Supporting Information

January 6 Arrests and Media Coverage Do Not Remobilize Conservatives on Social Media

Ross Dahlke and Jennifer Pan

Correspondence to: rdahlke@stanford.edu

Contents

S1 Social Media Data

First, we identified communities of elite right-wing Twitter users found by past research [3], namely the "#MAGA," "Christian Constitutionalists," "Conservative Media," "Tea Party Conservatives," and "White Nationalists" groups. We searched for all posts from these individuals via their Twitter usernames on Twitter and Gab from January 6, 2021, to April 30, 2022. Data from Twitter, now called "X," were collected via the Academic API v2.[S1] Data were collected before recent changes to Twitter/X's API and academic research data access policies, potentially making replication of the data collection difficult. Gab data were collected via the 'gabber'[S2] tool developed by the Stanford Internet Observatory [2]. This tool uses Gab's open API to access an authenticated end-point while respecting rate limits requested by Gab. Data were collected during the summer of 2023 on a rolling basis as daily rate limits were reached.

Due to relatively low username match rates on Gab for elite ring-wing users, we also analyzed two communities of Gab users who may remobilize: individuals who created a Gab account on January 6 or posted about January 6 between January 6, 2021, and April 30, 2022 (posts with the keywords "January 6," "Jan 6" or "J6"). Since the volume of posts from these individuals spiked on January 6, 2021, we conducted non-parametric testing on posts from April 2, 2021 to April 30, 2022, with cut dates are drawn from May 1, 2021 to March 31, 2022 (a 30-day window before/after earliest cut date for the non-parametric analysis), so that posting volumes would not be affected by this large burst. During this time period, we collected 10,657,702 tweets and 3,174,796 gab posts from these individuals.

To classify whether these posts were about January 6, two independent coders hand-coded a subset of posts to determine whether the posts were about January 6 (12,609 tweets, 4,698 gab posts, Cohen's Kappa = .84). Then, we trained two RoBERTa-based classifiers using these data (one for Twitter and Gab) with five-fold cross-validation, selecting the model with the best performance (F1) to apply to the remainder of the posts. These models achieved high accuracy (Twitter model: accuracy = 99.6%, precision = 81.6%, recall = 88.6%, F1 = .85; Gab model: accuracy = 99.7%, precision = 77.8%, recall = 87.5%, F1 = .82). We classified posts with a greater than 50% probability of being about January 6 as about January 6. This step left us with 102,552 tweets and 12,268 Gab posts about January 6 from May 1, 2021, to April 30, 2022.

[S1]https://developer.twitter.com/en/docs/twitter-api/early-access
[S2]https://github.com/stanfordio/gabber

S2   Coding Instructions & Scheme

To identify social media posts about StoptheSteal and January 6 for the classifier training data, human coders labeled posts as "1" if they were about StoptheSteal and January 6 and "0" if not. Specifically, the human coders were given the following coding scheme.

> Thank you for helping us with our project on the January 6 defendants. In this project, we examine the extent to which political influencers continued talking about January 6 after the event. We need your help hand-coding social media messages (Twitter and Gab) about January 6. You will go through these messages in a Google Sheet, and in the "J6" column, code a "1" if the message pertains to the events of January 6 and a "0" if it does not.
>
> Some of these messages explicitly mention "J6," "Jan. 6," or "January 6." However, many do not, especially messages sent shortly after January 6, 2021. Some other common themes to look out for as indications of speaking about January 6 (code as "1") include:
>
> - **Mentions of "the mob," "protesters," an event "at the capitol," or a "breach"**; e.g., "Some say Antifa planned to infiltrate &; cause problems."
>
> - **Mentions of "the insurrection," or "insurrectionists"** (But only in relation to J6 or StoptheSteal, not if the post is just using the word to describe Democrats or other people)
>
> - **False comparison with other movements**; e.g., "why are people mad when we protest but not at antifa?"
>
> - **Mentions of a rally**; e.g., "Say what you will. If the tables were turned, the Left would not be rallying by the millions, fighting for Joe Biden."
>
> - **Mentions of the January 6 congressional committee**; particularly targeted towards Liz Cheney and Adam Kinzinger or the DOJ, indictments of J6 defendants and other notable participants/inciters of the violence.
>
> - **Mentions of voter fraud or StoptheSteal**; the J6 movement is a continuation of the broader StoptheSteal movement, which is protesting the election outcome due to accusations of voter fraud.

