# The Critical Role of Social Leaders in the Spread of Social Movements against Gender-Based Violence [a]

Britta Rude[b]

November 8, 2022

**Abstract.** This paper asks how social movements against gender-based violence (GBV) spread on social media. To this end, I construct a novel data set measuring 10 large social movements against GBV on Twitter. I show that these movements start suddenly and fade out quickly and that there is considerable variation at the sub-national level in the US. Twitter users are more likely to share content created by other users instead of creating original content. Polarization is low and most users express fear and sadness. Neither polarized nor emotional content generates more traction in form of likes, retweets, replies or quotes. I develop a novel instrumental variable strategy and show that Twitter users with an established network play a major role in the spread of tweets. Social inclusion is low. Users are on average female, young, and White. Tweets posted by non-white users generate less traction. Moreover, women are more prone to reference content by women, while the reverse applies to men.

**Keywords:** Economics of Gender; Non-labor Discrimination; Demographic Economics: Public Policy; Social Choice; Clubs; Committees; Associations; Economic Sociology
**JEL Codes:** J16, J18, D71, Z13

[a]I thank Michele Battisti, Ilpo Kauppinen, Panu Poutvaara, Helmut Rainer, Claudia Steinwender for helpful comments and remarks. I thank participants of the Cesifo Junior Workshop on Big Data 2022, the ifo Lunchtime Seminar 2022, the PhD Idea Seminar at the University of Munich, the Berlin workshop on Gender Economics 2022 and the Webinar in Gender and Family Economics 2022 for helpful comments and suggestions.

[b]Contact: rude@ifo.de, ifo Institute - Leibniz Institute for Economic Research at the University of Munich and Ludwig Maximilian University Munich.

# 1 Introduction

History shows that social movements, such as the French Revolution, can result in radical political and social change. Yet, while some of these historical movements disrupted the way societies organize, others failed their purpose, or were nipped in the bud. Recent examples are the Arab Spring or the 2019 Venezuelan protests. Data scarcity hampered research studying social movements and, to date, it is not clear what drives success and failure of these movements.

Recent advances in technologies, such as social media platforms, allow for large-scale and publicly accessible conversations online and thus ease network building and the formation of social movements. One example is the famous *#metoo* movement[1]. The formation and spread of these movements in the online space generate new opportunities to study social movements and their drivers more systematically.

In this paper, I generate a new data set which mirrors 10 social movements against gender-based violence (GBV) on Twitter. The focus on GBV is interesting, as it is surrounded by stigmatization, tabooing and silencing, which leads to many victims to remain silent (Overstreet and Quinn, 2013). Thus, historical levels of GBV have been especially immune against social change and the rate of GBV has barely decreased over time. GBV, which is violence grounded on gender, is still one of the most severe problems of contemporaneous societies. According to UN Women (2021), one out of three women experience GBV along their lifetime. Although there is stigmatization and tabooing around GBV, people have discussed it extensively on social media, which is surprising and raises the case for studying these conversations in more detail.

I use data on Twitter tweets that mention hashtags which clearly relate to 10 social movements against GBV and study how they spread on Twitter.[2] I investigate the influence of emotions and polarized content by analyzing the text of tweets. Moreover, I study how social leaders influence the visibility of these movements on Twitter. To this end, I develop a novel instrumental variable strategy by making use of the fact that certain Twitter accounts are officially verified by Twitter.

I show that social movements against GBV on Twitter start suddenly and fade out quickly. Besides, there is considerable variation of Twitter activity at the federal state

[1] *#metoo* is a movement which widely spread on social media in October 2017 as a response to the *Weinstein scandal*. This scandal consisted of a number of sexual abuse allegations against *Harvey Weinstein* in October 2017. The expression *Me Too* was first used by activist Tarana Burke back in 2006 with the goal to empower victims of sexual violence and rape culture.

[2] I rely on the following 10 hashtags: *#bringbackourgirls*, *#yesallwomen*, *#rapecultureiswhen*, *#WhyIStayed*, *#everydaysexism*, *#NoWomanEver*, *#notokay*, *#metoo*, *#MeAt14*, *#whyididntreport*.

level. I demonstrate that Twitter users retweet more than that they create original content. Moreover, most tweets do not generate any considerable traction in form of likes, retweets, replies or quotes. My sentiment analysis reveals that tweets are mainly neutral and do not reflect extreme or polar content. Greater polarity of written text does not generate more traction in form of likes, replies, retweets, or quotes. My emotion detection analysis demonstrates that Twitter users mainly express fear and sadness. I do not find evidence in favor of emotions driving the spread of Twitter tweets. Both results from ordinary least square and instrumental variable regressions show that Twitter users with an established network play an important role in the dissemination of social movements against GBV on Twitter. Tweets with media attachments also generate more traction.

My findings are in line with previous research showing that leadership is a crucial mechanism in addressing GBV (Iyer et al. (2012); Wen (2021); Delaporte and Pino (2022)) as well as stereotypes (Besley and Ghatak, 2018). In addition, it confirms that social networks are important to advance gender equality (Agarwal et al., 2016), as are group dynamics (Gagliarducci and Paserman, 2012) and gender norms (González and Rodríguez-Planas, 2020). A related body of literature shows that it is often female leaders who drive change in the interest of women (Bhalotra and Clots-Figueras (2014); Brollo and Troiano (2016); Besley and Ghatak (2018) Bhalotra et al. (2018); Flabbi et al. (2019); Bertrand et al. (2019)). Based on these findings, I investigate if Twitter tweets by female social leaders generate more traction than Twitter tweets by male social leaders. My evidence does not support this thesis. So, an established network is more important than the gender of social leaders on Twitter.

The fact that media attachments in tweets generate higher visibility of tweets is in line with research by Cooper et al. (2020). Their paper demonstrates that exposure to video dramatizing of violence against women and girls increases the likelihood of reporting such crimes. Similarly, Haraldsson and Wängnerud (2019) find a significant relationship between media sexism and women's political ambition.

The second part of the analysis in this paper focuses on the inclusiveness of the movements studied. To this end, I generate socio-demographic information on Twitter users by applying a Face Recognition Tool, namely the *DeepFace* framework developed by Serengil and Ozpinar (2020). Twitter users who contribute to the social movements studied are mainly female but there are also many male users engaging. In addition, they are mainly White and on average 28.8 years old. Ethnic minorities are underrepresented among users, raising doubt on the inclusiveness of these movements. Tweets posted by non-white users as part of the social movements spread to a lower extent than those posted by white users. I test the accuracy of the face recognition by a validation exercise, relying on 10,000 labeled photos provided by Zhang and Qi (2017).

Next, I compare the gender of tweets' authors to the gender of referenced tweets' authors. I find that there are clear gender patterns in the spread of information across Twitter. Women are more likely to reference content by women, while men are more likely to reference content by men. These dynamics are in line with research by Roden et al. (2021) which shows that gender dynamics play an important role on social media. Men's support for content on gender equality increases with the extent to which they perceive endorsers to be male.

This paper contributes to the literature studying GBV on social media.[3] My paper also talks to the literature developing techniques to detect misogyny and sexism online (Pamungkas et al. (2020); Chowdhury et al. (2019); Rodríguez-Sánchez et al. (2020); Fersini et al. (2019); Pandey et al. (2018)). My work builds on a body of literature that analyzes how people behave on social media.[4] This paper also relates to the literature studying the interaction between how humans behave online and offline.[5]

My findings are robust to using alternative measures for a tweet's visibility, such as the number of retweets, quotes and replies. They hold under different model specifications and estimation strategies, such as a poisson regression, or the inclusion of time fixed effects. They also persist when restricting the sample to original tweets and abstracting from retweets, quotes and replies.

[3]Most closely, ElSherief et al. (2017) analyze the *#notokay* movement on Twitter and describe it descriptively. My paper is also closely related to work by Khatua et al. (2018) who extracted 0.7 million tweets during the *#metoo* social movement and employ deep learning techniques to extract information on types of GBV-related crimes committed, perpetrators and places of assault. Similarly, Pamungkas et al. (2020) explore the *#metoo* movement to study expression of GBV-related abuse. Related work by Garrett and Hassan (2019) study reasons for the silencing around GBV by exploring the *#whyididntreport* movement.

[4]Bakshy et al. (2015), for instance, study how news spread on Facebook. Several papers discuss the proliferation of fake news (Grinberg et al. (2019); Guess et al. (2019); Bovet and Makse (2019)) and related work by Gentzkow and Shapiro (2011) analyzes the effect of the internet on ideological segregation. The authors find no evidence in favor of an increased segregation through the internet.

[5]Korda and Itani (2013), for example, study the role of social media for health promotion and education. They conclude that social media has the potential to affect these outcomes, but that there is a need for careful applications and evaluations. Similarly, Zobeidi et al. (2022) investigate the impact of social media on renewable energy usage and find significant results. Patroni et al. (2020) demonstrate a significant link between social media conversations and organizational innovations. A related study by Allcott et al. (2020) demonstrates that deactivating Facebook for a period of four weeks increases well-being. Similarly, Braghieri et al. (2022) find a negative relationship between social media usage and mental health. Recent work by Chetty et al. (2022a) relies on Facebook data to study the interaction between social capital and economic mobility as well as the determinants of cross-class interactions (Chetty et al., 2022b). Battisti et al. (2022) show that social movements against GBV on Twitter decrease reported GBV-related crime rates in the United States.

The evidence generated in this paper points to the importance of social leaders in driving social change. In addition, the lack of inclusiveness, both in gender and ethnicity, is of concern. Given the importance of social leaders, the inclusion of male leaders into the agenda to address GBV is crucial. Similar recommendations apply to the media. Involving the media actively in the case against GBV is recommendable.

# 2 Conceptual Background on Social Networks and Social Leaders

This paper asks how information spreads across social networks. Although I investigate this question in the context of social media and online platforms, I can draw from common concepts on social networks developed more broadly. Formally, a social network consists of a set of actors and a set of relations between them (Knoke and Yang, 2019). Actors can be individuals, but also entities (such as organizations, or a team) and relationships can be linear or non-linear. Social networks as a whole can be causes and consequences of human actions and perception. Knoke and Yang (2019) mention that they generalize expectations, rules, and norms among their members, leading to a system of trust and sanctions.

Social leaders play an important role in social networks. Following the literature on leadership, I define leadership as a social influence process (Anca et al., 2014). Leadership can influence the motivation, goals, thinking, attitudes, culture and behavior of a social network, in this case a group of people who are followers of a given leader.

The concept of social networks has been studied empirically and theoretically in a broad range of disciplines. As Knoke and Yang (2019) point out correctly, social network analysis is an interdisciplinary field, and the availability of new data sources, such as social media data, create new avenues to study how social networks evolve and develop. The fact that social networks play a crucial role in building social norms and behavior makes it especially interesting to apply the question to the topic of GBV, given the little progress made in this area. Economists have studied the dynamic growth strategy of social networks (Veiga et al., 2017), word-of-mouth communication (Campbell, 2013), the importance of social networks for electoral outcomes (Cruz et al., 2017) and job referrals (Beaman and Magruder, 2012) and the formation of social networks (Jackson and Rogers, 2007).

In the following I shed light on how tweets that contribute to movements against GBV spread on Twitter. I first describe these movements descriptively and then investigate potential drivers behind the visibility of individual tweets. I proxy the visibility of tweets by the number of likes a given tweet receives, but also analyze alternative measures, such as the number of retweets, quotes, and replies. I pay special attention to the role social leaders play in the spread of social movements against GBV on Twitter.

Given that the similarity and homogeneity of its actors influence group dynamics and consequently social networks (Öberg et al., 2011), I also investigate the inclusiveness of the social movements studied in this paper. For this purpose, I first analyze the share of male and female users who contribute to these movements. I then also investigate users by their race. Lastly, I ask if the role which social leaders play in the spread of these movements differs by gender and race.

# 3 Twitter Data Set Creation

## 3.1 Data Extraction

To study the spread of social movements against GBV on Twitter, I create a novel data set of Twitter tweets. To this end, I filter for a subset of the full Twitter database via the Twitter Full Archive API.[6] This means that I define a customized search query. I do so by restricting the period investigated to the years 2014 to 2017 and by only considering tweets posted in English and mentioning certain keywords.

These keywords consist of a list of hashtags which clearly signal a contribution to 10 of the most widespread social movements against GBV on Twitter. I focus on the following 10 hashtags: *#bringbackourgirls*, *#yesallwomen*, *#rapecultureiswhen*, *#WhyIStayed*, *#everydaysexism*, *#NoWomanEver*, *#notokay*, *#metoo*, *#MeAt14*, and *#whyididntreport*. All of the social movements related to these hashtags emerged as a response to acts of violence against women and girls, or in order to raise awareness about GBV and related harmful gender norms. The *#bringbackourgirls* movement emerged as a reaction to the abduction of 276 female students by the terrorist group Boko Haram in Nigeria. *#yesallwomen* is a hashtag which was frequently used after the *Isla Vista Killings*, a series of misogynistic killings in California. Similarly, *#rapecultureiswhen* was a hashtag that started to trend after the suicide of Rehtaeh Parsons who had been raped and bullied afterwards, and *#WhyIStayed* was trending after the intimate partner violence case

[6] I retrieve data from the Twitter database by creating an Academic Developer Account, which gives me access to the Twitter Full Archive Search API V2. This API allows me to access the full universe of Twitter tweets posted since the creation of Twitter. The API has a rate limit of 10 million tweets per month and 150,000 tweets per 15 minutes. I take advantage of the Twarc2 command line tool and Python library.

of football player Ray Rice and his then-fiancee gained public attention. *#whyididntreport* was a hashtag used in support of Christine Blasey Ford and her allegations against Supreme Court nominee Brett Kavanaugh. The social movement related to *#notokay* became viral as a response to misogynistic talk by Donald Trump, and *#metoo* as a reaction to allegations against Harvey Weinstein. Both the movements using the hashtag *#everydaysexism* and *#NoWomanEver* aimed to raise awareness about harmful believes and gender norms driving GBV.

Through the usage of these hashtags, victims of GBV openly shared their personal experiences with this form of violence on Twitter. In addition, Twitter users contributing to these movements showed their support of and solidarity with victims of GBV cases which had gained wide public attention. Given that these movements broke with the silencing, tabooing and stigmatization which traditionally surrounds the topic of GBV (Overstreet and Quinn, 2013), it is of interest to study them in more detail.

I filter for all tweets using at least one of these hashtags as part of the tweets' written text. The generated final data set consists of 1.1 million tweets. The resulting data is organized in Json (JavaScript Object Notation) Objects, such as a *User* object or a *Tweet* object. There are four overall JSON Keys, which are the *Data*, *Includes*, *Error*, and *Meta* keys. Each key consists of several nested JSON Objects. Each object comes along with attributes describing the Json Object, such as information on the tweet's author, the actual text of the tweet, a unique tweet ID, a timestamp of when a tweet was posted, and sometimes geographic information about the location of the user or tweet itself. 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. I normalize the JSON data set in Python and transform it to a pandas data frame. The final data set is at the tweet level and consists of several variables which describe a given tweet and its author. Appendix A describes the process in more detail.

## 3.2 Analyzing the Text of Tweets

I apply text mining methods to extract information about the tweets' text. To this end, I first apply standard text analysis methods, such as the tokenization of the text of tweets, lemmatization and removing stop words. Appendix B details the procedure step by step.

To analyze sentiments behind what is written on Twitter as part of the social movements against GBV, I 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[7].

The VADER Sentiment Analysis tool deduces both the intensity and polarity of sentiments. 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.[8]

I also investigate the influence of emotions play in the spread of Twitter tweets. For this purpose I employ the emotion recognition tool developed by Colnerič and Demšar (2018). This tool was trained explicitly on Twitter data and detects six different emotions (Anger, Disgust, Fear, Joy, Sadness, Surprise).

### 3.3 Final Data Set

Table 1 details the summary statistic of the variables used in this paper. The average Twitter user mentioning one of the 10 hashtags in an original tweet, retweet, quote, or reply during the period 2014-2017 has 9 followers, follows 2 Twitter accounts, and posted 51,180 different tweets. The average tweet was retweeted 774 times, but has less than one reply or quote and only two likes. The percentile values presented in Row 5 and 6 show that most tweets do not generate any traction, while a small number of tweets drives the visibility of the social movements investigated in this paper on Twitter.

In general, there are more retweets than original tweets. Only 42 percent of the sample are original tweets. Consequently, Twitter users signaling a contribution to the

[7]Examples are emoticons (such as ":-)"), acronyms (such as "LOL"), and slang (such as "nah" or "giggly").

[8]For examples on the Compound score see the VADER Github Repository. Link: https://github.com/cjhutto/vaderSentiment. I 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, "GBV is horrible." has a compound score of -0.8225, "GBV is HORRIBLE." a compound score of -0.8531, and "GBV is really horrible." a compound score of -0.8357.

social movements investigated in this paper mainly share information by others instead of creating content by themselves. This finding could mean that certain individuals play a more significant role in the spread of information and ideas within social movements against GBV on Twitter.

The emotion detection analysis reveals that one fourth of the written text is marked by fearful emotions, while more than 15 percent can be characterized as sad as well as surprised. Only around one tenth of the written text is happy and approximately 5 percent angry (see Table 1). The sentiment analysis demonstrates that more than 70 percent of all tweets are classified as neutral (see Figure 1 and Figure 2). Figure 1 illustrates that a slightly larger share of tweets is positive than negative. Related to this pattern of results, the average compound score for all tweets analyzed in this paper is 0.004 and close to zero. Consequently, the written content is on average of low polarity.

Figure 1: Histogram of sentiment scores

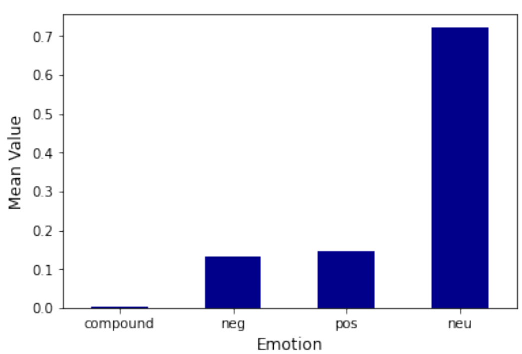
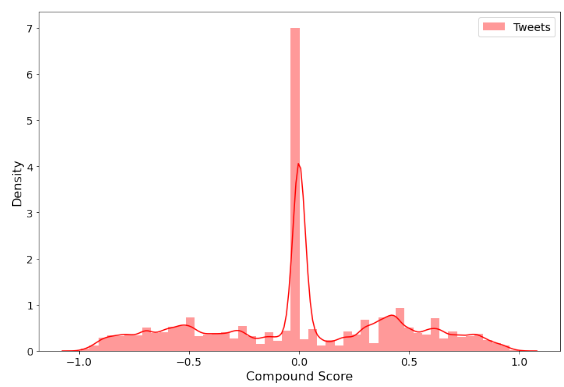



Figure 2: Distribution of compound score




Notes: The left figure depicts the average compound score as well as the share of tweets classified as negative, positive, or neutral. The right figure plots the distribution of the compound score. Source: Twitter (2014-2017).

