Twitter and Crime: The Effect of Social Movements on Gender-Based Violence [a]

Michele Battisti[b] Ilpo Kauppinen[c] Britta Rude[d]

November 15, 2022

Keywords: Economics of Gender, US, Domestic Abuse, Public Policy, Criminal Law, Illegal Behavior and the Enforcement of Law
JEL Codes: J12, J16, J78, K14, K42, O51

Abstract. This paper asks whether social movements taking place on Twitter affect gender-based violence (GBV). Using Twitter data and machine learning methods, we construct a novel data set on the prevalence of Twitter conversations about GBV. We then link this data to weekly crime reports at the federal state level from the United States. We exploit the high-frequency nature of our data and an event study design to establish a causal impact of Twitter social movements on GBV. Our results point out that Twitter tweets related to GBV lead to a decrease in reported crime rates. The evidence shows that perpetrators commit these crimes less due to increased social pressure and perceived social costs. The results indicate that social media could significantly decrease reported GBV and might facilitate the signaling of social norms.

[a]We thank Davide Cantoni, Lelys Dinarte, Jon Fiva, Eleonora Guarnieri, Ines Helm, Andreas Peichl, Panu Poutvaara, Helmut Rainer, Monika Schnitzer, Claudia Steinwender, Madhinee Valeyatheepillay for helpful comments and remarks. We thank participants of the Society of Family and Gender Economics Webinar 2022, the CESifo/ifo Junior Workshop on Big Data, the 2nd Berlin Workshop on Empirical Public Economics: Gender Economics 2022, the ifo Lunchtime Seminar, the Scottish Economic Society 2022 Annual Conference and the PhD Idea Seminar at the University of Munich for helpful comments and suggestions.

[b]Contact: michele.battisti@glasgow.ac.uk, University of Glasgow, IZA, CESifo, CReAM
[c]Contact: ilpo.kauppinen@vatt.fi, VATT Institute
[d]Contact: rude@ifo.de, ifo Institute - Leibniz Institute for Economic Research at the University of Munich and Ludwig Maximilian University Munich.

1 Introduction

Gender-based violence (GBV), which refers to violence against individuals based on their gender, continues to be a problem in today's societies. According to UN Women (2021) every third woman falls victim to some sort of GBV at least once in her lifetime. In addition to personal costs, such as physical, mental or material harm, GBV also generates costs through the use of health services, costs to the justice system, lost economic output, social welfare expenses and the need for specialized support services (Walby and Olive, 2014). It is therefore crucial to study GBV, its drivers and potential methods to reduce it. However, GBV is associated with social stigma, shame, discriminatory and stereotypical attitudes and other factors that silence many victims.[1] In contrast, the emergence of social movements such as the *#metoo* movement has led many to openly share their experiences of GBV on social media.

This paper examines whether social movements on Twitter affect GBV-related crime rates. We focus on Twitter as one of the most well known social movements, the *#metoo* movement, took place on this social media platform. Our key hypothesis is that these movements increase the social costs and social pressure of committing GBV, thereby reducing the prevalence of GBV. In addition, we hypothesize that victims feel empowered and increasingly report GBV and given this, perpetrators may perceive increased costs also via this channel. Overall, we argue that Twitter acts as a facilitator for the signaling of gender norms. We generate a novel dataset on conversations about GBV using data from Twitter. We use a set of machine learning techniques applied to hashtags used on Twitter to construct a weekly measure of the prevalence of conversations about GBV on Twitter across federal states in the United States and over time. Our Twitter-based dataset consists of around 11.4 million tweets from the period 2014-2016. We focus on this period as we match the data to crime-incidence level data gathered by the FBI, which is only available up to 2016 at the time of this study. We take advantage of the high frequency of our datasets, and conduct regressions and employ an event study at the federal state by week level.

Our results show that the number of tweets per 100 cellphone internet subscription decreases GBV-related reported crime rates per 100,000 inhabitants to authorities. We provide evidence that behavioral changes among perpetrators of GBV most likely drive our results. If social movements on Twitter also empower victims and lead to more

[1]For instance, a study by Palermo et al. (2014) on GBV in 24 developing countries finds that only seven percent of female victims of GBV had reported to the authorities or other formal institutions. The problem of under-reporting in GBV is well established in the literature (see for example Joseph et al. (2017) or Fernández-Fontelo et al. (2019)).

reporting, our coefficients are lower bound estimates of the true underlying effect of social movements on GBV on Twitter. Our findings confirm our hypothesis and show that Twitter facilitates the signaling of social norms and generates social pressure and social costs for perpetrators of GBV. According to the standard economic framework on crime by Becker et al. (1995), criminal behavior depends on the costs and benefits, which result from a criminal offense. Our findings are in line with work by Falk and Fischbacher (2002) and Fry et al. (2019) who show that these costs increase with peer pressure and neighborhood effects as well as a number of papers demonstrating direct effects of social pressure on GBV (Standish (2014); Fry et al. (2019)). In our case, the number of Twitter tweets would be a signal of peer pressure. Moreover, perpetrators might observe the social prosecution of those alike online, and interpret this as neighborhood effects. Twitter then acts as a facilitator for the signaling of shifting social norms. This channel is in line with previous evidence on the erosion of existing social norms, such as work by Bursztyn et al. (2020) showing significant effects of Donald Trump's rise in polarity on publicly expressed xenophobic views.

This paper is the first to conduct an in-depth analysis of the impact of social movements on Twitter on crime rates related to GBV. While there has been recent interest in this research question, no paper so far has used social media data. Closest to our paper is Levy and Mattsson (2021), who analyze Google Trends data. Their analysis focuses on the extent to which individuals try to get informed on this topic, rather than on online conversations. Moreover, we take advantage of the tweets' text and study sentiments around social movements related to GBV on Twitter. We therefore believe that our analysis goes one step further in analyzing the conversation taking place on social media.[2]

Our paper advances the current understanding of potential strategies to reduce GBV. Previous research studies focus on public transfer programs (Bobonis et al., 2013), anti-poverty programs (Amaral et al., 2015), employment of female police officer (Miller and Segal (2019); Amaral et al. (2021)), or the exposure to video dramas (Cooper et al., 2020). Morrison et al. (2007) give an early literature review on potential interventions. Our paper is innovative in the sense that it explores social movements which take place on social media platforms as a potential channel to reduce GBV.

[2]The paper by Levy and Mattsson (2021) focuses on an international setup while our paper takes place in the US and considers lower geographic variation. Lastly, differently from their paper, which solely relies on the *#metoo* movement, we focus on social movements taking place on Twitter in earlier periods. We believe that focusing on the years prior to the *#metoo* movement is a more appropriate setup for our underlying research question based on lower awareness on GBV, lower digitalization, and less exposure to confounding factors such as the election of President Trump.

Based on this rationale, we also contribute to the literature on social movements by showing that social movements in online spaces translate into offline behavioral changes. These findings are in line with a limited number of studies illustrating a significant association between social media usage and hate crimes (Müller and Schwarz, 2020) as well as political outcomes (Levy (2021); Zhuravskaya et al. (2020)).[3]

We conduct spatial regressions to validate the relevance of our research design. Twitter social movements might spread quickly across federal states and the introduction of spatial spillover effects allows us to account for this concern. Results from spatial regressions enforce our findings from the main empirical specification. In addition, we investigate if Twitter users who engage in the GBV debate and victims of GBV differ systematically from each other. To do so, we employ a face recognition technique to Twitter users' profile pictures and deduce their age and ethnicity. In addition, we infer the authors' gender from their first name. Our results show that victims of GBV and those tweeting about it are, on average, of similar age and ethnicity, but there are some systematic differences in gender. A larger share of victims of GBV than Twitter users are female.

To ensure that our findings are not driven by simultaneous unobserved shocks, we conduct placebo regressions. For this purpose, we use non-GBV related crime rates as outcome variables. These placebo regressions validate our findings.

To validate that the decrease in crime rates in response to social movements on Twitter is indeed due to perpetrators committing these crimes less frequently, we explore three alternative channels, which could potentially drive our results. First, we analyze the impact of social movements on Twitter on Google search activities for informal support networks. The underlying rationale is that crime reporting rates could decrease when victims replace formal support networks with informal ones. These regressions reveal no clear pattern of results.

Second, we investigate if tweets in favor of conservative gender norms, such as tweets using the hashtag *#alphamale*, impact GBV-related crime rates differently than tweets about GBV. If the number of tweets using *#alphamale* has a positive effect on GBV-related crime rates, this might point towards perpetrators changing their behavior and committing these crimes more. Therefore, analyzing the impact of Twitter tweets in favor of conservative gender norms can help to shed light on the reporting or perpetrator channel. The analysis demonstrates that there is limited evidence in favor of a positive

[3]One particular type of social movements studied more extensively in the political economy literature are political protests (Bremer et al. (2020); Bursztyn et al. (2021); Matta et al. (2021)). Recently, several paper have analyzed the impact of the *#blacklivesmatter* movement (Dave et al. (2020); Agarwal and Sen (2022)).

effect.

Lastly, we restrict our outcome variable to GBV-related violent crimes, namely homicides or aggravated assault, as victims of violent crimes are less likely to report these crimes. Hence, violent crime rates are more likely to reflect crime perpetration rather than reporting behavior. The pattern of results suggests a negative impact of social movements on Twitter on GBV-related violent crime rates.

A change in police behavior could also drive our results. If social pressure generated on Twitter spills over to the authorities, law enforcement with respect to GBV-related cases might increase. Increased law enforcement might then further raise the costs of committing GBV and lower the GBV-related crime rate. To investigate this further, we analyze the impact of social movements on Twitter on GBV-related arrest per crime rates. We find limited evidence on a positive effect.

The second stage of our analysis differentiates between different types of GBV. We believe that we can shed light on the role of social stigma and tabooing by making this distinction. We find that the impact of Twitter use on crime rates is strongest in the case of sexual violence. We interpret these findings as stigmatization, tabooing and silencing being especially persistent in the case of sexual violence. Lastly, this paper asks whether the polarity of the tweets' text plays a role in affecting people's offline behavior. To this end, we apply text analysis methods to the tweets in our Twitter sample to study the sentiments involved in the overall conversation. Our analysis of the tweets' written content shows that the polarity of tweets does not play a significant role in changing the crime rates. Consequently, what matters is the sheer magnitude of social movements on Twitter.

This paper makes a significant contribution to the economic literature studying GBV. To the best of our knowledge, this is the first paper using Twitter data to study the effect of online social movements on GBV-related crime rates. Our work contributes to economic papers studying potential drivers of GBV[4] as well as the impact of experiencing GBV.[5]

[4]To name a few examples, Aizer (2010) shows that a decreasing wage gap comes along with a decrease in domestic violence at the household level in the US. Related work by Bhalotra et al. (2021a) illustrates that an increase in male unemployment or a decrease in female unemployment increases intimate partner violence (IPV). Similarly, Brassiolo (2016) demonstrate that a decrease in divorce costs leads a decrease in IPV. Closely related work by González and Rodríguez-Planas (2020) finds that gender norms are important drivers of IPV. Related work analyzes the association between polygony and IPV (Cools and Kotsadam, 2017), family structures and IPV (Tur-Prats, 2019), and colonialism and IPV (Guarnieri and Rainer, 2021). There is also an increasing literature studying the impact of COVID-19 related lock-downs on IPV (Agüero (2021); Berniell and Facchini (2021); Bullinger et al. (2021)).

[5]Example studies within this stream of literature are studies by Welsh (1999), Fitzgerald and Cortina (2018), or Folke et al. (2020) finding negative effects for victims of sexual harassment. A number of

Our findings have several important policy implications. Those who are interested in decreasing the prevalence of GBV should explore the potential of social media platforms. Our results also point to the important role of stigmatization, tabooing and silencing. Policymakers should design strategies to facilitate the reporting of and conversation about GBV. Moreover, they should design policies that address harmful gender norms and lead to long-term changes in beliefs and attitudes concerning GBV. Our findings point to the importance of social networks in driving social change.

The rest of the paper proceeds as follows. Section 2 presents our economic rationale, the definition of GBV used in this paper, and a review of the related literature. Section 3 outlines the creation of our Twitter dataset and describes additional datasets used in this paper. Section 4 explains our methodology and section 5 presents our main results. Section 6 analyzes some of the mechanisms behind our main findings. Section 7 concludes.

2 Economic Relevance and Motivation

2.1 The Economics of Gender-Based Violence

The United Nations defines GBV as violence against individuals based on their gender (UNHCR, 2022). It can be of physical, sexual, psychological or economic nature. GBV can take place in the private sphere, for instance in the case of child abuse or intimate partner violence, or in the public sphere, such as rape or street harassment. Victims of GBV can be male, female, or non-binary. The root cause is gender inequality, the abuse of power and harmful social norms. Child marriage, female genital mutilation, and honor crimes are also part of GBV. Globally, one third of all women fall victim of GBV at least once in their lives.

GBV has economic relevance as it generates large costs both for individuals and society as a whole. The World Bank estimates that costs related to GBV amount to one to two percent of GDP (Duvvury et al., 2013). Nevertheless, it is still a largely unexplored topic. GBV is related to the economic field through three channels. First, one form of GBV is economic violence. One example of economic violence is the control of female-owned property or financial resources, as well as women's exclusion from those resources, for example when women are excluded from inheritance or property rights. This, in turn, affects female empowerment in general and creates serious barriers for countries to realize

related papers shows negative effects of GBV on economic activities (Duvvury et al. (2013); Ouedraogo and Stenzel (2021)). Similarly, GBV leads to work deterioration (Chakraborty et al. (2018); Siddique (2022)).

their full economic potential.[6]

Second, GBV leads to serious economic harm for those who fall victim to it. For example, Ouedraogo and Stenzel (2021) explore the economic consequences of violence against women in Sub-Saharan Africa. They show that an increase in GBV decreases economic activity. In fact, an increase of one percentage point in the share of women experiencing GBV, leads to a decrease in economic activity of up to eight percentage points. A related study by Chakraborty et al. (2018) studies the effect of crime against women on female labor force participation in India. The authors show that this leads to serious work deterioration. They also demonstrate that the impact is larger for women from more conservative families and at the lower end of the wage distribution. Similarly, Siddique (2022) shows that an increase in media coverage of sexual assault in India decreases female labor force participation. On the contrary, an increase in crime rates leads to an increase in male labor force participation (Mishra et al., 2021). Further studies on the negative effects of sexual harassment have been conducted by Welsh (1999), Fitzgerald and Cortina (2018) and Folke et al. (2020).

Third, several studies show that economic circumstances affect the occurrence of GBV. For example, a study by Li et al. (2019) shows that global economic integration, as measured by foreign direct investment, leads to an increase in female economic empowerment and a decrease in rape incidences in India. Household bargaining power also plays an important role. Related work by Aizer (2010) shows that a decreasing wage gap os associated with a decrease in household-level domestic violence in the US. Similarly, an increase in male unemployment or a decrease in female unemployment increases intimate-partner violence (IPV) (Bhalotra et al., 2021a). Cools and Kotsadam (2017) illustrate that resource inequality is associated with an increase in incidents of abuse. A decrease in divorce costs leads to a decrease in IPV (Brassiolo, 2016). Moreover, recent work by Bullinger et al. (2021) shows that COVID-19 lock-downs led to an increase in domestic violence-related calls to the police.

In addition, a recent body of literature analyzes how to decrease GBV.[7] Recently,

[6]See for example Swamy (2014) on the positive effect of female financial inclusion on poverty reduction.

[7]For example, Amaral et al. (2021) show that opening of women's police stations leads to an increase in police reports of crimes against women. Miller and Segal (2019) show similar evidence on an increased share of female police officers in the US. In addition, Cooper et al. (2020) conclude that exposure to videos that dramatize violence against women and girls (VAWG) increases their reporting. Amaral et al. (2015) study the effect of an anti-poverty program in India on GBV and find that an increase in female labor force participation leads to an increase in GBV. In contrast, Bobonis et al. (2013) show that women who benefit from a public transfer program in Mexico are less likely to experience physical abuse. Others have shown a significant effect of municipal female political leaders on GBV in the US (Wen, 2021), Brazil (Delaporte and Pino, 2022) and India (Iyer et al., 2012).

more studies explore new data sources, such as big data, to analyze GBV.[8] To the best of our knowledge, there has been very limited use of social media data to study GBV in the economic literature. Levy and Mattsson (2021) explore a similar research question to ours, but they rely on google trends data as their main measure of the *#metoo* movement's intensity by country. While they use a pre-existing Twitter dataset to validate this approach, our paper uses Twitter data per se. Their work also refrains from using social media to analyze emotions or sentiments around this topic. Additionally, they restrict their Twitter dataset to geotagged tweets posted in October 2017. Consequently, our paper is an important contribution to their work, as it goes one step further in addressing the underlying research question. We leverage Twitter data for a broader time period, make use of the tweets' text, and investigate the dynamics at a more dis-aggregated level, namely at the week by state level in the United States.

2.2 The Economics of Crime

Economists mostly base their understanding of criminal behavior on a cost-benefit-approach. The first to apply theoretical considerations to criminology was Becker et al. (1995). The author based his understanding of criminal behavior on an individual choice model. In this model, individuals commit crimes as soon as the benefits of the potential criminal act exceed the costs. Benefits can be of monetary or non-monetary nature, such as feeling a sense of danger, excitement, entitlement, or satisfaction. The latter two aspects might be especially important when applying crime theory to GBV. The costs of committing a crime, on the other hand, can take different forms: material costs, psychological costs, such as fear, guilt, anxiety, social sanctions, and opportunity costs. Opportunity costs could be lost income due to time spent in prison. Lastly, there are direct punishment costs, i.e., legal fees and formal and informal sanctions.

Our paper tests the model's empirical implications. We ask whether social movements on Twitter create potential costs to perpetrators due to a perceived increase in peer pressure and informal sanctions. Moreover, judicial and police authorities might also feel pressured by society and increase the level of control and punishment of these type of crimes which could increase the potential punishment costs. Based on this rationale, we would expect a negative effect of social movements on GBV-related crimes.

[8] One example is the work by ElSherief et al. (2017) in computational science. They create a one percent sample of the public Twitter stream and then filter for certain keywords or hashtags related to GBV. They then measure certain attributes of these tweets, such as user engagement, linguistic properties and sentiments. Similarly, Khatua et al. (2018) gain insights into the occurrence of different forms of sexual violence by analyzing 700,000 tweets from the *#metoo* movement.