S3 Focal Points

To determine focal points where we would expect to see remobilization, we examine arrests of January 6 participants and media coverage of January 6. We collected arrest data from the United States Department of Justice's official list of "Capitol Breach Cases."[S3] Focal points (arrests and media stories) were calculated between May 1, 2021, and March 31, 2022, to ensure that there were comparable 30-day windows for all potential focal points. We calculated the number of arrests of January 6 defendants per day. Then, we calculated which days were outliers in terms of the number of arrests by taking days that were at least two standard deviations above the mean number of arrests per day. We collected media coverage data from organizations identified as "U.S. major news organizations" by MediaCloud [S4] using the search terms "January 6," "J6," or "Jan 6." Similarly, we calculated days with an elevated number of stories by calculating outlier days that were two standard deviations above the mean in terms of stories per day (see Figure 2 in main text). As seen in Figure 1 in the main text, the volume of posts spiked on the one-year anniversary of the January 6 U.S. Capitol attack (January 6, 2022). Thus, we do not test calculated focal points that are within 30 days of January 6, 2022.

[S3]justice.gov/usao-dc/capitol-breach-cases
[S4]https://www.mediacloud.org/

S4   Non-Parametric Testing

We conduct non-parametric testing following past work on social media mobilization in response to government action [1]. We first calculate the change in the mean daily volume of posts ±30 days around the identified focal points. Then, we conducted placebo tests to generate a null distribution of the mean daily changes in volume by choosing a placebo intervention date at random (between May 1, 2021, and March 31, 2022, to ensure a comparable 30-day window for all draws in the distribution) and repeating this procedure 1,000 times. We use this null distribution to non-parametrically test whether the changes in post volume around the focal points fall outside the mass of the distribution of changes in volume generated by choosing placebo dates at random. Specifically, we calculate two-sided p-values with 95% confidence intervals (allowing for the hypotheses of remobilization and demobilization) representing the proportion of placebo tests that are at least the size of the actual observed changes in volume around the focal points (see Figure 3 in the main text).

Formally, let ΔV denote the change in the mean daily volume of posts in the ±30-day window surrounding a given focal point, where V_{post} is the mean volume post-intervention and V_{pre} is the mean volume pre-intervention. This value is calculated as:

$$\Delta V = V_{\text{post}} - V_{\text{pre}}$$

For the placebo tests, let $\Delta V_{\text{null},i}$ represent the change in mean daily volume for the i-th placebo intervention date. The set of $\Delta V_{\text{null},i}$ over $i = 1, \ldots, 1000$ random dates between May 1, 2021, and March 31, 2022, comprises our null distribution, $\mathcal{D}_{\text{null}}$.

To test the significance of ΔV, we compute the two-tailed empirical p-value, which is the proportion of the null distribution that is as extreme as or more extreme than our observed ΔV. Mathematically, this is represented as:

$$p\text{-value} = \frac{|\{\Delta V_{\text{null},i} \in \mathcal{D}_{\text{null}} : |\Delta V_{\text{null},i}| \geq |\Delta V|\}|}{1000}$$