Table 1: Summary statistics of Twitter tweets

| VARIABLES | (1) mean | (2) sd | (3) min | (4) max | (5) p25 | (6) p75 |
|---|---|---|---|---|---|---|
| Verified account | 0.03 | 0.17 | 0 | 1 | 0 | 0 |
| Has attachment | 0.09 | 0.29 | 0 | 1 | 0 | 0 |
| Follower (in 1,000s) | 8.82 | 248.90 | 0 | 60,544.17 | 0.18 | 1.82 |
| Following (in 1,000s) | 1.72 | 6.72 | 0 | 1,548.99 | 0.25 | 1.65 |
| No. of likes | 1.65 | 95.09 | 0 | 81,233 | 0 | 0 |
| No. of quotes | 0.04 | 3.64 | 0 | 2,583 | 0 | 0 |
| Has user location | 0.75 | 0.43 | 0 | 1 | 1 | 1 |
| Has geo location | 0.03 | 0.16 | 0 | 1 | 0 | 0 |
| original | 0.42 | 0.49 | 0 | 1 | 0 | 1 |
| No. of replies | 0.17 | 3.63 | 0 | 2,305 | 0 | 0 |
| No. of retweets | 773.56 | 2,908.62 | 0 | 17,160 | 0 | 61 |
| No. of tweets | 51,180.50 | 136,660.48 | 0 | 4,007,516 | 3,513 | 42,845.50 |
| Has mentions | 0.62 | 0.48 | 0 | 1 | 0 | 1 |
| Happy Score | 0.10 | 0.26 | 0 | 1 | 0 | 0 |
| Angry Score | 0.05 | 0.18 | 0 | 1 | 0 | 0 |
| Surprise Score | 0.16 | 0.30 | 0 | 1 | 0 | 0.33 |
| Sad Score | 0.18 | 0.32 | 0 | 1 | 0 | 0.33 |
| Fear Score | 0.26 | 0.37 | 0 | 1 | 0 | 0.50 |
| Negativity score | 0.13 | 0.17 | 0 | 0.95 | 0 | 0.25 |
| Neutrality score | 0.72 | 0.21 | 0.04 | 1 | 0.56 | 1 |
| Positivity score | 0.15 | 0.18 | 0 | 0.95 | 0 | 0.26 |
| Compound Sentiment | 0 | 0.45 | -0.99 | 1 | -0.32 | 0.36 |

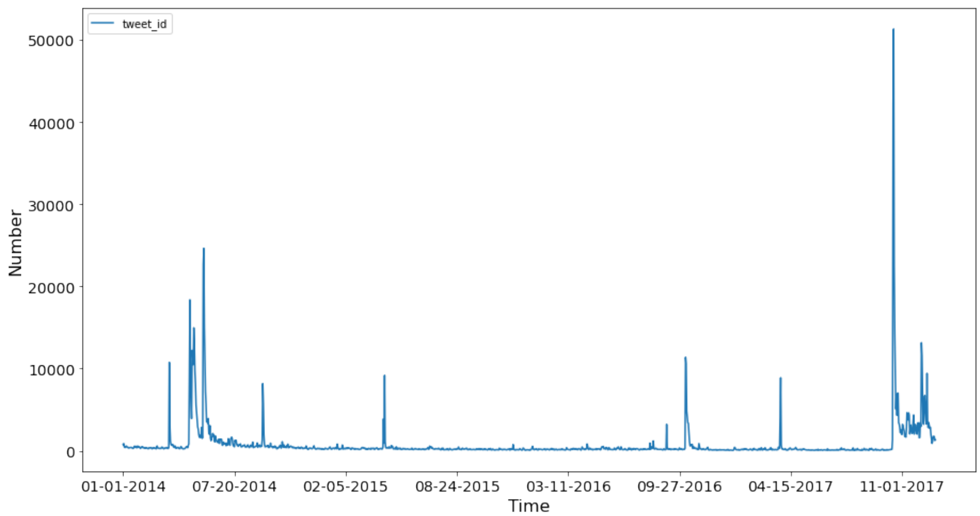
Notes: The table presents summary statistics of the data set generated via the Twitter API. I apply a hashtag-based approach, relying on 10 hashtags which identify GBV-related movements on Twitter, to filter on the full universe of Twitter tweets. I only consider the period 2014-2017. To extract the compound sentiment scores as well as the share of negative, neutral and positive tweets I employ the VADER Sentiment Analysis tool by Hutto and Gilbert (2014). To analyze emotions of written text I employ the emotion recognition tool developed by Colnerič and Demšar (2018). N=1,085,336. Source: Twitter (2014-2017).

# 4 The Spread of Twitter Tweets over Time and Space

In the following, I describe how the social movements against GBV spread on Twitter over time and space. I analyze their development over time, by type of tweets, as well as across federal states in the United States. I also investigate the content of what users write over time.

Figure 3 depicts the daily number of Twitter tweets generated via the hashtag-based approach developed in this paper. The figure reveals that social movements on Twitter emerge suddenly and fade out quickly. The first spike with more than 10,000 daily tweets refers to the *#bringbackourgirls* movement. The second spike with more than 25,000 tweets per day depicts the *#yesallwomen* movement and the last spike in 2017 represents the famous *#metoo* movement, which gained significant international visibility. Overall, on the majority of days, there are very little tweets.

Figure 3: Daily number of Twitter tweets (2014-2017)



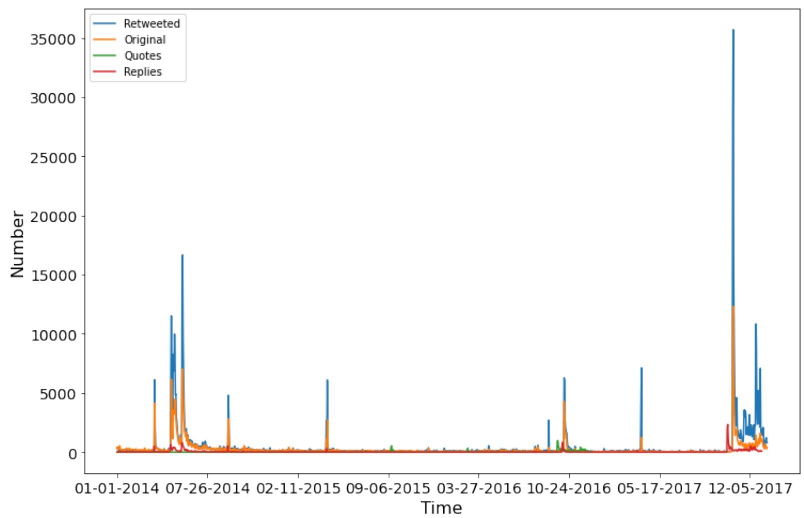

Notes: The graph plots the number of English tweets per day posted as part of the 10 social movements studied in this paper. Source: Twitter (2014-2017).

I next analyze how these movements spread across Twitter. For this purpose, I ask whether Twitter users mainly create original content or share content created by other authors. This question is of interest, given that it sheds light on the extent to which individual social leaders potentially drive these movements.

Figure 4 depicts the number of Twitter tweets by tweet type (original tweets, retweets, quotes, and replies). The graph reveals that tweets of different types move parallel to each other and that there are more retweets than original tweets on the majority of days. Consequently, the observation of users mostly sharing content instead of creating original content on their own persists over time. Likes are an important additional metric measuring a tweet's traction, as they increase a tweet's visibility on Twitter. Figure 5 reveals that the number of likes moves parallel to the number of tweets, similar to the observation made on retweets, quotes, and replies.

To shed light on the number of authors contributing to the social movements studied in

Figure 4: Number of daily tweets by tweet type (2014-2017)



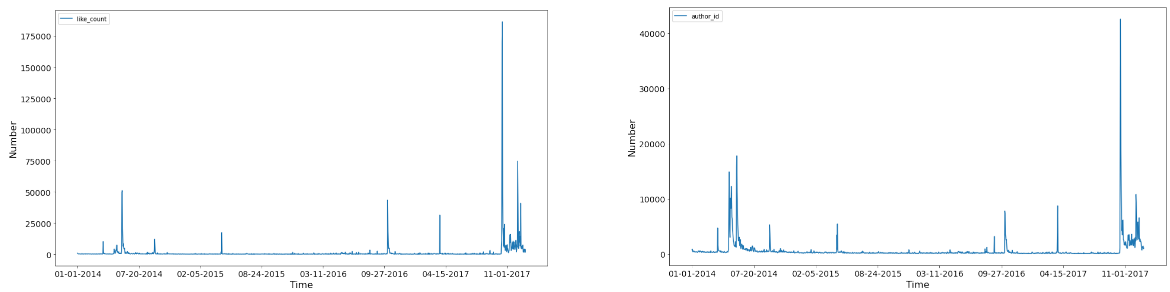

Notes: The graph plots the number of daily tweets by type of tweet (original tweet, retweet, quote, and replies). Source: Twitter (2014-2017).

this paper, I illustrate the daily number of authors (see Figure 5). The figure demonstrates that there is a parallel movement of the number of authors compared to the number of tweets. Consequently, social movements against GBV are not driven by individual authors suddenly engaging more actively by posting more content, but by a larger number of people contributing to these movements. To investigate if social movements are driven by a small number of authors tweeting many tweets I divide the number of tweets per day by the number of unique authors. The tweets per author rate fluctuates between one and two tweets for all days, which confirms that the movements are not driven by few authors posting the majority of tweets.

Figure 6 plots the average compound score per day as well as the daily share of sentiments of the underlying Twitter data set. The red line is the average compound score which ranges from -1 (very negative) to 1 (very positive). The green line represents the share of tweets classified as positive, the blue line the share classified as negative, and the yellow line the share classified as neutral. The graph reveals that more than half of the tweets are neutral on the majority of days.

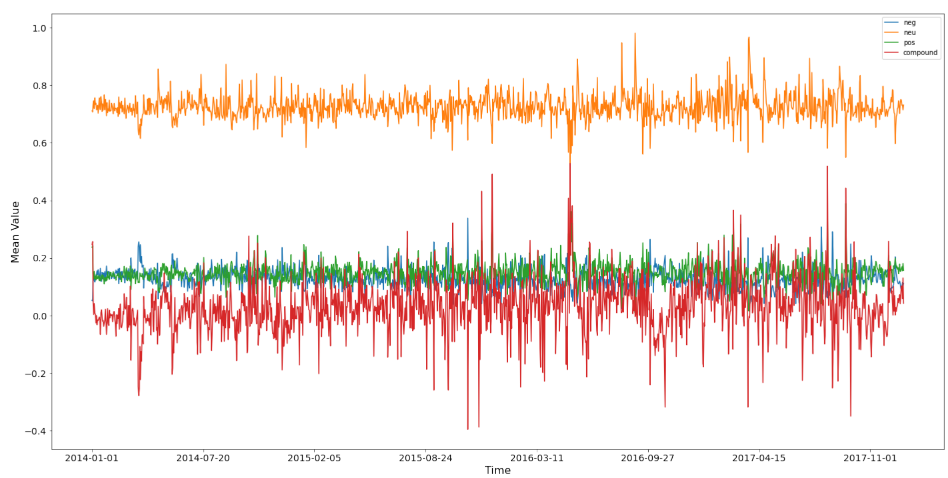
Figure 7 illustrates the daily share of emotions used within the 10 social movements. Based on this figure, emotions are mainly fearful, followed by sadness. There are also several days, during which people mainly express surprise. Anger and happiness seem to play a more negligible role.

Figure 5: Number of daily likes and unique authors posting tweets (2014-2017)




Notes: The left graph plots the number of daily likes. The right graph plots the number of unique authors posting tweets. Source: Twitter (2014-2017).

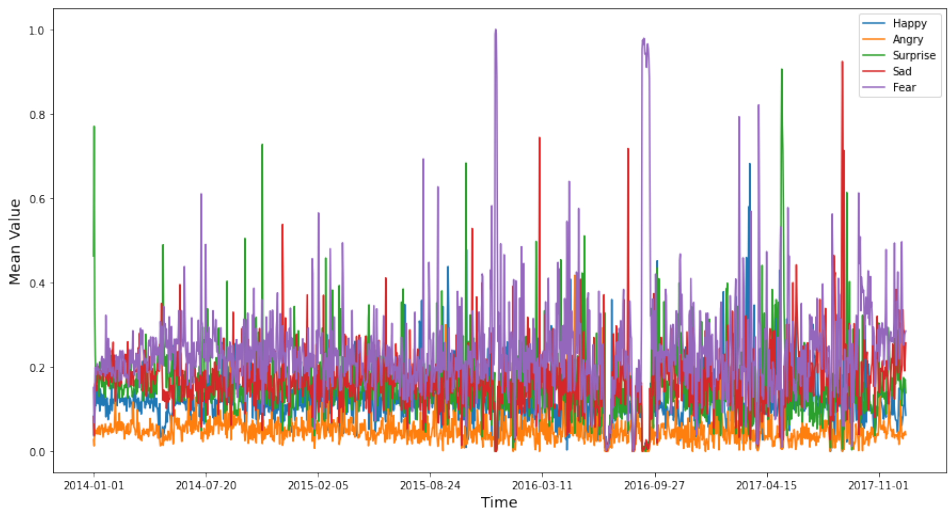
Figure 6: Daily compound score and sentiment shares of Twitter tweets (2014-2017)




Notes: The graph plots the average compound score and the share of sentiments per day for the period 2014-2017. Sentiments are retrieved via the VADER Sentiment Analysis Tool by Hutto and Gilbert (2014). The red line is the average compound score which ranges from -1 (very negative) to 1 (very positive). The green line represents the share of tweets classified as positive, the blue line the share classified as negative, and the yellow line the share classified as neutral. Source: Twitter (2014-2017).

To investigate if there are regional differences in social movements on Twitter, I make use of the fact that 75 percent of tweets have a Twitter user location. Only 3 percent of tweets have a tweet location (see Table 1). I, therefore, rely on the user location and not tweet location to extract geographic information. The Twitter user location is not available in a pre-processed format. While some users indicate their federal state, county, city, or even zip code, others mention their country of residence. Thus, I need to harmonize the data. I use information on all census-recognized cities/towns provided by

Figure 7: Daily emotions of tweets (2014-2017)

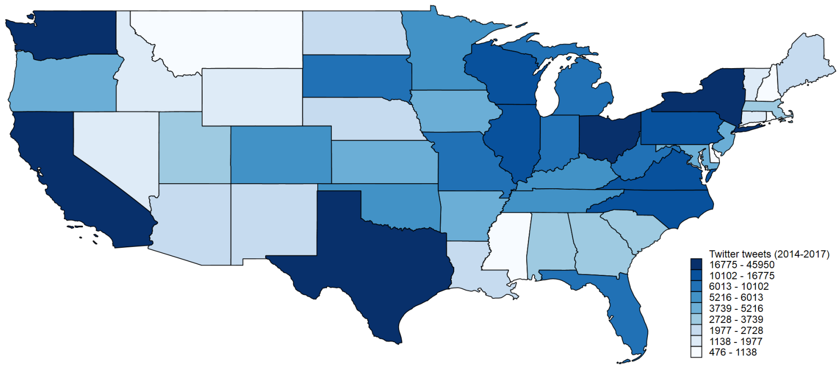



Notes: The graph plots the share of emotions detected by the Twitter Emotion Recognition Tool developed by Colnerič and Demšar (2018). The purple line represents the daily share of written content which is marked by fear, the red line the one marked by sadness, the green line the one marked by anger, while the blue line refers to the share of happy content. Source: Twitter (2014-2017).

SimpleMaps (2012). This data set contains information on the federal state, the federal state ID, the county and county ID, the city ID and zip codes. I can assign 33.0 percent of tweets a federal state in the US. I 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 million out of 206 million users worldwide) (Statista, 2022). Appendix C details the harmonization process.

Figure 8 presents the aggregated number of tweets per federal state for the period 2014-2017. Darker colors indicate a larger absolute number of Twitter tweets. States with the highest number of tweets are California, Washington, Texas, Ohio and New York. Those with the lowest number of tweets are Mississipi, Montana, and New Hampshire. The map reveals that there is considerable variation in the number of tweets at the federal state level.

Figure 8: Total number of Twitter tweets per federal state (2014-2017)




Notes: The map plots the number of English tweets per federal state for the period 2014-2017 relying on Twitter users' location and the methodology described in this section. The graph excludes Alaska, Hawaii, and Puerto Rico. Darker colors indicate a higher number of Twitter tweets. Source: Twitter (2014-2017).

# 5 Drivers of Tweets' Traction

## 5.1 Empirical Strategy

To investigate what drives the spread of Twitter tweets more closely, I estimate a simple linear regression. More specifically, I apply the following ordinary least square (OLS) regression:

$$Y_{it} = \alpha_0 + \alpha_1 \times \sum_{i=n}^{N} X_{n,it} + \gamma_t + \epsilon_{it} \tag{1}$$

$\sum X_{n,it}$ is a set of explanatory variables N which could potentially explain the spread of tweets on Twitter. $Y_{it}$ is the number of likes of tweet i on day t. $\gamma_t$ is a time fixed effect. I control for month as well as year fixed effects. These time fixed effects capture unobservable factors which affect all tweets at the same time, such as a holiday season or economic downturns. $\epsilon_{it}$ is an unobservable error term which I assume to be normally distributed. Standard errors are robust.

I also estimate alternative regression specifications to test the validity of my findings. I start by estimating an OLS regression which abstracts from time fixed effects. Next, I run a poisson regression. Poisson models are often used when the outcome variable is count data. The underlying assumption is that the number of likes can be described by a Poisson distribution. Lastly, I replace my outcome variable by a dummy variable which is equal to one as soon as a tweet has at least one like. I then estimate a linear probability model and a probit model. These regressions can shed light on underlying drivers of Twitter tweets' visibility at the extensive margin.

## 5.2 Results

Table 2 shows results for linear regressions on the number of likes of each tweet. In Column 1, I run a simple OLS regression without time fixed effects. In Column 2 to 5, I present robustness tests to validate the findings of my main empirical specification. In Column 2, I run ordinary least square regressions and include year and month fixed effects. In Column 3, I presents results for a poisson model. In Column 4 an 5, I report results from a linear probability and probit model.

The findings in Table 2 point to the crucial role of social leaders in the spread of social movements. The number of likes of a respective tweet is significantly correlated with the number of followers of its authors. An increase of one follower per 1,000 followers increases the number of likes of a given tweet by 0.0137 likes, which is equivalent to a 1 percent increase over the mean value. The coefficient in Row 1 barely changes when including fixed effects in Column 2. The coefficient on the number of followers remains significant under the alternative model specifications detailed in Column 3 to 5. There is also a significant association between the number of accounts a Twitter user, who posts a tweet, follows. One additional account followed by a tweet's author increases the number of likes of a given tweet by 0.0349, which is equivalent to 2 percent over the mean. Similar to my findings on the number of followers, the coefficient is robust across model specifications. These findings show that Twitter users with an established network play an important role in disseminating the social movements against GBV on Twitter.

