Tweeting negative emotion: An investigation of Twitter data in the aftermath of violence on college campuses.

Article in Psychological Methods · December 2016

DOI: 10.1037/met0000099

Nickolas Michael Jones
University of California, Irvine

Roxane Cohen Silver
University of California, Irvine

Tweeting Negative Emotion: An Investigation of Twitter Data in the Aftermath of Violence on College Campuses

Nickolas M. Jones
University of California, Irvine

Sean P. Wojcik
Upworthy

Josiah Sweeting
University of California, Irvine

Roxane Cohen Silver
University of California, Irvine

Studying communities impacted by traumatic events is often costly, requires swift action to enter the field when disaster strikes, and may be invasive for some traumatized respondents. Typically, individuals are studied after the traumatic event with no baseline data against which to compare their postdisaster responses. Given these challenges, we used longitudinal Twitter data across 3 case studies to examine the impact of violence near or on college campuses in the communities of Isla Vista, CA, Flagstaff, AZ, and Roseburg, OR, compared with control communities, between 2014 and 2015. To identify users likely to live in each community, we sought Twitter accounts local to those communities and downloaded tweets of their respective followers. Tweets were then coded for the presence of event-related negative emotion words using a computerized text analysis method (Linguistic Inquiry and Word Count, LIWC). In Case Study 1, we observed an increase in postevent negative emotion expression among sampled followers after mass violence, and show how patterns of response appear differently based on the timeframe under scrutiny. In Case Study 2, we replicate the pattern of results among users in the control group from Case Study 1, after a campus shooting in that community killed 1 student. In Case Study 3, we replicate this pattern in another group of Twitter users likely to live in a community affected by a mass shooting. We discuss conducting trauma-related research using Twitter data and provide guidance to researchers interested in using Twitter to answer their own research questions in this domain.

Keywords: Twitter; collective trauma; Isla Vista massacre; social media, school shootings; LIWC

Citation: Jones, N. M., Wojcik, S. P., Sweeting, J., Silver, R. C. (2016). Tweeting negative emotion: An investigation of Twitter data in the aftermath of violence on college campuses. *Psychological Methods*. *21*, 526-541. doi: http://dx.doi.org/10.1037/met0000099

Past research finds that experiences of community-wide traumas, especially of mass violence, are associated with adverse mental health outcomes (e.g., psychological distress) and physical health outcomes (Holman et al., 2008; Norris et al., 2002; Schlenger et al., 2002). However, studying the psychological impact of community-wide trauma is no small task. Researchers need to employ longitudinal methods to make inferences about an event's immediate impact and long-term psychological ramifications on affected individuals. Nonetheless, there are a number of challenges associated with this type of research. Principally, data collection usually begins after a traumatic event occurs, precluding the researcher's ability to control for respondents' pre-event functioning. Moreover, if data are not collected rapidly after disaster strikes, valuable insights about its impact can be lost (North & Pfefferbaum, 2002), as some victims tend to forget the symptoms they experienced closer to the event (North, Smith, & Spitznagel, 1997). But collecting these data is especially challenging due to difficulty in quickly securing ethics board approval and funding to send a research team into impacted communities (Silver, 2004). Even if funding and approval are obtained, community-based organizations might restrict access to potential respondents. And assuming these challenges can be surmounted, it may still be difficult to garner participation from traumatized community members who are distressed.

The sudden, unexpected nature of many traumatic events makes employing rigorous, controlled methods for studying their impact virtually impossible. Infrequently, traditional methods for studying these events after the fact can be used by researchers who have already collected data from the affected community before the traumatic event occurred Bravo, Rubio-Stipec, Canino, Woodbury, and Ribera (1990). Although this offers the methodological benefit of having pre-event data, these researchers rarely have a control group against which to compare trends in their data (see Norris et al., 2002, for a review).

Social Media Data on Twitter

Recently, some researchers have circumvented these prohibitive methodological challenges by turning to content of social media data on Twitter to ascertain the impact of unexpected or ongoing community traumas. Twitter is a widely used social media platform that allows a user to post status updates (tweets), of 140 characters or less, to his or her timeline. These posts are instantly shared across a personal network of followers (individuals who subscribe to a user's account), who can then reply to the tweets or post them to their own timelines (retweeting), thus sharing tweets across many networks. Fortunately, researchers can freely connect to Twitter, download public tweets from any users who have public accounts, and analyze the content of those tweets.

The analysis of text has principally been the domain of corpus linguists, who have been interested in understanding the meaning, usage, systematic co-occurrence of words, and the structure of language (Biber, Conrad, & Reppen, 1998). It has only been recently that psychologists, interested in the psychological underpinnings of word usage, have begun using similar analyses of text to understand what words reveal about how people think and feel. Indeed, a strong empirical tradition of psychological research finds (a) that expressive writing about traumatic events is associated with a number of psychological consequences; and (b) the content of these writings reveals psychological information about its author (for a review, see Pennebaker, Mehl, & Niederhoffer, 2003; Tausczik & Pennebaker, 2010). For example, past research indicates a positive association between experimentally induced expressive writing about traumatic events and psychological health (Frattaroli, 2006). However, in a longitudinal study of the impact of the September 11, 2001 (9/11) terrorist attacks, Seery, Silver, Holman, Ence, and Chu (2008) found that when individuals in a nationally representative sample of respondents were prompted to express their thoughts and feelings about the attacks in an online survey, those who responded to the prompt exhibited more distress in later waves of data collection than those who chose not to respond. Moreover, of the individuals who did respond, those who wrote the most words exhibited greater psychological distress and physical health problems over time.

In addition to exploring incidence and length of expressive writing after a collective trauma, researchers have also explored the content of these writings. In a study of online language use before and after the 9/11 attacks, Cohn, Mehl, and Pennebaker (2004) found that post-September 11th blog entries written by livejournal.com users exhibited increased psychological distancing (i.e., words having an abstract, impersonal, rational tone) when compared with pre-event entries and remained elevated for more than 6 weeks before returning to baseline. Furthermore, they found that individuals who wrote blog posts about the attacks suffered a large drop in positive emotion expression and exhibited a slower rate of return to baseline after the attacks.

Researchers have also sought to understand how general word usage patterns relate to traditionally measured psychological constructs. Rude, Gortner, and Pennebaker (2004) found that clinically depressed individuals expressed more negative emotion and used more first person singular words than non-depressed individuals in a series of expressive writing tasks. Furthermore, researchers have correlated word use with motivation (Pennebaker & King, 1999), personality (Pennebaker & King, 1999; Schwartz et al., 2013), community indicators of physical health (Eichstaedt et al., 2015; Pennebaker & Francis, 1996), and have demonstrated that emotion expression on social media networks can be transmitted to others in those networks via emotional contagion (Guillory et al., 2011; Hancock, Gee, Ciaccio, & Lin, 2008).

When dealing with text-based social media sources (e.g., Twitter data), however, the tremendous amount of available data precludes any single researcher's ability to read, manually code, analyze, and summarize the psychological information they contain. Thus, automatic text analysis programs, like the Linguistic Inquiry and Word Count (LIWC; Tausczik & Pennebaker, 2010) tool, have been developed to help researchers makes sense of these vast corpora and glean the psychological information contained therein. LIWC is a closed-vocabulary approach in that it comes with multiple dictionaries of words that are associated with psychological phenomena like emotion, mood, and cognitive states. The words contained in a corpus are compared with words in LIWC dictionaries and the tool counts the number of matches it finds.

This method for capturing psychological information from words has been validated multiple times across a number of psychological domains including emotion, cognition, and social interactions (Pennebaker, 2011; Tausczik & Pennebaker, 2010), and has proven useful in detecting actual emotional states. For example, Kahn, Tobin, Massey, and Anderson (2007) found that when participants were asked to write essays about an amusing and sad time in their lives, they used positive and negative emotion words, identified by LIWC, accordingly. LIWC was a valid tool with which to identify the valence of disclosure in the essays people wrote. Moreover, in a subsequent study, these researchers induced negative and positive emotions among their participants using film clips and asked participants to speak into a microphone and describe how they felt. They found that LIWC counts of negative emotion words in the audio transcripts of participants who watched a funeral scene were higher than in film conditions used to elicit other emotions. Thus, the words a person uses can provide researchers with a window into his or her emotional state, and archival social media data, like those available on Twitter, are good candidates for understanding how people feel.

Trauma Studies Using Twitter

The content of a user's tweets, which span across time and distance, represent an enormous opportunity for trauma researchers to study people and communities through their language use when disaster strikes. Although some researchers have used Twitter to study disaster communication (Murthy & Longwell, 2013; Sutton, League, Sellnow, & Sellnow, 2015; Sutton et al., 2014), few have seized this opportunity to tap into the psychological data that Twitter freely offers.

Doré, Ort, Braverman, and Ochsner (2015) examined Twitter content to identify shifts in emotion-specific word usage over time across the U.S. following the December, 2012 Sandy Hook Elementary School massacre in Newtown, CT. Although their goal was not to examine community-level effects of this traumatic event, they examined sentiment in tweets containing the keywords "Newtown" and "Sandy Hook" created immediately after the shootings, and two more times in the 6 months thereafter. They found a shift in word usage such that sadness words occurred more frequently in tweets around the Newtown area shortly after the shooting, and shifted toward anxiety words in those created at increasing temporal and spatial distance from Newtown. In another Sandy Hook-related study, Glasgow, Fink, and Boyd-Graber (2014) identified tweets with words about "death" from users living in communities proximal to Newtown and found that "death" words were used more frequently following the shootings than before the shootings, and this effect was more pronounced for communities closer to Newtown. While the primary goal of this study was to employ machine learning to improve upon the LIWC dictionary application, researchers nonetheless found evidence that a community-level effect of the shooting (i.e., use of more death words) was present in Twitter data, and that LIWC was a sufficient tool with which to identify this effect.

