From Tweets to the Streets. Twitter and Protest Participation in the United States

Gisli Gylfason

PARIS SCHOOL OF ECONOMICS
ECOLE D'ECONOMIE DE PARIS

MASTER THESIS N° 2022 – 03

From Tweets to the Streets.
Twitter and Protest Participation in the United States

Gisli Gylfason

JEL Codes: D70, L86, L82, P16.
Keywords: political economy; collective action; protests; social media; information technology.

L'ÉCOLE DES HAUTES ÉTUDES EN SCIENCES SOCIALES
École des Ponts ParisTech
ENS
PSL RESEARCH UNIVERSITY PARIS
PARIS SCHOOL OF ECONOMICS ECOLE D'ECONOMIE DE PARIS
CEPREMAP
CENTRE POUR LA RECHERCHE ECONOMIQUE ET SES APPLICATIONS
IRD

From Tweets to the Streets

Twitter and Protest Participation in the United States

Master Thesis: Public Policy and Development

Paris School of Economics

Submitted by:
Gísli Gylfason
Year
2021-2022

Supervision:
Ekaterina Zhuravskaya
Referee:
Oliver Vanden Eynde

Paris, 31/05/2022

From Tweets to the Streets: Twitter and Protest Participation in the United States*

Gísli Gylfason

May 31, 2022

Abstract

This paper investigates whether social media can influence the landscape of political protests beyond the size and frequency of protest events. Using early adoption of Twitter at the 2007 South by Southwest (SXSW) festival as a plausibly exogenous source of variation in county-level Twitter penetration, I find that Twitter penetration increases protest frequency in the United States substantially. I show that the effect is stronger for protest movements where coordination is more challenging, due to the scale of the coordination problem or when organizational capacities are low. However, these heterogeneous effects are not large enough to imply drastic changes in the protest landscape. I do not find evidence for heterogeneous effects depending on the topic of protest. Finally, I find that Twitter penetration increases the relative frequency of the use of violence or attempts at suppression by non-government groups during protest events. I find evidence suggesting that social media increases violence during protest events by easing coordination among those opposing the protest.

Keywords: Political Economy, Collective action, Protests, Social media, Information Technology

JEL codes: D70, L86, L82, P16.

*I am grateful to Ekaterina Zhuravskaya for invaluable supervision, guidance and support on this project. I thank Oliver Vanden Eynde for accepting to be my referee. I am also grateful to participants at the GPET dissertation seminar, Hillel Rapoport, Sandra Poncet, and fellow master students, for their excellent feedback and advice. I thank Alvin Opler for help with coding in a time of need. Finally, I am deeply thankful to my fellow PPD students for all their support and inspiration.

Contents

1 Introduction

In recent years, the globe has witnessed increasing waves of protest movements such as the Arab Spring, food riots or the Black Lives Matter Movement (Ortiz et al., 2013). In the politically polarized environment of the United States, recent reports document not only rising protest activity, but also that demonstrations are increasingly met with interference or violence from the state as well as from non-state actors (ACLED, 2020c). These developments have coincided with the rapid expansion of Information and Communication Technologies (ICTs) in general and social media in particular, creating widespread perception about their integral role in facilitating mobilization for these movements. This perception is illustrated by traditional media's reporting of social media as instrumental for deeply dividing protest movements such as the "Stop the Steal" movement that culminated in the Capital Riots during the 2021 US presidential transition (Brewster, 2021). The academic literature has explored theoretical mechanisms through which ICTs and social media can facilitate protest movements and confirmed empirically the causal effect of social media on protests. However, the best available empirical evidence is either limited to case-studies of a particular protest movement (Enikolopov, Makarin, & Petrova, 2020; Amorim et al., 2018) or studies taking a more global approach, treating protests as a black box (Manacorda & Tesei, 2020; Fergusson & Molina, 2019). Does social media impact interference or violence during protests in democratic states? If so, who commits violent acts? Do social movements of specific characteristics respond to social media? While increased violence or interference during demonstrations are worrying, understanding whether social media influences particular protest movements is important as evidence shows that protests may not only have persuasive power over policy or bring about social change (Madestam et al., 2013; Skoy, 2021; Klein Teeselink & Melios, 2021), but can induce polarization around the topic of demonstration (Caprettini et al., 2022). In this paper, I use detailed georeferenced demonstration data from the United States to explore whether social media can influence the landscape of political demonstrations beyond the size and frequency of demonstrations.

I contribute to the understanding of these issues by studying the effect of Twitter penetration on various protest movements, depending on the characteristics the movements and of the demonstration events themselves, in contemporary United States. The United States provide a great setting to test whether social media disproportionately affects some "types" of protest movements, or violence levels during demonstration events, as the country has long been a vibrant protest environment,

while demonstrations have been surging. US citizens demonstrate for various separate causes and events vary widely in terms of organization, reactions or violence levels. Following Müller & Schwarz (2020) and Fujiwara et al. (2021), my identification relies on a plausibly exogenous shock to Twitter's early adoption, the South By Southwest festival (SXSW) in Austin, Texas in 2007. Promotion of Twitter at the festival sparked it's popularity among attendees and, through network effects, in their home counties. Still to this day, proxies of county-level attendance at the festival has predictive power over regional variation in Twitter penetration. I use this shock, while controlling for pre-2007 general interest in the South By Southwest festival, in an instrumental variable framework.

I find that the magnitude of the effect of Twitter penetration on protest frequency is substantial. I estimate that an additional individual at the 2007 South by Southwest festival (proxied by the number of Twitter users who follow the account of the South by Southwest festival (@SXSW) and created their accounts around the time of the festival in 2007) increases Twitter penetration (proxied by the number of tweets from randomly drawn sample of tweets) by 5% and the number of demonstrations between January 2020 to 12th of November 2021 by 13 (from the mean level of 11) in their home counties. The two stage least squares estimates indicates that a percentage increase in Twitter penetration implies 2.75 additional demonstration events between January 2020 to 12th of November 2021.

I use the Armed Conflict Location and Event Data, aggregated at the county-level, as the information source on demonstration activity. The dataset includes information on attempts at suppression, the use of violence, and organizations associated with each demonstration event as well as qualitative information on the cause for demonstrating. I use this information to perform three heterogeneity analyses. First, I use information on the organizations associated with a demonstration event along with whether the event was part of a broad social movement calling for change at the national (or even pan-national) level to approximate "coordination complexity." Assuming that events not associated with any organization are more difficult to coordinate due to lack of organizational capacity, and that broad social movements are harder to coordinate due to the scale of coordination needed for them to succeed, I do indeed find larger effects of Twitter penetration on demonstrations that without social media would be more complex to coordinate. However, these heterogeneities are not large enough to imply fundamental changes in the demonstration landscape. An additional follower implies that the share of demonstrations with lower organizational capacity or higher coordination complexity among all demonstrations

increases by 0.6 percentage points (from a mean of 72%). Second, I use qualitative information on each demonstration event to filter out those demonstrations that concern particularly polarizing topics or hateful sentiments. I do find that Twitter penetration, proxied by the number of tweets from a large sample of randomly drawn tweets, increases demonstrations concerning polarizing topics, as well as those associated with groups with affiliations to white supremacy ideology. However, increased Twitter penetration does not imply an increase in the relative frequency of these demonstrations. Third, I use information on non-peaceful activities during demonstration events to disentangle whether demonstrators themselves, government entities such as police or military units, or non-government outside parties exhibit disruptive behaviour, engage in violent acts or attempt to suppress the demonstration event. I find no evidence that Twitter penetration affects the relative frequency of non-peaceful acts by demonstrators, nor the relative frequency of government violence or attempts at suppression. However, Twitter penetration seems to increase the relative frequency of violent acts or attempts at suppression during demonstration events by non-government parties. An additional follower implies that the share of demonstrations where peaceful protesters are met with attempts of suppression or violence by non-government parties among all demonstrations increases by 0.04 percentage points (from a mean of 0.3%). Although this effect is not large enough to explain reported increases in violence during demonstration events, it should not be disregarded.

It is important to note that, as I use an instrumental variable framework, my estimates refer to a Local Average Treatment Effect for complier counties. That is, counties where Twitter penetration responds strongly to being home to an individual who attended the SXSW Festival in 2007. Early adoption of Twitter among festival attendees is likely to increase overall Twitter penetration in attendees home counties if the SXSW attendees have large real-life social networks in their home counties and their networks are receptive to early adoption of a new social media platform. This is likely the case in counties with a nonnegligible population "similar" to SXSW enthusiasts in terms of socio-demographics or cultural interest, and in counties with a large population of "tech savvy" individuals interested in adopting a new and relatively unknown social media platform. However, as discussed below, the festival offers a wide variety of popular culture attractions and should attract the interest of diverse groups of people. The Local Average Treatment effects should thus not be excessively particular, although one should not overgeneralize these results.

Broadly speaking, this research contributes to the literature on media's effect on real-life political outcomes (Enikolopov et al., 2011; Gentzkow et al., 2011; Gentzkow,

2006; DellaVigna & Gentzkow, 2010), and especially the role of the Internet and social media (Falck et al., 2014; Guriev et al., 2021; Zhuravskaya et al., 2020). By examining specially demonstrations concerning particularly partisan topics on which the US population holds polarized views on, I contribute to understanding on the links between the internet and social media, and political polarization (Gentzkow & Shapiro, 2011; Halberstam & Knight, 2016; Boxell et al., 2017; Barberá, 2014; Yanagizawa-Drott et al., 2020; Melnikov, 2021; Lelkes et al., 2017). Further, by examining if Twitter penetration increases the relative frequency of violence, and the relative frequency of protests associated with white supremacists, I contribute to the literature on media's, in particular social media's, role in propagating violence and hateful sentiments (Müller & Schwarz, 2021, 2020; Bursztyn et al., 2019; Yanagizawa-Drott, 2014; Adena et al., 2015; DellaVigna et al., 2014).

This paper is directly related to the growing literature on the information and communication technologies', and social media's, ability to facilitate protest activity. Amorim et al. (2018) finds robust evidence of a causal effect of broadband Internet availability on the probability of an Occupy Movement protest event, where identification comes from topographic elevation as an exogenous determinant of the provision of Internet service providers. More specifically on social media, Enikolopov, Makarin, & Petrova (2020) show that VKontakte (VK), the most popular social media application in Russia, facilitated protests during a protest wave in 2011 triggered by electoral fraud benefiting the incumbent party, United Russia. For identification, they use a similar source of geographical variation as I do, steming from early adoption of VK among students who studied in the same school at the same time as VK's founder. They show that while VK increased protest activity, it did not increase government disapproval nor support for the opposition, and that the effect grows with city size, both results theoretically consistent with social media impacting protest activity through easing coordination rather than through increasing availability of information that cause protests. Avetian et al. (2021) show that social media not only helps mobilize individuals already sympathetic to a political cause, but can play an important role in mobilizing new protesters and broadening coalitions. They show that increased county-level uptake of Twitter prior to murder of George Floyd increases the likelihood of a Black Lives Matter (BLM) protest taking place in the 3 weeks following George Floyd's murder. They find that this effect is only detectable in counties where the salience of racial inequality was low initially, where no BLM protests had occurred before, as well as in more rural, high-income and whiter counties. However, all of these papers study a single protest movement and thus cannot

explore differential effects depending on the characteristics of movements, as I do. Other studies take a more aggregate approach, Manacorda & Tesei (2020) study the effect of mobile phone coverage signal on protests from 1998 to 2012 across the African continent. They find that mobile phone connection increases protest activity, but that this only holds during economic downturns. Further, they find suggestive evidence that individuals are more likely to participate in protests when a larger share of others in society will participate, and that this effect is enhanced by mobile phones. Christensen & Garfias (2018) study the effect of cell phone coverage on protest participation between 2007 and 2014 on a global scale. Using a difference-in-difference strategy they find that cell phone coverage increases the probability of a protest event, an effect driven by democratic countries or those with relative freedom of media. They also find that, in Africa, gaining cell phone coverage decreases the likelihood of government repression during demonstration events. They attribute this effect to to government accountability mechanisms. In particular, violence or repression might be more visible where cell phone connection is accessible to the population. This increased visibility could deter violent action or repression (Durante & Zhuravskaya, 2018). Finally, the paper most closely linked to this research is that of Fergusson & Molina (2019). They use Facebook's release in a new language as an exogenous variation in access to social media where the language is spoken to estimate it's effect on protests globally. They find a sizable effect of Facebook access on protests and show that the response to social media depending on democracy takes a U-shaped form, being strongest for very low and very high levels of democracy. This holds especially for protests against the current regime, while those against the opposition react most strongly to social media in low democracy countries. Further, they find that access to Facebook decreased violent conflict and provide evidence for two mechanisms. First, they show that Facebook increases perceived political freedom of expression and decreases violent conflict more in countries with characteristics that make them more conflict-prone, consistent with protests providing a way to voice discontent that would otherwise turn more violent. Second, consistent with the "increased visibility effect," they find that Facebook decreases violent conflict less in areas with more freedom of press, where Facebook should be less important for visibility of violent actions.

My contribution to the literature is to examine directly the effects of social media on protest activity, depending on the characteristics of protests movements themselves. In a democratic setting, I utilize qualitative information on demonstration events and information on organizations or groups associated with events to cate-

gorize protests movements depending on coordination complexity, and depending on the topic of demonstration. Further, compared to Christensen & Garfias (2018) and Fergusson & Molina (2019) who examine government repression and violent conflict, I can distinguish between non-peaceful behaviour of demonstrators themselves, non-governmental outside parties, or police and military forces, during demonstration events, where I find differential effects.

1.1 Conceptual framework

Theory and evidence suggests that social media may effect the demonstration landscape in various ways. Like traditional media, social media can provide citizens with politically relevant *information* increasing their willingness to demonstrate for a given cause. However, in addition to such one-way information transmission, social media can facilitate *coordination* through horizontal interaction between users, alleviating collective-action problems (Olson, 1965). These coordination effects can stem from reduced costs of acquiring and exchanging logistical and tactical information on demonstration events (Little, 2016; Enikolopov, Makarin, & Petrova, 2020), from updating beliefs about how many others are willing do demonstrate (Edmond, 2013; Barbera et al., 2020; González, 2020; Passarelli & Tabellini, 2017) or by altering how one can project their social image via protest participation (Enikolopov, Makarin, Petrova, & Polishchuk, 2020; Cantoni et al., 2019).

The importance of social media for logistical and tactical coordination, and for strategic coordination based on beliefs about others' participation, likely varies depending on the complexity of coordination. First of all, one might expect the coordination effect of social media to increase with the scale of the coordination problem. Social media might be less important for inferring others' willingness to participate in protests concerning primarily local matter, while it might be crucial for belief formation about willingness to participate in broader protests movements concerning national matters. This holds especially if costs and benefits of participation do not only depend on the size of an individual protest event but on the size of the movement as a whole, comprising a set of events occurring within some geographic boundaries and time-frame. Second, the coordination effect might depend on the organizational structure of the protest movements. Protest events backed by political parties, unions or any other organizations may deploy organizational capacities that facilitate coordination without social media, while protest movements that spur from the bottom-up without clear organizational leadership or hierarchy might depend more on social

media for coordination. While Fergusson & Molina (2019) rationalize their results along these lines—they find that the effect of Facebook on protests fades around electoral periods and hypothesise that political parties or other organizations might deploy organizational capacities to facilitate coordination during these periods—they do not provide direct evidence of these heterogeneities. I directly compare the effect of Twitter penetration on demonstrations that are either part of loosely organized, broad social movements or where no association to an organization can be identified on one hand, and on demonstrations that are not part of broad nationwide movements and are associated with an organization on the other. These heterogeneities are confirmed, I indeed find that Twitter penetration disproportionately fuels demonstrations with lower institutional support.

Social media may not only have differential effect depending on organizational structures. The effect could also vary depending on the topic for protesting. Focusing on beliefs about others' willingness to protest, it is often theorized that individuals' protest participation are strategic compliments, that the cost of participation is lower when the protest is larger (due to, for example, the likelihood of being arrested being lower in a larger crowd) or that the likelihood of success increases with size (Edmond, 2013; Passarelli & Tabellini, 2017; Barbera et al., 2020).[1] Although the theoretical insights are explored in a setting of autocratic regimes and revolts, they could be extended to a democratic setting where the "success of protest" refers to, e.g., the extent to which demonstrators can influence policy. In a setting of strategically complimentary participation, Barbera et al. (2020) explore the role of homophily, i.e., when one is surrounded by individuals with similar political interests and preferences. With homophily, the learning value of seeking information about others' willingness to protest is reduced. Homophily thus makes it harder to hold protests in cases where learning was necessary to enable mobilization, but makes it easier in cases where learning would demotivate protest movements. To the extent that social media induces homophily in interactions, one may hypothesize that individuals holding strong views on polarized topics might be less likely to update beliefs according to

[1]It is worth noting that the assumption of strategic complimentary is not always verified in the data. Cantoni et al. (2019) find experimental evidence of strategic substitutability. In particular, find that positively updated beliefs of others' protest participation decreases the likelihood of protest participation in the context of Hong Kong's ongoing anti-authoritarian movement, and the opposite effect for negatively updated beliefs. They provide suggestive evidence of 3 sources of strategic substitutability: Public Good charachteristics of protest participation, greater perceived likelihood of government crackdown in larger protests, and that subjects perceive participating in smaller protests as a stronger signal of their ideological "type." However, empirical evidence for strategic complements in protest participation has been found in other settings, e.g., in Manacorda & Tesei (2020).

the popularity of opposing views, thus facilitating mobilization concerning polarized topics.

Despite the widely held concern that social media creates echo chambers, preventing people to learn about political views opposing their own (Sunstein, 2017), a consensus has not been reached on its role in fuelling homophily in interactions and political polarization. Gentzkow & Shapiro (2011) find that interactions online are less ideologically segregated than off-line interactions, while Halberstam & Knight (2016) find that homophily in interactions on Twitter is closer to that of off-line interaction, than to the segregation in the consumption of online political news. Boxell et al. (2017) find that recent growth in political polarization in the United States is driven by demographic groups that are the least likely to use the Internet and social media and Barberá (2014) finds that Twitter users in Germany, Spain and the United States have ideologically diverse networks, arguing that social media strengthens exposure to their "weak ties" instead of creating echo chambers. On the opposite side, Yanagizawa-Drott et al. (2020) find that areas in the United States with greater political homophily of Facebook connections also have more homogeneous voting patterns, indicating that online homophily could drive political polarization. Deactivation of Facebook accounts has also been found to reduce political polarization of account owners (Allcott et al., 2020). Further, Barberá & Rivero (2015) show that Twitter users who post about politics are predominately male and tend to have more extreme ideological preferences than the general population. Melnikov (2021) also finds that, while access to 3G internet seems to have contributed to increased political polarization in the United States, the increased polarization largely did not occur *among* social media, but *between* experienced internet and social media users (becoming more pro-Democratic), and inexperienced users (becoming more pro-Republican). Finally, Lelkes et al. (2017) find that broadband internet availability increases segregation in the consumption of partisan media, and partisan hostility, which both could be strong forces for increased political polarization. In sum, the literature does not clearly demonstrate that social media induces homophily in interactions, thus potentially affecting the political landscape through the channels explored by Barbera et al. (2020). However, available evidence does not exclude the possibility either.

Another plausible reason for social media to facilitate political protests movements concerning polarized topics is related to emotions. Passarelli & Tabellini (2017) present a model where, in addition to beliefs about how many others are willing to act, the strength of a group's emotional response to a perceived unfair policy or injustice increases the likelihood of mobilization. If social media increases emotional

responses to some particular topics, it might disproportionately affect mobilization concerning them. Partisan hostility (Lelkes et al., 2017) is likely associated with increased emotional response to highly partisan/dividing topics. Further, Vosoughi et al. (2018) show that false news on Twitter, especially political news, spread both faster and more broadly, and are more emotionally charged than news stories that are true. The emotional nature of false news, coupled with recent evidence suggesting that political false news are predominately shared on Twitter by hostile partisans who selectively share content (Osmundsen et al., 2021), could increase Twitter's ability to fuel strong emotional responses to partisan topics and protests related to them. However, Grinberg et al. (2019) find that although false news represent a large share of news consumption on Twitter, their circulation was highly concentrated. Therefore, they conclude that most users likely get their political news from mainstream news outlets, while a small share of users' news feeds are dominated by fake-news.

In light of this, I differentiate between protests associated with partisan issues or otherwise deeply dividing topics in United States on one hand, and other protests on the other. I define the following demonstrations as demonstrations associated partisan issues or otherwise deeply dividing topics: demonstrations for or against public health measures related to the Coronavirus pandemic; demonstrations which call for action against alleged Democratic voter fraud and "illegal" ballot counting after the 2020 Presidential elections, and those calling for votes to be counted and election results to be respected; demonstration for or against abortion rights; and demonstrations concerning racial issues (including the Black Lives Matter Movement, protesting police brutality against black people, and counter-movements such as "Blue Lives Matter"). However, I do not find strong evidence supporting the hypothesis that social media disproportionately fuels protests concerning such polarizing topics.

Further, social media can act as a propagating mechanism for hateful sentiments. High prevalence of exposure to hate speech online has been documented (Oksanen et al., 2014), and there is evidence that social media can facilitate hate crime (Müller & Schwarz, 2021). This effect may be driven by persuasion (Bursztyn et al., 2019) or, and especially relevant in the United States, by spreading posts with hateful messages from influential opinion makers to the general public (Müller & Schwarz, 2020), thus influencing norms about what is socially acceptable (Bursztyn et al., 2017). As social image concerns likely influence protest participation decisions (Enikolopov, Makarin, Petrova, & Polishchuk, 2020; Cantoni et al., 2019), I test whether social media increases demonstration associated with known US white supremacist groups such as the Proud Boys, the Ku Klux Klan or Super Happy Fun America. While I do find a

positive effect for these demonstrations, Twitter does not seem to disproportionately increase these demonstrations relative to others.