My findings are in line with previous research showing that leadership is a crucial mechanism in addressing GBV (Iyer et al. (2012); Wen (2021); Delaporte and Pino (2022)) as well as stereotypes (Besley and Ghatak, 2018). In addition, it confirms that social networks are important to advance gender equality (Agarwal et al., 2016), as are group dynamics (Gagliarducci and Paserman, 2012) and gender norms (González and Rodríguez-Planas, 2020).

A related body of literature shows that it is often female leaders who drive change in the interest of women (Bhalotra and Clots-Figueras (2014); Brollo and Troiano (2016); Besley and Ghatak (2018) Bhalotra et al. (2018); Flabbi et al. (2019); Bertrand et al. (2019)). Based on these findings, I investigate if the impact of social leaders differs by their gender. I present the results in Appendix E.2. These regressions demonstrate that, while tweets posted by female authors generate on average more traction, the impact of leadership does not differ significantly by gender. Consequently, a user's network is more important than their gender.

Table 2 reveals that tweets that have attachments, such as media attachments, gain higher traction in terms of likes, which could point towards a significant relationship between social media and the traditional media in the spread of information. Having a media attachment included in a given tweet increases the number of likes of this tweet by 1.197. In terms of magnitude, the number of likes doubles compared to the mean value. Having a look at the results in Column 2 to 5 makes clear that the coefficient remains significant across model specifications.

The fact that media attachments in tweets generate higher traction of tweets is in line with research by Cooper et al. (2020). Their paper demonstrates that exposure to video dramatizing of violence against women and girls increases the likelihood of reporting such crimes. Similarly, Haraldsson and Wängnerud (2019) find a significant relationship between media sexism and women's political ambition.

Interestingly, revealing geographic information on the tweet itself or the tweet's author is also significantly correlated with the number of likes tweets receive. The coefficient in Column 1 and Row 5 demonstrates that a tweet with geographic information increases the number of likes by 1.002 likes. Compared to the mean value, this is an increase of 60 percent. The fact of indicating a user's location also increases the number of likes posted by a given user. The coefficient in Row 6 and Column 1 indicates that revealing a user's location increases the number of likes by 0.435. Having geographic information in one's tweet is more effective in generating traction than having geographic information in one's profile picture. A possible explanation could be that users who are located in a given location relate more to a tweet which refers to that same location. In addition, if a tweet relates to a certain location, the tweet's content might be less abstract and users who see the tweet might relate more to it on a personal level.

Neither sentiments nor emotions expressed in a tweet's text are significantly related to the number of likes of a tweet. The coefficients presented in Column 1 to 3 are insignificant at the common significance levels. Appendix E.1 presents additional descriptive evidence on the lack of a significant relationship between sentiments and tweets' traction. The tweets' content does play a role at the extensive margin though. The coefficients on the compound score and emotional scores reported in Column 4 and 5 are significant at the 1 percent significance level. While an increase in the compound score decreases the probability of having at least one like, happy, angry, sad and fearful emotions lead to a higher probability of having at least one like. Surprise, on the other hand, decreases that probability. Based on these findings, I conclude that emotions and polarized content play a role in generating traction of tweets in the first place, but once they generated a certain extent of traction, they do not lead to higher visibility.

The insignificant relationship between emotions and likes is not in line with previous research. Stieglitz and Dang-Xuan (2013), for example, find that emotionally charged Twitter tweets are retweeted more often. The fact that this is not the case for tweets about GBV could be related to the shaming, silencing and taboo around the topic.

Table 2: Regression for No. of likes

| | OLS | OLS | POISSON | LPM | PROBIT |
|---|---|---|---|---|---|
| main | | | | | |
| Follower (in 1,000s) | 0.0137*** | 0.0137*** | 0.000182*** | 0.0000325*** | 0.000114*** |
| | (0.00413) | (0.00413) | (0.0000207) | (0.00000476) | (0.0000226) |
| Following (in 1,000s) | 0.0349*** | 0.0303** | 0.00650*** | 0.00132*** | 0.00524*** |
| | (0.0127) | (0.0125) | (0.000616) | (0.0000884) | (0.000327) |
| Has attachment=1 | 1.197*** | 1.240*** | 1.248*** | 0.0416*** | 0.158*** |
| | (0.224) | (0.217) | (0.227) | (0.00128) | (0.00477) |
| Has mentions=1 | -2.729*** | -2.868*** | -2.729*** | -0.234*** | -0.852*** |
| | (0.241) | (0.250) | (0.245) | (0.000833) | (0.00292) |
| Has geo location=1 | 1.002** | 0.829* | 0.619** | 0.198*** | 0.588*** |
| | (0.441) | (0.449) | (0.297) | (0.00308) | (0.00827) |
| Has user location=1 | 0.435*** | 0.345** | 0.543*** | 0.0118*** | 0.0534*** |
| | (0.149) | (0.143) | (0.162) | (0.000821) | (0.00341) |
| Compound Sentiment | -0.199 | -0.293 | -0.182 | -0.00265*** | -0.00803** |
| | (0.257) | (0.264) | (0.252) | (0.000837) | (0.00337) |
| Happy Score | 0.190 | 0.276 | 0.184 | 0.0125*** | 0.0495*** |
| | (0.293) | (0.285) | (0.293) | (0.00157) | (0.00619) |
| Angry Score | -0.358* | -0.252 | -0.425* | 0.0123*** | 0.0507*** |
| | (0.194) | (0.174) | (0.235) | (0.00216) | (0.00845) |
| Surprise Score | 0.316 | 0.513 | 0.315 | -0.0166*** | -0.0635*** |
| | (0.683) | (0.723) | (0.627) | (0.00129) | (0.00535) |
| Sad Score | 0.0764 | -0.0477 | 0.0524 | 0.00739*** | 0.0300*** |
| | (0.200) | (0.194) | (0.207) | (0.00124) | (0.00500) |
| Fear Score | 0.130 | -0.0685 | 0.111 | 0.00647*** | 0.0254*** |
| | (0.202) | (0.197) | (0.205) | (0.00110) | (0.00443) |
| Constant | 2.598*** | 1.138*** | | 0.315*** | -0.511*** |
| | (0.221) | (0.174) | | (0.00117) | (0.00426) |
| Mean (Dep. Var) | 1.645 | 1.645 | 1.645 | 0.192 | 1.645 |
| St. Dv. (Dep. Var.) | 95.093 | 95.093 | 95.093 | 0.394 | 95.093 |
| Time fixed effects | No | Yes | No | No | No |
| N | 1085336 | 1085336 | 1085336 | 1085336 | 1085336 |

Notes: The table presents estimated coefficients from linear regressions of tweet characteristics on the number of likes of a given tweet. The level of analysis is a tweet. The first two columns show results from ordinary least square regressions (without and with year and month fixed effects) and the third column the ones from a poisson regression. The fourth column presents results from a linear probability model and the fifth one from a probit model. In Column 4 and 5, the outcome variable is a dummy variable which is equal to one if a given tweet received at least one like. The *Compound Score* is a score which ranges from -1 to 1 and is generated via text mining methods to mirror the sentiments of a given tweet. Similarly, emotional scores are deduced by text mining methods and reflect the average share of emotions in a given tweet's text. Standard errors are robust and in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Source: Twitter (2014-2017).

# 6 The Role of Social Leaders on Social Media

## 6.1 Empirical Strategy

My results from ordinary least square regressions point to the importance of social leaders in the spread of GBV-related movements on Twitter. However, the number of followers might be related to other factors which also drive the number of likes of a tweet. If there is an omitted variable driving both the number of likes of a tweet and the number of followers of the tweet's author, the coefficients presented in Table 2 might be biased. One example of a potential omitted variable would be a Twitter bot. In this case, the omitted variable would be the status of a Twitter account (being a Twitter bot or not). In the presence of these omitted variables, it is not possible to conclude that leaders on social media play an important role in the spread of GBV-related social movements on Twitter just by relying on ordinary least square regressions that estimate the number of followers on the number of a tweet's likes.

To address these concerns, I develop a novel instrumental variable. To this end, I exploit the fact that Twitter verifies certain accounts. The verification of Twitter accounts is subject to several conditions.[9] To be verified, a Twitter account needs to be authentic, notable, and active. When Twitter users apply for verification, Twitter requests proof from the applicant which shows that all of these three criteria are full-filled. Proof can be provided by sending links to an official website and evidence on news coverage or google searches, among others. The following entities are eligible to a Twitter account verification: governments; news organizations; individuals and journalists; companies, brands, and organizations; entities working in entertainment or sports and gaming; activists and organizers; content creators and influential individuals. Additional selection criteria apply to each one of these categories.

Based on these conditions, I conclude that the verification status of an account might better approximate the status of a social leader on Twitter. I then replicate the ordinary least square regressions from before but include a dummy variable for being verified as the main explanatory variable. In addition, I use the verification status as an instrumental variable for the number of Twitter users. I then estimate a regression as follows:

$$Y_{it} = \alpha_0 + \alpha_1 \times \sum_{i=n}^{N} X_{n,it} + \alpha_2 \times \hat{L_{it}} + \gamma_t + \epsilon_{it} \tag{2}$$

[9]For details see: https://help.Twitter.com/en/managing-your-account/about-Twitter-verified-accounts

$\sum X_{n,it}$ is the same set of explanatory variables N, which could potentially explain the spread of tweets on Twitter and $Y_{it}$ is the number of likes of tweet t. $\gamma_t$ is a time fixed effect. Similar to the previous regressions, I control for month fixed effects and year fixed effects. $\hat{L_{it}}$ is the number of followers of a given Twitter account, instrumented by its verification status as follows:

$$\hat{L_{it}} = \beta_0 + \beta_1 \times V_{it} + \omega_{it} \tag{3}$$

Table 3 shows the results of the first-stage regression. The coefficient is positive and significant at the 1 percent significance level and the F-statistic is 650 and well above 10. Consequently, the instrument is a strong instrument.

Table 3: First-stage regression

| | First-stage regression |
|---|---|
| | (1) |
| VARIABLES | Follower (in 1 000s) |
| | |
| Verified account | 197.2*** |
| | (7.731) |
| Constant | 2.803*** |
| | (0.0226) |
| | |
| Observations | 1,085,336 |
| R-squared | 0.019 |
| F-statistic | 650.6 |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: The table presents the first stage results of the verification status of a given Twitter account and the number of followers of that account. Verification is a dummy which is equal to one if the account is officially verified by Twitter. The unit of analysis is the tweet level. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Source: Twitter (2014-2017).

## 6.2 Results

Table 4 presents the results from the regressions specified previously and demonstrates that the significant effect of social leadership, approximated by the verification status of a Twitter account, persists. Column 1 and 2 present findings from ordinary least square regressions, using the verification status of an account as the main approximation for social leadership. The coefficients on being a verified account are significant and positive. When a tweet is posted by a verified account, the number of likes increases by 21.9, which is equivalent to a 1,333 percent increase over the mean. Differently from previous regressions, revealing a tweet's geography or a user's geography does not play a significant role in the number of likes a tweet receives. Attachments continue to boost the number of likes of a tweet and there is a negative relationship with mentions in a tweet's text.

Column 3 and 4 report the results from instrumental variable regressions. The coefficient on the number of followers is significant at the 1 percent significance level, similar to previous observations. In terms of magnitude, coefficients from IV regressions are approximately 10 times larger, which indicates that there might be an omitted variable bias in the main specification. Similarly to findings from the OLS regressions, the introduction of a dummy for being female reveals that tweets posted by female authors generate higher traction than the ones by male authors, but the impact of leadership does not differ by gender. Appendix E.2 details the results.

Table 4: Instrumental variable regressions

| | OLS | OLS | IV | IV |
|---|---|---|---|---|
| Verified account=1 | 21.92*** | 21.86*** | | |
| | (2.927) | (2.927) | | |
| Has attachment=1 | 1.064*** | 1.115*** | 0.211 | 0.237 |
| | (0.223) | (0.215) | (0.332) | (0.324) |
| Has mentions=1 | -2.691*** | -2.822*** | -2.546*** | -2.686*** |
| | (0.239) | (0.247) | (0.226) | (0.237) |
| Has geo location=1 | 0.882* | 0.721 | 1.543*** | 1.301** |
| | (0.449) | (0.458) | (0.421) | (0.429) |
| Has user location=1 | -0.0117 | -0.0940 | 0.443** | 0.351* |
| | (0.119) | (0.117) | (0.169) | (0.164) |
| Compound Sentiment | -0.217 | -0.303 | -0.201 | -0.299 |
| | (0.258) | (0.265) | (0.265) | (0.272) |
| Happy Score | 0.167 | 0.250 | 0.243 | 0.337 |
| | (0.293) | (0.284) | (0.310) | (0.301) |
| Angry Score | -0.398* | -0.290 | -0.106 | 0.0255 |
| | (0.199) | (0.177) | (0.191) | (0.175) |
| Surprise Score | 0.329 | 0.521 | 0.202 | 0.476 |
| | (0.687) | (0.725) | (0.678) | (0.724) |
| Sad Score | 0.0597 | -0.0595 | 0.123 | 0.0320 |
| | (0.201) | (0.195) | (0.212) | (0.203) |
| Fear Score | 0.113 | -0.0769 | 0.0437 | -0.128 |
| | (0.202) | (0.197) | (0.220) | (0.215) |
| Following (in 1,000s) | 0.00991 | 0.00597 | -0.129*** | -0.133*** |
| | (0.0121) | (0.0120) | (0.0291) | (0.0291) |
| Follower (in 1,000s) | | | 0.113*** | 0.113*** |
| | | | (0.0157) | (0.0157) |
| Constant | 2.431*** | 1.053*** | 1.974*** | 0.670** |
| | (0.209) | (0.171) | (0.205) | (0.208) |
| Mean (Dep. Var) | 1.645 | 1.645 | 1.645 | 1.645 |
| St. Dv. (Dep. Var.) | 95.093 | 95.093 | 95.093 | 95.093 |
| Time fixed effects | No | Yes | No | Yes |
| N | 1085336 | 1085336 | 1085336 | 1085336 |

Notes: The table presents estimated coefficients from regressions on the number of likes. The level of analysis is a tweet. The first two columns show the results from ordinary least square regressions (without and with year and month fixed effects). The third and fourth column present results from an instrumental variable regression, in which I instrument the number of followers by a dummy variable which is equal to one for all accounts that are officially verified by Twitter. I first present results without time fixed effects. I then include year and month fixed effects as additional controls. The *Compound Score* is a score which ranges from -1 to 1 and is generated via text mining methods to mirror the sentiments of a given tweet. Similarly, emotional scores are deduced by text mining methods and reflect the average share of emotions in a given tweet's text. Standard errors are robust and in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

# 7 Inclusiveness of Social Movements against GBV

I next investigate who contributes to the 10 social movements studied in this paper based on gender and ethnicity. This analysis can shed light on the inclusiveness of social movements on Twitter. I apply two different tools to retrieve information on Twitter users. First, I 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[10]. It allows me to retrieve users' age, gender, emotion, and race from their profile pictures. I 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 so-called "very deep" network. I employ the VGG-Face model to a 20 percent random sample of my overall data set.

Figure 9 and 10 depict the results. Twitter users who engage in GBV-related topics are on average 28.8 years old. In addition, the share of users above 50 years old is negligible. Moreover, most of users are White (78.9 percent), followed by Asian (10.5 percent), Middle-Eastern (5.8 percent), Black (2.8 percent), and Hispanic (1.5 percent). A small share is Indian (0.005 percent). The share of White contributing to the 10 social movements studied in this paper is higher than recent numbers describing the overall population in the US reported by the US Census Bureau (a share of 61.6 percent in 2020).[11]

The social movements against GBV studied in this paper seem to lag behind in their social inclusiveness. The views and voices of minority groups might be underrepresented within these movements, or might take place on separate Twitter spaces through relying on specialized hashtags, such as the ones examined by Peterson-Salahuddin (2022). To further investigate a potential under-representation and lag of agency in these movements, I replicate previous regressions but include an indicator variable for being non-white. The results in Appendix E.3 show that tweets posted by non-white Twitter users spread to a lower extent than tweets posted by white Twitter users. In addition, there is some evidence pointing to a lower influence of non-white social leaders when compared to white social leaders, but the results are less conclusive. Overall, these regressions confirm that the social movements studied in this paper lag social inclusiveness.

To shed light on the gender of those who contribute to the social movements studied

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

[11] For the link to the underlying study by the US Census Bureau see: https://www.census.gov/library/stories/2021/08/improved-race-ethnicity-measures-reveal-united-states-population-much-more-multiracial.html

Figure 9: Histogram of Twitter users' age

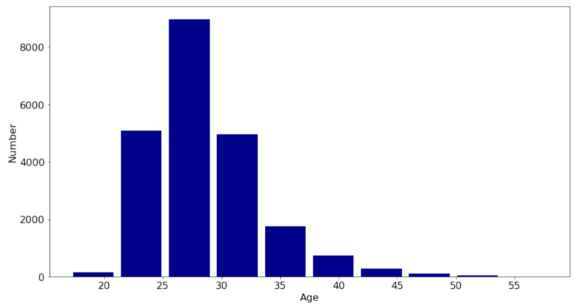
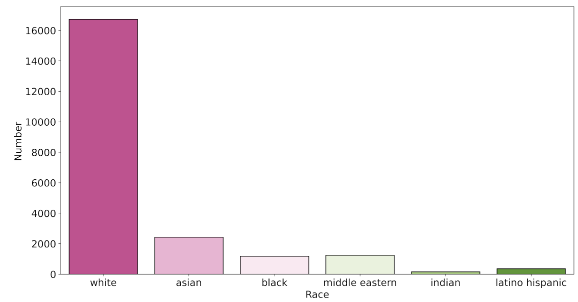



Figure 10: Distribution of Twitter users' race




Notes: The left figure depicts the age distribution of Twitter users who contribute to the social media movements analyzed in this paper. The right figure plots the distribution of their race. I employ a VGG-Face model to detect users' socioeconomic characteristics. I restrict the data set to a 20 percent sample of the generated data set. Source: Twitter (2014-2016).

in this paper, I employ the *GenderGuesser*.[12] This package allows for the detection of users' 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. Figure 11 illustrates the distribution of users' gender, which is mostly unknown (49.9 percent), followed by female (28.3 percent), male (14.5 percent), mostly female (4.0 percent) and mostly male (2.0 percent). A small share is classified as *andy* (1.4 percent). This pattern of results reveals that there are twice as many women contributing to the social movements studied in this paper as men, casting further doubt on the inclusiveness of these social movements. The larger engagement of women is expected given that they are more affected by GBV than men.

