Article

Subdued but Unbroken: The Cohesion of Far-Right Extremist Followers after Deplatforming

Wei Zhong and Maggie Mengqing Zhang

While deplatforming has become an increasingly common strategy to combat online harm and far-right extremism, its effects on the followers of extremist groups—who are key supporters and play a crucial role in spreading and sustaining these ideologies—remain underexplored. On August 10, 2018, Twitter (now X) deplatformed one such far-right extremist group, the Proud Boys, along with their affiliated accounts. Leveraging this intervention, our research addresses a key knowledge gap by examining the impact of deplatforming on the cohesion of extremist group followers. Specifically, we investigate whether deplatforming leads to fragmentation or reinforces unity among the group's followers. We assess cohesion through three theoretical lenses: task commitment, social commitment, and sense of belonging. By analyzing over 12 million tweets from approximately nine thousand Proud Boys supporters between August 1, 2017, and September 1, 2019, we find that deplatforming had a limited effect on reducing group cohesion. Instead, it may have prompted followers to seek broader networks and external interactions, leaving overall cohesion largely intact. This study offers important insights into the resilience of online extremist communities and the limitations of deplatforming as a strategy to disrupt them. Understanding these dynamics is essential for developing more effective approaches to counter online extremism and promote safer digital spaces.

The dynamic nature of social media has not only redefined how we communicate and share information but has also significantly influenced the landscape of political and societal discourse. While social media gives a voice to a wide array of individuals and groups, it has also inadvertently provided a platform and megaphone for those who promulgate far-right extremism (Zhang and Davis 2024). Extremism, broadly defined, refers to ideologically motivated beliefs and actions that reject democratic norms and seek to achieve political, social, or religious objectives through radical or coercive means (Jones 2022). Right-wing extremism, in particular, encompasses movements that promote racial, ethnic, or religious supremacy, oppose government authority, and resist progressive social policies (Hoffman 2017). These groups often position themselves as defenders of tradition against perceived societal decline, using narratives centered on national identity, cultural preservation, and opposition to perceived threats such as immigrants, minority groups, or political adversaries.

Psychological research highlights key factors that draw individuals to extremism. Van Prooijen and Krouwel (2019) identify traits such as cognitive simplicity (black-and-white thinking), belief in conspiracy theories, and a strong need for certainty and structure. These features make extremist ideologies appealing by offering clear, unambiguous answers to complex issues and by satisfying a psychological need for order and predictability. In the context of right-wing extremism, these psychological traits are further exploited and amplified by specific narratives.

Corresponding author: Wei Zhong (wz490@nyu.edu, United States) is a postdoctoral research scientist at the Center for Social Media and Politics at New York University. She holds a PhD in Political Science and an MS in Statistics from the University of Illinois Urbana-Champaign. Her research focuses on political communication, extremism, and platform governance, using computational methods to examine how digital platforms shape political behavior and group dynamics. She is an incoming assistant professor in the Department of Political Science at Stony Brook University, beginning in fall 2025.

Maggie Mengqing Zhang (ffx3ab@virginia.edu, United States) is a postdoctoral research associate at the McIntire School of Commerce, University of Virginia. She holds a PhD in Communication and Media and an MS in Computer Science from the University of Illinois Urbana-Champaign. Her research investigates artificial intelligence and social media through a sociotechnical systems lens, focusing on how information technologies impact individuals and society. She employs big-data analytics and experimental methods to uncover patterns and causal relationships in digital environments.

doi:10.1017/S1537592725101941

For instance, narratives of cultural and national decline heighten perceptions of societal threat, reinforce the group's adversarial identity, and deepen the appeal of extremist ideologies by aligning with individuals' preexisting cognitive and emotional vulnerabilities.

With the rise of digital platforms, extremism has evolved into "e-extremism," a decentralized network of online actors who leverage social media to spread conspiracy theories, white supremacist narratives, and anti-government ideologies. Unlike traditional extremist organizations with formal hierarchies, e-extremism thrives in loosely connected online communities, where users engage in ideological reinforcement, recruitment, and mobilization. These digital networks blur the boundaries between mainstream and fringe discourse, allowing extremist rhetoric to reach broader audiences while evading traditional forms of detection and suppression (Zhang and Davis 2024).

How to mitigate the potential harm posed by extremist groups has become a challenge to the entire society. In response, social media companies have adopted a strategy commonly known as "deplatforming" as a form of content moderation. Content moderation broadly refers to the governance mechanisms that regulate participation in online communities, ensuring compliance with platform policies while preventing abuse and misinformation (Grimmelmann 2015). This includes a range of enforcement actions, from flagging and restricting content to more severe measures like deplatforming (West 2018). Deplatforming, the most extreme form of content moderation, involves the removal of an account, page, or group for violating a platform's terms of service or community standards (Innes and Innes 2023). It serves as a form of social control aimed at curbing hate speech, disinformation, and other policy-violating behaviors (Moynihan 2021). Deplatforming can take various forms, ranging from temporary suspensions lasting from hours to weeks to permanent removals (West 2018). Platforms enforce these actions through a combination of automated detection systems, user-driven reporting, and human moderation (Iqbal et al. 2022). By prohibiting problematic users and groups from utilizing their services, social media companies aim to limit the reach and influence of harmful actors (Jhaver et al. 2021).

Over recent years, major platforms like Facebook, Instagram, YouTube, and Twitter (now X) have suspended or expelled a diverse array of individuals and groups, including white nationalists, anti-Semites, alt-right followers, neo-Nazis, and hate groups (Kraus 2018). Many deplatformed figures belong to the extreme right of the ideological spectrum, with some being prominent internet personalities. Notable cases include Milo Yiannopoulos and Alex Jones, whose removal significantly diminished their public visibility, fan base, and financial support. Yiannopoulos, for instance, cited deplatforming as a key factor in his bankruptcy, following the loss of a book deal and the cancellation of college campus events (Beauchamp 2018; Maurice 2019). Similarly, Jones experienced a sharp decline in viewership and the perceived influence of his content (Wong 2018).

Despite the increasing use of deplatforming as a content moderation strategy and a growing body of research on the subject, its broader consequences remain insufficiently understood—particularly its impact on the followers of deplatformed entities. Existing research has examined the direct effects of deplatforming on prominent far-right influencers, assessing declines in their visibility, engagement, and revenue following removal from major platforms (Beauchamp 2018; Wong 2018). More recently, studies have also begun to explore the deplatforming of everyday users and broader online communities, shedding light on how deplatforming impacts rank-and-file members, content creators, and marginalized groups (Are and Briggs 2023; Mekacher, Falkenberg, and Baronchelli 2023; Vu, Hutchings, and Anderson 2024). However, prior studies have largely focused on deplatforming's direct effects on platform-wide discourse, such as its effectiveness in reducing hate speech and online toxicity (Jhaver et al. 2021), or on tracking where deplatformed groups and individuals migrate post-removal (Rogers 2020). These approaches often emphasize general public reactions rather than examining how deplatforming affects the internal structure of extremist groups and their digital resilience.

A critical gap remains in understanding how deplatforming affects the cohesion and adaptation of extremist digital networks, particularly among followers who continue to engage in content dissemination, recruitment, and mobilization. Given that extremist movements do not rely solely on high-profile figures (Jasko and LaFree 2020), further research is needed to assess whether deplatforming disrupts or inadvertently strengthens these networks. This shift in focus is crucial because the followers of extremist groups are not mere bystanders; they form the core network that sustains these entities. Their role is critical in spreading ideological and political doctrines, significantly amplifying the influence of extremist groups (Awan 2017; O'Callaghan et al. 2013). Moreover, they serve as vital channels for recruitment, drawing new members to the cause (Chatfield et al. 2015; Malthaner 2018). Importantly, solidarity among followers is key to preserving the group's unity and operational longevity, ensuring that the ideological essence of the group persists even in the absence of leaders or figureheads (Benard 2012).

To address this gap, our study shifts the focus to the cohesion of deplatformed groups' followers, specifically investigating whether the removal of official accounts weakens their social ties or instead strengthens solidarity among remaining members. By examining these overlooked dynamics, we provide new insights into how deplatforming shapes extremist networks and their ability to persist in online spaces.

In this study, we assess the effects of Twitter's August 10, 2018, removal of accounts linked to the Proud Boys, a group widely recognized for promoting far-right and neofascist ideology. The Proud Boys, founded in 2016 by media figure Gavin McInnes, describe themselves as "Western chauvinists" (McBain 2020). They have been designated as a hate group by the Southern Poverty Law Center and identified by the Federal Bureau of Investigation as having connections to white nationalist movements (Rosenberg 2018; Wilson 2018a). The group has been associated with multiple violent incidents and arrests, including their involvement in the 2017 Unite the Right rally in Charlottesville and the University of California, Berkeley, protests (MacFarquhar 2021). These characterizations and events have been well documented in prior research (Zhong et al. 2024).

This group has cultivated a strong digital footprint, strategically using social media platforms such as Twitter, YouTube, and Facebook to spread propaganda, engage in recruitment and coordination, disseminate ideology, and maintain group cohesion (DeCook 2018; Hatmaker 2018; Reid, Valasik, and Bagavathi 2020). This dual reliance on online platforms and offline mobilization makes them a particularly relevant case for studying the effects of deplatforming. Unlike groups that operate primarily in digital spaces, the Proud Boys actively translate their online networks into real-world actions, including street-level confrontations and organized rallies (Bailard et al. 2024; Reid and Valasik 2020). Twitter played a crucial role in this process, enabling the group to circulate calls for action and amplify their rhetoric while navigating content moderation policies to avoid immediate removal (Klein 2019).

Given their hybrid structure, deplatforming raises important questions about whether removing the group's main accounts disrupts their online cohesion or merely fragments their network. If deplatforming weakens their digital network and cohesion, it may hinder their ability to maintain unity and support future offline actions. Conversely, if a sense of persecution strengthens cohesion, deplatforming may have limited effects in curbing their influence. Our research examines this dynamic, focusing on whether Twitter's deplatforming of the Proud Boys led to fragmentation among their followers or, conversely, reinforced their unity. By assessing the response of their online follower network, we contribute to a broader understanding of how extremist groups adapt to platform restrictions and whether deplatforming effectively disrupts their capacity for both digital and real-world mobilization.

Our analysis focuses on three key dimensions of cohesion among extremist followers: task commitment, social commitment, and sense of belonging. First, we examine task commitment, which reflects the group's collective effort in content creation and their unified dedication to achieving shared goals. Second, we evaluate social commitment by analyzing both direct interactions between members, such as retweets, and their indirect collaborations, such as collective information sharing through co-shared hashtags and links. These interactions reveal the strength of their supportive networks and active participation. Third, we assess the sense of belonging, which measures the depth of shared identity and individual members' identification with the group. By exploring these dimensions, we aim to understand how deplatforming affects the internal cohesion of extremist networks.

Related Works

Deplatforming

The effectiveness of deplatforming as a content moderation strategy to curb harmful online behaviors has increasingly become a focal point in recent research. Studies have shown that deplatforming can reduce the visibility and engagement of extremist content by disrupting online communities and limiting their ability to reach broader audiences. For instance, Reddit's removal of subreddits like r/fatpeoplehate and r/coontown effectively curtailed hate speech, leading to a sustained decline in user engagement, with many members ceasing participation altogether or failing to reestablish their communities elsewhere (Chandrasekharan et al. 2017; Saleem and Ruths 2018). Similarly, quarantining offensive communities on Reddit—which restricts access rather than outright banning them—was found to hinder recruitment efforts and limit the growth of harmful groups (Chandrasekharan et al. 2022). Research on individual influencers further supports these findings, as the deplatforming of figures like Alex Jones and Milo Yiannopoulos resulted in a substantial decline in discussions about them on Twitter, as well as a reduction in the posting activity and toxicity levels of their supporters (Jhaver et al. 2021). Additionally, Thomas and Wahedi (2023) demonstrate that Facebook's removal of hate-based organizations reduced both the consumption and production of hateful content among peripheral members, reinforcing the argument that deplatforming disrupts harmful online ecosystems.

While deplatforming often reduces toxic activity on mainstream platforms, studies also highlight its unintended consequences and limitations. Research has shown that deplatformed users frequently migrate to alternative platforms, where they attempt to rebuild their networks. For example, Parler's deplatforming by Google, Apple, and Amazon, which hosted the social media platform's app and website, after the US Capitol

https://doi.org/10.1017/S1537592725101941

riot led to increased activity on other fringe platforms like Gab and Rumble, raising concerns about the strengthening of alternative ecosystems (Horta Ribeiro et al. 2023). Similarly, Rauchfleisch and Kaiser (2021) find that although YouTube's removal of far-right channels limited the spread of disinformation, alternative platforms could not fully compensate for the loss of mainstream visibility. In some cases, deplatforming has also led to financial gains and increased engagement on fringe platforms, as seen when YouTube content creators moved to BitChute and experienced a surge in bitcoin donations (Klinenberg 2024). Mekacher, Falkenberg, and Baronchelli (2023) also document the migration of banned Twitter users to Gettr, noting that while these users exhibited lower toxicity on Gettr, they remained ideologically aligned with far-right movements. Additionally, broader network disruptions, such as the takedown of the online forum Kiwi Farms, show that even sustained and coordinated deplatforming efforts may only partially succeed. Despite losing many casual users, the forum's core members remained active, new users with heightened toxicity emerged, and the community ultimately reconstituted itself across alternative platforms (Vu, Hutchings, and Anderson 2024). These findings suggest that while deplatforming can be effective in disrupting harmful networks, its long-term impact depends on the resilience of core members, the availability of alternative platforms, and the consistency of enforcement across multiple digital ecosystems.[1]

Social Media Followers

In both digital and organizational contexts, followers have extended their influence beyond hierarchical structures, actively shaping their environments and fostering shared community identities. Traditional organizational theory viewed followers as subordinates who supported and responded to leadership, thereby reinforcing hierarchical power structures (Achua and Lussier 2013; Kelley 1988). However, this perspective has evolved, with scholars recognizing that followers actively shape decision making, reinforce group cohesion, and even assume leadership roles when necessary (Paunova 2015; Uhl-Bien et al. 2014). Theories of shared leadership emphasize that leadership is a dynamic process in which followers contribute to group direction and success (Shamir 2007; Uhl-Bien and Maslyn 2000).

Insights from organizational theory are highly applicable to digital spaces, where social media followers exhibit behaviors that mirror those in traditional organizations. Chaleff (1995) distinguishes between passive subordinates and active followers in offline contexts, a distinction that aligns with social media users ranging from silent observers to active participants shaping discussions. Similarly, Kelley's (1992) classification of followers—exemplary (highly engaged), pragmatic (context dependent), and passive (minimally participatory)—provides a framework for understanding varying levels of engagement online. In both settings, followers play a proactive role in sustaining group cohesion and driving collective action. Moreover, the fluidity of leadership and followership roles, highlighted by Newstrom (2000) in offline organizations, is even more pronounced in digital spaces, where users fluidly shift between consuming, engaging with, and initiating discussions in real time (Carsten et al. 2010; Sy 2010).

The unique affordances of digital environments have further amplified followers' influence, making them central to the flow of information and the shaping of public discourse. Social media platforms decentralize power, enabling users to collectively determine visibility, spread narratives, and mobilize around shared ideologies in real time (Dolan et al. 2016; Kapoor et al. 2018). This participatory nature fosters cohesion comparable to or stronger than traditional organizations, as digital communities can rapidly adapt to external disruptions and sustain collective action (Bliuc et al. 2020). Such dynamics position social media followers as critical actors in digital networks, actively shaping narratives, political movements, and collective action within an increasingly networked information landscape.