On the contrary, the perceived benefits could also increase according to the model when applied to our underlying research question. First, the level of thrill and excitement might increase as the perceived costs increase. Next, perpetrators might increase their engagement in GBV as they might see a long-term benefit in protecting the status-quo. This type of backlash has been outlined in previous studies (see for example Amaral et al. (2015), or Bandyopadhyay et al. (2020)). This would lead to a positive impact of social movements on Twitter on GBV.

Another possibility for a positive effect of Twitter tweets on GBV is a scenario, in which victims of GBV feel empowered and report these types of crimes more often. While the economic literature on social movements is scarce, one specific form of social movements has been studied more in depth: the impact of political protests. To name a few examples, Bremer et al. (2020) study the impact of social movement on electoral outcomes. Similarly, Bursztyn et al. (2021) investigate how political protests and political engagement interact. Related work by Matta et al. (2021) estimates the overall economic impact of mass protests. Our paper brings a new aspect to this literature through analyzing the role of social movements in the online space. We ask whether these online movements lead to real behavioral change.[9] Recently, the importance of social norms and movements as drivers of lasting change have gained more attention (see for example, Agranov et al. (2021) or Viscusi et al. (2011)), and the work at hand provides additional evidence for this stream of literature. Moreover, our paper examines the relevance of social media platforms for public policies (for a detailed review on this literature see Zhuravskaya et al. (2020)). A limited number of studies illustrate a significant association between social media usage and hate crimes (Müller and Schwarz, 2020) as well as political outcomes (Levy, 2021).

3 Data

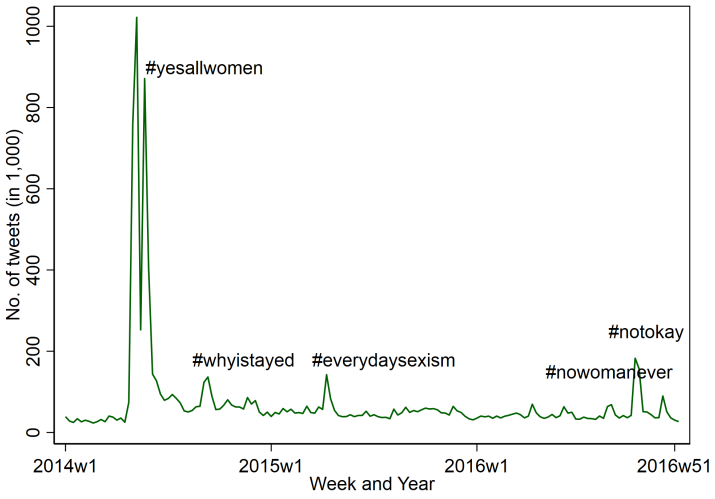
The paper at hand uses a novel dataset on GBV-related Twitter tweets. To generate our dataset, we take advantage of the Twitter API for Academic Research. This API gives academic researchers access to the entire universe of Twitter tweets since the first tweet in 2006. We use hashtags to filter the full universe of Twitter tweets for those tweets talking about GBV. Due to the character limit of the Twitter API, we only include 62 hashtags. We define a list of hashtags, which represent GBV-related Twitter movements as closely as

[9]Related work is by Dave et al. (2020) and Agarwal and Sen (2022) who analyzed the impact of the *#blacklivesmatter* movement.

possible, by applying supervised learning methods and several machine learning classifiers. We evaluate the performance of the prediction made to hand-coded classifications by an independent research assistant. See Annex B for details on the data generation process and performance measures.

After generating our final list of 62 hashtags, we retrieve all tweets, including retweets, quotes, and replies from 2014-2016, filtering by these 62 hashtags. We restrict our dataset to the years 2014-2016, as we have data on crime incidents in the US for this period. This results in 6,175,643 tweets for 2014, 2,685,019 tweets for 2015 and 2,474,767 tweets for 2016. The unit of observation of our generated dataset is the tweet level. For each tweet, we have information on its author, its text, the inclusion of other content (such as fotos, media, or videos), the number of times it was shared, liked, quoted, or replied to, and the time it was originally posted. Figure 1 shows the number of tweets in our dataset per week.

Figure 1: Number of weekly Twitter tweets about GBV (2014-2016)



Notes: The figure shows the average number of English language tweets related to GBV generated from our Twitter API via the hashtag based approach for the period 2014 to 2016. The first spike refers to the Twitter movements *#yesallwomen*. For details behind the hashtag-based approach see Appendix B. The x-axis shows the respective week in a respective year. The y-axis shows the number of Twitter tweets (in 1,000). The graph does not include the *#metoo* movement, as it only took place in October 2017. Source: Twitter (2014-2016).

We next exploit geographic variation in the extent to which people engange in this conversation on Twitter. Only a small share of Twitter tweets (approximately 1.5 percent) are geo-located. This is why we rely on users' location information. Around 76.5 percent of tweets identify a user location in their authors' user profile. However, when accessing

data through an academic account, the location information is not available in a unified format. The administrative level varies largely, from neighborhoods to cities to states. Additionally, some of the locations provided by users are fictitious. For this reason, we conduct a location mapping of all Twitter tweets that are part of our dataset. For this purpose, we match users' location with administrative data on locations in the US.[10] We can assign a location to 29.9 percent of all tweets in our dataset. To assess whether this is a suitable share, we rely on information about the location distribution of Twitter users by countries. In 2021, 37.7 percent of Twitter users were from the United States (77.75 out of 206 million users worldwide) (Statista, 2022). As our dataset consists of English-language tweets, the share of US users is probably slightly higher. This would mean that we cannot fully attribute the complete share of US tweets to a location. Moreover, many US cities have the same name. If Twitter users only indicate the city they live in, it is not possible for us to distinguish duplicate names. In these cases, we assign observations to the city with the largest population. This is an important data limitation. Although using population sizes as a decision criteria reflects a probability distribution, where larger cities and/or counties have a larger probability, a user's location might still differ from our assignment. Although we can only assign 29.9 percent of Twitter tweets to a location, there are no missing state by week combinations.[11]

To scale a federal state's Twitter activity, we leverage estimates on the number of people with a cellular data plan for a smartphone or other mobile device. The American Community Survey (ACS) elevates this data at the year by federal state level. We then divide the weekly number of Twitter tweets per federal state by these yearly estimates to scale our Twitter tweets according to our geographic level.[12]

Figure 2 shows the aggregate number of Twitter tweets over the period 2014-2016 per 100 cellphone internet subscriptions in 2014.There is a clear spatial pattern in Figure 2. Especially the southeastern federal states as well as Arizona have low Twitter rates. Structural differences between federal states, such as a lower number of Twitter users or lower use of social media in general, might drive these results. Lower internet connectivity or a higher median age in these federal states might also account for these patterns. We

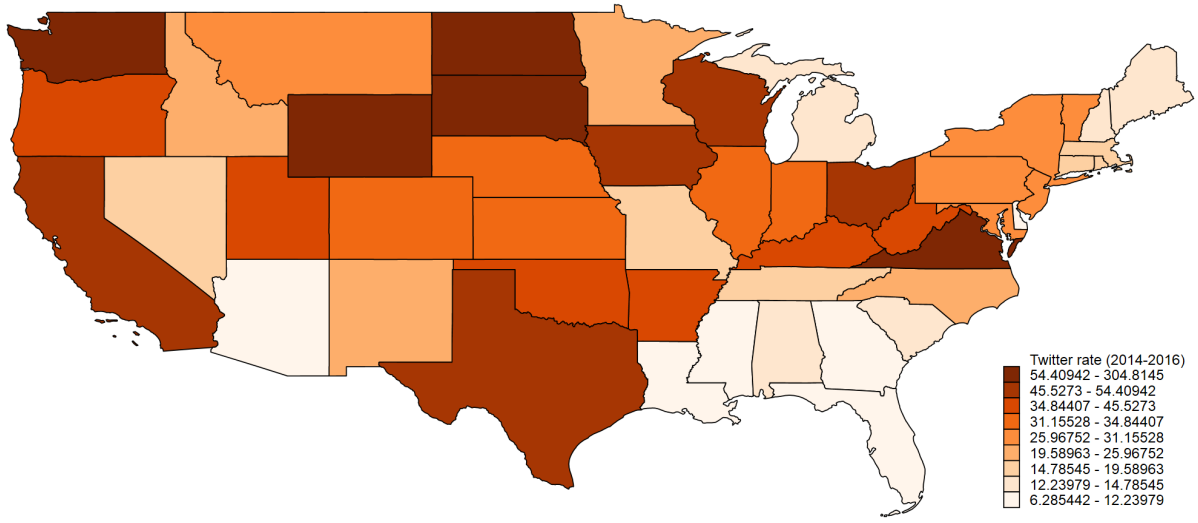
[10]For a detailed overview of the location generation see Annex B.2.

[11]There are on average 52 weeks in each year. This means that our dataset consists of roughly 156 weeks in total, as we consider three different years. When multiplying this by 50 federal states plus DC and Puerto Rico, we would theoretically end up with a dataset of 8,112 observations. We verify that there are indeed 8,112 observations in our dataset. This speaks for the quality of the data at hand.

[12]While it would be better to scale the number of Twitter tweets by the number of Twitter users, we do not dispose of this data at the state-week level. While one could theoretically generate the number of Twitter users via the API, this would exceed our monthly rate limit as well as storage space. We believe that smartphone internet plans are a good enough proxy for the number of Twitter users.

consider these potential structural factors by including federal state fixed effects in our regressions.

Figure 2: Number of GBV-related tweets over cellphone internet subscriptions by federal states (aggregate of 2014-2016)



Notes: The map depicts the aggregate number of Twitter tweets for the years 2014-2016 divided by the number of cellphone internet subscriptions in 2014 at the federal state level in the US. The graph excludes Alaska, Hawaii, and Puerto Rico. Darker colors indicate higher aggregated GBV-related crime rates. Source: Twitter data and US Census Bureau.

To measure crime reporting we use data from the National Incident-Based Reporting System (NIBRS). The NIBRS is an incidence-based reporting system managed by the FBI for police-reported crimes in the US. The system collects a variety of information on each incident reported to the police, such as the nature of the offense, characteristics of the victim(s) and offender(s), and the date and location of the incident. We use data processed by the Inter-university Consortium for Political and Social Research (ICPSR). We use the crime incident dataset for the period 2014-2016, which is at the incident-level and also contains information on each incident's offender and victim. Appendix A provides details on the dataset and its limitations.

We use the NIBRS crime classification to identify crime relevant to our research question, namely sexual violence (rape, sodomy, sexual assault with an object, fondling, statutory rape), physical violence (murder/intentional manslaughter, aggravated assault, simple assault, kidnapping/abduction) and emotional violence (intimidation). For physical violence, we use information about the circumstances of the crime and restrict the cases to those related to an argument or dispute between lovers[13]. Additionally, for physical

[13]While this measure is not perfect, given that it may also include cases of violence between opposite

and emotional violence, we limit ourselves to cases in which victim and offender are of opposite sex, as we are only interested in GBV. We do not classify other crime types, in which the victim and offender are of opposite sexes as GBV. We also consider information on the victim's sex, age, race and residence status, as well as the victim's relationship to the offender and the offender's sex, age and race.

We make use of this data by aggregating it from the crime-incidence to the week by federal state level, to gain insights into the crime activity over time as well as at the regional level.[14]

To generate data on crime rates, we leverage data on the population per federal state and year provided by the US Census Bureau. We use this data by dividing the crime reporting data by population estimates to determine the crime rate at the federal state level. To the best of our knowledge, the population data is only available at the yearly level. Therefore, we divide our high-frequency data on crimes by yearly population estimates to generate the crime rate. As this would result in very small numbers, we report the crime rate per 100,000 inhabitants.

Figure 3 plots the aggregate number of GBV-related crime reports in 2014-2016 as a share of the 2014 population at the federal state level. The map depicts significant variation in the aggregated GBV rate across states. The rate varies from near zero to 0.028. While there seems to be a clear spatial agglomeration of aggregated GBV rates in the northwest and central parts of the country, there is no clear spatial pattern in the eastern parts of the United States. This could be due to the fact that political factors, such as police and law enforcement policies, differ across federal states. We demonstrate that these spatial patterns are unlikely driven by the GBV categories in Appendix A.

We match our Twitter and crime data sets at the week by state level.[15]

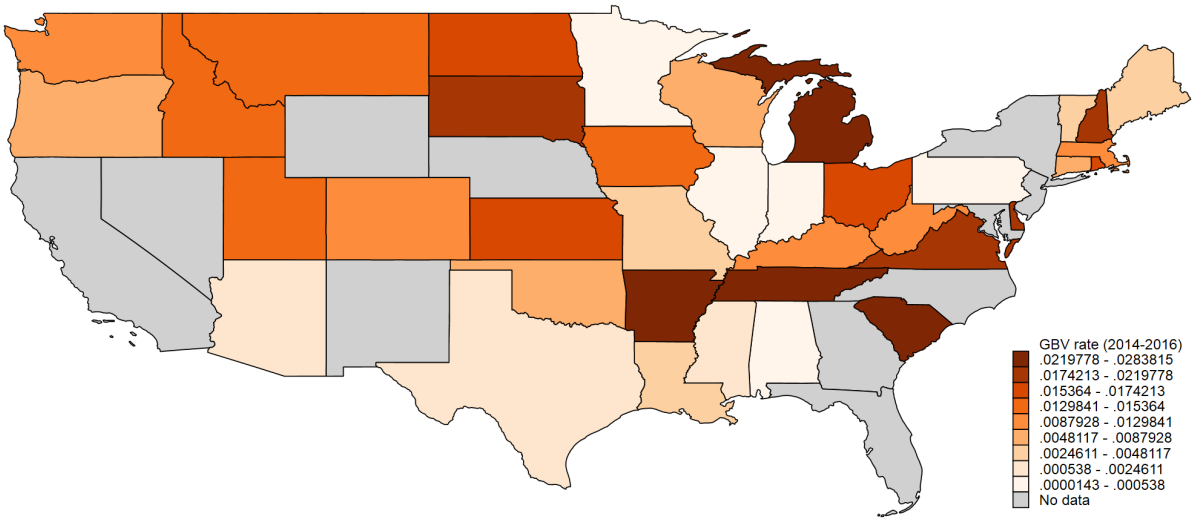
Table 1 shows the summary statistics. The table indicates that there is significant variation in the variables investigated in this paper. The GBV-related crime rate varies

sexes, such as an argument between neighbors, we are confident that it is a good approximation for GBV.

[14]Figure F2 plots the weekly reports of GBV to the police in the United States. The graph illustrates that crime reports are subject to considerable seasonal variation. We take this into consideration by controlling for respective time fixed effects in our regressions.

[15]Regressions at the national level could lead to endogeneity concerns, such as reversed causality or simultaneity bias. For example, more GBV-related crimes could lead to more tweets about GBV. Moreover, the level of variation might not be sufficient to truly understand the impact of Twitter tweets on crime reports. Because of these concerns, we dis-aggregate our two datasets to the week by state level. Each line of data then represents a different week in a different federal state. Due to the data limitations outlined in Section A we end up with missing observations. While the combination of approximately 156 weeks and 50 federal states should theoretically lead to 7,800 lines of code, the missing observations in the crime data lead to a dataset of only 5,751 observations. This is due to 2,361 week-state cells missing in the crime data and 156 missing week-state cells in the Twitter data.

Figure 3: Number of GBV-related crime reports in relation to the total population by federal states (aggregate of 2014-2016)



Notes: The map depicts the aggregate number of GBV-related crimes reported to the police for the years 2014-2016 divided by population estimates from 2014 at the federal state level in the US. The graph excludes Alaska, Hawaii, and Puerto Rico. Darker colors indicate higher aggregated GBV-related crime rates. Source: NIBRS and US Census Bureau.

between 0 and 21.833 crime reports per 100,000 people in a given week and federal state. On average, 7.233 GBV-related crimes are reported to the police per 100,000 people at the week by state level. The average crime rate is lowest in the case of sexual violence and highest for physical violence. In general, the GBV-related crime rate is much lower than that for non-GBV-related crimes. There are on average 29.364 crime reports of theft and robbery per 100,000 people at the week by state level. Our main explanatory variable, the number of Twitter tweets per 100 cellphone internet subscriptions, varies between 0.003 and 18.95, with an average of 0.148 Twitter tweets per 100 cellphone internet subscriptions.

If Twitter users who engage in the GBV-related debate differ significantly from those who fall victim of GBV, the link between both might be questionable. To investigate this further, we analyze the average characteristics of those reporting GBV to the police to those engaging in GBV-related Twitter tweets. To this end, we employ the *DeepFace* framework developed by Serengil and Ozpinar (2020) for Twitter users' profile pictures. This framework is a lightweight face recognition and facial attribute analysis package in Python.[16] It enables the determination of age, gender, emotion, and race from profile pictures. We apply this code to a 10 percent random sample of Twitter users whose

[16]Its accuracy is above 97.53 percent (Serengil and Ozpinar, 2020).

Table 1: Summary Statistics of crime data and Twitter tweets at the week by federal state level (2014-2016)

VARIABLES	mean	sd	min	max	p25	p75
GBV	5.961	5.508	0.000	21.833	1.289	10.038
Physical violence	4.000	3.715	0.000	15.638	1.035	6.178
Sexual violence	0.765	0.720	0.000	5.061	0.134	1.269
Emotional violence	1.197	1.418	0.000	6.984	0.100	1.847
Homicides	0.018	0.028	0.000	0.527	0.000	0.030
Violent crimes	0.963	1.026	0.000	4.907	0.148	1.373
Theft and Robbery	23.863	20.903	0.054	76.519	5.220	41.235
Twitter tweets	0.148	0.301	0.003	18.950	0.043	0.141

Notes: The table shows the summary statistics of the main variables of interest at the week by federal state level. For each crime type, the variable s the average crime rate per 100,000 inhabitants by calendar year and federal state. GBV refers to all crimes related to Gender-Based Violence (sexual, physical, and emotional crime). We define sexual violence as rape, sodomy, sexual assault with an object, fondling, and statutory rape. We define physical violence as murder/intentional manslaughter, aggravated assault, simple assault, kidnapping/abduction. We define emotional violence as intimidation. In the case of physical violence, we use information provided on the circumstances of the crime and restrict the cases to those related to an argument or lovers quarrel. Additionally, we restrict physical and emotional violence to cases, in which victim and offender are of opposite sexes, as we are only interested in GBV. The Twitter tweets are the number of tweets per 100 cellphone internet subscriptions in a respective year and federal state. The period under consideration is 2014 to 2016. Source: NIBRS, Twitter data, and ACS (2014-2016).

tweets, retweets, or quotes are part of our dataset. Our results show that the average Twitter user in our sample is 31.5 years old. Users are mainly white (63.8 percent). Around 9.2 percent are black. 59.6 percent of Twitter users are female.[17]

When analyzing the basic socioeconomic characteristics of victims of GBV in the period 2014-2016, it is noticeable that a large share is female (76.8 percent) (see Table 2). Moreover, victims of GBV are relatively young. The average age is 31.8 years, and only one quarter of victims are over 40 years old. The majority of victims are White (67.7 percent), 27.1 percent are Black. Only a small share are of American Indian, Alaskan (1.0 percent), Asian (1.0 percent), and Native Hawaiian origin (0.02 percent). 3.4 percent do not report race. 9.4 percent of all victims are Hispanic.[18]

In conclusion, victims of GBV and those tweeting about GBV are similar in terms of

[17]We apply the *GenderGuesser* tool to the first name of Twitter users to detect the gender of tweet authors. Appendix B.4 presents the details of this open-source Python package.