There are also clear gender patterns when analyzing the gender of tweets' authors relative to referenced tweets' authors (see Figure 12). When female Twitter users retweet, quote, or reply to previous tweets that are part of the social movements studied in this paper, they are more likely to engage with tweets from female authors. 54.0 percent of referenced tweets were posted by women (62.5 percent when also considering those classified as "mainly female"). Referenced tweets by women are mainly referenced by women. Only 29.2 percent of tweets posted by female authors are retweeted, quotes or replied to by male authors. In the case of male authors, the share of men referencing these tweets is higher, although there are overall less men engaging in the full mass of

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

Figure 11: Gender of Twitter users (2014-2017)

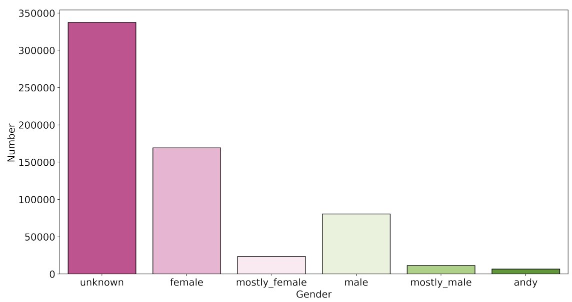
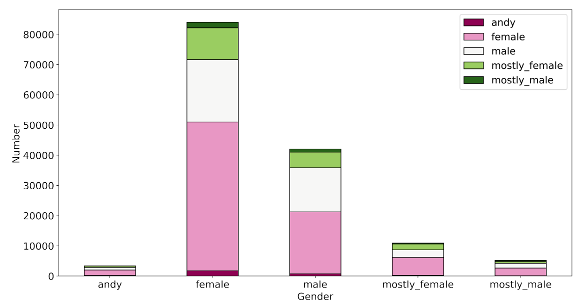



Figure 12: Referenced tweets and gender patterns




Notes: The left figure plots the gender of Twitter users included in my data set generated via the *GenderGuesser*. The right figure plots the cross-tabulation of Twitter users' gender by tweets and referenced tweets (considering retweets, quotes, and replies). The y-axis reports the gender of tweets' authors while the legend reports the gender of referenced tweets' authors. The right graph relies on the subset of my data set which refers to previously posted tweets. I drop all tweets and referenced tweets for which a gender detection based on user names was not possible. Source: Twitter (2014-2017).

tweets analyzed in this paper. 40.1 percent of those who retweet, quote, or reply to a man's tweet that contributes to social movements against GBV are men. This pattern of results indicates that women are more likely to reference content by women, while men are more likely to reference content by men. These dynamics could reinforce the lagging inclusiveness of social movements against GBV.

# 8 Further Robustness Checks

The inclusion of retweets, quotes, and replies using the 10 hashtags investigated in this paper into the main sample might underestimate the true spread of Twitter tweets if non-original tweets have a lower visibility than original tweets. Therefore, I repeat earlier analyses for a sample, which only consists of original tweets. Appendix E.4 presents the results. These are in line with the ones on the main sample.

The number of likes is not the only way through which tweets gain visibility on Twitter. Hence, I use alternative measures to proxy a tweet's spread: the number of retweets, the number of quotes, and the number of replies to a given tweet. These regressions confirm my findings from using likes as a measure for the spread of Twitter tweets. Appendix E.5 shows the results.

I test the validity of the Face Recognition Tool I apply in this paper by using data provided by Zhang and Qi (2017). Their data set consists of more than 20,000 photos of faces. Each photo is labeled with respect to the age, gender and race. Age is an integer from 0 to 116, gender a dummy variable which is equal to one for women, race is an indicator variable which ranges from 0 to 4, denoting White, Black, Asian, Indian, and Others (like Hispanic, Latino, Middle Eastern). I apply the Face Recognition Tool to a subset of 10,000 photos of their data set and compare the predicted age, gender and ethnicity to the labeled information.

Overall, predictions are similar. Still, the tool seems to under-predict the share of black people in the sample.[13] Consequently, my results on the ethnic under-representation of Twitter users who engage in the social movements against GBV should be taken with some caution. They could be confounded by imprecise or inaccurate predictions. Although it is important to flag this caveat, it is unlikely that inaccuracies in the prediction are serious enough to cancel out the under-representation of ethnic minorities. See Appendix D.1 for the details on the validation exercise.

# 9 Conclusion

In this paper, I investigate the spread of social movements against GBV on social media. For this purpose, I generate a new data set measuring 10 social movements against GBV which spread on Twitter. I retrieve data from the Twitter API using a hashtag-based approach and filter for all tweets that mention hashtags which clearly signal a contribution to 10 of the most well-known social movements against GBV during the period 2014-2017.

My findings indicate that social movements on Twitter start suddenly and fade out quickly. In addition, there are more retweets than original tweets, and most tweets do not generate considerable traction. Moreover, I study Twitter users' location and find that there is significant variation in the number of tweets by federal state. To further investigate potential drivers between the tweets' traction, I analyze the written text of tweets. Through a sentiment analysis, I show that tweets are mostly neutral. In addition, more extreme tweets do not result in higher traction. Twitter users who contribute to the movements mainly express fear and sadness, but emotions are uncorrelated with the spread of tweets. I apply ordinary least square regressions and a novel instrumental variable strategy to demonstrate that Twitter users with an established network play an important role in disseminating the social movements against GBV on Twitter. These results speak for the importance of social leaders in driving social change. The influence of

[13] The tool also under-predicts the share of women in the sample, but given that I deduce users' gender from their names, I do not discuss this weakness here.

social leaders does not significantly differ by their gender. Media attachments also result in higher traction of tweets.

My findings raise concerns about the inclusiveness of the social movements studied in this paper. A face recognition exercise of profile pictures and a gender detection analysis of Twitter users' names reveals that a larger share of those who contribute are female. Over half of users are White, and they are young on average (28.8 years old). Tweets posted by non-white users as part of the social movements spread to a lower extent than those posted by white users. Comparing the gender of authors who retweet, quote, and reply, to the gender of referenced tweets' authors reveals clear gender patterns. Women are more likely to reference tweets by women, while men engage to a larger extent with information posted by men.

In this paper, I generate valuable evidence on the drivers behind social movements online. In addition, my paper makes an important contribution to the literature which uses social media data to advance our understanding of GBV, especially with respect to the silencing and stigmatization around it. My work also builds on a growing body of literature studying how information spreads on social media. Based on the findings of this paper, policymakers should engage male leaders and the media in the case against GBV. They should also find ways to increase the social inclusivness of these movements. Future studies could explore if these online social movements against GBV impact offline behavior.

# References

Agarwal, Sumit et al. (2016). “Playing the boys game: Golf buddies and board diversity”. *American Economic Review* 106 (5), pp. 272–76.

Allcott, Hunt et al. (2020). “The welfare effects of social media”. *American Economic Review* 110 (3), pp. 629–76.

Anca, Vacar et al. (2014). “Leadership–a necessity in projects”. *Studies in Business & Economics* 9 (2), pp. 128–134.

Bakshy, Eytan, Solomon Messing, and Lada A Adamic (2015). “Exposure to ideologically diverse news and opinion on Facebook”. *Science* 348 (6239), pp. 1130–1132.

Battisti, Michele, Ilpo Kauppinen, and Britta Rude (2022). *Twitter and Crime: The Effect of Social Movements on GenderBased Violence*. Tech. rep. ifo Institute-Leibniz Institute for Economic Research at the University of Munich.

Beaman, Lori and Jeremy Magruder (2012). “Who gets the job referral? Evidence from a social networks experiment”. *American Economic Review* 102 (7), pp. 3574–93.

Bertrand, Marianne et al. (2019). “Breaking the glass ceiling? The effect of board quotas on female labour market outcomes in Norway”. *The Review of Economic Studies* 86 (1), pp. 191–239.

Besley, Timothy and Maitreesh Ghatak (2018). “Prosocial motivation and incentives”. *Annual Review of Economics* 10, pp. 411–438.

Bhalotra, Sonia and Irma Clots-Figueras (2014). “Health and the political agency of women”. *American Economic Journal: Economic Policy* 6 (2), pp. 164–97.

Bhalotra, Sonia, Irma Clots-Figueras, and Lakshmi Iyer (2018). “Pathbreakers? Women’s electoral success and future political participation”. *The Economic Journal* 128 (613), pp. 1844–1878.

Bovet, Alexandre and Hernán A Makse (2019). “Influence of fake news in Twitter during the 2016 US presidential election”. *Nature communications* 10 (1), pp. 1–14.

Braghieri, Luca, Ro’ee Levy, and Alexey Makarin (2022). “Social media and mental health”. *American Economic Review* 112 (11), pp. 3660–93.

Brollo, Fernanda and Ugo Troiano (2016). “What happens when a woman wins an election? Evidence from close races in Brazil”. *Journal of Development Economics* 122, pp. 28–45.

Campbell, Arthur (2013). “Word-of-mouth communication and percolation in social networks”. *American Economic Review* 103 (6), pp. 2466–98.

Chetty, Raj et al. (2022a). “Social capital I: measurement and associations with economic mobility”. *Nature* 608 (7921), pp. 108–121.

Chetty, Raj et al. (2022b). “Social capital II: determinants of economic connectedness”. *Nature* 608 (7921), pp. 122–134.

Chowdhury, Arijit Ghosh et al. (2019). “Speak up, fight back! detection of social media disclosures of sexual harassment”. *Proceedings of the 2019 conference of the North American chapter of the Association for Computational Linguistics: Student research workshop*, pp. 136–146.

Colnerič, Niko and Janez Demšar (2018). “Emotion recognition on twitter: Comparative study and training a unison model”. *IEEE transactions on affective computing* 11 (3), pp. 433–446.

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.

Cruz, Cesi, Julien Labonne, and Pablo Querubin (2017). “Politician family networks and electoral outcomes: Evidence from the Philippines”. *American Economic Review* 107 (10), pp. 3006–37.

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

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*.

Fersini, Elisabetta, Francesca Gasparini, and Silvia Corchs (2019). “Detecting sexist MEME on the Web: A study on textual and visual cues”. *2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*. IEEE, pp. 226–231.

Flabbi, Luca et al. (2019). “Do female executives make a difference? The impact of female leadership on gender gaps and firm performance”. *The Economic Journal* 129 (622), pp. 2390–2423.

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.

Garrett, Abigail and Naeemul Hassan (2019). “Understanding the silence of sexual harassment victims through the# whyididntreport movement”. *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 649–652.

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

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

Grinberg, Nir et al. (2019). "Fake news on Twitter during the 2016 US presidential election". *Science* 363 (6425), pp. 374–378.

Guess, Andrew, Jonathan Nagler, and Joshua Tucker (2019). "Less than you think: Prevalence and predictors of fake news dissemination on Facebook". *Science advances* 5 (1), eaau4586.

Haraldsson, Amanda and Lena Wängnerud (2019). "The effect of media sexism on women's political ambition: evidence from a worldwide study". *Feminist media studies* 19 (4), pp. 525–541.

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.

Jackson, Matthew O and Brian W Rogers (2007). "Meeting strangers and friends of friends: How random are social networks?" *American Economic Review* 97 (3), pp. 890–915.

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.

Knoke, David and Song Yang (2019). *Social network analysis*. SAGE publications.

Korda, Holly and Zena Itani (2013). "Harnessing social media for health promotion and behavior change". *Health promotion practice* 14 (1), pp. 15–23.

Öberg, Perola, Sven Oskarsson, and Torsten Svensson (2011). "Similarity vs. homogeneity: Contextual effects in explaining trust". *European Political Science Review* 3 (3), pp. 345–369.

Overstreet, Nicole M and Diane M Quinn (2013). "The intimate partner violence stigmatization model and barriers to help seeking". *Basic and applied social psychology* 35 (1), pp. 109–122.

Pamungkas, Endang Wahyu, Valerio Basile, and Viviana Patti (2020). "Misogyny detection in twitter: a multilingual and cross-domain study". *Information Processing & Management* 57 (6), p. 102360.

Pandey, Rahul et al. (2018). "Distributional semantics approach to detect intent in twitter conversations on sexual assaults". *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*. IEEE, pp. 270–277.

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

Patroni, Joanne, Frederik von Briel, and Jan Recker (2020). "Unpacking the social media–driven innovation capability: How consumer conversations turn into organizational innovations". *Information & Management*, p. 103267.

Peterson-Salahuddin, Chelsea (2022). "Posting Back: Exploring Platformed Black Feminist Communities on Twitter and Instagram". *Social Media+ Society* 8 (1).

Roden, Jessica, Matea Mustafaj, and Muniba Saleem (2021). "Who else likes it? Perceived gender of social endorsers predicts gender equality support". *Computers in Human Behavior* 118, p. 106696.

Rodríguez-Sánchez, Francisco, Jorge Carrillo-de Albornoz, and Laura Plaza (2020). "Automatic classification of sexism in social networks: An empirical study on twitter data". *IEEE Access* 8, pp. 219563–219576.

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`.

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

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/`.

Stieglitz, Stefan and Linh Dang-Xuan (2013). "Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior". *Journal of management information systems* 29 (4), pp. 217–248.

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

Veiga, André, E Glen Weyl, and Alexander White (2017). "Multidimensional platform design". *American Economic Review* 107 (5), pp. 191–95.

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

Zhang Zhifei, Song Yang and Hairong Qi (2017). "Age Progression/Regression by Conditional Adversarial Autoencoder". *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE.

Zobeidi, Tahereh, Nadejda Komendantova, and Masoud Yazdanpanah (2022). "Social media as a driver of the use of renewable energy: The perceptions of instagram users in Iran". *Energy Policy* 161, p. 112721.

# Online Appendix

## A Data Generation Process

I generate my final data set through the Twitter API for Academic researchers. I retrieve data from the Twitter database by creating an Academic Developer Account, which gives me access to the Twitter Full Archive Search API V2. This API allows me to access the full universe of Twitter tweets posted since the creation of Twitter. The API has a rate limit of 10 million tweets per month and 150,000 tweets per 15 minutes. I take advantage of the Twarc2 command line tool and Python library.

The retrieved data set is a nested pandas data frame consisting of the following variables: a conversation id, a dictionary with public metrics (such as the number of times the tweet was retweeted), a dictionary of referenced tweet information (with information of the referenced tweet and its author), the reply settings, the tweet author's id, the tweet's text, a unique ID, the source (such as Twitter for iPhone or Android), if the tweet is possibly sensitive, a dictionary with entities (such as mentions or hashtags included in the tweet), the exact date and time the tweet was created, the language of the tweet's text, a dictionary with information about the tweet's author (such as the user name or profile picture), information on the twarc interface, a dictionary with information on the context of the tweet, in case of the tweet being a reply to somebody to whom the author replied to, a dictionary with geographic settings (such as the the geolocation of the tweet), a dictionary with attachments, and withheld information. Figure A.1 shows the first 5 lines of the generated pandas data frame.

I normalize the remaining dictionaries and end up with 48 variables related to each individual tweet, which are: the context annotation, the date the tweet was created at, the author id, the conversation id, if the tweet is possibly sensitive, information on referenced tweets (which I keep as a dictionary at this point), reply settings, language, the source, the edit history of tweet ids, the tweet id, the tweet's text, information on the twarc interface, a dictionary with attachments, information on the user replied to (the user's name and ID, if it applies), the place id, the full name of the user, information on the geography of the tweet, a geographic ID, the user name, protection settings of the user, the pinned tweet of the user, if the user is verified, a dictionary with entities, the location of the user, the number of followers of the user, the number of people the user follows, the number of tweets the user has posted, the number of lists, the number of retweets, likes, replies, and quotes of the tweet, the hashtags mentioned in the tweet's text, urls included

Figure A.1: Pandas data frame generated via the Twarc Interface

```
data.head()
```

| | context_annotations | public_metrics | created_at | entities | author_id | conversation_id | possibly_sensitive | referenced_tweets | reply_setting |
|---|---|---|---|---|---|---|---|---|---|
| **0** | [{'domain': {'id': '30', 'name': 'Entities [En... | {'retweet_count': 4, 'reply_count': 0, 'like_c... | 2017-12-25T02:56:15.000Z | {'hashtags': [{'start': 15, 'end': 33, 'tag': ... | 602977481 | 945126011305431040 | False | [{'type': 'retweeted', 'id': '9449595828157808... | everyon |
| **1** | [{'domain': {'id': '30', 'name': 'Entities [En... | {'retweet_count': 4, 'reply_count': 0, 'like_c... | 2017-12-24T15:55:33.000Z | {'hashtags': [{'start': 15, 'end': 33, 'tag': ... | 3000154492 | 944959739645038592 | False | [{'type': 'retweeted', 'id': '9449595828157808... | everyon |
| **2** | [{'domain': {'id': '30', 'name': 'Entities [En... | {'retweet_count': 4, 'reply_count': 0, 'like_c... | 2017-12-24T15:54:55.000Z | {'urls': [{'start': 179, 'end': 202, 'url': 'h... | 261311170 | 944959582815780865 | False | [{'type': 'quoted', 'id': '944954460060647424'... | everyon |
| **3** | NaN | {'retweet_count': 0, 'reply_count': 3, 'like_c... | 2017-12-24T03:55:13.000Z | {'urls': [{'start': 277, 'end': 300, 'url': 'h... | 4330544475 | 944778462392438784 | False | NaN | everyon |
| **4** | NaN | {'retweet_count': 0, 'reply_count': 0, 'like_c... | 2017-12-19T00:49:39.000Z | {'urls': [{'start': 135, 'end': 158, 'url': 'h... | 339381808 | 942919826573422592 | False | NaN | everyon |

Notes: First five lines of the generated pandas data frame from the hashtag-based approach and Twarc Interface. Source: Twitter (2014-2017).

in the tweet, and annotations of the tweet (see Figure A.2.