Finally it is possible that social media might effect the level of repression or violence during demonstration events. On one hand, consistent with prior research (Fergusson & Molina, 2019; Christensen & Garfias, 2018; Durante & Zhuravskaya, 2018), social media can enhance the visibility of violent or repressive acts, deterring violent actions. This "increased visibility" effect might be particularly relevant for government actions during demonstrations, as opposed to non-government actors, as they are more accountable for their actions to informed voters (Snyder Jr & Strömberg, 2010). On the other hand, social media might increase violence or inference with demonstrations by non-government actors. First, social media could disseminate logistical information about demonstration events not only among supporters, but also to those opposing the demonstration cause, and allow "the opposite side" to coordinate. Further, social media's potential effects on partisan hostility and emotional reaction to political information discussed above might compel those opposing a demonstration to interfere. I test these ideas by distinguishing between: a) "Two sided demonstrations," i.e., events where two protests occur at the same time, in the same place, which are essentially "counter" one another[2] and "one sided demonstrations," b) Protests met with violence or an attempt to suppress the protest by government-entities, by non-government actors, or protests where demonstrators are allowed to carry on without interference. I find that Twitter penetration does not induce a relative decrease in government intervened protest nor does it increase "two sided" demonstrations disproportionately. I find some evidence of a Twitter induced increase in the share of protests that are met with interference or violence by non-government actors. However, the magnitude of this effect is very small.

2 Data and Background

2.1 Protests in the United States

Contemporary United States provide a good setting to test heterogeneities in the effect of social media on different types of protest movements. The United States has been a vibrant protest environment throughout the years, and recently demonstration activity has been surging. As an example, ACLED (2020c) recorded more

[2]E.g., A pro-mask mandate demonstration and an anti-mask mandate demonstration occurring simultaneously

demonstrations in the United States than any other country, except for India, covered in their almost-global data set during the initial 3 months of data collection in the United States. US citizens do not only protest a lot, they also engage in protests around a diverse set of issues. In 2020, three issues were a particularly common cause for protests: a) Demonstrations associated with the Black Lives Matter (BLM) movement, protesting police brutality against Black people in the United States, following the police murder of George Floyd. b) Demonstrations related to the Presidential elections in 2020. These include rallies in support of or against one of the presidential candidates, as well as protests against the election results, alleged Democratic voter fraud and "illegal" ballot counting, or protests demanding the results to be respected. c) Demonstration linked to the Coronavirus pandemic. These protests mainly involved demands that can be grouped into two categories. First, demands for stricter or more lenient public health policy to combat the spread of the virus, such as vaccine mandates, mask-mandates, or restrictions of business or institutional activity of sectors and services deemed non-essential. Second, demands of financial or economic government support for specific businesses, workers, or households due to the economic stress caused by the public health measures or the pandemic in general (ACLED, 2020a). In addition to these three highly visible and oftentimes dividing topics, the ACLED data records numerous other causes for protests in the United States, ranging from demands concerning laborer's rights or compensation, local or federal environmental policy, human or animal rights topics, or immigration issues. Further, there are increasing signs of political instability and agression during demonstration events. Comparing 2020 to 2019, the start of their data collection in the United States, ACLED reports that counter-protests became more frequent, government authorities increasingly engaged with and exhibited force against protesters, and armed non-state actors were increasingly visible in demonstration events, either participating in protests, engaging in "peacekeeping," or intimidating demonstrators (ACLED, 2020b).

2.1.1 Data and main variables

I use the Armed Conflict Location and Event Data (ACLED), which is introduced in the descriptive analysis in the section above, for protest variables (Raleigh et al., 2010). ACLED is an event-level dataset that documents the location, date, and characteristics of each event. ACLED researchers manually compile events from regional, national, and international media outlets deemed reliable and are supplemented by reports by international institutions or non-governmental organizations (NGOs). Un-

fortunately, no information on the number of individual participants is available in the data set. I use data from January 2020 to 12th of November 2021, the latest available data at the time of analysis.

ACLED documents various political violent events and conflict, defined as use of force by a group with a political purpose or motivation. It also documents events that are "potential pre-cursors or critical junctures of a violent conflict," such as protests and riots, these events will be the focus of my study. Demonstration events are defined as all physical congregations of three or more people when they are directed against a political entity, government institution, policy, group or individual, tradition or event, businesses, or other private institutions. As a baseline measure of county-level demonstration activity, I simply count all demonstration events in each county occurring between January 2020 and November 2021. Figure 1 shows the spacial distribution of demonstration events per capita across continental United States during the period. In total, 43% of counties had no demonstration events over the period of analysis, Figure A1 in the Appendix shows the distribution of demonstrations among counties where at least 1 demonstration occurred.

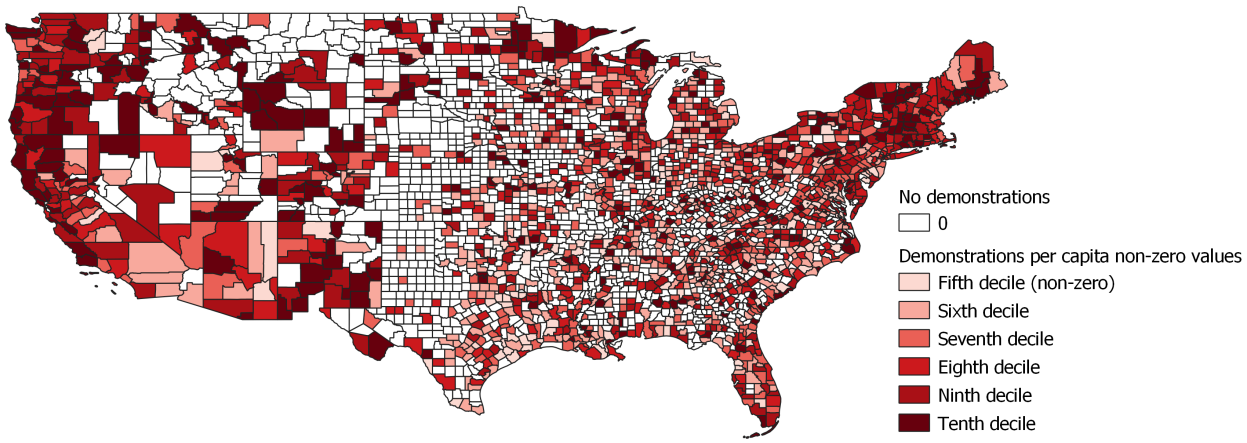


Figure 1: Spacial distribution of demonstration per capita

There are 3 variable-types coded for each event which I use to categorize protests. First, each event is described by "actor types" and "associated actors." Actor types for demonstration events in the United States are either generic terms for protesters or rioters (defined below), civilians, police or military forces, or armed non-government militia groups. Further, for each event the associated groups, organizations, or (where relevant) the salient racial/ethnic identity of demonstrators are documented as associated actors. There can be many such associated actors associated with one event.

There are 574 unique associated actors coded in the data, where each event is associated to 0-18 of these actors. A specificity in the US data is that associated actors for demonstration events can also represent "Broad Social Movements", these actors are not meant to suggest that the event is directly affiliated with an organization, but that demonstrators are demanding something as part of a broader movement.[3] I use this information to proxy the complexity of coordination for a protest. First, for 8,369 out of 34,189 demonstrations in the data, no associated actor can be identified. Theses demonstrations are likely to be without organizational leadership or capabilities to facilitate coordination; where no key organization leads the event and spreads logistical and tactical information. Therefore, a priori I assume that these events are more difficult to coordinate than those associated with specific organizations. Second, broad social movements might not only be difficult to logistically coordinate due to lack of organizational capabilities (indeed, some events associated with broad social movements are also associated with specific organizations or activist groups), but the scale of the movement implies that estimating others' willingness to participate in the movement is a challenge—a challenge social media could help solve. I therefore count all events associated with a broad social movement in each county, and those not associated with a broad social movement and define "broad social movement demonstrations" and "non-broad social movement demonstrations" as outcome variables. Further, I count all demonstrations where no associated actor can be identified and define the outcome variable "broad social movement or not organized demonstrations" as the number of events either associated with a broad social movement or with no associations, and the mirror image: Demonstrations not associated with a broad social movement but otherwise with associated actors, "Not broad social movement and organized demonstrations."

Each event is further accompanied by a "note" with qualitative information about the event.[4] I use this information, combined with information on associated actors, to classify demonstration events associated with specific topics. In particular, I aim to code protests associated with deeply dividing topics in the United States during the period. I count events which include a word or a phrase from a set of "buzz-words" associated with a given topic. First, I code demonstrations concerning abortion rights

[3]Broad Social Movements identified are: Black Lives Matter (BLM), Cancel the Rents Movement, Abolish ICE, Blue Lives Matter, Stop Asian Hate, Back the Blue, Occupy Movement, Native Lives Matter (NLM), Save Our Children, Fridays For Future.

[4]An example of a note accompanying an event is the following: "On 12 November 2021, a group of protesters gathered outside of Ascension Via Christi hospital in Wichita (Kansas) to protest against coronavirus vaccine mandates for health workers. Unvaccinated employees joined the protest after finishing their final shift before the mandate took effect that day."

with buzz-words such as "abortion" or "pro-life." Second, ACLED data includes a "tag" in the note for all demonstrations associated with the 2020 post-presidential election unrest with either "stop the steal" or "count every vote," i.e., demonstrations for or against accepting valid mail-in/absentee voting ballots and respecting the election results. I code all demonstrations including this tag. Third, I code demonstration for or against public health measures related to the corona virus pandemic with buzz-words such as "vaccine," "mask mandate," or "coronavirus prevention measures." The goal is to disentangle demonstrations strictly related to public health measures and those related to economic relief of specific groups *due to* the pandemic. General public health measures were a deeply dividing and politicized topic in the United States, while demands of economic relief, better working conditions, higher wages or banning of evictions for specific groups due to the pandemic might be seen as simply advocating for specific interest groups and not as polarizing a topic. Fourth, I code topics related to racial issues, but do this using associated actors. I manually coded the 574 unique associated actors and identified organizations or movements based around racial issues by examining the notes to their protest events. These organizations or movements include for example "Black Lives Matter," "Revolutionary Black Panther Party," "Stop Asian Hate," "Aryan Nations," "National Association for the Advancement of Colored People." I also specifically coded demonstrations where an associated actor has known affiliations with white nationalism or white supremacists ideology, groups such as "Aryan Nations" or "Proud Boys." Finally, I group all events that fit into any of the above defined topics and define those events as demonstrations concerning "polarizing topics." A full list of buzz-words, phrases, and associated actors used to code each of the topics is provided in the Appendix. The "note" further includes information on whether the demonstration event was "two-sided." If two demonstrations occur at the same place at the same time, and they are essentially counter to one another, ACLED codes these demonstrations as one event, and codes a tag indicating that the event included two demonstrations counter to one another in the note accompanying the event. I also code events including this tag.

Finally, events are described by the *event type* and *sub-event type*. Demonstration event types are Protests or Riots. Protests are events where demonstrators do not engage in violence or disruptive acts such as property destruction, although violence may be used against them. Riots are events where demonstrators engage in disruptive acts such as property destruction or violence without the *use* of lethal weapons. Protest are further categorized into three sub-types: *Peaceful protests* are events where demonstrators are not engaging in violence or any other form of rioting

behaviour, and are not faced with any sort of force or engagement. *Protests with intervention* are events where demonstrators are peaceful, but there is an attempt to disperse or suppress the protest without serious/lethal injuries being reported. *Protests with excessive force against protesters* are events where peaceful protesters are targeted with violence leading to (or if it could lead to) serious/lethal injuries. Riots on the other hand are categorized into two sub-types: *Violent demonstrations* are events where demonstrators engage in disruptive or violent behaviour, examples of disruptive behaviour would include vandalism, road-blocking using barricades, or burning tires. *Mob violence* are events when rioters violently interact with other rioters, another armed groups or civilians, outside of demonstrations and without the *use* of lethal weapons. I examine specifically the number of each sub-events as an outcome variable, and further group sub-events: I look at peaceful protests versus all other sub-types, i.e., "non-peaceful demonstrations." Next, I split non-peaceful demonstrations into i) "Peaceful protests met with intervention or violence" (i.e., Protests with interventions or protests with excessive force), and ii) "Violent protesters" (i.e., Violent demonstrations or mob violence). Finally, for demonstrations met with intervention or violence by an outside actor, I count events where the "intervening actor's" type is police or military forces, versus events where the intervening actor is a non-government group.

2.2 Twitter and the South by Southwest festival

In 2021, surveys indicate that a large majority of Americans self-report ever using any social media site, and that their share has remained relatively stable over half a decade. This data suggests that that the most popular platforms are YouTube, Facebook and Instagram, used by 81%, 69% and 40% of those who respond to survey questions. 23% of U.S. adults use Twitter, a similar number of users as for Snapchat (25%) and WhatsApp (23%). Like with most other social media platforms, Twitter is most popular among young, educated and urban adults (Pew Research Center, 2021).

Although Twitter was founded in March 2006, it remained largely unknown before the South by Southwest festival (SXSW hereafter) in March 2007. The importance of the festival for early adoption has been explicitly stated by Evan Williams, a co-founder of the platform. At the festival, Twitter had screens in hallways and created an event-specific feature where one could easily create an account and broadcast posts on these screens. Those who created accounts through this feature were automatically made to follow "Twitter ambassadors," festival attenders already on the platform

(Quora, 2011). Fujiwara et al. (2021) and Müller & Schwarz (2020) document quite thoroughly the immediate growth in Twitter penetration after the festival in 2007, and show that this growth occurs primarily in the counties of festival goers. However, of course the platform did not reach its later popularity immediately. It went from an average of 5,000 tweets per day in 2007, to 300,000 in 2008, numbers dwarfing in comparison to the 500 million tweets sent each day on average in 2019.

The SXSW festival itself is an ambitious project combining film, interactive media and music festivals with professional conferences with a focus on creative industries and technology. These bundles of events take place annually in mid-March in Austin, Texas. The festival advertises itself as an essential destination for global professionals where, in addition to film screenings, music and comedy shows, there are ample opportunities for professional development and networking (see the official homepage of the festival: https://www.sxsw.com/about/). The festival is wildly popular, in 2016 it attracted around 300,000 official attendees (Theis, 2016). Already by 2007, the festival attracted speakers, performers and attendees from all around the United States and beyond. Broadly speaking, the festival's content can be described as combining a wide range of western pop-culture entertainment with professional themes of technological innovation.[5]

2.2.1 Twitter data

As a proxy for **Twitter penetration** at the county level, I utilize a sample of 475 million geo-coded tweets collected by Kinder-Kurlanda et al. (2017). These tweets are all from within the United States, and were collected from 1st of June to the end of November, in 2014 and 2015 through the Twitter Streaming API, which returns a 1% sample of daily tweets as long as it is called. Only tweets that can be geo-located are included in the sample. Twitter allows users to "geo-tag" their tweets, adding the latitude/longitude information from one's device's GPS sensor to the post. This information therefore refers to the location from which each tweet was posted. The tweets in the dataset are already assigned to counties. I simply approximate Twitter penetration by the number of Tweets in each county.

[5]For example, in 2007 the festival included performances from musicians from all around the United States and international acts such as Peter, Bjorn & John from Sweeden, Amy Winehouse from the UK, and YB from South Korea. The festival premiered Judd Apatow's wildly popular comedy film "Knocked Up" to the world, and held panel sessions on topics such as "Virtual Worlds and Virtual Humans" with more than 450 speakers from the media and information technology sector. Examining the program for 2006 is suggestive of similar scope, size and themes, and can be found here: https://www.sxsw.com/about/history/.

There are two issues worth noting concerning this measure. First, as geo-tagging tweets is optional, the observed sample is highly self-selected. In fact, only about 1% of users chose to geo-locate their tweets. Further selection issues arise as during the data-collection period, Twitter introduced a new way for users to share their location by tagging tweets with the name of their location from a dropdown list, as opposed to the latitude/longitude pair. The data collection did not catch tweets sharing this location information, but this new option substantially reduced the popularity of "geo-tagging" tweets (Kinder-Kurlanda et al., 2017). However, Fujiwara et al. (2021) show that measures using these tweets closely resemble other measures of Twitter penetration, such as estimates of the number of Twitter users over time from Statista or survey estimates of county-level number of Twitter users, alleviating some concerns about self-selection biases. A second concern is that of the relevance of this measure. The theoretical channels through which Twitter should influence protest participation are all likely to rely more heavily on Twitter *penetration* than Twitter *activity*, that is, how many people tweets can reach, not how many are posted. Malik et al. (2015) show that in this dataset, the distribution of tweets per user is very long-tailed to the right (few users who tweet exceptionally often), something they claim is a common feature of data from Twitter. Although this does not raise confidence in the ability of the measure I use to capture Twitter penetration, it is worth noting that the relevant variation used to elicit causal relationship stems from county-level attendance at the South by Southwest festival. This instrumental variable has been shown to be a good predictor of county-level Twitter *users* (Müller & Schwarz, 2020; Fujiwara et al., 2021).

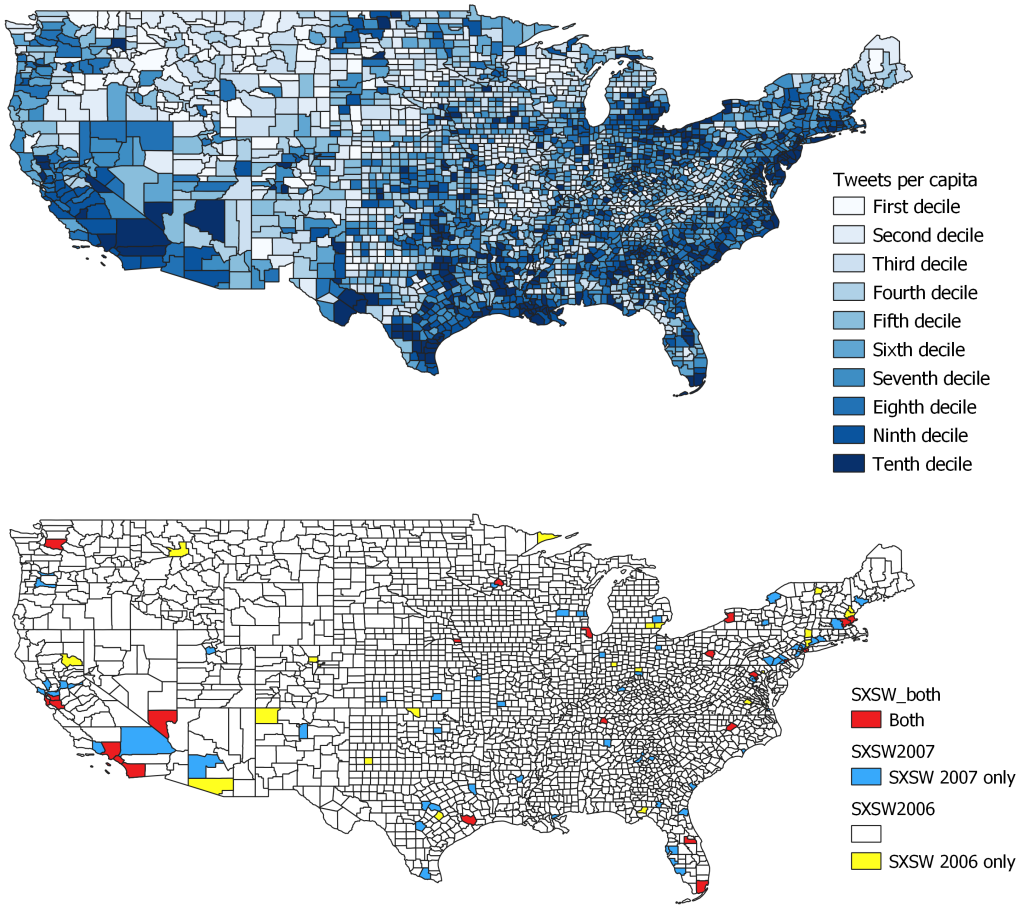


Figure 2: Spacial distribution of Tweets per capita and SXSW followers

2.2.2 South by Southwest festival attendance proxy

To measure county-level attendance at the 2007 SXSW festival I collected data through the Twitter API. I scraped data on each account following the official Twitter account of the SXSW festival (@SXSW) at the time of data collection (February 2022). I then manually geo-coded each follower based on the location users report in their user profile. Of 639,000 followers, 460,000 report a location. The two crucial variables I construct with this data are: a) The county-level number of SXSW followers who created their Twitter account in March 2007, 2007 SXSW followers hereafter, and b) the number of SXSW followers that created their Twitter account in 2006 (the founding year of Twitter), 2006 SXSW followers hereafter. The number of SXSW followers who created accounts in March 2007 are taken as a proxy for county-level attendance at the festival, my excluded instrument, while those creating accounts in 2006 are used to control for general interest in the festival. In March

2007, 1,234 SXSW followers created their accounts, of which I am able to locate 292 within the United States, into a total of 87 counties. Of the 330 SXSW followers who created their accounts in 2006, I am able to locate 121 within the United States, into a total of 52 counties. Figure 2 shows the spacial distribution of Tweets per capita and of the SXSW followers across continental United States, while Figures A3 and A5 in the Appendix show the frequency distribution of Tweets and SXSW followers.