Therefore, followers are not peripheral figures but pivotal actors in sustaining and propagating extremist ideologies. Their cohesion underpins group resilience and impact, making them more resistant to external disruptions like deplatforming. To fully understand extremism, it is essential to examine the role of followers and the dynamics of their group cohesion in depth.

Group Cohesion

Social Identity Theory and Group Cohesion. Social identity theory (SIT), developed by Tajfel and Turner (1979), provides a foundational framework for understanding how individuals derive their self-concept from membership in social groups. At its core, SIT posits that people categorize themselves and others into social groups (e.g., nationality, political affiliation, or organizational membership) and strive to maintain a positive social identity by favorably comparing their in-group to relevant out-groups. This process, known as in-group favoritism, is driven by two key mechanisms: self-categorization (seeing oneself as part of a group) and social comparison (evaluating one's group relative to others). Through these mechanisms, individuals enhance their self-esteem by aligning with groups that provide a sense of belonging and positive distinctiveness (Brown 2000; Tajfel and Turner 1979). Importantly, in-group favoritism is not limited to high-status groups; even low-status or marginalized groups can exhibit this bias as a way of asserting their identity and resisting external threats (Ellemers, Spears, and Doosje 2002).

https://doi.org/10.1017/S1537592725101941

A key outcome of these social identity processes is group cohesion, or emotional and psychological bonds that hold a group together. When individuals self-categorize and exhibit in-group favoritism, they feel more connected to the group and more motivated to work toward its goals. This dynamic is particularly evident in the context of external threats, such as exclusion or marginalization, which can activate social identity processes and strengthen cohesion (Huddy 2001). For example, when a group faces external challenges—such as deplatforming—its members may perceive the threat as an attack on their identity. This perception can intensify in-group favoritism and self-categorization, reinforcing their sense of belonging and solidarity. As a result, external interventions like deplatforming may paradoxically strengthen group cohesion rather than dismantle it (Bornstein 2003; Bourhis et al. 1979; Branscombe et al. 1999).

Thus, SIT provides a valuable framework for understanding how external interventions, such as deplatforming, impact group cohesion. By activating social identity processes like in-group favoritism and self-categorization, deplatforming can deepen members' commitment to the group and enhance their collective identity, even in the face of external pressures.

Conflict Studies and Group Cohesion. The emphasis on group cohesion has also resonated in the broader field of conflict studies, which has extensively explored how external conflict and internal group dynamics shape cohesion (Coser 1956; Sherif 1966; Stein 1976; Sumner 1906). Research has demonstrated that conflicts between groups strengthen internal unity by fostering a sense of shared struggle, increasing members' positive perceptions of their group, and enhancing their emotional bonds (Blake, Shepard, and Mouton 1964; Bornstein 2003; Brewer 2001; Sherif 1966). Evans and Dion (2012) further highlight how collective victories reinforce group identity and efficacy, strengthening the group's cohesion and willingness to pursue shared objectives. Additionally, conflict clarifies in-group and out-group distinctions, reinforcing identity and mobilization against a perceived adversary (Schmid and Muldoon 2015; Stott and Drury 2017).

Beyond intergroup conflict, internal socialization mechanisms also sustain cohesion within extremist and armed groups. Cohen (2013; 2017) argues that violence functions as a key socialization tool, particularly for forcibly recruited members who lack preexisting ties. Participation in collective violence fosters trust, loyalty, and shared experiences, replacing voluntary social bonds with enforced cohesion. Gates (2002) examines how rebel groups maintain unity through ideological indoctrination, ethnic solidarity, and economic incentives, noting that those operating far from government control tend to be more cohesive, while those closer to state forces fragment more easily. Checkel (2017) further notes that both formal indoctrination and peer reinforcement contribute to long-term ideological commitment, with initiation rituals and shared hardships deepening internal bonds.

Applying these concepts to Twitter's deplatforming policy, the removal of extremist groups from digital spaces can be understood as an external conflict that reinforces internal cohesion among supporters. For the Proud Boys, known for their violent street confrontations and far-right ideology (Kutner 2020), deplatforming is not just a regulatory action but also a direct challenge to their identity and organizational structure. After their official accounts and leadership presence were removed, socialization could continue informally, as Checkel (2017) highlights, through peer reinforcement and sustained engagement with ideological content, allowing group norms to persist despite deplatforming.

Thus, rather than dismantling the Proud Boys' network, deplatforming may instead activate the very mechanisms that sustain it. By fostering a sense of collective struggle and reinforcing their adversarial identity, deplatformed groups may rally together and strengthen their ideological commitments. This understanding, drawn from conflict studies, provides a valuable framework for analyzing how the Proud Boys and similar extremist groups adapt, reorganize, and sustain their networks despite platform interventions.

Measuring Group Cohesion: Task Commitment, Social Commitment, and Sense of Belonging. Measuring cohesion is challenging due to its multifaceted nature. To address its complexities, scholars such as Carron, Widmeyer, and Brawley (1985) and Festinger, Schachter, and Back (1963) have introduced the task dimension as a critical aspect of cohesiveness, emphasizing the bonds formed through shared commitment to group objectives, which may include both concrete tasks and the promotion of common ideologies and narratives (Knott and Lee 2020). In ideological or extremist communities, cohesion is often rooted not in accomplishing tangible tasks but in sustaining and propagating a collective worldview (Mellor and Shilling 2010; Youngblood 2020). This highlights how task commitment can extend beyond physical actions to the reinforcement of shared beliefs and identities. Crucially, cohesion is not just about having common goals but also about the collaborative process of advancing them. The collective effort involved in promoting ideological narratives can be as significant as traditional task-based collaboration in strengthening group bonds. Jenson (2010) further argues that cohesion intensifies when communities develop shared values, confront common challenges, and construct collective interpretations. Task commitment, therefore, is a key expression of cohesion, whether through goal-directed collaboration on concrete tasks or through the continuous reinforcement of ideological narratives that unify the group.

https://doi.org/10.1017/S1537592725101941

Expanding on these concepts, Carron, Widmeyer, and Brawley (1985) and Seashore (1954) explore the social dimension of cohesion, often referred to as social commitment, emphasizing the intricate network of interpersonal connections that sustain group cohesion. Within this framework, interaction and communication play a central role, as members engage in discussions, exchange viewpoints, and acknowledge one another's perspectives constructively. The frequency and depth of these interactions help to strengthen social bonds, fostering a shared sense of identity and belonging among group members. This dimension not only underscores the importance of relationships within the group but also illustrates how these connections serve as the foundation for collective unity and strength. For instance, Bliuc and colleagues (2019) illustrate the importance of social cohesion through the practice of members contributing to the same discussion thread, showcasing how such interactions are instrumental in reinforcing the group's cohesion. This example underlines the functional significance of these exchanges in fortifying the group's collective unity.

Cohesiveness is also deeply intertwined with a profound sense of belonging, playing an instrumental role in fostering and preserving communal ties (Tsai, Yang, and Cheng 2014). From a SIT perspective, belonging emerges through self-categorization and in-group identification, whereby individuals define themselves as part of a group and internalize its values and norms (Tajfel and Turner 1979). Unlike task commitment, which is rooted in action and goal-oriented collaboration, a sense of belonging stems from emotional connection and group identity. Members develop a deep attachment to the group not just through participation but also through shared values, beliefs, and a collective purpose that reinforce their social identity (Beal et al. 2003). Moreover, belonging is not merely a subjective feeling but a structural component of group cohesion. Chan, Chiu, and Chiu (2010) emphasize that it encompasses shared values, unwavering commitment, and a collective identity, aligning with SIT's assertion that individuals seek positive distinctiveness through group membership. When individuals strongly identify with a group, they exhibit in-group favoritism, fostering trust, solidarity, and resilience against external threats (Ellemers, Spears, and Doosje 2002). Such cohesive dynamics consistently catalyze positive attitudes and behaviors within groups, reinforcing members' loyalty and motivation to uphold their collective identity (Friedkin 2004).

While task commitment, social commitment, and sense of belonging all contribute to group cohesion, they do so in distinct ways. Task commitment is fundamentally goal oriented, emphasizing the collaborative effort required to achieve shared objectives. It binds members through their active participation in common tasks and the pursuit of collective success. In contrast, a sense of belonging is identity driven, rooted in the emotional connection members feel toward the group, shaped by shared values, beliefs, and purpose. It fosters cohesion through psychological attachment rather than shared work. Social commitment, on the other hand, focuses on the interpersonal relationships and communication within the group. It highlights how frequent and meaningful interactions create a strong network of social ties, reinforcing the group's cohesion through the exchange of ideas and mutual recognition. While task commitment strengthens cohesion through collaboration, and a sense of belonging does so through identity and shared purpose, social commitment ensures cohesion through interpersonal interaction and ongoing engagement. Together, these dimensions provide a comprehensive understanding of how groups remain cohesive across different contexts.

Although these three forms of cohesion are conceptually distinct, they are often interconnected in practice. Individuals do not operate in isolated modes of cohesion; rather, they may simultaneously pursue shared goals, express group identity, and engage in social interaction. The key distinction lies in the primary function of their engagement—whether it is driven by shared tasks, collective identity, or interpersonal relationships.

Building upon this framework, this paper evaluates the impact of deplatforming on the online cohesion of Proud Boys followers by examining three key dimensions. First, we assess task commitment, reflected in followers' active content creation, including the consistent use of Proud Boys-related hashtags, circulation of group-aligned links, and promotion of ideological narratives. These behaviors demonstrate their dedication to advancing the group's objectives and maintaining its online presence (Carron, Widmeyer, and Brawley 1985; Festinger, Schachter, and Back 1963). Second, we examine social commitment, distinguishing between direct and indirect engagement. Direct interaction, such as retweeting, explicitly connects users by amplifying content and reinforcing interpersonal bonds (Metaxas et al. 2015). We analyze internal retweets to assess how followers engage with each other, indicating the group's internal cohesion. Additionally, we examine external retweets to determine how often followers interact with non-Proud Boys users, reflecting their outreach beyond the group. Indirect engagement, including co-sharing hashtags and domain links, signals ideological alignment without direct interaction, sustaining networked ties within the group. Though passive, these behaviors reinforce shared narratives and contribute to the group's social fabric (Selim and Popovac 2024; Wang, Liu, and Gao 2016). Lastly, we explore followers' sense of belonging, reflected in group-specific slang, symbols, and references in shared content. These linguistic markers strengthen identity and reinforce exclusivity among followers (Friedkin 2004).

Following our examination of group cohesion dimensions among Proud Boys followers, we formulate our

https://doi.org/10.1017/S1537592725101941

research question: how does the deplatforming of Proud Boys-affiliated accounts on Twitter impact the cohesion of their followers, particularly in terms of their task commitment, social commitment, and their expression of a sense of belonging?

Data and Method

Twitter took action on August 10, 2018, by suspending accounts linked to the Proud Boys for infringing on its policies against "violent extremist groups" (Mac and Montgomery 2018), resulting in the deletion of 133 accounts related to the group from the platform. Prior to this enforcement, on July 15, 2018, we archived follower data from these accounts, capturing a total of 83,126 followers. This number represents the cumulative count of all followers across the suspended accounts, meaning some users were counted multiple times if they followed more than one of these accounts. To ensure analytical accuracy, we deduplicated this dataset, identifying 43,677 unique follower accounts—each representing an individual who followed at least one of the 133 suspended accounts. Of these, 16,289 accounts were either subsequently suspended after the Proud Boys' ban[2] or were already protected,[3] leaving a refined dataset of 27,388 unique users for our investigation.

To understand the impact of deplatforming, we have collected all historical tweets posted by these followers of the Proud Boys from August 1, 2017 (one year prior to the intervention), through to September 1, 2019 (one year postintervention), which resulted in the acquisition of over 12 million tweets posted by 9,728 accounts that remained active. An account is considered active if it posted at least once in either the year before or after the ban and remained on Twitter without being suspended. While we retrieved follower data from all 133 suspended accounts, our dataset represents a single snapshot of followership as of July 15, 2018. That means it captures only the followers present at that exact moment and does not include users who had previously followed and later unfollowed before our collection date.[4] All replication data and code used in this study are publicly available on Harvard Dataverse (Zhong and Zhang 2025).

To investigate the impact of suspending Proud Boys' accounts on their followers—specifically changes in group cohesion over time—we employ an interrupted time series (ITS) analysis, a quasi-experimental design suited for evaluating the longitudinal effects of interventions (Broniatowski et al. 2023; Dahlke and Pan 2024; Kontopantelis et al. 2015; Schaffer, Dobbins, and Pearson 2021). ITS is particularly useful for analyzing observational data where randomization or a case-control design is not feasible, as it leverages preintervention trends to assess postintervention changes (Kontopantelis et al. 2015). This method allows us to examine shifts in task commitment, social commitment, and sense of belonging among Proud Boys followers, providing a structured approach to identifying significant changes following deplatforming.

To address challenges posed by autocorrelation and nonstationarity—common in time series data—we apply seasonal autoregressive integrated moving average (SARIMA) models to the preintervention data.[5] Once fitted, these models generate predictions of what the trend would have been postintervention had the deplatforming not occurred, representing the counterfactual scenario. To enhance robustness, we conduct a thousand simulations of this counterfactual scenario, creating a distribution of possible outcomes based on preintervention trends (Broniatowski et al. 2023). By comparing the actual postintervention data to this counterfactual distribution, we test whether the differences between the observed postintervention results and the predicted trend are statistically significant under the null hypothesis of no intervention effect. Statistical analyses are performed using Python's `pmdarima` and `statsmodels` packages. All tests are two sided, and a p-value of 0.05 or less is considered statistically significant. Detailed methodologies are provided in the online supplementary materials.

The ITS analysis is conducted on a weekly dataset, using weekly aggregates to measure the frequency and proportion of relevant tweets. Weekly intervals are chosen over daily aggregates to reduce volatility and minimize noise from external factors. Prior research on deplatforming (e.g., Chandrasekharan et al. 2017; Horta Ribeiro et al. 2023) demonstrates that weekly or 10-day time windows effectively capture behavioral shifts while smoothing out short-term fluctuations. Similarly, Jhaver and colleagues (2021) find that while posting activity declines sharply after deplatforming, users do not immediately alter their long-term engagement patterns. Instead, they gradually adjust their behavior over time, reinforcing the need for an analytical window that captures sustained shifts rather than temporary fluctuations. Therefore, using weekly data provide a more stable foundation for analysis, allowing us to identify meaningful behavioral trends while minimizing short-term noise.

Measurement

Task Commitment

Hashtag Use as an Indicator of Task Commitment. In our study focusing on task commitment, we explore the thematic consistency of content produced by followers, encompassing hashtags, news links, and ideological narratives. Our methodology's initial step involves extracting a comprehensive dataset of 396,270 unique hashtags from the tweets of 9,728 followers, which appear collectively 4,528,218 times. To ensure the relevance of these hashtags and the communal engagement they receive, we refine our dataset to include only hashtags shared more than once, ultimately identifying 238,107 unique hashtags. This refinement is crucial for assessing the significance of the hashtags and confirming their repeated use and sharing among followers.

