results of the presidential election this November, how important do you think it is for a losing candidate to publicly acknowledge the winner as the legitimate president of the country?	1=Not at all important; 2=Not too important; 3=Somewhat important; 4=Very important; 999=Not sure; .=Respondent Skipped	Dummy: 1=Not important ; 0=Important
11	Concerned Twitter Bans	phase3: 01/12/21	How worried are you that social media companies like Twitter and Facebook are going to ban content from people like you, or haven't you thought much about this?	1=Very worried; 2=Somewhat worried; 3=Not at all worried; 999=Haven't thought much about this; .=Respondent Skipped	Dummy: 1=Worried ; 0=Not worried

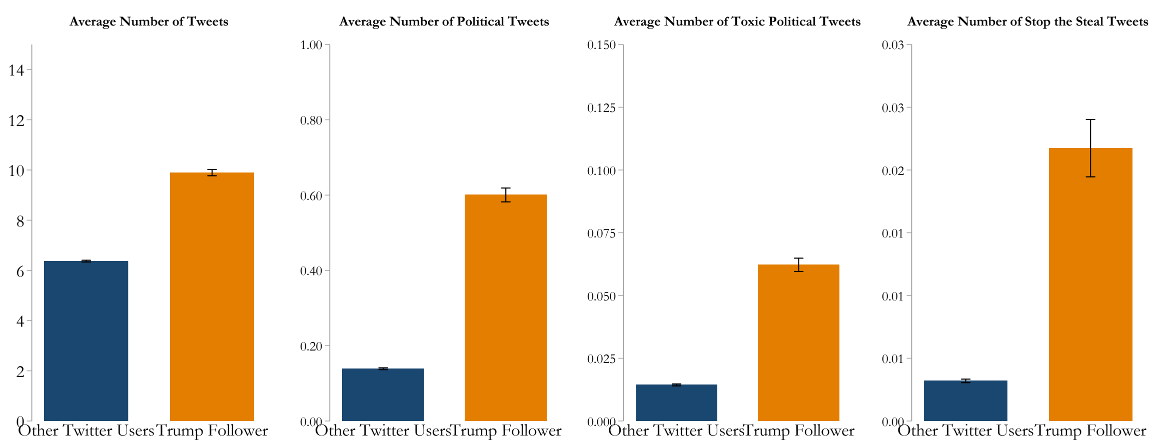
Notes: This table describes the main variables from the Nationscape survey used in our analysis. The Nationscape Project, conducted by the Democracy Fund and UCLA (Holliday et al., 2021), is one of the largest public opinion surveys during the 2020 U.S. election cycle, surveying approximately 6,250 respondents per week online between July 2020 and February 2021. Our analysis focuses on Phase 3 of the survey, which covers the post-election period from November 2020 to January 2021. The table reports the time coverage (survey waves), exact survey prompt or question text, available response choices, and the regression coding scheme for each variable. Variable definitions and question wording are drawn from the codebook files provided for each survey wave.

Representative Twitter User Data

Table A.4: Examples of Toxic Tweets

Tweet	Toxicity	Severe Toxicity	Identity Attack	Insult	Profanity	Threat
@RepAbaham You are an asshole too. Eat Trump's shit	0.99	0.87	0.62	0.99	0.98	0.25
RT @MasonWattles: Trump and Hillary are both shit, but trump is the better shit	0.96	0.81	0.64	0.95	0.97	0.27
@RandPaul You are the dumbest motherfucker	0.99	0.926	0.56	0.99	0.99	0.235
@laurenboebert Lol, goddamn are you a fucking imbecile	0.99	0.92	0.71	0.99	0.99	0.231
@Jim_Jordan Gym Jordan you're an asshole #gymjordanisanasshole	0.994	0.89	0.48	0.98	0.98	0.18

Figure A.2: Twitter Activity of Trump's Twitter Followers



Notes: This figure compares tweeting behavior between Trump followers (orange bars) and non-followers (blue bars) prior to Trump's account suspension. The five panels display: average tweets per user, average political tweets, average toxic political tweets, and average *Stop the Steal* tweets. Error bars represent 95% confidence intervals. Toxic content is identified using the Perspective API. Political and *Stop the Steal* content are classified using keyword filtering and machine learning methods.