2.3 County characteristics

Additionally, I collect county-level demographic, socio-economic, and voting result variables as controls. I use variables from 2009 and 2010, which can be considered a "pre-treatment" period as Twitter did not gain immense popularity until later.[6] I use data on population size, age, race, and ethnic composition from the 2010 US Census and estimates of poverty rates and median household income from the 2009 US Census Bureau's SAIPE program. I use estimates of educational attainment (in particular, share of adult population with at least a high school diploma and with at least a Bachelor degree) in 2010 from the American Community Survey, and Unemployment and Employment-to-Population rates in 2009 from the Bureau of Labor Statistics. Finally, I use data on county-level presidential election results from the MIT Election Lab. In particular, I use the share of votes toward the Republican candidate in 2000 (George W. Bush), 2004 (George W. Bush), and 2008 (John McCain). Finally, as Fujiwara et al. (2021) find that Twitter penetration may have persuaded voters with moderate views to vote against Trump in 2016 and 2020, one could worry that effects found in this paper might be driven by changes in voters' attitudes and thus not representing any additional effect to the one found in Fujiwara et al. (2021). Therefore, I replicate all results presented in the main text while additionally controlling for the votes towards the Republican candidate in 2012 (Mitt Romney), 2016 (Donald J. Trump) and 2020 (Donald J. Trump). Results remain almost completely unchanged by the addition of these controls, shown in Tables A15 to A20 in the Appendix.

3 Empirical Strategy

If Twitter were an important facilitator of protests, we would expect to see greater demonstration a activity in areas where Twitter penetration is higher. Our relation-

[6] In 2010, worldwide Twitter users were around 30-50 million, about 10-15% of the 300 million users estimated in 2015. Since 2015, the growth of Twitter users has remained relatively modest, with between 300 and 350 million users until 2019 (Statista, 2019)

ship of interest is:

$$Y_i = \beta_0 + \beta_1 Twitter\ penetration_i + \boldsymbol{X_i'\beta} + \epsilon_i \tag{1}$$

Where Y_i represents some demonstration variable in county i, $Twitter\ penetration_i$ represents a measure of Twitter penetration, in our case the natural logarithm number of tweets in the Kinder-Kurlanda et al. (2017) data set, in county i, $\boldsymbol{X_i'}$ is a vector of covariates, β's are coefficients to be estimated and ϵ_i is the error term. When looking at overall demonstration activity, we would expect β_1 to be positive, representing the positive effect of Twitter penetration on demonstration activity. However, Twitter penetration may well be endogenous to demonstration activity for various reasons, giving alternative interpretations to β_1. First, unobserved county characteristics might drive both Twitter adoption and demonstration activity. As an example, as Twitter allows users to follow politicians and express their political opinions, differences in inherent interest in political activism might drive county-level differences in Twitter adoption—leading to a spurious correlation and overestimation of the impact of Twitter penetration on demonstration activity. Second, as social medias such as Twitter facilitate protest event coordination, one might expect counties high regional protest activity to drive individuals to join Twitter, to ease coordination for demonstrations they would have attended even without the existence of Twitter.[7]

To circumvent these concerns, I exploit the plausibly exogenous shock to early Twitter adoption in the United States connected to the 2007 SXSW festival, inspired by Fujiwara et al. (2021) and Müller & Schwarz (2020). As discussed above, the 2007 conferences provided a tipping point in the rise of Twitter's popularity. Moreover, consistent with the literature on path dependence in technology adoption (Arrow, 2000; Liebowitz & Margolis, 1995), following Fujiwara et al. (2021) and Müller & Schwarz (2020), I find that the festival had persistent effects on the spacial distribution of Twitter penetration. This setting lends itself well to an instrumental variable strategy where county-level Twitter penetration is instrumented for by the county-level attendance at SXSW in 2007 and subsequent adoption, measured by SXSW Twitter account followers who created their account around the time of the festival. However, it is not immediately obvious that county-level festival attendance and subsequent early adoption of Twitter is exogenous to protest participation. One crucial concern is that counties with festival goers or general interest in the festival might be

[7]Note that this reverse causality problem can only occur in my data if demonstration activity pre-2014 drives Twitter adoption, and pre-2014 demonstration activity is predictive of demonstrations in 2020 and 2021.

systematically different from other counties, in ways that correlate with demonstration activity. To alleviate these concerns, I utilize information on the exact timing of Twitter adoption of SXSW followers. Consistent with prior evidence, I find that SXSW followers that created their accounts in March 2007, presumably after attending the festival, are robustly and persistently predictive of the spacial distribution of Twitter penetration, while the county-level number of SXSW followers who created their accounts in 2006 are not predictive of Twitter penetration. This allows me to disentangle the effect of general interest in the SXSW festival and early adoption of Twitter from the effect of current Twitter penetration. By including the number of SXSW followers who created their account prior to the festival (in 2006), we effectively control for general interest in SXSW and early adoption of Twitter. Moreover, this provides an placebo test by testing whether inherent interest in the festival has an effect on protest activity. It turns out that throughout my analysis, SXSW followers who created their accounts in 2006 rarely correlate with demonstration related outcomes—in the few cases where they correlate to demonstration outcomes, the correlation is of the opposite sign compared to 2007 followers, i.e., negative. In addition, I show that conditional on followers who created their accounts in 2006 and population, the excluded instrument is essentially uncorrelated with a host of county characteristics.

Formally, the estimations can be represented as follows:

$$\begin{aligned} Twitter\ penetration_i &= \alpha_0 + \alpha_1 SXSW^{March2007} + \alpha_2 SXSW^{2006} + \boldsymbol{X_i'\alpha} + u_i \quad (2) \\ Y_i &= \gamma_0 + \gamma_1 SXSW^{March2007} + \gamma_2 SXSW^{2006} + \boldsymbol{X_i'\gamma} + \mu_i \quad (3) \end{aligned}$$

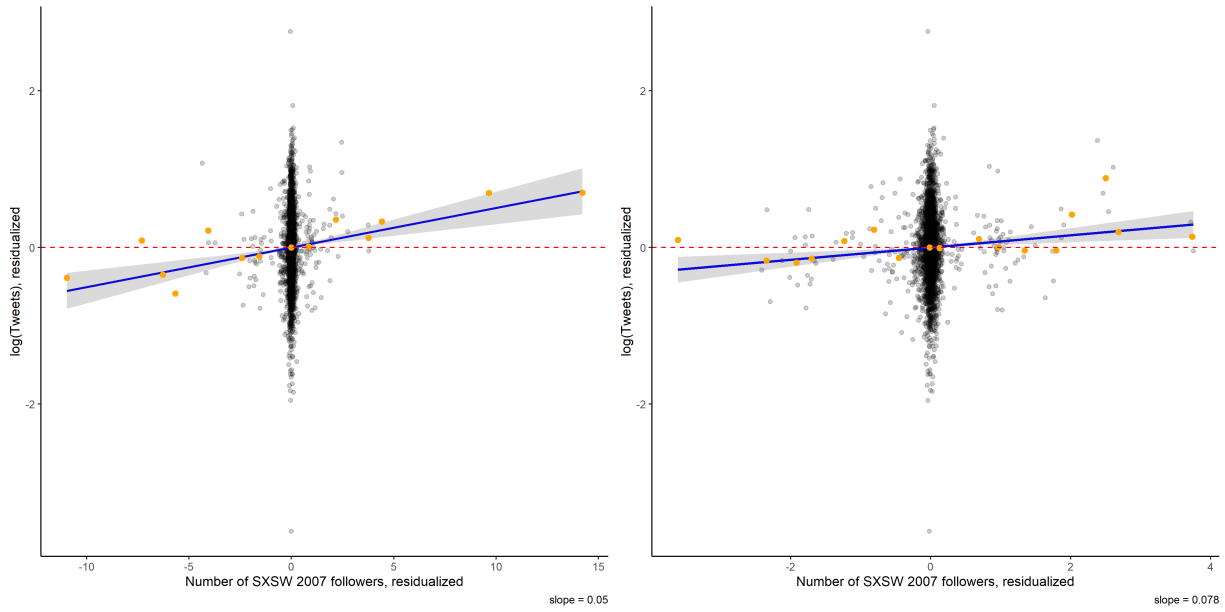
Where equation (2) represents the First stage and equation 2 the reduced-form estimation. $SXSW^{March2007}$ represents the number of 2007 SXSW followers and $SXSW^{2006}$ represents the number of 2006 SXSW followers in county i. α's and γ's are coefficients to be estimated, and u_i and μ_i are error terms. Other notation remains the same as for eq (1). It is worth noting that population size proves to be an crucial control variable as Twitter penetration, demonstration outcomes and SXSW follower are highly correlated with population, therefore I control quite flexibly for population using a set of 50 2-percentile dummies, included in X_i'. The accompanying second stage equation can be written as follows:

$$Y_i = \lambda_0 + \lambda_1 Twitter\ \hat{penetration}_i + \lambda_2 SXSW^{2006} + \boldsymbol{X_i'\lambda} + \phi_i \quad (4)$$

Where $Twitter\ \hat{penetration}_i$ refers to the predicted value of the county-level measure

of Twitter penetration from equation (2), λ's are coefficients to be estimated, ϕ is the error term, and other notation is the same as before. Here, λ_1 represents the Local Average Treatment Effect (LATE), the effect of increased Twitter penetration for counties where Twitter adoption would be effected by being the home of an SXSW festival goer in 2007. Intuitively, early adoption of Twitter by festival goers likely impacts the overall long-run Twitter adoption through the extended networks of friends of the festival attendants back home. Therefore, we can expect that the "complier" counties are those where SXSW attendees have large local networks of friends. This is likely to be the case in counties where there are many individuals "similar" to SXSW attendees—where festival goers and newly minted Twitter users show this new platform to tech-savvy individuals from their networks who are receptive to becoming users themselves. Although the LATE interpretation of the estimated effects warrants caution in terms of overgeneralizing the results, and that the festival does specifically target "global professionals from creative industries and technology," the cultural offerings at the festival are fairly general pop-culture events likely appealing to a wide variety of US citizens.

Tables A4 and A5 in the Appendix show correlations between the excluded instrument SXSW 2007 followers and the observed county characteristics in 2009 and 2010, conditional on SXSW 2006 followers, or on SXSW 2006 followers and population size (controlling flexibly for population via 2-percentile dummies). Even before controlling for population, the instrument rarely correlates with covariates. When controlling for population, SXSW 2007 followers are essentially uncorrelated with a host of observed county characteristics. Next, Figure 3 visualizes the first stage of the main specification. The figure presents a binned scatter plot of the relationship between the natural logarithm of Tweets and the number of SXSW 2007 followers, rediduralized by partialling out the full set of controls. Grey dots represent individual observations, yellow dots represent bins, the blue line represents the regression line and the shaded area shows 95% confidence intervals, calculated using heteroscedacity robust standard errors clustered at the state-level, the red dotted line simply plots a zero-line to facilitate reading of the figure. The left panel plots all observations used in my main specifications. We see a statistically significant relationship between SXSW 2007 followers and Twitter penetration, one additional SXSW 2007 follower increases Twitter penetration by 5%. The left panel includes outliers, the vast majority of observations have an SXSW 2007 follower residual between -4 and 4, while a handful of observations have residuals outside of this range. Therefore, to test if the relationship is driven by outliers I plot in the right panel the same relationship in



(a) Full sample (b) Subsample excluding outliers

Figure 3: First stage visualization

Note: This figure presents binned scatter plots of the relationship between county-level Twitter users in 2014-2015 and the number of SXSW followers who joined Twitter in March 2007. Variables are residualized by partialling out SXSW followers who joined in 2006, population 2-percentiles, demographic, socio-economic, and pre-2010 election controls (see description of control variables in Table A1 in the Appendix). The left hand side figure shows the full sample, while the right hand side figure shows a subsample excluding outier counties with abnormally high or low values for the SXSW followers who joined Twitter in March 2007 residual. Grey dot's represent individual observations and yellow dots represent average values of both variables within a bin. The blue line represents a line of best fit using the unbinned data, and the shaded blue area are 95% confidence intervals calculated using robust standard errors clustered at the state-level.

a sample of observations with residuals between -4 and 4. We see that the relationship is just as strong when outliers are dropped, the estimated slope is even slightly steeper. Further, in Figure A7 in the Appendix I restrict the sample even further and show that the relationship still holds among only observations with SXSW 2007 residuals between -2 and 2. Finally, Figure A9 in the Appendix shows a similar graphs as Figure 3, but for the relationship between SXSW 2006 followers and the natural logarithm of Tweets. The figure shows clearly that no relationship exists between these variables.

4 Baseline results

4.1 First stage

I start by examining the relationship between the county-level number of SXSW 2007 followers and Twitter penetration, measured as the natural logarithm of tweets from 2014-2015. Table 1 shows regression results for the first stage. It shows that, as discussed above, the SXSW 2007 followers correlate quite robustly to the logarithm of tweets, while the SXSW 2006 followers do not. The relationship between the instrument and tweets becomes considerably more precisely estimated once main controls are included. In the more precise estimations, I estimate that an additional SXSW 2007 follower increases Twitter penetration by about 5%. To gauge at how likely it is that the relationship is driven by selection on unobservables, I apply the approach from Oster (2019) of using coefficient stability and changes in explained variation between specifications to see how strong selection into SXSW 2007 based on unobservables would have to be to "explain away the relationship." I compare the specification with the fewest controls (column 1) to the one with all controls (column 5) and, assuming unobserved factors could explain all variation in demonstration frequency, I get an "Oster-δ" of 4.42. This suggest, that for the true effect of SXSW 2007 followers on the logarithm of tweets to be zero, unobservable variables would have to be about four times as important as the controls added between columns (1) and (5) in terms of selection into "treatment".[8] Given the inclusion of numerous controls, the scale of selection on unobservables needed to explain away the relationship is implausibly high, which is reassuring for causal interpretation in what follows.

Examining the robust F-statistic, we see that it is well below any traditional rule-of-thumb threshold in column (1) and (2). However, for the specifications including more county characteristic controls, it takes values around 18-19. We note that because the model is just-identified (i.e., does not include more excluded instruments than instrumented variables of interest), the robust F-statistic (sometimes also called the Kleibergen Paap) corresponds to the effective F-statistic developed by Olea & Pflueger (2013). Although the F-statistics are below the threshold for a potential bias of 10% and a significance level of 5%, which is 23, they are only slightly so. Further, my F-statistics are well above the conventional rule-of-thumb threshold of

[8]Note that assuming unobservables can fully explain variation in Twitter penetration is very conservative. This might not be the case in the presence of outcome measurement error. Assuming for example, that 5% of variation in Twitter penetration is left unexplained by unobservables and doing the same comparison gives an "Oster-δ" of 34.

10, and Andrews et al. (2019) show that for higher-than-10 F-statistics the weak-instrument problem is unlikely to affect the validity of conventional t-statistics in the case of clustered standard errors. This alleviates concerns that the results presented while controlling properly for county characteristics severely suffer from a weak first stage.

Table 1: First stage analysis

	(1)	(2)	(3)	(4)	(5)
	Dependent variable: Log(Tweets)				
SXSW March2007	0.168*	0.032**	0.049***	0.051***	0.050***
	(0.101)	(0.015)	(0.011)	(0.011)	(0.011)
SXSW 2006	0.114	0.037	−0.021	−0.008	−0.011
	(0.257)	(0.041)	(0.033)	(0.033)	(0.032)
R^2	0.326	0.923	0.937	0.942	0.943
Cluster robust F-stat	2.31	4.47	17.83	19.49	19.31
State FE	Yes	Yes	Yes	Yes	Yes
Population 2-percentile controls		Yes	Yes	Yes	Yes
Demographic controls			Yes	Yes	Yes
Socio-economic controls				Yes	Yes
Pre 2010 election controls					Yes
Observations	3,108	3,107	3,107	3,106	3,105

Note: This table presents county-level regressions where the natural logarithm of a sample of Tweets from 2014-2015 is the dependent variable. SXSW March2007 is the number of Twitter users who joined Twitter in March 2007 and follow South by Southwest (@SXSW). SXSW 2006 is the number of Twitter users who joined in 2006 and follow SXSW. Regressions include the indicated control variables (see Table A1 in the Appendix for their descriptions). Robust standard errors in parentheses are clustered by state. *p<0.1; **p<0.05; ***p<0.01

4.2 IV results

Table 2 presents estimations from the formal TSLS model using the same 5 specifications as before, the table provides 3 sets of results. Panel A shows introductory county-level OLS correlation between tweets and the number of demonstrations, Panel B shows the reduced-form, and Panel C reports the second stage results, the local average treatment effect of Twitter penetration on demonstration frequency.

In Panel A we see that there is a positive, significant, but quantitatively unimportant correlation between tweets and the number of demonstrations. After adding the main controls, a percentage increase in tweets is associated with approximately 0.03 additional demonstration events.

Panel B presents reduced-form relationship between the county-level number of SXSW 2007 followers and the total number of demonstrations. Across all reported specifications, I find a positive relationship between the number of SXSW 2007 followers and the number of demonstrations significant at the 5% level. After the inclusion of state fixed-effects and population controls, coefficient estimates remain largely stable across specifications: An additional SXSW 2007 follower is associated with approximately 13 additional demonstration events between January 2020 and 12th November 2021. This is a sizable effect, given that there are 11 demonstration events in a county on average over the period. On the contrary, the coefficient estimate of the effect of SXSW 2006 followers is only about one half of the effect of SXSW 2007 followers, it is also imprecisely estimated and statistically indistinguishable from zero. Again, to gauge at how likely it is that the relationship is driven by selection on unobservables, I calculate "Oster-δ's" comparing columns (1) and (5). Assuming that unobservables would fully explain variation in demonstration frequency, unobservable variables would have to be about 3.85 times as important as the controls added between the columns in terms of selection into treatment to explain away the results. As the demonstration data is computed manually from secondary sources we can expect some measurement error, assuming 5% of variation in demonstration frequency would be left unexplained by unobservable variables, they would have to be about 5 times as important as the observables bewteen the columns. Again, these numbers indicate a need for an implausibly high degree of selection to explain away the relationship. Figure 5 visualizes the regression from column (5) using binned scatter plots. The left hand panel shows the full sample and again, to verify that the relationship is not solely driven by outliers, the right hand panel shows the relationship on a sample excluding SXSW 2007 follower residual outliers outside of the range of -4 to 4. The bins (yellow dots) show a clear positive relationship between SXSW 2007 followers and the number of demonstrations. Although some part of the relationship seems to be driven by the outliers, dropping them reduces the average effect from 13 to about 9, there still remains a clear and sizable effect of SXSW 2007 followers on demonstration frequency. Figure A11 in the Appendix shows similar graphs for the number of SXSW 2006 followers and the total number of demonstrations. The graphs show quite clearly that no relationship exists between these two variables.