To precisely determine the association of these hashtags with the Proud Boys, we conduct an in-depth analysis. Utilizing datasets from Bailard and colleagues (2024) and Zhong and colleagues (2024), which catalog all historical hashtags used by 92 core Proud Boys Telegram channels from November 2018 to July 2022, we match 1,776 hashtags from our dataset to theirs. This matching process is vital for ensuring the direct relevance of these hashtags to the Proud Boys. Through detailed examination and manual annotation of 1,776 hashtags and their associated tweets, we identify 716 unique hashtags[6] that are clearly associated with the group.[7] Notably, a significant portion of these hashtags, such as "proudboys," "pb," "uhuru," "poyb," and "istandwiththeproudboys," directly reflect the group's identity, promoting a sense of unity and solidarity among its members. Another key set of hashtags, including "fightfortrump" and "maga," underscores the group's political allegiance, particularly its support for Donald Trump, a stance commonly shared by far-right factions. Hashtags such as "freespeech," "patriot," "whitegenocide," "whitepower," "defundantifa," and "nohomo" highlight the group's key ideologies, merging nationalism, free speech advocacy, white identity politics, and opposition to antifa (the antifascist movement) and LGBTQ+ inclusivity. Additionally, hashtags referencing conspiracy theories and media distrust, such as "qanon," "thegreatawakening," "deepstate," "fakenews," "censorship," "enemyofthepeople," and "mainstreammedia," are also prevalent.

News Link Sharing as an Indicator of Task Commitment. Beyond analyzing hashtags, our study explores how social platforms facilitate the dissemination of information through links to external content, such as news articles, circulated among Proud Boys followers. To pinpoint links related to the Proud Boys, we extract and analyze the domain names from each link within our dataset. For instance, from a web address like www.yahoo.com/news/MAGA-news-story, we extract "yahoo.com" for a more focused examination. Our examination then narrows to focus specifically on domains from news websites shared within their posts. We classify these news websites based on political labels assigned by the Media Bias/Fact Check (MBFC) website, a fact-checking organization that evaluates the ideological bias and factual reporting credibility of various media sources.[8] Despite MBFC's methodology including subjective elements, it employs a numerical scoring system to assign labels of political bias. These labels span a spectrum from "far left" to "far/extreme right" and include "left," "left-center," "least biased/pro-science," "right-center," "right," "right conspiracy," "extreme-right," and "satire." These labels are applied at the domain level, ensuring consistency across all articles from a specific source.

To precisely focus on content relevant to the Proud Boys, we employed a filtering approach that targets domains categorized by MBFC as "right," "right conspiracy," "extreme-right," or "satire." Zhong and colleagues (2024) find that Proud Boys members frequently share content from right-wing, far-right, and conspiracy-aligned sources, making this filtering essential for capturing the external content most commonly shared and endorsed by their followers. Although "satire" may seem distinct from ideological categories, it plays a strategic role in far-right media ecosystems. Far-right-aligned satire functions as a tool for shaping narratives, influencing discourse, and normalizing extremist views under the guise of humor (Fielitz and Ahmed 2021). It employs irony and comedic framing to make extremist rhetoric more palatable while maintaining plausible deniability (Phillips and Milner 2018). Research has shown that far-right communities use satire to reframe racist, xenophobic, and conspiratorial rhetoric as humor, lowering barriers to engagement with extremist ideas (Marwick and Lewis 2017). Schwarzenegger and Wagner (2018) further argue that satire facilitates ideological infiltration, making radical positions appear more legitimate and mainstream. Empirical evidence has further supported satire's role in far-right information networks. Zhong and colleagues (2024) find that 2% of the top hundred news websites shared by Proud Boys public channels on Telegram were classified as "satire" by MBFC. This confirms that satire is not an outlier but an integral part of far-right media consumption. Given this evidence, we include satire in our classification to provide a comprehensive analysis of the media shaping Proud Boys discourse.

From our dataset, we have pinpointed 45,991 unique domains, appearing a total of 5,178,859 times. Among this vast array, 815 news domains fall under MBFC's designated categories of "right," "right conspiracy," "extreme-right," or "satire," collectively accounting for 228,013 instances. This detailed examination allows us to identify the role that related news domains play in disseminating information within Proud Boys circles, underscoring their impact on shaping the group's collective understanding and actions.

Ideological-Narrative Sharing as an Indicator of Task Commitment. The Proud Boys are recognized for their far-right extremist views, encompassing Islamophobia, anti-Semitism, homophobia, transphobia, misogyny, and xenophobia. Their ideological narratives primarily propagate xenophobia toward immigrants, racism, misogyny, and transphobia, reflecting a deep-seated inclination to spread hatred based on gender, immigration status, race, and sexual orientation (Stern et al. 2019). To systematically detect and analyze these ideological narratives in social media discourse, we adopt a lexicon-based approach.

https://doi.org/10.1017/S1537592725101941

Research has demonstrated that lexicon-based methods are effective for systematically tracking hate speech across large datasets, particularly in online environments where hateful content is repetitive and structured (Davidson, Bhattacharya, and Weber 2019). These methods are widely used in hate-speech detection because they allow for consistent, high-coverage identification of key hate terms, avoiding the subjectivity and labor intensiveness of manual annotation (MacAvaney et al. 2019; Watanabe, Bouazizi, and Ohtsuki 2018). For example, Wiegand and colleagues (2018) have developed a lexicon of abusive terms, while Chandrasekharan and colleagues (2017) have created a specialized dictionary for detecting hate speech in Reddit posts. Further supporting the efficacy of lexicon-based approaches, Liang, Ng, and Tsang (2023) demonstrate that these methods outperform more complex machine-learning techniques in identifying instances of incivility. Building on these foundations, our approach integrates existing lexicons from prior research with an interactive human review process to ensure accuracy and relevance.

We first construct a specialized lexicon targeting four primary types of ideological hate speech: gender-based, immigration-related, race-based, and sexuality-related hate. This is based on lexicon resources developed by hate-speech researchers.[9] Next, we apply the word-embedding enrichment method proposed by Liang, Ng, and Tsang (2023) to expand our initial hate lexicon. This method identifies words that are similar in meaning to the given lexicon entries. To maintain high similarity to our data, we only use data from the Measuring Hate Speech Corpus (Sachdeva et al. 2022), which comprises 50,070 social media comments from YouTube, Reddit, and Twitter, labeled by 11,143 annotators. We then manually review each word in the enriched lexicon and remove any irrelevant terms. The four sets of lexicons are merged into a single dictionary.[10] For each tweet in our dataset, we check if it contains any words from this ideological hate-speech dictionary. If so, we label the tweet as containing Proud Boys-related ideological-narrative hate speech.

To validate the dictionary, we manually label two thousand tweets as either Proud Boys ideological-narrative hate-related or unrelated, using the same criteria as in the dictionary construction (i.e., identifying hate speech targeting on the basis of gender, immigration status, race, or sexuality). Comparing the human labels to the dictionary-based labels yields an overall accuracy of 0.81.[11] Finally, we apply the hate-speech lexicon to all tweets in our dataset to determine which ones contain Proud Boys-related ideological hate narratives.

Social Commitment

Retweeting as an Indicator of Social Commitment. We are also interested in investigating whether Proud Boys followers directly and actively share content posted by in-group members, demonstrating their endorsement and promotion efforts. Consequently, we construct a retweet network, where each node represents an individual follower account, and edges indicate the retweet relationship. These networks are directed, meaning the edges show the direction of retweeting (i.e., who is retweeting whom), and weighted, meaning the edges reflect the frequency or number of retweets between users.

To capture the dynamic nature of retweeting behavior and evolving interaction patterns, we construct and analyze the retweet networks on a weekly basis. Each week, we compile all tweets and retweets made by Proud Boys followers, creating a network where edges represent retweets and are weighted by their frequency. Since our dataset includes retweets from both Proud Boys followers and nonfollowers, we distinguish between internal retweets (retweets from within the group) and external retweets (retweets from nonfollowers). This differentiation helps us to determine whether followers primarily engage in an echo chamber, amplifying in-group content, or actively seek and disseminate information from external sources.

There are multiple ways to measure and describe the structure of a network. In the literature, three primary dimensions of social network structure dominate: centralization, clustering, and density (Stokman 2001). Centralization identifies the most influential nodes based on their connections, but is less effective for understanding overall network cohesion (Caldarelli 2007; Newman 2018). Clustering measures the tendency of nodes to form tight-knit groups, providing insights into small, cohesive subgroups but not the entire network's cohesion (Venna, Gottumukkala, and Raghavan 2016). Density, by contrast, quantifies the ratio of actual connections to all possible connections (Wasserman and Faust 1994), making it a key indicator of structural interconnectedness. A higher density suggests a more cohesive network where followers are extensively linked (Moody and White 2003).

In high-density networks, nodes are highly interconnected, fostering strong internal linkages that facilitate information flow and contribute to the spread of shared narratives and ideologies (Hu and Racherla 2008; Jaramillo et al. 2021; Makagon, McCowan, and Mench 2012). Conversely, lower-density networks are more fragmented, with weaker connections that may limit content diffusion and group coordination (Haythornthwaite 1996). Given this, density serves as a key metric for assessing the structural connectivity of Proud Boys followers, providing insights into their overall interconnectedness and potential for cohesion.[12]

Co-Sharing of Hashtags and Domains. Another key aspect of studying deplatforming's impact on Proud Boys followers' cohesion is indirect engagement, particularly the co-sharing of group-specific content. This seemingly passive act serves as a mutual acknowledgment among group members, signifying their collective efforts to elevate specific content within the public discourse (Selim and Popovac 2024; Wang, Liu, and Gao 2016). When members share Proud Boys-specific hashtags and domains, they make a deliberate choice to engage with one another (Al-Rawi 2024; Darius and Stephany 2019). This public acknowledgment of each other's contributions fosters a supportive environment where individual actions strengthen the group's cohesion (Forsyth 2018; Severt and Estrada 2015). Such strategic engagement highlights the group's social fabric, as members indirectly express their allegiance and solidarity through shared content (Simpson 2018). This focused intent sets their interactions apart from more generalized social interactions. To assess whether content coproduction was affected by the deplatforming of Proud Boys accounts, we construct networks based on the co-sharing of Proud Boys-related hashtags and domains.

To examine the hashtag co-sharing network related to the Proud Boys, we begin by filtering tweets containing hashtags associated with the group, based on those hashtags related to the Proud Boys identified earlier. For each user posting these tweets, we document every Proud Boys-related hashtag they have used within a week, forming a user-hashtag matrix. This matrix enables us to create a user co-occurrence matrix that shows the frequency at which user pairs share the same Proud Boys-related hashtags. Through this method, we construct the Proud Boys hashtag co-sharing network. In this network, nodes represent individual user accounts, while edges link accounts that have shared identical hashtags. The network is both weighted and undirected, meaning that the connections (edges) can represent multiple shared hashtags between user pairs, with the weight indicating the shared hashtag frequency on a weekly basis. Adopting a similar methodology, we also develop networks based on the co-sharing of followers' news domains, which is defined by the news domains identified earlier that have been shared by Proud Boys followers. These networks are analyzed weekly and are both weighted and undirected, with the weights quantifying the number of times user pairs shared the same Proud Boys-related domains. The structured development of both hashtag and domain co-sharing networks sheds light on how Proud Boys followers coproduce and disseminate content, especially in light of the effects of deplatforming. For the co-sharing of both hashtags and domains, we calculate the weekly network density and track the changes over time.

Sense of Belonging

To accurately identify tweets associated with the Proud Boys' identity, we adopt a keyword-based methodology, drawing on extensive research to compile a comprehensive list of terms and phrases deeply tied to the group's ideologies and distinctive activities, as documented by Bailard and colleagues (2024) and Zhong and colleagues (2024). Our selection of keywords spans a wide array of specific slogans, mottos, and other unique expressions predominantly utilized by the Proud Boys and their adherents. These identity-related keywords include terms such as "Proud Boys," "uhuru," "sons of liberty," "POYB" (proud of your boy), "MAGA" (Make America Great Again), "Western chauvinist," "the West is the best," "Pinochet did nothing wrong," "OK hand sign," "6MWE" (6 million wasn't enough), "stand back and stand by," "Fred Perry," "no apologies," and "fraternal order of the Proud Boys."[13] Using these predefined keywords and their variations, we systematically scan our dataset to identify tweets that contain these terms. This structured approach allows us to sift through extensive data and isolate tweets that are explicitly associated with the group. By employing this method, we curate a collection of tweets that provide insights into the Proud Boys' self-identification among their followers.

Results

Our analysis begins with the evaluation of two outcome variables:

1. Weekly counts, which encompass the aggregate of Proud Boys-related hashtags, news domains, and ideological-narrative posts. To normalize the data, these counts undergo a log transformation. This process effectively reduces skewness from outliers and extreme values, and helps to stabilize variance across the dataset (Manning and Mullahy 2001).
2. Weekly proportions, highlighted in our main results to adjust for fluctuations in overall online activity and to allow for temporal comparisons.

We define these proportions based on different types of engagement within the Proud Boys' online discourse. The hashtag proportion is calculated as the number of tweets containing Proud Boys-related hashtags divided by the total number of tweets with hashtags posted by Proud Boys followers each week. Similarly, the domain proportion is determined by the number of tweets sharing links to right-wing or extremist domains relative to the total number of tweets containing links from Proud Boys followers in a given week. The narrative proportion captures the prevalence of tweets referencing key Proud Boys-related narratives, again standardized by the total weekly tweet volume of Proud Boys followers. Lastly, the sense-of-belonging proportion measures the frequency of tweets containing Proud Boys-specific identity markers, such as slogans or symbols, relative to the group's total weekly tweet output. By centering our attention on proportions,

https://doi.org/10.1017/S1537592725101941

we obtain a more precise measure of the shifts in the relative prominence and influence of content related to the Proud Boys, which remains unaffected by the total volume of online discourse. In total, our analysis includes nine distinct weekly time series measures, including both proportions and frequency counts.

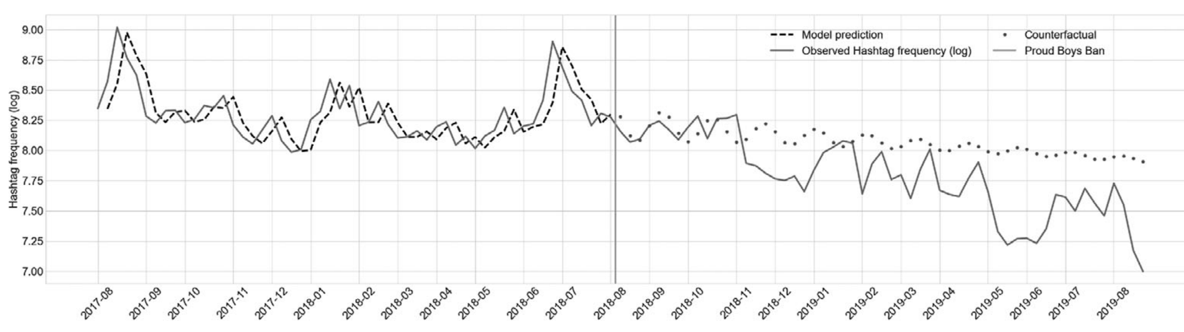
Group Cohesion in Task Commitment

Figures 1–3 depict the time series both prior to and subsequent to the intervention. In these figures, the observed weekly frequencies (presented in log format) and proportions (presented as decimal format) of Proud Boys-related hashtags, news domains, and ideological narratives are traced by the solid dark gray line. The implementation of Twitter's deplatforming or ban of the Proud Boys is demarcated by the vertical gray line, which bifurcates the timeline into pre- and postintervention phases. The SARIMA model's predictions are represented by the black dashed line for the pre-ban period and the black dotted line for the post-ban period. The segment of the dotted line extending beyond the intervention serves as the counterfactual trend—indicating what the model predicts would have occurred in the absence of the intervention, based on preintervention data. The divergence between the counterfactual and the observed postintervention data is indicative of the ban's impact.