Summary Statistics County-Day Level Panel

Table A.5: Summary Statistics (County-Day Panel)

	Mean	Std. Dev.	Min.	Median	Max.	N
Outcome Variables and Tweets Variables						
Number of Trump Tweets on StopSteal	8.90	7.41	0.00	7.00	31.00	261,072
Number of Stop the Steal tweets	0.01	0.23	0.00	0.00	45.00	261,072
Save America donations Amount	58.56	628.97	0.00	0.00	90076.75	261,072
Log(Twitter users)	5.29	1.76	0.00	5.13	12.35	261,072
Log(SXSW followers, March 2007)	0.06	0.32	0.00	0.00	4.98	261,072
Log(SXSW followers, Pre)	0.02	0.18	0.00	0.00	3.61	261,072

Notes: This table presents descriptive statistics for the estimation in Table 8, Table C.1, and Table C.2.

Summary Statistics User-level Panel Account Deletion

Table A.6: Summary Statistics: User-level Panel Account Deletion

Variable	Mean	SD	p50	Min	Max	N
Outcomes Variables						
Number of Tweets	7.42	39.98	0.00	0.00	3249.00	10,822,050
Political Tweet	0.17	2.69	0.00	0.00	1163.00	10,822,050
Toxic Political Tweet	0.02	0.37	0.00	0.00	121.00	10,822,050
Pro-Trump Tweet	0.00	0.20	0.00	0.00	281.00	10,822,050
Toxic Pro-Trump Tweet	0.00	0.03	0.00	0.00	25.00	10,822,050
User Variables						
Trump Follower	0.10	0.30	0.00	0.00	1.00	10,822,050
2nd Degree Trump Foll.	0.81	0.40	0.00	1.00	1.00	6,700,300

Notes: This table presents the mean, standard deviation, median, minimum, maximum, and number of observations of our main outcome variables, main variables of interest, and control variables for the full estimation sample.

B Additional Cross-sectional Evidence

B.1 Additional Evidence Instrumental Variable

Table B.1: Instrument Balancedness

	March 2007 *and* Pre (1)	March 2007 *only* (2)	Pre *only* (3)	Difference in means (2) - (3)	p-value	Šidàk p-value
Population density	5192.27	1021.39	1998.35	-976.96	0.07*	0.91
Log(County area)	6.30	6.63	6.54	0.09	0.73	1.00
Distance from Austin, TX (in miles)	1775.99	1749.38	1626.64	122.74	0.48	1.00
Distance from Chicago (in miles)	1439.45	1329.47	1214.42	115.05	0.53	1.00
Distance from NYC (in miles)	1685.31	1594.99	1510.05	84.94	0.78	1.00
Distance from San Francisco (in miles)	2751.83	2900.11	2833.01	67.10	0.83	1.00
Distance from Washington, DC (in miles)	1558.55	1450.23	1397.05	53.18	0.85	1.00
% aged 20-24	0.07	0.08	0.08	0.00	0.92	1.00
% aged 25-29	0.09	0.07	0.07	-0.00	0.51	1.00
% aged 30-34	0.08	0.07	0.07	-0.00	0.58	1.00
% aged 35-39	0.07	0.06	0.06	-0.00	0.82	1.00
% aged 40-44	0.06	0.06	0.06	0.00	0.82	1.00
% aged 45-49	0.07	0.06	0.06	0.00	0.89	1.00
% aged 50+	0.32	0.35	0.35	-0.00	0.97	1.00
Population growth, 2000-2016	0.18	0.18	0.15	0.03	0.56	1.00
% white	0.50	0.65	0.67	-0.02	0.62	1.00
% black	0.18	0.12	0.08	0.04	0.20	1.00
% native American	0.01	0.01	0.02	-0.02	0.02**	0.45
% Asian	0.10	0.05	0.05	-0.01	0.55	1.00
% Hispanic	0.20	0.16	0.15	0.01	0.80	1.00
% unemployed	4.86	5.05	4.51	0.54	0.07*	0.91
% below poverty level	15.71	15.82	13.69	2.14	0.17	1.00
% employed in IT	0.04	0.02	0.02	-0.00	0.98	1.00
% employed in construction/real estate	0.06	0.07	0.07	0.01	0.39	1.00
% employed in manufacturing	0.07	0.09	0.07	0.02	0.16	0.99
% adults with high school degree	21.76	25.99	25.77	0.22	0.88	1.00
% adults with college degree	18.89	21.16	21.20	-0.04	0.97	1.00
% watching Fox News	0.25	0.26	0.26	-0.00	0.91	1.00
% watching prime time TV	0.42	0.43	0.43	0.00	0.91	1.00
Exposure to Chinese import competition	2.55	2.46	2.79	-0.32	0.54	1.00
Share of routine occupations	32.47	31.38	31.25	0.13	0.82	1.00
Average offshorability index	0.37	-0.07	-0.05	-0.02	0.84	1.00
Republican two-party vote share (2020)	0.30	0.44	0.46	-0.02	0.58	1.00