[18]Ethnicity is reported independently of race in the crime-incidence level reporting system. Most Hispanics are classified as White (95.9 percent), followed by Black.

age and ethnicity, but a smaller proportion of tweet writers than victims of GBV, who report to the police, are female. Overall, we conclude that Twitter users and victims of GBV are sufficiently similar in terms of observable characteristics to establish the relevance of the underlying research question. Still, there are some systematic differences with respect to ethnicity and gender.

Table 2: Descriptive statistics of victims of GBV (2014-2016)

VARIABLES	Mean	Std. Dev.	Min	Max	p25	p75
Age	31.81744	14.53767	0	99	22	41
White	0.67708	0.46759	0	1	0	1
Black	0.27078	0.44436	0	1	0	1
Hispanic	0.06969	0.25462	0	1	0	0
Female	0.76760	0.42236	0	1	1	1

Notes: The table shows descriptive statistics of characteristics of victims of GBV during the period 2014 to 2016 in the United States. *Age* is in years, while the rest of variables reports the share of those belonging to the respective group. While the *White* and *Black* variable is based on information gathered on victims' race, the Hispanic variable is drawn from information gathered on victims' ethnicity. Source: NIBRS (2014-2016).

To shed light on the potential mechanisms behind our results, we use a number of additional data sets. First, to analyze whether social movements on Twitter affect searches for informal support networks, we rely on data from Google Trends. More specifically, we gather data on the weekly search activity for the term *National Domestic Violence Hotline* for each federal state. We then merge this data at the federal state by week level to our Twitter dataset. Second, we generate data on a backlash movements on Twitter. For this purpose, we use a hashtag, which stands for conservative gender norms. We choose the hashtag *#alphamale*. We then retrieve all tweets in 2014-2016, which used this hashtag and aggregate it to the week by state level. To study the impact of social movements on Twitter on police behavior, we exploit the arrestee-level extract file of the NIBRS data. The arrestee-level extract file contains one record for each arrestee recorded in NIBRS for arrest dates in a given year, regardless of the date of the incident. We aggregate the number of GBV-related arrests at the week by federal state level. This allows us to combine the data on arrests with our Twitter data.[19]

[19]Figure F3 shows the number of GBV-related arrests at the weekly level in the United States. The graph shows that in line with the observations on crime reporting, there is considerable seasonality in the number of weekly GBV-related arrests. We again take this into account by controlling for time fixed effects later in our regressions.

4 Empirical Strategy

4.1 Ordinary-least Square Regressions

We run regressions at the week-state level as follows:

$$Y_{ws} = \alpha_0 + \beta_1 * T_{ws} + MY + S + \epsilon \tag{1}$$

where Y_{ws} is the crime rate of GBV per 100,000 inhabitants, or one of its subcategories, at the week by federal state level. T_{ws} is the number of GBV-related tweets per 100 cellphone internet subscriptions at the week by federal state level. MY are month of the year fixed effects which control for monthly trends, such as holiday seasons, at the national level. Federal state fixed effects (S) control for state characteristics which are constant over time. Examples are population compositions, or internet connectivity. Although our fixed effects model eliminates omitted variable bias from unobservables that are constant over time at the state level, or constant across states at the monthly level, they still leave room for confounding factors that occur at the month by federal state level, such as economic downturns or policy instruments. We cluster standard errors at the federal state by month level to account for within-group dependencies. We weight each cell by the population size of federal states to account for the relative importance of each federal state in the United States.

People might not react immediately to social movements on Twitter. They might reflect on what they observe online before internalizing this information and changing certain behavioral patterns. To account for these potential behavioral delays, we introduce lags of Twitter tweets from the previous weeks as alternative regression specifications. We consider one to two different lags, but do not go back further than one month due to our month of the year fixed effects. The introduction of lagged coefficients can also shed light on the causal interpretation of our estimates. It is unlikely that future tweets affect current or past crime rates as the Twitter movements investigated in this paper emerged suddenly. Consequently, it is unlikely that crime victims anticipate them. We can therefore ensure that our estimates represent the effect of social movements on Twitter on GBV and not the other way around.

The introduction of lagged coefficients results in the following final equation:

$$Y_{ws} = \alpha_0 + \beta_1 * T_{ws} + \beta_2 * T_{w-1s} + \beta_3 * T_{w-2s} + MY + S + \epsilon \tag{2}$$

T_{w-1s} represents the number of Twitter tweets in the previous week while T_{w-2s} is the

number of Twitter tweets two weeks previously to the one investigated.

One possible threat to the empirical estimation strategy outlined above is the existence of unobserved factors affecting both the number of total crimes committed and the number of Twitter tweets. To investigate the robustness of our empirical strategy to this potential confounding factor, we estimate placebo regressions, in which our main outcome variable is the number of crimes per week and state not related to GBV over 100,000 inhabitants: theft and robberies.

Table 3 shows that GBV-related social movements on Twitter do not have a significant impact of theft and robbery crime rates. All point estimators in the table are insignificant. We are therefore confident that our empirical strategy is robust to potential confounding factors that drive both social movements on Twitter and crime rates.

Table 3: The effect of social movements on GBV on crime rates per 100,000 inhabitants (Theft and Robbery)

	(1) Theft, Robbery	(2) Theft, Robbery	(3) Theft, Robbery
Twitter tweets	-0.170	-0.165	-0.103
	(0.167)	(0.125)	(0.133)
L.Twitter tweets		-0.0219	0.0668
		(0.147)	(0.123)
L2.Twitter tweets			-0.235
			(0.160)
Constant	23.89***	23.91***	23.94***
	(0.128)	(0.129)	(0.131)
Mean (Dep. Var)	23.86	23.89	23.90
St. Dv. (Dep. Var.)	20.90	20.92	20.92
State-Month fixed-effects	Yes	Yes	Yes
N	5751	5712	5673

Notes: The table shows the results from a linear regression of the number of Twitter tweets on the crime rate. The outcome variable is the crime rate of thefts and robberies per 100,000 inhabitants per week and federal state. The explanatory variable is the number of GBV-related tweets in the federal state during the week, divided by 100 cellphone internet plan subscriptions in the federal state in that year. The unit of analysis is the week by federal state. The first column only considers the impact of Twitter tweets on the contemporaneous crime rate. Column 2 adds Twitter tweets in the following week, while Column 3 also considers Twitter tweets two weeks later. We weight each cell by the population size of each federal state in the respective year. We control for month of the year and state fixed effects. Month by state level clustered standard errors are reported in parenthesis. Source: NIBRS, Twitter and ACS data. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4.2 Spatial regressions

Our empirical strategy relies on the assumption that geography plays a significant role in the way in which information spreads on Twitter. While previous research shows that geographic networks play an important role on Twitter (Comito (2021); Hawelka et al. (2014)), information spreads quickly across regions. This might confound our empirical strategy, which relies on geographic variations. Consequently, our results could be subject to spillover effects of Twitter tweets between federal states. To account for this possibility, we conduct spatial regressions and investigate if our results hold when allowing for spillovers between neighboring states. More concretely speaking, we follow Lee and Yu (2010) and apply a spatial autoregressive model for panel data of the form:

$$y_{fw} = \lambda W y_{fw} + c_f + u_{fw} \tag{3}$$

$$u_{fw} = \rho M u_{fw} + v_{fw} \tag{4}$$

, where y_{fw} is the crime rate or arrest per crime rate in week w and federal state f, c_f is the area fixed effect, u_{fw} is the spatially lagged error and v_{fw} is an error term, which is assumed to be independent and identically distributed. M and W are spatial weighting matrices. We assume a random effects model and include a spatial lag of our independent variable, which is the number of Twitter tweets in the respective week and federal state.

4.3 Event Study Design

Our main regressions rely on several hashtags related to GBV, and while we believe in its validity, it is certainly not the only way one could assess the impact of social movements on GBV on Twitter. In order to test the robustness of our results, we conduct an event study. Through this, we can also further establish the causality of our estimates. Event studies have also been used to study the effects of civil unrest.[20] We follow this literature and define an event as a social movement triggered by a specific hashtag. We measure the event by counting the number of English-language tweets that use the hashtag in question.

For this purpose, we only consider the hashtag *#yesallwomen*. The *#yesallwomen* movement was a reaction to the 2014 Isla Vista killings, a series of misogynistic murders

[20]To name a few examples, Dave et al. (2020) use an event study design to study the effect of Black Lives Matter protests on risk avoidance, while Chernin and Lahav (2014) analyze the impact of social protests on the financial market.

that occurred in May, 2014, close to the Santa Barbara Campus in California. The movement also partly arose in response to the hashtag *#notallmen*. Therefore, a key empirical assumption of the event study, namely that there are no other confounding events taking place at the same time as the event occurring, might be more likely. One important underlying assumption of event studies is that the event is unexpected. Figure 5 shows that the number of tweets using *#yesallwomen* increased sharply on May 24 2014, the day that officially marked the start of the social movement on Twitter.

We estimate our event study at the federal state by week level. This allows us to exploit the fact that the hashtag *#yesallwomen* was trending in different states at different points in time. We determine the treatment status of a state in a given week by its relative rank in the tweet rate compared to all other states. Once a state ranks in the top third of states with the highest tweet rate, we consider the state as treated. We consider the first week that a state ranked in the top third as the moment that state starts to be exposed to the Twitter social movement. We then calculate the relative event time relative to this cutoff date and set all event times to zero. We follow Clarke and Tapia-Schythe (2021) and implement our event study design by estimating the following regression:

$$Y_{gt} = \beta_0 + \sum_{j=2}^{J} \beta_j \times (Leadj)_{gt} + \sum_{k=1}^{k} \gamma_k \times (Lagk)_{gk} + \mu_g + \lambda_t + \epsilon_{gt} \tag{5}$$

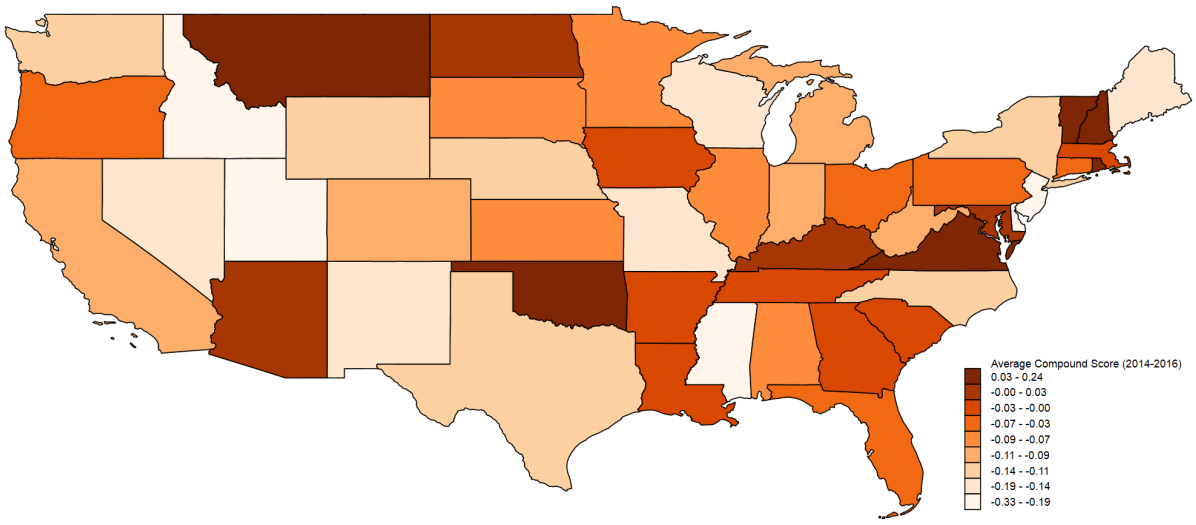
In the above equation, g is the federal state, t is the week of the year, μ_g are federal state fixed effects and λ_t are week of the year fixed effects. We consider 21 pre-treatment periods, as the *#yesallwomen* movement started in week 21 of 2014 and our sample starts in week one of 2014. We consider 52 post-treatment periods in order to include one full year after treatment exposure into our analysis. This means that $J = 11$ and $K = 52$.

4.4 Sentiment Analysis of the text of tweets

The impact of Twitter tweets could vary greatly depending on what is written. If many tweets disagree with GBV-related movements, the potential effect on crime reports might be different than in a scenario, in which people agree with them. To investigate this further, we take our processed tweet text and conduct a sentiment analysis for each tweet. A sentiment analysis is a text analysis method that determines the polarity of the underlying text. For this purpose, we rely on the Valence Aware Dictionary and Sentiment Reasoner (VADER), which was explicitly trained on social media data (Hutto and Gilbert, 2014).[21] The VADER relies on a pre-defined dictionary, which relates lexical

[21]For the details on the VADER Analysis see the Annex B.3.

Figure 4: Average Compound Score for the period 2014-2016 at the federal state level



Notes: The map depicts the average compound score at the federal state level for the period 2014-2016 in the US. We derive sentiment scores by applying the VADER Sentiment Analysis Tool (see Appendix B.3 for details). The graph excludes Alaska, Hawaii, and Puerto Rico. Darker colors indicate higher compound scores. Source: Twitter data.

features to intensities of emotions. The sentiment score of the final text is the sum of the sentiment intensity of each word in the text, and expresses the degree to which a text is positive, neutral, or negative.

We first identify the sentiment score of each tweet, and then calculate the average sentiment score for each week of each year. Figure F1 shows the average weekly sentiment score. The graph indicates that the compound score fluctuates, particularly below 0. This means that tweets had, on average, negative sentiment scores in most weeks. Calculating the average sentiment score for the period 2014-2016, yields a value of -0.133. Consequently, negative sentiments dominate the conversation.

Figure 4 shows that there is significant variation in the average compound score for the period 2014-2016 at the federal state level.

5 The Impact of Social Movements

5.1 Ordinary-least Square Regressions

The following section presents our main results. If social movements increase the social costs of GBV and deter perpetrators from committing these crimes, we expect a negative effect of Twitter tweets on crimes. This would be in line with the model by Becker et al.

(1995) described in Section 2. At the same time, reporting might increase due to victims being more empowered. If the effect on perpetration outweighs the effect on reporting we expect to see negative overall effects.

Table 4: The effect of social movements on Twitter on GBV-related crime rates per 100,000 inhabitants (GBV)

	(1) GBV	(2) GBV	(3) GBV
Twitter tweets	-0.0608*	-0.0257	-0.00476
	(0.0367)	(0.0376)	(0.0384)
L.Twitter tweets		-0.0698**	-0.0378
		(0.0340)	(0.0277)
L2.Twitter tweets			-0.0842**
			(0.0425)
Constant	5.970***	5.977***	5.987***
	(0.0266)	(0.0271)	(0.0273)
Mean (Dep. Var)	5.961	5.963	5.968
St. Dv. (Dep. Var.)	5.508	5.508	5.512
State-Month fixed-effects	Yes	Yes	Yes
N	5751	5712	5673

Notes: The table shows the results from a linear regression of the number of Twitter tweets on crime rates. The unit of analysis are week-state combinations. The outcome variable is the crime rate per 100,000 inhabitants in a respective week and federal state, considering all GBV-related crimes. We define GBV-related crimes as physical, sexual, and emotional crimes, in which the perpetrator and victim are of a different gender. The explanatory variable is the number of GBV-related tweets in a given federal state and week, divided by 100 cellphone internet plan subscriptions in the federal state in that year. In the first column we only considers the impact of Twitter tweets on the contemporaneous crime rate. In Column 2, we add Twitter tweets of the previous week, while in Column 3, we also considers Twitter tweets two weeks previously. We weight each cell by the population size of each federal state in the respective year. We control for month of the year and state fixed effects. Month by state level clustered standard errors are reported in parenthesis. Source: NIBRS, Twitter and ACS. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4 shows the results of our main empirical specification.[22] Our findings indicate that social movements do indeed deter GBV-related crimes at the state by week level. In Column 1, we do not include the number of GBV-related Twitter tweets in the previous week. The reported coefficient on GBV is -0.068 and significant at the 10 percent significance level. This means that social movements decrease the crime rate. One additional

[22]Results are sensitive to the inclusion of fixed effects and the level of clustering, given the seasonality in the underlying data observed in Figure F2. For detailed results on variations in fixed effects and clustering see Table E15.

tweet per 100 total cellphone internet subscriptions decreases the number of reported GBV-related crimes by 0.068 per 100,000 people. Compared to the mean, this effect is roughly a 1 percent decrease.

Column 2 and 3 account for a lagged effect of Twitter tweets on GBV as people might not react immediately to what they observe online. Unlike Column 1, the coefficient in Row 1 is insignificant when lags are introduced, while the lagged coefficients are significant at the 5 percent significance level. In terms of magnitude, the coefficients are similar to those observed in Column 1, but increase over time. The point estimate in Column 3 indicates that the crime rate per 100,000 people decreases by 1.4 percent relative to the mean.

Our results should be taken with caution as they might be subject to empirical limitations. First, there might be a simultaneity bias. GBV-related social movements on Twitter might be triggered by an increase in GBV-related crimes committed or arrest per crime rates in a given federal state. Given that our results persist when lagged coefficients are introduced, we believe that this is unlikely. We confirm the causality of our findings through an event study later in this section.