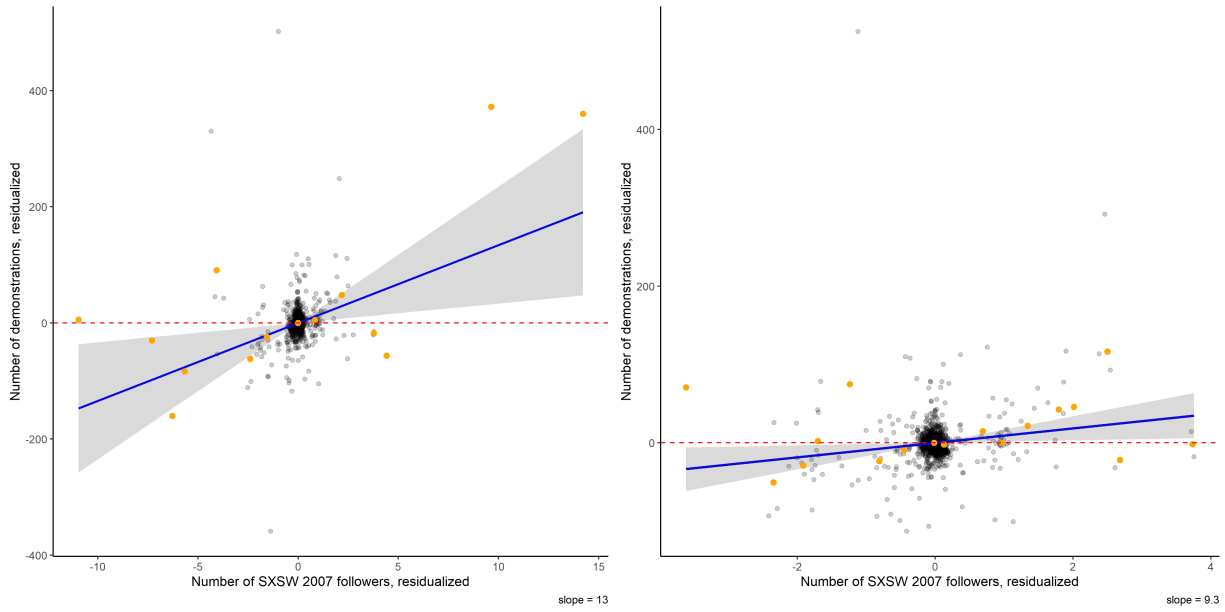
Table 2: Effect of Twitter penetration on demonstrations

	(1)	(2)	(3)	(4)	(5)
	Dependent variable: Number of demonstrations				
Panel A: OLS					
Log(Tweets)	8.960***	5.265***	3.025**	3.114**	2.903**
	(1.264)	(1.475)	(1.337)	(1.426)	(1.450)
R^2	0.549	0.715	0.733	0.736	0.736
Panel B: Reduced-form					
SXSW March2007	16.849***	12.356**	13.327**	13.395**	13.402**
	(5.777)	(5.506)	(5.283)	(5.254)	(5.237)
SXSW 2006	7.295	6.201	4.646	4.764	4.689
	(19.552)	(14.846)	(13.690)	(13.646)	(13.613)
R^2	0.482	0.736	0.762	0.762	0.764
Panel C: TSLS Second Stage					
log(Tweets)	100.236*	403.806***	280.442***	273.212***	275.770***
	(54.181)	(121.798)	(91.954)	(88.054)	(90.283)
SXSW 2006	−4.092	−8.928	10.493	6.734	7.661
	(19.368)	(11.877)	(8.645)	(8.847)	(8.839)
State FE	Yes	Yes	Yes	Yes	Yes
Population 2-percentile controls		Yes	Yes	Yes	Yes
Demographic controls			Yes	Yes	Yes
Socio-economic controls				Yes	Yes
Pre 2010 election controls					Yes
Mean number of demonstrations	11.00	11.00	11.00	11.00	11.00
Observations	3,108	3,107	3,107	3,106	3,105

Note: This table presents county-level regressions for an OLS and a TSLS model of the effect of Twitter penetration on demonstration frequency. Log(Tweets) is the natural logarithm of a sample of tweets from 2014-2015. SXSW March2007 is the number of Twitter users who joined Twitter in March 2007 and follow South by Southwest (@SXSW). SXSW 2006 is the number of Twitter users who joined in 2006 and follow SXSW. Regressions include the indicated control variables (see Table A1 in the Appendix for their descriptions). Robust standard errors in parentheses are clustered by state. *p<0.1; **p<0.05; ***p<0.01

As presented in Panel C, moving from left to right we see that coefficients vary quite a lot until in column (3) where they stabilize. In columns (3) to (5) the estimated effect of Twitter penetration on demonstrations is positive and significant: A percentage increase in Twitter penetration implies almost 3 additional demonstrations. This is, again, a sizable impact as the mean number of demonstrations per county is 11. On the contrary, SXSW 2006 followers do not seem to correlate with demonstration frequency. Comparing Panel A and Panel C, we note that the coefficients estimates are much lower in the OLS estimates where endogeneity is not taken care of. The comparison of OLS and TSLS results from Enikolopov, Makarin, & Petrova (2020), who use a similar identification strategy—a shock to early adoption as a plausibly exogenous source of variation in regional VK penetration—to estimate the effect of social media on protest participation in Russia gave similar results. This differences could be due to large negative selection of counties into Twitter penetration, for example that people who are in general less interested in societal issues or political activism and thus less likely to participate in demonstrations are more likely to join Twitter. Alternatively, the effects estimated in Panel C could reflect particularities in the local average treatment (LATE), i.e., that the effect of Twitter on demonstration activity is larger in counties where the effect of SXSW followers on Twitter penetration was also stronger.

Taken together, these results can be taken as evidence that Twitter may facilitate demonstration activity. These correlations are unlikely to be driven by selection of SXSW enthusiasts and attendees into areas with high demonstration activity as I control for the number of SXSW followers who created their accounts in 2006, and they also do not correlate with demonstration activity. Further, selection into "treatment" based on unobservables needs to be implausibly strong to explain away the effect.



(a) Full sample (b) Subsample excluding outliers

Figure 5: Reduced-form visualization

Note: This figure presents binned scatter plots of the relationship between county-level number of demonstrations between Jan 2020 and Nov 2021, and the number of SXSW followers who joined Twitter in March 2007. Variables are residualized by partialling out SXSW followers who joined in 2006, population 2-percentiles, demographic, socio-economic, and pre-2010 election controls (see Table A1 in the Appendix for their descriptions). The left hand side figure shows the full sample, while the right hand side figure shows a subsample excluding outier counties with abnormally high or low values for the SXSW followers who joined Twitter in March 2007 residual. Grey dot's represent individual observations and yellow dots represent average values of both variables within a bin. The blue line represents a line of best fit using the unbinned data, and the shaded blue area are 95% confidence intervals calculated using robust standard errors clustered at the state-level.

When interpreting these results, two conceptual questions arise. First, the large differences between the OLS estimates and IV estimates could be due to selection in the former estimations. Alternatively, they could be due to particularities of the LATE in the TSLS estimations, or both. However, as discussed above, the festival is likely to attract a wide variety of individuals, so the LATE should not be excessively particular, although caution with respect to overgeneralization is of course needed. Second, one may wonder to what extent the causal effects estimated can be exclusively attributed to Twitter, as opposed to other social media platforms. It is quite possible that, while the initial diffusion through SXSW in 2007 was likely specific to Twitter, there existed significant spillovers in adoption of other social media platforms in the medium-run. Therefore, the estimated effect might not be pure "Twitter effect."

These spillovers should however not change the interpretation of estimated effects, namely that these estimates refer to the causal effect of social media diffusion on demonstration frequency.

5 Heterogeneity analysis

5.1 Coordination complexity

Now that I have established a causal relationship between Twitter penetration and the frequency of demonstrations, I move on to explore heterogeneities in this effect depending on the nature of the protest movement. I start by looking at protests by coordination complexity. As explained in section 2.1, we approximate coordination complexity by assuming that demonstrations that are part of larger "broad social movements", or events which cannot be associated with any organization, are relatively complex to coordinate. Table 3 shows the results of this heterogeneity analysis. I count the number of demonstrations that either a) are a part of a broad social movement or cannot be associated with an organization or b) are not a part of a broad social movement and cannot be associated with an organization. In Panel A I run the main TSLS specification with the full set of controls on the number of demonstrations that fit the mutually exclusive categories a) or b) separately. Panel B restricts the sample to only counties where at least 1 demonstration event occurred over the period, and runs the same specifications on the share of demonstrations, among all demonstration in the county, that fit category a) or b).

In Panel A we see that Twitter penetration has a significant and positive effect on demonstrations of both categories: A percent increase in Twitter penetration implies 1.8 additional demonstrations that are either part of a broad social movement or have no organization attached to them, and implies 1.2 additional other demonstrations. Although the coefficient on the demonstrations assumed to be more complex to coordinate is larger, it is not clear that this implies an disproportionate effect for these demonstrations as they are on average more numerous. However, Panel B seems to confirm that the effect is indeed larger for protests part of broad social movements or otherwise not associated with any organizations. We see that, among counties that do have some demonstration activity, Twitter penetration increases the share of the demonstrations assumed more complex to coordinate significantly at the 10% level. However, this effect is not majorly transformative of the demonstration landscape, an percentage increase in Tweets implies only an increase in the share of these

protests of 0.14 percentage points (from a mean of 0.72%). Alternatively, as one extra SXSW 2007 follower increases Twitter penetration by 5%, this one follower implies an increase in the share of 0.6 percentage points.[9]

In Tables A6 and A7 in the Appendix I show that this disproportionate effect is relatively robust to different approximations of complexity of coordination. I compare the effect of Twitter penetration on demonstrations that are associated with a broad social movement versus all others; on demonstrations that are *only* associated with a broad social movement versus all others; on demonstrations that cannot be associated with an organization nor a broad social movement versus all others; or on demonstrations that are *only* associated with a broad social movement *or* that cannot be associated with an organization nor a broad social movement versus all other. I also define broad social movements slightly differently and do the same comparisons. The effect on the shares of demonstrations that fit each of these categories is in general smaller and in many cases indistinguishable from zero. One plausible explanation is that these alternative definitions do not capture all of the "hard to coordinate" demonstrations captured in my preferred specification, the effect on those "hard to coordinate" demonstrations not captured runs through the denominator, but not the numerator of the share, biasing the estimate downwards. However, in all cases the point estimate of the effect of Twitter penetration on the shares is of the expected sign, suggesting that the share of demonstrations of the category that is likely to be more complex to coordinate increases with increased Twitter penetration.

[9] $0.141 \times log(\frac{105}{100})$

Table 3: TSLS demonstrations by coordination complexity

	(1)	(2)	(3)
Type of demonstrations:		Broad Social Movements + No Org.	Other demonstrations
Specification:	First Stage	Second Stage	
Panel A			
Dependent variables:	Log(Tweets)	*Number of demonstrations*	
SXSW March 2007	0.050***		
	(0.011)		
log(Tweets)		181.692***	126.108***
		(61.430)	(41.405)
SXSW 2006	−0.011	0.685	6.766
	(0.032)	(5.827)	(4.386)
R^2	0.943		
Cluster robust F-stat	19.31		
Mean dep. var.		6.89	4.11
N	3,105	3,105	3,105
Panel B			
Dependent variables:	Log(Tweets)	*Share among all demonstrations*	
SXSW March 2007	0.050***		
	(0.011)		
log(Tweets)		0.141*	−0.141*
		(0.073)	(0.073)
SXSW 2006	−0.004	−0.016**	0.016**
	(0.032)	(0.007)	(0.007)
R^2	0.946		
Cluster robust F-stat	18.19		
Mean dep. var.		0.72	0.28
N	1,782	1,782	1,782

Note: This table presents county-level regressions. Log(Tweets) is the natural logarithm of a sample of tweets from 2014-2015. SXSW March2007 is the number of Twitter users who joined Twitter in March 2007 and follow South by Southwest (@SXSW). SXSW 2006 is the number of Twitter users who joined in 2006 and follow SXSW. Column (1) presents first stage regressions, while columns (2) and (3) present second stages. "No Org." refers to demonstrations that cannot be associated with a group or an organization. Panel A presents TSLS regressions where the number of demonstrations that are "BSMs" or "No Org." is the dependent variable (column 2), or where the number of other demonstrations is the dependent variable (column 3). Panel B presents regression results on a subsample of counties where at least 1 demonstration event occurred between January 2020 and November 2021, where the dependent variable is the share of demonstrations that fit the category. All regressions control for population via 2-percentile dummies, and the full set of controls (as in column 5 in Table 1), see Table A1 in the Appendix for their descriptions. Robust standard errors in parentheses are clustered by state.
*p<0.1; **p<0.05; ***p<0.01

Overall, these results points to an disproportionate effect of Twitter penetration on demonstration movements with a more challenging coordination problem. In particular, I find evidence that social media particularly facilitates demonstrations that are a part of "broad social movements," national movements demanding social or policy change on a large scale. However, these heterogeneities are quantitatively small, suggesting that Twitter is unlikely to have dramatically changed the nature of demonstration activity in the United States along these dimensions.

5.2 Demonstration topics

Next, I examine whether Twitter penetration influences demonstration activity concerning some specific topics disproportionately. In particular, I test whether the effect of Twitter penetration is larger for demonstrations concerning topics that the US population holds particularly polarized views on. As explained in section 2.1, I filter out events associated with a given topic. Then I perform similar analysis as in section 5.1, comparing the effect of Twitter penetration on the total number and shares of demonstrations associated with a given topic. The results of this analysis is shown in Table 4. Column (1) shows the corresponding first stage and in columns (2) to (5) look specifically at demonstrations concerning one specific topic: a) Public health measures due to the Coronavirus pandemic, b) Demonstrations contesting presidential election results, or those demanding results to be respected (represented in column (3), as "Stop the Steal" related demonstrations), c) Demonstrations concerning abortion rights, d) Demonstrations concerning racial issues. Column (6) examines a subgroup of demonstrations concerning racial issues—those associated with groups affiliated with white supremacy ideologies. Finally, in column (7) I group demonstrations associated with any of the topics mentioned under the umbrella term "Polarizing topics."[10] Panel A presents TSLS estimates of the effect of Twitter acitvity on the number of demonstrations of a given category, and Panel B shows the estimated effect on the demonstrations of a given category as a share of all demonstrations. Both panels are estimated with the full set of controls.

Panel A shows that Twitter penetration seems to have an effect on demonstrations concerning all topics. The precision of the estimates vary somewhat, only demonstrations concerning racial issues are statistically different from zero at the 1% level, but

[10]Note that the categories in columns (2) to (5) are not mutually exclusive. For example, some protests associated with "Stop the Steal" might also demand the end of mask-mandates as a public health measure due to the Coronavirus pandemic. Demonstrations that fit more than one topic category are not double counted in column (7).

all estimates are at least statistically significant at the 10% level. I estimates that a percentage increase in Twitter penetration implies 0.3 additional Coronavirus demonstrations, 0.02 additional "Stop the Steal" related demonstrations, 0.03 additional demonstrations concerning abortion rights, 1.2 additional demonstrations racial issue related demonstrations and 0.07 additional demonstrations associated with white supremacists. Overall, a percentage increase in Twitter penetration is estimated to increase demonstrations concerning these "Polarizing topics" by 1.6. The effect sizes are in general quite equal relative to the frequency of demonstrations in each topic category, but it is worth noting that the coefficient on protests associated with white supremacists is both larger and performs better on standard t-tests than the coefficients for "Stop the Steal" or abortion rights related protests—both of which are more numerous than protests associated with white supremacists.

However, moving to Panel B, we find almost no evidence suggesting that Twitter penetration increases the share of protests fitting any of the topic categories. Coefficients are in most cases quite close to zero and imprecisely estimated. The only exception is that there is some slight evidence of an Twitter penetration induced increase in the share of demonstrations that concern racial issues. The coefficient estimate for this effect is very small, a percentage increase in Twitter penetration is estimated to increase the share of demonstrations that concern racial issues by 0.125 percentage points while these demonstrations represent on average half of all demonstrations. However, Table A8 in the Appendix shows that this effect is solely driven by the Black Lives Matter Movement, a "Broad Social Movement." In the table, I calculate the share of demonstrations that fit the "Racial issues" or the overall "Polarizing topics" categories, but are not associated with Black Lives Matter. The table clearly shows that the effect of Twitter penetration on these shares is non-existent. The fact that I find relatively robustly heterogeneous effects depending on coordination complexity, but no disproportionate effects for any "Polarizing topic" that is not convoluted with Broad Social Movements is suggestive that the positive effect on the share of protests concerning racial issues is driven by the fact that most protests concerning racial issues were part of the nationwide Black Lives Matter movement (or counter-movements), i.e., driven by coordination complexity of these protests rather than the nature of the topic itself. Finally, one might be worried that, even if the effect of Twitter penetration on protests of a given category are larger-than-average, we could find zero-effects on the share of these protests among all protests, if the numerator includes other protest categories that are also disproportionately effected.[11]

[11]Imagine for example an unusually large effect on protests concerning abortion rights, and on

Therefore, in Table A9 in the Appendix I calculate the share of demonstrations in the categories shown in columns (2) to (6) among the number of protests in *that specific category* and demonstrations on "non-polarizing topics," i.e., protests fitting none of these categories, and run the main specification with these shares as outcome variables on a sample of counties with at least one demonstration in that category or in the "non-polarizing topics" category. By doing this, the denominator stays unaffected by effects on demonstrations in other "Polarizing topics" categories. This table shows broadly the same picture, protests of none of these topic categories seem to be particularly effected by Twitter penetration.

Taken together, these results seem to suggest that Twitter penetration does not seem to influence *what* is being protested, at least not in terms of the polarizing nature of topics of demonstration events.

protests associated with white supremacists. We could find zero-effects on the share of protests that are associated with white supremacists simply because the denominator of the share is doubly disproportionately effected as both white supremacists and abortion right demonstrations increase with increased Twitter penetration.

Table 4: TSLS demonstrations by topic

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Type of demonstrations:		Coronavirus	"Stop the Steal"	Abortion rights	Racial issues	White Supremacist	Polarizing topics
Specification:	First Stage	Second Stage					
Panel A							
Dependent variables:	Log(Tweets)	*Number of demonstrations*					
SXSW March 2007	0.050***						
	(0.011)						
log(Tweets)		33.553**	2.402*	2.734**	120.860***	7.547**	156.271***
		(14.679)	(1.234)	(1.126)	(42.400)	(3.023)	(50.250)
SXSW 2006	−0.011	0.201	−0.068	0.169	−0.844	−0.253	−0.487
	(0.032)	(1.180)	(0.152)	(0.128)	(4.223)	(0.313)	(4.709)
R^2	0.943						
Cluster robust F-stat	19.31						
Mean dep. var.		1.10	0.20	0.25	4.28	0.14	5.72
N	3,105	3,105	3,105	3,105	3,105	3,105	3,105
Panel B							
Dependent variables:	Log(Tweets)	*Share among all demonstrations*					
SXSW March 2007	0.050***						
	(0.011)						
log(Tweets)		−0.037	0.019	0.015	0.125*	−0.001	0.124
		(0.033)	(0.014)	(0.023)	(0.073)	(0.013)	(0.082)
SXSW 2006	−0.004	0.0003	−0.003*	−0.001	−0.014	0.0002	−0.017*
	(0.032)	(0.004)	(0.001)	(0.003)	(0.009)	(0.002)	(0.009)
R^2	0.946						
Cluster robust F-stat	18.19						
Mean dep. var.		0.096	0.012	0.021	0.495	0.007	0.622
N	1,782	1,782	1,782	1,782	1,782	1,782	1,782

Note: This table presents county-level regressions. Log(Tweets) is the natural logarithm of a sample of tweets from 2014-2015. SXSW March2007 is the number of Twitter users who joined Twitter in March 2007 and follow South by Southwest (@SXSW). SXSW 2006 is the number of Twitter users who joined in 2006 and follow SXSW. Column (1) presents first stage regressions, while columns (2)-(7) present second stages. Coronavirus refers to demonstrations related to public health measures due to the Covid pandemic. "Stop the Steal" refers to demonstrations related to the 2020 post-presidential election unrest (including those calling for results to be respected). Abortion rights refers to demonstrations related to reproductive rights. Racial issues refers to demonstrations concerning racial tensions while white supremacists are a subset of those where demonstrating groups have known affiliations with white supremacy ideology. Polarizing topics include all other topics in the table. A detailed description of topic categorization is presented in the Appendix. Panel A presents TSLS results where the number of demonstrations fitting the given column category is the dependent variable. Panel B presents TSLS results on a subsample of counties where at least 1 demonstration event occurred between Jan 2020 and Nov 2021, where the dependent variable is the share of demonstration that fit the column-category. All regressions control for population via 2-percentile dummies, and the full set of controls (as in column 5 in Table 1), see Table A1 in the Appendix for their descriptions. Robust standard errors in parentheses are clustered by state. *p<0.1; **p<0.05; ***p<0.01

5.3 Attempts of suppression or the use of violence during demonstration events

This section analyses whether Twitter plays a role in reported increases of hostility, violence, and repression during demonstration events in the United States. I start by contrasting the effect of Twitter penetration on demonstrations events that are "Two sided" versus those that are "One sided." Next I examine whether there are heterogeneous effects on protests that are peaceful versus those characterized by attempts of suppression or violent behaviour on behalf of demonstrators or outside parties, i.e., non-peaceful demonstrations. Finally, I examine separately demonstrations where the demonstrators themselves are violent or exhibiting disruptive behaviour, and demonstrations where outside groups or individuals interact non-peacefully with demonstrators. Table 5 presents TSLS estimates of the effect of Twitter penetration on these different demonstration categories. Panel A shows estimates where the total number of demonstrations in each category is the dependent variable, and Panel B shows estimate where the share of each category among all demonstrations is the dependent variable.

Columns (2) and (3) show that Twitter penetration has a significant effect on both one and two sided demonstrations. A percentage increase in Twitter penetration implies around 2.5 additional one sided demonstrations and 0.2 additional two sided demonstrations. As there are on average substantially fewer two sided demonstrations, the effect size as a share of the mean is larger for two sided demonstrations. However, Panel B shows that Twitter penetration does not significantly affect the relative number of two versus one sided demonstrations, the coefficient to the effect of Twitter penetration on the share of two sided demonstrations is close to zero and very imprecisely estimated. Moving to columns (4) and (5) which compare peaceful versus non-peaceful protests, we see that the coefficient is precisely estimated for peaceful protests implying 2 additional protests for a percentage increase in Twitter penetration, while the coefficient is less precisely estimated for non-peaceful protests. However, as the majority of demonstrations in the sample are peaceful, the (imprecisely estimated) effect size for non-peaceful protests is much larger. Panel B shows that we cannot rule out a null-effect on the share of non-peaceful protests due to large standard errors. The noise in the effect on non-peaceful protests warrants further disaggregation, provided in columns (6) and (7). In column (6) we see that estimated effects on protests characterised by violent protesters is very imprecisely estimated. On the other hand, column (7) shows that the effect on those where demonstrators

themselves are peaceful, but are met with interference or violence by outside actors is distinguishable from zero at the 1% level, a percentage increase in Twitter penetration increases such protests by 0.2 on average. However, Panel B shows that we cannot rule out that Twitter penetration has no effect on the share of demonstrations in either sub-group of non-peaceful demonstrations, although point estimates for both subgroups are positive. The categories in columns (2) and (3) on one hand, and in columns (4) and (5) on the other, split the total number of demonstrations in two. The number of demonstrations from either columns (6) and (7) however cannot be added to the number of demonstration from any one category to make up all demonstrations. Therefore, for the same reason as in section 5.2, I calculate the share of demonstrations in the categories in columns (6) or (7) among the number of demonstrations in that specific category and peaceful protests, and run the main specification on these outcome variables, on a sample of counties with either peaceful protests or protests of that specific category. These results are presented in Table A10 in the Appendix, coefficient estimates are a bit larger but as imprecisely estimated and statistically indistinguishable from zero. In other words, this alternative test does not affect the broad picture. Finally, Table A11 in the Appendix disaggregates the categories reported in column (6) and (7), as both comprise two "sub-event types" as reported in the ACLED data. The broad story holds and this additional dissagregation of events does not introduce further subtleties.