Notes: This table presents the averages for the main control variables separately for the three types of counties relevant for our identification strategy: 1) the 47 counties with SXSW followers that joined Twitter both in March 2007 and the "pre-period"; 2) the 108 counties with SXSW followers that joined in March 2007 (but none in the "pre-period"); and 3) the 20 counties with SXSW that joined in the "pre-period" (but none in March 2007). The demographic and socioeconomic controls are measured in 2016. We report p-values from a two-sided t-test for the equality of means between the counties with the key identifying variation, as well as Šidàk-corrected values to account for multiple hypothesis testing. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

B.2 Robustness: January 6th and Save America Donations

Table B.2: Additional Robustness Checks

	Population regression weights (1)	Pre-period control polynomial (2)	Pre-period control deciles (3)	No zero Jan 6 counties (4)	No zero SXSW user counties (5)	No zero donation counties (6)
Panel A: January 6 Cases						
	Dep. Var.: Nr. Jan 6 Cases					
Log(Twitter users)	6.380***	3.212***	6.477***	3.072***	7.309***	3.077***
	(1.167)	(0.769)	(1.068)	(1.038)	(1.907)	(0.736)
Observations	3,064	3,064	3,064	609	158	2,666
Mean of DV	4.01	0.47	3.65	2.33	3.13	0.53
Robust F-stat.	91.08	61.82	75.54	58.72	29.84	37.79
Panel B: Save America Donations						
	Dep. Var.: log(Save America Donations)					
Log(Twitter users)	0.782***	0.697***	0.666***	0.957***	1.000***	0.657***
	(0.114)	(0.199)	(0.079)	(0.154)	(0.191)	(0.105)
Population deciles	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Pre-period Control Deciles			Yes			
Geographical controls	Yes	Yes	Yes	Yes	Yes	Yes
Demographic controls	Yes	Yes	Yes	Yes	Yes	Yes
Socioeconomic controls	Yes	Yes	Yes	Yes	Yes	Yes
China shock controls	Yes	Yes	Yes	Yes	Yes	Yes
2020 election control	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,064	3,064	3,064	609	158	2,666
Mean of DV	10.19	6.67	10.00	9.04	10.42	7.67
Robust F-stat.	91.08	61.82	75.54	58.72	29.84	37.79

Notes: This table presents additional robustness checks for the 2SLS results as in Equation (2). The dependent variable is the number of January 6th court cases in Panel A and log(Save America Donations) (with 1 added inside) in Panel B. Log(Twitter users) is instrumented using the number of users who started following SXSW in March 2007 (in logs with 1 added inside). Column 1 uses population regression weights. Column 2 includes a polynomial for pre-period SXSW Twitter users. Column 3 includes pre-period control deciles. Column 4 excludes counties with no January 6th cases. Column 5 excludes counties with no SXSW users. Column 6 uses spatial standard errors based on the method proposed in Colella, Lalive, Sakalli and Thoenig (2019), implemented in Stata as `acreg`, using a 200-mile cutoff. The regressions include population decile and state fixed effects, and controls as indicated. Standard errors in parentheses are clustered by state. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B.3: Robustness: Functional Form