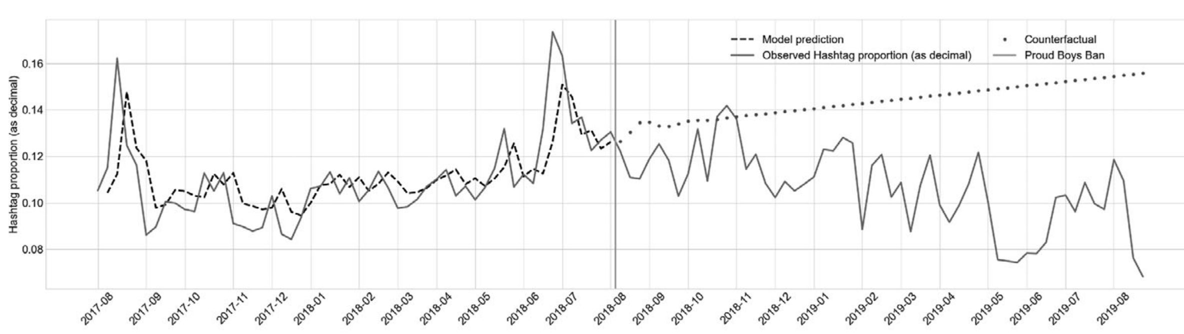
Furthermore, table 1 displays the results from the thousand-simulation analyses of the impact of Twitter's ban on the Proud Boys followers, focusing on the mean differences in both frequency and proportion before and after the intervention. The table is structured to provide a statistical summary of these two key variables, presenting the average change (log mean difference), the average change in likelihood (risk ratio), the variability of the changes (standard deviation), the range within which we can be confident the true mean difference lies (lower and upper confidence interval), the *z*-score indicating the number of standard deviations from the mean, and the *p*-value to test the hypothesis that there is no difference (null hypothesis).

As we can see, figure 1 illustrates the observed weekly frequencies and proportions of Proud Boys-related hashtags, analyzed over a period including Twitter's ban on the group. The top panel shows the frequency of these hashtags in log-transformed format, highlighting notable volatility and a peak in activity before the intervention. After the ban, the frequency exhibits a more stabilized, downward trend, suggesting the ban may have contributed to a reduction in hashtag usage. The bottom panel presents the

Figure 1
Interrupted Time Series of Proud Boys-Related Hashtags



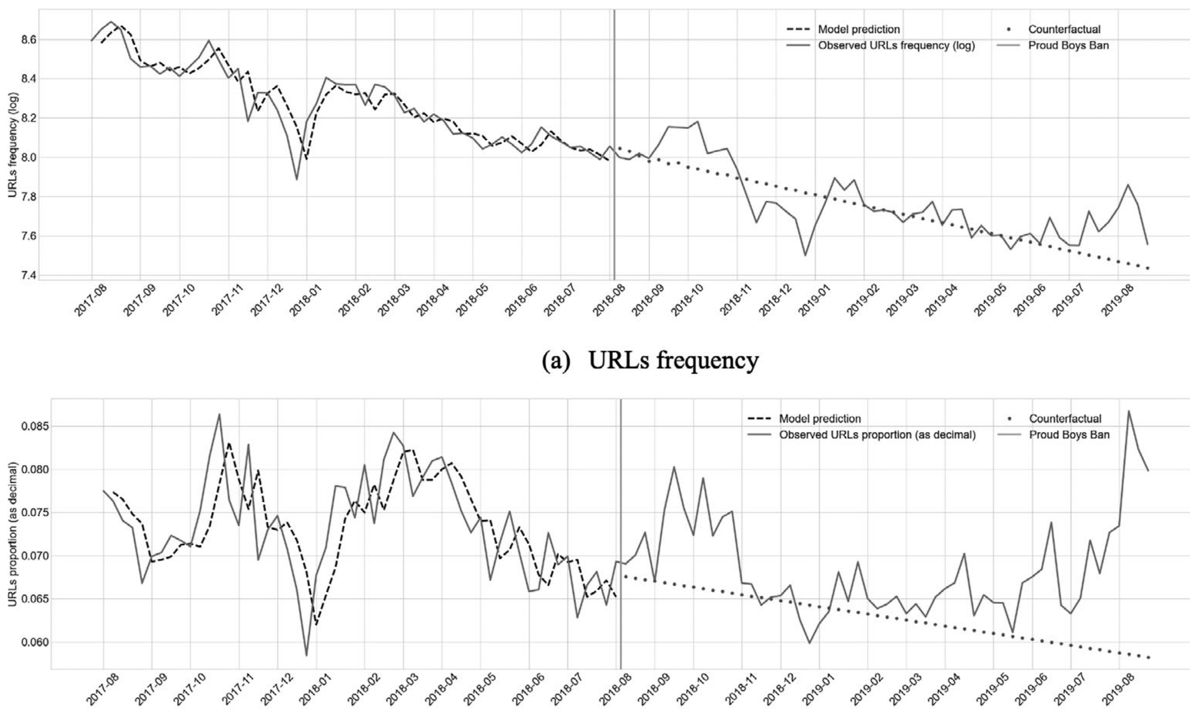
(a) Hashtag frequency



(b) Hashtag proportion

https://doi.org/10.1017/S1537592725101941

Figure 2
Interrupted Time Series of Proud Boys-Related News Domains



(a) URLs frequency

(b) URLs proportion

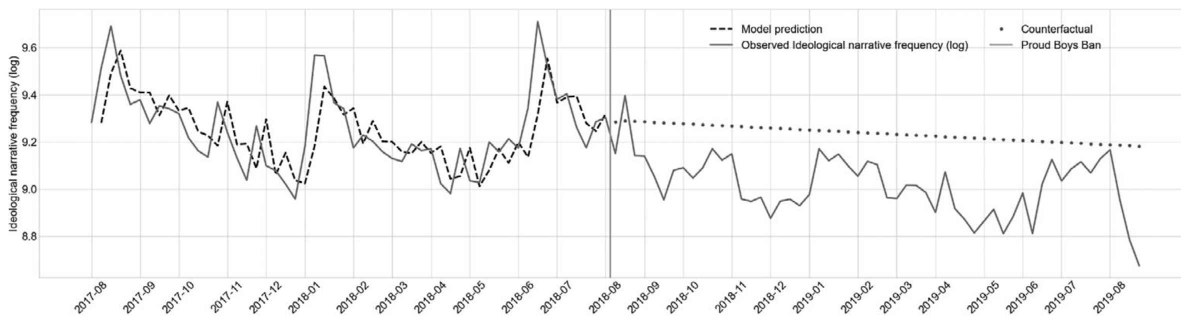
proportion of Proud Boys-related hashtags as a decimal fraction of total online activity. This proportion also fluctuates significantly before the ban, with a peak that aligns with increased activity. Postintervention, the proportion demonstrates a clear decline and stabilizes at a lower level, indicating a reduction in the relative prominence of Proud Boys-related content. The dotted lines in both panels represent the counterfactual scenarios, showing the predicted values if the ban had not occurred, further supporting the observed reductions in both frequency and proportion post-ban.

Figure 2 shows the observed weekly frequencies and proportions of Proud Boys-related news domains. In the top panel, postintervention, the observed frequencies closely align with the counterfactual predictions (dashed line and dotted points), indicating that the ban did not lead to a significant change in the frequency of Proud Boys-related news domains. In the bottom panel, the observed proportions, shown by the solid dark gray line, fluctuate considerably before the intervention, with several prominent peaks. After the ban, the observed proportions follow the counterfactual predictions until August 2019, suggesting that deplatforming did not significantly impact the relative prominence of these domains for around a year after the ban.

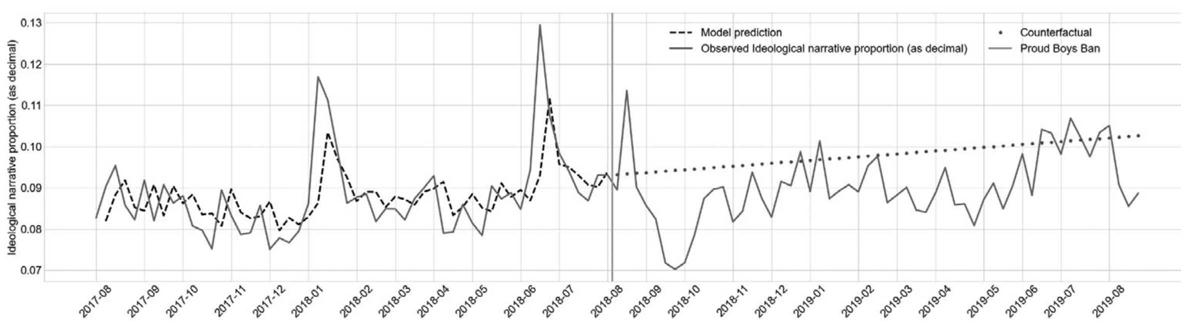
Figure 3 depicts the observed weekly frequencies and proportions of Proud Boys-related ideological narratives. After the intervention, there is a small peak, but overall, the observed frequencies remain mostly below the counterfactual predictions, indicating a slight reduction in the frequency of Proud Boys-related ideological narratives. The observed proportions after the ban, however, relatively closely align with the counterfactual predictions, suggesting that the ban did not significantly impact the relative prominence of these ideological narratives. Overall, these figures suggest that Twitter's ban on the Proud Boys might have had a slight effect on the frequency but did not have a substantial effect on the proportion of related ideological narratives.

The findings in table 1 reveal a statistically significant decrease in the proportion of Proud Boys-related hashtags post-ban, with a mean difference of −0.037 (p = 0.0257). This significant reduction indicates that the ban effectively diminished the relative visibility of and engagement with Proud Boys-related content in terms of hashtags. Conversely, the change in the proportion of domains linked to the Proud Boys is not statistically significant, with a mean difference of 0.006 (p = 0.6393). This suggests that, while the ban impacted the usage and spread of specific hashtags, it did not significantly alter the overall share of domains

Figure 3
Interrupted Time Series of Proud Boys-Related Ideological Narratives



(a) Ideological narrative frequency



(b) Ideological narrative proportion

Table 1
Impact of Twitter Ban on Group Cohesion in Task Commitment of Proud Boys Followers (Production of Hashtags, News Domains, and Ideological Narratives)

Time series	Log mean difference	Risk ratio	Std	Lower log CI	Upper log CI	Lower CI	Upper CI	*Z*-score	*P*-value
Hashtag frequency	−0.257	0.773	0.44	−1.12	0.606	0.326	1.834	−0.583	0.5598
Hashtag proportion	−0.037	0.964	0.016	−0.069	−0.004	0.933	0.996	−2.231	0.0257
Domain frequency	0.06	1.062	0.235	−0.401	0.521	0.67	1.683	0.255	0.7991
Domain proportion	0.006	1.006	0.012	−0.018	0.03	0.982	1.03	0.469	0.6393
Narrative frequency	−0.222	0.801	0.276	−0.763	0.32	0.466	1.377	−0.803	0.4219
Narrative proportion	−0.008	0.992	0.005	−0.017	0.001	0.983	1.001	−1.732	0.0833

associated with the group in the same time frame. Similarly, the proportion of ideological narratives, while showing a decrease (p = 0.0833), does not reach the conventional threshold for statistical significance. This result suggests a trend toward a slight reduction in the proportion of ideological narratives (including hateful content) related to the Proud Boys, which, although not statistically conclusive, indicates a potential decrease in the intensity of such content on Twitter following the ban.

These findings suggest that Twitter's intervention was effective in reducing the visibility of specific types of toxic content, particularly Proud Boys-related hashtags, as evidenced by the statistically significant decrease in their proportion post-ban. This indicates that the ban successfully diminished the engagement and spread of content associated with the Proud Boys through hashtags, which are a crucial mechanism for content discovery and dissemination on the platform. However, the lack of a statistically

https://doi.org/10.1017/S1537592725101941

significant change in the proportion of domains linked to the Proud Boys suggests that the ban's impact was not uniformly effective across all types of content. Domains, which may represent more established and potentially less easily altered sources of information, do not show a significant decrease in relative presence. This implies that while Twitter's ban curtailed the spread of hashtags, it did not significantly affect the distribution of content from Proud Boys-related domains. Similarly, the observed decrease in the proportion of ideological narratives does not reach statistical significance, although there is an indication of a potential reduction. This trend suggests that the intensity of ideological speech related to the Proud Boys may have lessened, even if not conclusively so. Overall, these discrepancies highlight that while the intervention curtailed the spread of certain forms of content, it did not uniformly affect all types of Proud Boys-related online activity.

Group Cohesion in Social Commitment

In this section, we analyze the effect of the ban on the social commitment of Proud Boys followers by examining changes in their retweet network density over time. Given the sparsity of these networks, the density values have been rescaled to facilitate a clearer analysis. The rescaling process adjusts the density values to a range between zero and one, accounting for the variability in network density and allowing for a more standardized comparison.[14]

Figure 4 offers a detailed look at the evolving dynamics of retweet networks among Proud Boys followers on Twitter. The figure is segmented into three panels, each representing different participant groups to assess the impact of Twitter's ban on the Proud Boys. The top panel shows the retweet network density among all participants, including both Proud Boys followers and non-Proud Boys followers. The observed density (solid dark gray line) continues to rise post-ban, closely following or slightly exceeding the counterfactual predictions (dashed line and dotted points). This indicates that the overall retweet network density may not have been significantly disrupted by the intervention. The middle panel focuses on the retweet network density between Proud Boys followers and nonfollowers. Here, the observed density exhibits continued growth, aligning closely with or surpassing the counterfactual predictions. Similarly, this suggests that interactions between these two groups may not have been significantly impacted by the ban.

The bottom panel illustrates the retweet network density among Proud Boys followers exclusively. Unlike the previous two panels, the observed density generally remains below the counterfactual estimates, indicating that internal retweet activity within the group declined following the ban. Table 2 further confirms this disruption, showing a significant decrease in internal retweet engagement among Proud Boys followers (mean decrease: −0.451, p = 0.0). This suggests that the ban weakened internal cohesion within the group, reducing engagement among members. At the same time, the ban led to a significant increase in retweet activity between Proud Boys followers and non-Proud Boys followers. Table 2 shows a significant increase in retweet activity between Proud Boys followers and nonfollowers (mean increase: 0.27, p = 0.0039), indicating a shift in engagement toward a broader audience. Rather than interacting primarily within their own group, Proud Boys followers redirected their engagement toward a broader audience.

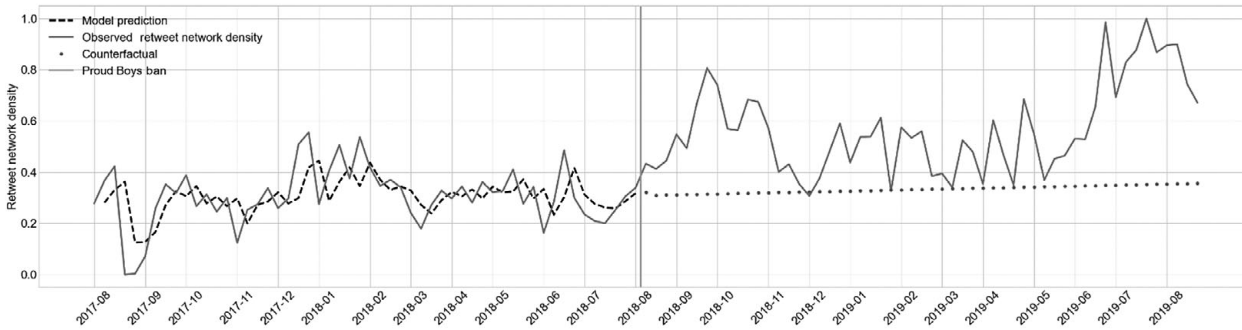
Taken together, these results suggest that Twitter's ban on the Proud Boys correlates with a significant increase in their followers' engagement with the wider community, while concurrently, their internal group interactions significantly decreased. This dichotomy points to a pivot in social strategy among the Proud Boys, characterized by a loss of internal cohesion and a surge in outward-facing communications post-ban.