5.2 Spatial Regressions

We next report the results from spatial regressions in Table 5. These findings confirm our main results. The impact of Twitter tweets on GBV-related crime rates is negative. Importantly, while the direct effect of Twitter tweets on crime rates in the same federal state is insignificant, the spillover effect is significant at the 1 percent significance level. The overall effect reported in the third row of the Table is significant and larger than the direct effect presented in the first row. Consequently, an increase in the number of Twitter tweets at the state-week level is associated with a decrease in the GBV-related crime rate. Accounting for spatial spillover effects of Twitter tweets across federal states confirms our findings. The overall impact of social movements on Twitter on GBV-related crime rates is larger when accounting for spillover effects between neighboring federal states.

Table 5: Spatial regression of Twitter tweets on crime rates at the state-week level

	Coefficient	P-Values
direct Twitter tweets	-.0006908	0.985
indirect Twitter tweets	-.1745702	0.001
total Twitter tweets	-.175261	0.000
Observations	5616	

Notes: The table shows the results from a spatial regression of the number of Twitter tweets on GBV-related crime rates. The unit of analysis is the week by federal state. The direct effect presents results for the impact of Twitter tweets in the same federal state, while the indirect effect presents the results on spatial lags. The total effect presents results for the combination of both. We control for state fixed effects and month fixed effects. Column 1 presents the regression coefficient, while Column 2 presents the coefficient p-value. Source: Twitter data (2014-2016).

5.3 Event Study Regression

Figure 6 shows the event study graph for Twitter tweet exposure to tweets using the hashtag *#yesallwomen*. The graph confirms our findings from previous analyses. There is a crime deteriorating effect of social movements on Twitter on GBV with a significant drop in the coefficients after week 11. Importantly, the event study assumption is satisfied. There is no clear pretrend in the event study graph.

It is worth mentioning that our main results from the ordinary-least square regressions and the event study analysis are not fully comparable. While we look at short-term effects in our main analysis, the event study design considers the weekly impact of Twitter tweets on crime rates up to one year later. Moreover, the event study graph controls for week of year effects, while our main specification relies on month of year fixed effects. In addition, our event study uses data on only one hashtag. We validate that our OLS results hold for a data set that only consists of tweets mentioning the hashtag *#yesallwomen* as part of their written text in Appendix C.3.

Figure 5: Weekly number of tweets using the hashtag *#yesallwomen* (2014-2016)

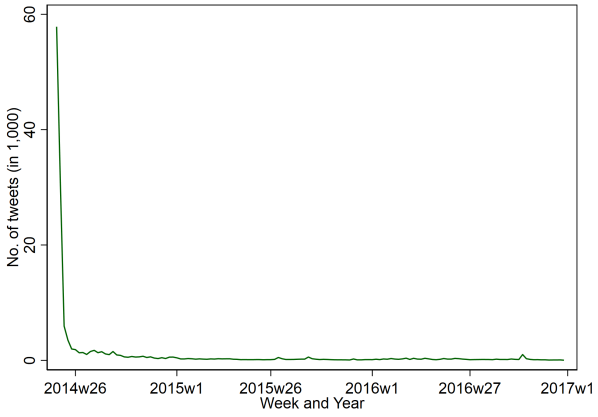
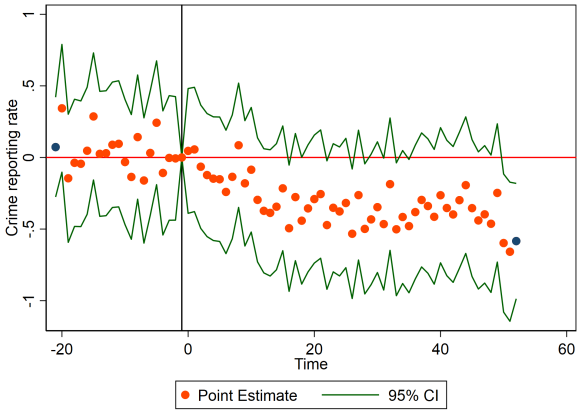


Figure 6: Event study graph at the week by state level



Notes: The figure on the left shows the weekly number of tweets using the hashtag *#yesallwomen*. The figure on the right shows the event study graph for exposure to the hashtag *#yesallwomen* at the state by week level. The vertical black line indicates the most recent period, in which states were not treated. Treated states are those states that are in the top third of states with the highest tweet rate for at least one period prior to the week under consideration. We control for federal state and week of the year fixed effects. We consider 21 pre-treatment periods, as the *#yesallwomen* movement begins in week 21 of 2014 and our sample starts in week one of 2014. We consider 52 post-treatment periods to account for a full year after treatment. For details on the estimation procedure see Clarke and Tapia-Schythe (2021). Source: Twitter, NIBRS, and ACS.

5.4 Interpretation of Main Results

Based on our main findings, we conclude that online conversations on Twitter have important implications for offline behavior. Twitter most likely facilitates the signaling of shifting social norms. This is in line with previous studies demonstrating a significant association between social norms and GBV (Linos et al. (2013); Yilmaz (2018)), as well as work by Bursztyn et al. (2020) demonstrating significant effects of Donald Trump's rise in polarity on publicly express xenophobic views. Our findings demonstrate that perpetrators are increasingly aware of the social pressure resulting from online movements and fear social punishment. They might also become more aware of punishments experienced by other perpetrators due to the increased visibility of these cases on social media platforms. Our estimates might be subject to reporting bias. If victims of GBV feel empowered by social movements on Twitter, the reporting of GBV might increase. Our coefficients might then reflect lower bound estimates of the true underlying effect on crime.

A change in perpetrators' behavior as a result of increased social pressure is in line

with the standard economic perspective on crime. From an economic viewpoint, criminal activity varies with the price of committing a crime. This price increases with peer pressure and neighborhood effects (Falk and Fischbacher, 2002). Previous research studying the role of social pressure in the prevention of GBV-related crimes confirm the role of potential social costs. Standish (2014), for example, find that social pressure plays an important role in the prevention of dowry murder. Likewise, a literature review on the role of gender norms in GBV in forced displacement settings finds that social pressure plays a significant role in the perpetration of GBV by male youth (Fry et al., 2019). In a slightly different setting, Balestrino (2008) explain that people who do not normally commit illegal acts became digital pirates as there was no social stigma and, consequently, no social costs associated with doing so.

While our interpretation of a negative impact of social movements on Twitter on GBV-related crime rates is convincing and in line with the theoretical rationale, there could be alternative explanations for a negative effect, such as the substitution of formal for informal support networks, or backlash. We next explore these potential alternative channels in detail and show that it is unlikely that they drive our results.

6 Drivers of GBV-Related Social Movements

6.1 Reporting versus Committing Crimes

In this section, we investigate potential alternative channels which could explain a negative effect of social movements on Twitter on GBV-related crime rates, next to behavioral changes in perpetrators.

One might argue that our results reflect a decrease in the reporting behavior of victims of GBV. This could be due to their increasing access to informal support networks on Twitter, or to potential backlash by perpetrators. Either would result in a decrease in the crime rate. While this seems counter-intuitive, it could well be that people who have been victim of GBV have found informal online support networks, or alternative ways to express their outrage or pain about these experiences thanks to online platforms, such as Twitter. They might therefore feel less urged to report their experiences to the authorities. McCart et al. (2010) find that only a small fraction of crime victims seek help from formal support networks while many seek help from informal sources. This pattern may have increased due to GBV-related Twitter tweets and easier access to these informal networks as many people identify as victims of GBV.

To investigate this further, we analyze the impact of social movements on Twitter for

Google search activities on the term "*National domestic violence hotline*". If an increase in access to informal support networks drives the observed decrease in crime rates, we would expect a positive impact of the number of Twitter tweets on the Google search activity for this term.

Table 6 shows that there is no clear pattern for the interaction between Twitter tweets and Google search activity for informal support. The lagged coefficients presented in Column 3, although significant at the 10 percent significance level, are in opposite directions. This finding implies that an increase in the number of Twitter tweets per 100 cellphone internet subscriptions initially leads to an increase in the search for informal support. One week later, it leads to a decrease. Consequently, it is unlikely that this channel is the dominant driver behind our results.

Table 6: The effect of social movements on Google searches on informal support networks

	(1) Google Trends	(2) Google Trends	(3) Google Trends
Twitter tweets	0.482	-0.245	0.309
	(2.199)	(2.275)	(2.208)
L.Twitter tweets		1.434	2.410*
		(1.563)	(1.364)
L2.Twitter tweets			-2.270*
			(1.294)
Constant	8.931***	8.839***	8.952***
	(0.433)	(0.457)	(0.472)
Mean (Dep. Var)	8.996	9.000	9.012
St. Dv. (Dep. Var.)	19.14	19.17	19.20
State-Month fixed-effects	Yes	Yes	Yes
N	4212	4185	4158

Notes: The table shows the results from a linear regression of the number of Twitter tweets on google search activity for the term *National domestic violence hotline*. The unit of analysis are week-state combinations. The explanatory variable is the number of GBV-related Twitter tweets in a respective week by federal state, per 100 cellphone internet plan subscriptions in a respective year and federal state. In Column 1, we only consider the impact of Twitter tweets on the contemporaneous crime rate. In Column 2, we add Twitter tweets in the previous week, while in Column 3, we also considers Twitter tweets two weeks previously. We weight each cell by the population size of each federal state in the respective year. We control for month of the year by state fixed effects. Month by state level clustered standard errors are reported in parenthesis. Source: Google Trends and Twitter. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Alternatively, the number of crimes reported by victims could decrease because they experience backlash. A backlash is a sudden and violent backward movement. Backlash has been identified in the political economic literature as a response to female political empowerment (example studies include Gangadharan et al. (2019), Gagliarducci and Paserman (2012)), and several identified it in response to female economic empowerment (such as Bobonis et al. (2013), Erten and Keskin (2018), Guarnieri and Rainer (2021), Bhalotra et al. (2021b)). Backlashes by male partners might increase as a response to GBV-related social movements on Twitter with the aim to protect the status quo. Perpetrators might feel threatened by online social movements and, consequently, further intimidate their victims.

We investigate the existence of backlash by testing the impact of a Twitter social movement identified as a backlash movement. If our results are driven by victims reporting fewer GBV-related crimes to the police due to backlash, we would expect an even more negative effect of Twitter backlash movements on GBV-related crime rates. To investigate this possible channel, we retrieve all tweets with the hashtag *#alphamale* from the Twitter API. We believe that this hashtag embraces traditional gender norms and a traditional understanding of masculinity. We then estimate our regressions using the number of tweets with the hashtag *#alphamale* per 100 cellphone internet subscriptions as our main explanatory variable.

The evidence presented in Table 7 may confirm our hypothesis that behavioral changes by perpetrators drive our main findings. The number of Twitter tweets belonging to Twitter backlash movements does not affect crime rates. While the coefficient in Column 1 is significant at the 10 percent significance level, it becomes insignificant when accounting for lagged coefficients in Column 2 and 3. If anything, the coefficient in Column 1 would indicate an increase in GBV, as it is unlikely that reporting would increase in response to such a movement. Consequently, the results presented in Table 7 confirm that our main findings are based on changes in offender behavior.

To shed further light on the question of the extent to which reporting behavior affects our overall estimator, we investigate the impact of Twitter tweets on violent crimes, namely homicides and aggravated assault. The underlying idea is that homicides cannot be driven by a change in reporting behavior. In addition, aggravated assaults often involve relationships that require protection, such as a caregiver and a mentally ill person. The victims who require protection might be less unlikely to report crimes by themselves. Therefore, the crime rate of violent crimes might be less susceptible to reporting bias. Consequently, if there is a significant impact of social movements on Twitter on the

Table 7: The effect of tweets using the hashtag *#alphamale* on crime rates per 100,000 inhabitants (GBV)

	(1) GBV	(2) GBV	(3) GBV
Twitter tweets	19.61*	16.68	12.27
	(10.59)	(14.81)	(15.67)
L.Twitter tweets		3.967	-7.770
		(14.41)	(14.15)
L2.Twitter tweets			19.36
			(12.69)
Constant	5.948***	5.948***	5.951***
	(0.0264)	(0.0264)	(0.0263)
Mean (Dep. Var)	5.961	5.963	5.968
St. Dv. (Dep. Var.)	5.508	5.508	5.512
State-Month fixed-effects	Yes	Yes	Yes
N	5751	5712	5673

Notes: The table shows the results from a linear regression of the number of Twitter tweets on the crime rate. The unit of analysis are week-state combinations. The outcome variable is the respective crime rate per 100,000 inhabitants per week and federal state, considering all GBV-related crimes. The explanatory variable is the number of GBV-related Twitter tweets using the hashtag *#alphamale* in a given week and federal state, per 100 cellphone internet plan subscriptions in a given year and federal state. In Column 1, we only consider the impact of Twitter tweets on the contemporaneous arrest per crime rate. In Column 2, we add Twitter tweets in the previous week, while in Column 3, we also consider Twitter tweets posted two weeks previously. We weight each cell by the population size of each federal state in the respective year. We control for month of the year and state fixed effects. Month by state level clustered standard errors are reported in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

violent crime rate, it is likely driven by a change in perpetrators' behavior rather than victims' reporting behavior.

Table 8 demonstrates that Twitter tweets reduce the crime rate of violent crimes. The coefficient on the lagged number of Twitter tweets in Row 2 is significant at the 5 percent significance level. Moreover, when abstracting from lagged coefficients, Column 1 reports a significant estimator at the 1 percent significance level. Importantly, the coefficients indicate that the number of Twitter tweets lowers the violent crime rate. One additional tweet per 100 cellphone internet subscriptions leads to a decrease of 0.02 to 0.03 violent crimes per 100,000 people. This evidence suggests that perpetrators' behavior drive our results.

In summary, the evidence provided in this section bolsters the case for changes in

perpetrators' behavior driving our results. While victims' reporting behavior might also change in response to social movements on Twitter, it is unlikely that this channel dominates the overall estimator.

Table 8: The effect of social movements on GBV on crime rates per 100,000 inhabitants (Violent crimes)

	(1) Violent crimes	(2) Violent crimes	(3) Violent crimes
Twitter tweets	-0.0258***	-0.0105	-0.00652
	(0.00859)	(0.00795)	(0.00889)
L.Twitter tweets		-0.0311**	-0.0249**
		(0.0133)	(0.0118)
L2.Twitter tweets			-0.0158
			(0.0129)
Constant	0.967***	0.970***	0.973***
	(0.00639)	(0.00666)	(0.00670)
Mean (Dep. Var)	0.963	0.964	0.966
St. Dv. (Dep. Var.)	1.026	1.026	1.028
State-Month fixed-effects	Yes	Yes	Yes
N	5751	5712	5673

Notes: The table shows the results from a linear regression of the number of Twitter tweets on the crime rate. The unit of analysis are week-state combinations. The outcome variable is the respective crime rate per 100,000 inhabitants per week and federal state, considering all violent crimes. We define violent crimes as homicides and aggravated assault, in which the perpetrator and victim are of a different gender. The explanatory variable is the number of GBV-related Twitter tweets in a respective week and federal state, per 100 cellphone internet plan subscriptions in a given year and federal state. In Column 1, we only consider the impact of Twitter tweets on the contemporaneous arrest per crime rate. In Column 2, we add Twitter tweets in the previous week, while in Column 3 we also consider Twitter tweets posted two weeks previously. We weight each cell by the population size of each federal state in the given year. We control for month of the year and state fixed effects. Month by state level clustered standard errors are reported in parenthesis. Source: NIBRS, Twitter and ACS. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

6.2 The Role of Stigmatization and Tabooing

To better understand what drives our results we analyze to which extent social stigmatization and tabooing play a role in our findings. This is an important question, as it can provide further guidance to policymakers on how best to address the underlying causes of GBV. To this end, we conduct a more granular analysis distinguishing between sexual,

physical and emotional GBV.[23] We subdivide by these types of GBV as the degree of stigmatization might differ depending on the type of GBV.[24]

Our hypothesis is that stigmatization and tabooing will be greatest for sexual violence, followed by emotional violence, and lastly physical violence.[25] If social stigmatization and tabooing are important drivers of GBV, we would expect a lower effect of social movements on sexual GBV. A greater impact on sexual GBV, on the other hand, could be due to many of the social movements on Twitter investigated in this paper focusing on sexual violence. Additionally, previous reporting rates of sexual violence might have been especially low.

Table 9 presents our findings on the impact of Twitter tweets on sexual violence. The table shows a clear negative impact of social movements on Twitter on the crime rate in the same week and one and two weeks after the social movements on Twitter. These findings are stable across model specifications and suggest that perpetrators committ these crimes at lower rates. The coefficient in Row 1 and Column 3 is -0.0242. In terms of magnitude, an increase in the number of Twitter tweets per 100 cellphone internet subscriptions decreases sexually violent crimes per 100,000 people by approximately 3.158 percent compared to the mean value. Two weeks later, the crime-reducing effect increases to 5.979 percent compared to the average.

Interestingly, the coefficient on the first lag reported in Row 2 is significant and positive. One possible interpretation is that social stigmatization and tabooing is likely higher for sexual violence than for other forms of GBV. Hence, the reporting effect triggered by social movements on Twitter could be especially large in this case and might dominate the crime reducing effect.Our results imply that social movements on Twitter potentially counteract stigma and taboo surrounding sexual violence. Thus, they encourage victims to report these types of crimes more often. This line of thought bolsters a potential interpretation of social movements on Twitter serving as a signaling mechanism for shifting social norms.

In the case of physical violence, social movements on Twitter significantly decrease the crime rate one week later. The coefficients reported in Row 2 of Table 10 are significant

[23]While one might argue that there is no clear definition of emotional violence, we base our definition on data gathered by the FBI. This means that emotional violence is intimidation between opposite sexes. While the classification might not represent the full universe of emotional violence, we believe that it is a close enough approximation to capture its occurrence.

[24]To date, there is only limited evidence of the degree of stigma by type of GBV. Work by Harris (2017) demonstrates that the type of violence does not alter the relationship between stigma and reporting GBV in the case of homosexual men.

[25]Scholars from other fields have shown that social stigma around sexual violence is especially high (see for example Delker et al. (2020)).