Figure A.2: Pandas data frame generated via the Twarc Interface after normalization

```
data.head()
```

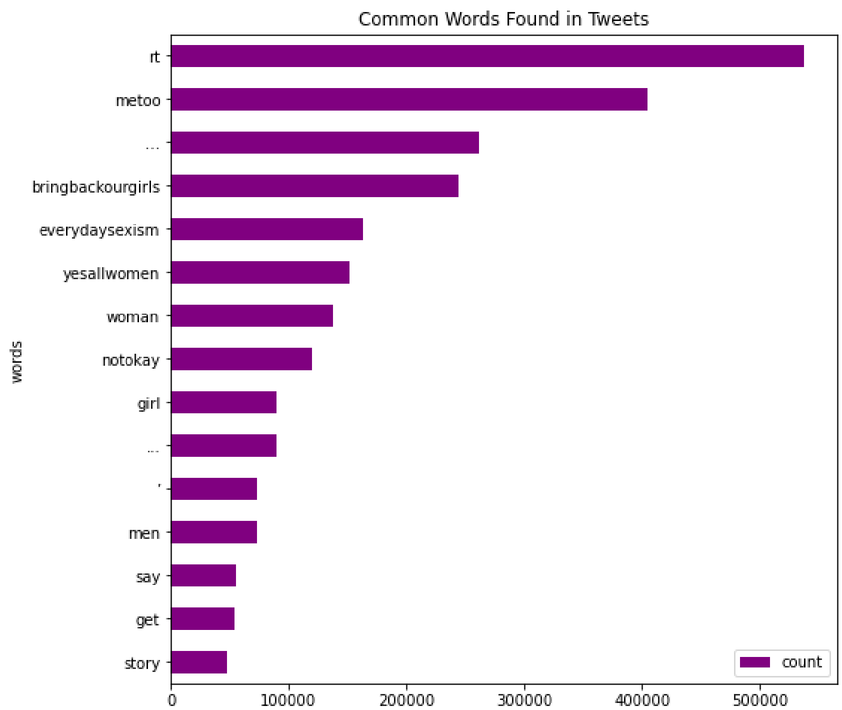
| | context_annotations | public_metrics | created_at | entities | author_id | conversation_id | possibly_sensitive | referenced_tweets | reply_setting |
|---|---|---|---|---|---|---|---|---|---|
| **0** | [{'domain': {'id': '30', 'name': 'Entities [En... | {'retweet_count': 4, 'reply_count': 0, 'like_c... | 2017-12-25T02:56:15.000Z | {'hashtags': [{'start': 15, 'end': 33, 'tag': ... | 602977481 | 945126011305431040 | False | [{'type': 'retweeted', 'id': '9449595828157808... | everyon |
| **1** | [{'domain': {'id': '30', 'name': 'Entities [En... | {'retweet_count': 4, 'reply_count': 0, 'like_c... | 2017-12-24T15:55:33.000Z | {'hashtags': [{'start': 15, 'end': 33, 'tag': ... | 3000154492 | 944959739645038592 | False | [{'type': 'retweeted', 'id': '9449595828157808... | everyon |
| **2** | [{'domain': {'id': '30', 'name': 'Entities [En... | {'retweet_count': 4, 'reply_count': 0, 'like_c... | 2017-12-24T15:54:55.000Z | {'urls': [{'start': 179, 'end': 202, 'url': 'h... | 261311170 | 944959582815780865 | False | [{'type': 'quoted', 'id': '944954460060647424'... | everyon |
| **3** | NaN | {'retweet_count': 0, 'reply_count': 3, 'like_c... | 2017-12-24T03:55:13.000Z | {'urls': [{'start': 277, 'end': 300, 'url': 'h... | 4330544475 | 944778462392438784 | False | NaN | everyon |
| **4** | NaN | {'retweet_count': 0, 'reply_count': 0, 'like_c... | 2017-12-19T00:49:39.000Z | {'urls': [{'start': 135, 'end': 158, 'url': 'h... | 339381808 | 942919826573422592 | False | NaN | everyon |

Notes: First five lines of the normalized pandas data frame. Source: Twitter (2014-2017).

# B Text Analysis

I follow standard procedures to prepare the Twitter tweets' text in a way that make their processing easier and less time consuming. I start by tokenizing the text of each tweet by splitting it into words and punctuation. For this purpose, I apply the *Tweet Tokenizer* method from the Natural Language Toolkit package from Python (*NLTK*). Figure B.1 shows the most frequent tokens in my data set.

Figure B.1: Word count of tokenized tweets




Source: Twitter (2014-2017)

I then remove urls, mentions, hashtags, and stop words from the resulting list of tokens. I then lemmatize the tokens by reducing words to their stem word. To give an example, the word "kidnapped" is reduced to its lemma "kidnap". I do not remove punctuation from the list of tokens, given that they identify smileys, such as in the case of a happy smiley: ":-)". I keep these smileys for the emotion and sentiment analysis of the tweets.

I next put the cleaned list of tokens back into a pandas data frame and then apply text mining methods to each tweet. Both in the case of the sentiment analyzer as in the emotion detection analyzer the result is a dictionary consisting of four different keys. Figure B.2 gives an example. I normalize these and generate four different columns in the final pandas data frame, each one describing one of the four keys. Table B.1 and B.2 present some examples of the resulting scores.

Figure B.2: Example of emotion detection

**Apply function to dataframe**

```
# creating new column 'polarity' in clean_df
df['emotion'] = df['cleaned_tweets'].apply(emo)
df.head()
```

| | cleaned_tweets_list | cleaned_tweets | emotion |
|---|---|---|---|
| **0** | [rt, bringbackourgirls, crazy, two, week, sinc... | rt bringbackourgirls crazy two week since kidn... | {'Happy': 0.0, 'Angry': 0.0, 'Surprise': 0.0, ... |
| **1** | [rt, please, sign, petition, retweet, bringbac... | rt please sign petition retweet bringbackourgi... | {'Happy': 0.0, 'Angry': 0.0, 'Surprise': 0.0, ... |
| **2** | [bringbackourgirls, please, speak, kidnapping,... | bringbackourgirls please speak kidnapping nige... | {'Happy': 0.0, 'Angry': 0.0, 'Surprise': 0.25,... |
| **3** | [rt, two, week, since, kidnapping, 234, nigeri... | rt two week since kidnapping 234 nigerian girl... | {'Happy': 0.0, 'Angry': 0.0, 'Surprise': 0.33,... |
| **4** | [rt, two, week, since, kidnapping, 234, nigeri... | rt two week since kidnapping 234 nigerian girl... | {'Happy': 0.0, 'Angry': 0.0, 'Surprise': 0.33,... |

Source: Twitter (2014-2017)

Table B.1: Example of emotion classification exercise

| text | Happy | Angry | Surprise | Sad | Fear |
|---|---|---|---|---|---|
| "@SanRomeo: Who knows maybe your future wife is among the 234 girls that was kidnapped #bringbackourgirls http://t.co/0KETJVTBTf" | 0.00 | 0.0 | 0.50 | 0.00 | 0.50 |
| "@rotexonline: Rain soaked, women put their shoes on the head and marched on #bringbackourgirls http://t.co/Fe1THLr2JC l see men too" | 0.00 | 0.0 | 0.00 | 0.33 | 0.67 |
| Download Video Chelsea 1 - 3 Atletico Madrid [Champions League] Highlights: http://t.co/OehZKLOFHV #bringbackourgirls | 0.50 | 0.0 | 0.00 | 0.00 | 0.50 |
| where is #Gej 's daughter that Got married. she wasn't kidnapped oh. #bringbackoursisters #bringbackourgirls | 0.00 | 0.0 | 0.00 | 0.00 | 1.00 |
| I encourage you to pay more attention to what is happening in west #Africa. There is a lot of hope, we must fight #bringbackourgirls #girls | 0.00 | 0.0 | 0.25 | 0.50 | 0.25 |
| It's been two weeks since the kidnapping of 234 Nigerian girls and they still aren't home #bringbackourgirls http://t.co/Er3wI4k9sy | 0.00 | 0.0 | 0.33 | 0.33 | 0.33 |
| Protesters urge Nigeria to step up hunt for girls abducted by Islamists http://t.co/4pw1vmpX7y via @reuters #bringbackourgirls | 0.33 | 0.0 | 0.67 | 0.00 | 0.00 |
| 234 #girls were abducted in #Nigeria. #bringbackourgirls.Make it headline news. @BBCNews @SkyNews @CNN @Reuters @itn @itvnews @ABC @CBSNews | 0.00 | 0.0 | 0.50 | 0.50 | 0.00 |
| I have nothing against GEJ as a person or anyone in his cabinet. I however also have low tolerance for incompetence. #bringbackourgirls" | 0.00 | 0.0 | 0.00 | 1.00 | 0.00 |
| "@eLDeeTheDon: I hv notin against GEJ as a person or any1 in his cabinet. I however also hv low tolerance 4 incompetence. #bringbackourgirls | 0.00 | 0.0 | 0.00 | 1.00 | 0.00 |
| It's been two weeks since the kidnapping of 234 Nigerian girls and they still aren't home #bringbackourgirls http://t.co/0KETJVTBTf" | 0.00 | 0.0 | 0.33 | 0.33 | 0.33 |
| @GirlsontheBall@AVLFCOfficial #bringbackourgirls please support the fight to get the kidnapped Nigeria school girls. | 0.00 | 0.0 | 0.33 | 0.33 | 0.33 |
| Too sad "@maryodia: Very sad "@iamwardlaw: 234 school girls, wow that's sad"" #bringbackourgirls | 0.00 | 0.0 | 0.40 | 0.60 | 0.00 |
| It's been two weeks since the kidnapping of 234 Nigerian girls and they still aren't home #bringbackourgirls :o http://t.co/2OQElimrd5 | 0.00 | 0.0 | 0.33 | 0.33 | 0.33 |
| 234 #girls were abducted in #Nigeria. #bringbackourgirls.Make it headline news. @BBCNews @SkyNews @CNN @Reuters @facebook @twitter | 0.00 | 0.0 | 0.50 | 0.50 | 0.00 |
| #bringbackourgirls these young ladies just want the same education we had. . http://t.co/8vDaJR4qLY | 0.00 | 0.0 | 0.00 | 0.00 | 0.00 |
| #Repost from youlovebullet with repostapp #bringbackourgirls \r\r—\r\rLet's pause the Donald Sterling. . . http://t.co/nB5RoXSxbL | 0.00 | 0.0 | 0.00 | 0.00 | 0.00 |
| If these girls are released today it's not because of #Nigeria Gov't but because of care Nigerians that participated... #bringbackourgirls | 0.00 | 0.0 | 1.00 | 0.00 | 0.00 |
| #bringbackourgirls this is crazy. It's been two weeks since they were kidnapped. #rp http://t.co/4L0JYeDPz9 | 0.00 | 0.0 | 0.00 | 0.00 | 1.00 |
| this is crazy #bringbackourgirls #gullygirls http://t.co/wbNJ9Td1GS | 0.00 | 0.0 | 0.00 | 0.00 | 1.00 |

Source: Twitter (2014-2017)

Table B.2: Example of sentiment analysis

| text | neg | neu | pos | compound |
|---|---|---|---|---|
| RT @singlikediamond: #bringbackourgirls this is... | 0.211 | 0.789 | 0.000 | -0.3400 |
| RT @VokePetra: Please sign petition and Retweet... | 0.000 | 0.723 | 0.277 | 0.3182 |
| @TheEllenShow #bringbackourgirls please speak o... | 0.134 | 0.672 | 0.193 | 0.1779 |
| RT @maryjblige: It's been two weeks since the k... | 0.000 | 1.000 | 0.000 | 0.0000 |
| RT @maryjblige: It's been two weeks since the k... | 0.000 | 1.000 | 0.000 | 0.0000 |
| RT @maryjblige: It's been two weeks since the k... | 0.000 | 1.000 | 0.000 | 0.0000 |
| RT @maryjblige: It's been two weeks since the k... | 0.000 | 1.000 | 0.000 | 0.0000 |
| RT @maryjblige: It's been two weeks since the k... | 0.000 | 1.000 | 0.000 | 0.0000 |
| RT @maryjblige: It's been two weeks since the k... | 0.000 | 1.000 | 0.000 | 0.0000 |
| RT @EBONYMag: Terror group still holding Nigeri... | 0.274 | 0.726 | 0.000 | -0.5267 |
| "@SanRomeo: Who knows maybe your future wife is... | 0.000 | 1.000 | 0.000 | 0.0000 |
| RT @maryjblige: It's been two weeks since the k... | 0.000 | 1.000 | 0.000 | 0.0000 |
| RT @common: We must help with this! #bringbacko... | 0.000 | 0.526 | 0.474 | 0.4019 |
| RT @maryjblige: It's been two weeks since the k... | 0.000 | 1.000 | 0.000 | 0.0000 |
| RT @maryjblige: It's been two weeks since the k... | 0.000 | 1.000 | 0.000 | 0.0000 |
| RT @mzT08: spread the word #bringbackourgirls i... | 0.000 | 1.000 | 0.000 | 0.0000 |
| RT @maryjblige: It's been two weeks since the k... | 0.000 | 1.000 | 0.000 | 0.0000 |
| "@rotexonline: Rain soaked, women put their sho... | 0.000 | 0.769 | 0.231 | 0.4588 |
| Download Video Chelsea 1 - 3 Atletico Madrid ... | 0.000 | 0.746 | 0.254 | 0.5267 |
| RT @common: We must help with this! #bringbacko... | 0.000 | 0.526 | 0.474 | 0.4019 |

Source: Twitter (2014-2017)

# C Harmonization of Twitter user locations

To harmonize Twitter users' location information, I first split the Twitter user location into different columns, based on commas. I 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 data set. I next merge both data sets on the city name and state name. Next, I use the county name and state name. I then use the county zip code and state name. As a next step, I merge both data sets 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, I 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). I repeat this procedure for the city name, county name, city zip and county zip code respectively.

In a second step, I address duplicated values. Duplicated values occur, as many cities in the US have the same name. If a city is duplicated, I keep the value with the largest population. I can assign 33.0 percent of tweets a federal state in the US. I believe that this captures a large enough share of all Twitter users. In 2021, 37.7 percent were from the United States (77.75 million out of 206 million users worldwide) (Statista, 2022).

# D Additional Material on Methods

## D.1 Face Recognition Exercise

I apply the Face Recognition Tool developed by Serengil and Ozpinar (2020) to a 20 percent sample of my data set. I choose the SSD detector backend. For details on the different backend options see: https://github.com/serengil/deepface. I do not implement a force option which requires the algorithm to hand out demographic information under all means. The face recognition tool is able to retrieve demographic information on users' fotos in 56,321 out of 217,067 cases, a share of 25.9 percent.

The pictures below show some case studies of the Face Recognition Tool I am applying in this paper. Especially the picture on Barack Obama shows that the tool faces some important empirical limitations. While Barack Obama is black, the algorithm identifies him as Asian. The mismatch could be due to the black-and-white image.

To investigate potential shortfalls of the face recognition tool by Serengil and Ozpinar (2020), I take advantage of a large data set of pictures provided by Zhang and Qi (2017),

Figure D.1: Face Analysis: 37 years old, Black, happy

Figure D.2: Face Analysis: 34 years old, White, angry

which includes labels on each photo's age, gender and race. I use a subset of 10,719 labeled photos and compare the predictions made by the algorithm to the labels provided by Zhang and Qi (2017). Table D.1 shows that the values are close to each other, but that there are still some significant deviations. The tool clearly under-predicts the share of women in the sample. While the share of women is 46.5 percent, the tool predicts a share of only 35.2 percent. Due to these deviations, I deduce a person's gender from their first names. Next, the tool slightly over-predicts the share of white people in the same (by 4 percentage points) and the share of Asians (by 4 percentage points). There is a significant deviation in the share of Black people. While the algorithm predicts a share of 22.0 percent, the true share of Black people in the sample is 35.6 percent.

Consequently, the results on ethnicity presented in this paper should be taken with some caution, as they might suffer from in-precise predictions made by the algorithm applied in this paper. However, as the share of White people only deviates by 4 percentage points, it is unlikely that the main messages of the principle analysis refers due to prediction errors.

Figure D.3: Face Analysis: 47 years old, Asian, happy

Figure D.4: Face Analysis: 52 years old, White, happy

Table D.1: Comparison of labeled characteristics and face recognition analysis

| | Age (predicted) | Age (true) | Gender (predicted) | Gender (true) | White (predicted) | White (true) | Black (predicted) | Black (true) | Asian | Asian (true) | Indian | Indian (true) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| mean | 32.63 | 34.62 | 0.35 | 0.47 | 0.41 | 0.37 | 0.22 | 0.36 | 0.21 | 0.17 | 0.03 | 0.05 |
| std | 5.63 | 13.82 | 0.48 | 0.50 | 0.49 | 0.48 | 0.41 | 0.48 | 0.41 | 0.38 | 0.16 | 0.16 |
| min | 16 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 25% | 29 | 26 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 50% | 32 | 30 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 75% | 35 | 40 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| max | 64 | 116 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

Notes: The table evaluates the face analysis of 10,000 pictures provided by Zhang and Qi (2017). I predict age, gender and ethnicity of the person in a given picture. I apply the face recognition tool by Serengil and Ozpinar (2020). I first report the predicted average value or share and then the true one.

## D.2 Gender Guesser

I show some examples on the gender deduced from users' names in the Table below. The table also shows how I first split the full names of users based on whitespaces. I then apply the *GenderGuesser* Tool to the first name and then the second name. In case the tool is able to detect a user's gender based on the first name, I rely on this deduction. In case the tool can only detect a user's gender based on the second name, I use this deduction.

Table D.2: Example of gender guesser

| author_id | user | FirstLastName | FirstName | SecondName | ThirdName | Gender | Gender2 | GenderFinal |
|---|---|---|---|---|---|---|---|---|
| 232721515 | SUPER STAR | SUPER STAR | SUPER | STAR | | unknown | unknown | unknown |
| 2397928686 | Falaiye henry | Falaiye henry | Falaiye | henry | | unknown | unknown | unknown |
| 335588690 | Andrea Bent | Andrea Bent | Andrea | Bent | | female | male | female |
| 32937031 | Darling Nky | Darling Nky | Darling | Nky | | unknown | unknown | unknown |
| 605099377 | Abubakar | Abubakar | Abubakar | | | unknown | unknown | unknown |
| 297132278 | Laura Coyle | Laura Coyle | Laura | Coyle | | female | unknown | female |
| 607970783 | olumighty | olumighty | olumighty | | | unknown | unknown | unknown |
| 180389078 | Juliette Gash | Juliette Gash | Juliette | Gash | | female | unknown | female |
| 193885259 | Tiffany N. Moore | Tiffany N. Moore | Tiffany | N. | Moore | female | unknown | female |
| 257515824 | H3R 4 LĂB™ | H3R 4 LĂB™ | H3R | 4 | LĂB™ | unknown | unknown | unknown |
| 104815677 | Adegoke A. Oluwatobi | Adegoke A. Oluwatobi | Adegoke | A. | Oluwatobi | unknown | unknown | unknown |
| 617453909 | jenny | jenny | jenny | | | unknown | unknown | unknown |
| 17686306 | Sonia Meggie | Sonia Meggie | Sonia | Meggie | | female | unknown | female |
| 203895204 | ABH | ABH | ABH | | | unknown | unknown | unknown |
| 1393560841 | JEN | JEN | JEN | | | unknown | unknown | unknown |
| 419586269 | Layinka | Layinka | Layinka | | | unknown | unknown | unknown |
| 347885761 | EA | EA | EA | | | unknown | unknown | unknown |
| 2314113700 | TheTops | TheTops | TheTops | | | unknown | unknown | unknown |
| 127496087 | E Pluribus Unum (Ayobami) | E Pluribus Unum (Ayobami) | E | Pluribus | Unum | unknown | unknown | unknown |
| 1023131827 | Rashonda James | Rashonda James | Rashonda | James | | unknown | male | male |

Notes: 20 examples of the gender deduction applied in this paper. I first apply the tool to the first name and then second name of a user's name. The resulting gender variable is a combination of both deductions. Source: Twitter data (2014-2017).