In a very different context, De Choudhury, Monroy-Hernandez, and Mark (2014) used Twitter data in conjunction with countrylevel homicide data from the Mexican government to examine whether residents of four major cities in Mexico displayed desensitization to protracted violence resulting from the Mexican Drug War. Over a 2-year period, they found that negative affect expressed in tweets declined despite increases in homicides.

Shared features of the studies conducted by Glasgow et al. (2014) and De Choudhury et al. (2014) that make them strong designs include the presence of pre-event Twitter data and the use of control communities to which affected communities were compared. As stated previously, traditional studies of community-wide trauma using such designs are nearly impossible.

Identifying Twitter Users in an Affected Community

In the case of studying a community-wide trauma, it is vitally important to locate users in affected communities. Although tweets can be geocoded (e.g., longitude/latitude coordinates) with a reasonable estimate of where the user was when he or she posted the tweet, very few users opt-in to this option. Moreover, tweets can originate in any location, and it is possible that tweets originating outside of a target community might actually belong to a resident who is temporarily outside the community. Such tweets might be removed from a sample because their geocoordinates are outside of the target area. Thus, it is very challenging to ensure one is downloading tweets from users within target areas and researchers have employed various strategies to identify individuals within a geographic location.

For example, Glasgow et al. (2014) used tweets with geolocation data producing 460,000 tweets across two communities over a 4-year time period. They do not indicate how many tweets they downloaded to obtain their final sample of tweets, nor do they indicate whether they relied on geocodes or the "location" field users can fill in. Nonetheless, it can be surmised that if they created a dataset with geocoded tweets, they started with many more tweets, perhaps millions more. In contrast, De Choudhury et al. (2014) identified over 3 million tweets across a 2-year time period, representing over 200,000 users, from four major Mexican cities. Their strategy was to use tweets that contained hashtags referring to each city in their analysis (e.g., #monterrey). Although there is no guarantee that a user who tweeted with that hashtag lives in the city, it is probable. In either case, the problem is clear: A researcher can either employ geocoded tweets at the expense of vast amounts of data, or keep as much data as possible, with lower certainty that the user resides in the community under study, and rely on keywords and hashtags.

The Present Case Studies

In the present research, we provide another method for locating Twitter users in a target community following three community-wide traumatic events: identifying Twitter accounts likely to be followed by local residents of the communities under study and downloading tweets from those users' accounts. In Case Study 1, we chose to test this method in a study of the Isla Vista killings of May, 2014, where a 22 year-old college student from the University of California, Santa Barbara (UCSB) killed six, injured 13 others, and took his own life after a fire-fight with law enforcement officers. We examined pre- and post-negative emotion expression alone, and in conjunction with the use of event-related words among Twitter users who followed Twitter accounts local to Santa Barbara versus a control community, Flagstaff, Arizona. In Case Study 2, we replicated the pattern of results observed in Study 1 with tweets from users in our original control sample, as a school shooting with one fatality occurred at Northern Arizona University (NAU; in Flagstaff) 17 months after the Isla Vista violence. These results represent a rare quasi-experimental design: an interrupted time series design with a control group and a reversal (cf. Shadish, Cook, & Campbell, 2002). Finally, in Case Study 3, we replicated the pattern of results observed in Case Studies 1 and 2 among Twitter users who followed accounts local to Roseburg, Oregon, where a mass shooting at Umpqua Community College on the morning of October 1, 2015 claimed the lives of 10 students and injured seven more.

Case Study 1: Isla Vista Violence

Given that prior work suggests communities experience negative outcomes following a community-wide traumatic event, especially following a school shooting (Lowe & Galea, 2015), we were interested in determining whether Twitter could serve as a viable source from which to identify negative emotion expression among followers of accounts local to Santa Barbara following the violence in Isla Vista (IV). Specifically, we sought to ascertain whether (a) there was an increase in general and event-related negative emotion expression among followers of Twitter accounts local to Santa Barbara in the wake of the Isla Vista violence compared with a control community, (b) this increase was more pronounced among campus account followers versus community account followers, and (c) whether the time frame under scrutiny is relevant to identifying negative emotion expression.

Method

Comparison community selection. Guided by the procedures used in (Wicke & Silver, 2009) to identify our comparison community, we examined city data from the U.S. Census. We generated a list of comparable communities and selected Flagstaff, Arizona because it matched Santa Barbara closely on a number of relevant indicators (e.g., population size, ethnic makeup, university population, median household income, and crime rate; see Table 1)[1].

Twitter account selection and data collection.

Community account followers. Although Twitter does allow for searching based on geographic coordinates, it is not currently possible to perform geographic searches for tweets that fall within a specific time frame if that time frame is more than 3 days prior to the date of the search. Thus, we identified tweets of Santa Barbara and Flagstaff community account followers around the time of the shooting using a novel methodology. For each community, we identified local organizations' Twitter accounts that were likely to be followed almost exclusively by community residents. Across communities, we selected similarly sized accounts (e.g., similar numbers of followers) that disseminate similar content. For example, we identified Twitter accounts from local radio stations, local city Twitter accounts, and Twitter accounts dedicated to spreading information about local events or issues. It should be noted that for Flagstaff, multiple radio accounts were selected due to the low number of followers each had compared with the radio station account selected in Santa Barbara. Table 2 provides the list of local Twitter accounts we identified in each community.

Our objective was to obtain tweets from 1,000 individuals in each community. After connecting Twitter's Application Programming Interface (API) through R statistical software, we identified the user names of the 700 most recent users who follow each local Twitter account. Although most Twitter user accounts are public, we purposefully oversampled the number of users following each account because some follower accounts are private, which prevents access to their Twitter timeline. This ensured that we would obtain at least 350 public follower accounts per local community Twitter account. For local community accounts with fewer than 700 followers, all follower usernames were recorded. As there are no established standards for sample size decisions, we sought to match sample sizes typically seen in community disaster studies (for similar methods, see Ritter, Preston, & Hernandez, 2013; Sylwester & Purver, 2015; Wojcik, Hovasapian, Graham, Motyl, & Ditto, 2015

Using the twitteR package (Gentry, 2015) for R (R Core Team, 2015), we connected to the Twitter API to download user-level account information, along with the 750 most recent tweets from each of these user accounts in August, 2014. This ensured that we would acquire tweets generated at least 6 weeks prior to the IV violence for most users. For users with fewer than 750 tweets, all tweets from their timelines were downloaded, and for users whose profiles were protected by privacy settings, no tweets were downloaded. We downloaded a total of 860,334 tweets. We retained tweets created during the 6 weeks before the violence, and those created within the 6 weeks after the violence. All other tweets were deleted. We also removed tweets from user accounts created the day of or after the violence, because these newer accounts would have no tweets during the 6 weeks prior to the violence. We were left with 151,520 tweets from 2,295 users (see Table 3).

We took measures to ensure that we only examined tweets that were written in English. Using the Compact Language Detector for R (CLDR), we identified the language of each tweet, and excluded those for which the CLDR package indicated less than 60% confidence that a given tweet was written in English. As a result, 64,579 tweets and 234 users were removed from the analysis.

Campus account followers. In a subsequent data collection effort, we used a more systematic approach to collect tweets from followers of the main university Twitter

[1] A third community was also identified to address a conceptual question that is not relevant to the issues addressed in this report and will not be discussed further.

Table 1
Comparison of community characteristics based on most current U.S. Census data.

Community	Population	White	Hispanic/ Latino	Median Household Income	County seat	University student population	Crime rate
Santa Barbara	90,412	75.1%	38.0%	$65,034	Yes	23,051	261.9
Flagstaff	68,667	73.4%	18.4%	$49,771	Yes	25,590	269.1

accounts forUC Santa Barbara (@ucsantabarbara) and for Northern Arizona State University (@nau), respectively. We used Twitter's API to download the list of all followers for each university account. We deleted user accounts from these lists based two criteria: (a) the account was created the day of or after the shooting, and (b) the account language code was something other than English (EN). This left us with a total of 15,747 user accounts in Santa Barbara, and 11,061 in Flagstaff[2]. From each of these lists, we randomly selected 2,500 user accounts. This ensured that we would obtain roughly 1,000 public follower accounts per campus.

In this second round of data collection, we used the Twitter API to download user-level information along with the 1,500 most recent tweets from each of these users. We elected to download this number of tweets for each user because we conducted our search approximately 8 months after the shooting and felt this would ensure that we would obtain tweets created at least 6 weeks prior to the shooting. For users with fewer than 1,500 tweets, all tweets were downloaded. In all, we collected 1,623,701 tweets. We then removed tweets with a timestamp outside of the 12-week window specified earlier (6 weeks before and after the violence). Data from 1,870 users, including 77,611 tweets, remained. Next, using CLDR, we identified non-English tweets and removed them. As a result, 8,787 tweets and 32 users were removed from the analysis (see Table 3). All procedures were approved by the institutional review board at the University of California, Irvine.

Table 2
Twitter accounts used to sample followers from each community

Santa Barbara	Flagstaff
@TheVibe1033	@CityofFlagstaff
@SBCity	@Flagstaff2012
@SantaBarbaraBuzz	@Flagstaff365
	@KAFFCountry
	@EagleRocks1037

Geocoded tweets. To illustrate how much data would be lost if we only used geocoded tweets, we removed all tweets without geocodes from the data set. In all, 3,994 tweets from campus account followers were geocoded, or nearly 6% of the total campus sample. Of these, only 600 tweets (or .008% of the entire campus account sample), representing 130 users (71 UCSB followers, and 59 NAU followers), were generated the week before and after the violence. Across a 2-week window around the violence, the average number of tweets per day was never greater than 40 among UCSB followers, and never greater than 25 among NAU followers. On average, these tweets represent five twitter users a day in Santa Barbara and four in Flagstaff. We then removed all tweets with geocodes outside of the greater Santa Barbara, Ventura, Los Angeles region and were left with unique tweets from 10 followers of UCSB and 16 followers of NAU. Because culling nongeocoded tweets would severely hinder our ability to make any inference, all tweets were retained in both community and campus account samples.