Table 9: The effect of social movements on GBV on crime rates per 100,000 inhabitants (Sexual violence)

	(1) Sexual violence	(2) Sexual violence	(3) Sexual violence
Twitter tweets	-0.0258***	-0.0356***	-0.0242**
	(0.00810)	(0.0116)	(0.0104)
L.Twitter tweets		0.0231*	0.0407***
		(0.0119)	(0.0138)
L2.Twitter tweets			-0.0456***
			(0.0123)
Constant	0.768***	0.765***	0.768***
	(0.00442)	(0.00425)	(0.00440)
Mean (Dep. Var)	0.765	0.763	0.764
St. Dv. (Dep. Var.)	0.720	0.717	0.717
State-Month fixed-effects	Yes	Yes	Yes
N	5751	5712	5673

Notes: The table shows the results from a linear regression of the number of Twitter tweets on the crime rate. The unit of analysis are state-week combinations. The outcome variable is the respective crime rate per 100,000 inhabitants per week and federal state, considering all crimes related to sexual violence. We define crimes related to sexual violence as rape, sodomy, sexual assault with an object, fondling, statutory rape, in which the perpetrator and victim are of a different gender. The explanatory variable is the number of GBV-related tweets in a given federal state and week, divided by 100 cellphone internet plan subscriptions in a given federal state and year. In Column 1, we only consider the impact of Twitter tweets on the contemporaneous arrest per crime rate. In Column 2, we add Twitter tweets in the previous week, while in Column 3 we also consider Twitter tweets posted two weeks previously. We weight each cell by the population size of each federal state in the respective year. We control for month of the year and state fixed effects. Month by state level clustered standard errors are reported in parenthesis. Source: NIBRS, Twitter and ACS. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

at the 1 percent significance level for all model specifications. Therefore, while Twitter tweets do not affect crime rates in the same week, it decreases the number of physical GBV-related crimes in the following week. The coefficient in Row 2 is close to -0.07, meaning that one more tweet per 100 cellphone internet subscriptions leads to a decrease of 0.07 in the number of physical crimes per 100,000 people. In terms of magnitude, the point estimator reflects a decrease of 1.705 percent compared to the average value for physical violence per 100,000 people. The effect then seems to fade out in the following weeks.

Lastly, Table 11 demonstrates a lagged effect of Twitter tweets on crime rates for emotional GBV. The point estimates presented in Table 11 fluctuate between -0.02 and

Table 10: The effect of social movements on GBV on crime rates per 100,000 inhabitants (Physical violence)

	(1) Physical violence	(2) Physical violence	(3) Physical violence
Twitter tweets	-0.0132	0.0215	0.0236
	(0.0269)	(0.0302)	(0.0312)
L.Twitter tweets		-0.0714***	-0.0683***
		(0.0271)	(0.0259)
L2.Twitter tweets			-0.00897
			(0.0279)
Constant	4.002***	4.009***	4.013***
	(0.0178)	(0.0182)	(0.0181)
Mean (Dep. Var)	4.000	4.002	4.005
St. Dv. (Dep. Var.)	3.715	3.716	3.718
State-Month fixed-effects	Yes	Yes	Yes
N	5751	5712	5673

Notes: The table shows the results from a linear regression of the number of Twitter tweets on the crime rate. The unit of analysis are week-state combinations. The outcome variable is the respective crime rate per 100,000 inhabitants per week and federal state, considering all crimes related to physical violence. We define physical violence as crimes related to murder/intentional manslaughter, aggravated assault, simple assault, kidnapping/abduction, in which the perpetrator and victim are of a different gender. The explanatory variable is the number of GBV-related tweets in a given federal state and week, divided by 100 cellphone internet plan subscriptions in a given federal state and year. In Column 1, we only consider the impact of Twitter tweets on the contemporaneous arrest per crime rate. In Column 2, we add Twitter tweets in the previous week, while in Column 3 we also consider Twitter tweets posted two weeks previously. We weight each cell by the population size of each federal state in the given year. We control for month of the year and state fixed effects. Month by state level clustered standard errors are repoted in parenthesis. Source: NIBRS, Twitter and ACS. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

-0.03 and are significant at the 10 and 5 percent significance levels respectively. A one unit increase in the number of tweets per 100 cellphone internet subscriptions decreases the emotional violence crime rate per 100,000 people by 0.02. These estimates correspond to a 2.469 percent increase for the second lag in Column 3. In conclusion, social movements have a significant effect on emotional GBV, similarly to our results on sexual and physical violence.

In summary, the previously reported overall impact on GBV persists for all three forms of GBV-related crimes investigated in this paper. The effect is largest in the case of sexual violence. Not only do Twitter social movements appear to decrease the prevalence of these crimes, but they also seem to be especially effective in altering the reporting rate

Table 11: The effect of social movements on GBV on crime rates per 100,000 inhabitants (Emotional violence)

	(1) Emotional violence	(2) Emotional violence	(3) Emotional violence
Twitter tweets	-0.0218*	-0.0116	-0.00416
	(0.0112)	(0.0110)	(0.0113)
L.Twitter tweets		-0.0216**	-0.0102
		(0.00985)	(0.0101)
L2.Twitter tweets			-0.0296**
			(0.0125)
Constant	1.200***	1.203***	1.206***
	(0.00791)	(0.00798)	(0.00801)
Mean (Dep. Var)	1.197	1.198	1.199
St. Dv. (Dep. Var.)	1.418	1.419	1.420
State-Month fixed-effects	Yes	Yes	Yes
N	5751	5712	5673

Notes: The table shows the results from a linear regression of the number of Twitter tweets on the crime rate. The unit of analysis are week-state combinations. The outcome variable is the respective crime rate per 100,000 inhabitants in a given week and federal state, considering all crimes related to emotional violence. We define emotional violence as intimidation, in which the perpetrator and victim are of a different gender. The explanatory variable is the number of GBV-related tweets in a given federal state and week, divided by 100 cellphone internet plan subscriptions in a given federal state and year. In Column 1, we only consider the impact of Twitter tweets on the contemporaneous arrest per crime rate. In Column 2, we add Twitter tweets in the previous week, while in Column 3 we also consider Twitter tweets posted two weeks previously. We control for month of the year and state fixed effects. Month by state level clustered standard errors are reported in parenthesis. Source: NIBRS, Twitter and ACS. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

of sexual violence. These results suggest that stigmas are more persistent in the case of sexual violence and that social movements address these stigmas. This insight aligns with the fact that many of the social movements investigated in this paper had a strong focus on sexual violence. Overall, our findings show that stigmatization and tabooing play an important role when analyzing the effect of GBV-related social movements on GBV-related crime rates.

Consequently, policymakers interested in decreasing the prevalence of GBV should address stigmatization, tabooing and silencing of this form of violence. It is recommended to design interventions addressing harmful gender norms.

6.3 Arrest per Crime Rates and the Role of Social Stigma

If social movements on Twitter increase social pressure for the authorities, they might result in better law enforcement, which again might then encourage victims to report GBV-related crimes more. To shed light on this potential channel, we analyze the impact of social movements on Twitter on arrest per crime rates. If the social pressure generated by these movements trickles down to the authorities, we would expect a positive and significant impact of Twitter tweets on arrests per crime rates. Analyzing this channel is relevant, as it has important policy implications for law enforcement.

Table 12 shows that there is limited evidence on a significant impact of social movements on Twitter on arrest per crime rates. Only the coefficient of the second lag in Column 3 is positive and significant at the 10 percent significance level. In terms of magnitude, the effect is equivalent to a 2.1 percent increase when compared to the mean.

Table 12: The effect of social movements on GBV on arrests per crime (GBV)

	(1) Arrests	(2) Arrests	(3) Arrests
No. of Twitter tweets	0.00190	0.00233	0.000266
	(0.00532)	(0.00439)	(0.00487)
L.No. of Twitter tweets		-0.00193	-0.00514
		(0.00360)	(0.00392)
L2.No. of Twitter tweets			0.00796*
			(0.00476)
Constant	0.388***	0.388***	0.388***
	(0.00155)	(0.00167)	(0.00167)
Mean (Dep. Var)	0.388	0.388	0.388
St. Dv. (Dep. Var.)	0.134	0.134	0.134
State-Month fixed-effects	Yes	Yes	Yes
N	5748	5708	5668

Notes: The table shows the results from a linear regression of the number of tweets on the arrest per crime rate. We define GBV-related crimes as physical, sexual, and emotional crimes, in which the perpetrator and victim are of opposite gender. The explanatory variable is the number of GBV-related tweets in the federal state during the week, divided by 100 cellphone internet plan subscriptions in the federal state in that year. The unit of analysis is the week by federal state. The first column only considers the impact of Twitter tweets on the contemporaneous arrest per crime rate. Column 2 adds Twitter tweets in the previous week, while Column 3 also considers Twitter tweets two weeks previously. We weight each cell by the population size of each federal state in the respective year. We control for month of the year and state fixed effects. Month by state level clustered standard errors are reported in parenthesis. Source: NIBRS, Twitter and ACS. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The positive coefficient on arrest per crime rates in Table 12 could mean that the social pressure generated via social movements on Twitter on GBV does trickle down to the authorities and leads to an increase in the arrests made relative to the crimes reported. Thus, law enforcement might increase, but the evidence is not strong. However, our results indicate that police are not responding to social movements on Twitter with backlash. Backlashes are a concern as less than 13 percent of full-time police officers in the United States are women. Consequently, the police force in the United States consists mostly of male and might consciously or subconsciously feel threatened by social movements about GBV. Backlash has been observed against gay police officers disclosing their homosexuality (Rumens and Broomfield, 2012). Mandatory and preferential arrest policies related to domestic violence cases have also resulted in backlash for victims who were arrested along with their batterers (Finn and Bettis, 2006). In a similar fashion, Amaral et al. (2021) demonstrate that while the introduction of women's police stations in India might be efficient in reporting more cases of GBV to the authorities, this might be hampered by backlash from male police officers.

We also investigate this channel by subtype of GBV, as it can shed light on the extent to which social stigma and taboo play a role in law enforcement. If arrests made by the police are affected by these factors and social movements on Twitter impact them, we would expect a varying effect of social movements on Twitter on arrest per crime rates. We detail the results in Appendix C.4. We do not find compelling evidence that social movements on Twitter alter the arrest per crime rate in the case of physical or emotional violence. While coefficients on sexual violence are significant, they go into opposite directions.

6.4 Analyzing the Text of Tweets

We next explore the extent to which the text of the tweets plays a role in our findings. The impact of tweets in favor of GBV-related movements might differ from tweets opposing them. In addition, the polarity of written text might also play a role. More extreme tweets, for example, could have a stronger impact than less extreme tweets. To investigate this in more detail, we analyze how the average polarity of what is written affects crime rates as well as arrest per crime rates. For this purpose, we estimate our main regression specification using the average compound score in a given week and federal state as our main explanatory variable and the crime rate and arrest per crime rate as our outcome variables. If the polarity of tweets matters, we expect the coefficient to be significant. The results presented in Table 13 indicate that the polarity of tweets do not play a

significant role in their impact on crime rates. This means that the sheer magnitude of social movements on Twitter is more important than their content.

Table 13: The effect of the polarity of GBV-related tweets on crime rates per 100,000 inhabitants (GBV)

	(1) Crime rate	(2) Crime rate	(3) Crime rate
Compound Score	-0.000648	-0.0214	-0.0210
	(0.000976)	(0.0312)	(0.0357)
L.Compound Score		0.0210	0.0327
		(0.0315)	(0.0271)
L2.Compound Score			-0.0121
			(0.0344)
Constant	5.962***	5.963***	5.968***
	(0.0258)	(0.0259)	(0.0258)
Mean (Dep. Var)	5.961	5.963	5.968
St. Dv. (Dep. Var.)	5.508	5.508	5.512
State-Month fixed-effects	Yes	Yes	Yes
N	5751	5712	5673

Notes: The table shows the results from a linear regression of the average polarity of Twitter tweets on the crime rate. The unit of analysis is the week-state combination. We deduce the average polarity of Twitter tweets by employing a VADER Sentiment Analysis. This text analysis method is a social media sentiment analysis method and approximates the average polarity of social media text by a compound score. The compound score is a score with values ranging from -1 to 1. A value of -1 represents text in complete disagreement while a value of 1 represents text in complete agreement. For methodological details of the VADER Sentiment Analysis see Appendix B.3. The outcome variable is the respective crime rate per 100,000 inhabitants per week and federal state, considering all GBV-related crimes. We define GBV-related crimes as physical, sexual, and emotional crimes, in which the perpetrator and victim are of a different gender. In Column 1, we only consider the impact of Twitter tweets on the contemporaneous crime rate. In Column 2, we add Twitter tweets of the previous week, while in Column 3 we also consider Twitter tweets posted two weeks previously. We weight each cell by the population size of each federal state in a given year. We control for month of the year and state fixed effects. Month by state level clustered standard errors are reported in parenthesis. Source: NIBRS, Twitter and ACS data. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

7 Conclusion

This paper examines whether social movements on Twitter impact crime rates of gender-based violence (GBV). We utilize text analysis and machine learning methods to create a novel dataset that measures online conversations about GBV on Twitter. We take

advantage of the high frequency of our data and conduct regressions at the state by week level in the US. We introduce a number of fixed effects to account for potential confounding factors. We also include lagged coefficients to allow for potential adjustment times of human behavior and to allow a causal interpretation of our results.

We find that social movements on Twitter lead to a decrease in GBV-related crime rates of about 1 percent. We provide evidence that behavioral changes among perpetrators of GBV most likely drive our results. If, in addition, victims feel empowered and would be more likely to report GBV, our coefficients are lower bound estimates of the true underlying effect. Moreover, we show that the impact of Twitter tweets on crime rates is most pertinent in the case of sexual violence. This could mean that stigmatization, tabooing and silencing of sexual violence were especially persistent. Social movements on Twitter most likely removed some of these barriers.

Furthermore, analysis of the tweets' text using sentiment analysis shows that the polarity of what is written does not play a significant role. Consequently, sheer scale of social movements is more important than their content. Moreover, we find limited evidence on significant effects of social movements on GBV-related arrest per crime rates. The police force possibly increases its law enforcement as a result of rising social pressure, but the evidence supporting this channel is limited.

We conduct an event study to shed further light on the causality of our main findings. The analysis confirms that social movements on Twitter have a crime-decreasing impact. Moreover, we run placebo regressions to investigate the potential existence of unobservable confounding factors. Our results are robust to using non-GBV related crime rates as outcome variables. Lastly, there is no evidence of selection bias with respect to observable personal characteristics of Twitter users and victims of GBV.

One important limitation of this paper is that there might be spillover effects of Twitter tweets between states. While previous research shows that geographic networks play an important role in Twitter tweets (Comito (2021); Hawelka et al. (2014)), information spreads quickly across regions. This might confound our empirical strategy, which relies on geographic variations. Still, allowing for spillover effects between neighboring states confirms our main findings, and reinforces them for crime rates.

The pattern of results presented in this paper makes the case for Twitter platforms facilitating the signaling of social norms and having significant effects on offline behaviors. The evidence in this paper points towards Twitter tweets increasing social pressure and costs. These findings have important policy implications. Institutions interested in decreasing the prevalence of GBV should explore the potential of social media platforms

for this purpose. They can take advantage of these platforms to create informal support networks for those who experience GBV or signal social norms. Our findings port to the importance of social networks in creating real societal changes. Moreover, our paper generates novel insights on stigmatization, tabooing and silencing playing an important role when it comes to reporting related to GBV. Hence, policymakers should design strategies to address these barriers and facilitate reporting and conversation on GBV.

Further research should investigate how to fully disentangle the effect of crime perpetration and crime reporting on reported crime rates. Moreover, it would be interesting to examine whether our results, focusing exclusively on social movements on Twitter, differ from other social media platforms, such as Facebook. Future research could also explore alternative levels of variation than geographic variation for the empirical design. Lastly, future studies could investigate the extent to which traditional media coverage of social movements on Twitter influences our results.

References

Abrams, David S (2021). "COVID and crime: An early empirical look". *Journal of public economics* 194, p. 104344.

Agarwal, Saharsh and Ananya Sen (2022). "Antiracist Curriculum and Digital Platforms: Evidence from Black Lives Matter". *Management Science* 68 (4), pp. 2932–2948.

Agranov, Marina, Matt Elliott, and Pietro Ortoleva (2021). "The importance of Social Norms against Strategic Effects: The case of COVID-19 vaccine uptake". *Economics Letters* 206, p. 109979.

Agüero, Jorge M (2021). "COVID-19 and the rise of intimate partner violence". *World development* 137, p. 105217.

Aizer, Anna (2010). "The gender wage gap and domestic violence". *American Economic Review* 100 (4), pp. 1847–59.

Amaral, Sofia, Siddhartha Bandyopadhyay, and Rudra Sensarma (2015). "Employment programmes for the poor and female empowerment: The effect of NREGS on gender-based violence in India". *Journal of interdisciplinary economics* 27 (2), pp. 199–218.

Amaral, Sofia, Sonia Bhalotra, and Nishith Prakash (2021). "Gender, crime and punishment: Evidence from women police stations in india".

Balestrino, Alessandro (2008). "It is a theft but not a crime". *European Journal of Political Economy* 24 (2), pp. 455–469.

Bandyopadhyay, Debasis, James Allan Jones, and Asha Sundaram (2020). "Gender Bias and Male Backlash as Drivers of Crime Against Women: Evidence from India". *The University of Auckland Business School Research Paper Forthcoming*.

Becker, Gary S et al. (1995). "The economics of crime". *Cross Sections* 12 (Fall), pp. 8–15.

Berniell, Inés and Gabriel Facchini (2021). "COVID-19 lockdown and domestic violence: Evidence from internet-search behavior in 11 countries". *European Economic Review* 136, p. 103775.

Bhalotra, Sonia et al. (2021a). "Intimate partner violence: The influence of job opportunities for men and women". *The World Bank Economic Review* 35 (2), pp. 461–479.

Bhalotra, Sonia et al. (2021b). "Job displacement, unemployment benefits and domestic violence". *CEPR Discussion Paper No. DP16350*.

Bobonis, Gustavo J, Melissa González-Brenes, and Roberto Castro (2013). "Public transfers and domestic violence: The roles of private information and spousal control". *American Economic Journal: Economic Policy* 5 (1), pp. 179–205.

Brassiolo, Pablo (2016). "Domestic violence and divorce law: When divorce threats become credible". *Journal of Labor Economics* 34 (2), pp. 443–477.

Bremer, Björn, Swen Hutter, and Hanspeter Kriesi (2020). "Dynamics of protest and electoral politics in the Great Recession". *European Journal of Political Research* 59 (4), pp. 842–866.

Bullinger, Lindsey Rose, Jillian B Carr, and Analisa Packham (2021). "COVID-19 and crime: Effects of stay-at-home orders on domestic violence". *American Journal of Health Economics* 7 (3), pp. 249–280.