Summing up the above, we find evidence that Twitter penetration does indeed increase demonstration events characterized by hostile interactions, or violence or repression by outside parties. However, we find no particular evidence suggesting that Twitter plays a role in the relative frequency of hostile or violent demonstration events. In what follows, I distinguish peaceful protests met with intervention or violence by an outside party, depending on if the outside party is a government entity or not.

Table 5: TSLS demonstrations by "peacefulness"

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Type of demonstration:		One Sided	Two Sided	Peaceful	Non-Peaceful	Violent Protesters	Peaceful Protesters met with intervention or violence
Specification:	First Stage	Second Stage					
Panel A							
Dependent variables:	Log(Tweets)	*Number of demonstrations*					
SXSW March2007	0.050***						
	(0.011)						
log(Tweets)		248.589***	17.024***	206.608***	59.004*	37.506	21.498***
		(81.247)	(6.124)	(73.486)	(31.946)	(25.885)	(7.357)
SXSW 2006	−0.011	7.984	−0.345	9.965	−2.326	−1.911	−0.415
	(0.032)	(7.956)	(0.533)	(7.036)	(3.223)	(2.565)	(0.751)
R^2	0.943						
Cluster robust F-stat	19.31						
Mean dep. var.		10.47	0.53	10.30	0.70	0.33	0.37
Effect size as % of mean		0.24	0.32	0.20	0.84	1.15	0.58
Panel B							
Dependent variables:	log(Tweets)	*Shares among all demonstrations*					
SXSW March2007	0.050***						
	(0.011)						
log(Tweets)			−0.010		0.094	0.066	0.030
			(0.025)		(0.067)	(0.056)	(0.024)
SXSW 2006	−0.004		0.0004		−0.009	−0.006	−0.002
	(0.032)		(0.003)		(0.008)	(0.006)	(0.003)
R^2	0.946						
Cluster robust F-stat	18.19						
Mean dep. var.			0.003		0.034	0.012	0.022

Note: This table presents county-level regressions. Log(Tweets) is the natural logarithm of a sample of tweets from 2014-2015. SXSW March2007 is the number of Twitter users who joined Twitter in March 2007 and follow South by Southwest (@SXSW). SXSW 2006 is the number of Twitter users who joined in 2006 and follow SXSW. Column (1) presents first stage regressions, while columns (2)-(7) present second stages. "Two sided" refers to events where two demonstrations who are essentially counter one another occur at the same time in the same place. "One sided" refers to all demonstrations that are not "Two sided." Violent protesters are events where demonstrators engage in violent or disruptive behaviour and Peaceful Protesters met with intervention or violence are events where demonstrators are peaceful but there is an attempt of suppression or the use of violence by outside parties. Non-Peaceful refers to the sum of the protests in columns (6) and (7) while Peaceful refers to all other demonstrations. Panel A presents TSLS results where the number of demonstrations fitting the given column category is the dependent variable. Panel B presents TSLS results on a subsample of counties where at least 1 demonstration event occurred between Jan 2020 and Nov 2021, where the dependent variable is the share of demonstration that fit the column-category. All regressions control for population via 2-percentile dummies, and the full set of controls (as in column 5 in Table 1), see Table A1 in the Appendix for their descriptions. Robust standard errors in parentheses are clustered by state. *p<0.1; **p<0.05; ***p<0.01

5.3.1 Who intervenes?

As explained in the conceptual framework, one could expect different effects of social media on violence or attempts of suppression by outside actors during demonstrations, depending on whether the outside actor is a government entity or not. We could expect Twitter penetration to constraint government willingness to perform violent or repressive acts through accountability mechanisms, while Twitter penetration might fuel violence or interference by non-government actors through facilitating coordination, or by fuelling hostility or emotional reactions by those opposing of those opposing the demonstration cause. To test this, I split demonstrations where peaceful protesters are met with intervention or violence (those counted in column (7) of Table 5) in two depending on whether the intervening actor is a government entity or not. Table 6 presents TSLS estimates of the effect of Twitter penetration on these demonstration categories. Panel A shows estimates of the effect on the total number of these demonstrations, while Panel B shows estimates of the effect on the share of demonstrations of these categories among all demonstrations.

Panel A shows that Twitter penetration has a significant positive effect on both types of demonstrations, although for both these effects are quantitatively small: A percentage increase in Twitter penetration implies 0.16 additional demonstrations met with intervention or violence by the government, and 0.05 additional demonstrations met with intervention or violence by non-government entities. These categories of demonstrations are not too common, so these effects amount to about half of the average number of demonstrations of that category. In Panel B, we see that the effect of Twitter penetration on the share of peaceful demonstrations that are met with intervention or violence by the government is very imprecisely estimated and indistinguishable from zero. Twitter penetration does however increase the share of peaceful demonstrations met with intervention or violence by non-government actors, this effect is significant at the 5% level. Although the effect is rather precisely estimated, it is quantitatively small, an percentage increase in Twitter penetration implies that these protests represent 0.008 percentage points more of total demonstration activity (from an average of 0.3 percent). Alternatively, an additional SXSW follower implies an increase in the share by 0.04 percentage points[12] As before, in Table A12 in the Appendix I present TSLS results using the same right hand side specification where the outcome variable is share of protests fitting the category in column (2) or (3), among peaceful demonstrations or those fitting the given category, on a sample of

[12] $0.008 \times ln(\frac{105}{100})$

counties only including counties where either a peaceful protest or a protest of the category occurred, results remain qualitatively similar.

Therefore, the evidence reported here does not confirm that Twitter constraints governments willingness to use violence or try to suppress demonstration activity, on the contrary I find that Twitter penetration increases the occurrence of such events equally to other demonstration events. On the other hand, we do find some evidence suggesting that Twitter penetration increases violence or other interruptions during demonstration events by non-government actors. Although this effect is minuscule and thus not a plausible contender to explain the reported overall rise in violence and hostility during demonstration events, the occurrence of non-governmental attempts at suppression or the use of violence during demonstration event is an extreme outcome, so the effect should not be disregarded.

To shed light on these results, I present two pieces of information. First, who are these "non-government" intervening actors? In 113 out of 177 events fitting this category, the intervening actor is a sole perpetrator, while the rest of the events are met with intervention or violence from Rioters, private security forces, or armed groups. Sadly, for 135 out of 177 events fitting this category there is no information on the groups or organizations that intervene or use violence. For the remaining 25% of events in this category, the groups or organizations that intervene are often associated with the Black Lives Matter movement, the counter movements Blue Lives Matter and Back the Blue, or groups associated with white supremacy ideologies. A full list of groups and organizations associated with interventions or the use of violence against peaceful protesters is reported in Table A13 in the Appendix. Finally, to gauge at mechanisms behind this increase, I code separately the share of demonstrations where a sole non-government perpetrator makes attempts at suppression or uses violence, and the share of demonstrations where a non-government group makes attempts at suppression or uses violence. Table A14 shows results of the effect of Twitter penetration on these shares, where we see that the effect found in Table 6 is driven solely by interventions or violence by non-government *groups*. From these pieces of information we can infer that Twitter penetration disproportionately increases non-peaceful demonstrations that concern the highly polarized topic of racial issues—where hostile groups of individuals from "the other side" attend and exhibit disruptive or violent behaviour. Finally, as this effect is driven by violent *groups* as opposed to individuals, we can infer that coordination among those who aim to disrupt the otherwise peaceful demonstration event plays an important role.

Table 6: TSLS demonstrations by intervening actor

	(1)	(2)	(3)
Intervening actor:		Intervention by Gov.	Intervention by non-Gov. Actor
Specification:	First Stage	Second Stage	
Panel A			
Dependent variables:	Log(Tweets)	*Number of demonstrations where peaceful protesters are met with intervention or violence*	
SXSW March 2007	0.050***		
	(0.011)		
log(Tweets)		16.667***	4.831***
		(5.744)	(1.677)
SXSW 2006	−0.011	−0.234	−0.181
	(0.032)	(0.586)	(0.173)
R^2	0.943		
Cluster robust F-stat	19.31		
Mean dep. var.		0.31	0.06
Effect size as % of mean		0.54	0.51
N	3,105	3,105	3,105
Panel B			
Dependent variables:	Log(Tweets)	*Share among all demonstrations*	
SXSW March 2007	0.050***		
	(0.011)		
log(Tweets)		0.024	0.008**
		(0.022)	(0.004)
SXSW 2006	−0.004	−0.002	−0.001**
	(0.032)	(0.003)	(0.0004)
R^2	0.946		
Cluster robust F-stat	18.19		
Mean dep. var.		0.019	0.003
Effect size as % ..		0.012	0.031
N	1,782	1,782	1,782

Note: This table presents county-level regressions. Log(Tweets) is the natural logarithm of a sample of tweets from 2014-2015. SXSW March2007 is the number of Twitter users who joined Twitter in March 2007 and follow South by Southwest (@SXSW). SXSW 2006 is the number of Twitter users who joined in 2006 and follow SXSW. Column (1) presents first stage regressions, while columns (2) and (3) present second stages. Column (2) looks at demonstrations where peaceful where peaceful protesters are met with an attempt of suppression or the use of violence by government entities, while column (3) looks at demonstrators where peaceful protesters are met with an an attempt of suppression or the use of violence by non-government actors. Panel A presents TSLS results where the number of demonstrations fitting the given column category is the dependent variable. Panel B presents TSLS results on a subsample of counties where at least 1 demonstration event occurred between Jan 2020 and Nov 2021, where the dependent variable is the share of demonstration that fit the column-category. All regressions control for population via 2-percentile dummies, and the full set of controls (as in column 5 in Table 1), see Table A1 in the Appendix for their descriptions. Robust standard errors in parentheses are clustered by state. *p<0.1; **p<0.05; ***p<0.01

6 Conclusion

In recent years, the United States have witnessed skyrocketing levels of demonstration activity and political unrest. Reports also indicate that these demonstration events have become more violent. Researchers have convincingly documented the power of social media with regard to facilitating protest movements (Enikolopov, Makarin, & Petrova, 2020; Manacorda & Tesei, 2020; Amorim et al., 2018; Fergusson & Molina, 2019). Further, social media's potential effect on political polarization, hostility towards those holding different political views than oneself or it's ability to propagate hateful sentiments raise serious concerns about how social medias can impact societies. In this paper, I investigate whether social media can effect the landscape of political protests beyond the size and frequency of protest events.

I leverage a plausibly exogenous shock in early adoption of Twitter stemming from attendance at the South by Southwest festival in 2007 to predict variation in county-level Twitter penetration later on. Utilizing this shock in an instrumental variable strategy, I find that Twitter penetration increases demonstration activity considerably, consistent with prior research. By examining the underlying organizational structure of protests, I find that protest movements characterized by lack of centralized organization and a call for a broad national social or policy change are disproportionally affected by Twitter penetration, relative to other protest movements that are less reliant on social media for coordination. However, the scale of this heterogeneity is not large and does not point towards a dramatic Twitter induced change in organizational structure of protest movements. Next, I find evidence that Twitter penetration increases protests concerning especially polarizing topics or hateful sentiments, but no strong evidence that the relative frequency of such protests is affected. I also find that Twitter penetration increases non-peaceful demonstrations, those characterized by violent demonstrators or peaceful protesters met with interference or violence by outside parties. These non-peaceful demonstrations do not seem to be affected relatively more than peaceful protests, with the exception of protests where peaceful protesters who are met with interference or violence by non-government actors. I find evidence that coordination among non-government actors who aim to disrupt otherwise peaceful protests, or use violence against protesters, plays an important role in this Twitter induced relative increase. Although Twitter penetration implies an increase in the share of these protests relative to other demonstrations, the effect is again not large enough to explain a substantial shift in the violence level during protest events. However, even a small increase in the relative frequency of

violence during demonstration events is worrying.

Overall, these results imply that while Twitter facilitated protest movements during the period, its effect did not vary dramatically by the characteristics of the demonstration, leaving the protest landscape largely unaffected along dimensions other than overall activity.

References

ACLED. (2020a). *Cdt spotlight: Covid-19 us protest patterns.* Retrieved from https://acleddata.com/2020/11/30/cdt-spotlight-covid-19-us-protest-patterns/ (Last accessed 21 May 2022)

ACLED. (2020b). *Demonstration trends in the united states.* Retrieved from https://acleddata.com/2020/09/23/demonstration-trends-in-the-united-states/ (Last accessed 21 May 2022)

ACLED. (2020c). *Mid-year update: 10 conflicts to worry about in 2020.* Retrieved from https://acleddata.com/2020/08/18/mid-year-update-10-conflicts-to-worry-about-in-2020/#1597759650330-f3890636-9943 (Last accessed 21 May 2022)

Adena, M., Enikolopov, R., Petrova, M., Santarosa, V., & Zhuravskaya, E. (2015). Radio and the rise of the nazis in prewar germany. *The Quarterly Journal of Economics*, *130*(4), 1885–1939.

Allcott, H., Braghieri, L., Eichmeyer, S., & Gentzkow, M. (2020). The welfare effects of social media. *American Economic Review*, *110*(3), 629–76.

Amorim, G., Costa Lima, R., & Sampaio, B. (2018). Broadband internet and protests: evidence from the occupy movement. *Available at SSRN 2764162*.

Andrews, I., Stock, J. H., & Sun, L. (2019). Weak instruments in instrumental variables regression: Theory and practice. *Annual Review of Economics*, *11*, 727–753.

Arrow, K. J. (2000). Increasing returns: historiographic issues and path dependence. *The European Journal of the History of Economic Thought*, *7*(2), 171–180.

Avetian, V., Artís, A. C., Sardoschau, S., & Saxena, K. (2021). Going viral in a pandemic: Social media and allyship in the black lives matter movement.

Barbera, S., Jackson, M. O., et al. (2020). A model of protests, revolution, and information. *Quarterly Journal of Political Science*, *15*(3), 297–335.

Barberá, P. (2014). How social media reduces mass political polarization. evidence from germany, spain, and the us. *Job Market Paper, New York University*, *46*.

Barberá, P., & Rivero, G. (2015). Understanding the political rep- resentativeness of twitter users. *Social Science Computer Review*, *33*(6).

Boxell, L., Gentzkow, M., & Shapiro, J. M. (2017). Greater internet use is not associated with faster growth in political polarization among us demographic groups. *Proceedings of the National Academy of Sciences*, *114*(40), 10612–10617.

Brewster, T. (2021). Sheryl sandberg downplayed facebook's role in the capitol hill siege—justice department files tell a very different story. *Retrieved from https://www.forbes.com/sites/thomasbrewster/2021/02/07/ sheryl-sandberg-downplayed-facebooks-role-in-the-capitol-hill-siege -justice-department-files-tell-a-very-different-story/ (Accessed: 2020- 03-19)*.

Bursztyn, L., Egorov, G., Enikolopov, R., & Petrova, M. (2019). *Social media and xenophobia: evidence from russia* (Tech. Rep.). National Bureau of Economic Research.

Bursztyn, L., Egorov, G., & Fiorin, S. (2017). *From extreme to mainstream: How social norms unravel* (Tech. Rep.). National Bureau of Economic Research.

Cantoni, D., Yang, D. Y., Yuchtman, N., & Zhang, Y. J. (2019). Protests as strategic games: experimental evidence from hong kong's antiauthoritarian movement. *The Quarterly Journal of Economics*, *134*(2), 1021–1077.

Caprettini, B., Caesmann, M., Voth, H.-J., & Yanagizawa-Drott, D. (2022). *Going viral: Propaganda, persuasion and polarization in 1932 hamburg* (Tech. Rep.). Technical Report, Working Paper.

Christensen, D., & Garfias, F. (2018). Can you hear me now? how communication technology affects protest and repression. *Quarterly journal of political science*, *13*(1), 89.

DellaVigna, S., Enikolopov, R., Mironova, V., Petrova, M., & Zhuravskaya, E. (2014). Cross-border media and nationalism: Evidence from serbian radio in croatia. *American Economic Journal: Applied Economics*, *6*(3), 103–32.

DellaVigna, S., & Gentzkow, M. (2010). Persuasion: empirical evidence. *Annu. Rev. Econ.*, *2*(1), 643–669.

Durante, R., & Zhuravskaya, E. (2018). Attack when the world is not watching? us news and the israeli-palestinian conflict. *Journal of Political Economy*, *126*(3), 1085–1133.

Edmond, C. (2013). Information manipulation, coordination, and regime change. *Review of Economic studies*, *80*(4), 1422–1458.

Enikolopov, R., Makarin, A., & Petrova, M. (2020). Social media and protest participation: Evidence from russia. *Econometrica*, *88*(4), 1479–1514.

Enikolopov, R., Makarin, A., Petrova, M., & Polishchuk, L. (2020). Social image, networks, and protest participation. *Networks, and Protest Participation (April 26, 2020)*.

Enikolopov, R., Petrova, M., & Zhuravskaya, E. (2011). Media and political persuasion: Evidence from russia. *American Economic Review*, *101*(7), 3253–85.

Falck, O., Gold, R., & Heblich, S. (2014). E-lections: Voting behavior and the internet. *American Economic Review*, *104*(7), 2238–65.

Fergusson, L., & Molina, C. (2019). Facebook causes protests. *Documento CEDE*(41).

Fujiwara, T., Müller, K., & Schwarz, C. (2021). *The effect of social media on elections: Evidence from the united states* (Tech. Rep.). National Bureau of Economic Research.

Gentzkow, M. (2006). Television and voter turnout. *The Quarterly Journal of Economics*, *121*(3), 931–972.

Gentzkow, M., & Shapiro, J. M. (2011). Ideological segregation online and offline. *The Quarterly Journal of Economics*, *126*(4), 1799–1839.

Gentzkow, M., Shapiro, J. M., & Sinkinson, M. (2011). The effect of newspaper entry and exit on electoral politics. *American Economic Review*, *101*(7), 2980–3018.

González, F. (2020). Collective action in networks: Evidence from the chilean student movement. *Journal of Public Economics*, *188*, 104220.

Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019). Fake news on twitter during the 2016 us presidential election. *Science*, *363*(6425), 374–378.

Guriev, S., Melnikov, N., & Zhuravskaya, E. (2021). 3g internet and confidence in government. *The Quarterly Journal of Economics*, *136*(4), 2533–2613.

Halberstam, Y., & Knight, B. (2016). Homophily, group size, and the diffusion of political information in social networks: Evidence from twitter. *Journal of public economics*, *143*, 73–88.

Kinder-Kurlanda, K., Weller, K., Zenk-Möltgen, W., Pfeffer, J., & Morstatter, F. (2017). Archiving information from geotagged tweets to promote reproducibility and comparability in social media research. *Big Data & Society*, *4*(2), 2053951717736336.

Klein Teeselink, B., & Melios, G. (2021). Weather to protest: The effect of black lives matter protests on the 2020 presidential election. *Available at SSRN 3809877*.

Lelkes, Y., Sood, G., & Iyengar, S. (2017). The hostile audience: The effect of access to broadband internet on partisan affect. *American Journal of Political Science*, *61*(1), 5–20.

Liebowitz, S. J., & Margolis, S. E. (1995). Path dependence, lock-in, and history. *Journal of Law, Economics, & Organization*, 205–226.

Little, A. T. (2016). Communication technology and protest. *The Journal of Politics*, *78*(1), 152–166.

Madestam, A., Shoag, D., Veuger, S., & Yanagizawa-Drott, D. (2013). Do political protests matter? evidence from the tea party movement. *The Quarterly Journal of Economics*, *128*(4), 1633–1685.

Malik, M., Lamba, H., Nakos, C., & Pfeffer, J. (2015). Population bias in geotagged tweets. In *proceedings of the international aaai conference on web and social media* (Vol. 9, pp. 18–27).

Manacorda, M., & Tesei, A. (2020). Liberation technology: Mobile phones and political mobilization in africa. *Econometrica*, *88*(2), 533–567.

Melnikov, N. (2021). Mobile internet and political polarization. *Available at SSRN 3937760*.

Müller, K., & Schwarz, C. (2020). From hashtag to hate crime: Twitter and anti-minority sentiment. *Available at SSRN 3149103*.

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

Oksanen, A., Hawdon, J., Holkeri, E., Näsi, M., & Räsänen, P. (2014). Exposure to online hate among young social media users. In *Soul of society: a focus on the lives of children & youth.* Emerald Group Publishing Limited.

Olea, J. L. M., & Pflueger, C. (2013). A robust test for weak instruments. *Journal of Business & Economic Statistics*, *31*(3), 358–369.

Olson, M. (1965). *The logic of collective action.* Harvard University Press, MA.

Ortiz, I., Burke, S., Berrada, M., & Cortés, H. (2013). World protests 2006-2013. *Initiative for Policy Dialogue and Friedrich-Ebert-Stiftung New York Working Paper*(2013).

Osmundsen, M., Bor, A., Vahlstrup, P. B., Bechmann, A., & Petersen, M. B. (2021). Partisan polarization is the primary psychological motivation behind political fake news sharing on twitter. *American Political Science Review*, *115*(3), 999–1015.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, *37*(2), 187–204.

Passarelli, F., & Tabellini, G. (2017). Emotions and political unrest. *Journal of Political Economy*, *125*(3), 903–946.

Pew Research Center. (2021). *Social media use in 2021.* Technical report.

Quora. (2011). *What is the process involved in launching a start-up at sxsw.* Retrieved from https://www.quora.com/What-is-the-process-involved-in-launching-a-start-up-at-SXSW (Last accessed 21 May 2022)

Raleigh, C., Linke, A., Hegre, H., & Karlsen, J. (2010). Introducing acled: an armed conflict location and event dataset: special data feature. *Journal of Peace Research*, *47*(5), 651–660.