Measures.

Negative emotion. We analyzed the linguistic content of each tweet in the data set. Using a custom R script, we tallied the frequency with which words in each tweet matched words from the linguistic dictionaries of the Linguistic Inquiry and Word Count (LIWC) software. Each tweet was coded dichotomously such that if it contained at least one negative emotion word from the negative emotion LIWC dictionary, it was coded 1 (all others coded 0).

Event-related words and negative emotion. Based on the tweets posted the day after the violent rampage, an event-related word list of 22 words was created to capture tweets referencing the Isla Vista event (see Table 4). This word list was created by the first author who reviewed the tweets from followers of @UCSantaBarbara created the day after the violence. Words commonly used in tweets about the incident were recorded and were grouped into general (e.g., shooting) and context-specific (e.g., IslaVistaShooting) words tied to this event (for additional information about this list, see Supplemental Material). The goal was to generate a list of face-valid words related to the IV violence.

To assess the performance of this word list, two inde-

[2]Roughly 1,000 to 1,500 users were eliminated from each list as a result of this protocol.

Table 3
Number of users and tweets in the 12-Week window before and after removing non-English tweets

	All available tweets		After removing non-English tweets and accounts	
Sample	#users (%)	#tweets (%)	#users (%)	#tweets (%)
Community account followers				
Santa Barbara	1,155 (50.3)	96,943 (64.0)	947 (45.9)	38,012 (56.3)
Flagstaff	1,142 (49.7)	54,577 (36.0)	1,116 (54.1)	48,929 (43.7)
Total	2,297	151,520	2,063	86,941)
Campus account followers				
Santa Barbara	1,000 (53.2)	48,673 (62.7)	984 (53.2)	43,035 (62.5)
Flagstaff	880 (46.8)	28,938 (37.3)	864 (46.8)	25,789 (37.5)
Total	1,880	77,611	1,848	68,824

pendent coders were given a random sample of 500 tweets generated by followers of @UCSantaBarbara the day after the violence. They were then asked to code whether each tweet was about the IV violence (no / yes coded 0 and 1, respectively). These raters were blind to the date the tweets were created. We then compared ratings between both the human coders and the automated counter by calculating a kappa statistic, which represents chance-corrected agreement Landis and Koch (1977) between all raters. This analysis of inter-rater agreement yielded a kappa of .67 (96.98% average chance-uncorrected agreement). In Table 5, we provide example tweets and report the performance of the automated coder compared to human coders.

Having achieved acceptable concordance between the human coders and the automated counter, the remaining tweets were coded dichotomously by the automated counter such that if they contained at least one event-related word, and at least one negative emotion word, they were coded as 1 (all others coded 0). This allowed us to examine negative emotion expression related to the violence.

Table 4
Event-related words derived from tweets created the day of–and day after–the Isla Vista violence

General event-related words		Context specific words
.bullet	.injur	guacho.
dead	mass	islavist.
death.	massacre.	islavis.
deceased	rampage.	isla vis.
die	rip	IV.
die	shoot.	UCSB
dies	shot.	
dying	spree.	
gun.		

Analytic strategy. Analyses were conducted in Stata version, 13.1 (Stata Corp.). For each sample of tweets from followers of local and campus Twitter accounts, we calculated weekly mean proportions of tweets containing negative emotion, and of tweets having both event-related words and negative emotion. Using procedures outlined by (Mitchell, 2012), piecewise regression analyses with a notch placed at the week before and after the violence, respectively, were conducted on tweets from both follower samples 6 weeks before and after the violence (12-week window). Piecewise regression is a suitable modeling technique for examining nonlinear change in time-series data where a known point of change occurs Kim, Fay, Feuer, Midthune, et al. (2000). Thus, this technique allowed us to observe and compare changes in regression-line slopes before the violence (Segment A: Weeks -6 through -1), the week of the violence (Segment B: Weeks -1 and 1), and after the violence (Segment C: Weeks 1 through 6), both between and within followers of Twitter accounts from each community.

Many users tweeted multiple times each day throughout the 12-week time frame. If a user is more prolific than others, and expresses negative emotion across many of his or her tweets, that user could have undue influence on the results due to the correlated nature of his or her tweets. These dependencies violate the assumption of independence of residuals for an ordinary least squares regression analysis. To compensate, tweets were clustered within users in all regression analyses (Primo, Jacobsmeier, & Milyo, 2007; UCLA Statistical Consulting Group, 2015). In all contrasts, Flagstaff served as the reference group.

To obtain a more fine-grained analysis before and after the violence, we focused in on a 2-week window around the violence. For both samples, we calculated daily proportions of tweets containing negative emotion, and both event-related words and negative emotion, respectively. We then conducted similar analyses on tweets created 7 days before (Segment A) and 7 days after the event (Segment B). For these analyses we used only one notch (the day after the violence) in an approach common to regression discontinuity analyses (Thistlethwaite & Campbell, 1960). For example, this type of analysis uses pre-event daily proportions of negative emotion expression to derive an estimate of the proportion of tweets that would contain negative emotion on the day after, had the event not occurred. This estimate can then be compared to an adjusted proportion of negative emotion that was actually expressed the day after.

Results

We report our findings below for each sample of followers (community and campus accounts) in each window of time. Standardized coefficients are presented for all contrasts.

12-week window.

Community account followers. The raw proportions of negative emotion expression in each week and piecewise regression lines across the 12-week window are presented in Figure 1. Comparisons of all slopes between communities across all segments were not significantly different. Additionally, comparing slopes across segments within each community revealed no significant differences.

Weekly proportions of tweets containing both negative emotion and event-related words are depicted in Figure 2. A contrast between community slopes in Segment B revealed a significant difference (b = .10, SE = .03, p = .002) such that in the week after the violence, Santa Barbara community account followers expressed more event-related negative emotion over time relative to followers of Flagstaff community accounts. Across time within Santa Barbara, we found a significant difference in slopes between Segments A and B (b = .11, SE = .03, p < .001). Likewise, the difference in slopes between Segments B and C was significantly different from zero (b = -.15, SE = .03, p < .001), an indication that following the IV incident, event-related negative emotion expression decreased over time, relative to Flagstaff. No significant changes between slopes over time were observed among followers of accounts local to Flagstaff.

Campus account followers. Similar analyses were conducted on the followers of each campus account, the graph for which is depicted in Figure 3. Comparisons of all slopes between campus accounts across segments were not significantly different. However, comparing slopes across time

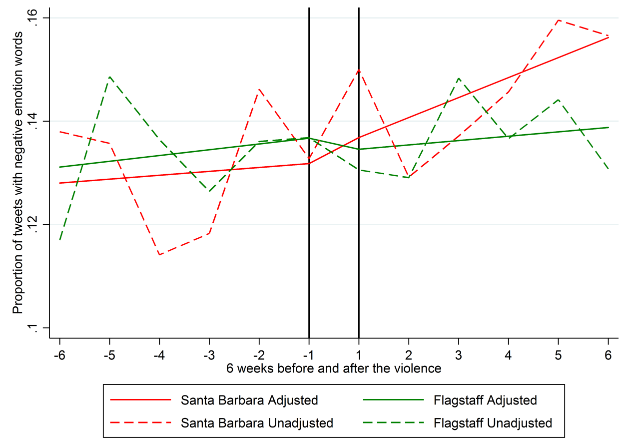

Figure 1. Weekly proportions of tweets containing negative emotion words 6 weeks before and after the Isla Vista violence among **community account followers**, with piecewise regression lines.

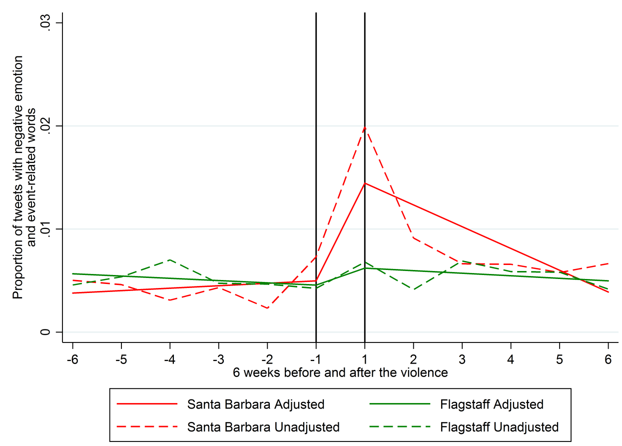

Figure 2. Weekly proportions of tweets containing both negative emotion words and event-related words 6 weeks before and after the Isla Vista violence among **community account followers**, with piecewise regression lines.

withing each campus account revealed significant differences between segments only for UCSB followers (Segments A vs. B: b = .07, SE = .03, p = .04; Segments B vs. C: b = -.10, SE =.03, p = .006).