Bursztyn, Leonardo, Georgy Egorov, and Stefano Fiorin (2020). "From extreme to mainstream: The erosion of social norms". *American economic review* 110 (11), pp. 3522–48.

Bursztyn, Leonardo et al. (2021). "Persistent political engagement: Social interactions and the dynamics of protest movements". *American Economic Review: Insights* 3 (2), pp. 233–50.

Chakraborty, Tanika et al. (2018). "Stigma of sexual violence and women's decision to work". *World Development* 103, pp. 226–238.

Chernin, Yulia and Yaron Lahav (2014). ""The people demand social justice" a case study on the impact of protests on financial markets". *Accounting, Economics and Law* 4 (2), pp. 99–121.

Clarke, Damian and Kathya Tapia-Schythe (2021). "Implementing the panel event study". *The Stata Journal* 21 (4), pp. 853–884.

Comito, Carmela (2021). "How covid-19 information spread in us the role of twitter as early indicator of epidemics". *IEEE Transactions on Services Computing*.

Cools, Sara and Andreas Kotsadam (2017). "Resources and intimate partner violence in Sub-Saharan Africa". *World Development* 95, pp. 211–230.

Cooper, Jasper, Donald P Green, and Anna M Wilke (2020). "Reducing Violence against Women in Uganda through Video Dramas: A Survey Experiment to Illuminate Causal Mechanisms". *AEA Papers and Proceedings*. Vol. 110, pp. 615–19.

Cullen, Claire (2020). "Method matters: Underreporting of intimate partner violence in Nigeria and Rwanda". *World Bank Policy Research Working Paper* (9274).

Dave, Dhaval M et al. (2020). *Black lives matter protests and risk avoidance: The case of civil unrest during a pandemic*. Tech. rep. National Bureau of Economic Research.

Delaporte, Magdalena and Francisco Pino (2022). "Female Political Representation and Violence Against Women: Evidence from Brazil". *IZA Discussion Paper*.

Delker, Brianna C et al. (2020). "Who has to tell their trauma story and how hard will it be? Influence of cultural stigma and narrative redemption on the storying of sexual violence". *Plos one* 15 (6), e0234201.

Duvvury, Nata et al. (2013). "Intimate partner violence: Economic costs and implications for growth and development". *World Bank. License: CC BY 3.0 IGO.*

ElSherief, Mai, Elizabeth Belding, and Dana Nguyen (2017). "# notokay: Understanding gender-based violence in social media". *Eleventh International AAAI Conference on Web and Social Media*.

Erten, Bilge and Pinar Keskin (2018). "For better or for worse?: Education and the prevalence of domestic violence in turkey". *American Economic Journal: Applied Economics* 10 (1), pp. 64–105.

Falk, Armin and Urs Fischbacher (2002). ""Crime" in the lab-detecting social interaction". *European Economic Review* 46 (4-5), pp. 859–869.

Fernández-Fontelo, Amanda et al. (2019). "Untangling serially dependent underreported count data for gender-based violence". *Statistics in medicine* 38 (22), pp. 4404–4422.

Finn, Mary A and Pamela Bettis (2006). "Punitive action or gentle persuasion: Exploring police officers' justifications for using dual arrest in domestic violence cases". *Violence against women* 12 (3), pp. 268–287.

Fitzgerald, Louise F and Lilia M Cortina (2018). "Sexual harassment in work organizations: A view from the 21st century."

Folke, Olle et al. (2020). "Sexual harassment of women leaders". *Daedalus* 149 (1), pp. 180–197.

Fry, M Whitney, Asheley C Skinner, and Stephanie B Wheeler (2019). "Understanding the relationship between male gender socialization and gender-based violence among refugees in Sub-Saharan Africa". *Trauma, Violence, & Abuse* 20 (5), pp. 638–652.

Gagliarducci, Stefano and M Daniele Paserman (2012). "Gender interactions within hierarchies: evidence from the political arena". *The Review of Economic Studies* 79 (3), pp. 1021–1052.

Gangadharan, Lata et al. (2019). "Female leaders and their response to the social environment". *Journal of Economic Behavior & Organization* 164, pp. 256–272.

González, Libertad and Núria Rodríguez-Planas (2020). "Gender norms and intimate partner violence". *Journal of Economic Behavior & Organization* 178, pp. 223–248.

Guarnieri, Eleonora and Helmut Rainer (2021). "Colonialism and female empowerment: A two-sided legacy". *Journal of Development Economics* 151, p. 102666.

Harris, Wesley Eugene (2017). "The Effect of Stigma on Intimate Partner Violence Reporting Among Men Who Have Sex with Men". *East Tennessee State University*.

Hawelka, Bartosz et al. (2014). "Geo-located Twitter as proxy for global mobility patterns". *Cartography and Geographic Information Science* 41 (3), pp. 260–271.

Hutto, Clayton and Eric Gilbert (2014). "Vader: A parsimonious rule-based model for sentiment analysis of social media text". *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 8. 1.

Iyer, Lakshmi et al. (2012). "The power of political voice: women's political representation and crime in India". *American Economic Journal: Applied Economics* 4 (4), pp. 165–93.

Joseph, George et al. (2017). "Underreporting of gender-based violence in Kerala, India: An application of the list randomization method". *World Bank Policy Research Working Paper* (8044).

Khatua, Aparup, Erik Cambria, and Apalak Khatua (2018). "Sounds of silence breakers: Exploring sexual violence on twitter". *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, pp. 397–400.

LaFree, Gary and Kriss A Drass (1996). "The effect of changes in intraracial income inequality and educational attainment on changes in arrest rates for African Americans and whites, 1957 to 1990". *American Sociological Review*, pp. 614–634.

Lee, Lung-fei and Jihai Yu (2010). "Estimation of spatial autoregressive panel data models with fixed effects". *Journal of econometrics* 154 (2), pp. 165–185.

Levitt, Steven D (1998). "Why do increased arrest rates appear to reduce crime: deterrence, incapacitation, or measurement error?" *Economic inquiry* 36 (3), pp. 353–372.

Levy, Ro'ee (2021). "Social media, news consumption, and polarization: Evidence from a field experiment". *American economic review* 111 (3), pp. 831–70.

Levy, Roee and Martin Mattsson (2021). "The effects of social movements: Evidence from# MeToo". *Available at SSRN 3496903*.

Li, Tianshu, Sonal Pandya, and Sheetal Sekhri (2019). *Repelling Rape: Foreign Direct Investment Empowers Women*. Tech. rep. Working Paper.

Li, Susan (2018). *Multi-Class Text Classification Model Comparison and Selection*. `https://towardsdatascience.com/multi-class-text-classification-model-comparison-and-selection-5eb066197568`. Accessed: 2018-25-09.

Linos, Natalia et al. (2013). "Influence of community social norms on spousal violence: a population-based multilevel study of Nigerian women". *American journal of public health* 103 (1), pp. 148–155.

Matta, Samer, Michael Bleaney, and Simon Appleton (2021). "The economic impact of political instability and mass civil protest". *Economics & Politics*.

McCart, Michael R, Daniel W Smith, and Genelle K Sawyer (2010). "Help seeking among victims of crime: A review of the empirical literature". *Journal of Traumatic Stress: Official Publication of The International Society for Traumatic Stress Studies* 23 (2), pp. 198–206.

Miller, Amalia R and Carmit Segal (2019). "Do female officers improve law enforcement quality? Effects on crime reporting and domestic violence". *The Review of Economic Studies* 86 (5), pp. 2220–2247.

Mishra, Ankita, Vinod Mishra, and Jaai Parasnis (2021). "The asymmetric role of crime in women's and men's labour force participation: Evidence from India". *Journal of Economic Behavior & Organization* 188, pp. 933–961.

Morrison, Andrew, Mary Ellsberg, and Sarah Bott (2007). "Addressing gender-based violence: a critical review of interventions". *The World Bank Research Observer* 22 (1), pp. 25–51.

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

Ouedraogo, Rasmane and David Stenzel (2021). "The Heavy Economic Toll of Gender-based Violence: Evidence from Sub-Saharan Africa". *IMF Working Papers* 2021 (277).

Palermo, Tia, Jennifer Bleck, and Amber Peterman (2014). "Tip of the iceberg: reporting and gender-based violence in developing countries". *American journal of epidemiology* 179 (5), pp. 602–612.

Parkhi, Omkar M, Andrea Vedaldi, and Andrew Zisserman (2015). "Deep face recognition". *British Machine Vision Association*.

Rumens, Nick and John Broomfield (2012). "Gay men in the police: Identity disclosure and management issues". *Human Resource Management Journal* 22 (3), pp. 283–298.

Serengil, Sefik Ilkin and Alper Ozpinar (2020). "LightFace: A Hybrid Deep Face Recognition Framework". *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*. IEEE, pp. 23–27. DOI: 10.1109/ASYU50717.2020.9259802.

Siddique, Zahra (2022). "Media-Reported Violence and Female Labor Supply". *Economic Development and Cultural Change* 70 (4), pp. 000–000.

SimpleMaps (2012). *United States Cities Database*. data retrieved from SimpleMaps, https://simplemaps.com/data/us-cities.

Standish, Katerina (2014). "Understanding cultural violence and gender: honour killings; dowry murder; the zina ordinance and blood-feuds". *Journal of Gender Studies* 23 (2), pp. 111–124.

Statista (2022). *Leading countries based on number of Twitter users as of October 2021.* data retrieved from Statista, `https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/`.

Swamy, Vighneswara (2014). "Financial inclusion, gender dimension, and economic impact on poor households". *World development* 56, pp. 1–15.

Tur-Prats, Ana (2019). "Family types and intimate partner violence: A historical perspective". *Review of Economics and Statistics* 101 (5), pp. 878–891.

UN Women (June 21, 2021). "Facts and figures: Ending violence against women". *UN Women.*

UNHCR (Jan. 29, 2022). "Gender-based Violence". *UNHCR.*

Viscusi, W Kip, Joel Huber, and Jason Bell (2011). "Promoting recycling: private values, social norms, and economic incentives". *American Economic Review* 101 (3), pp. 65–70.

Walby, Sylvia and Philippa Olive (2014). "Estimating the costs of gender-based violence in the European Union". *European Institute for Gender Equality.*

Welsh, Sandy (1999). "Gender and sexual harassment". *Annual review of sociology* 25 (1), pp. 169–190.

Wen, Jinglin (2021). "Female Mayors and Violence Against Women: Evidence from the US".

Yilmaz, Okan (2018). "Female autonomy, social norms and intimate partner violence against women in Turkey". *The Journal of Development Studies* 54 (8), pp. 1321–1337.

Zhuravskaya, Ekaterina, Maria Petrova, and Ruben Enikolopov (2020). "Political effects of the internet and social media". *Annual Review of Economics* 12, pp. 415–438.

Appendix A Detailed Description of Crime Data

The NIBRS is an incidence-based reporting system managed by the FBI for police-reported crimes in the US. The system collects a variety of information on each incident reported to the police, such as the nature of the offense, characteristics of the victim(s) and offender(s), and the date and location of the incident. The system collects information on 22 offense categories covering 46 specific crimes. The data is collected through reports submitted to the FBI by city, county and state law enforcement agencies. The submission is voluntary and takes place monthly. There are 6,251 law enforcement agencies included in the data for 2014, 6,278 in the data for 2015, and 6,570 in the data for 2016. This shows that there was only a marginal increase in the number of agencies that report to the FBI over time. We are therefore confident that our results are robust to significant changes in the pattern of reporting law enforcement agencies. In total, approximately 30 percent of all law enforcement agencies in the United States report to the NIBRS. There is an estimated total of 18,000 agencies in the US (Link: https://ucr.fbi.gov/nibrs/2019).

We use data processed by the Inter-university Consortium for Political and Social Research (ICPSR). The ICPSR provides the NIBRS data in four different formats, namely at the crime incident, victim, offender, and arrestee level. We use the crime incident dataset for the period 2014-2016. This dataset consists of one record per incident and a total of 4,919,278 cases in 2014, 4,986,608 cases in 2015, and 5,293,536 cases in 2016. The ICPSR merges information on the victim(s) and offender(s) for each incident based on the uniquely identified incidence number, resulting in a total of 390 variables.

The crime data has several data limitations. First, the reporting to the FBI by the federal states is voluntary. As a consequence, not all federal states are part of the underlying dataset.[26] Moreover, similar to other papers using crime data, it is not possible for us to distinguish between reporting rates and actual underlying crime rates. Crime rates reported to the police can change due to changes in crime penetration or changes in reporting behavior. The first channel reflects behavioral adjustments by perpetrators. The second channel relies on victims changing their behavior. This is a limitation for the research question at hand, as an identification of channels is crucial to the interpretation of our findings. If social movements cause perpetrators to adapt their behavior, social

[26]More concretely speaking, only 39 out of 50 federal states form part of the dataset. Additionally, while the dataset covers all weeks (156), several of the week-state combinations are missing from the dataset. While a dataset consisting of 156 weeks and 39 states should result in a total number of 6,084 observations, there are only 5,907 observations in the underlying dataset. This means that 177 week-state combinations are missing. Adding the missing observations from the 11 federal states not reporting their crime data to the FBI, we end up with 1,894 missing week-state combinations.

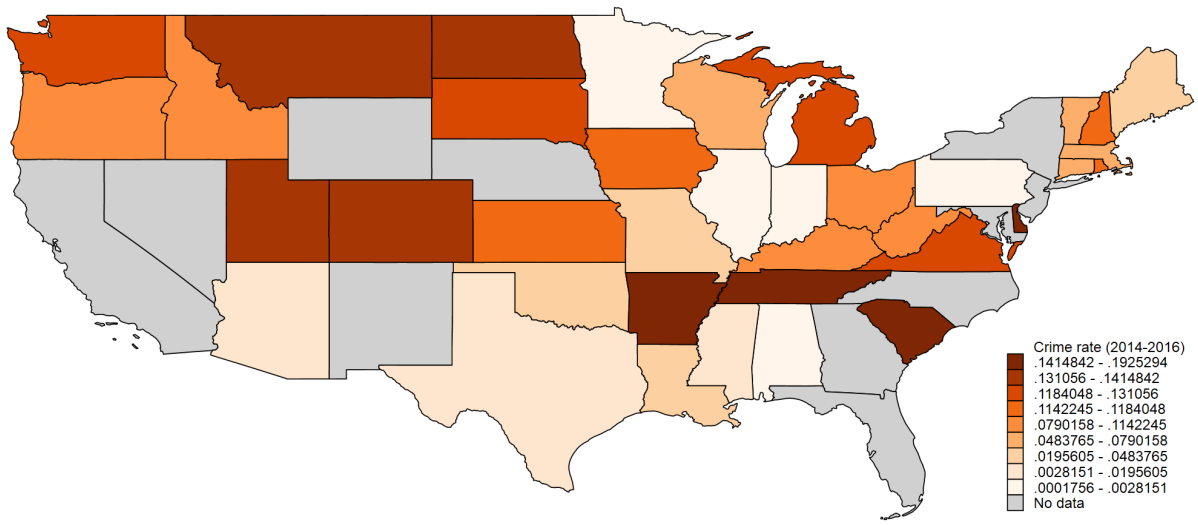
movements on Twitter mirror social pressure. If, on the other hand, social movements change victims' behavior, this would be evidence of victims' empowerment. We address these concerns by additional analyses in Section 6.

The fact that the agglomerated GBV rate is lower in conservative states, like Texas, may be unexpected. There could be two reasons for this. First, in conservative settings, the rate of unreported GBV might be especially high, as victims might be less empowered. Authorities might also be less likely to respond to reports on GBV, which could create further incentives to not report these crimes. Similar patterns have been observed in Nigeria and Rwanda for the reporting of IPV (Cullen, 2020). Next, the spatial patterns could emerge due to the FBI's data gathering process. As indicated in Section 3, the NIBRS relies on voluntary and monthly crime report transmission by law enforcement agencies at the city, county, and state level. It could be that a lower number of law enforcement agencies in states with lower aggregated GBV rates participate in the FBI's Uniform Crime Reporting Program. A concern would be that the number of law enforcement agencies in certain states develops differently than in other federal states, which could bias our results. To show that the lack of spatial pattern is not related to the GBV categories investigated, we present a similar map for non-GBV categories in Figure 7. The distribution of aggregated non-GBV-related crime rate is very similar to that of aggregated GBV-related crime rates. This suggests that the spatial distribution is due to reporting. We conclude that these spatial patterns are unlikely driven by the GBV categories.

We are also interested in the effect of social movements on Twitter on behavioral changes by authorities. Accordingly, we investigate the impact of Twitter tweets on GBV-related arrests. Although there are limitations to using arrests to measure behavior changes by authorities, this indicator is widely for this purpose in the literature (examples are LaFree and Drass (1996), Levitt (1998), Bullinger et al. (2021), and Abrams (2021)). An important empirical limitation, however, is that arrests could be driven by pure behavioral changes by authorities, or by a change in the number of crimes committed. To account for this limitation, we look at arrest per crime rates instead of absolute numbers of arrests.

To study the impact of social movements on Twitter on arrests, we exploit the arrestee-level extract file of the NIBRS data. The arrestee-level extract file contains one record for each arrestee recorded in NIBRS for arrest dates in a given year, regardless of the date of the incident. There are 3,174,660 records in 2014, 3,174,815 records in 2015, and 3,308,784 records in 2016. We then prepare our variables of interest in a similar manner

Figure 7: Aggregated non-GBV-related crime rate (2014-2016) over the population in 2014



Notes: The map depicts the aggregate number of non-GBV-related crimes reported to the police for the years 2014-2016 divided by population estimates from 2014 at the federal state level in the US. The graph excludes Alaska, Hawaii, and Puerto Rico. Darker colors indicate higher aggregated GBV-related crime rates. Source: NIBRS and US Census Bureau.

to how the variables from the crime-incident roster of the NIBRS are formatted. That is, we uniquely identify arrests related to GBV-related crimes, as well as the respective subcategories (physical, sexual and emotional violence).