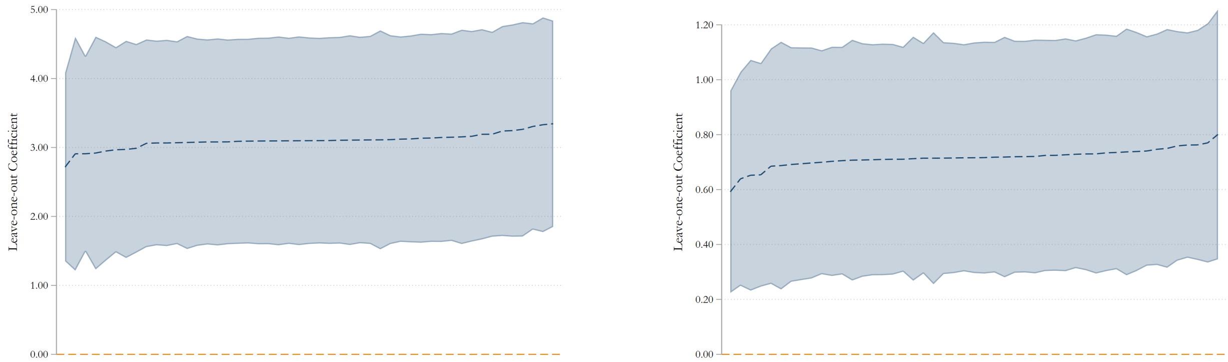
Dep. Var.:	*Nr. January 6th Court Cases*			*Save America Donations*		
	Nr. Cases (1)	I[Cases] (2)	asinh(Cases) (3)	Donations (4)	Nr.Donors (5)	asinh(Donations) (6)
Log(Twitter users)	3.097***	0.244***	0.677***	8.541***	201.234***	0.721***
	(0.743)	(0.079)	(0.166)	(1.822)	(42.035)	(0.224)
Population deciles	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Geographical controls	Yes	Yes	Yes	Yes	Yes	Yes
Demographic controls	Yes	Yes	Yes	Yes	Yes	Yes
Socioeconomic controls	Yes	Yes	Yes	Yes	Yes	Yes
China shock controls	Yes	Yes	Yes	Yes	Yes	Yes
2020 election control	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,064	3,064	3,064	3,064	2,667	3,064
Mean of DV	0.47	0.20	0.27	0.87	25.09	7.28
Robust F-stat.	50.14	50.14	50.14	50.14	52.76	50.14

Notes: This table examines the robustness of the relationship between Twitter usage and political mobilization to alternative functional form specifications of the dependent variables. Columns (1)-(3) analyze participation in the January 6$^{\text{th}}$ Capitol riot, while columns (4)-(6) analyze donations to the *Save America PAC*. Column 1 uses the number of court cases in levels. Column 2 uses a binary indicator equal to one if the county had any cases and zero otherwise. Column 3 uses the inverse hyperbolic sine (asinh) transformation of the number of cases. Column 4 uses total donation amounts in levels. Column 5 uses the number of individual donors. Column 6 uses the asinh transformation for total donations. The key explanatory variable is Log(Twitter users), measuring the natural logarithm of the number of Twitter users in each county. The regressions include population decile and state fixed effects and controls as indicated. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Figure B.1: Leave-one-Out Estimates

(a) Jan 6th Participation

(b) Save America Donations



Notes: This figure presents leave-one-out estimates based on the 2SLS specification in Equation (2). Panel (a) shows results for January 6th participation (number of court cases), while Panel (b) shows results for Save America donations (in logs, with 1 added inside). Each point represents the estimated coefficient on Log(Twitter users) when excluding one state from the sample. The shaded area represents the 95% confidence interval. Log(Twitter users) is instrumented using the number of users who started following SXSW in March 2007 (in logs with 1 added inside). All specifications include the same control variables as in column (5) of Table 3 and Table 4. Standard errors are clustered at the state level.