Weekly proportions of tweets containing both negative emotion and event-related words are depicted in Figure 4. A contrast between campus account slopes in Segment B revealed a significant difference (b = .16, SE = .03, p < .001), such that the slope for UCSB followers was steeper than that of NAU followers. The contrast of slopes between campus accounts was also significant in Segment C (b = -.04, SE = .01, p < .001), reflecting a return to baseline among UCSB followers. Within UCSB, slopes were significantly different

Table 5
Example tweets and agreement between human coders and automated coder (1 = tweet about the Isla Vista violence; 0 = Not about the violence)

Tweet content	Coder 1	Coder 2	Automated Coder
Total agreement			
My prayers go out to all of the victims and people affected by the shooting at UCSB	1	1	1
UCSB housing I think its pushed back because they found 3 more bodies at the shooters apt complex Capri apts	1	1	1
Agreement with one human coder and automated coder			
Showing some Gaucho love in San Antonio I'm with all of the Isla Vista community in SPI	0	1	1
OMITTED hopefully you already know but there was a shooting tonight be careful	0	1	1
Automated coder misses			
It's unreal that this could happen in a place I call home but I've never been more proud to call it that prayers to all those affected	1	1	0
Say it father of Isla Vista victim enough condolences act	1	1	0
Automated coder errors			
Bayarea mass transit no acetrain service today all other systems incl muni Bart are on modified scheds SF Bay totaltrafficsf	0	0	1
4 ways to mobile optimize online video from shoot to playback	0	0	1

between all segments (Segments A vs. B: b = .17, SE = .03, p < .001; Segments B vs. C: b = -.22, SE = .03, p < .001), such that during the week after, event-related negative emotion expression increased, but returned to baseline in the weeks thereafter. No significant differences between slopes over time were observed for followers of NAU.

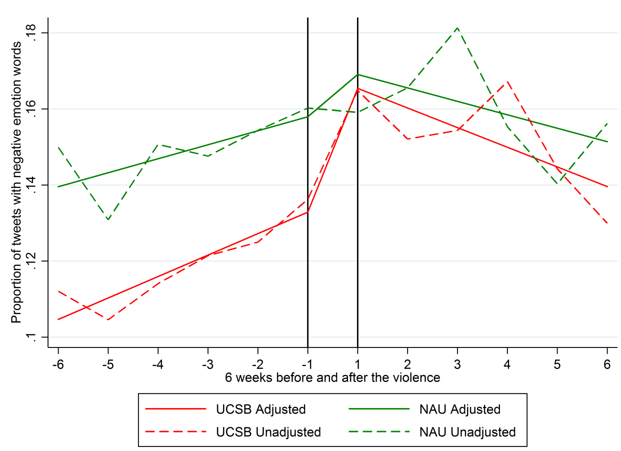

Figure 3. Weekly proportions of tweets containing negative emotion words 6 weeks before and after the Isla Vista violence among **campus account followers**, with piecewise regression lines..

Narrowing in: The 2-week window.

Community account followers. The proportions of tweets with daily negative emotion across the 2-week window are presented in Figure 5. Comparisons of all slopes between community accounts across both segments were not significantly different. Additionally, comparing slopes across segments among community account followers revealed no significant differences. However, there was a significant increase in negative emotion expression (5.7%) among followers of Santa Barbara community accounts following the violence (b = .16, SE = .06, p = .01, Cohen's d = .09). No differences were found among followers of Flagstaff community accounts.

Daily proportions of tweets containing both negative emotion and event-related words are depicted in Figure 6. A contrast between community account slopes in Segment B revealed a marginal difference (b = -.03, SE = .01, p = .06). Within each community, slopes across segments were not significantly different. However, as expected, there was a significant increase in the proportion of tweets containing event-related words and negative emotion (3.2%) among followers of Santa Barbara community accounts following the violence (b = .35, SE = .07, p < .001, Cohen's d = .19). No difference in post-event negative emotion expression about the violence was observed among followers of community accounts local to Flagstaff.

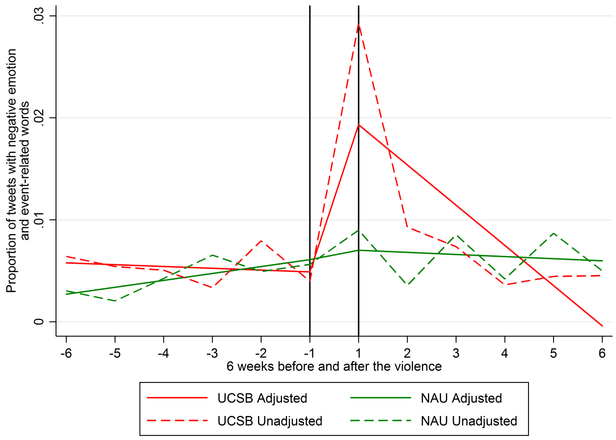

Figure 4. Weekly proportions of tweets containing both negative emotion words and event-related words 6 weeks before and after the Isla Vista violence among **campus account followers**, with piecewise regression lines.

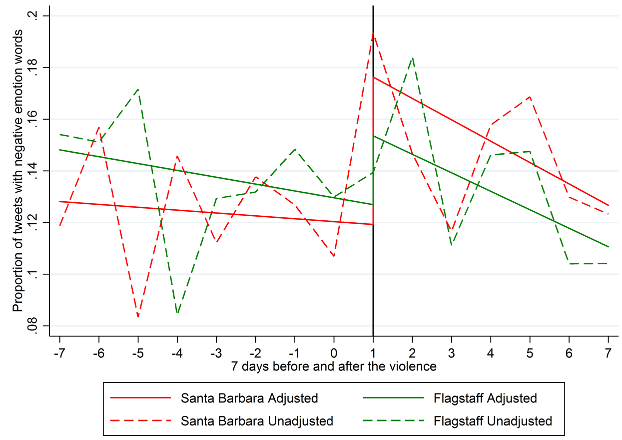

Figure 5. Daily proportions of tweets containing negative emotion words 7 days before and after the Isla Vista violence among **community account followers**, with piecewise regression lines.

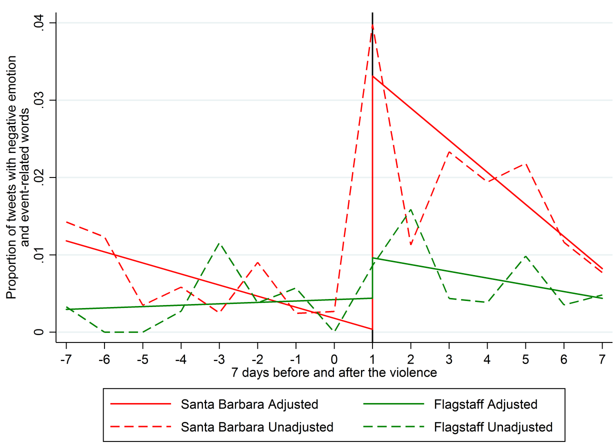

Figure 6. Daily proportions of tweets containing both negative emotion words and event-related words seven days before and after the Isla Vista violence among **community account followers**, with piecewise regression lines.

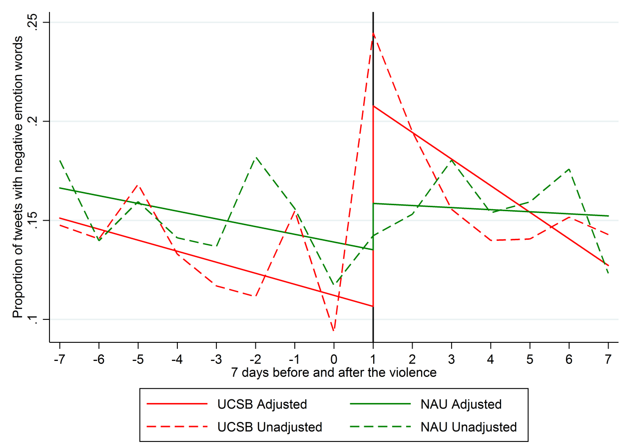

Figure 7. Daily proportions of tweets containing negative emotion words 7 days before and after the Isla Vista violence among **campus account followers**, with piecewise regression lines.

Campus account followers. The proportions of tweets with daily negative emotion across the 2-week window are presented in Figure 7. A similar pattern was observed for tweets in this window. There were no differences in slopes between andwithin community accounts across segments. However, the estimated intercepts before and after the violence were significantly different for UCSB followers, where the proportion of tweets containing negative emotion increased by 10.1% (b = .28, SE = .07, p < .001, Cohen's d = .17). No differences were found among NAU followers.

Daily proportions of tweets containing both negative emotion and event-related words are depicted in Figure 8. While no differences among NAU followers were found, significant findings emerged for followers of UCSB. We found that the difference between slopes before and after the violence was significantly different from zero (b = -.10, SE = .02, p < .001). Additionally, we found that after the violence, the proportion of tweets containing both negative emotion and event-related words increased by 6.4% (b = .59, SE = .09, p < .001, Cohen's d = .28).

Consistent versus non-consistent Twitter users. Not all users posted tweets during each regression segment. Therefore, we examined whether the pattern of results between UCSB followers who tweeted across the entire 12-week timeframe (n = 324) and those who did not (n = 296) varied when excluding each group from the analysis of tweets in the 2-week window. Separate piecewise regressions conducted for each of these groups revealed similar patterns of event-related negative emotion expression. Among followers who tweeted consistently across the timeframe, we observed a 6.8% increase in event-related negative emotion word use compared to the estimated proportion had the event never occurred (b = .57, SE = .09, p < .001). Among non-consistent users, we observed a 6.3% increase in eventrelated negative emotion word use, which was statistically significant (b = .54, SE = .18, p = .01).

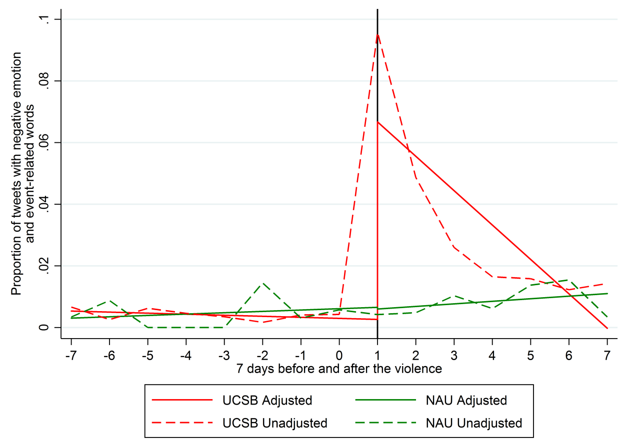

Figure 8. Daily proportions of tweets containing both negative emotion words and event-related words 7 days before and after the Isla Vista violence among **campus account followers**, with piecewise regression lines.