The numerator of the p-value formula, $|\{\Delta V_{\text{null},i} \in \mathcal{D}_{\text{null}} : |\Delta V_{\text{null},i}| \geq |\Delta V|\}|$, counts the number of times the change in volume from the placebo tests is at least as large as the change we observed around our focal point. This count is then divided by the total number of placebo tests (1000 in this case), which is the denominator. If this proportion is small, it suggests that such an extreme change is rare when there is no actual intervention, leading us to believe that our observed change is likely not due to random chance. Conversely, a large p-value indicates that the observed change is not unusual and could easily occur without any actual intervention. We define an observation as statistically significant if it falls outside the 95% confidence interval derived from $\mathcal{D}_{\text{null}}$, corresponding to a p-value less than 0.05. The two-tailed nature of the test considers both directions of change—increases and decreases—since we are interested in deviations of any kind from the null scenario. This allows us to account for both potential remobilization, where $\Delta V > 0$, and demobilization, where $\Delta V < 0$.

To visualize the test results and the significance of the observed changes, we refer to Figure 3 in the main text, which plots ΔV against the backdrop of $\mathcal{D}_{\text{null}}$, highlighting areas of statistical significance where the observed data diverge from what might be expected by chance alone.

S5 Reanalysis of Full Data

We primarily analyzed the time period of April 2, 2021, to April 30, 2022. Focal points (arrests and media stories) and the null distribution of cut dates were drawn from May 1, 2021 (30 days after April 2, 2021, to maintain a 30-day window before/after the cut dates) to March 31, 2022 (30 days before April 30, 2022, to maintain a 30-day window before/after the cut dates). We do not begin analysis on January 6, 2021 because volumes of social media posts about the StoptheSteal movement and January 6 declined dramatically after their peak on January 6, 2021. Therefore, it is possible that analyses examining immediately after January 6, 2021, could be confounded by this decline back to "normal" levels of mobilization on social media. In other words, declines detected in our analysis may not genuinely represent demobilization but instead represent the ongoing decline after January 6, 2021. However, the period between the events of January 6 and the beginning of our primary analysis period of April 2, 2021, remains important for the movement. Sustained mobilization after focal points (i.e., arrests and media coverage) during this period of time could potentially translate to long-term movement sustainability.

Therefore, we analyze all data from, January 6, 2021 to April 30, 2022. Focal points (arrests and media stories) and the null distribution of cut dates were drawn from February 5, 2021 (30 days after January 6, 2021) to March 31, 2022 (30 days before April 30, 2022). Then, we repeated our non-parametric analysis). Similar to the primary analysis, we observe mostly null results. A few of the newly-added focal points were significant and negative (indicating a demobilization): February 8, 2021 (Gab: Posted About January 6, Tea Party Conservatives; Twitter: #MAGA, Christian Constitutionalists, Conservative Media, White Nationalists), February 9, 2021 (Gab: Posted About January 6, Tea Party Conservatives; Twitter: #MAGA, Tea Party Conservatives), February 10, 2021 (Gab: Posted About January 6; Twitter: #MAGA), February 11, 2021 (Gab: Tea Party Conservatives; Twitter: #MAGA), February 12, 2021 (Gab: Posted About January 6, Tea Party Conservatives; Twitter: #MAGA), February 13, 2021 (Gab: Tea Party Conservatives), February 15, 2021 (Gab: Tea Party Conservatives), February 16, 2021 (Gab: Tea Party Conservatives), February 19, 2021 (Gab: Tea Party Conservatives), February 23, 2021 (Gab: Christian Constitutionalists), February 25, 2021 (Gab: #MAGA, Christian Constitutionalists), March 4, 2021 (Gab: #MAGA, Christian Constitutionalists, Created Account on January 6), March 9, 2021 (Gab: #MAGA, Created Account on January 6). However, all of these significant results are **declines** in mobilization before the start of our primary analysis period (i.e., before May 1, 2021). These results may suggest that mobilization on social media declined after focal points (i.e., bursts in arrests or media coverage) in the months immediately after January 6, 2021. However, these declines could also be due to a general downward trend after the initial events of January 6. In other words, since the test is examining absolute change, instead of relative change, these points may look large compared to the rest of the points because they are part of the decline that starts at a very high point of activity.