Table B.4: Additional January 6^{th}-Related Questions in Nationscape Survey

	Trump won 2020 Election (1)	Approve Rioters Jan 6 (2)	Approve Trump Jan 6 (3)	No-Confidence Fair Election (4)	Concerned Twitter Bans (5)
Social Media User	0.032***	0.033***	0.039***	0.016***	0.078**
	(0.006)	(0.008)	(0.012)	(0.006)	(0.030)
Cong. District FE	Yes	Yes	Yes	Yes	Yes
Gender	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes
Hispanic	Yes	Yes	Yes	Yes	Yes
Race	Yes	Yes	Yes	Yes	Yes
Income	Yes	Yes	Yes	Yes	Yes
Education	Yes	Yes	Yes	Yes	Yes
Observations	56,519	10,600	11,455	68,185	3,747
Mean of DV	0.22	0.09	0.21	0.44	0.42

Notes: This table presents regression estimates using data from ten waves of the Nationscape survey conducted weekly between November 5, 2020 and January 21, 2021. The dependent variables capture respondents' beliefs and attitudes: Column (1) measures belief that Trump won the 2020 presidential election; Column (2) measures approval of the actions of people who stormed the U.S. Capitol on January 6^{th}; Column (3) measures approval of Trump's actions on January 6^{th}; Column (4) measures confidence that the election was conducted fairly and accurately; Column (5) measures concern about social media companies banning content from people like the respondent. Exact question wording is provided in Appendix Table A.3. The key explanatory variable, Social Media User, equals one if respondents have seen or heard news about politics on social media (e.g., Facebook, Twitter) in the past week, and zero otherwise. All specifications include controls for gender, age, Hispanic ethnicity, race, income, education, and congressional district fixed effects. Standard errors in parentheses are clustered at the state level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

C Additional Evidence from Trump's Tweets

Table C.1: Falsification Tests – Panel Regressions Using Other Festivals

	SXSW (1)	Burning Man (2)	Coachella (3)	Pitchfork (4)	EDC (5)	ACL (6)
Panel A: Reduced Form: Nr.Stop the Steal Tweets						
	Dep. Var.: Log(Stop the Steal Tweets)					
Log(Followers Festival Month) × Log(Trump StopSteal Tweets)	0.018***	0.004	0.004	0.002	0.003	0.003
	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes
Population deciles	Yes	Yes	Yes	Yes	Yes	Yes
Log(Total donations)	Yes	Yes	Yes	Yes	Yes	Yes
Demographic controls	Yes	Yes	Yes	Yes	Yes	Yes
Socioeconomic controls	Yes	Yes	Yes	Yes	Yes	Yes
China shock controls	Yes	Yes	Yes	Yes	Yes	Yes
2020 election control	Yes	Yes	Yes	Yes	Yes	Yes
Observations	257,460	257,460	257,460	257,460	257,460	257,460
Panel B: Reduced Form: Save America Donations						
	Dep. Var.: Log(Save America Donations)					
Log(Followers Festival Month) × Log(Trump StopSteal Tweets)	0.052***	-0.017	0.002	0.015	0.000	0.018
	(0.013)	(0.016)	(0.014)	(0.010)	(0.010)	(0.014)
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes
Population deciles	Yes	Yes	Yes	Yes	Yes	Yes
Log(Total donations)	Yes	Yes	Yes	Yes	Yes	Yes
Demographic controls	Yes	Yes	Yes	Yes	Yes	Yes
Socioeconomic controls	Yes	Yes	Yes	Yes	Yes	Yes
China shock controls	Yes	Yes	Yes	Yes	Yes	Yes
2020 election control	Yes	Yes	Yes	Yes	Yes	Yes
Observations	257,460	257,460	257,460	257,460	257,460	257,460