Discussion

In this case study, we demonstrate a novel approach to acquiring Twitter users for the purposes of studying the impact of a community-wide trauma. Rather than relying on geocodes or hashtags to locate users, we identified Twitter accounts local to an impacted community that were likely followed by residents. We used this method to acquire tweets from two samples: a sample of individuals who followed community accounts in Santa Barbara and Flagstaff (our control community) and a random sample of individuals who followed the Twitter accounts for UCSB and NAU, respectively. Overall, our results demonstrate that negative emotion expression was (a) visible at the weekly level when examining tweets with event-related words and negative emotion, (b) most apparent among campus account followers when compared to the community account followers, and (c) amplified in the 2-week time frame compared with the 12-week window. Within the 2-week window, we observed that for community and campus account followers, there was a significant increase in event-related negative emotion expression and negative emotion remained above baseline levels for many days following the violence. Further, by identifying tweets with negative emotion expression and event-related words, we were able to capture negative emotion specific to the event. The pattern of results we observed is similar to other studies of post-trauma word usage. For example, following a traumatic event at a Texas university, researchers found that school newspapers demonstrated increased negative emotion word usage in the weeks following the tragedy. The authors suggest that in an emergency phase (immediate aftermath of an event), emotion expression is heightened and after some time has passed, that expression wanes (inhibition phase; Gortner & Pennebaker, 2003). A similar pattern was observed among individuals in an online chatroom following the death of Princess Diana: negative emotion expression occurred with more frequency after her death, but tapered off in the weeks following her death because individuals were sanctioned or discussing the matter (Stone & Pennebaker, 2002).

Similar to De Choudhury et al. (2014), we made the choice to approximate a user's community of residence, rather than rely on geocoded tweets. Of course, there is no guarantee that all sampled users resided in the communities we targeted. Nonetheless, we can be confident, especially with respect to the community samples we drew, that most of the users whose tweets we examined probably do live in these communities. It is our opinion that it is very unlikely that someone living across the country subscribes to the Santa Barbara City Hall Twitter account. The advantage of these local Twitter accounts is that their reach is likely only local.

However, the same cannot be said for the campus account followers. While we believe that a number of the users we obtained are likely to be students, we acknowledge that many alumni (who may live elsewhere), and local and national businesses, might follow Twitter accounts of these large universities. Given that those most likely to be directly affected by this event were students, finding ways to identify them on a platform like Twitter is crucial. Few alternative methods to gathering lists of students from other, non-university affiliated Twitter accounts exist. Even if such accounts were known, it is likely that the number of followers of those accounts would be eclipsed by the follower count of a university Twitter account. Moreover, we would have no way of knowing whether followers of those alternative accounts would vary systematically from those who follow a university account. When dealing with accounts that simultaneously serve as a geographical indicator of a resident and an indicator of psychological distance of a follower not geographically proximal (e.g., alumni), no single approach will perfectly inhibit the noise in the sample. The goal with using "Big Data" is to cast a wide (but somewhat systematically targeted) net and hope that the signal is louder than the noise. Despite this limitation with university accounts, we attempted to replicate our findings across two more studies of mass shootings at institutions of higher learning to demonstrate the utility of relying on these larger Twitter accounts.

Case Study 2: Northern Arizona University Shooting

During the early hours of October 9, 2015, an altercation unfolded in the parking lot of a fraternity house on campus at Northern Arizona University. In the end, one freshman student was shot dead and three others were taken to the hospital with injuries. This tragic event gave us a unique opportunity to replicate the pattern we observed among UCSB followers in our previous case study in the same sample of NAU followers we originally used as a control group. We anticipated a reversal effect, such that around the NAU event we expected to see a significant spike in event-related negative emotion word use among NAU followers but not among followers of UCSB.

Method

Tweet collection and data cleaning. Nearly a week after the NAU shooting, and using the same method for connecting to the Twitter API used in Case Study 1, we pulled the 500 most recent tweets from campus account followers that were part of our final sample described in the previous study. Tweets not written in English or that were generated before the week preceding the Isla Vista violence were deleted and 176,129 tweets were retained. This new batch of tweets was then merged with the data from Study 1. Duplicate tweets between these batches were identified and 15,826 were deleted. The time frame for this merged data set began seven days prior to the Isla Vista violence in May, 2014 and continued until 4 days after the shooting at NAU (~17 months). In all, this merged data set contained 160,303 tweets representing 960 UCSB followers and 832 NAU followers.

Measures.

Negative emotion. As in Case Study 1, we counted the number of negative emotion words in each tweet using words from the LIWC dictionary and coded tweets dichotomously (0 = no negative emotion, 1 = contains negative emotion).

Event-related words and negative emotion. A list of nine context-specific words related to the shooting was generated from an examination of tweets that originated the day of the shooting (e.g., NAUshooting, FlagstaffStrong, etc.). These words were combined with the general event-related words used in the previous case study and programmed into the custom word counter script we created in R. A random sample of 500 tweets from the day of the shooting was given to two independent coders, blind to the date of the tweets, who were instructed to independently indicate whether a tweet was about the NAU shooting. An analysis of human and automated counter ratings yielded a chance-corrected agreement kappa score of .71 (97.41% average chance-uncorrected agreement). The remaining tweets were coded dichotomously such that if they contained at least one event-related word, and at least one negative emotion word, they were coded as 1 (all others coded 0). This allowed us to examine negative emotion expression related to the NAU shooting.

Analytic strategy. To capture changes in event-related negative emotion, we employed a piecewise regression analysis where we segmented the 17-month timeframe into four time frames: the 7 days before the IV violence (Segment A), the day after the violence to a year later (Segment B), a year before the NAU shooting, up to the day of (Segment C), and the four days following the NAU shooting (Segment D). Notches in the timeframe were placed the day after the IV violence (as we did in Case Study 1) and the day of the NAU violence (which occurred in the early morning hours).

It is important to note that tweets occurring between the eighth day after the IV violence and before the fourth day preceding the shooting at NAU were placed into three groups spanning roughly 160 days each. For example, tweets occurring 4 months after the IV violence were placed in the first group, while tweets generated 14 months after were placed in the third group. We also placed a notch on the middle group which represented roughly a year after IV and a year before NAU. This was done to allow for negative emotion expressed after the IV shooting among UCSB followers to return to baseline at the end of Segment B, thus providing us with a more accurate starting estimate at the beginning of Segment C. Because we were primarily interested in determining whether there was a significant increase in event-related negative emotion expressed as a result of the NAU event, we did not compare slopes between or within communities across time. Rather we compared the estimated and observed proportions of negative emotion words, and negative emotion words with event-related words, respectively.

Results

General negative emotion expression among NAU followers increased significantly following the shooting (7.4%; b = .21, SE = .07, p = .004, Cohen's d = .09). Graphs depicting the raw and adjusted proportions of tweets containing event-related negative emotion for each sample of campus followers across a 17-month time period are presented in Figure 9. We found that after the NAU shooting, the adjusted proportion of tweets containing event-related negative emotion increased by 7.0% among NAU followers (b = 1.01, SE = .18, p = .001, Cohen's d = .43). In contrast, no significant increase was observed among UCSB followers (.02%, b = -.04, SE = .06, p = .50).

Discussion

The results of this study provide a compelling replication of the pattern we observed in Case Study 1 within the community of followers that were used as a control in that study.