While observing that Proud Boys followers interacted more broadly with external audiences, it is important to determine whether they engaged with outsiders about Proud Boys-related ideological content or unrelated topics. To investigate this, we have conducted a latent dirichlet allocation topic-modeling analysis (Blei, Ng, and Jordan 2003) on the content retweeted by these followers when interacting with non-Proud Boys followers, summarized in table 3 (the specific keywords for each topic are detailed in the online supplementary materials). The predominant topic, "Trump and presidential politics," accounts for 11.55% of the retweets, focusing on ongoing support for Trump. This is followed by "patriotism and conservative politics," which comprises 8.75% of the retweets, suggesting a narrative centered on national pride and conservative values. "Media criticism," at 8.21%, highlights a critical stance toward mainstream media and political opponents. Other notable topics include engagement with multimedia content and public figures ("videos and public figures," at 7.85%), discussions on racial issues and figures like Candace Owens ("race and public figures," at 7.65%), electoral processes ("voting and campaigns," at 6.95%), societal issues involving the police ("police and society," at 6.66%), and interaction with controversial news stories ("news and controversy," at 6.59%).

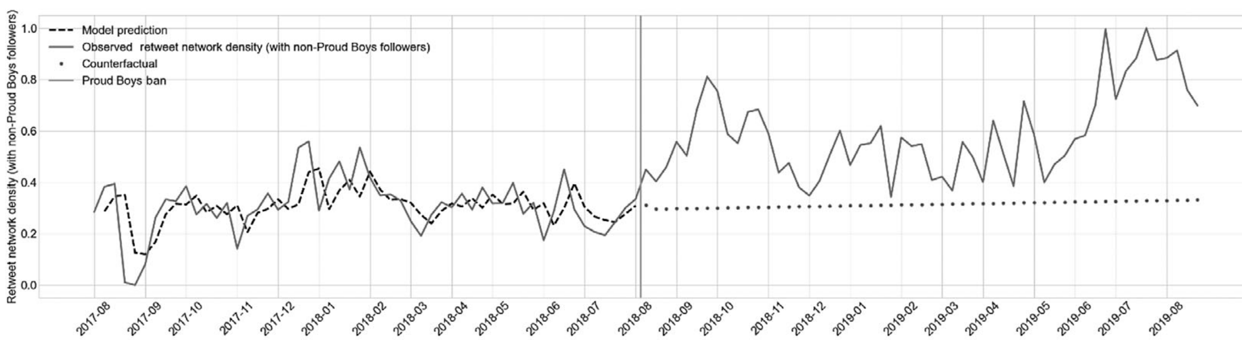
When these followers broadened their engagement and interacted with a wider audience, they predominantly focused on topics related to politics and conservative ideology. This suggests that Proud Boys followers adapted to the ban by diversifying their interactions while heavily concentrating on political issues that aligned with their ideological perspectives. The predominant topics of their interactions, such as support for Trump, patriotism, and media criticism, indicate that these followers were reaching out to broader audiences that may share or be sympathetic

https://doi.org/10.1017/S1537592725101941

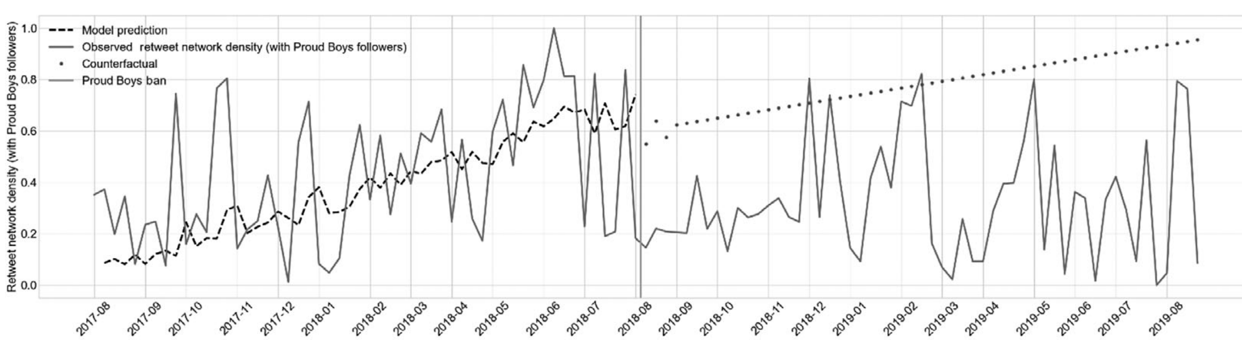
Figure 4
Interrupted Time Series of Proud Boys Retweet-Engagement Network Density



(a) Retweet network density for all Proud Boys-connected users



(b) Retweet network density for Proud Boys followers interacting with non–PB followers



(c) Retweet network density for Proud Boys followers only

Table 2
Impact of Twitter Ban on Group Cohesion in Social Commitment of Proud Boys Followers

Time series	Mean difference	Std	Lower CI	Upper CI	*Z*-score	*P*-value
Retweet network density	0.234	0.074	0.088	0.379	3.147	0.0016
Retweet network density (with non-Proud Boys followers)	0.270	0.093	0.087	0.452	2.89	0.0039
Retweet network density (with Proud Boys followers)	–0.451	0.059	–0.567	–0.336	–7.66	0.0

Table 3
Topics between Proud Boys Followers with Non-Proud Boys Followers

Topic	%	N
Trump and presidential politics	11.55%	128,211
Patriotism and conservative politics	8.75%	97,216
Media criticism	8.21%	91,223
Videos and public figures	7.85%	87,205
Race and public figures	7.65%	84,974
Voting and campaigns	6.95%	77,227
Police and society	6.66%	73,902
News and controversy	6.59%	73,128
Brexit and tech giants	6.54%	72,580
Gender and children	6.31%	70,036
American identity and politics	6.13%	68,035
Clinton and current events	5.89%	65,423
General tweets and politics	5.5%	61,069
Media and investigations	5.42%	60,243

to their views. Despite the ban, they continued to seek support and solidarity from like-minded individuals or those with relatively similar ideologies. This adaptive behavior highlights the complexity of moderating online extremist content. Direct interventions like deplatforming can disrupt specific network dynamics but can also lead to broader outreach efforts. This potentially expands the influence of groups like the Proud Boys by connecting them with a larger audience that resonates with their far-right ideologies and conservative perspectives. Therefore, while the ban disrupted internal communication, it also inadvertently encouraged Proud Boys followers to align their messaging with broader, yet ideologically similar, political conversations, enhancing their reach and impact.

Beyond retweet engagement, we also examine the collaborative generation of hashtags and domains by the group's followers, utilizing the network identity framework as described in the "Measurement" section. In defining network identity, we consider exclusively the original posts by these followers, as retweets have been previously examined. Figure 5 illustrates the co-sharing network density for Proud Boys-related content, focusing on hashtags and domains, before and after Twitter's ban on the group. The top panel shows the co-sharing network density for hashtags. Post-ban, the observed density aligns with or slightly falls below the counterfactual predictions, indicating a possible reduction after the ban. The bottom panel depicts the co-sharing network density for domains associated with the Proud Boys. Following the ban, the observed density remains above or aligns with the counterfactual predictions.

Although the panels display different patterns, neither shows a clear increase or decrease in observed densities, as the trends largely overlap with the counterfactual values. This suggests that the intervention had no clear or significant impact on the joint sharing of hashtags and news domains in original tweets. Table 4 further supports this finding, showing that the co-sharing of neither hashtags ($p = 0.7673$) nor domains ($p = 0.8205$) in original tweets exhibits a statistically significant difference post-intervention.

In summary, these results suggest that while Twitter's ban impacted the internal dynamics of Proud Boys followers, it did not silence the group's online joint activity. Instead, the group adapted by engaging more with the broader community and maintaining its use of hashtags and domains for content sharing. This might indicate the challenges of using platform bans as a tool to reduce the influence of such groups, as they can adapt and find new strategies to maintain their presence and influence online.

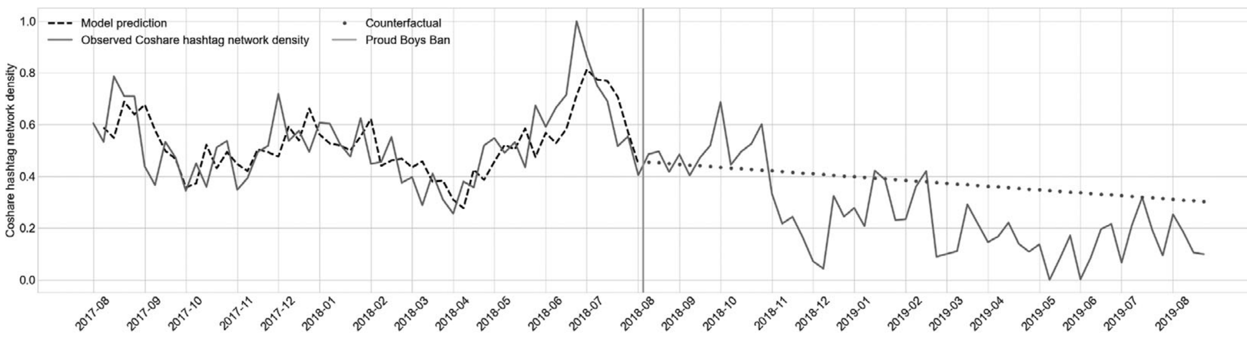
Group Cohesion in Sense of Belonging

Lastly, we investigate the shifts in the expression of a sense of belonging among Proud Boys followers before and after Twitter's intervention. Figure 6 examines the frequency and proportion of Proud Boys-related identity content on Twitter before and after the implementation of the platform's ban. In the top panel, after the intervention, the observed frequency tends to fall below the counterfactual predictions, suggesting a possible reduction in the frequency of such content. The bottom panel depicts the proportion of identity-related content as a fraction of total Twitter activity. Post-ban, the observed proportions generally align with the counterfactual predictions, indicating a nonsignificant decrease in the relative prominence of Proud Boys-related identity content. Corroborating this, simulated outcomes presented in table 5 support these findings, showing mean differences in identity frequency and proportion that are negligible, with p-values of 0.1105 and 0.3279, respectively. This reinforces the conclusion that the ban did not have a lasting effect on the group's sense of belonging as reflected in online expression.

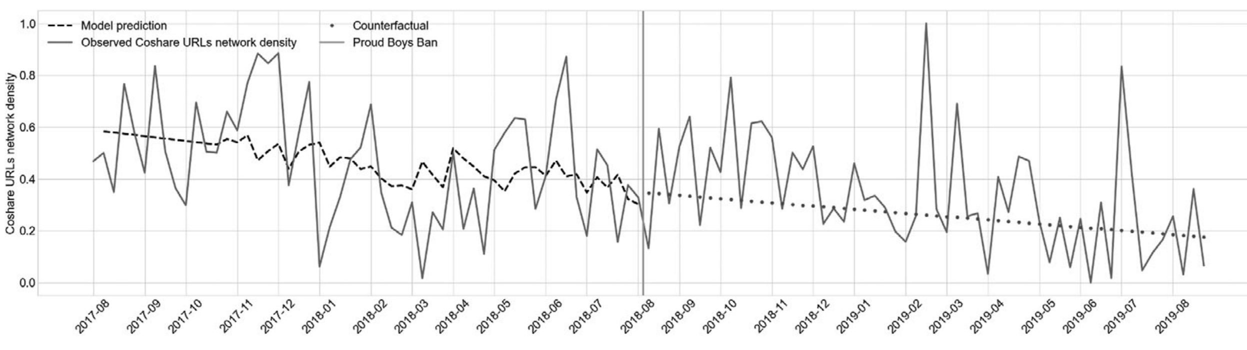
Discussion

The deplatforming of extremist groups remains a critical issue in platform governance, with ongoing debates about its effectiveness and unintended consequences. While previous studies have shown that deplatforming can significantly reduce the visibility and engagement of banned accounts (Chandrasekharan et al. 2017; Saleem and Ruths 2018), its impact on the broader networks of extremist group followers remains less understood. Our findings indicate that while deplatforming decreased the use of Proud Boys-related hashtags and disrupted internal retweet engagement, it did not reduce the promotion of extremist content, weaken the interconnectedness of followers' information-sharing networks, or diminish their attachment to the group. Instead, engagement with broader audiences increased, suggesting adaptation rather

Figure 5
Interrupted Time Series of Proud Boys Joint Sharing Proud Boys-Related Hashtags and News Domains



(a) Co-share hashtag network density



(b) Co-share URLs network density

Table 4
Impact of Twitter Ban on Group Cohesion in Task Commitment of Proud Boys Followers (Coproduction of Hashtags, News Domains)

Variable	Mean difference	Std	Lower CI	Upper CI	*Z*-score	*P*-value
Co-share hashtag network density	−0.106	0.358	−0.807	0.595	−0.296	0.7673
Co-share domain network density	0.068	0.298	−0.516	0.651	0.227	0.8205

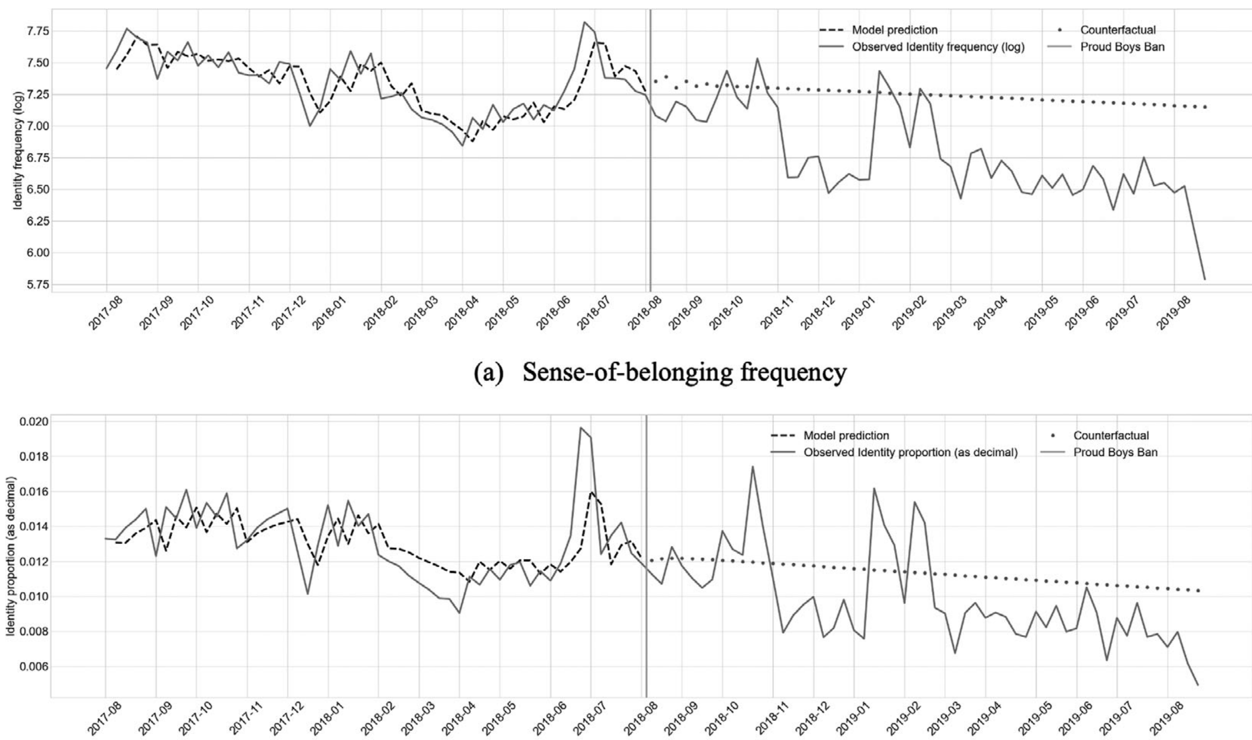
than fragmentation. These results align with SIT (Tajfel and Turner 1979) and conflict studies (Cohen 2013; Coser 1956; Gates 2002), which suggest that external threats can strengthen group cohesion and attachment. They also support research on extremist group resilience and adaptation (Bliuc et al. 2020; Horta Ribeiro et al. 2023), highlighting the need for more nuanced moderation strategies.