Notes: This table presents falsification tests using other festivals in 2007 rather than SXSW as placebo treatments. Panel A presents reduced form results with the number of "Stop the Steal" tweets (in logs, with 1 added inside) as the dependent variable. Panel B shows similar results but with donations to the *Save America PAC* (in logs, with 1 added inside) as the dependent variable. The exposure variable "Followers Festival Month" measures the log number of users who started following each respective festival during their main month (with 1 added inside). Column 1 presents the baseline results using SXSW (March 2007) as exposure, while columns 2-6 test alternative festivals: Burning Man, Coachella, Pitchfork, Electric Daisy Carnival (EDC), and Austin City Limits (ACL). All regressions include the full set of control variables interacted with Trump tweets. Standard errors in parentheses are clustered by state.

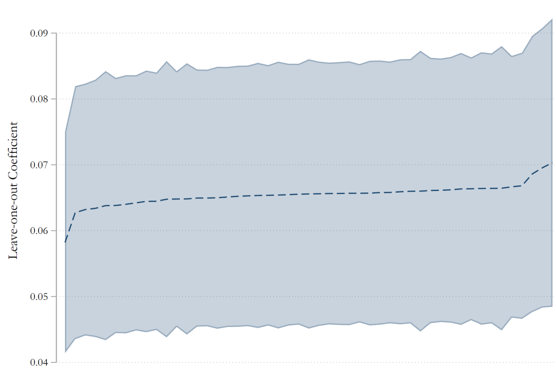
Table C.2: Robustness: Functional Form

	Dep. Var.: Nr.Stop Steal Tweets			*Dep. Var.: Save America Donations*		
	Nr. Cases	I[Cases]	asinh(Cases)	Donations	Nr.Donors	asinh(Donations)
	(1)	(2)	(3)	(4)	(5)	(6)
Log(Twitter users) × Log(Trump StopSteal Tweets)	0.161***	0.052***	0.085***	478.205***	3.299***	0.206***
	(0.033)	(0.008)	(0.013)	(115.042)	(0.726)	(0.045)
Log(SXSW followers, Pre) × Log(Trump StopSteal Tweets)	0.053	0.001	0.012	85.682	0.493	-0.031
	(0.035)	(0.006)	(0.011)	(103.157)	(0.689)	(0.030)
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes
Population deciles	Yes	Yes	Yes	Yes	Yes	Yes
Log(Total donations)	Yes	Yes	Yes	Yes	Yes	Yes
Demographic controls	Yes	Yes	Yes	Yes	Yes	Yes
Socioeconomic controls	Yes	Yes	Yes	Yes	Yes	Yes
China shock controls	Yes	Yes	Yes	Yes	Yes	Yes
2020 election control	Yes	Yes	Yes	Yes	Yes	Yes
Observations	257,460	257,460	257,460	257,460	257,460	257,460
Mean of DV	0.01	0.01	0.01	58.95	0.42	0.51
Robust F-stat.	43.25	43.25	43.25	43.25	43.25	43.25

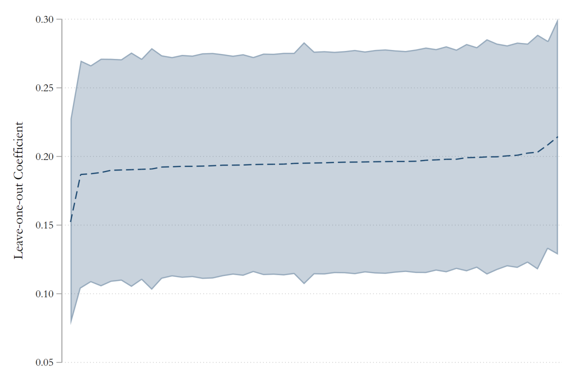
Notes: This table presents robustness checks examining the relationship between county-level exposure to Trump's "Stop the Steal" tweets and political outcomes using alternative functional form specifications. The dependent variable is the number of "Stop the Steal" tweets in columns (1)-(3) and Save America donations in columns (4)-(6). Column 1 uses levels. Column 2 uses a binary indicator for any activity. Column 3 uses the inverse hyperbolic sine (asinh) transformation. Column 4 uses donation levels. Column 5 uses the number of donors. Column 6 uses asinh transformation of donations. The key explanatory variable is Log(Twitter users) × Log(Num Trump Tweets on StopSteal), measuring exposure to Trump's "Stop the Steal" messaging. All specifications include county fixed effects, date fixed effects, population deciles, log(total donations), demographic controls, socioeconomic controls, China shock controls, and 2020 election controls. The robust F-statistic tests the significance of the interaction coefficient. Standard errors in parentheses are clustered at the state level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Figure C.1: Leave-one-Out Estimates