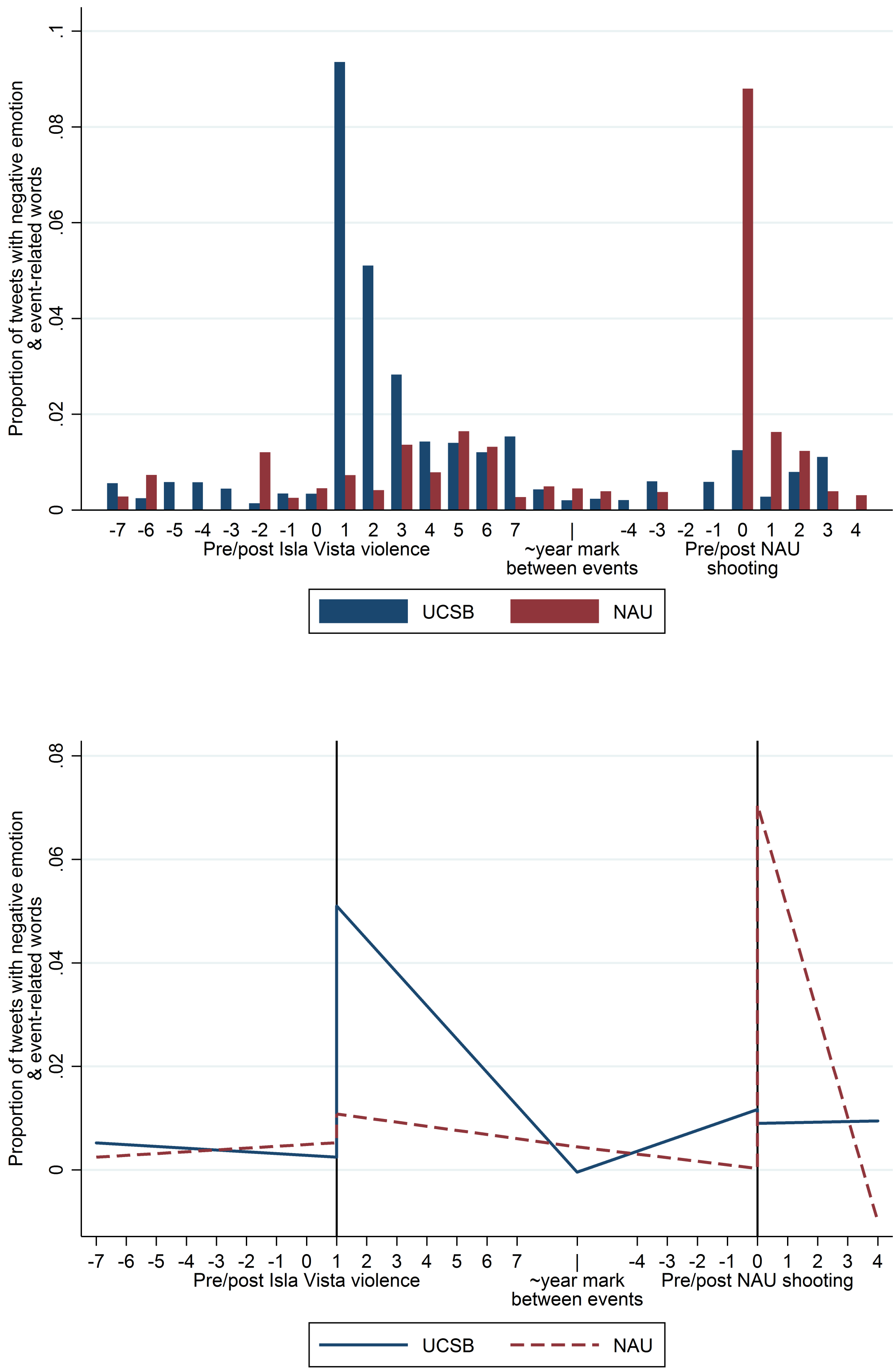

Figure 9. Seventeen-month timeline of raw (top) and piecewise adjusted proportions (bottom) of tweets containing both negative emotion words and event-related words in the days around the Isla Vista killings and the NAU shooting.

The shooting at NAU provided us with a unique opportunity to test one of the strongest quasi-experimental designs possible: an interrupted time series with a control group and a reversal (cf. Shadish et al., 2002). In this design, both groups received the treatment (violence) at different points in time, thereby providing strong evidence that increases in negative emotion expression we observed can be attributed to these violent acts. The reversal we observed also suggests that the pattern of results we saw in Study 1 was not particular to the public followers who were selected to be in the UCSB follower sample. It also provides evidence that our method is robust to the severity of the act of violence. For example, Isla Vista residents experienced an act of mass violence where many people died, potentially affecting many individuals connected to the victims, whereas people affiliated with NAU woke up to news about a targeted act of violence after an altercation on campus where only a single person was killed.

Alternatively, the pattern observed among NAU followers could represent a cumulative effect of college-campus shootings proximal in time to the NAU incident: there was a mass-shooting at Umpqua Community College (UCC) in Roseburg, Oregon days before and a shooting at Texas Southern University later the same day as the shooting at NAU. We concede that it is possible that some tweets we counted as being tweets about the NAU shooting could be tweets about those other shootings, or any combination of them. This potential history effect weakens the claim that the NAU shooting alone caused the observed increase in negative emotion among NAU followers. Thus, we attempted to replicate this pattern in another event of mass violence.

Case Study 3: Umpqua Community College Shooting

On October 1, 2015, an armed student fatally shot an assistant professor and eight students, and injured nine others inside a classroom at Umpqua Community College (UCC). After a brief shootout with police, the gunman committed suicide. This shooting was the first in a series of school shooting incidents to unfold in the month of October, 2015. Our aim in this study was to replicate the pattern we observed among UCSB followers in the wake of the Isla Vista violence and among NAU followers after the shooting at a NAU campus fraternity house. Principally, we were concerned with using our method to examine the effect of a school shooting event that had not been contaminated by other similar events proximal in time.

Method

Comparison community selection. Roseburg, Oregon is the county seat of Douglas County, with a population of 21,903. We sought to identify a comparison community based on these features of Roseburg. As in Case Study 1, we examined city data from the U.S. Census to identify our comparison community. We compared Roseburg with a number of candidates on the west coast, but no good matches were found. We then expanded our search into the midwest and southern states, finally landing on Kerrville, Texas, a city similar in size, population demographics, and political leaning that is also the county seat. Kerrville does not have a community college, but a small private college, Schreiner University, is located there. Full-time enrollment is roughly one-third of the size of UCC enrollment.

Twitter account selection and data collection. We identified the main Twitter accounts for UCC (@umpquacc) and Schreiner University (SU; @schreiner). Roughly 2 hours after the shooting occurred, we downloaded the list of all followers of UCC (n = 466). Four days after the shooting, we downloaded the user names of all followers of SU (n = 1,011) and began downloading the most recent 500 tweets from all public users in each list. Because the number of followers for each of these accounts was far below the selection thresholds we employed in previous data collection efforts (e.g., random sample of 2,500 users), we opted to download tweets from all users. We collected a total of 148,026 tweets. We screened for and deleted (a) tweets from accounts created after the shooting (58 tweets); (b) tweets that were generated more than 4 days before the shooting (137,339); (c) tweets the CLDR package identified as being non-English language tweets; and (d) tweets duplicated within or across users (63 tweets). This left us with a total of 9,668 tweets, representing 611 individuals (294 UCC followers and 327 SU followers) in the 4 days before and after the shooting.

Measures.

Negative emotion. As in Case Studies 1 and 2, we counted the number of negative emotion words in each tweet using words from the LIWC dictionary and coded tweets dichotomously (0 = no negative emotion, 1 = contains negative emotion).

Event-related words and negative emotion. A list of nine context-specific words related to the shooting was generated from an examination of tweets that originated the day of the shooting (e.g., UCCstrong, UmpquaShooting, etc.). These words were combined with the general event-related words used in the previous two case studies and programmed into the custom word counter. A random of sample of 500 tweets from the day of the shooting was given to two independent coders, blind to the date of the tweets, who were instructed to independently indicate whether a tweet was about the UCC shooting. An analysis of human and automated counter ratings yielded a chance-corrected agreement kappa score of .82 (93.12% average chance-uncorrected agreement). Each tweet in the dataset containing at least one of these words and a negative emotion word was coded as 1 (all others coded 0). This allowed us to examine negative emotion expression related to the UCC shooting.

Analytic strategy. We used a piecewise regression analysis with two segments: 4 days before the shooting (Segment A) and 4 days afterward (Segment B), with a notch placed on the day of the shooting. Because we were primarily interested in determining whether there was a significant increase in negative emotion as a result of the UCC shooting, we compared the adjusted proportion of event-related negative emotion expression with an estimated proportion had the shooting not occurred, for followers of UCC and for followers of SU, our control group.

Results

Negative emotion expression among UCC followers increased significantly following the shooting (8.1%; $b = .21$, $SE = .06$, $p < .001$, Cohen's $d = .19$), whereas no increase among SU followers was observed (0%, $b = -.13$, $SE = .08$, $p = .13$). Graphs depicting the raw and adjusted proportions of tweets containing negative emotion and event-related words for each sample of campus followers across the 8-day window around the UCC shooting are presented in Figure 10. Among followers of UCC, we found that after the violence, the proportion of tweets containing event-related negative emotion increased by 10.7% ($b = .59$, $SE = .07$, $p < .001$, Cohen's $d = .54$). No increase was observed among SU followers (<1%, $b = .01$, $SE = .02$, $p = .45$).

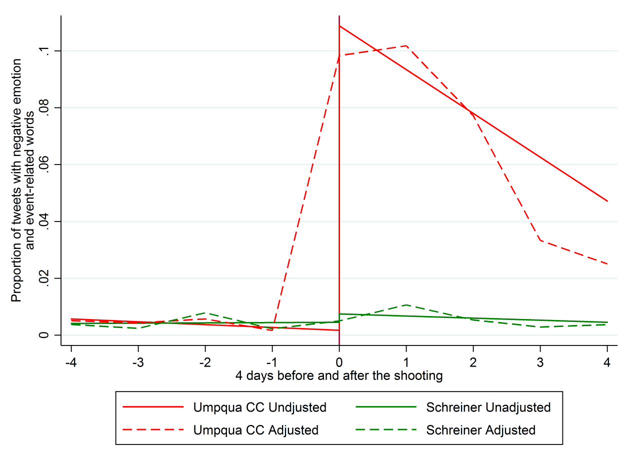

Figure 10. Raw and piecewise adjusted proportions of tweets containing both negative emotion words and event-related words in the days around the Umpqua Community College shooting.

Discussion

Within days of the UCC shooting we were able to collect thousands of tweets from hundreds of followers of the UCC Twitter account. Our aim was to replicate our findings from the previous studies among a group of users who were not affected by the history effects present in Case Study 2. We demonstrated that following the shooting at UCC, event-related negative emotion expression increased significantly among UCC followers, relative to a control community. These findings provide evidence of the external validity of our methodology, as its utility in identifying event-related negative emotion spanned across multiple situations and people.

One important feature of this study is the rapidity with which we downloaded the UCC follower list. As soon as we saw the story unfolding in the media, we downloaded the list of users who followed UCCâĂŹs Twitter account. It is important to do this as soon as possible for one very important reason: the information Twitter provides about each user does not include when a user began following the account. It is therefore impossible to filter out individuals who followed the account because they were interested in keeping abreast of the latest updates from the campus Twitter account or wanting to show support for the affected community. We recommend acquiring a follower list very shortly after an event unfolds to mitigate contamination of the list after media coverage becomes widespread.

General Discussion

Across three case studies, we demonstrated the utility of a novel method for capturing negative emotion expression among individuals likely to be part of communities that have experienced a traumatic event (e.g., violence). We averted many of the challenges researchers using traditional methods to study community trauma might face. We avoided the prohibitive monetary cost that would have been incurred to collect data from the number of people we obtained in our Twitter samples. Using free tools available online (e.g., R statistical software; twitteR package) to download tweets, we were able to collect data provided by users before the onset of violence in each community, allowing us to compare post-event patterns with these pre-event data. We also demonstrated how quickly a researcher can enter the field to collect data after a target event.