A key insight from our study is that social identity mechanisms and conflict dynamics help to explain the continued cohesion of extremist group followers post-deplatforming. SIT posits that group membership is a core aspect of individuals' self-concept, and external threats—such as deplatforming—can strengthen in-group solidarity rather than weaken it (Abrams and Hogg 1990; Tajfel and Turner 1979). Political identities, particularly those tied to ideological or extremist movements, tend to be highly stable and resistant to disruption (Huddy 2001). When groups perceive punitive or exclusionary actions from external sources, members—particularly those with strong group identities—often respond by reinforcing their attachment to the in-group. This response can lead to greater internal cohesion and, in some cases, heightened hostility toward the out-group, rather than withdrawal from their own group (Bourhis et al. 1979; Branscombe et al. 1999).

Conflict studies further explain why deplatforming may reinforce rather than weaken extremist group cohesion.

https://doi.org/10.1017/S1537592725101941

Figure 6
Interrupted Time Series of Proud Boys' Sense of Belonging



(a) Sense-of-belonging frequency

(b) Sense-of-belonging proportion

Table 5
Impact of Twitter Ban on Group Cohesion in Sense of Belonging of Proud Boys Followers

Time series	Log mean difference	Risk ratio	Std	Lower log CI	Upper log CI	Lower CI	Upper CI	*Z*-score	*P*-value
Identity frequency	−0.463	0.629	0.29	−1.032	0.106	0.356	1.111	−1.596	0.1105
Identity proportion	−0.001	0.999	0.001	−0.004	0.001	0.996	1.001	−0.978	0.3279

Research shows that external threats often unify groups by strengthening in-group identity, a feeling of shared struggle, and emotional bonds (Blake, Shepard, and Mouton 1964; Coser 1956; Sherif 1966). The Proud Boys, whose ideology thrives on conflict, may have framed deplatforming as political suppression, deepening their collective identity. Even when formal structures are disrupted, groups often maintain cohesion through shared narratives and peer reinforcement (Checkel 2017; Cohen 2013; Gates 2002). Thus, while deplatforming may disrupt centralized coordination, it likely sustains ideological loyalty through decentralized digital networks.

Additionally, our analysis of social commitment—particularly retweet networks—further highlights the adaptability of extremist groups. The resilience of these networks suggests that banning official accounts may unintentionally push extremist followers to expand their messaging beyond their immediate community, potentially increasing their reach to broader audiences. This aligns with prior research showing that deplatforming disrupts organizational structures in the short term but does not prevent users from finding alternative ways to sustain their networks (Horta Ribeiro et al. 2023; Mekacher, Falkenberg, and Baronchelli 2023; Vu, Hutchings, and Anderson 2024).

Taken together, these findings suggest that deplatforming alone is unlikely to fully dismantle the broader networks of extremist communities. Instead, it may serve as one factor within a larger process of organizational adaptation, where followers adjust their engagement strategies to maintain group cohesion and influence. Future research should further explore multiplatform effects of deplatforming, the long-term ideological trajectories of deplatformed groups, and the role of cross-platform enforcement in mitigating extremist content more effectively.

At the same time, it is important to recognize the limitations of this study in assessing the full scope of deplatforming's effects on extremist group followers. First, our dataset may not capture the complete set of followers for each account, as follower lists are dynamic and constantly changing due to new follows, unfollows, and account suspensions during the data collection period. These fluctuations could affect the composition of the follower network, potentially influencing our observations and interpretations. Second, our analysis relies on publicly available Twitter data, which may not capture the full extent of Proud Boys followers' activities across other platforms. After deplatforming, users often migrate to alternative or fringe platforms, such as Gab or Telegram (Horta Ribeiro et al. 2023), where they may continue engaging in ways that are not observable in our dataset. Future research could incorporate multiplatform analyses to provide a more comprehensive picture of the impact of deplatforming. Third, the effectiveness of deplatforming is shaped by factors beyond the scope of our study, including the broader organizational and legal challenges that may have also influenced the group's trajectory. For example, Gavin McInnes's resignation as leader in November 2018 (Wilson 2018b) and the prosecution of several Proud Boys members following violent altercations in October 2018 (Kriner and Lewis 2021) could have played a role in shaping the group's activity and cohesion during this period. While our study focuses on deplatforming as a key mechanism, these concurrent developments highlight the complexity of disentangling the effects of platform interventions from other external pressures. Finally, our analysis focuses on the Proud Boys as a case study, which, while offering a detailed examination of one group's response to deplatforming, may limit the generalizability of our findings to other extremist organizations or online communities. Consequently, our findings may not fully capture the broader efficacy of deplatforming as a universal content moderation strategy.

Supplementary material

To view supplementary material for this article, please visit http://doi.org/10.1017/S1537592725101941.

Data replication

Data replication sets are available in Harvard Dataverse at: https://doi.org/10.7910/DVN/DVFCT3

Notes

1 To contextualize these findings, offline analogies offer valuable insights. Like state-imposed censorship, deplatforming restricts access to certain ideas, shaping discourse by removing harmful or controversial content (Crawford and Pilanski 2014; Fisher et al. 1999). Similarly, banning political parties can weaken movements by limiting their visibility, though in some cases it fuels resistance and underground mobilization (Lust-Okar 2005; Wegner 2011). These parallels highlight how controlling access to communication—whether online or offline—can reshape political behavior and mobilization in complex ways.

2 In this paper, we use the terms "ban" and "deplatforming" interchangeably to refer to Twitter's removal of the Proud Boys' main accounts from the platform.

3 A protected Twitter account means that the account owner has chosen to make their tweets and profile information private. This setting restricts who can see their tweets and who can follow them. See details at https://help.twitter.com/en/safety-and-security/public-and-protected-posts. Although we archived the follower data before the August 10, 2018, suspension, we were unable to retrieve tweet histories for 16,289 users due to subsequent account suspensions or privacy settings. Some accounts were suspended after the Proud Boys ban, making their past tweets inaccessible via the Twitter application programming interface (API). Others had protected (private) accounts, either before or after the ban, preventing access to their historical tweets. As a result, our final dataset includes 27,388 unique users whose accounts remained public and active throughout our data collection period.

4 While this dataset provides a comprehensive view of followership at that time, it does not reflect prior fluctuations. Given the dynamic nature of Twitter followership, we acknowledge that this snapshot may not fully represent the historical follower base of these accounts.

5 ITS analysis often relies on segmented regression to estimate intervention effects by modeling changes in level and trend. However, this method assumes a linear trend and independently distributed residuals, which are often unrealistic for social media data, where user activity is highly correlated over time (Tenkanen et al. 2017). In our case, users who post actively tend to continue posting, creating temporal dependencies that segmented regression fails to account for. To address this, we use SARIMA models, which explicitly

model autocorrelation by regressing the outcome on its own past values rather than on time alone. This approach is particularly effective for ITS studies with complex preintervention trends, correlated residuals, or seasonality, providing more accurate counterfactual predictions and a more robust estimation of intervention effects (Schaffer, Dobbins, and Pearson 2021).

6 See the full list of hashtags in the online supplementary materials.

7 While our analysis initially drew from the 1,776 hashtags identified by Bailard and colleagues (2024) and Zhong and colleagues (2024), we included only a subset that met specific criteria for relevance to the Proud Boys' discourse. Some hashtags were excluded due to their generic nature or because they fell outside the 2017–19 study period. Generic hashtags unrelated to the Proud Boys, such as #superheroes, #ufcfightnight, #netflix, #hollywood, and #pepsi, were omitted. Additionally, hashtags tied to events after September 2019, including #covid19, #covidvaccine, #stopasianhate, #freekylerittenhouse, #rittenhousedidnothingwrong, #hernameisashlibabbitt, and #j6truth, were excluded as they emerged in response to the COVID-19 pandemic, the 2020 election, or the January 6, 2021, Capitol riot.

8 See the full list of MBFC domains in the online supplementary materials.

9 We use Weaponized Word (2025) and Hurtlex (Bassignana, Basile, and Patti 2019).

10 See the full list of hateful terms in the online supplementary materials.

11 This is calculated as the proportion of correctly classified instances (both true positives and true negatives) out of the total number of instances in our dataset.

12 Although our retweet network is directed and weighted, we use the traditional (binary) density measure, which calculates the ratio of observed connections to all possible connections in the network. This measure captures the overall structural interconnectedness of the network but does not account for the frequency or strength of interactions (i.e., edge weights representing retweet counts). While a weighted density approach could incorporate interaction intensity, our analysis focuses on the presence of connections rather than their strength, making traditional density the most appropriate measure for assessing structural cohesion.

13 See the full list of identity keywords in the online supplementary materials.

14 We applied a min-max rescaling transformation to normalize density values between zero and one, ensuring comparability across time periods. This transformation involves normalizing the density values by subtracting the minimum value and dividing by the range (the difference between the maximum and minimum values). Specifically, the lowest observed density across all weeks is mapped to zero, representing the period when the network was most fragmented with the fewest connections relative to possible ties, while the highest observed density is mapped to one, representing the period when the network was most interconnected. All other values are scaled proportionally within this range, reflecting their relative position between these two extremes. Since network size and structure fluctuate weekly, raw density values do not offer a stable reference for assessing changes in network cohesion. Small absolute differences may seem negligible but can reflect meaningful shifts in connectivity, particularly in sparse networks. Rescaling addresses this issue by ensuring that observed changes are evaluated relative to the full range of variations within the dataset rather than in isolation. By normalizing density in this way, we ensure that observed changes reflect genuine variations in connectivity rather than fluctuations in network size, allowing for clearer assessments of cohesion trends over time. This transformation provides a more interpretable picture of how engagement patterns among Proud Boys followers evolved after the ban.

References

Abrams, Dominic, and Michael A. Hogg. 1990. *Social Identity Theory: Constructive and Critical Advances*. New York: Springer.

Achua, Christopher F., and Robert N. Lussier. 2013. *Effective Leadership*, 5th edition. Mason, OH: South-Western Cengage Learning.

Al-Rawi, Ahmed. 2024. "The Far-Right Online Communities and the Dissemination of Hate." In *Online Hate on Social Media*, 23–57. Cham: Springer Nature Switzerland. DOI: 10.1007/978-3-031-61819-2_2.

Are, Carolina, and Pam Briggs. 2023. "The Emotional and Financial Impact of De-Platforming on Creators at the Margins." *Social Media + Society* 9 (1): 20563051231155103. DOI: 10.1177/20563051231155103.

Awan, Imran. 2017. "Cyber-Extremism: ISIS and the Power of Social Media." *Society* 54 (2): 138–49. DOI: 10.1007/s12115-017-0114-0.

Bailard, Catie Snow, Rebekah Tromble, Wei Zhong, Federico Bianchi, Pedram Hosseini, and David Broniatowski. 2024. "'Keep Your Heads Held High Boys!' Examining the Relationship between the Proud Boys' Online Discourse and Offline Activities." *American Political Science Review* 18 (4): 2054–71. DOI: 10.1017/S0003055423001478.

Bassignana, Elisa, Valerio Basile, and Viviana Patti. 2019. "Hurtlex: A Multilingual Lexicon of Words to Hurt." In *Proceedings of the Fifth Italian Conference on Computational Linguistics CLiC-it 2018, 10–*

12 December 2018, Torino, eds. Elena Cabrio, Alessandro Mazzei, and Fabio Tamburini, 51–56. Turin: Accademia University Press. DOI: 10.4000/books.aaccademia.3085.

Beal, Daniel J., Robin R. Cohen, Michael J. Burke, and Christy L. McLendon. 2003. "Cohesion and Performance in Groups: A Meta-Analytic Clarification of Construct Relations." *Journal of Applied Psychology* 88 (6): 989–1004. DOI: 10.1037/0021-9010.88.6.989.

Beauchamp, Zack. 2018. "Milo Yiannopoulos's Collapse Shows that No-Platforming Can Work." *Vox*, December 5. https://www.vox.com/policy-and-politics/2018/12/5/18125507/milo-yiannopoulos-debt-no-platform.

Benard, Stephen. 2012. "Cohesion from Conflict: Does Intergroup Conflict Motivate Intragroup Norm Enforcement and Support for Centralized Leadership?" *Social Psychology Quarterly* 75 (2): 107–30. DOI: 10.1177/0190272512442397.

Blake, Robert R., Herbert A. Shepard, and Jane S. Mouton. 1964. *Managing Intergroup Conflict in Industry*. Houston: Gulf Publishing.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3: 993–1022. https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf.

Bliuc, Ana-Maria, John M. Betts, Matteo Vergani, Muhammad Iqbal, and Kevin Dunn. 2019. "Collective Identity Changes in Far-Right Online Communities: The Role of Offline Intergroup Conflict." *New Media & Society* 21 (8): 1770–86. DOI: 10.1177/1461444819831779.

Bliuc, Ana-Maria, John M. Betts, Nicholas Faulkner, Matteo Vergani, Rui Jie Chow, Muhammad Iqbal, and David Best. 2020. "The Effects of Local Socio-Political Events on Group Cohesion in Online Far-Right Communities." *PLOS One* 15 (3): e0230302. DOI: 10.1371/journal.pone.0230302.

Bornstein, Gary. 2003. "Intergroup Conflict: Individual, Group, and Collective Interests." *Personality and Social Psychology Review* 7 (2): 129–45. DOI: 10.1207/S15327957PSPR0702_129-145.

Bourhis, Richard Y., Howard Giles, Jacques P. Leyens, and Henri Tajfel. 1979. "Psycholinguistic Distinctiveness: Language Divergence in Belgium." In *Language and Social Psychology*, eds. Howard Giles and Robert St. Clair, 158–85. Oxford, UK: Blackwell.

Branscombe, Nyla R., Naomi Ellemers, Russell Spears, and Bertjan Doosje. 1999. "The Context and Content of Social Identity Threat." In *Social Identity: Context, Commitment, Content*, eds. Naomi Ellemers, Russell Spears, and Bertjan Doosje, 35–58. Oxford, UK: Blackwell.

Brewer, Marilynn B. 2001. "Ingroup Identification and Intergroup Conflict: When Does Ingroup Love Become Outgroup Hate?" In *Social Identity, Intergroup Conflict, and Conflict Reduction*, eds. Richard D. Ashmore, Lee Jussim, and David Wilder, 17–41. Oxford, UK: Oxford University Press. DOI: 10.1093/oso/9780195137422.003.0002.

Broniatowski, David A., Joseph R. Simons, Jiayan Gu, Amelia M. Jamison, and Lorien C. Abroms. 2023. "The Efficacy of Facebook's Vaccine Misinformation Policies and Architecture during the COVID-19 Pandemic." *Science Advances* 9 (37): eadh2132. DOI: 10.1126/sciadv.adh2132.