(a) Log(Number of Stop the Steal Tweets)



(b) Log(Save America Donations)



Notes: This figure presents leave-one-out estimates based on the 2SLS specification in Equation (7). The dependent variable is the number of "Stop the Steal" tweets (in logs, with 1 added inside) sent by users from each county in Panel (a) and Save America donations (in logs, with 1 added inside) in Panel (b). Each point represents the coefficient on the interaction term Log(Twitter users) × Log(Num Trump Tweets on StopSteal) when one state is excluded from the estimation. The dashed line indicates the baseline estimate using the full sample. The shaded area represents the 95% confidence interval. All specifications include county fixed effects, date fixed effects, population deciles, log(total donations), demographic controls, socioeconomic controls, China shock controls, and 2020 election controls. Standard errors are clustered at the state level.

D Additional Evidence on Trump's Account Suspension and the Crackdown on Election Denial

Table D.1: User Behavior Heterogeneity by Political Party

	Political Tweets		Toxic Political Tweets		Stop the Steal Tweets	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Republicans						
Trump Follower × Account Deletion	-0.870***	-0.702***	-0.068***	-0.063***	-0.115***	-0.090***
	(0.135)	(0.154)	(0.016)	(0.020)	(0.021)	(0.018)
User FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
User Linear Time Trend	Yes	Yes	Yes	Yes	Yes	Yes
User × Month of Year FE		Yes		Yes		Yes
User Quadratic Time Trend		Yes		Yes		Yes
Observations	296,825	296,825	296,825	296,825	296,825	296,825
Pre-Period Mean of DV	0.971	0.971	0.078	0.078	0.052	0.052
R^2	0.41	0.68	0.40	0.66	0.15	0.58
Panel B: Independents						
Trump Follower × Account Deletion	-0.230***	-0.223***	-0.028***	-0.028***	-0.010***	-0.007***
	(0.032)	(0.034)	(0.006)	(0.006)	(0.002)	(0.002)
User FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
User Linear Time Trend	Yes	Yes	Yes	Yes	Yes	Yes
User × Month of Year FE		Yes		Yes		Yes
User Quadratic Time Trend		Yes		Yes		Yes
Observations	8,068,500	8,068,500	8,068,500	8,068,500	8,068,500	8,068,500
Pre-Period Mean of DV	0.046	0.046	0.004	0.004	0.001	0.001
R^2	0.40	0.68	0.36	0.64	0.21	0.60
Panel C: Democrats						
Trump Follower × Account Deletion	-0.138***	-0.139**	-0.019**	-0.017*	0.000	0.004
	(0.049)	(0.058)	(0.008)	(0.010)	(0.002)	(0.003)
User FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
User Linear Time Trend	Yes	Yes	Yes	Yes	Yes	Yes
User × Month of Year FE		Yes		Yes		Yes
User Quadratic Time Trend		Yes		Yes		Yes
Observations	2,456,725	2,456,725	2,456,725	2,456,725	2,456,725	2,456,725
Pre-Period Mean of DV	0.552	0.552	0.061	0.061	0.013	0.013
R^2	0.42	0.67	0.38	0.65	0.28	0.63

Notes: This table presents the point estimates for β based on Equation (8), where the dependent variable is the number of toxic tweets sent by a user in a given month. Standard errors are clustered by user. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table D.2: Robustness: Variable Transformations