# E Additional Results

## E.1 The Role of Sentiments in the Spread of Twitter Tweets

I ask whether more extreme tweets drive the spread of social movements against GBV. In order to investigate this question, I plot the daily number of tweets against the average daily compound score. If polarity drives the spread of tweets, I would expect to find a U-shaped distribution of the relationship. Figure E.1 illustrates an agglomeration of data points around the compound score of 0, meaning that on days with tweets which are on average more neutral there are also more tweets. The correlation coefficient of both variables is -0.089, showing that there is a slightly negative correlation between the number of daily tweets and the sentiment score. Still, the correlation coefficient is close to zero. Therefore, polarity does not seem to play a significant role in the spread of Twitter tweets which contribute to social movements against GBV.

I repeat this analysis at the tweet level and plot the average sentiment score against the number of replies in Figure E.2, the number of likes in Figure E.3 and the number of retweets in Figure E.4. These figures confirm the lack of a clear relationship between the polarity of written content and the degree to which these tweets are shared and distributed by users. The related correlation coefficients are slightly negative but close to zero, showing that there is no strong correlation. These analyses at the tweet level confirm that polarity plays a minor role in the spread of social movements against GBV on Twitter.

Figure E.1: Daily compound scores and No. of tweets

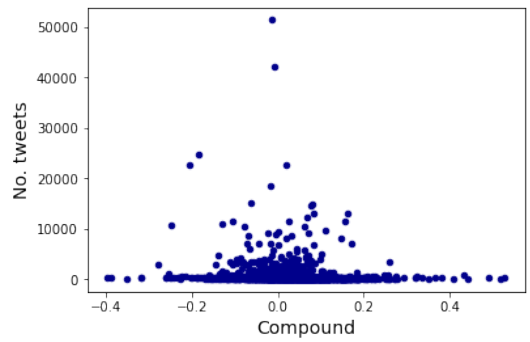
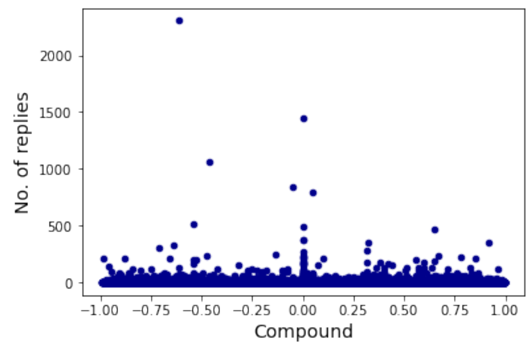



Figure E.2: Scatter plot of compound scores and replies




Notes: The left figure shows a scatter plot of the average daily compound score and the daily number of tweets for the period 2014-2017. The right figure shows a scatter plot of the compound score of each tweet and the number of replies to each tweet. Source: Twitter (2014-2017).

Figure E.3: Scatter plot of compound scores and number of likes

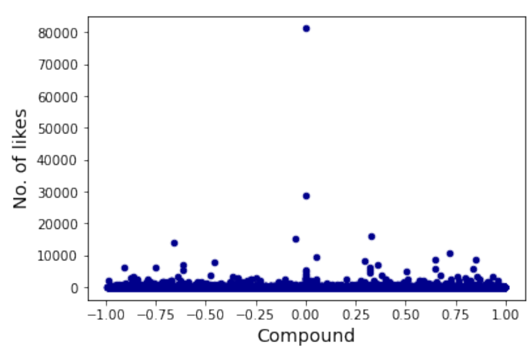
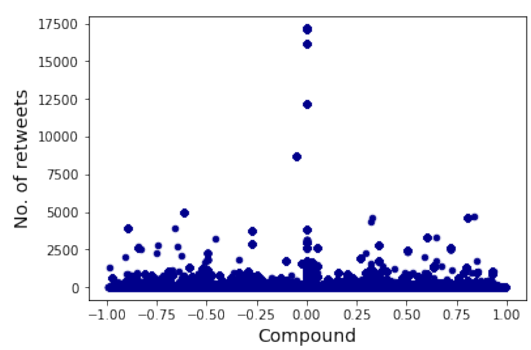



Figure E.4: Scatter plot of compound scores and retweets




Notes: The left figure shows a scatter plot of the compound score and the number of likes of each tweet. The right figure shows a scatter plot of the compound score and the number of retweets of each tweet. Source: Twitter (2014-2017).

## E.2 Social Leaders by Gender

Table E.1: Summary statistics of Twitter tweets (with gender)

| VARIABLES | (1) mean | (2) sd | (3) min | (4) max | (5) p25 | (6) p75 |
|---|---|---|---|---|---|---|
| Verified account | 0.03 | 0.17 | 0 | 1 | 0 | 0 |
| Has attachment | 0.09 | 0.29 | 0 | 1 | 0 | 0 |
| Follower (in 1,000s) | 8.92 | 246.66 | 0 | 60,544.17 | 0.19 | 1.87 |
| Following (in 1,000s) | 1.76 | 6.70 | 0 | 1,548.99 | 0.26 | 1.69 |
| No. of likes | 1.65 | 94.30 | 0 | 81,233 | 0 | 0 |
| No. of quotes | 0.04 | 3.61 | 0 | 2,583 | 0 | 0 |
| Has user location | 0.76 | 0.43 | 0 | 1 | 1 | 1 |
| Has geo location | 0.03 | 0.16 | 0 | 1 | 0 | 0 |
| original | 0.41 | 0.49 | 0 | 1 | 0 | 1 |
| No. of replies | 0.18 | 3.60 | 0 | 2,305 | 0 | 0 |
| No. of retweets | 765.78 | 2,893.14 | 0 | 17,160 | 0 | 60 |
| No. of tweets | 53,622.97 | 139,431.02 | 0 | 4,007,516 | 3,629 | 45,399 |
| Has mentions | 0.62 | 0.48 | 0 | 1 | 0 | 1 |
| Happy Score | 0.10 | 0.26 | 0 | 1 | 0 | 0 |
| Angry Score | 0.05 | 0.18 | 0 | 1 | 0 | 0 |
| Surprise Score | 0.16 | 0.30 | 0 | 1 | 0 | 0.33 |
| Sad Score | 0.18 | 0.32 | 0 | 1 | 0 | 0.33 |
| Fear Score | 0.26 | 0.37 | 0 | 1 | 0 | 0.50 |
| Negativity score | 0.13 | 0.17 | 0 | 0.95 | 0 | 0.25 |
| Neutrality score | 0.72 | 0.21 | 0.04 | 1 | 0.56 | 1 |
| Positivity score | 0.15 | 0.18 | 0 | 0.95 | 0 | 0.26 |
| Compound Sentiment | 0 | 0.45 | -0.99 | 1 | -0.32 | 0.36 |
| Female | 0.67 | 0.47 | 0 | 1 | 0 | 1 |
| Male | 0.33 | 0.47 | 0 | 1 | 0 | 1 |
| Unknown | 0.56 | 0.50 | 0 | 1 | 0 | 1 |

Notes: The table presents summary statistics of the data set generated via the Twitter API. I apply a hashtag-based approach, relying on 10 hashtags which identify GBV-related movements on Twitter, to filter on the full universe of Twitter tweets. I only consider the period 2014-2017. To extract the compound sentiment scores as well as the share of negative, neutral and positive tweets I employ the VADER Sentiment Analysis tool by Hutto and Gilbert (2014). To analyze emotions of written text I employ the emotion recognition tool developed by Colnerič and Demšar (2018). N=1,085,336. The female and male dummy variable have missing observations for the case that gender is unknown. This applies to 626,200 lines of data. The sample consequently drops significantly for regressions which include gender as a control variable. I apply the *genderguesser* tool to users' names to extract information on people's gender. Source: Twitter (2014-2017).

Table E.2: Regression for No. of likes

| | OLS | OLS | LPM |
|---|---|---|---|
| Follower (in 1,000s) | 0.0563** | 0.0564** | 0.000105*** |
| | (0.0244) | (0.0244) | (0.0000269) |
| Female=1 | 0.654*** | 0.670*** | 0.0587*** |
| | (0.175) | (0.178) | (0.00118) |
| Female=1 × Follower (in 1,000s) | -0.0372 | -0.0373 | -0.0000389 |
| | (0.0250) | (0.0251) | (0.0000326) |
| Following (in 1,000s) | 0.0279 | 0.0228 | 0.00146*** |
| | (0.0312) | (0.0306) | (0.000168) |
| Has attachment=1 | 1.654*** | 1.699*** | 0.0310*** |
| | (0.360) | (0.356) | (0.00199) |
| Has mentions=1 | -3.033*** | -3.208*** | -0.254*** |
| | (0.248) | (0.273) | (0.00129) |
| Has geo location=1 | 1.376** | 1.255* | 0.186*** |
| | (0.668) | (0.668) | (0.00412) |
| Has user location=1 | 0.515*** | 0.347** | 0.0228*** |
| | (0.174) | (0.171) | (0.00132) |
| Compound Sentiment | 0.167 | 0.0500 | -0.00153 |
| | (0.330) | (0.337) | (0.00131) |
| Happy Score | -0.357 | -0.181 | 0.0113*** |
| | (0.419) | (0.377) | (0.00245) |
| Angry Score | -0.291 | -0.132 | 0.0122*** |
| | (0.361) | (0.322) | (0.00337) |
| Surprise Score | -0.287 | -0.113 | -0.0131*** |
| | (0.361) | (0.312) | (0.00204) |
| Sad Score | 0.0659 | -0.0605 | 0.00652*** |
| | (0.323) | (0.290) | (0.00194) |
| Fear Score | 0.242 | 0.0393 | 0.00619*** |
| | (0.353) | (0.311) | (0.00172) |
| Constant | 2.543*** | 0.833*** | 0.309*** |
| | (0.439) | (0.302) | (0.00200) |
| Mean (Dep. Var) | 1.649 | 1.649 | 0.193 |
| St. Dv. (Dep. Var.) | 94.298 | 94.298 | 0.394 |
| Time fixed effects | No | Yes | No |
| N | 482915 | 482915 | 482915 |

Estimated coefficients from Poisson and Probit not shown due to a convergence error.
Standard errors in parentheses
Source: Twitter data. Standard errors are robust.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table E.3: Instrumental variable regressions

| | OLS | OLS | IV | IV |
|---|---|---|---|---|
| Verified account=1 | 18.10*** | 18.03*** | | |
| | (5.183) | (5.189) | | |
| Female=1 | 0.225** | 0.237** | 0.297* | 0.298* |
| | (0.0727) | (0.0758) | (0.147) | (0.148) |
| Verified account=1 × Female=1 | 3.085 | 3.060 | | |
| | (5.482) | (5.483) | | |
| Has attachment=1 | 1.487*** | 1.540*** | 1.257*** | 1.310*** |
| | (0.362) | (0.358) | (0.376) | (0.373) |
| Has mentions=1 | -3.019*** | -3.183*** | -2.862*** | -3.021*** |
| | (0.249) | (0.274) | (0.248) | (0.273) |
| Has geo location=1 | 1.257 | 1.150 | 1.906** | 1.730** |
| | (0.669) | (0.670) | (0.654) | (0.655) |
| Has user location=1 | -0.0210 | -0.171 | 0.533** | 0.365 |
| | (0.158) | (0.159) | (0.190) | (0.187) |
| Compound Sentiment | 0.146 | 0.0410 | 0.0910 | -0.0285 |
| | (0.330) | (0.338) | (0.334) | (0.339) |
| Happy Score | -0.368 | -0.198 | -0.393 | -0.202 |
| | (0.429) | (0.387) | (0.486) | (0.441) |
| Angry Score | -0.258 | -0.0972 | -0.183 | 0.0229 |
| | (0.359) | (0.320) | (0.389) | (0.348) |
| Surprise Score | -0.229 | -0.0625 | -0.457 | -0.208 |
| | (0.356) | (0.309) | (0.388) | (0.338) |
| Sad Score | 0.0525 | -0.0630 | 0.114 | 0.0114 |
| | (0.323) | (0.290) | (0.339) | (0.305) |
| Fear Score | 0.234 | 0.0513 | 0.145 | -0.0319 |
| | (0.353) | (0.311) | (0.366) | (0.326) |
| Following (in 1,000s) | 0.0409* | 0.0369* | -0.214*** | -0.217*** |
| | (0.0172) | (0.0169) | (0.0635) | (0.0631) |
| Follower (in 1,000s) | | | 0.191*** | 0.190*** |
| | | | (0.0546) | (0.0547) |
| Female=1 × Follower (in 1,000s) | | | -0.0239 | -0.0239 |
| | | | (0.0561) | (0.0561) |
| Constant | 2.612*** | 1.028*** | 2.099*** | 0.708** |
| | (0.367) | (0.245) | (0.304) | (0.229) |
| Mean (Dep. Var) | 1.649 | 1.649 | 1.649 | 1.649 |
| St. Dv. (Dep. Var.) | 94.298 | 94.298 | 94.298 | 94.298 |
| Time fixed effects | No | Yes | No | Yes |
| N | 482915 | 482915 | 482915 | 482915 |

Standard errors in parentheses

Source: Twitter data. Standard errors are robust.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

## E.3 Social Leaders by Ethnicity

Table E.4: Summary statistics of Twitter tweets (with race)

| VARIABLES | (1) mean | (2) sd | (3) min | (4) max | (5) p25 | (6) p75 |
|---|---|---|---|---|---|---|
| Verified account | 0.06 | 0.23 | 0 | 1 | 0 | 0 |
| Has attachment | 0.10 | 0.30 | 0 | 1 | 0 | 0 |
| Follower (in 1,000s) | 9.10 | 166.75 | 0 | 17,103.47 | 0.19 | 2.16 |
| Following (in 1,000s) | 2.35 | 6.78 | 0 | 278.81 | 0.26 | 2.27 |
| No. of likes | 2.25 | 40.38 | 0 | 6,236 | 0 | 0 |
| No. of quotes | 0.06 | 1.32 | 0 | 183 | 0 | 0 |
| Has user location | 0.77 | 0.42 | 0 | 1 | 1 | 1 |
| Has geo location | 0.02 | 0.15 | 0 | 1 | 0 | 0 |
| original | 0.44 | 0.50 | 0 | 1 | 0 | 1 |
| No. of replies | 0.26 | 2.31 | 0 | 196 | 0 | 0 |
| No. of retweets | 410.40 | 2,062.20 | 0 | 17,160 | 0 | 18 |
| No. of tweets | 45,257.15 | 108,383.75 | 1 | 2,497,642 | 3,812 | 48,149 |
| Has mentions | 0.62 | 0.49 | 0 | 1 | 0 | 1 |
| Happy Score | 0.11 | 0.26 | 0 | 1 | 0 | 0 |
| Angry Score | 0.05 | 0.18 | 0 | 1 | 0 | 0 |
| Surprise Score | 0.16 | 0.30 | 0 | 1 | 0 | 0.33 |
| Sad Score | 0.18 | 0.32 | 0 | 1 | 0 | 0.33 |
| Fear Score | 0.26 | 0.37 | 0 | 1 | 0 | 0.50 |
| Negativity score | 0.13 | 0.17 | 0 | 0.90 | 0 | 0.25 |
| Neutrality score | 0.72 | 0.21 | 0.07 | 1 | 0.56 | 1 |
| Positivity score | 0.15 | 0.18 | 0 | 0.94 | 0 | 0.26 |
| Compound Sentiment | 0 | 0.46 | -0.99 | 0.99 | -0.34 | 0.36 |
| White | 0.72 | 0.45 | 0 | 1 | 0 | 1 |

Notes: The table presents summary statistics of the data set generated via the Twitter API. I apply a hashtag-based approach, relying on 10 hashtags which identify GBV-related movements on Twitter, to filter on the full universe of Twitter tweets. To extract the compound sentiment scores as well as the share of negative, neutral and positive tweets I employ the VADER Sentiment Analysis tool by Hutto and Gilbert (2014). To analyze emotions of written text I employ the emotion recognition tool developed by Colnerič and Demšar (2018). N=56,324. I apply the Face Recognition tool by Serengil and Ozpinar (2020) to a 20 percent sample of users' profile pictures to extract information on people's race. Source: Twitter (2014-2017).