Skoy, E. (2021). Black lives matter protests, fatal police interactions, and crime. *Contemporary Economic Policy*, *39*(2), 280–291.

Snyder Jr, J. M., & Strömberg, D. (2010). Press coverage and political accountability. *Journal of political Economy*, *118*(2), 355–408.

Statista. (2019). *Number of monthly active twitter users worldwide from 1st quarter 2010 to 1st quarter 2019.* Retrieved from https://www.statista.com/

statistics/282087/number-of-monthly-active-twitter-users/ (Last accessed 21 May 2022)

Sunstein, C. R. (2017). # republic. In *# republic.* Princeton University Press.

Theis, M. (2016). *Sxsw economic impact up slightly in 2016; hotel rates hit new high.* Austin Buisness Journal. Retrieved from https://www.bizjournals.com/austin/news/2016/09/07/sxsw-economic-impact-up-slightly-in-2016-hotel.html (Last accessed 21 May 2022)

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151.

Yanagizawa-Drott, D. (2014). Propaganda and conflict: Evidence from the rwandan genocide. *The Quarterly Journal of Economics*, *129*(4), 1947–1994.

Yanagizawa-Drott, D., Rao, A., Petrova, M., Enikolopov, R., & L., B. (2020). Echo chambers: Does online network structure affect political polarization? *Working Paper University of Zurich*.

Zhuravskaya, E., Petrova, M., & Enikolopov, R. (2020). Political effects of the internet and social media. *Annual Review of Economics*, *12*, 415–438.

A Appendix

A.1 Data description and summary stats

Table A1: Description of control variables

Variables	**Description**	**Source**
Demographic controls	Include total population, the share of people in age buckets 20-24, 25-29, 30-34, 35-39, 40-44, 45-49 and 50+, the share of Women, the share of Black or African American, American Indian and Alaska Native, Asian, Native Hawaiian or other Pacific Islander and the share of Hispanic among the total population.	2010 US Census
Socio-economic controls 1	Poverty rates and median household income	2009 US Census Bureau's SAIPE program
Socio-economic controls 2	Share of adult population with at least high school diploma, share of adult population with at least Bachelor degree	2010 American Community Survey
Socio-economic controls 3	Unemployment rate and Employment-to-Population rate in 2010	Bureau of Labor Statistics
Pre-2010 election controls	Share of votes towards the Republican candidate in 2000 (George W. Bush), 2004 (George W. Bush), and 2008 (John McCain)	MIT Election Lab
Election controls	In addition to pre-2010 election controls, these include the share of votes towards the Republican candidate in 2012 (Mitt Romney), 2016 (Donald J. Trump), and 2020 (Donald J. Trump)	MIT Election Lab

Table A2: Description of filtering processes for Broad Social Movement categorization

Category	**Filter by**
Broad Social Movements	Associated actors include one of the following: Black Lives Matter (BLM), Cancel the Rents Movement, Abolish ICE, Blue Lives Matter, Stop Asian Hate, Back the Blue, Occupy Movement, Native Lives Matter (NLM), Save Our Children, Fridays For Future.
Broad Social Movements Alternative definition	Additionally to main definition, events where notes include: "stop the steal" tag or "count every vote" tag

Table A3: Description of filtering processes for topic categorization

Topic	**Filter by**
Coronavirus public health measures	Notes include: "vaccine" *or* "vaccinated" *or* "vaccination" *or* "mask mandate" *or* "mask-mandate" *or* "must wear masks" *or* "coronavirus mandate" *or* "coronavirus prevention measures" *or* "COVID-related restrictions" *or* "COVID restrictions" *or* "Covid restrictions" *or* "public health protocols" *or* "coronavirus protocols" *or* "health protection" *or* "slow the spread" *or* "wearing masks" *or* "mask wearing" *or* "anti-mask" *or* "pro-mask" *or* "universal masking" *or* "coronavirus related safety" *or* "mask policy" *or* ("mask" *and* "coronavirus") *or* "coronavirus pandemic restrictions" *or* "restrictions related to the coronavirus" *or* "contact tracing" *or* "health and safety measures" *or* "coronavirus protection" *or* "coronavirus-related policies" *or* "public health measures" *or* "coronavirus test" *or* "coronavirus precautions" *or* "quarantine" *or* "Quarantine" *or* ("safety protocols" *and* "corona") *or* "coronavirus-related mandates" *or* "stricter measures" *or* (("mandates" *or* "measures") *and* ("corona" *or* "virus" *or* "covid")) *or* "coronavirus restrictions" *or* "policies on coronavirus" *or* "protections against the spread" *or* ("restrictions" *and* ("corona" *or* "pandemic" *or* "covid" *or* "COVID")) *or* "in-person learning" *or* "remote learning" *or* "Reopen Our Cities" *or* "reopen businesses" *or* "reopening businesses" *or* "business closure" *or* "school closure" *or* ("lockdown" *and* "corona") *or* "reopening schools" *or* "reopen schools" *or* "in-person school" *or* "reopening of schools" *or* "in-person classes" *or* "reopen bars" *or* "reopening of bars" *or* "closing of bars" *or* "restaurant closure" *or* ("stay-at-home order" *and* "corona")
Stop the Steal	Notes include: "stop the steal" tag *or* "count every vote" tag
Abortion rights	Notes include: "abortion" *or* "pro-life" *or* "Planned Parenthood" *or* "planned parenthood" *or* "reproductive rights" *Or* Associated actors include "PP: Planned Parenthood"
Racial Issues	Associated actors include: *or* "BLM: Black Lives Matter" *or* "Back the Blue" *or* "Blue Lives Matter" *or* "Antifa" *or* "BU: Black Unity" *or* "RBPP: Revolutionary Black Panther Party" *or* "Gastonia Watchmen" *or* "West Virginia 34 Mountain Milita" *or* "NFAC: Not Fucking Around Coalition" *or* "Expect US" *or* "NOI: Nation of Islam" *or* "TPR: The People's Revolution" *or* "Stop Asian Hate" *or* "Take 'Em Down" *or* "BVM: Black Voters Matter" *or* "NAACP: National Association for the Advancement of Colored People" *or* "NBPP: New Black Panther Party" *or* "Black Panthers" *or* "Refuse Fascism" *or* "BLM757" *or* "White Nationalists" *or* "Boogaloo Boys" *or* "Proud Boys" *or* "Groypers" *or* "White Defence Force" *or* "Aryan Nations" *or* "ABC: Aryan Cowboys Brotherhood" *or* "QAnon" *or* "Super Happy Fun America" *or* "GDL: Goyim Defense League" *or* "NSC: Nationalist Social Club" *or* "KKK: Ku Klux Klan" Fun America" *or* "Minnesota Patriot Alliance" *or* "NSM: National Socialist Movement" *or* "Patriot Front" *or* ("Washington State III%ers" *or* "Patriot Prayer" *and* not "coronavirus" or "gun regulations" in notes) *or* "UDAF: United American Defense Force"
White supremacists	Associated actors include: "White Nationalists" *or* "Boogaloo Boys" *or* "Proud Boys" *or* "Groypers" *or* "White Defence Force" *or* "Aryan Nations" *or* "ABC: Aryan Cowboys Brotherhood" *or* "QAnon" *or* "Super Happy Fun America" *or* "GDL: Goyim Defense League" *or* "NSC: Nationalist Social Club" *or* "KKK: Ku Klux Klan" Fun America" *or* "Minnesota Patriot Alliance" *or* "NSM: National Socialist Movement" *or* "Patriot Front" *or* ("Washington State III%ers" *or* "Patriot Prayer" *and* not ("coronavirus" or "gun regulations") in notes) *or* "UDAF: United American Defense Force"

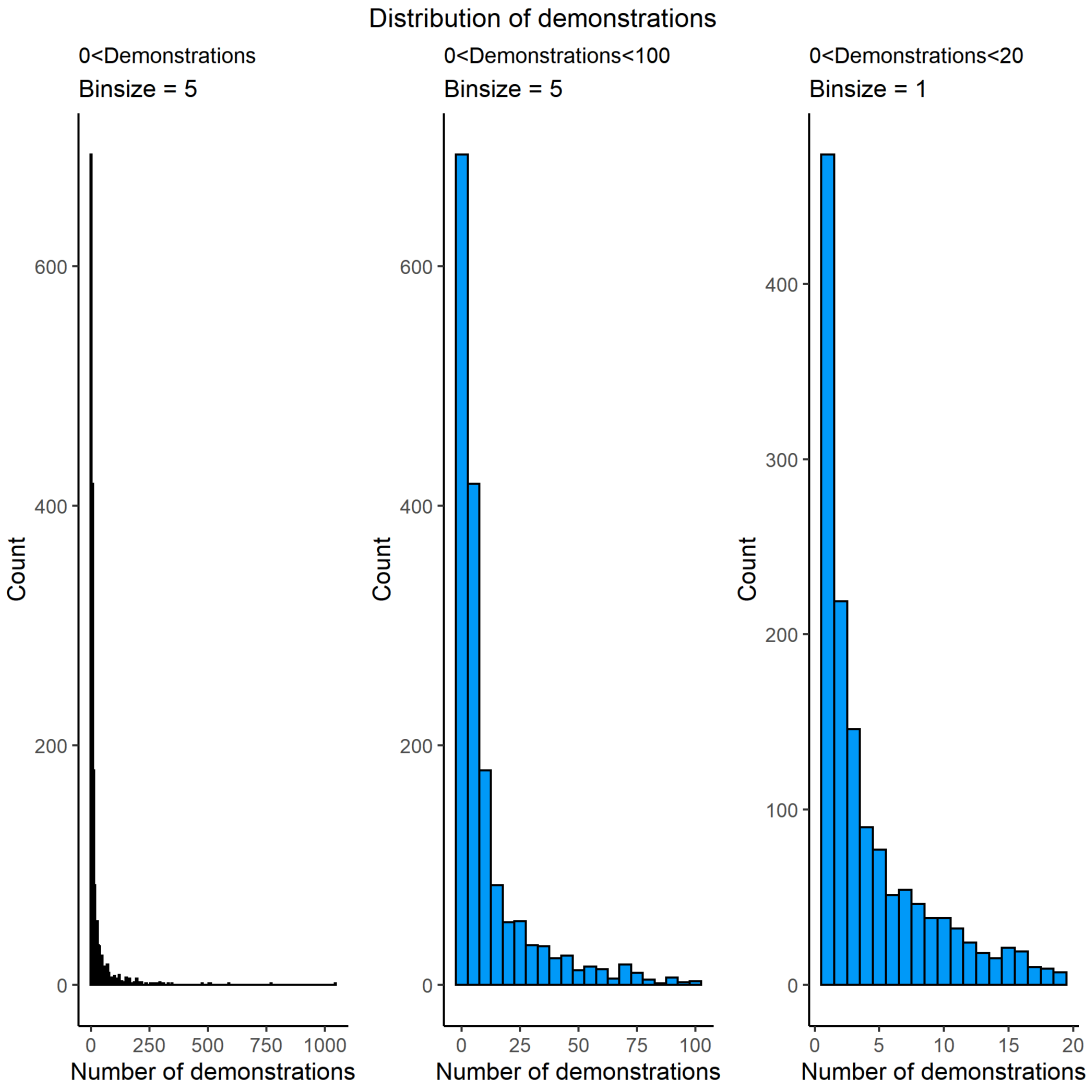


Figure A1: Distribution of demonstrations

Note: The figures show the distribution of demonstrations. The panel to the left shows the distribution among counties with at least 1 demonstration. The middle panel shows counties with at least 1, but less than 100 demonstrations. The right panel shows counties with at least 1, but less than 20 demonstrations.

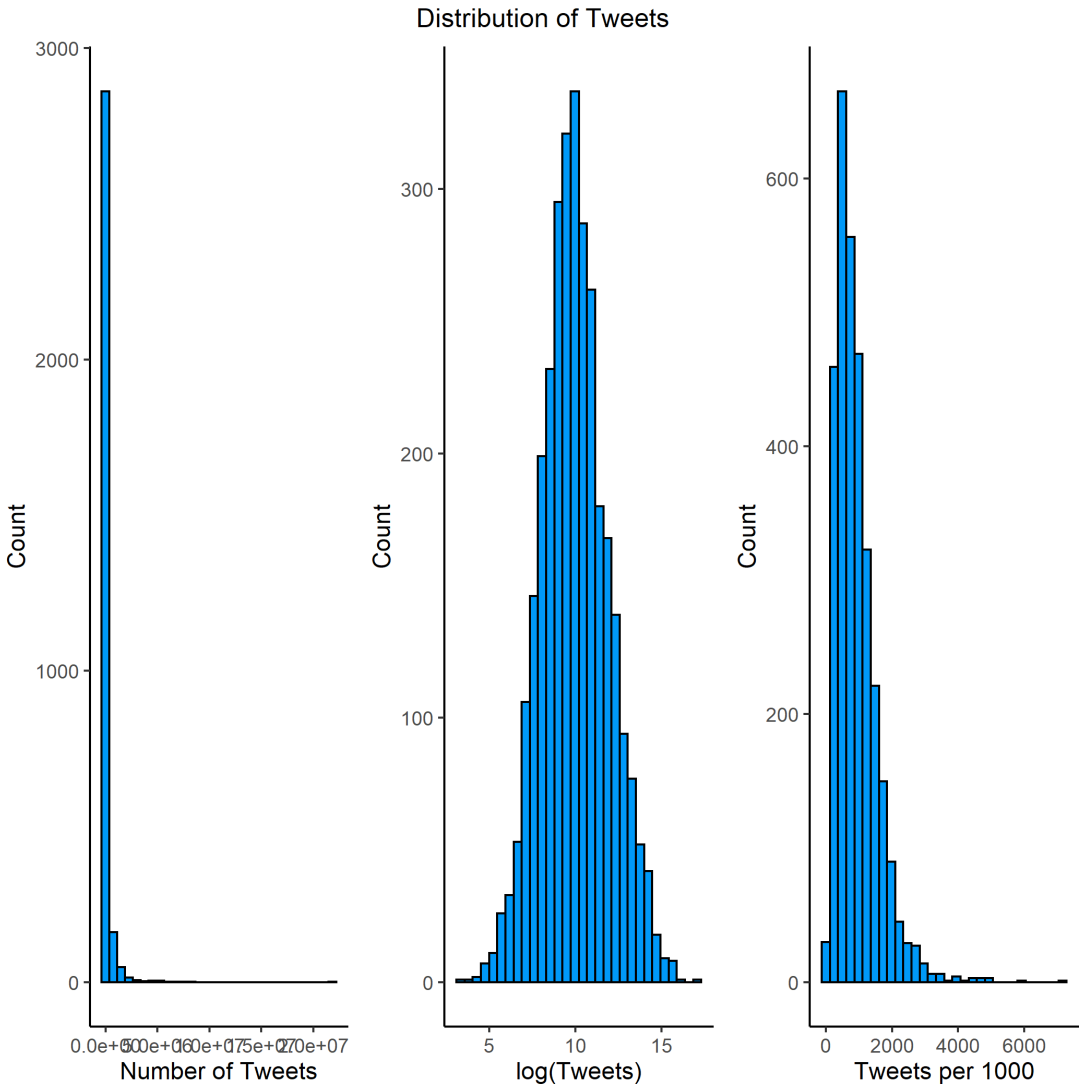


Figure A3: Distribution of tweets

Note: The figures show the distribution of Tweets in our dataset. Tweets per 1000 refers to Tweets per 1000 inhabitants in 2019

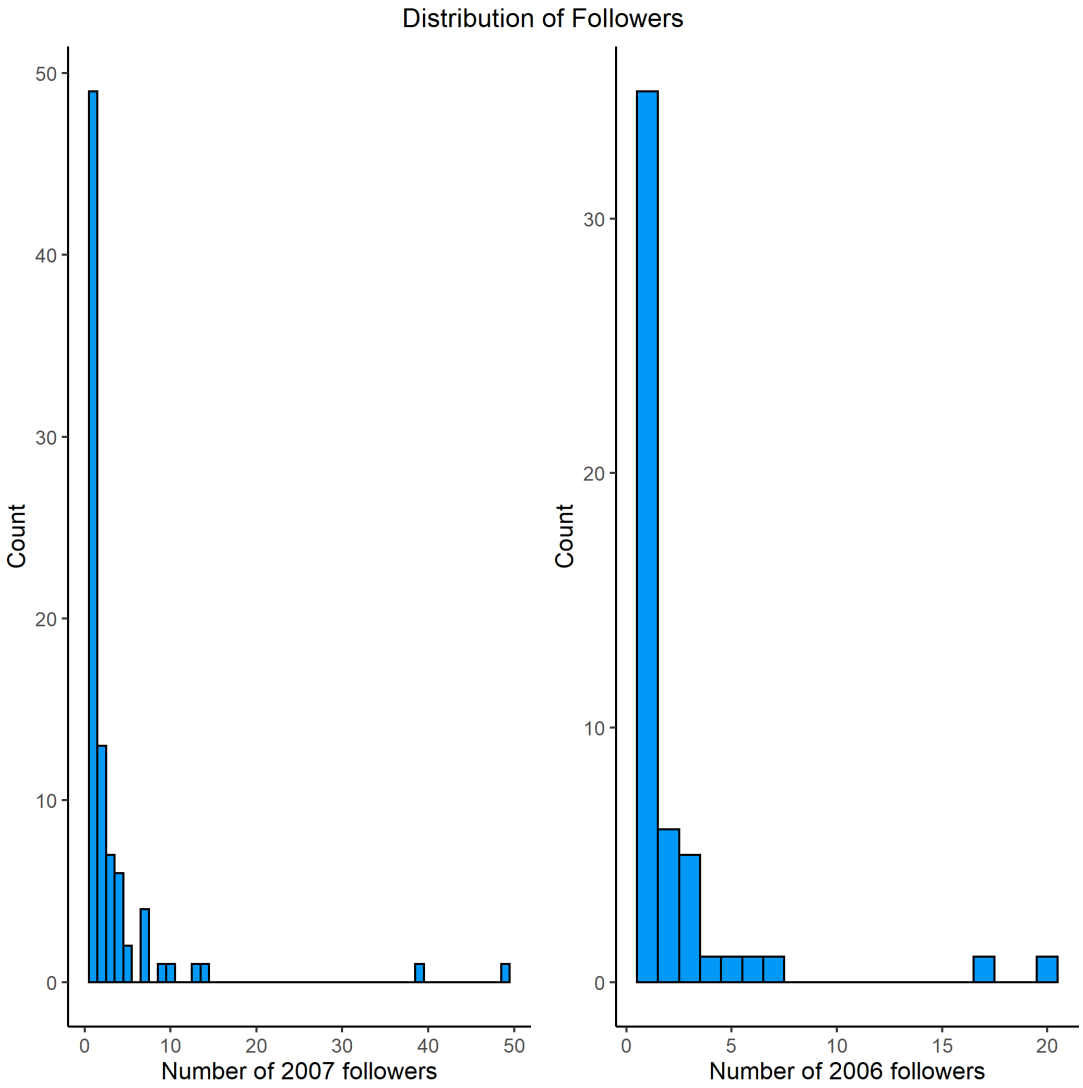


Figure A5: Distribution of SXSW followers

Note: The figures show the distribution of Twitter users following the account of SXSW, who created their accounts in March 2007 (left panel) or in 2006 (right panel). The figures only shows counties with at least 1 follower.