Brown, Rupert. 2000. "Social Identity Theory: Past Achievements, Current Problems and Future Challenges." *European Journal of Social Psychology* 30 (6): 745–78. DOI: 10.1002/1099-0992(200011/12)30:6%3C745::aid-ejsp24%3E3.0.co;2-o.

Caldarelli, Guido. 2007. *Scale-Free Networks: Complex Webs in Nature and Technology*. Oxford, UK: Oxford University Press. DOI: 10.1093/acprof:oso/9780199211517.001.0001.

Carron, Albert V., W. Neil Widmeyer, and Lawrence R. Brawley. 1985. "The Development of an Instrument to Assess Cohesion in Sport Teams: The Group Environment Questionnaire." *Journal of Sport Psychology* 7 (3): 244–66. DOI: 10.1123/jsp.7.3.244.

Carsten, Melissa K., Mary Uhl-Bien, Bradley J. West, Jaime L. Patera, and Rob McGregor. 2010. "Exploring Social Constructions of Followership: A Qualitative Study." *Leadership Quarterly* 21 (3): 543–62. DOI: 10.1016/j.leaqua.2010.03.015.

Chaleff, Ira. 1995. *The Courageous Follower: Standing Up to and for Our Leaders*. San Francisco: Berrett-Koehler.

Chan, Simon T. M., Sammy W. S. Chiu, and Marcus Y. L. Chiu. 2010. "One Country, Two Cities: A Comparison of Perceived Cohesion in Guangzhou and Hong Kong." In *Social Cohesion in Greater China: Challenges for Social Policy and Governance*, eds. Ka Ho Mok and Yeun-Wen Ku, 121–46. Singapore: World Scientific. DOI: 10.1142/9789814291934_0005.

Chandrasekharan, Eshwar, Shagun Jhaver, Amy Bruckman, and Eric Gilbert. 2022. "Quarantined! Examining the Effects of a Community-Wide Moderation Intervention on Reddit." *ACM Transactions on Computer–Human Interaction* 29 (4): Article 29. DOI: 10.1145/3490499.

Chandrasekharan, Eshwar, Umashanthi Pavalanathan, Anirudh Srinivasan, Adam Glynn, Jacob Eisenstein, and Eric Gilbert. 2017. "You Can't Stay Here: The Efficacy of Reddit's 2015 Ban Examined through Hate Speech." *Proceedings of the ACM on Human–Computer Interaction* 1 (CSCW): Article 31. DOI: 10.1145/3134666.

Chatfield, Akemi Takeoka, Christopher G. Reddick, and Uuf Brajawidagda. 2015. "Tweeting Propaganda, Radicalization, and Recruitment: Islamic State

Supporters' Multi-Sided Twitter Networks." In *Proceedings of the 16th Annual International Conference on Digital Government Research, Phoenix, AZ, May 27–30*, eds. Jing Zhang and Yushim Kim, 239–49. New York: Association for Computing Machinery. DOI: 10.1145/2757401.2757408.

Checkel, Jeffrey T. 2017. "Socialization and Violence: Introduction and Framework." *Journal of Peace Research* 54 (5): 592–605. DOI: 10.1177/0022343317721813.

Cohen, Dara Kay. 2013. "Explaining Rape during Civil War: Cross-National Evidence (1980–2009)." *American Political Science Review* 107 (3): 461–77. DOI: 10.1017/s0003055413000221.

——. 2017. "The Ties that Bind: How Armed Groups Use Violence to Socialize Fighters." *Journal of Peace Research* 54 (5): 701–14. DOI: 10.1177/0022343317713559.

Coser, Lewis A. 1956. *The Functions of Social Conflict*. London: Routledge.

Crawford, Jarret T., and Jane M. Pilanski. 2014. "Political Intolerance, Right and Left." *Political Psychology* 35 (6): 841–51. DOI: 10.1111/j.1467-9221.2012.00926.x.

Dahlke, Ross, and Jennifer Pan. 2024. "January 6 Arrests and Media Coverage Do Not Remobilize Conservatives on Social Media." *Proceedings of the National Academy of Sciences* 121 (23): e2401239121. DOI: 10.1073/pnas.2401239121.

Darius, Philipp, and Fabian Stephany. 2019. "Hashjacking the Debate: Polarisation Strategies of Germany's Political Far-Right on Twitter." *Social Informatics: Proceedings of the 11th International SocInfo Conference, Doha, Qatar, November 18–21*, eds. Ingmar Weber, Kareem M. Darwish, Claudia Wagner, Emilio Zagheni, Laura Nelson, Samin Aref, and Fabian Flöck, 298–308. Lecture Notes in Computer Science, Vol. 11864. Cham: Springer. DOI: 10.1007/978-3-030-34971-4_21.

Davidson, Thomas, Debasmita Bhattacharya, and Ingmar Weber. 2019. "Racial Bias in Hate Speech and Abusive Language Detection Datasets." arXiv preprint arXiv: 1905.12516, version 1, updated May 29, 2019. DOI: 10.48550/arXiv.1905.12516.

DeCook, Julia R. 2018. "Memes and Symbolic Violence: #ProudBoys and the Use of Memes for Propaganda and the Construction of Collective Identity." *Learning, Media and Technology* 43 (4): 485–504. DOI: 10.1080/17439884.2018.1544149.

Dolan, Rebecca, Jodie Conduit, John Fahy, and Steve Goodman. 2016. "Social Media Engagement Behaviour: A Uses and Gratifications Perspective." *Journal of Strategic Marketing* 24 (3–4): 261–77. DOI: 10.1080/0965254X.2015.1095222.

Ellemers, Naomi, Russell Spears, and Bertjan Doosje. 2002. "Self and Social Identity." *Annual Review of Psychology* 53 (1): 161–86. DOI: 10.1146/annurev.psych.53.100901.135228.

Evans, Charles R., and Kenneth L. Dion. 2012. "Group Cohesion and Performance: A Meta-Analysis." *Small Group Research* 43 (6): 690–701. DOI: 10.1177/1046496412468074.

Festinger, Leon, Stanley Schachter, and Kurt W. Back. 1963. *Social Pressures in Informal Groups: A Study of Human Factors in Housing*. Stanford, CA: Stanford University Press.

Fielitz, Maik, and Reem Ahmed. 2021. "It's Not Funny Anymore: Far-Right Extremists' Use of Humour." *Radicalisation Awareness Network technical report*. Brussels: European Commission. https://web.archive.org/web/20210909171114/https://ec.europa.eu/home-affairs/system/files/2021-03/ran_ad-hoc_pap_fre_humor_20210215_en.pdf.

Fisher, Randy, Stuart Lilie, Clarice Evans, Greg Hollon, Mary Sands, Dawn Depaul, Christine Brady, et al. 1999. "Political Ideologies and Support for Censorship: Is It a Question of Whose Ox Is Being Gored?" *Journal of Applied Social Psychology* 29 (8): 1705–31. DOI: 10.1111/j.1559-1816.1999.tb02047.x.

Forsyth, Donelson R. 2018. *Group Dynamics*, 7th edition. Boston: Cengage Learning.

Friedkin, Noah E. 2004. "Social Cohesion." *Annual Review of Sociology* 30 (1): 409–25. DOI: 10.1146/annurev.soc.30.012703.110625.

Gates, Scott. 2002. "Recruitment and Allegiance: The Microfoundations of Rebellion." *Journal of Conflict Resolution* 46 (1): 111–30. DOI: 10.1177/0022002702046001007.

Grimmelmann, James. 2015. "The Virtues of Moderation." *Yale Journal of Law and Technology* 17: 42–109. https://openyls.law.yale.edu/handle/20.500.13051/7798.

Hatmaker, Taylor. 2018. "Facebook Is the Recruiting Tool of Choice for Far-Right Group the Proud Boys." *TechCrunch*, August 10. https://techcrunch.com/2018/08/10/proud-boys-facebook-mcinnes.

Haythornthwaite, Caroline. 1996. "Social Network Analysis: An Approach and Technique for the Study of Information Exchange." *Library & Information Science Research* 18 (4): 323–42. DOI: 10.1016/s0740-8188(96)90003-1.

Hoffman, Bruce. 2017. *Inside Terrorism*. New York: Columbia University Press. DOI: 10.7312/hoff17476.

Horta Ribeiro, Manoel, Homa Hosseinmardi, Robert West, and Duncan J. Watts. 2023. "Deplatforming Did Not Decrease Parler Users' Activity on Fringe Social Media." *PNAS Nexus* 2 (3): pgad035. DOI: 10.1093/pnasnexus/pgad035.

Hu, Clark, and Pradeep Racherla. 2008. "Visual Representation of Knowledge Networks: A Social Network Analysis of Hospitality Research Domain."

International Journal of Hospitality Management 27 (2): 302–12. DOI: 10.1016/j.ijhm.2007.01.002.
Huddy, Leonie. 2001. "From Social to Political Identity: A Critical Examination of Social Identity Theory." *Political Psychology* 22 (1): 127–56. DOI: 10.1111/0162-895x.00230.
Innes, Helen, and Martin Innes. 2023. "De-Platforming Disinformation: Conspiracy Theories and Their Control." *Information, Communication & Society* 26 (6): 1262–80. DOI: 10.1080/1369118x.2021.1994631.
Iqbal, Waleed, Muhammad Haseeb Arshad, Gareth Tyson, and Ignacio Castro. 2022. "Exploring Crowdsourced Content Moderation through Lens of Reddit during COVID-19." In *AINTEC '22: Proceedings of the 17th Asian Internet Engineering Conference, Hiroshima, Japan, December 19–21*, eds. Gareth Tyson and Simone Ferlin, 26–35. New York: Association for Computing Machinery. DOI: 10.1145/3570748.3570753.
Jaramillo, Ana María, Felipe Montes, Olga L. Sarmiento, Ana Paola Ríos, Lisa G. Rosas, Ruth F. Hunter, Ana Lucía Rodríguez, and Abby C. King. 2021. "Social Cohesion Emerging from a Community-Based Physical Activity Program: A Temporal Network Analysis." *Network Science* 9 (1): 35–48. DOI: 10.1017/nws.2020.31.
Jasko, Katarzyna, and Gary LaFree. 2020. "Who Is More Violent in Extremist Groups? A Comparison of Leaders and Followers." *Aggressive Behavior* 46 (2): 141–50. DOI: 10.1002/ab.21865.
Jenson, Jane. 2010. "Defining and Measuring Social Cohesion." Social Policies in Small States No. 2310-2136, October 25. London: Commonwealth Secretariat. DOI: 10.14217/9781848590724-en.
Jhaver, Shagun, Christian Boylston, Diyi Yang, and Amy Bruckman. 2021. "Evaluating the Effectiveness of Deplatforming as a Moderation Strategy on Twitter." *Proceedings of the ACM on Human–Computer Interaction* 5 (CSCW2): Article 381. DOI: 10.1145/3479525.
Jones, Seth G. 2022. *Rise of Far-Right Extremism in the United States*. Washington, DC: Center for Strategic and International Studies.
Kapoor, Kawaljeet Kaur, Kuttimani Tamilmani, Nripendra P. Rana, Pushp Patil, Yogesh K. Dwivedi, and Sridhar Nerur. 2018. "Advances in Social Media Research: Past, Present, and Future." *Information Systems Frontiers* 20 (3): 531–58. DOI: 10.1007/s10796-017-9810-y.
Kelley, Robert E. 1988. "In Praise of Followers." *Harvard Business Review*, November 1. https://hbr.org/1988/11/in-praise-of-followers.
——. 1992. *The Power of Followership: How to Create Leaders People Want to Follow and Followers Who Lead Themselves*. New York: Doubleday Currency.
Klein, Adam. 2019. "From Twitter to Charlottesville: Analyzing the Fighting Words between the Alt-Right and Antifa." *International Journal of Communication* 13: 297–318. https://ijoc.org/index.php/ijoc/article/view/10076.
Klinenberg, Danny. 2024. "Does Deplatforming Work?" *Journal of Conflict Resolution* 68 (6): 1199–1225. DOI: 10.1177/00220027231188909.
Knott, Kim, and Benjamin J. Lee. 2020. "Ideological Transmission in Extremist Contexts: Towards a Framework of How Ideas Are Shared." *Politics, Religion & Ideology* 21 (1): 1–23. DOI: 10.1080/21567689.2020.1732938.
Kontopantelis, Evangelos, Tim Doran, David A. Springate, Iain Buchan, and David Reeves. 2015. "Regression Based Quasi-Experimental Approach when Randomisation Is Not an Option: Interrupted Time Series Analysis." *BMJ* 350: h2750. DOI: 10.1136/bmj.h2750.
Kraus, Rachel. 2018. "2018 Was the Year We (Sort of) Cleaned Up the Internet." *Mashable*, December 26. https://mashable.com/article/deplatforming-alex-jones-2018.
Kriner, Matthew, and Jon Lewis. 2021. "Pride & Prejudice: The Violent Evolution of the Proud Boys." *CTC Sentinel* 14 (6): 26–38.
Kutner, Samantha. 2020. "Swiping Right: The Allure of Hyper Masculinity and Cryptofascism for Men Who Join the Proud Boys." ICCT Research Paper 2468-0664, May 26. The Hague: International Centre for Counter-Terrorism. DOI: 10.19165/2020.1.03.
Liang, Hai, Ng, Margaret Yee Man, and Nathan L. T. Tsang. 2023. "Word Embedding Enrichment for Dictionary Construction: An Example of Incivility in Cantonese." *Computational Communication Research* 5 (1): 1–26. DOI: 10.5117/ccr2023.1.10.lian.
Lust-Okar, Ellen. 2005. *Structuring Conflict in the Arab World: Incumbents, Opponents, and Institutions*. Cambridge, UK: Cambridge University Press. DOI: 10.1017/cbo9780511491009.
Mac, Ryan, and Blake Montgomery. 2018. "Twitter Suspended Proud Boys' and Founder Gavin McInnes' Accounts Ahead of the 'Unite the Right' Rally." *BuzzFeed News*, August 11. https://www.buzzfeednews.com/article/ryanmac/twitter-suspends-proud-boys-and-founder-gavin-mcinnes.
MacAvaney, Sean, Hao-Ren Yao, Eugene Yang, Katina Russell, Nazli Goharian, and Ophir Frieder. 2019. "Hate Speech Detection: Challenges and Solutions." *PLOS One* 14 (8): e0221152. DOI: 10.1371/journal.pone.0221152.
MacFarquhar, Neil. 2021. "Jury Finds Rally Organizers Responsible for Charlottesville Violence." *New York Times*, November 23. https://www.nytimes.com/2021/11/23/us/charlottesville-rally-verdict.html.

Makagon, Maja M., Brenda McCowan, and Joy A. Mench. 2012. "How Can Social Network Analysis Contribute to Social Behavior Research in Applied Ethology?" *Applied Animal Behaviour Science* 138 (3–4): 152–61. DOI: 10.1016/j.applanim.2012.02.003.

Malthaner, Stefan. 2018. "Spaces, Ties, and Agency: The Formation of Radical Networks." *Perspectives on Terrorism* 12 (2): 32–43. https://www.jstor.org/stable/26413312.

Manning, Willard G., and John Mullahy. 2001. "Estimating Log Models: To Transform or Not to Transform?" *Journal of Health Economics* 20 (4): 461–94. DOI: 10.1016/s0167-6296(01)00086-8.

Marwick, Alice, and Rebecca Lewis. 2017. "*Media Manipulation and Disinformation Online*." Technical report, May 15. New York: Data & Society Research Institute. https://datasociety.net/library/media-manipulation-and-disinfo-online.

Maurice, Emma Powys. 2019. "Milo Yiannopoulos 'Can't Put Food on the Table' and Might 'Retire from Social Media Entirely.'" *PinkNews*, September 9. https://www.thepinknews.com/2019/09/09/milo-yiannopoulos-cant-put-food-on-table-might-retire-social-media-entirely.