Limitations

Despite the creativity and speed of this technique, we must acknowledge that working with Twitter data is messy at best. Researchers interested in using Twitter should be mindful of biases at each step of the process. Below we consider some possible limitations of our approach.

We know from past work that Twitter users are not representative of the general population, as they are typically younger (18–29), and are socioeconomically advantaged (Duggan, Ellison, Lampe, Lenhart, & Madden, 2014). Although this does limit the extent to which we can generalize our findings to the communities we studied as a whole, we can say with some confidence that the pattern of results observed among campus account followers are likely representative of the students at the universities we targeted in this article. Concerns over representativeness will always be woven into studies of individuals using technologies that are not yet widely adopted in the general population.

User location. We acknowledge that our method of identifying Twitter users who are likely members of a community is not without limitations. The first is that there is no guarantee that all of the users we identified live in the communities we targeted. Moreover, lists of users who follow large campus accounts contain more opportunities for error than of those who follow community accounts, as university accounts may be followed by students, alumni, faculty, university departments, and local and national businesses. One imperfect way to mitigate this error would be to filter out users who indicate living somewhere other than the target community. User lists downloaded from the Twitter API provides the text a user entered into the location field for his or her profile. For example, in our study of the IV violence, we could have chosen to delete any user with a specified location other than Santa Barbara, Isla Vista, Goleta, or UCSB. However, doing so would have come at great cost. Recall that we pulled a list of 15,747 users who follow UCSB; had we filtered out users based on the presence of these city names in the location field, we would have lost 83% of the original sample (with 2,564 users remaining) from which we drew our random sample of 2,500. We also would have deleted any users who attend college at UCSB but entered their hometown into the location field.

The strategy of only retaining users who indicate living in a target area is not a bad one, if the ratio of loss is not uncomfortably high. In a subsequent data collection period, we identified a community account (@edhat) in Santa Barbara previously overlooked. This local news account, when we found it, had 3,666 public followers. Removing accounts with a city other than a Santa Barbara area community name listed in the location field would have cut that sample by 56%[3]. We suggest that doing this only makes sense insofar as it does not impede a researcher's ability to run a meaningful analysis. However, rather than creating opportunity for bias, we recommend that researchers opt to retain all individuals and run a sensitivity analysis to determine whether the pattern of results changes as a result of excluding users in this fashion.

Selection bias. We also recognize that there are selection biases at work using this method. First, we do not have a list of all Twitter accounts in a geographic area from which to randomly select accounts in each community and pull users. That is, perhaps the community Twitter accounts we selected have followers that vary systematically (e.g., by age, gender) from followers of other community accounts that were not selected. Unfortunately, there is no way for social scientists to know whether this is the case because data for user demographics per account do not exist. The use of identifying similar accounts, perhaps followed by similar types of users in a control community, is the only means by which to minimize this problem.

Second, selection biases also played a role before tweets were collected. Across our case studies, the percentage of private accounts among our samples ranged from 9% to 13%. Because data cannot be downloaded from these users, we cannot determine whether these users systematically differed in negative emotion expression from those with public accounts. Biases were also introduced after tweets were downloaded. Recall that in all case studies, we started with a very large number of tweets. The count of tweets decreased incrementally at various steps of our data cleaning process based on various inclusion criteria (e.g., certainty of English language), and by the time we completed our cleaning procedures, we were left with a small portion of the original sample of tweets. Moreover, not all users tweet every day or every week. In limiting the time frames under scrutiny, large amounts of data were excluded.

Other Considerations and Future Research

Given that much data were lost in the selection and cleaning procedures outlined above, sample sizes associated with Big Data research are an important consideration for researchers interested in using a similar approach. In Big Data studies in which data are collected from social media APIs, such as our own, there are virtually no costs associated with increasing sample sizes. As a result, we chose to use sample sizes similar to those in past research on related topics that were large enough to detect even small effects, enhancing our confidence in their reliability. We advise other researchers approaching this problem to consult traditional data collection processes undertaken in previous, related research as guidelines, but also to consider the (usually low) costs of data collection along with their statistical power needs. Additionally, when reporting findings from Big Data studies, researchers should provide appropriate measures of effect size to make clear the magnitude of the effects reported. We expect that most psychological researchers collecting data from Big Data sources will find it convenient to oversample when collecting data, resulting in increased confidence in the reliability of their findings.

Another consideration in the use of this technique is the manner in which event-related word lists are generated. We took an approach we believe was face valid, but other more systematic approaches could be used, and might be more

[3]Although the data are not presented here, we found that the patterns of event-related negative emotion expression observed in all case studies were also observed among 492 randomly selected @edhat followers, who indicated living in Santa Barbara, and tweeted during the 2-week window around the IV violence (see Figure A in Supplement).

useful for analyzing complex social contexts where the narratives present in social media data might be more dynamic (e.g., political rallies, civil unrest). For example, a researcher could build a specialized corpus of text comprised of news articles or blog posts that discuss a specific event in time and use linguistic software (e.g., WordSmith Tools; Scott, 2016) to analyze which words are used more frequently to write about the event (i.e., concordance) and/or examine clusters of words that statistically hang together in the corpus (i.e., collocation; for a good introduction to common methods used in corpus linguistics and discourse analysis, see Baker et al., 2008; Mautner, 2007). These words, or their clusters, could then be compiled into a word-list and applied to another corpus (e.g., social media text data).

How tweets are coded and manipulated to answer research questions is an important part of any empirical endeavor using social media data. Twitter data coded for emotion are versatile in that they can be manipulated in a variety of ways. For example, we reported the proportion of tweets containing negative emotion in each community across weeks and days in time; however, we could instead have analyzed the proportion of users who expressed negative emotion across time, or simply analyzed the raw counts of negative emotion words used each day, week or hour. Each strategy has its purpose and it is up to the researcher to determine the best route for telling the story. In our case, the patterns were similar across all of these strategies. However, had we gone the route of modeling a raw count of negative emotion words used, the interpretation of our findings might not have been as meaningful.

It is worth noting that although we focused exclusively on the expression of negative emotion in our analyses, researchers interested in using social media data to understand the impact of an event in time could examine other psychological states using available validated LIWC dictionaries, of which there are dozens. For example, in a study of the impact of a collective trauma, a researcher could examine a range of emotions (e.g., anger, anxiety, fear) among individuals in the affected community. Similarly, for future studies of the impact of other timely events (e.g., policy changes, political/ideological shifts), researchers have at their disposal a number of dictionaries with which they can evaluate other psychological processes and emotional states. In both cases, however, it would be important to ensure that the signal from those individual processes or emotions is detectable within a vast array of thousands (or sometimes millions) of entries in a social media-based corpus.

In addition to the closed-vocabulary approach we used in this article, there are other ways to analyze tweets to extract meaningful information from sampled users. For example, some researchers have used an open vocabulary, topic-based approach (e.g., differential language analysis) to analyze Twitter data in which topics are created by grouping semantically related words to garner insights from very large corpora of text (see Eichstaedt et al. 2015; Schwartz et al. 2013). It is unknown whether this type of approach would be useful for identifying topical clusters around an event in time (e.g., a community trauma) affecting a relatively small proportion of Twitter users in a given region, but it is worth exploring.

Future researchers could connect patterns of word usage on Twitter with subjective data collected from the users themselves. For example, in the case of trauma researchers, it might be feasible to contact Twitter users who have been identified as living in a particular community and ask them to participate in a study, ideally in the acute phase of a collective trauma. In this subsequent data collection effort, researchers could explore underlying motivations for expressing negative emotion on Twitter and, if using a longitudinal design, ascertain whether using Twitter to express these emotions is psychologically detrimental in the long run (Seery et al., 2008). In this way, researchers could reap the benefits of having archival data across time for each person that could be examined in conjunction with demographics and validated measures of distress, to triangulate on the effect the event had on community members. Moreover, data like those we collected in our research could be connected to county, city, or university-level measures of mental health, once those data become available in the future.

The versatility of Twitter data is evident in how researchers have tapped into its scalability to understand large-scale phenomena relevant to psychologists. For example, Twitter data have been used to examine global happiness (Dodds, Harris, Kloumann, Bliss, & Danforth, 2011), to predict county level health outcomes (Eichstaedt et al., 2015), and to understand how gangs in an urban area communicate (Patton, Eschmann, & Butler, 2013). Analyses conducted at these large scales are useful and informative. However, we see the method we have presented as a Big Data approach to studying topics that exist on a smaller, community scale.

Although we present a picture of challenges with working with these data (e.g., the lack of demographic details about users), researchers are finding creative ways to infer important details about Twitter users they sample. For example, using techniques like machine learning, computer scientists are using Twitter data to estimate user political affiliation (Conover, Gonçalves, Ratkiewicz, Flammini, & Menczer, 2011), and demographic information (Volkova, Bachrach, Armstrong, & Sharma, 2015). We suspect that the tools for overcoming these challenges will continue to be developed so long as researchers are interested in accessing Twitter data and recognize the excitement they provide. Moreover, social scientists interested in working with text-based social media data should keep abreast of the modeling techniques and programmatic tools being developed, and used, by computer scientists in field of natural language processing to determine the best way to answer their substantive research questions.

Conclusion

Based on our findings, we believe that Twitter is a creative tool with which to identify negative emotion in the aftermath violence on college campuses, in particular because students are more likely to be Twitter users than others in the general population. Free, easy access to Twitter data, as it stands today, offers researchers a unique ability to engage in naturalistic observation of a large number of individuals that is difficult to achieve with other social media platforms that require a user's permission to access timeline content. However, whether Twitter remains a relevant technological feature of everyday life remains to be seen, and researchers interested in using a Big Data approach to study community-level effects should keep abreast of adoption trends for new technologies. It could be the case that a few years from now, connecting to Twitter as we have outlined it here, will not be possible. We believe, however, that regardless of the social media platform du jour, a researcher interested in using a Big Data approach to study how trauma affects a community, should make every effort to begin his or her inquiry with the users who are likely to live there.