Appendix B Twitter Dataset Creation

B.1 The hashtag-based Approach

We access our Twitter tweets by creating an Academic Developer Account and accessing tweets via the Twitter Full Archive Search API V2. The API has a rate limit of 10 million tweet per month and 150,000 tweets per 15 minutes. We take advantage of the Twarc2 command line tool and Python library. We define a customized search query, restricting our time frame to the year 2017 and certain keywords. The resulting data is organized in Json (JavaScript Object Notation) Objects, such as a *User* object or a *Tweet* object. Each object comes along with *attributes* describing the Json Object, such as the author, the actual message, a unique ID, a timestamp of when it was posted, and sometimes geo metadata about the location of the user or tweet. There are also *entity* objects associated

with some tweets, such as hashtags, mentions, media, and links. A single tweet can have up to 150 attributes coming along with the actual text. There are four overall JSON Keys: Data, Includes, Error, Meta. Each comes with several nested JSON Objects.

In order to narrow down our keywords on which we filter our API query, we ask ourselves the question of how to best proxy the conversation on GBV on Twitter. For this purpose, we identify 10 of the biggest movements related to GBV on Twitter. We then extract all tweets on the first 4 weeks of each of the 10 movements. We select the following movements:

- *#aufschrei*: 24th of Jan. 2013, 39,130 tweets in 1 month
- *#yesallwomen*: 24th of May 2014, 330,428 tweets in 1 month
- *#rapecultureiswhen*: 25th of May 2014, 6,698 tweets in 1 month
- *#WhyIStayed*: 8th of Sept. 2014, 37,915 tweets in 1 month
- *#everydaysexism*: 16th of April 2015, 9,175 tweets in 1 month
- *#NoWomanEver*: 18th of June 2016, 15,643 tweets in 1 month
- *#notokay*: 7th of October 2016, 18,885 tweets in 1 month
- *#metoo*: 14th of October 2017, 278,769 tweets in 1 month
- *#MeAt14*: 10th of November 2017, 7,653 tweets in 1 month
- *#whyididntreport*: 21st of September 2018, 68,707 tweets in 1 month

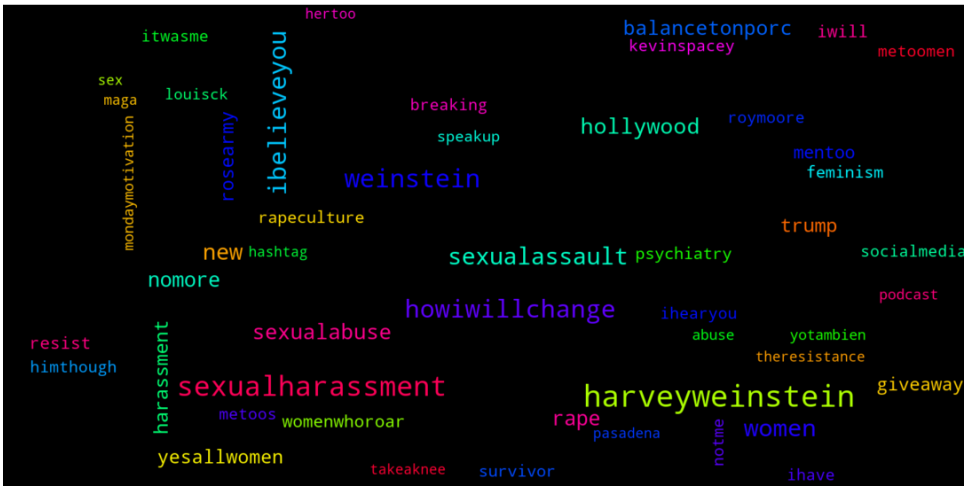
We end up with a total of 813,003 tweets. From this sample, we extract a list of all hashtags mentioned in these close to one million tweets and end up with a list of 73,430 unique hashtags. Figure A1 shows the 500 most used hashtags during the first month of each of these 10 debates. Figure A2 further restrict this list to the 50 most used hashtags.

The figures illustrates that not all of these hashtags uniquely identify topics related to GBV. One example is the hashtag *#hollywood*. If we would include this hashtag in our query as a keyword, we would get a large variety of tweets not related to GBV but other topics. We therefore decide to only include hashtags in our query, which answer the following question: "*looking at this hashtag standing on its own, does it uniquely identify as a hashtag related to GBV?*" with yes. This question is therefore our first selection criteria. Going through all 73,430 unique hashtags and answering the above question for each one of them would take a very long time. This is why we decide to train an algorithm, which conducts this task for us. We train our algorithm based on all hashtags mentioned in the first month of the *metoo* movement, a total of 32,487 different hashtags. This means that we go through each one of the 32,487 different hashtags (and 278,769 tweets) and manually code the data. When doing this, we notice that we can come up with 10 overall topics categorizing these hashtags, which are:

Figure A1: Top 500 most mentioned hashtags in the first month of 10 Twitter movements

Notes: The graph plots a word cloud of the Top 500 most mentioned hashtags in the first month of 10 social movements on Twitter. The size of each word represents its relative occurrence when compared to all other words in the sample. Source: Twitter data.

Figure A2: Top 50 most mentioned hashtags in the first month of 10 Twitter movements



Notes: The graph plots a word cloud of the Top 50 most mentioned hashtags in the first month of 10 social movements on Twitter. The size of each words represents its relative occurrence when compared to all other words in the sample. Source: Twitter data.

- **GBV/Sexism:** #sexualharassment, #YesAllWomen, #SexualAssault, #rapeculture, #sexism, #nomeansno, #vaw, #domesticviolence
- **Misogynie:** #misogyny, #misogynist, #MisogynyIsReal, #antimisogynoir, #endmisogyny
- **Support:** #IBelieveYou, #IHearYou, #BelieveWomen, #believesurvivors, #IStandWithYou
- **Gender Equality:** #GenderEquality, #GenerationEquality, #inequality, #Womensrights, #FuckBoysWillBeBoys
- **Feminism:** #Feminism, #feminist, #howtospotafeminist, #ThisIsWhyINeedFeminism, #FeministFriday
- **Non-White Feminism:** #solidarityisforwhitewomen, #whitefeminism, #Womanofcolor, #BelieveBlackWomen
- **Anti-Feminism:** #womenagainstfeminism, #antifeminismus, #NotMyFeminism, #FeminismIsCancer
- **Masculinity:** #teachourboys, #menneed2dobetter, #mencanendrape, #iwillteachmysonbetter, #fragilemasculinity
- **Misogynist, silencing:** #GoGetRaped, #ShutUp, #AttentionWhores, #GetOverIt, #Boyswillbeboys
- **Others:** #Hollywood, #news, #assault, #women, #psychiatry, #giveaway

We also apply an unsupervised machine learning algorithm, the *Latent Dirichlet Allocation* (LDA), in order to check if manual coding is needed or can be done by an algorithm in the first place. We conclude that the unsupervised machine learning algorithm does a fairly poor job, also when varying the number of topics. Therefore, we rely on our supervised machine learning algorithm.

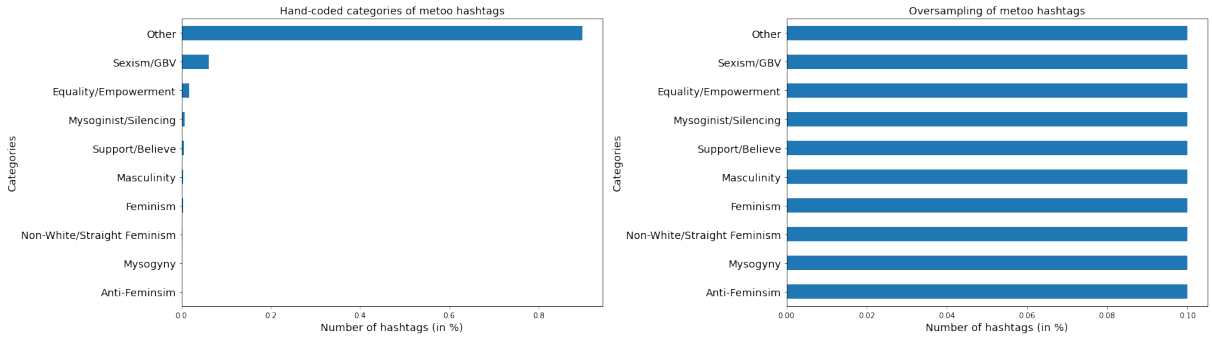
As a next step we use the manually coded *#metoo* hashtags to train the remaining 534,234 tweets (40,943 hashtags) on the remaining 9 Twitter movements. We split our Training data (the 32,487 manually coded *metoo* hashtags) into a training and testing dataset by a 50-50-ratio. We feed the training data (16,243 hashtags) into the model for training. We test the performance of the model through running the model on the test data (the remaining 16,243 hashtags). We then measure the performance of the model through comparing the *predicted* categories of the *testing* data with the manually coded categories (Model Accuracy), but also based on alternative performance measures, such as the Precision Score, Recall Score, F1-Score, and Confusion Matrix. We then run the model on the new, unseen, not hand-coded data (the 40,943 remaining hashtags). We also conduct a cross validation (10-Fold Cross Validation) and tune our model parameters

through a grid search. We spot-check a variety of different algorithms and compare their performance to each other:

- Multinomial Naive Bayes classifier
- K-Nearest Neighbor Classification
- Regularized Logistic Models
- Support Vector Machine
- Stochastic Gradient Descent
- Decision Tree Classifier
- Ensemble Methods:
  - Random Forest Classifier
  - AdaBoost Classifier

We encounter an Imbalanced Classification Problem as more than 90 percent of our hashtags belong to one category, our Garbage Category (see Figure A3). In order to address this problem, we oversample the minority classes (all but "Other"). This means that we draw random samples (Fraction > 1) from each of the minority classes with equal probability weighting (see Figure A4).

Figure A3: Handcoded tweets by category Figure A4: Oversampled tweets by category



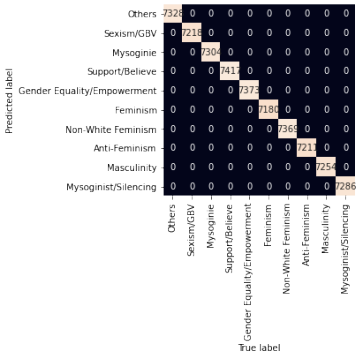
Notes: The left panel shows the share of hashtags used within the first month of the *#metoo* movement in October 2017 in 10 manually defined categories. The right panel shows the same share after oversampling the minority categories. Source: Twitter data.

Figure A5 to A12 present the confusion matrices and accuracy scores. The fastest algorithms are (in descending order) the Stochastic Gradient Descent (SGD), Decision Tree, Regularized logistic regression and Naive Bayes Classifier. Based on our performance measures, we choose two different classifier for our prediction: SGD alias linear SV and Logistic Regression. The SGD Classifier is characterized by high accuracy, high speed and

is widely accepted as one of the best text data classifiers (Li, Susan, 2018). The Logistic Regression Classifier is characterized by high accuracy, high speed, easy to interpret and use, and is widely used in economics.

Based on our different performance measures, the Linear Support Vector Machine is the best performing classifier for our underlying classification problem. We then run the linear SVM algorithm on the remaining 40,943 hashtags that are not hand-coded. We evaluate the performance of the prediction made by the algorithm on the unlabeled data by comparing a ten percent sample of the predicted classes to manually coded classifications of this same sample done by an independent research assistant.

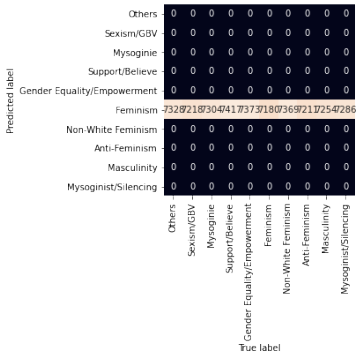
Figure A5: Regularized logistic regression



a

[a]Penalty is elastic net with L-ratio of 0.5. The solver is the saga optimizer. Accuracy is 100.

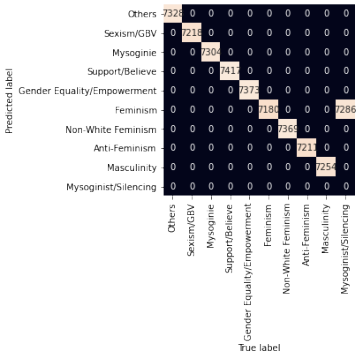
Figure A6: Support Vector Machine



a

[a]Kernel is Rbf, gamma is auto, L-2 panelty. Accuracy is 9.8.

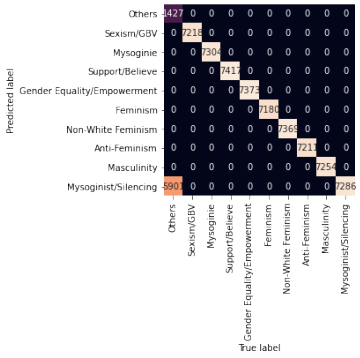
Figure A7: Naive Bayes



a

[a]All default values. Accuracy is 90.

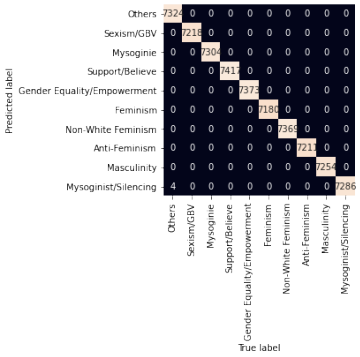
Figure A8: KNN Classifier



a

[a]N=27 (Square root of test data lenght), Metric is the Euclidean distance. Rest is default values. Accuracy is 91.9

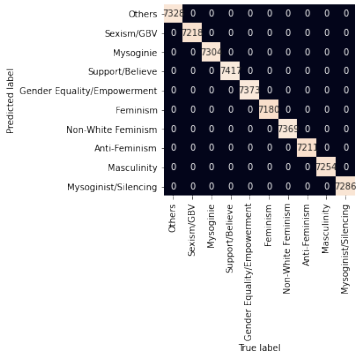
Figure A9: Stochastic Gradient Classifier



a

[a]Linear SVM with L2-penalty and default settings. Accuracy is 100.

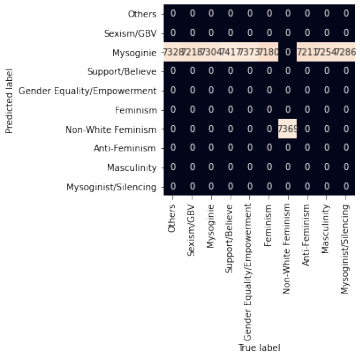
Figure A10: Decision Tree Classifier



a

[a]Quality of split measured by entropy. Rest is default. Accuracy is 100.

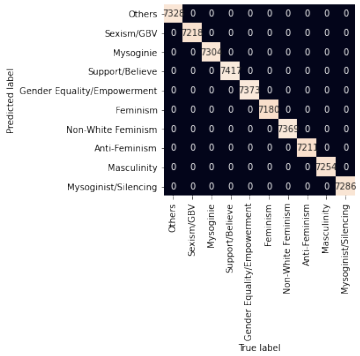
Figure A11: AdaBoost Classifier



a

[a]Accuracy is 20.11.

Figure A12: Random Forest Classifier



a

[a]All specifications are default (100 parallel trees, criteria for split is Gini impurity, no pruning and 2 parallel jobs). Accuracy is 100.

After applying our supervised learning algorithm, we restrict our list of hashtags to category 1 (GBV). Lastly, we combine the hand-coded hashtags on GBV with the machine-coded hashtags on GBV, resulting in a total of 2,009 hashtags complying with our first selection criteria. Due to the character limit of 1,024 characters per request of the Twitter API query, we develop a second selection criteria. We decide to use relevance as our second selection criteria. Therefore, we order our sample of 2,009 hashtags by relevance (number of occurrences) and only include the top 62 most used hashtags in our final query. These hashtags are as follows: #yesallwomen, #metoo, #whyididntreport, #aufschrei, #whyistayed, #notokay, #everydaysexism, #meat14, #rapecul-

tureiswhen, #yesallmen, #rapeculture, #sexismus, #sexualharassment, #sexualassault, #nomoore, #domesticviolence, #rape, #vaw, #bringbackourgirls, #sexualabuse, #harassment, #sexism, #balancetonporc, #abuse, #whywomendontreport, #nomeansno, #metoos, #yotambien, #roymoorechildmolester, #notok, #hertoo, #metoodebatte, #metoomovement, #consent, #violenceagainstwomen, #weinsteinscandal, #domesticabuse, #victimblaming, #moiaussi, #sexualviolence, #enoughisenough, #endrapeculture, #streetharassment, #abusefreeindia, #csa, #endvaw, #youtoo, #ustoo, #sexualharrassment, #wetoo, #metoomarch, #vawg, #sexualabise, #childabuse, #domesticviolenceawareness, #lockerroomtalk, #freethenipple, #stopvictimblaming, #gbv, #outrage, #dvam, #stopabuse', #sexist, #raped, #metoocampaign.

We extract all tweets (including retweets, quotes and replies) for the period 2014-2016, resulting in a total of 11,335,429 tweets measuring the conversation around GBV on Twitter during that year.

B.2 Assigning Location Information to Twitter Data

To take advantage of regional differences in social movements on Twitter as well as crime reports, we make use of the fact that 76.5 percent of tweets have a Twitter user location. Only 1.5 percent of tweets have a tweet location. We, therefore, decide to rely on the user location and not tweet location to extract geographic information. The Twitter user location is not available in a preprocessed format. This means that some users mention their country of residence, while others indicate their federal state, county, city, or even zip code. In order to match the Twitter data to crime data at the federal state level, we need to unify the data. We do so through using information on all census-recognized cities/towns provided by SimpleMaps (2012). This dataset contains information of the federal state, the federal state ID, the county and county ID, the city ID and zip codes. We then proceed as follows.

We first split the Twitter user location into different columns, based on commas. We then merge the city data to the Twitter data based on the first two columns identifying the user location and the city name and state ID from the city dataset. We next merge both datasets on the city name and state name. Next, we use the county name and state name. We then use the county zip code and state name. As a next step, we merge both datasets on the county name and state ID, and later on the state ID only (using first the first column of the user location and then the second column of the user's location). Similarly, we use the state name only (using first the first column of the user location and then the second column of the user's location only), and then the city name only

(using first the first column of the user location and then the second column of the user location only). We repeat this for the city name, county name, city zip and county zip respectively.