A.2 Empirical strategy tests

Table A4: Instrument correlation with "pre-treatment" county characteristics, conditional on SXSW 2006 followers

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Dependent variables:						
	Share 20-24yo	Share 24-29yo	Share 30-34yo	Share 34-39yo	Share 40-44yo	Share 44-49yo	Share 50 or older
SXSW March 2007	−0.0004	0.0003	0.0005	0.001	0.0004**	0.00002	−0.003
	(0.001)	(0.001)	(0.001)	(0.0003)	(0.0002)	(0.0002)	(0.002)
SXSW 2006	0.004	0.003	0.002	0.001	0.0003	−0.0003	−0.004
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.006)
	Share Women	Share White	Share Black	Share Native Am.	Share Asian	Share Pac. Isl.	Share Hispanic
SXSW March 2007	0.001	−0.006	0.003	−0.001	0.003***	0.0001**	0.010**
	(0.0005)	(0.006)	(0.006)	(0.001)	(0.001)	(0.0001)	(0.004)
SXSW 2006	−0.0004	−0.012	−0.002	0.002	0.011***	0.00003	−0.002
	(0.001)	(0.012)	(0.012)	(0.002)	(0.002)	(0.0002)	(0.014)
	Poverty rate	log(Med. hh. Inc.)	Share High school	Share Bachelor	Unempl.	Empl. to pop	Rep. Share 2000
SXSW March 2007	−0.149	0.015**	−0.021	0.402	0.001**	−0.002	−0.004
	(0.107)	(0.007)	(0.176)	(0.520)	(0.0004)	(0.002)	(0.007)
SXSW 2006	−0.024	0.014	0.288	1.751	−0.002*	0.008**	−0.023
	(0.274)	(0.022)	(0.331)	(1.237)	(0.001)	(0.003)	(0.015)
	Rep. Share 2004	Rep. Share 2008	Population				
SXSW March 2007	−0.004	−0.005	160,661.700***				
	(0.007)	(0.007)	(59,205.810)				
SXSW 2006	−0.027	−0.026	−20,853.870				
	(0.016)	(0.017)	(182,631.000)				

Note: This table represents county-level regressions. SXSW March2007 is the number of Twitter users who joined Twitter in March 2007 and follow South by Southwest (@SXSW). SXSW 2006 is the number of Twitter users who joined in 2006 and follow SXSW. No additional controls are included in these regressions. Robust errors in parentheses are clustered by state. *p<0.1; **p<0.05; ***p<0.01

Table A5: Instrument correlation with "pre-treatment" county characteristics, conditional on population

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Dependent variables:						
	Share 20-24yo	Share 24-29yo	Share 30-34yo	Share 34-39yo	Share 40-44yo	Share 44-49yo	Share 50 or older
SXSW March 2007	−0.001	−0.0003	−0.00000	0.0001	0.0001	−0.00001	0.0001
	(0.001)	(0.001)	(0.0005)	(0.0002)	(0.0002)	(0.0003)	(0.001)
SXSW 2006	0.003	0.003*	0.001	0.001	0.0001	−0.0003	−0.002
	(0.002)	(0.002)	(0.001)	(0.0005)	(0.0004)	(0.001)	(0.003)
	Share Women	Share White	Share Black	Share Native Am.	Share Asian	Share Pac. Isl.	Share Hispanic
SXSW March 2007	0.00001	−0.0003	0.0001	−0.001	0.001	0.0001	0.005
	(0.0003)	(0.005)	(0.005)	(0.001)	(0.001)	(0.0001)	(0.004)
SXSW 2006	−0.001	−0.008	−0.005	0.002	0.010***	−0.00000	−0.005
	(0.001)	(0.011)	(0.012)	(0.002)	(0.003)	(0.0002)	(0.010)
	Poverty rate	log(Med. hh. Inc.)	Share High school	Share Bachelor	Unempl.	Empl. to pop	Rep. Share 2000
SXSW March 2007	−0.022	0.002	−0.139	−0.190	0.001*	−0.002	0.002
	(0.126)	(0.005)	(0.115)	(0.303)	(0.0004)	(0.001)	(0.006)
SXSW 2006	0.078	0.006	0.139	1.306**	−0.002*	0.007**	−0.020*
	(0.296)	(0.011)	(0.300)	(0.658)	(0.001)	(0.003)	(0.011)
	Rep. Share 2004	Rep. Share 2008					
SXSW March 2007	0.003	0.002					
	(0.006)	(0.005)					
SXSW 2006	−0.023*	−0.021*					
	(0.012)	(0.011)					

Note: This table represents county-level regressions. SXSW March2007 is the number of Twitter users who joined Twitter in March 2007 and follow South by Southwest (@SXSW). SXSW 2006 is the number of Twitter users who joined in 2006 and follow SXSW. Additionally, all regressions control for population via 2-percentile dummies. Robust errors in parentheses are clustered by state. *p<0.1; **p<0.05; ***p<0.01

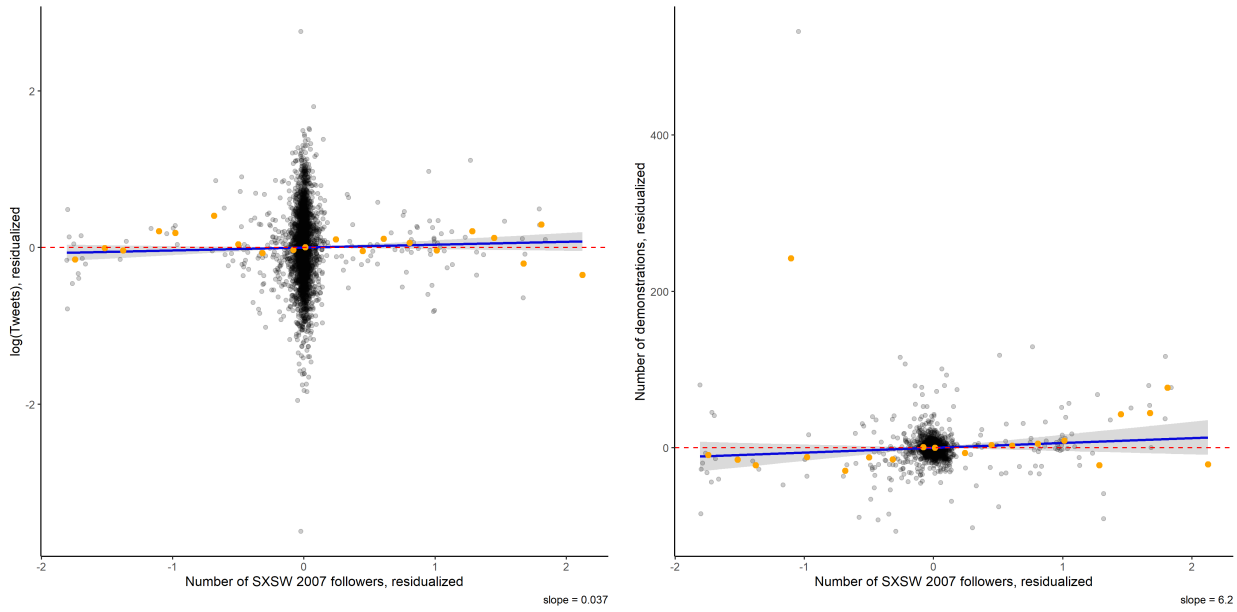


Figure A7: First stage and reduced-form visualization, without SXSW 2007 residual outliers

Note: This figure presents binned scatter plots of the relationship between county-level the number of SXSW followers who joined Twitter in March 2007, and a) Twitter users in 2014-2015 (left panel) or b) the number of demonstrations between Jan 2020 and Nov 2021 (right panel). Variables are residualized by partialling out SXSW followers who joined in 2006, population 2-percentiles, demographic, socio-economic, and pre-2010 election controls (see description of control variables in Table A1 in the Appendix). The left hand side figure shows the full sample, while the right hand side figure shows a subsample excluding outier counties with abnormally high or low values for the SXSW followers who joined Twitter in March 2007 residual. Grey dot's represent individual observations and yellow dots represent average values of both variables within a bin. The blue line represents a line of best fit using the unbinned data, and the shaded blue area are 95% confidence intervals calculated using robust standard errors clustered at the state-level.

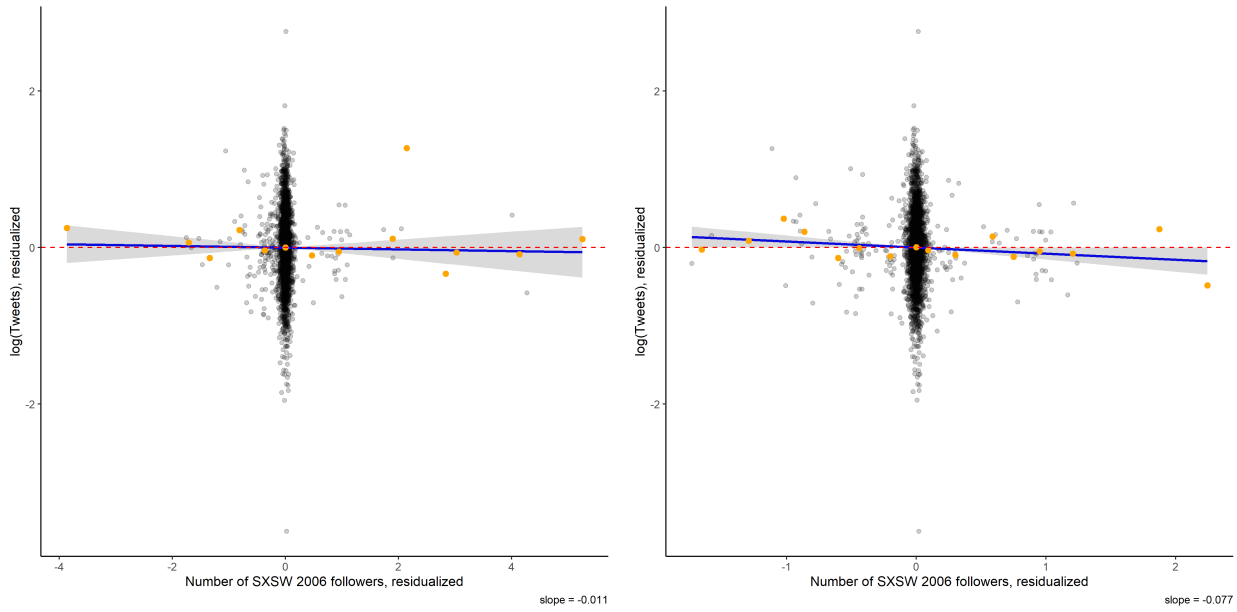


Figure A9: Residual plots for SXSW 2006 followers and log(Tweets)

Note: This figure presents binned scatter plots of the relationship between county-level Twitter users in 2014-2015 and the number of SXSW followers who joined Twitter in 2006. Variables are residualized by partialling out SXSW followers who joined in March 2007, population 2-percentiles, demographic, socio-economic, and pre-2010 election controls (see description of control variables in Table A1 in the Appendix). The left hand side figure shows the full sample, while the right hand side figure shows a subsample excluding outier counties with abnormally high or low values for the SXSW followers who joined Twitter in March 2007 residual. Grey dot's represent individual observations and yellow dots represent average values of both variables within a bin. The blue line represents a line of best fit using the unbinned data, and the shaded blue area are 95% confidence intervals calculated using robust standard errors clustered at the state-level.

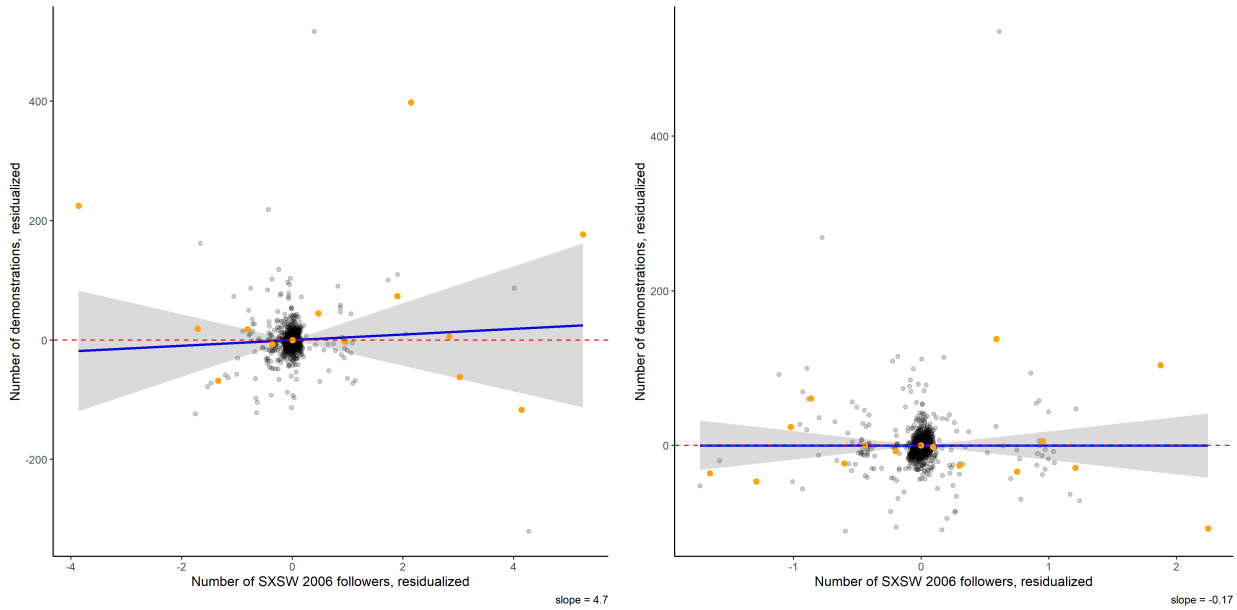


Figure A11: Residual plots for SXSW 2006 followers and demonstrations

Note: This figure presents binned scatter plots of the relationship between county-level number of demonstrations between Jan 2020 and Nov 2021, and the number of SXSW followers who joined Twitter in 2006. Variables are residualized by partialling out SXSW followers who joined in March 2007, population 2-percentiles, demographic, socio-economic, and pre-2010 election controls (see description of control variables in Table A1 in the Appendix). The left hand side figure shows the full sample, while the right hand side figure shows a subsample excluding outier counties with abnormally high or low values for the SXSW followers who joined Twitter in March 2007 residual. Grey dot's represent individual observations and yellow dots represent average values of both variables within a bin. The blue line represents a line of best fit using the unbinned data, and the shaded blue area are 95% confidence intervals calculated using robust standard errors clustered at the state-level.

A.3 Robustness for heterogeneity analyses

Table A6: Alternative definition of "hard to coordinate"

	(1)	(2)	(3)	(4)	(5)
	FS	Broad Social Movements	Broad Social Movements only	Broad Social Movements only + No Org.	No Org.
Specification:	First Stage	Second Stage			
Dep var:	Log(Tweets)	*Share among all demonstrations*			
SXSW March 2007	0.050***				
	(0.011)				
log(Tweets)		0.125*	0.070	0.086	0.016
		(0.073)	(0.081)	(0.102)	(0.061)
SXSW 2006	−0.004	−0.014*	−0.013	−0.015	−0.002
	(0.032)	(0.008)	(0.009)	(0.011)	(0.008)
R^2	0.946				
Cluster robust F-stat	18.19				
Mean dep. var.		0.49	0.36	0.59	0.23
N	1,782	1,782	1,782	1,782	1,782

Note: This table presents county-level regressions. Log(Tweets) is the natural logarithm of a sample of tweets from 2014-2015. SXSW March2007 is the number of Twitter users who joined Twitter in March 2007 and follow South by Southwest (@SXSW). SXSW 2006 is the number of Twitter users who joined in 2006 and follow SXSW. Column (1) presents first stage regressions, while columns (2) to (5) present second stages. "Broad Social Movements only" refers to demonstrations where there the only associated actor is a "Broad Social Movement." "No Org." refers to demonstrations that cannot be associated with a group or an organization. The table presents TSLS results on a subsample of counties where at least 1 demonstration event occurred between Jan 2020 and Nov 2021, where the dependent variable is the share of demonstrations that fit the category. All regressions control for population via 2-percentile dummies, and the full set of controls (as in column 5 in Table 1), see Table A1 in the Appendix for their descriptions. Robust standard errors in parentheses are clustered by state.
*p<0.1; **p<0.05; ***p<0.01

Table A7: Alternative definition of "Broad Social Movements"

	(1)	(2)	(3)	(4)	(5)
	FS	Broad Social Movements	Broad Social Movements only	Broad Social Movements only + No Org.	No Org.
Specification:	First Stage	Second Stage			
Dep var:	Log(Tweets)	*Share among all demonstrations*			
SXSW March 2007	0.050***				
	(0.011)				
log(Tweets)		0.144*	0.084	0.100	0.160**
		(0.076)	(0.085)	(0.106)	(0.076)
SXSW 2006	−0.004	−0.017*	−0.015*	−0.017	−0.018**
	(0.032)	(0.009)	(0.009)	(0.011)	(0.008)
R^2	0.946				
Cluster robust F-stat	18.19				
Mean dep. var.		0.50	0.37	0.60	0.73
N	1,782	1,782	1,782	1,782	1,782

Note: This table presents county-level regressions. Log(Tweets) is the natural logarithm of a sample of tweets from 2014-2015. SXSW March2007 is the number of Twitter users who joined Twitter in March 2007 and follow South by Southwest (@SXSW). SXSW 2006 is the number of Twitter users who joined in 2006 and follow SXSW. Column (1) presents first stage regressions, while columns (2) to (5) present second stages. "Broad Social Movements only" refers to demonstrations where the only associated actor is a "Broad Social Movement." Here I use an alternative definition of Broad Social Movements, see see Table A2 for the definitions. "No Org." refers to demonstrations that cannot be associated with a group or an organization. The table presents TSLS results on a subsample of counties where at least 1 demonstration event occurred between Jan 2020 and Nov 2021, where the dependent variable is the share of demonstrations that fit the category. All regressions control population via 2-percentile dummies, and the full set of controls (as in column 5 in Table 1), see Table A1 in the Appendix for their descriptions. Robust standard errors in parentheses are clustered by state.
*p<0.1; **p<0.05; ***p<0.01

Table A8: Racial or polarizing issues, excluding BLM

	(1)	(2)	(3)	(4)	(5)
Type of demonstration:		Racial Issues	"Polarizing Topics"	Racial Issues Excl. BLM	"Polarizing Topics" Excl. BLM
Specification:	First Stage	Second Stage			
Dependent variables:	Log(Tweets)	*Share among all demonstrations*			
SXSW March 2007	0.050***				
	(0.011)				
log(Tweets)		0.125*	0.124	−0.003	0.014
		(0.073)	(0.082)	(0.003)	(0.043)
SXSW 2006	−0.004	−0.014	−0.017*	0.0001	−0.003
	(0.032)	(0.009)	(0.009)	(0.001)	(0.005)
R^2	0.946				
F-stat	18.19				
Mean dep. var.		0.495	0.622	0.017	0.140
N	1,782	1,782	1,782	1,782	1,782

Note: This table presents county-level regressions. Log(Tweets) is the natural logarithm of a sample of tweets from 2014-2015. SXSW March2007 is the number of Twitter users who joined Twitter in March 2007 and follow South by Southwest (@SXSW). SXSW 2006 is the number of Twitter users who joined in 2006 and follow SXSW. Column (1) presents first stage regressions, while columns (2)-(5) present second stages. Racial issues refers to demonstrations concerning racial tensions, while "Polarizing topics" refers to all topic categories presented in Table A3. Excl. BLM means that we do not count events where the Broad Social Movement "Black Lives Matter" is reccorded as an associated actor. The table presents TSLS results on a subsample of counties where at least 1 demonstration event occurred between Jan 2020 and Nov 2021, where the dependent variable is the share of demonstrations that fit the category. All regressions control for population via 2-percentile dummies, and the full set of controls (as in column 5 in Table 1), see Table A1 in the Appendix for their descriptions. Robust standard errors in parentheses are clustered by state.
*p<0.1; **p<0.05; ***p<0.01

Table A9: Share of topics relative to "non-polarizing topics"

	(1)	(2)	(3)	(4)	(5)
Type of demonstration:	Coronavirus	"Stop the Steal"	Abortion rights	Racial issues	White Supremacist
Panel A: First stage					
Dependent variable:			*Log(Tweets)*		
SXSW March 2007	0.048***	0.050***	0.049***	0.050***	0.049***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
SXSW 2006	−0.001	−0.004	−0.003	−0.003	−0.003
	(0.034)	(0.034)	(0.035)	(0.033)	(0.034)
R^2	0.949	0.950	0.950	0.946	0.950
F-stat	16.77	18.21	17.91	17.86	17.93
Panel B: Second stage					
Dependent variable:			*Share relative to "non-polarizing topics"*		
log(Tweets)	−0.020	0.047	0.053	0.128	0.002
	(0.069)	(0.033)	(0.040)	(0.080)	(0.020)
SXSW 2006	0.0001	−0.005	−0.004	−0.014	−0.001
	(0.009)	(0.004)	(0.005)	(0.009)	(0.002)
Mean dep. var.	0.206	0.033	0.058	0.565	0.018
N	1,353	1,270	1,283	1,712	1,266

Note: This table presents county-level regressions. Log(Tweets) is the natural logarithm of a sample of tweets from 2014-2015. SXSW March2007 is the number of Twitter users who joined Twitter in March 2007 and follow South by Southwest (@SXSW). SXSW 2006 is the number of Twitter users who joined in 2006 and follow SXSW. Coronavirus refers to demonstrations related to public health measures due to the Covid pandemic. "Stop the Steal" refers to demonstrations related to the 2020 post-presidential election unrest (including those calling for results to be respected). Abortion rights refers to demonstrations related to reproductive rights. Racial issues refers to demonstrations concerning racial tensions while white supremacists are a subset of those where demonstrating groups have known affiliations with white supremacy ideology. A detailed description of topic categorization is presented in Table A3. Each column shows regression results using a sample of counties where at least 1 demonstration that fits none of the topic categories in the table occurred, or where a demonstration in the topic category of that column occurred. Panel A shows first stage regressions, while Panel B shows second stages where the dependent variable is the number of demonstrations of a given topic category, divided by the sum of demonstrations fitting the category and demonstrations fitting none of the categories in the table. All regressions control for population via 2-percentile dummies, and the full set of controls (as in column 5 in Table 1), see Table A1 in the Appendix for their descriptions. Robust standard errors in parentheses are clustered by state. *p<0.1; **p<0.05; ***p<0.01

Table A10: Share of non-peaceful demonstrations relative to peaceful demonstrations

	(1)	(2)
Type of demonstrations:	Violent Demonstrators	Peaceful protesters met with intervention or violence
Panel A: First stage		
Dependent variable:	*log(Tweets)*	
SXSW March 2007	0.050***	0.050***
	(0.011)	(0.011)
SXSW 2006	−0.004	−0.003
	(0.032)	(0.032)
R^2	0.947	0.946
Cluster robust F-stat	18.44	18.38
Panel B: Second stage		
Dependent variable:	*Share relative to peaceful protests*	
log(Tweets)	0.071	0.038
	(0.061)	(0.028)
SXSW 2006	−0.007	−0.003
	(0.007)	(0.003)
Mean dep. var.	0.012	0.023
N	1,744	1,750