McBain, Sophie. 2020. "The Rise of the Proud Boys in the US." *New Statesman*, October 7. https://www.newstatesman.com/long-reads/2020/10/rise-proud-boys-us.

Mekacher, Amin, Max Falkenberg, and Andrea Baronchelli. 2023. "The Systemic Impact of Deplatforming on Social Media." *PNAS Nexus* 2 (11): pgad346. DOI: 10.1093/pnasnexus/pgad346.

Metaxas, Panagiotis, Eni Mustafaraj, Kily Wong, Laura Zeng, Megan O'Keefe, and Samantha Finn. 2015. "What Do Retweets Indicate? Results from User Survey and Meta-Review of Research." In *Proceedings of the International AAAI Conference on Web and Social Media* 9 (1): 658–61. DOI: 10.1609/icwsm.v9i1.14661.

Mellor, Philip A., and Chris Shilling. 2010. "Body Pedagogics and the Religious Habitus: A New Direction for the Sociological Study of Religion. ." *Religion* 40 (1): 27–38. DOI: 10.1016/j.religion.2009.07.001.

Moody, James, and Douglas R. White. 2003. "Structural Cohesion and Embeddedness: A Hierarchical Concept of Social Groups." *American Sociological Review* 68 (1): 103–27. DOI: 10.2307/3088904.

Moynihan, Harriet. 2021. "Deplatforming Trump Puts Big Tech under Fresh Scrutiny." Expert Comment, January 22. London: Chatham House. https://www.chathamhouse.org/2021/01/deplatforming-trump-puts-big-tech-under-fresh-scrutiny.

Newman, Mark. 2018. *Networks*, 2nd edition. Oxford, UK: Oxford University Press. DOI: 10.1093/oso/9780198805090.001.0001.

Newstrom, John W. 2000. *Comportamiento humano en el trabajo* [Human behavior at work], 10th edition. New York: McGraw Hill.

O'Callaghan, Derek, Derek Greene, Maura Conway, Joe Carthy, and Pádraig Cunningham. 2013. "An Analysis of Interactions within and between Extreme Right Communities in Social Media." In *Ubiquitous Social Media Analysis*, eds. Martin Atzmueller, Alvin Chin, Denis Helic, and Andreas Hotho, 88–107. Cham: Springer. DOI: 10.1007/978-3-642-45392-2_5.

Paunova, Minna. 2015. "The Emergence of Individual and Collective Leadership in Task Groups: A Matter of Achievement and Ascription." *Leadership Quarterly* 26 (6): 935–57. DOI: 10.1016/j.leaqua.2015.10.002.

Phillips, Whitney, and Ryan M. Milner. 2018. *The Ambivalent Internet: Mischief, Oddity, and Antagonism Online*. Cambridge, UK: Polity.

Rauchfleisch, Adrian, and Jonas Kaiser. 2021. "Deplatforming the Far-Right: An Analysis of YouTube and BitChute." SSRN preprint, updated June 25, 2021. DOI: 10.2139/ssrn.3867818.

Reid, Shannon E., and Matthew Valasik. 2020. *Alt-Right Gangs: A Hazy Shade of White*. Berkeley, CA: University of California Press. DOI: 10.1525/9780520971844.

Reid, Shannon E., Matthew Valasik, and Arunkumar Bagavathi. 2020. "Examining the Physical Manifestation of Alt-Right Gangs: From Online Trolling to Street Fighting." In *Understanding Gangs in the Era of Internet and Social Media*, eds. Chris Melde and Frank Weerman, 105–34. Cham: Springer. DOI: 10.1007/978-3-030-47214-6_6.

Rogers, Richard. 2020. "Deplatforming: Following Extreme Internet Celebrities to Telegram and Alternative Social Media." *European Journal of Communication* 35 (3): 213–29. DOI: 10.1177/0267323120922066.

Rosenberg, Eli. 2018. "FBI Considers Proud Boys Extremists with White-Nationalist Ties, Law Enforcement Officials Say." *Washington Post*, November 19. https://www.washingtonpost.com/nation/2018/11/20/fbi-says-proud-boys-have-white-nationalist-ties-law-enforcement-officials-say.

Sachdeva, Pratik, Renata Barreto, Geoff Bacon, Alexander Sahn, Claudia Von Vacano, and Chris Kennedy. 2022. "The Measuring Hate Speech Corpus: Leveraging Rasch Measurement Theory for Data Perspectivism." In *Proceedings of the 1st Workshop on Perspectivist Approaches to NLP @LREC2022*, eds. Gavin Abercrombie, Valerio Basile, Sara Tonelli, Verena Rieser, and Alexandra Uma, 83–94. Marseille: European Language Resources Association. https://aclanthology.org/2022.nlperspectives-1.11.

Saleem, Haji Mohammad, and Derek Ruths. 2018. "The Aftermath of Disbanding an Online Hateful Community." arXiv preprint arXiv:1804.07354, version 1, updated April 19, 2018. DOI: 10.48550/arXiv.1804.07354.

Schaffer, Andrea L., Timothy A. Dobbins, and Sallie-Anne Pearson. 2021. “Interrupted Time Series Analysis Using Autoregressive Integrated Moving Average (ARIMA) Models: A Guide for Evaluating Large-Scale Health Interventions.” *BMC Medical Research Methodology* 21 (1): Article 58. DOI: 10.1186/s12874-021-01235-8.

Schmid, Katharina, and Orla T. Muldoon. 2015. “Perceived Threat, Social Identification, and Psychological Well-Being: The Effects of Political Conflict Exposure.” *Political Psychology* 36 (1): 75–92. DOI: 10.1111/pops.12073.

Schwarzenegger, Christian, and Anna J. M. Wagner. 2018. “Can It Be Hate if It Is Fun? Discursive Ensembles of Hatred and Laughter in Extreme Right Satire on Facebook.” *Studies in Communication and Media* 7 (4): 473–98. DOI: 10.5771/2192-4007-2018-4-473.

Seashore, Stanley Emanuel. 1954. *Group Cohesiveness in the Industrial Work Group*. Ann Arbor, MI: University of Michigan Institute for Social Research.

Selim, Heyla A., and Maša Popovac. 2024. “Social Cohesion in an Online Era: Opportunities and Challenges on Social Media.” In *Handbook of Social Media in Education, Consumer Behavior, and Politics*, Vol. 1, ed. Vladlena Benson, 279–98. Cambridge, MA: Academic Press. DOI: 10.1016/b978-0-323-90237-3.00013-8.

Severt, Jamie B., and Armando X. Estrada. 2015. “On the Function and Structure of Group Cohesion.” In *Team Cohesion: Advances in Psychological Theory, Methods and Practice*, Vol. 17, eds. Eduardo Salas, William B. Vessey, and Armando X. Estrada, 3–24. Leeds: Emerald Group. DOI: 10.1108/s1534-085620150000017002.

Shamir, Boas. 2007. “From Passive Recipients to Active Co-Producers: Followers’ Roles in the Leadership Process.” In *Follower-Centered Perspectives on Leadership: A Tribute to the Memory of James R. Meindl*, eds. Boas Shamir, Rajnandini Pillai, Michelle C. Bligh, and Mary Uhl-Bien, 39–60. Charlotte, NC: Information Age.

Sherif, Muzafer. 1966. *In Common Predicament: Social Psychology of Intergroup Conflict and Cooperation*. Boston: Houghton Mifflin.

Simpson, Ellen. 2018. “Integrated & Alone: The Use of Hashtags in Twitter Social Activism.” In *CSCW ’18: Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing, Jersey City, NJ, November 3–7*, eds. Mauro Cherubini, Andrew Miller, and Alyson Young, 237–40. New York: Association for Computing Machinery. DOI: 10.1145/3272973.3274064.

Stein, Arthur A. 1976. “Conflict and Cohesion: A Review of the Literature.” *Journal of Conflict Resolution* 20 (1): 143–72. DOI: 10.1177/002200277602000106.

Stern, Yaakov, Anna MacKay-Brandt, Seonjoo Lee, Paula McKinley, Kathleen McIntyre, Qolamreza Razlighi, Emil Agarunov, Matthew Bartels, and Richard P. Sloan. 2019. “Effect of Aerobic Exercise on Cognition in Younger Adults: A Randomized Clinical Trial.” *Neurology* 92 (9): e905–16. DOI: 10.1212/WNL.0000000000007003.

Stokman, Frans N. 2001. “Networks: Social.” In *International Encyclopedia of the Social & Behavioral Sciences*, eds. Neil J. Smelser and Paul B. Baltes, 10509–14. Oxford, UK: Pergamon. DOI: 10.1016/B0-08-043076-7/01934-3.

Stott, Clifford, and John Drury. 2017. “Contemporary Understanding of Riots: Classical Crowd Psychology, Ideology, and the Social Identity Approach.” *Public Understanding of Science* 26 (1): 2–14. DOI: 10.1177/0963662516639872.

Sumner, William Graham. 1906. *Folkways: A Study of the Sociological Importance of Usages, Manners, Customs, Mores, and Morals*, ed. Albert Galloway Keller. Cambridge, MA: Ginn. https://archive.org/details/folkwaysstudyofssumn.

Sy, Thomas. 2010. “What Do You Think of Followers? Examining the Content, Structure, and Consequences of Implicit Followership Theories.” *Organizational Behavior and Human Decision Processes* 113 (2): 73–84. DOI: 10.1016/j.obhdp.2010.06.001.

Tajfel, Henri, and John C. Turner. 1979. “An Integrative Theory of Intergroup Conflict.” In *The Social Psychology of Intergroup Relations*, eds. William G. Austin and Stephen Worchel, 33–37. Monterey, CA: Brooks/Cole.

Tenkanen, Henrikki, Enrico Di Minin, Vuokko Heikinheimo, Anna Hausmann, Marna Herbst, Liisa Kajala, and Tuuli Toivonen. 2017. “Instagram, Flickr, or Twitter: Assessing the Usability of Social Media Data for Visitor Monitoring in Protected Areas.” *Scientific Reports* 7 (1): 17615. DOI: 10.1038/s41598-017-18007-4.

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. DOI: 10.1073/pnas.2214080120.

Tsai, Chia-Ching, Yung-Kai Yang, and Yu-Chi Cheng. 2014. “Does Relationship Matter? Customers’ Response to Service Failure.” *Managing Service Quality* 24 (2): 139–59. DOI: 10.1108/MSQ-06-2013-0113.

Uhl-Bien, Mary, and John M. Maslyn. 2000. “Examining the Exchange in Leader–Member Exchange (LMX): Identification of Dyadic Relational Styles and Their Association with Key Attitudes and Behaviors.” *Academy of Management Proceedings* 2000 (1): K1–6. DOI: 10.5465/apbpp.2000.5535191.

Uhl-Bien, Mary, Ronald E. Riggio, Kevin B. Lowe, and Melissa K. Carsten. 2014. "Followership Theory: A Review and Research Agenda." *Leadership Quarterly* 25 (1): 83–104. DOI: 10.1016/j.leaqua.2013.11.007.

Van Prooijen, Jan-Willem, and André P. M. Krouwel. 2019. "Psychological Features of Extreme Political Ideologies." *Current Directions in Psychological Science* 28 (2): 159–63. DOI: 10.1177/0963721418817755.

Venna, Siva R., Raju Gottumukkala, and Vijay V. Raghavan. 2016. "Visual Analytic Decision-Making Environments for Large-Scale Time-Evolving Graphs." In *Handbook of Statistics*, Vol. 35, eds. Venkat N. Gudivada, Vijay V. Raghavan, Venu Govindaraju, and C. R. Rao, 81–115. Amsterdam: Elsevier. DOI: 10.1016/bs.host.2016.07.002.

Vu, Anh V., Alice Hutchings, and Ross Anderson. 2024. "No Easy Way Out: The Effectiveness of Deplatforming an Extremist Forum to Suppress Hate and Harassment." In 2024 *IEEE Symposium on Security and Privacy (SP)*, San Francisco, *19–23 May*, eds. Ziming Zhao and Syed Hussain, 717–34. New York: Institute of Electrical and Electronics Engineers. DOI: 10.1109/sp54263.2024.00007.

Wang, Rong, Wenlin Liu, and Shuyang Gao. 2016. "Hashtags and Information Virality in Networked Social Movement: Examining Hashtag Co-Occurrence Patterns." *Online Information Review* 40 (7): 850–66. DOI: 10.1108/oir-12-2015-0378.

Wasserman, Stanley, and Katherine Faust. 1994. *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511815478.

Watanabe, Hajime, Mondher Bouazizi, and Tomoaki Ohtsuki. 2018. "Hate Speech on Twitter: A Pragmatic Approach to Collect Hateful and Offensive Expressions and Perform Hate Speech Detection." *IEEE Access* 6: 13825–35. DOI: 10.1109/access.2018.2806394.

Weaponized Word. 2025. Lexicons [lexicographic database]. Weaponized Word. https://weaponizedword.org/lexicons.

Wegner, Eva. 2011. *Islamist Opposition in Authoritarian Regimes: The Party of Justice and Development in Morocco*. Syracuse, NY: Syracuse University Press.

West, Sarah Myers. 2018. "Censored, Suspended, Shadowbanned: User Interpretations of Content Moderation on Social Media Platforms." *New Media & Society* 20 (11): 4366–83. DOI: 10.1177/1461444818773059.

Wiegand, Michael, Josef Ruppenhofer, Anna Schmidt, and Clayton Greenberg. 2018. "Inducing a Lexicon of Abusive Words: A Feature-Based Approach." In *Long Papers*, Vol. 1 of *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, eds. Marilyn Walker, Heng Ji, and Amanda Stent, 1046–56. Stroudsburg, PA: Association for Computational Linguistics. DOI: 10.18653/v1/N18-1095.

Wilson, Jason. 2018a. "FBI Now Classifies Far-Right Proud Boys as 'Extremist Group,' Documents Say." *Guardian*, November 19. https://www.theguardian.com/world/2018/nov/19/proud-boys-fbi-classification-extremist-group-white-nationalism-report.

——. 2018b. "Proud Boys Founder Gavin McInnes Quits 'Extremist' Far-Right Group." *Guardian*, November 22. https://www.theguardian.com/world/2018/nov/22/proud-boys-founder-gavin-mcinnes-quits-far-right-group.

Wong, Julia Carrie. 2018. "Don't Give Facebook and YouTube Credit for Shrinking Alex Jones' Audience." *Guardian*, September 5. https://www.theguardian.com/commentisfree/2018/sep/04/alex-jones-infowars-social-media-ban.

Youngblood, Mason. 2020. "Extremist Ideology as a Complex Contagion: The Spread of Far-Right Radicalization in the United States between 2005 and 2017." *Humanities and Social Sciences Communications* 7 (1): Article 49. DOI: 10.1057/s41599-020-00546-3.

Zhang, Xinyi, and Mark Davis. 2024. "E-Extremism: A Conceptual Framework for Studying the Online Far Right." *New Media & Society* 26 (5): 2954–70. DOI: 10.1177/14614448221098360.

Zhong, Wei, Catie Bailard, David Broniatowski, and Rebekah Tromble. 2024. "Proud Boys on Telegram." *Journal of Quantitative Description: Digital Media* 4: 1–47. DOI: 10.51685/jqd.2024.003.

Zhong, Wei, and Maggie Mengqing Zhang. 2025. "Replication Data for: Subdued but Unbroken: The Cohesion of Far-Right Extremist Followers after Deplatforming by Twitter." *Harvard Dataverse.* DOI: 10.7910/DVN/DVFCT3.