Through this procedure, we can assign 74.8 percent (8,566,786 out of 11,449,223) tweets a location (4,244,582 out of 6,236,539 tweets in 2014, 2,112,004 out of 2,685,019 tweets in 2015, 2,210,200 out of 2,527,665 tweets in 2016). In a second step, we address duplicated values. Duplicated values occur, as many cities in the US have the same name. There are 2,744,093 duplicated values in 2014, 1,284,143 in 2015, and 1,394,794 in 2016. If a city is duplicated, we keep the value with the largest population. This leaves us with 1,500,489 tweets in 2014, 827,861 in 2015, and 815,406 in 2016. Consequently, we are able to associate 24.1 percent of tweets in 2014 with a federal state in the US, 30.8 percent of tweets in 2015, and 32.3 percent of tweets in 2016. Through this procedure we can assign 27.5 percent of tweets (3,143,756 out of 11,449,223) a federal state in the US. We believe that this captures a large enough share of all Twitter users in 2021, as 37.7 percent were from the United States (77.75 out of 206 million users worldwide) (Statista, 2022).

B.3 The VADER Sentiment Analysis - Methodological Background

To analyze the content of what is written on Twitter within the conversation on GBV, we employ the VADER Sentiment Analysis tool (Hutto and Gilbert, 2014). The VADER Sentiment Analysis tool is a lexicon and rule-based sentiment analysis tool, which was trained on social media data. The lexicon has been validated by 10 independent human raters. It builds upon pre-existing, well-established sentiment word-banks (LIWC, ANEW, and GI) and adds common lexical features used on social media to these word-banks. Examples are emoticons (such as ":-)"), acronyms (such as "LOL"), and slang (such as "nah" or "giggly").

The VADER Sentiment Analysis tool deduces both the intensity and polarity of sentiments. The polarity refers to a binary classification into positive, neutral, or negative text. The tool reports the fraction of text, which is positive, neutral, and negative. Adding all three columns results in a value of 1. Importantly, the three columns do not account for contextual interplays of words. The contextual interplay is reflected in the compound score. The compound score is a single uni-dimensional measure of a text's sentiment. It accounts for the contextual connection of independent words through a variety of different methodologies, such as taking into account word-order sensitive relationships, or degree modifiers. The score ranges from -1 to 1. -1 is the most negative and 1 the most positive

classification possible.[27]

We showcase the resulting compound score by giving some artificial examples. The term "#metoo is great :-)" has a compound score of "0.7506" and is therefore overall positive. The sentence "#metoo is great." has a compound score of "0.6249". It is less positive as the previous example, as it lacks the smiley. In a similar fashion, "Gender-Based Violence is horrible." has a compound score of -0.8225, "Gender-Based Violence is HORRIBLE." a compound score of -0.8531, and "Gender-Based Violence is really horrible." a compound score of -0.8357.

B.4 Retrieving Socioeconomic Characteristics from Twitter Data

We apply two different tools to retrieve socioeconomic information from Twitter. Firstly, we make use of the *DeepFace* framework developed by Serengil and Ozpinar (2020). This framework is a lightweight face recognition and facial attribute analysis package in Python[28]. It allows to retrieve users' age, gender, emotion, and race from profile pictures. We make use of the default model, which is the VGG-Face model. The VGG-Face model was developed by Parkhi et al. (2015) and is a convolutional neural network (CNN) model. This model was trained using photos of two million faces and a "very deep" network.

To shed light on the gender of those tweeting within our dataset, we employ the *GenderGuesser*.[29] This package allows for the detection of authors' gender based on their first names. The resulting sexes are male, female, mostly male, mostly female, andy (having an equal probability to be male and female) as well as unknown.

Appendix C Additional Results

C.1 Summary Statistics

Table E14 shows summary statistics at the weekly level in the period 2014-2016, resulting in 156 observations. The number of Twitter tweets per week varies from 7,533 to 1.1 million tweets per week, while the number of crime reports ranges from 5,292 to 13,836 reports per week. The significant variation in the number of weekly tweets is in line with

[27]For examples on the Compound score see the VADER Github Repository. Link: https://github.com/cjhutto/vaderSentiment

[28]Its accuracy is above 97.53 percent (Serengil and Ozpinar, 2020).

[29]For the details and license information on the *GenderGuesser* package see `https://pypi.org/project/gender-guesser/`.

observations in Figure 1, which clearly shows that, while there are very few tweets in many weeks, the social movements are large and sudden.

Table E14: Summary statistics at the weekly level (2014-2016)

VARIABLES	Mean	Std. Dev.	Min	Max	p25	p75
GBV	11,333.03	1,294.70	5,292	13,836	10,602	12,259
Sexual violence	1,458.03	238.97	583	2,168	1,319	1,567
Physical violence	7,602.82	884.49	3,017	8,987	7,145	8,233
Emotional violence	2,272.18	294.06	751	2,681	2,156	2,455
Non-GBV crime	84,261.34	10,316.90	28,374	97,080	81,194	90,438
No. of tweets	71,000.92	127,427.41	7,533	1,132,676	37,546	58,858

Notes: The table shows the summary statistics of Twitter tweets and crime reports at the weekly level. For each crime type, the variable measuring crime is the number of crime reports in the United States at the weekly level. GBV refers to all crimes related to Gender-Based Violence (sexual, physical, and emotional crime). Sexual violence is defined as rape, sodomy, sexual assault with an object, fondling, and statutory rape. Physical violence includes murder/intentional manslaughter, aggravated assault, simple assault, kidnapping/abduction. Emotional violence is defined as intimidation. In the case of physical violence, we use information provided on the circumstances of the crime and restrict the cases to those related to an argument or lovers quarrel. Additionally, we restrict physical and emotional violence to cases, in which victim and offender are of opposite sexes, as we are only interested in GBV. The time period considered is 2014 to 2016. Source: NIBRS and Twitter data (2014-2016).

C.2 Robustness to Fixed Effects and Clustering

Table E15 reveals another interesting fact about the relationship between social movements on Twitter and GBV-related crime rates. The table presents the contemporaneous point coefficients of our regressions when including a number of different fixed effects. Column 1 abstracts from fixed effects, while Column 2 considers time fixed effects. Column 3 includes state fixed effects. Column 4 is our main regression specification, which considers month and state fixed effects. The table clearly shows that seasonality plays a crucial role in the relationship of social movements on Twitter and GBV-related crime rates. This is in line with the seasonality observed in Figure F2.

Table E15 also illustrates the impact of aggregating standard errors at different levels. Up to Column 5 we cluster standard errors at the month-state level. We then analyze the effect of varying the level of clustering. Column 4 to 6 demonstrate that standard errors increase with their level of aggregation. While coefficients are significant at the 10 percent significance level when clustering at the month-state level, they are significant at

the 5 percent significance level when clustering at the quarter level. Under a specification that clusters at the quarter-state level, our coefficient is insignificant. Consequently, estimates are sensitive to the level of clustering. We choose the second most conservative specification as our baseline model.

Table E15: The effect of social movements on GBV on crime reporting rates per 100,000 inhabitants (GBV)

	(1) GBV	(2) GBV	(3) GBV	(4) GBV	(5) GBV	(6) GBV
Twitter tweets	0.532	0.149**	0.390	-0.0610*	-0.0610**	-0.0610
	(0.501)	(0.0583)	(0.429)	(0.0367)	(0.0145)	(0.0399)
Constant	5.883***	5.939***	5.904***	5.970***	5.970***	5.970***
	(0.363)	(0.0322)	(0.363)	(0.0266)	(0.00214)	(0.0404)
Mean (Dep. Var)	5.961	5.961	5.961	5.961	5.961	5.961
St. Dv. (Dep. Var.)	5.508	5.508	5.508	5.508	5.508	5.508
State fixed effects	No	Yes	No	Yes	Yes	Yes
Month fixed effects	No	No	Yes	Yes	Yes	Yes
Clustered standard errors	Month-State	Month-State	Month-State	Month-State	Quarter	Quarter-State
N	5751	5751	5751	5751	5751	5751

Notes: The table shows the results from a linear regression of the number of Twitter tweets on crime rates under different empirical specifications. The outcome variable is the crime rate per 100,000 inhabitants in a respective week and federal state, considering all GBV-related crimes. We define GBV-related crimes as physical, sexual, and emotional crimes, in which the perpetrator and victim are of different gender. The explanatory variable is the number of GBV-related tweets in each week and state, divided by 100 cell-phone internet plan subscriptions in the federal state in that year. The unit of analysis is the week by state. Column 1 abstracts from fixed effects. Column 2 includes state fixed effects and Column 3 month fixed effects. Column 4 controls for both state and month fixed effects. Clustered standard errors are at the month-state level in Column 1 to 4, at the quarter level in Column 5 and at the quarter-state level in Column 6. Clustered standard errors are reported in parentheses. Source: NIBRS, Twitter and ACS. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

C.3 Robustness to *#yesallwomen*

To validate results from our Event Study design, we restrict our dataset to the movement *#yesallwomen* and employ our main regression specification to this restricted version of the dataset. The results on crime rates presented in Table E16 are significant at the 5 percent significance level. Compared to the specification which considers several Twitter movements the estimates are larger. Importantly, the lagged coefficients are negative, which is in line with our main results.

Table E16: The effect of social movements on GBV on crime rates per 100,000 inhabitants (GBV)

	(1) GBV	(2) GBV	(3) GBV
Twitter tweets	-0.716	2.560*	-0.835
	(0.462)	(1.505)	(3.107)
L.Twitter tweets		-1.826**	1.062
		(0.766)	(1.669)
L2.Twitter tweets			-2.266**
			(0.897)
Constant	5.528***	5.180***	4.912***
	(0.0281)	(0.0290)	(0.0289)
Mean (Dep. Var)	5.525	5.179	4.901
St. Dv. (Dep. Var.)	5.421	5.296	5.182
State-Month fixed-effects	Yes	Yes	Yes
N	2471	1707	1340

Notes: The table shows the results from a linear regression of the number of Twitter tweets using *#yesallwomen* on crime rates. The outcome variable is the crime rate per 100,000 inhabitants in a respective week and federal state, considering all GBV-related crimes. We define GBV-related crimes as physical, sexual, and emotional crimes, in which the perpetrator and victim are of opposite gender. The explanatory variable is the number of GBV-related tweets in the federal state during the week, divided by 100 cellphone internet plan subscriptions in the federal state in that year. The unit of analysis is the week by state. The first column only considers the impact of Twitter tweets on the contemporaneous crime rate. Column 2 adds Twitter tweets in the previous week, while Column 3 also considers Twitter tweets two weeks previously. We weight each cell by the population size of each federal state in the respective year. We control for month of the year by state fixed effects. Month by state level clustered standard errors are reported in parenthesis. Source: NIBRS, Twitter and ACS. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

C.4 Arrest per Crime Rates by Type of Crime

Table E17: The effect of social movements on GBV on arrests per crime (Physical violence)

	(1) Arrests	(2) Arrests	(3) Arrests
Twitter tweets	0.000269	-0.00526	-0.00832
	(0.00644)	(0.00668)	(0.00807)
L.Twitter tweets		0.00998*	0.00520
		(0.00562)	(0.00346)
L2.Twitter tweets			0.0118
			(0.0115)
Constant	0.473***	0.472***	0.471***
	(0.00224)	(0.00227)	(0.00238)
Mean (Dep. Var)	0.473	0.473	0.473
St. Dv. (Dep. Var.)	0.161	0.160	0.160
State-Month fixed-effects	Yes	Yes	Yes
N	5732	5692	5652

Notes: The table shows the results from a linear regression of the number of Twitter tweets on the arrest per crime rate. The outcome variable is the respective arrest per crime rate, considering all crimes related to physical violence. We define physical violence as crimes related to murder/intentional manslaughter, aggravated assault, simple assault, kidnapping/abduction, in which the perpetrator and victim are of opposite gender. The explanatory variable is the number of GBV-related tweets in the federal state during the week, divided by 100 cellphone internet plan subscriptions in the federal state in that year. The unit of analysis is the week by federal state. The first column only considers the impact of Twitter tweets on the contemporaneous arrest per crime rate. Column 2 adds Twitter tweets in the previous week, while Column 3 also considers Twitter tweets two weeks previously. We weight each cell by the population size of each federal state in the respective year. We control for month of the year and state fixed effects. Month by state level clustered standard errors are reported in parenthesis. Source: NIBRS, Twitter and ACS. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table E18: The effect of social movements on GBV on arrests per crime (Emotional violence)

	(1) Arrests	(2) Arrests	(3) Arrests
Twitter tweets	-0.00463	-0.000605	-0.00219
	(0.00767)	(0.00641)	(0.00680)
L.Twitter tweets		-0.00830	-0.0105*
		(0.00622)	(0.00635)
L2.Twitter tweets			0.00557
			(0.00848)
Constant	0.215***	0.216***	0.216***
	(0.00328)	(0.00346)	(0.00355)
Mean (Dep. Var)	0.214	0.215	0.215
St. Dv. (Dep. Var.)	0.204	0.204	0.205
State-Month fixed-effects	Yes	Yes	Yes
N	5480	5445	5408

Notes: The table shows the results from a linear regression of the number of Twitter tweets on the arrest per crime rate. The outcome variable is the respective arrest per crime rate per week and federal state, considering all crimes related to emotional violence. We define emotional violence as intimidation, in which the perpetrator and victim are of opposite gender. The explanatory variable is the number of GBV-related tweets in the federal state during the week, divided by 100 cellphone internet plan subscriptions in the federal state in that year. The unit of analysis is the week by federal state. The first column only considers the impact of Twitter tweets on the contemporaneous arrest per crime rate. Column 2 adds Twitter tweets in the previous week, while Column 3 also considers Twitter tweets two weeks previously. We weight each cell by the population size of each federal state in the respective year. We control for month of the year and state fixed effects. Month by state level clustered standard errors are reported in parenthesis. Source: NIBRS, Twitter and ACS. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

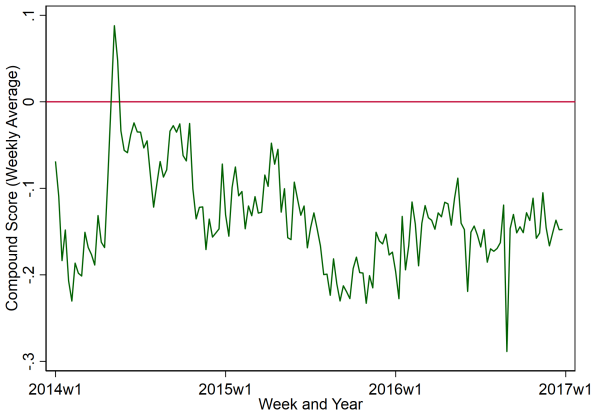
Table E19: The effect of social movements on GBV on arrests per crime (Sexual violence)

	(1) Arrests	(2) Arrests	(3) Arrests
Twitter tweets	0.00899	0.0128	0.00788
	(0.00608)	(0.00837)	(0.00765)
L.Twitter tweets		-0.00814	-0.0156*
		(0.00717)	(0.00868)
L2.Twitter tweets			0.0188**
			(0.00751)
Constant	0.187***	0.188***	0.187***
	(0.00309)	(0.00302)	(0.00306)
Mean (Dep. Var)	0.188	0.189	0.189
St. Dv. (Dep. Var.)	0.183	0.183	0.183
State-Month fixed-effects	Yes	Yes	Yes
N	5431	5394	5360

Notes: The table shows the results from a linear regression of the number of Twitter tweets on the arrest per crime rate. The outcome variable is the respective arrest per crime rate per 100,000 inhabitants per week and federal state, considering all crimes related to sexual violence. We define crimes related to sexual violence as rape, sodomy, sexual assault with an object, fondling, statutory rape, in which the perpetrator and victim are of opposite gender. The explanatory variable is the number of GBV-related tweets in the federal state during the week, divided by 100 cellphone internet plan subscriptions in the federal state in that year. The unit of analysis is the week by federal state. The first column only considers the impact of Twitter tweets on the contemporaneous arrest per crime rate. Column 2 adds Twitter tweets in the previous week, while Column 3 also considers Twitter tweets two weeks previously. We weight each cell by the population size of each federal state in the respective year. We control for month of the year and state fixed effects. Month by state level clustered standard errors are reported in parenthesis. Source: NIBRS, Twitter and ACS. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix D Additional Figures

Figure F1: Weekly Sentiment Scores (2014-2016)



Notes: The figure shows weekly average sentiment scores for all tweets in our dataset. We employ the VADER Sentiment Analysis tool to identify the sentiments in text of tweets and consider data from the period 2014-2016. The x-axis shows the respective week and the y-axis shows the compound score. The red line refers to a compound score of zero, which reflects neutrality. For the details behind the compound score see Appendix B.3. Source: Twitter (2014-2016).

Figure F2: Weekly development of GBV-related crime reports (2014-2016)

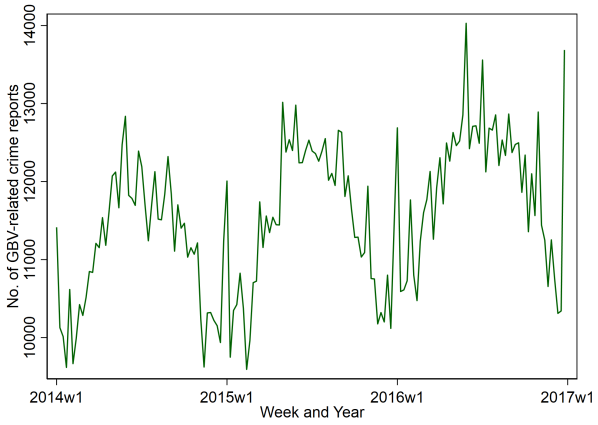
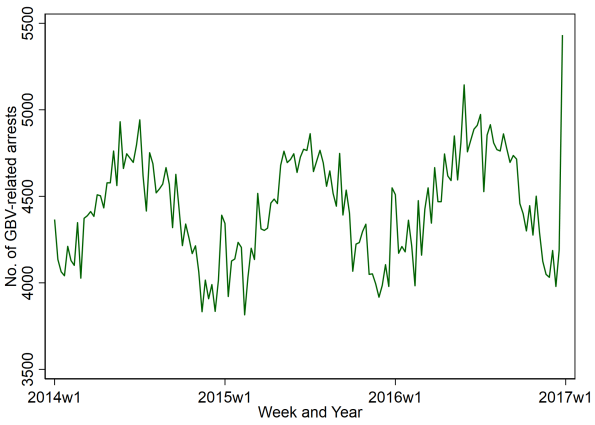


Figure F3: Weekly development of GBV-related crime arrest (2014-2016)



Notes: The left panel plots the number of weekly crime reports on GBV-related crimes in the United States reported in the NIBRS for the period 2014-2016. The right panel plots the weekly number of arrests of GBV-related crimes. Source: NIBRS (2014-2016).