Table E.5: Regression for No. of likes

| | OLS | OLS | POISSON | LPM | PROBIT |
|---|---|---|---|---|---|
| main | | | | | |
| Follower (in 1,000s) | 0.0169* | 0.0168* | 0.00113*** | 0.0000474** | 0.000131** |
| | (0.00956) | (0.00951) | (0.000142) | (0.0000226) | (0.0000517) |
| Not white=1 | -1.838*** | -1.538*** | 3.47434e+09 | -0.0846*** | -0.318*** |
| | (0.255) | (0.235) | (1.63824e+10) | (0.00362) | (0.0147) |
| Not white=1 × Follower (in 1,000s) | 0.00260 | 0.00344 | | 0.000517*** | 0.00276*** |
| | (0.0127) | (0.0126) | | (0.0000670) | (0.000490) |
| Following (in 1,000s) | -0.00822 | -0.0127 | 0.00145 | 0.00411*** | 0.0112*** |
| | (0.0129) | (0.0136) | (0.00984) | (0.000334) | (0.000958) |
| Has attachment=1 | 0.385 | 0.694 | 0.288 | 0.0445*** | 0.142*** |
| | (0.542) | (0.539) | (0.561) | (0.00582) | (0.0200) |
| Has mentions=1 | -2.879*** | -2.942*** | -2.951*** | -0.232*** | -0.775*** |
| | (0.365) | (0.366) | (0.400) | (0.00377) | (0.0123) |
| Has geo location=1 | 4.326*** | 4.296*** | 2.872*** | 0.191*** | 0.543*** |
| | (1.433) | (1.431) | (0.964) | (0.0140) | (0.0373) |
| Has user location=1 | 1.675*** | 1.108*** | 1.741*** | 0.0587*** | 0.229*** |
| | (0.251) | (0.191) | (0.250) | (0.00384) | (0.0153) |
| Compound Sentiment | 0.00883 | -0.251 | 0.0186 | 0.000930 | 0.00579 |
| | (0.501) | (0.514) | (0.486) | (0.00387) | (0.0141) |
| Happy Score | -0.180 | -0.0611 | -0.255 | 0.00669 | 0.0228 |
| | (0.346) | (0.348) | (0.460) | (0.00710) | (0.0262) |
| Angry Score | 0.660 | 0.538 | 0.799 | 0.0220** | 0.0819** |
| | (0.664) | (0.663) | (0.675) | (0.00997) | (0.0352) |
| Surprise Score | 0.627 | 0.636 | 0.766 | -0.00618 | -0.0172 |
| | (0.536) | (0.534) | (0.567) | (0.00603) | (0.0227) |
| Sad Score | 0.580* | 0.239 | 0.716* | 0.0197*** | 0.0731*** |
| | (0.350) | (0.358) | (0.392) | (0.00585) | (0.0212) |
| Fear Score | 1.633** | 1.278* | 1.594** | 0.0119** | 0.0416** |
| | (0.698) | (0.683) | (0.627) | (0.00510) | (0.0188) |
| Constant | 2.331*** | 0.307 | | 0.322*** | -0.513*** |
| | (0.340) | (0.371) | | (0.00546) | (0.0192) |
| Mean (Dep. Var) | 2.250 | 2.250 | 2.250 | 2.250 | 0.227 |
| St. Dv. (Dep. Var.) | 40.381 | 40.381 | 40.381 | 40.381 | 0.419 |
| Time fixed effects | No | Yes | No | No | No |
| N | 56324 | 56324 | 56324 | 56324 | 56324 |

Standard errors in parentheses

Source: Twitter data. Standard errors are robust.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table E.6: Instrumental variable regressions

| | OLS | OLS | IV | IV |
|---|---|---|---|---|
| Verified account=1 | 23.44*** | 22.78*** | | |
| | (3.268) | (3.147) | | |
| Not white=1 | -0.554*** | -0.369* | -0.122 | 0.0932 |
| | (0.139) | (0.146) | (0.228) | (0.239) |
| Verified account=1 × Not white=1 | -16.44*** | -15.93*** | | |
| | (3.606) | (3.501) | | |
| Has attachment=1 | 0.709 | 0.899 | 0.292 | 0.548 |
| | (0.527) | (0.528) | (0.923) | (0.898) |
| Has mentions=1 | -2.871*** | -2.904*** | -3.284*** | -3.206*** |
| | (0.356) | (0.359) | (0.392) | (0.384) |
| Has geo location=1 | 4.567** | 4.488** | 5.185*** | 4.972*** |
| | (1.396) | (1.395) | (1.413) | (1.410) |
| Has user location=1 | 0.709*** | 0.307* | 0.299 | -0.437 |
| | (0.168) | (0.145) | (0.199) | (0.264) |
| Compound Sentiment | -0.0187 | -0.213 | -0.736 | -0.952 |
| | (0.502) | (0.510) | (0.697) | (0.700) |
| Happy Score | -0.282 | -0.183 | 0.828 | 0.904 |
| | (0.344) | (0.345) | (0.571) | (0.569) |
| Angry Score | 0.538 | 0.484 | 0.884 | 0.704 |
| | (0.662) | (0.663) | (0.739) | (0.726) |
| Surprise Score | 0.504 | 0.579 | 0.0666 | 0.176 |
| | (0.526) | (0.525) | (0.909) | (0.899) |
| Sad Score | 0.325 | 0.106 | 0.985* | 0.623 |
| | (0.355) | (0.359) | (0.460) | (0.446) |
| Fear Score | 1.457* | 1.215 | 1.215 | 0.888 |
| | (0.680) | (0.676) | (0.831) | (0.815) |
| Following (in 1,000s) | -0.0519*** | -0.0545*** | -0.245*** | -0.247*** |
| | (0.0133) | (0.0140) | (0.0509) | (0.0502) |
| Follower (in 1,000s) | | | 0.226*** | 0.218*** |
| | | | (0.0430) | (0.0419) |
| Not white=1 × Follower (in 1,000s) | | | -0.0999* | -0.0960* |
| | | | (0.0482) | (0.0471) |
| Constant | 1.894*** | 0.106 | 1.970*** | 0.394 |
| | (0.363) | (0.394) | (0.520) | (0.496) |
| Mean (Dep. Var) | 2.250 | 2.250 | 2.250 | 2.250 |
| St. Dv. (Dep. Var.) | 40.381 | 40.381 | 40.381 | 40.381 |
| Time fixed effects | No | Yes | No | Yes |
| N | 56324 | 56324 | 56324 | 56324 |

Standard errors in parentheses

Source: Twitter data. Standard errors are robust.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

## E.4 Restricting the Sample to Original Tweets

Table E.7: Regression for No. of likes

| | OLS | OLS | POISSON | LPM | PROBIT |
|---|---|---|---|---|---|
| main | | | | | |
| Follower (in 1,000s) | 0.0220*** | 0.0220*** | 0.000566 | 0.0000507*** | 0.00174*** |
| | (0.00743) | (0.00744) | (.) | (0.00000912) | (0.000113) |
| Following (in 1,000s) | 0.0591** | 0.0545** | 0.00873 | 0.00201*** | 0.00522*** |
| | (0.0262) | (0.0258) | (.) | (0.000210) | (0.000481) |
| Has attachment=1 | 2.739*** | 2.283*** | 2.898 | 0.133*** | 0.341*** |
| | (0.525) | (0.541) | (.) | (0.00245) | (0.00627) |
| Has mentions=1 | -1.543*** | -1.733*** | -0.918 | -0.0437*** | -0.126*** |
| | (0.307) | (0.304) | (.) | (0.00162) | (0.00456) |
| Has geo location=1 | 0.587 | 0.218 | 0.477 | 0.121*** | 0.322*** |
| | (0.517) | (0.554) | (.) | (0.00347) | (0.00882) |
| Has user location=1 | 0.754** | 0.492 | 140.1 | -0.0140*** | -0.0439*** |
| | (0.359) | (0.338) | (.) | (0.00170) | (0.00460) |
| Compound Sentiment | -0.823 | -0.911* | -0.683 | -0.0181*** | -0.0504*** |
| | (0.537) | (0.541) | (.) | (0.00164) | (0.00446) |
| Happy Score | 0.763 | 0.807 | 52.68 | 0.0240*** | 0.0665*** |
| | (0.545) | (0.546) | (.) | (0.00303) | (0.00821) |
| Angry Score | -0.752** | -0.687** | 51.38 | 0.0217*** | 0.0602*** |
| | (0.326) | (0.315) | (.) | (0.00414) | (0.0111) |
| Surprise Score | 1.075 | 1.343 | 52.86 | -0.0384*** | -0.109*** |
| | (1.576) | (1.659) | (.) | (0.00250) | (0.00701) |
| Sad Score | 0.214 | -0.0798 | 52.14 | 0.00674*** | 0.0187*** |
| | (0.395) | (0.417) | (.) | (0.00245) | (0.00666) |
| Fear Score | 0.300 | -0.139 | 52.21 | 0.00460** | 0.0122** |
| | (0.374) | (0.399) | (.) | (0.00215) | (0.00587) |
| Constant | 2.011*** | 0.0803 | | 0.349*** | -0.394*** |
| | (0.360) | (0.432) | | (0.00200) | (0.00542) |
| Mean (Dep. Var) | 3.185 | 3.185 | 3.185 | 0.350 | 3.185 |
| St. Dv. (Dep. Var.) | 138.117 | 138.117 | 138.117 | 0.477 | 138.117 |
| Time fixed effects | No | Yes | No | No | No |
| N | 452221 | 452221 | 452221 | 452221 | 452221 |

Standard errors in parentheses
Source: Twitter data. Standard errors are robust.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table E.8: Instrumental variable regressions

| | OLS | OLS | IV | IV |
|---|---|---|---|---|
| Verified account=1 | 34.53*** | 34.40*** | | |
| | (5.541) | (5.534) | | |
| Has attachment=1 | 2.174*** | 1.776** | -0.229 | -0.764 |
| | (0.536) | (0.547) | (0.917) | (0.945) |
| Has mentions=1 | -2.134*** | -2.304*** | -1.634*** | -1.821*** |
| | (0.392) | (0.385) | (0.338) | (0.338) |
| Has geo location=1 | 0.603 | 0.235 | 1.573** | 1.021* |
| | (0.521) | (0.559) | (0.484) | (0.506) |
| Has user location=1 | -0.0877 | -0.323 | 1.099* | 0.933* |
| | (0.280) | (0.271) | (0.451) | (0.442) |
| Compound Sentiment | -0.826 | -0.901 | -0.863 | -0.914 |
| | (0.536) | (0.539) | (0.560) | (0.560) |
| Happy Score | 0.733 | 0.771 | 1.018 | 0.954 |
| | (0.543) | (0.545) | (0.568) | (0.564) |
| Angry Score | -0.761* | -0.692* | -0.366 | -0.353 |
| | (0.331) | (0.319) | (0.340) | (0.335) |
| Surprise Score | 1.066 | 1.336 | 0.720 | 1.148 |
| | (1.582) | (1.662) | (1.536) | (1.636) |
| Sad Score | 0.159 | -0.115 | 0.160 | -0.0719 |
| | (0.399) | (0.420) | (0.449) | (0.459) |
| Fear Score | 0.252 | -0.156 | -0.0721 | -0.423 |
| | (0.375) | (0.399) | (0.437) | (0.458) |
| Following (in 1,000s) | 0.0190 | 0.0152 | -0.179** | -0.183** |
| | (0.0253) | (0.0249) | (0.0564) | (0.0562) |
| Follower (in 1,000s) | | | 0.162*** | 0.161*** |
| | | | (0.0271) | (0.0271) |
| Constant | 1.923*** | 0.238 | 1.027* | -0.291 |
| | (0.350) | (0.410) | (0.439) | (0.486) |
| Mean (Dep. Var) | 3.185 | 3.185 | 3.185 | 3.185 |
| St. Dv. (Dep. Var.) | 138.117 | 138.117 | 138.117 | 138.117 |
| Time fixed effects | No | Yes | No | Yes |
| N | 452221 | 452221 | 452221 | 452221 |

Standard errors in parentheses

Source: Twitter data. Standard errors are robust.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

## E.5 Quotes and Replies

In this section, I estimate OLS and IV regressions but for alternative outcome variables. I first use the number of quotes as an outcome variable. I then proxy the spread of Twitter tweets by the replies a tweet receives. I first present results for ordinary least square regressions and then instrumental variable regressions.

Table E.9: Regression for No. of quotes

| | OLS | OLS | POISSON | LPM | PROBIT |
|---|---|---|---|---|---|
| main | | | | | |
| Follower (in 1,000s) | 0.000278*** | 0.000279*** | 0.00000358*** | 0.0000181*** | 0.0000967*** |
| | (0.0000647) | (0.0000648) | (0.000000359) | (0.00000247) | (0.0000146) |
| Following (in 1,000s) | 0.00321 | 0.00301 | 0.000190*** | 0.000372*** | 0.00343*** |
| | (0.00203) | (0.00201) | (0.0000373) | (0.0000472) | (0.000443) |
| Has attachment=1 | 0.0665*** | 0.0656*** | 0.0695*** | 0.0210*** | 0.473*** |
| | (0.00785) | (0.00811) | (0.00778) | (0.000549) | (0.00882) |
| Has mentions=1 | -0.0575*** | -0.0623*** | -0.0570*** | -0.0136*** | -0.425*** |
| | (0.00984) | (0.0101) | (0.00954) | (0.000247) | (0.00711) |
| Has geo location=1 | 0.0402** | 0.0317* | 0.0278** | 0.0157*** | 0.305*** |
| | (0.0160) | (0.0162) | (0.0112) | (0.00107) | (0.0158) |
| Has user location=1 | -0.00590 | -0.00913 | -0.00203 | 0.00426*** | 0.166*** |
| | (0.0144) | (0.0148) | (0.0135) | (0.000220) | (0.00888) |
| Compound Sentiment | 0.00660 | 0.00300 | 0.00720 | 0.000657*** | 0.0244*** |
| | (0.00731) | (0.00725) | (0.00751) | (0.000248) | (0.00816) |
| Happy Score | -0.0136 | -0.0103 | -0.0137 | -0.0000613 | 0.000641 |
| | (0.0123) | (0.0119) | (0.0124) | (0.000440) | (0.0148) |
| Angry Score | -0.0203* | -0.0171* | -0.0240** | -0.000617 | -0.0213 |
| | (0.0106) | (0.0100) | (0.0122) | (0.000596) | (0.0208) |
| Surprise Score | -0.0125 | -0.00431 | -0.0123 | -0.000894** | -0.0271** |
| | (0.0118) | (0.0112) | (0.0123) | (0.000365) | (0.0128) |
| Sad Score | 0.0120 | 0.00932 | 0.00951 | 0.00168*** | 0.0545*** |
| | (0.0211) | (0.0214) | (0.0168) | (0.000368) | (0.0119) |
| Fear Score | -0.00989 | -0.0155 | -0.0102 | 0.00162*** | 0.0515*** |
| | (0.0101) | (0.0106) | (0.00982) | (0.000324) | (0.0105) |
| Constant | 0.0667*** | 0.0247 | | 0.0136*** | -2.281*** |
| | (0.0221) | (0.0174) | | (0.000316) | (0.0108) |
| Mean (Dep. Var) | 0.037 | 0.037 | 0.037 | 0.012 | 0.037 |
| St. Dv. (Dep. Var.) | 3.642 | 3.642 | 3.642 | 0.110 | 3.642 |
| Time fixed effects | No | Yes | No | No | No |
| N | 1085336 | 1085336 | 1085336 | 1085336 | 1085336 |

Standard errors in parentheses
Source: Twitter data. Standard errors are robust.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table E.10: Regression for No. of quotes

| | OLS | OLS | POISSON | LPM | PROBIT |
|---|---|---|---|---|---|
| main | | | | | |
| Follower (in 1,000s) | 0.000452*** | 0.000453*** | 0.00000692*** | 0.000176*** | 0.0000289*** |
| | (0.000112) | (0.000112) | (0.000000757) | (0.0000328) | (0.00000394) |
| Following (in 1,000s) | 0.00642 | 0.00616 | 0.000405*** | 0.00368*** | 0.000586*** |
| | (0.00452) | (0.00450) | (0.0000865) | (0.000661) | (0.000102) |
| Has attachment=1 | 0.143*** | 0.112*** | 0.152*** | 0.572*** | 0.0466*** |
| | (0.0187) | (0.0223) | (0.0162) | (0.0102) | (0.00118) |
| Has mentions=1 | -0.0140 | -0.0252** | -0.0112 | 0.103*** | 0.00590*** |
| | (0.0111) | (0.0120) | (0.0104) | (0.00928) | (0.000562) |
| Has geo location=1 | 0.0351* | 0.0166 | 0.0324 | 0.166*** | 0.0105*** |
| | (0.0199) | (0.0205) | (0.0201) | (0.0174) | (0.00124) |
| Has user location=1 | -0.0234 | -0.0329 | -0.0168 | 0.140*** | 0.00646*** |
| | (0.0357) | (0.0371) | (0.0337) | (0.0108) | (0.000485) |
| Compound Sentiment | 0.00797 | 0.00373 | 0.00807 | 0.00597 | 0.000274 |
| | (0.0157) | (0.0153) | (0.0153) | (0.00980) | (0.000528) |
| Happy Score | -0.0268 | -0.0253 | -0.0280 | -0.000619 | -0.0000647 |
| | (0.0265) | (0.0263) | (0.0279) | (0.0180) | (0.000924) |
| Angry Score | -0.0391 | -0.0396* | -0.0478* | -0.0214 | -0.000990 |
| | (0.0241) | (0.0236) | (0.0287) | (0.0249) | (0.00124) |
| Surprise Score | -0.0218 | -0.00919 | -0.0215 | -0.0399*** | -0.00206*** |
| | (0.0270) | (0.0260) | (0.0283) | (0.0155) | (0.000766) |
| Sad Score | 0.0313 | 0.0213 | 0.0250 | 0.0473*** | 0.00243*** |
| | (0.0509) | (0.0512) | (0.0391) | (0.0144) | (0.000787) |
| Fear Score | -0.0192 | -0.0349 | -0.0190 | 0.0626*** | 0.00323*** |
| | (0.0227) | (0.0248) | (0.0225) | (0.0126) | (0.000690) |
| Constant | 0.0731* | 0.0155 | | -2.265*** | 0.00947*** |
| | (0.0437) | (0.0337) | | (0.0127) | (0.000571) |
| Mean (Dep. Var) | 0.075 | 0.075 | 0.075 | 0.075 | 0.023 |
| St. Dv. (Dep. Var.) | 5.497 | 5.497 | 5.497 | 5.497 | 0.151 |
| Time fixed effects | No | Yes | No | No | No |
| N | 452221 | 452221 | 452221 | 452221 | 452221 |

Standard errors in parentheses

Source: Twitter data. Standard errors are robust.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table E.11: Instrumental variable regressions for No. of quotes

| | OLS | OLS | IV | IV |
|---|---|---|---|---|
| Verified account=1 | 0.541*** | 0.538*** | | |
| | (0.111) | (0.110) | | |
| Has attachment=1 | 0.0626*** | 0.0619*** | 0.0416*** | 0.0403** |
| | (0.00837) | (0.00860) | (0.0120) | (0.0123) |
| Has mentions=1 | -0.0565*** | -0.0610*** | -0.0529*** | -0.0577*** |
| | (0.00959) | (0.00986) | (0.00897) | (0.00930) |
| Has geo location=1 | 0.0375* | 0.0293 | 0.0538*** | 0.0436** |
| | (0.0161) | (0.0163) | (0.0153) | (0.0154) |
| Has user location=1 | -0.0169 | -0.0199 | -0.00570 | -0.00898 |
| | (0.0164) | (0.0167) | (0.0145) | (0.0149) |
| Compound Sentiment | 0.00616 | 0.00275 | 0.00656 | 0.00284 |
| | (0.00726) | (0.00722) | (0.00747) | (0.00740) |
| Happy Score | -0.0142 | -0.0109 | -0.0123 | -0.00874 |
| | (0.0123) | (0.0120) | (0.0123) | (0.0119) |
| Angry Score | -0.0211* | -0.0179 | -0.0139 | -0.0101 |
| | (0.0106) | (0.0101) | (0.00965) | (0.00902) |
| Surprise Score | -0.0122 | -0.00414 | -0.0154 | -0.00524 |
| | (0.0118) | (0.0112) | (0.0124) | (0.0115) |
| Sad Score | 0.0116 | 0.00908 | 0.0131 | 0.0113 |
| | (0.0211) | (0.0214) | (0.0213) | (0.0216) |
| Fear Score | -0.0104 | -0.0158 | -0.0121 | -0.0170 |
| | (0.0102) | (0.0107) | (0.0107) | (0.0111) |
| Following (in 1,000s) | 0.00249 | 0.00231 | -0.000940 | -0.00112 |
| | (0.00193) | (0.00191) | (0.00166) | (0.00165) |
| Follower (in 1,000s) | | | 0.00280*** | 0.00278*** |
| | | | (0.000582) | (0.000580) |
| Constant | 0.0622** | 0.0224 | 0.0509** | 0.0129 |
| | (0.0211) | (0.0168) | (0.0191) | (0.0154) |
| Mean (Dep. Var) | 1.645 | 1.645 | 1.645 | 1.645 |
| St. Dv. (Dep. Var.) | 95.093 | 95.093 | 95.093 | 95.093 |
| Time fixed effects | No | Yes | No | Yes |
| N | 1085336 | 1085336 | 1085336 | 1085336 |