Note: This table presents county-level regressions. Log(Tweets) is the natural logarithm of a sample of tweets from 2014-2015. SXSW March2007 is the number of Twitter users who joined Twitter in March 2007 and follow South by Southwest (@SXSW). SXSW 2006 is the number of Twitter users who joined in 2006 and follow SXSW. Column (1) looks at events where demonstrators engage in violent or disruptive behaviour while column (2) looks at demonstration events where demonstrators are peaceful but there is an attempt of suppression or the use of violence by outside parties. Each column is estimated shows regression results using a sample of counties where at least 1 peaceful demonstration, or a demonstration fitting the column category occurred. Panel A shows first stage regressions, while Panel B shows second stages where the dependent variable is the number of demonstrations of a given column category divided by the sum of peaceful demonstrations and demonstrations fitting the category. All regressions control for population via 2-percentile dummies, and the full set of controls (as in column (5) in Table 1), see Table A1 in the Appendix for their descriptions. Robust standard errors in parentheses are clustered by state.
*p<0.1; **p<0.05; ***p<0.01

Table A11: TSLS demonstrations by sub-event

	(1)	(2)	(3)	(4)	(5)
	FS	Intervention	Excessive Force	Violent Dem	Mob Violence
Panel A					
Depenent variable:	Log(Tweets)	*Number of demonstrations*			
SXSW March 2007	0.050***				
	(0.011)				
log(Tweets)		17.547***	3.952**	34.954	2.552***
		(5.524)	(1.912)	(25.407)	(0.817)
SXSW 2006	−0.011	−0.267	−0.148	−1.914	0.003
	(0.032)	(0.566)	(0.196)	(2.508)	(0.078)
R^2	0.943				
Cluster robust F-stat	19.31				
Mean dep. var		0.32	0.05	0.30	0.03
N	3,105	3,105	3,105	3,105	3,105
Panel B					
Depenent variable:	Log(Tweets)	*Share among all demonstrations*			
SXSW March 2007	0.050***				
	(0.011)				
log(Tweets)		0.030	0.001	0.064	−0.002
		(0.021)	(0.004)	(0.050)	(0.005)
SXSW 2006	−0.004	−0.002	−0.0002	−0.006	0.0002
	(0.032)	(0.003)	(0.0004)	(0.005)	(0.001)
R^2	0.946				
Cluster robust F-stat	18.19				
Mean dep. var.		0.02	0.003	0.01	0.003
N	1,782	1,782	1,782	1,782	1,782

Note: This table presents county-level regressions. Log(Tweets) is the natural logarithm of a sample of tweets from 2014-2015. SXSW March2007 is the number of Twitter users who joined Twitter in March 2007 and follow South by Southwest (@SXSW). SXSW 2006 is the number of Twitter users who joined in 2006 and follow SXSW. Column (1) presents first stage regressions, while columns (2)-(5) present second stages. "Intervention" refers to events where demonstrators are peaceful, but there is an attempt to disperse or suppress the protest without serious/lethal injuries being reported. "Excessive force" refers to events where peaceful protesters are targeted with violence leading to (or could lead to) serious/lethal injuries. "Violent Dem." refers to events where demonstrators themselves engage in disruptive or violent behaviour. "Mob violence" refers to events where rioters interact violently with other rioters, another armed group or civilians, outside of demonstrations and without the *use* of lethal weapons. Panel A presents TSLS results where the number of demonstrations fitting the given column category is the dependent variable. Panel B presents TSLS results on a subsample of counties where at least 1 demonstration event occurred between Jan 2020 and Nov 2021, where the dependent variable is the share of demonstration that fit the column-category. All regressions control for population via 2-percentile dummies, and the full set of controls (as in column (5) in Table 1), see Table A1 in the Appendix for their descriptions. Robust standard errors in parentheses are clustered by state. *p<0.1; **p<0.05; ***p<0.01

Table A12: Share of demonstrations relative to peaceful protests by intervening actor

	(1)	(2)
	Intervention by Gov.	Intervention by non-Gov. Actor
Panel A: First stage		
Dependent variable:	*log(Tweets)*	
SXSW March 2007	0.050***	0.050***
	(0.011)	(0.011)
SXSW 2006	−0.003	−0.005
	(0.032)	(0.032)
R^2	0.946	0.947
Cluster robust F-stat	18.38	18.63
Panel B: Second stage		
Dependent variable:	*Share of demonstrations where peaceful protesters are met with intervention or violence, relative to peaceful protests*	
log(Tweets)	0.030	0.009
	(0.026)	(0.006)
SXSW 2006	−0.002	−0.001
	(0.003)	(0.001)
Mean dep. var.	0.020	0.003
N	1,750	1,739

Note: This table presents county-level regressions. Log(Tweets) is the natural logarithm of a sample of tweets from 2014-2015. SXSW March2007 is the number of Twitter users who joined Twitter in March 2007 and follow South by Southwest (@SXSW). SXSW 2006 is the number of Twitter users who joined in 2006 and follow SXSW. Column (1) looks at events where peaceful protesters are met with an attempt of suppression or the use of violence by government entities, while column (2) looks at demonstrations where peaceful protesters are met with an attempt of suppression or the use of violence by non-government actors. Each column is estimated on a sample of counties where at least 1 peaceful demonstration, or a demonstration fitting the column category occurred. Panel A shows first stage regressions, while Panel B shows second stages where the dependent variable is the number of demonstrations of a given column category divided by the sum of peaceful demonstrations and demonstrations fitting the category. All regressions control for population via 2-percentile dummies, and the full set of controls (as in column 5 in Table 1), see Table A1 in the Appendix for their descriptions. Robust standard errors in parentheses are clustered by state.
*p<0.1; **p<0.05; ***p<0.01

Table A13: Non-government groups and organizations associated with attempts at suppression or the use of violence against peaceful protesters

Category	**Filter by**
Groups:	Students, Teachers, Christian Group, Proud Boys, White Nationalists, American Guard, BLM: Black Lives Matter, Labour Group, Civilians, UCA: Utah Citizens' Alarm, Three Percenters (III%), Latinx Group, Protesters, Blue Lives Matter, Pro-Police Group, Unidentified Communal Militia, Back the Blue, Aryan Brotherhood, Oath Keepers, Keystone United, ADS: American, Defense Skins, KKK: Ku Klux Klan.

Table A14: TSLS demonstrations by intervening actor

	(1)	(2)	(3)
Intervening actor:		Intervention by non-Gov. Sole Perpetrator	Intervention by non-Gov. Group
Specification:	First Stage	Second Stage	
Dependent variables:	Log(Tweets)	*Share of all demonstrations where peaceful protesters are met with intervention or violence*	
SXSW March 2007	0.050***		
	(0.011)		
log(Tweets)		−0.003	0.009***
		(0.003)	(0.003)
SXSW 2006	−0.004	0.0001	−0.001*
	(0.032)	0.0001	(0.0005)
R^2	0.946		
Cluster robust F-stat	18.19		
Mean dep. var.		0.019	0.003
N	1,782	1,782	1,782

Note: This table presents county-level regressions. Log(Tweets) is the natural logarithm of a sample of tweets from 2014-2015. SXSW March2007 is the number of Twitter users who joined Twitter in March 2007 and follow South by Southwest (@SXSW). SXSW 2006 is the number of Twitter users who joined in 2006 and follow SXSW. Column (1) presents first stage regressions, while columns (2) and (3) present second stages. Column (2) looks at demonstrations where peaceful protesters are met with an attempt of suppression or the use of violence by a sole non-government perpetrator, while column (3) looks at demonstrators where peaceful protesters are met with an an attempt of suppression or the use of violence by non-governmental groups. The table presents TSLS results on a subsample of counties where at least 1 demonstration event occurred between Jan 2020 and Nov 2021, where the dependent variable is the share of demonstration that fit the column-category. All regressions control for population via 2-percentile dummies, and the full set of controls (as in column (5) in Table 1), see Table A1 in the Appendix for their descriptions. Robust standard errors in parentheses are clustered by state. *p<0.1; **p<0.05; ***p<0.01

A.4 Main results replicated with additional election controls

Table A15: Reduced-form analysis with additional election controls

	(1)	(2)	(3)	(4)	(5)
	Dependent variable: Number of demonstrations				
SXSW March2007	16.849***	12.356**	13.327**	13.395**	13.452**
	(5.777)	(5.506)	(5.283)	(5.254)	(5.240)
SXSW 2006	7.295	6.201	4.646	4.764	4.645
	(19.552)	(14.846)	(13.690)	(13.646)	(13.573)
State FE	Yes	Yes	Yes	Yes	Yes
Population 2-percentile controls		Yes	Yes	Yes	Yes
Demographic controls			Yes	Yes	Yes
Socio-economic controls				Yes	Yes
Election controls					Yes
R^2	0.482	0.736	0.762	0.762	0.754
Mean of dependent variable	11.00	11.00	11.00	11.00	11.00
Observations	3,108	3,107	3,107	3,106	3,103

Note: This table presents county-level regressions where the natural logarithm of a sample of Tweets from 2014-2015 is the dependent variable. SXSW March2007 is the number of Twitter users who joined Twitter in March 2007 and follow South by Southwest (@SXSW). SXSW 2006 is the number of Twitter users who joined in 2006 and follow SXSW. Regressions include the indicated control variables, including election controls for 2012, 2016 and 2020 (see Table A1 in the Appendix for their descriptions). Robust standard errors in parentheses are clustered by state. *p<0.1; **p<0.05; ***p<0.01

Table A16: Effect of Twitter on demonstrations with additional election controls

	(1)	(2)	(3)	(4)	(5)
Panel A: OLS	*Dependent variable: Number of demonstrations*				
Log(Tweets)	8.960***	5.265***	3.025**	3.114**	3.052**
	(1.264)	(1.475)	(1.337)	(1.426)	(1.450)
R^2	0.549	0.715	0.733	0.736	0.724
Panel B: TSLS First stage	*Dependent variable: Log(Tweets)*				
SXSW March2007	0.168*	0.032**	0.049***	0.051***	0.049***
	(0.101)	(0.015)	(0.011)	(0.011)	(0.011)
SXSW 2006	0.114	0.037	−0.021	−0.008	−0.001
	(0.257)	(0.041)	(0.033)	(0.033)	(0.033)
R^2	0.326	0.923	0.937	0.942	0.943
Cluster robust F-stat	2.31	4.47	17.83	19.49	18.28
Panel C: TSLS Second Stage	*Dependent variable: Number of demonstrations*				
log(Tweets)	100.236*	403.806***	280.442***	273.212***	277.195***
	(54.181)	(121.798)	(91.954)	(88.054)	(87.283)
SXSW 2006	−4.092	−8.928	10.493	6.734	4.871
	(19.368)	(11.877)	(8.645)	(8.847)	(8.795)
State FE	Yes	Yes	Yes	Yes	Yes
Population 2-percentile controls		Yes	Yes	Yes	Yes
Demographic controls			Yes	Yes	Yes
Socio-economic controls				Yes	Yes
Election controls					Yes
Mean number of demonstrations	11.00	11.00	11.00	11.00	11.00
Observations	3,108	3,107	3,107	3,106	3,103

Note: This table presents county-level regressions for an OLS and a TSLS model of the effect of Twitter on demonstration frequency. Log(Tweets) is the natural logarithm of a sample of tweets from 2014-2015. SXSW March2007 is the number of Twitter users who joined Twitter in March 2007 and follow South by Southwest (@SXSW). SXSW 2006 is the number of Twitter users who joined in 2006 and follow SXSW. Regressions include the indicated control variables, including election controls for 2012, 2016 and 2020 (see Table A1 in the Appendix for their descriptions). Robust standard errors in parentheses are clustered by state. *p<0.1; **p<0.05; ***p<0.01

Table A17: TSLS demonstrations by coordination complexity with additional election controls

	(1)	(2)	(3)
Type of demonstrations:		Broad Social Movements + No Org.	Other demonstrations
Specification:	First Stage	Second Stage	
Panel A			
Dependent variables:	Log(Tweets)	*Number of demonstrations*	
SXSW March 2007	0.049***		
	(0.011)		
log(Tweets)		189.266***	87.929***
		(62.372)	(29.837)
SXSW 2006	−0.001	−1.154	6.025*
	(0.033)	(6.179)	(3.360)
R^2	0.943		
Cluster robust F-stat	18.28		
Mean dep. var.		6.89	4.11
N	3,103	3,103	3,103
Panel B			
Dependent variables:	Log(Tweets)	*Share among all demonstrations*	
SXSW March 2007	0.048***		
	(0.011)		
log(Tweets)		0.154**	−0.154**
		(0.074)	(0.074)
SXSW 2006	0.004	−0.017**	0.017**
	(0.033)	(0.008)	(0.008)
R^2	0.948		
Cluster robust F-stat	16.84		
Mean dep. var.		0.72	0.28
N	1,753	1,753	1,753

Note: This table presents county-level regressions. Log(Tweets) is the natural logarithm of a sample of tweets from 2014-2015. SXSW March2007 is the number of Twitter users who joined Twitter in March 2007 and follow South by Southwest (@SXSW). SXSW 2006 is the number of Twitter users who joined in 2006 and follow SXSW. Column (1) presents first stage regressions, while columns (2) and (3) present second stages. "BSMs" stands for demonstrations that are part of "Broad Social Movements," while "No Org." refers to demonstrations that cannot be associated with a group or an organization. Panel A presents TSLS regressions where the number of demonstrations that are "BSMs" or "No Org." is the dependent variable (column 2), or where the number of other demonstrations is the dependent variable (column 3). Panel B presents regression results on a subsample of counties where at least 1 demonstration event occurred between Jan 2020 and Nov 2021, where the dependent variable is the share of demonstrations that fit the category. All regressions control for population via 2-percentile dummies, and the full set of controls (as in column 5 in Table A15), see Table A1 in the Appendix for their descriptions. Robust standard errors in parentheses are clustered by state. *p<0.1; **p<0.05; ***p<0.01

Table A18: TSLS demonstrations by topic with additional election controls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Type of demonstrations:		Coronavirus	"Stop the Steal"	Abortion rights	Racial issues	White Supremacist	Polarizing topics
Specification:	First Stage	Second Stage					
Panel A							
Dependent variables:	Log(Tweets)	*Number of demonstrations*					
SXSW March 2007	0.049***						
	(0.011)						
log(Tweets)		34.990**	2.502**	2.787**	125.719***	7.868**	162.609***
		(15.163)	(1.274)	(1.181)	(43.302)	(3.091)	(51.314)
SXSW 2006	−0.001	−0.131	−0.093	0.149	−2.028	−0.326	−2.017
	(0.033)	(1.286)	(0.162)	(0.139)	(4.511)	(0.335)	(5.008)
R^2	0.943						
Cluster robust F-stat	18.28						
Mean dep. var.		1.10	0.20	0.25	4.28	0.14	5.72
N	3,103	3,103	3,103	3,103	3,103	3,103	3,103
Panel B							
Dependent variables:	Log(Tweets)	*Share among all demonstrations*					
SXSW March 2007	0.048***						
	(0.011)						
log(Tweets)		−0.030	0.021	0.015	0.153**	−0.001	0.161*
		(0.033)	(0.014)	(0.023)	(0.074)	(0.014)	(0.085)
SXSW 2006	0.004	0.001	−0.003*	−0.001	−0.017*	0.0004	−0.020*
	(0.033)	(0.004)	(0.002)	(0.003)	(0.010)	(0.002)	(0.010)
R^2	0.948						
Cluster robust F-stat	16.84						
Mean dep. var.		0.096	0.012	0.021	0.495	0.007	0.622
N	1,753	1,753	1,753	1,753	1,753	1,753	1,753

Note: This table presents county-level regressions. Log(Tweets) is the natural logarithm of a sample of tweets from 2014-2015. SXSW March2007 is the number of Twitter users who joined Twitter in March 2007 and follow South by Southwest (@SXSW). SXSW 2006 is the number of Twitter users who joined in 2006 and follow SXSW. Column (1) presents first stage regressions, while columns (2)-(7) present second stages. Coronavirus refers to demonstrations related to public health measures due to the Covid pandemic. "Stop the Steal" refers to demonstrations related to the 2020 post-presidential election unrest (including those calling for results to be respected). Abortion rights refers to demonstrations related to reproductive rights. Racial issues refers to demonstrations concerning racial tensions while white supremacists are a subset of those where demonstrating groups have known affiliations with white supremacy ideology. Polarizing topics include all other topics in the table. A detailed description of topic categorization is presented in the Appendix. Panel A presents TSLS results where the number of demonstrations fitting the given column category is the dependent variable. Panel B presents TSLS results on a subsample of counties where at least 1 demonstration event occurred between Jan 2020 and Nov 2021, where the dependent variable is the share of demonstration that fit the column-category. All regressions control for population via 2-percentile dummies, and the full set of controls (as in column 5 in Table A15), see Table A1 in the Appendix for their descriptions. Robust standard errors in parentheses are clustered by state. *p<0.1; **p<0.05; ***p<0.01

Table A19: TSLS demonstrations by "peacefulness" with additional election controls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Type of demonstration:		One Sided	Two Sided	Peaceful	Non-Peaceful	Violent Protesters	Peaceful Protesters met with intervention or violence
Specification:	First Stage	Second Stage					
Panel A							
Dependent variables:	Log(Tweets)	*Number of demonstrations*					
SXSW March2007	0.049***						
	(0.011)						
log(Tweets)		259.467***	17.728***	215.839***	61.356*	38.959	22.397***
		(81.748)	(6.264)	(74.123)	(33.008)	(26.779)	(7.578)
SXSW 2006	−0.001	5.386	−0.516	7.807	−2.936	−2.294	−0.642
	(0.033)	(8.300)	(0.566)	(7.324)	(3.534)	(2.821)	(0.818)
R^2	0.943						
Cluster robust F-stat	18.28						
Mean dep. var.		10.47	0.53	10.30	0.70	0.33	0.37
Panel B							
Dependent variables:	log(Tweets)	*Shares among all demonstrations*					
SXSW March2007	0.048***						
	(0.011)						
log(Tweets)			−0.006		0.099	0.067	0.032
			(0.025)		(0.074)	(0.058)	(0.024)
SXSW 2006	0.004		−0.0001		−0.010	−0.007	−0.003
	(0.033)		(0.003)		(0.009)	(0.007)	(0.003)
R^2	0.948						
Cluster robust F-stat	16.84						
Mean dep. var.			0.003		0.034	0.012	0.022

Note: This table presents county-level regressions. Log(Tweets) is the natural logarithm of a sample of tweets from 2014-2015. SXSW March2007 is the number of Twitter users who joined Twitter in March 2007 and follow South by Southwest (@SXSW). SXSW 2006 is the number of Twitter users who joined in 2006 and follow SXSW. Column (1) presents first stage regressions, while columns (2)-(7) present second stages. "Two sided" refers to events where two demonstrations who are essentially counter one another occur at the same time in the same place. "One sided" refers to all demonstrations that are not "Two sided." Violent protesters are events where demonstrators engage in violent or disruptive behaviour and Peaceful Protesters met with intervention or violence are events where demonstrators are peaceful but there is an attempt of suppression or the use of violence by outside parties. Non-Peaceful refers to the sum of the protests in columns (6) and (7) while Peaceful refers to all other demonstrations. Panel A presents TSLS results where the number of demonstrations fitting the given column category is the dependent variable. Panel B presents TSLS results on a subsample of counties where at least 1 demonstration event occurred between Jan 2020 and Nov 2021, where the dependent variable is the share of demonstration that fit the column-category. All regressions control for population via 2-percentile dummies, and the full set of controls (as in column 5 in Table A15), see Table A1 in the Appendix for their descriptions. Robust standard errors in parentheses are clustered by state. *p<0.1; **p<0.05; ***p<0.01

Table A20: TSLS demonstrations by intervening actor with additional election controls

	(1)	(2)	(3)
Intervening actor:		Intervention by Gov.	Intervention by non-Gov. Actor
Specification:	First Stage	Second Stage	
Panel A			
Dependent variables:	Log(Tweets)	*Number of demonstrations where peaceful protesters are met with intervention or violence*	
SXSW March 2007	0.049***		
	(0.011)		
log(Tweets)		17.376***	5.021***
		(5.901)	(1.744)
SXSW 2006	−0.001	−0.411	−0.231
	(0.033)	(0.637)	(0.190)
R^2	0.943		
Cluster robust F-stat	18.28		
Mean dep. var.		0.31	0.06
N	3,103	3,103	3,103
Panel B			
Dependent variables:	Log(Tweets)	*Share among all demonstrations*	
SXSW March 2007	0.048***		
	(0.011)		
log(Tweets)		0.025	0.008**
		(0.022)	(0.004)
SXSW 2006	0.004	−0.002	−0.001**
	(0.033)	(0.003)	(0.0005)
R^2	0.948		
Cluster robust F-stat	16.84		
Mean dep. var.		0.019	0.003
N	1,753	1,753	1,753

Note: This table presents county-level regressions. Log(Tweets) is the natural logarithm of a sample of tweets from 2014-2015. SXSW March2007 is the number of Twitter users who joined Twitter in March 2007 and follow South by Southwest (@SXSW). SXSW 2006 is the number of Twitter users who joined in 2006 and follow SXSW. Column (1) presents first stage regressions, while columns (2) and (3) present second stages. Column (2) looks at demonstrations where peaceful protesters are met with an attempt of suppression or the use of violence by government entities, while column (3) looks at demonstrators where peaceful protesters are met with an an attempt of suppression or the use of violence by non-government actors. Panel A presents TSLS results where the number of demonstrations fitting the given column category is the dependent variable. Panel B presents TSLS results on a subsample of counties where at least 1 demonstration event occurred between Jan 2020 and Nov 2021, where the dependent variable is the share of demonstration that fit the column-category. All regressions control for population via 2-percentile dummies, and the full set of controls (as in column 5 in Table A15), see Table A1 in the Appendix for their descriptions. Robust standard errors in parentheses are clustered by state. *p<0.1; **p<0.05; ***p<0.01