What is important is that this analysis demonstrates that not even in this critical period when the movement could have found sustainability did arrests or media coverage remobilize these groups of conservatives.

S6 Time Window Robustness Check

The null effects in the manuscript's primary analysis use a ± 30 30-day window. However, it is possible that changes in mobilization levels are only measurable at a smaller time scale. To test for this possibility, we reran our analysis using time windows ranging from five days to thirty days daily). While two of the significant focal points at a 30-day time window remain significant at slightly shorter time windows, they are not consistently significant for all time windows. No other focal points are consistently significant at shorter time windows.

S7   Interrupted Time Series Robustness Check

We also conducted parametric interrupted time series analyses (ITSA) as a robustness check to the non-parametric method used in the main manuscript. We follow the method used by [1]. Formally, we estimate

$$Y_t = \beta_0 + \beta_1(T) + \beta_2(X_t) + \beta_3(X_t T)$$

.

    In this equation: Y_t represents the number of tweets or Gab posts at time t; T is the time (number of days) since the focal point; X_t is a dummy variable representing the occurrence of the focal point (coded as 0 before the event and 1 after); $X_t T$ is the interaction term of time (number of days since the focal point) and the occurrence of the focal point; β_0 represents the baseline volume of tweets or Gab posts at $t = 0$; β_1 shows the gradual per-day change in post volume, therefore representing the underlying daily trend; β_2 captures the change in intercept (level change) of the focal point on the volume of tweets or Gab posts; β_3 captures the slope change in the daily volume of tweets or Gab posts after the focal point, relative to the pre-focal point trend. This segmented regression model allows us to measure the pre-event trend, the immediate change in volume following the event, and the gradual change in the volume after the focal point. We used a first-order autoregressive (AR1) model instead of the standard Ordinary Least Squares (OLS) ITSA model to address serial autocorrelation in the data.

    Substantively, β_2 or β_3 could represent significant, albeit different, responses to the focal points. In other words, a significant β_2 represents the level change in mobilization levels immediately after the focal point, and β_3 represents the gradual but enduring change in mobilization levels after the focal point. A strength of this method is that it allows us to decompose these possible responses to understand the potential type of change in mobilization occurring after focal points more precisely. We fit this model for each focal point for each group analyzed.

    In terms of significant level changes, on May 14, 2021, there is a level increase for #MAGA and White Nationalists; on May 19 2021, there is a level increase for #MAGA and Christian Constitutionalists, and on May 27, 2021, there is a level increase for Christian Constitutionalists and Conserative Media. On May 28, 2021, there is a negative level change for Christian Constitutionalists, Conservative Media, and Tea Party Conservatives. On July 28, 2021 discussions among #MAGA, Christian Constitutionalists, Tea Party Conservatives, and White Nationlists experience a significant level change in the negative direction. July 28, 2021 is a focal point representing a burst in media stories about January 6. We do not observe a negative level change for Conservative Media on July 28, 2021 but do so on July 27, 2021.

    In terms of significant coefficients, on August 19, 2021, Christian Constitutionalists and anyone who posted about January 6 experienced a level increase. On July 27 and 28, 2021, #MAGA and Conservative Media experienced a negative slope change. On

December 1, 2021, Christian Constitutionalists and those who created an account on January 6 experienced a positive slope change, but White Nationalists experienced a negative slope change.

References

[1] J. Pan and A. A. Siegel. How saudi crackdowns fail to silence online dissent. *American Political Science Review*, 114(1):109–125, 2020.

[2] D. Thiel and M. McCain. Gabufacturing dissent: An in-depth analysis of gab, 2022.

[3] Y. Zhang, F. Chen, and K. Rohe. Social media public opinion as flocks in a murmuration. *Journal of Computer-Mediated Communication*, 27(1), 2022.