	Baseline	ln(1+Dep. Var)	asinh(Dep. Var)	I[Dep. Var>0]	Share of Dep. Var
	(1)	(2)	(3)	(4)	(5)
Panel A: Political Tweets					
Trump Follower × Account Deletion	-0.480***	-0.064***	-0.080***	-0.035***	-0.012***
	(0.032)	(0.002)	(0.003)	(0.001)	(0.000)
User FE	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes
User Linear Time Trend	Yes	Yes	Yes	Yes	Yes
Observations	10,822,050	10,822,050	10,822,050	10,822,050	10,822,050
Pre-Period Mean of DV	0.186	0.046	0.058	0.035	0.006
R^2	0.42	0.63	0.63	0.50	0.31
Panel B: Toxic Political Tweets					
Trump Follower × Account Deletion	-0.051***	-0.015***	-0.019***	-0.012***	-0.001***
	(0.005)	(0.001)	(0.001)	(0.001)	(0.000)
User FE	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes
User Linear Time Trend	Yes	Yes	Yes	Yes	Yes
Observations	10,822,050	10,822,050	10,822,050	10,822,050	10,822,050
Pre-Period Mean of DV	0.019	0.008	0.010	0.008	0.001
R^2	0.38	0.45	0.45	0.36	0.17
Panel C: Stop the Steal Tweets					
Trump Follower × Account Deletion	-0.032***	-0.010***	-0.013***	-0.009***	-0.000***
	(0.004)	(0.001)	(0.001)	(0.000)	(0.000)
User FE	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes
User Linear Time Trend	Yes	Yes	Yes	Yes	Yes
Observations	10,822,050	10,822,050	10,822,050	10,822,050	10,822,050
Pre-Period Mean of DV	0.005	0.002	0.003	0.003	0.000
R^2	0.19	0.26	0.26	0.22	0.10

Notes: This table presents the point estimates for β based on Equation (8), where the dependent variable is a measure of toxicity. Column 1 uses our baseline measure, the number of toxic tweets, defined as those with a Perspective API toxicity score of larger than 0.8. Column 2 uses the natural logarithm of the number of toxic tweets (with 1 added inside). Column 3 instead uses the inverse hyperbolic sine transformation, which naturally accommodates zero and negative values. Column 4 considers a dummy dependent variable equal to 1 for users who post at least one toxic tweet, and 0 otherwise. Column 5 uses the share of toxic tweets in all user posts. $Trump\ follower$ is an indicator variable equal to 1 for Twitter users who followed Trump before his account was suspended, and 0 otherwise. *Account Deletion* is an indicator variable equal to 1 for the months after the account deletion in January 2021. Standard errors are clustered by user. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table D.3: Robustness: Alternative Toxicity Thresholds

	Toxicity $\geq$ 0.5	Toxicity $\geq$ 0.6	Toxicity $\geq$ 0.7	Toxicity $\geq$ 0.8	Toxicity $\geq$ 0.9	Toxicity $\geq$ 0.95
	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var: Toxic Political Tweets						
Trump Follower × Account Deletion	-0.121***	-0.095***	-0.073***	-0.051***	-0.026***	-0.011***
	(0.010)	(0.008)	(0.006)	(0.005)	(0.003)	(0.001)
User FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
User Linear Time Trend	Yes	Yes	Yes	Yes	Yes	Yes
Observations	10,822,050	10,822,050	10,822,050	10,822,050	10,822,050	10,822,050
Pre-Period Mean of DV	0.044	0.035	0.027	0.019	0.010	0.004
R^2	0.39	0.39	0.38	0.38	0.37	0.34

Notes: This table presents the point estimates for β based on Equation (8), where the dependent variable is the number of toxic political tweets sent by a user in a given month. Toxic tweets are those with a Perspective API aggregate toxicity score larger than the cutoff in the top row. $Trump\ follower$ is an indicator variable equal to 1 for Twitter users who followed Trump before his account was suspended, and 0 otherwise. *Account Deletion* is an indicator variable equal to 1 for the months after the account deletion in January 2021. Standard errors are clustered by user. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.