Standard errors in parentheses

Source: Twitter data. Standard errors are robust.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table E.12: Instrumental variable regressions for No. of quotes

| | OLS | OLS | IV | IV |
|---|---|---|---|---|
| Verified account=1 | 0.923*** | 0.918*** | | |
| | (0.218) | (0.217) | | |
| Has attachment=1 | 0.125*** | 0.0952*** | 0.0606 | 0.0275 |
| | (0.0219) | (0.0256) | (0.0360) | (0.0406) |
| Has mentions=1 | -0.0299* | -0.0405** | -0.0165 | -0.0276* |
| | (0.0146) | (0.0153) | (0.0121) | (0.0130) |
| Has geo location=1 | 0.0365 | 0.0178 | 0.0624** | 0.0387* |
| | (0.0197) | (0.0203) | (0.0191) | (0.0192) |
| Has user location=1 | -0.0456 | -0.0543 | -0.0138 | -0.0207 |
| | (0.0404) | (0.0416) | (0.0340) | (0.0349) |
| Compound Sentiment | 0.00786 | 0.00399 | 0.00687 | 0.00365 |
| | (0.0157) | (0.0153) | (0.0161) | (0.0158) |
| Happy Score | -0.0274 | -0.0261 | -0.0198 | -0.0213 |
| | (0.0265) | (0.0264) | (0.0257) | (0.0260) |
| Angry Score | -0.0390 | -0.0394 | -0.0285 | -0.0303 |
| | (0.0239) | (0.0234) | (0.0223) | (0.0220) |
| Surprise Score | -0.0223 | -0.00957 | -0.0316 | -0.0146 |
| | (0.0272) | (0.0261) | (0.0287) | (0.0270) |
| Sad Score | 0.0298 | 0.0204 | 0.0298 | 0.0215 |
| | (0.0508) | (0.0511) | (0.0511) | (0.0515) |
| Fear Score | -0.0208 | -0.0356 | -0.0295 | -0.0427 |
| | (0.0231) | (0.0250) | (0.0252) | (0.0268) |
| Following (in 1,000s) | 0.00512 | 0.00488 | -0.000173 | -0.000401 |
| | (0.00428) | (0.00426) | (0.00378) | (0.00376) |
| Follower (in 1,000s) | | | 0.00432*** | 0.00430*** |
| | | | (0.00104) | (0.00104) |
| Constant | 0.0698 | 0.0193 | 0.0458 | 0.00523 |
| | (0.0425) | (0.0342) | (0.0380) | (0.0318) |
| Mean (Dep. Var) | 3.185 | 3.185 | 3.185 | 3.185 |
| St. Dv. (Dep. Var.) | 138.117 | 138.117 | 138.117 | 138.117 |
| Time fixed effects | No | Yes | No | Yes |
| N | 452221 | 452221 | 452221 | 452221 |

Standard errors in parentheses

Source: Twitter data. Standard errors are robust.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table E.13: Regression for No. of replies

| | OLS | OLS | POISSON | LPM | PROBIT |
|---|---|---|---|---|---|
| main | | | | | |
| Follower (in 1,000s) | 0.00135*** | 0.00135*** | 0.0000190*** | 0.0000369*** | 0.000141*** |
| | (0.000235) | (0.000235) | (0.000000882) | (0.00000515) | (0.0000257) |
| Following (in 1,000s) | 0.00436** | 0.00405** | 0.000704*** | 0.000515*** | 0.00227*** |
| | (0.00202) | (0.00200) | (0.0000790) | (0.0000700) | (0.000309) |
| Has attachment=1 | 0.142*** | 0.137*** | 0.147*** | 0.0210*** | 0.128*** |
| | (0.0126) | (0.0127) | (0.0128) | (0.000976) | (0.00568) |
| Has mentions=1 | -0.206*** | -0.209*** | -0.206*** | -0.0713*** | -0.445*** |
| | (0.00955) | (0.00990) | (0.00929) | (0.000605) | (0.00354) |
| Has geo location=1 | 0.128*** | 0.114*** | 0.0903*** | 0.0763*** | 0.363*** |
| | (0.0246) | (0.0248) | (0.0184) | (0.00238) | (0.00930) |
| Has user location=1 | 0.0199 | 0.0176 | 0.0305*** | 0.0117*** | 0.0795*** |
| | (0.0124) | (0.0126) | (0.0114) | (0.000596) | (0.00419) |
| Compound Sentiment | 0.000691 | -0.00555 | 0.00230 | 0.000110 | 0.00196 |
| | (0.00894) | (0.00907) | (0.00904) | (0.000613) | (0.00404) |
| Happy Score | -0.000564 | 0.0000420 | -0.00124 | 0.00508*** | 0.0331*** |
| | (0.0102) | (0.0101) | (0.0100) | (0.00117) | (0.00744) |
| Angry Score | -0.0198** | -0.0186** | -0.0225** | 0.00383** | 0.0264*** |
| | (0.00831) | (0.00831) | (0.00894) | (0.00160) | (0.0102) |
| Surprise Score | -0.0177 | -0.00572 | -0.0172 | -0.00663*** | -0.0419*** |
| | (0.0141) | (0.0146) | (0.0148) | (0.000954) | (0.00643) |
| Sad Score | 0.0169 | 0.0132 | 0.0138 | -0.00130 | -0.00882 |
| | (0.0185) | (0.0189) | (0.0176) | (0.000916) | (0.00605) |
| Fear Score | -0.00477 | -0.00930 | -0.00645 | -0.00145* | -0.0103* |
| | (0.00912) | (0.00906) | (0.00923) | (0.000813) | (0.00535) |
| Constant | 0.254*** | 0.243*** | | 0.117*** | -1.213*** |
| | (0.0142) | (0.0137) | | (0.000840) | (0.00521) |
| Mean (Dep. Var) | 0.174 | 0.174 | 0.174 | 0.085 | 0.174 |
| St. Dv. (Dep. Var.) | 3.627 | 3.627 | 3.627 | 0.279 | 3.627 |
| Time fixed effects | No | Yes | No | No | No |
| N | 1085336 | 1085336 | 1085336 | 1085336 | 1085336 |

Standard errors in parentheses

Source: Twitter data. Standard errors are robust.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table E.14: Regression for No. of replies

| | OLS | OLS | POISSON | LPM |
|---|---|---|---|---|
| Follower (in 1,000s) | 0.00229*** | 0.00229*** | 0.0000332*** | 0.0000608*** |
| | (0.000358) | (0.000358) | (0.00000184) | (0.0000102) |
| Following (in 1,000s) | 0.00849* | 0.00820* | 0.00136*** | 0.000902*** |
| | (0.00444) | (0.00442) | (0.000172) | (0.000156) |
| Has attachment=1 | 0.330*** | 0.297*** | 0.355*** | 0.0814*** |
| | (0.0282) | (0.0302) | (0.0277) | (0.00197) |
| Has mentions=1 | -0.0475*** | -0.0575*** | -0.0383*** | 0.00522*** |
| | (0.0134) | (0.0143) | (0.0130) | (0.00119) |
| Has geo location=1 | 0.0557* | 0.0345 | 0.0409 | 0.0345*** |
| | (0.0287) | (0.0291) | (0.0287) | (0.00259) |
| Has user location=1 | 0.0194 | 0.0139 | 0.0377 | 0.0116*** |
| | (0.0303) | (0.0312) | (0.0283) | (0.00118) |
| Compound Sentiment | -0.00831 | -0.0126 | -0.00616 | -0.00420*** |
| | (0.0205) | (0.0205) | (0.0204) | (0.00116) |
| Happy Score | -0.00950 | -0.0131 | -0.0124 | 0.00688*** |
| | (0.0217) | (0.0217) | (0.0220) | (0.00220) |
| Angry Score | -0.0498*** | -0.0524*** | -0.0568*** | 0.00253 |
| | (0.0181) | (0.0184) | (0.0204) | (0.00299) |
| Surprise Score | -0.0371 | -0.0189 | -0.0348 | -0.0130*** |
| | (0.0324) | (0.0335) | (0.0345) | (0.00180) |
| Sad Score | 0.0283 | 0.0207 | 0.0226 | -0.00576*** |
| | (0.0442) | (0.0445) | (0.0414) | (0.00175) |
| Fear Score | -0.0178 | -0.0282 | -0.0189 | -0.00630*** |
| | (0.0209) | (0.0209) | (0.0212) | (0.00155) |
| Constant | 0.245*** | 0.211*** | | 0.116*** |
| | (0.0261) | (0.0242) | | (0.00140) |
| Mean (Dep. Var) | 0.313 | 0.313 | 0.313 | 0.134 |
| St. Dv. (Dep. Var.) | 5.462 | 5.462 | 5.462 | 0.341 |
| Time fixed effects | No | Yes | No | No |
| N | 452221 | 452221 | 452221 | 452221 |

Standard errors in parentheses

Source: Twitter data. Standard errors are robust.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table E.15: Instrumental variable regressions for No. of replies

| | OLS | OLS | IV | IV |
|---|---|---|---|---|
| Verified account=1 | 1.405*** | 1.402*** | | |
| | (0.106) | (0.105) | | |
| Has attachment=1 | 0.138*** | 0.134*** | 0.0835*** | 0.0778*** |
| | (0.0131) | (0.0132) | (0.0160) | (0.0163) |
| Has mentions=1 | -0.205*** | -0.207*** | -0.195*** | -0.198*** |
| | (0.00938) | (0.00970) | (0.00925) | (0.00966) |
| Has geo location=1 | 0.118*** | 0.105*** | 0.160*** | 0.142*** |
| | (0.0247) | (0.0249) | (0.0241) | (0.0243) |
| Has user location=1 | -0.00877 | -0.0106 | 0.0204 | 0.0179 |
| | (0.0138) | (0.0140) | (0.0131) | (0.0133) |
| Compound Sentiment | -0.000444 | -0.00614 | 0.000598 | -0.00592 |
| | (0.00893) | (0.00908) | (0.00972) | (0.00985) |
| Happy Score | -0.00229 | -0.00192 | 0.00262 | 0.00362 |
| | (0.0102) | (0.0101) | (0.0121) | (0.0119) |
| Angry Score | -0.0235** | -0.0224** | -0.00484 | -0.00211 |
| | (0.00847) | (0.00846) | (0.00863) | (0.00860) |
| Surprise Score | -0.0163 | -0.00504 | -0.0244 | -0.00790 |
| | (0.0141) | (0.0146) | (0.0150) | (0.0156) |
| Sad Score | 0.0156 | 0.0120 | 0.0197 | 0.0179 |
| | (0.0185) | (0.0189) | (0.0190) | (0.0194) |
| Fear Score | -0.00545 | -0.00956 | -0.00989 | -0.0128 |
| | (0.00915) | (0.00909) | (0.0103) | (0.0102) |
| Following (in 1,000s) | 0.00353 | 0.00326 | -0.00537** | -0.00568*** |
| | (0.00190) | (0.00189) | (0.00169) | (0.00169) |
| Follower (in 1,000s) | | | 0.00727*** | 0.00725*** |
| | | | (0.000598) | (0.000595) |
| Constant | 0.246*** | 0.240*** | 0.217*** | 0.216*** |
| | (0.0135) | (0.0134) | (0.0127) | (0.0136) |
| Mean (Dep. Var) | 1.645 | 1.645 | 1.645 | 1.645 |
| St. Dv. (Dep. Var.) | 95.093 | 95.093 | 95.093 | 95.093 |
| Time fixed effects | No | Yes | No | Yes |
| N | 1085336 | 1085336 | 1085336 | 1085336 |

Standard errors in parentheses

Source: Twitter data. Standard errors are robust.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table E.16: Instrumental variable regressions for No. of replies

| | OLS | OLS | IV | IV |
|---|---|---|---|---|
| Verified account=1 | 2.412*** | 2.411*** | | |
| | (0.208) | (0.207) | | |
| Has attachment=1 | 0.307*** | 0.278*** | 0.139** | 0.100* |
| | (0.0302) | (0.0326) | (0.0441) | (0.0474) |
| Has mentions=1 | -0.0883*** | -0.0970*** | -0.0533*** | -0.0632*** |
| | (0.0162) | (0.0171) | (0.0154) | (0.0166) |
| Has geo location=1 | 0.0515 | 0.0313 | 0.119*** | 0.0864** |
| | (0.0287) | (0.0291) | (0.0279) | (0.0282) |
| Has user location=1 | -0.0413 | -0.0456 | 0.0416 | 0.0425 |
| | (0.0337) | (0.0345) | (0.0320) | (0.0327) |
| Compound Sentiment | -0.00827 | -0.0119 | -0.0109 | -0.0128 |
| | (0.0205) | (0.0206) | (0.0224) | (0.0225) |
| Happy Score | -0.0130 | -0.0164 | 0.00692 | -0.00362 |
| | (0.0219) | (0.0219) | (0.0238) | (0.0239) |
| Angry Score | -0.0525** | -0.0546** | -0.0249 | -0.0308 |
| | (0.0184) | (0.0186) | (0.0197) | (0.0200) |
| Surprise Score | -0.0358 | -0.0184 | -0.0600 | -0.0315 |
| | (0.0326) | (0.0336) | (0.0335) | (0.0345) |
| Sad Score | 0.0248 | 0.0182 | 0.0248 | 0.0212 |
| | (0.0442) | (0.0445) | (0.0455) | (0.0458) |
| Fear Score | -0.0191 | -0.0278 | -0.0418 | -0.0465 |
| | (0.0210) | (0.0211) | (0.0244) | (0.0243) |
| Following (in 1,000s) | 0.00698 | 0.00673 | -0.00686 | -0.00715 |
| | (0.00419) | (0.00418) | (0.00387) | (0.00387) |
| Follower (in 1,000s) | | | 0.0113*** | 0.0113*** |
| | | | (0.00109) | (0.00109) |
| Constant | 0.244*** | 0.224*** | 0.181*** | 0.187*** |
| | (0.0254) | (0.0246) | (0.0240) | (0.0248) |
| Mean (Dep. Var) | 3.185 | 3.185 | 3.185 | 3.185 |
| St. Dv. (Dep. Var.) | 138.117 | 138.117 | 138.117 | 138.117 |
| Time fixed effects | No | Yes | No | Yes |
| N | 452221 | 452221 | 452221 | 452221 |

Standard errors in parentheses

Source: Twitter data. Standard errors are robust.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table E.17: Regression for No. of retweets

| | OLS | OLS | LPM |
|---|---|---|---|
| Follower (in 1,000s) | 0.0183*** | 0.0183*** | 0.0000694*** |
| | (0.00577) | (0.00577) | (0.0000125) |
| Following (in 1,000s) | 0.0286** | 0.0284** | 0.00404*** |
| | (0.0140) | (0.0139) | (0.000478) |
| Has attachment=1 | 2.588*** | 2.612*** | 0.143*** |
| | (0.326) | (0.342) | (0.00230) |
| Has mentions=1 | -0.530*** | -0.533*** | 0.0278*** |
| | (0.111) | (0.107) | (0.00143) |
| Has geo location=1 | -0.223 | -0.167 | -0.0199*** |
| | (0.205) | (0.208) | (0.00269) |
| Has user location=1 | 0.423** | 0.377** | 0.0380*** |
| | (0.170) | (0.162) | (0.00140) |
| Compound Sentiment | -0.246* | -0.249* | -0.0107*** |
| | (0.147) | (0.150) | (0.00137) |
| Happy Score | 0.195 | 0.222 | 0.00793*** |
| | (0.211) | (0.213) | (0.00254) |
| Angry Score | -0.311* | -0.300* | 0.00883** |
| | (0.168) | (0.166) | (0.00346) |
| Surprise Score | 0.0919 | 0.0413 | 0.000938 |
| | (0.358) | (0.374) | (0.00211) |
| Sad Score | 0.113 | 0.0833 | -0.000734 |
| | (0.203) | (0.208) | (0.00204) |
| Fear Score | 0.0888 | 0.0317 | -0.00108 |
| | (0.203) | (0.211) | (0.00179) |
| Constant | 0.866*** | 0.361* | 0.145*** |
| | (0.190) | (0.208) | (0.00162) |
| Mean (Dep. Var) | 1.618 | 1.618 | 0.202 |
| St. Dv. (Dep. Var.) | 44.948 | 44.948 | 0.402 |
| Time fixed effects | No | Yes | No |
| N | 452221 | 452221 | 452221 |

Estimated coefficients from Poisson and Probit not shown due to a convergence error.
Standard errors in parentheses
Source: Twitter data. Standard errors are robust.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table E.18: Instrumental variable regressions for No. of retweets

| | OLS | OLS | IV | IV |
|---|---|---|---|---|
| Verified account=1 | 16.56*** | 16.52*** | | |
| | (1.625) | (1.624) | | |
| Has attachment=1 | 2.482*** | 2.537*** | 1.329*** | 1.316*** |
| | (0.348) | (0.366) | (0.379) | (0.389) |
| Has mentions=1 | -0.809*** | -0.802*** | -0.569*** | -0.570*** |
| | (0.132) | (0.127) | (0.124) | (0.122) |
| Has geo location=1 | -0.270 | -0.203 | 0.195 | 0.174 |
| | (0.204) | (0.208) | (0.196) | (0.199) |
| Has user location=1 | -0.0000746 | -0.0389 | 0.569** | 0.565** |
| | (0.188) | (0.181) | (0.193) | (0.188) |
| Compound Sentiment | -0.245 | -0.245 | -0.263 | -0.251 |
| | (0.147) | (0.150) | (0.162) | (0.164) |
| Happy Score | 0.166 | 0.196 | 0.303 | 0.284 |
| | (0.212) | (0.214) | (0.222) | (0.223) |
| Angry Score | -0.337* | -0.321 | -0.147 | -0.158 |
| | (0.170) | (0.168) | (0.175) | (0.175) |
| Surprise Score | 0.107 | 0.0485 | -0.0584 | -0.0418 |
| | (0.360) | (0.375) | (0.356) | (0.374) |
| Sad Score | 0.0896 | 0.0660 | 0.0900 | 0.0866 |
| | (0.204) | (0.209) | (0.223) | (0.224) |
| Fear Score | 0.0864 | 0.0394 | -0.0690 | -0.0889 |
| | (0.206) | (0.214) | (0.219) | (0.226) |
| Following (in 1,000s) | 0.0225* | 0.0227* | -0.0724*** | -0.0725*** |
| | (0.0105) | (0.0103) | (0.0182) | (0.0182) |
| Follower (in 1,000s) | | | 0.0775*** | 0.0775*** |
| | | | (0.00825) | (0.00825) |
| Constant | 0.878*** | 0.457* | 0.448* | 0.203 |
| | (0.191) | (0.212) | (0.191) | (0.213) |
| Mean (Dep. Var) | 3.185 | 3.185 | 3.185 | 3.185 |
| St. Dv. (Dep. Var.) | 138.117 | 138.117 | 138.117 | 138.117 |
| Time fixed effects | No | Yes | No | Yes |
| N | 452221 | 452221 | 452221 | 452221 |

Standard errors in parentheses

Source: Twitter data. Standard errors are robust.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$