References

Baker, P., Gabrielatos, C., Khosravinik, M., Krzyżanowski, M., McEnery, T., & Wodak, R. (2008). A useful methodological synergy? combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse & Society*, *19*, 273–306. doi: 10.1177/0957926508088962

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: New York, NY: Cambridge University Press. doi: 10.1017/CBO9780511804489

Bravo, M., Rubio-Stipec, M., Canino, G. J., Woodbury, M. A., & Ribera, J. C. (1990). The psychological sequelae of disaster stress prospectively and retrospectively evaluated. *American Journal of Community Psychology*, *18*, 661–680. doi: 10.1007/BF00931236

Cohn, M. A., Mehl, M. R., & Pennebaker, J. W. (2004). Linguistic markers of psychological change surrounding September 11, 2001. *Psychological science*, *15*, 687–693. doi: /10.1111/j.0956-7976.2004.00741.x

Conover, M. D., Gonçalves, B., Ratkiewicz, J., Flammini, A., & Menczer, F. (2011, March). Predicting the political alignment of Twitter users. In *2011 ieee international conference on privacy, security, risk and trust and ieee international conference on social computing* (pp. 192–199). Palo Alto, CA: Conference Publishing Services. doi: 10.1109/PASSAT/SocialCom.2011.34

De Choudhury, M., Monroy-Hernandez, A., & Mark, G. (2014, April). Narco emotions: Affect and desensitization in social media during the Mexican drug war. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 3563–3572). New York, NY: ACM. doi: 10.1145/2556288.2557197

Dodds, P. S., Harris, K. D., Kloumann, I. M., Bliss, C. A., & Danforth, C. M. (2011). Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter. *PloS ONE*, *6*, e26752.

Doré, B., Ort, L., Braverman, O., & Ochsner, K. N. (2015). Sadness shifts to anxiety over time and distance from the national tragedy in newtown, connecticut. *Psychological Science*, *26*, 363–373. doi: 10.1177/0956797614562218

Duggan, M., Ellison, N. B., Lampe, C., Lenhart, A., & Madden, M. (2014). *Social media update 2014.* Retrieved from http://www.pewinternet.org/files/2015/01/PI_SocialMediaUpdate20144.pdf

Eichstaedt, J. C., Schwartz, H. A., Kern, M. L., Park, G., Labarthe, D. R., Merchant, R. M., ... others (2015). Psychological language on twitter predicts county-level heart disease mortality. *Psychological science*, *26*, 159–169. doi: 10.1177/0956797614557867

Frattaroli, J. (2006). Experimental disclosure and its moderators: a meta-analysis. *Psychological Bulletin*, *132*, 823–865. doi: 10.1037/0033-2909.132.6.823

Gentry, J. (2015). *twitteR: R based Twitter client (Version 1.1.9) [Computer software].* Retrieved from https://cran.r-project.org/web/packages/twitteR/twitteR.pdf

Glasgow, K., Fink, C., & Boyd-Graber, J. L. (2014, May). "our grief is unspeakable": Automatically measuring the community impact of a tragedy. In *Proceedings of the 8th international AAAI Conference on Weblogs and Social Media.* Palo Alto, CA: AAAI Press. Retrieved from http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8086

Gortner, E.-M., & Pennebaker, J. W. (2003). The archival anatomy of a disaster: Media coverage and community-wide health effects of the Texas A&M bonfire tragedy. *Journal of Social and Clinical Psychology*, *22*, 580. doi: 10.1521/jscp.22.5.580.22923

Guillory, J., Spiegel, J., Drislane, M., Weiss, B., Donner, W., & Hancock, J. (2011). Upset now?: Emotion contagion in distributed groups. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 745–748). New York, NY, USA: ACM. doi: 10.1145/1978942.1979049

Hancock, J. T., Gee, K., Ciaccio, K., & Lin, J. M.-H. (2008, November). I'm sad you're sad: Emotional contagion in cmc. In *Proceedings of the 2008 ACM conference on computer supported cooperative work* (pp. 295–298). doi: 10.1145/1460563.1460611

Holman, E. A., Silver, R. C., Poulin, M., Andersen, J., Gil-Rivas, V., & McIntosh, D. N. (2008). Terrorism, acute stress, and cardiovascular health: A 3-year national study following the September 11th attacks. *Archives of General Psychiatry*, *65*, 73–80. doi: 10.1001/archgenpsychiatry.2007.6

Kahn, J. H., Tobin, R. M., Massey, A. E., & Anderson, J. A. (2007). Measuring emotional expression with the Linguistic Inquiry and Word Count. *The American Journal of Psychology*, *120*, 263–286. Retrieved from http://www.jstor.org/stable/20445398

Kim, H., Fay, M. P., Feuer, E. J., Midthune, D. N., et al.

(2000). Permutation tests for joinpoint regression with applications to cancer rates. *Statistics in Medicine*, *19*, 335–351. doi: 10.1002/(sici)1097-0258(20000215)19:3%3C335::aid-sim336%3E3.0.co;2-z

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174. doi: 10.2307/2529310

Lowe, S. R., & Galea, S. (2015). The mental health consequences of mass shootings. *Trauma, Violence, & Abuse*. (Advance online publication) doi: 10.1177/1524838015591572

Mautner, G. (2007). Mining large corpora for social information: The case of elderly. *Language in Society*, *36*, 51–72. doi: 10.1017/0S0047404507070030

Mitchell, M. N. (2012). *Interpreting and visualizing regression models using Stata*. College Station, TX: Stata Press.

Murthy, D., & Longwell, S. A. (2013). Twitter and disasters: The uses of Twitter during the 2010 Pakistan floods. *Information, Communication & Society*, *16*, 837–855. doi: 10.1080/1369118X.2012.696123

Norris, F. H., Friedman, M. J., Watson, P. J., Byrne, C. M., Diaz, E., & Kaniasty, K. (2002). 60,000 disaster victims speak: Part I. An empirical review of the empirical literature, 1981–2001. *Psychiatry*, *65*, 207–239. doi: 10.1521/psyc.65.3.207.20173

North, C. S., & Pfefferbaum, B. (2002). Research on the mental health effects of terrorism. *Journal of the American Medical Association*, *288*, 633–636. doi: 10.1001/jama.288.5.633

North, C. S., Smith, E. M., & Spitznagel, E. L. (1997). One-year follow-up of survivors of a mass shooting. *American Journal of Psychiatry*, *154*, 1696–1702. doi: 10.1176/ajp.154.12.1696

Patton, D. U., Eschmann, R. D., & Butler, D. A. (2013). Internet banging: New trends in social media, gang violence, masculinity and hip hop. *Computers in Human Behavior*, *29*, A54–A59.

Pennebaker, J. W. (2011). *The secret life of pronouns: What our words say about us*. New York, NY: Bloomsbury Press.

Pennebaker, J. W., & Francis, M. E. (1996). Cognitive, emotional, and language processes in disclosure. *Cognition & Emotion*, *10*, 601–626. doi: 10.1080/026999396380079

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology*, *77*, 1296-1312. doi: 10.1037/0022-3514.77.6.1296

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, *54*, 547–577. doi: 10.1037/0022-3514.77.6.1296

Primo, D. M., Jacobsmeier, M. L., & Milyo, J. (2007). Estimating the impact of state policies and institutions with mixed-level data. *State Politics & Policy Quarterly*, *7*, 446–459. doi: 10.1177/153244000700700405

R Core Team. (2015). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Ritter, R. S., Preston, J. L., & Hernandez, I. (2013). Happy tweets: Christians are happier, more socially connected, and less analytical than atheists on Twitter. *Social Psychological and Personality Science*, *5*, 243–249. doi: 10.1177/1948550613492345

Rude, S., Gortner, E.-M., & Pennebaker, J. (2004). Language use of depressed and depression-vulnerable college students. *Cognition & Emotion*, *18*, 1121–1133. doi: 10.1080/02699930441000030

Schlenger, W. E., Caddell, J. M., Ebert, L., Jordan, B. K., Rourke, K. M., Wilson, D., … Kulka, R. A. (2002). Psychological reactions to terrorist attacks: Findings from the National Study of Americans' Reactions to September 11. *Journal of the American Medical Association*, *288*, 581–588. doi: 10.1001/jama.288.5.581

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., … others (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS ONE*, *8*, e73791. doi: 10.1371/journal.pone.0073791

Scott, M. (2016). *WordSmith Tools version 7.* Stroud, UK: Lexical Analysis Software.

Seery, M. D., Silver, R. C., Holman, E. A., Ence, W. A., & Chu, T. Q. (2008). Expressing thoughts and feelings following a collective trauma: Immediate responses to 9/11 predict negative outcomes in a national sample. *Journal of Consulting and Clinical Psychology*, 657–667. doi: 10.1037/0022-006X.76.4.657

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference* (2nd ed.). Boston, MA: Houghton Mifflin Company.

Silver, R. C. (2004). Conducting research after the 9/11 terrorist attacks: Challenges and results. *Families, Systems, & Health*, *22*, 47–51. doi: 10.1037/1091-7527.22.1.47

Stone, L. D., & Pennebaker, J. W. (2002). Trauma in real time: Talking and avoiding online conversations about the death of Princess Diana. *Basic and Applied Social Psychology*, *24*, 173–183. doi: 10.1207/S15324834BASP2403_1

Sutton, J., League, C., Sellnow, T. L., & Sellnow, D. D. (2015). Terse messaging and public health in the midst of natural disasters: The case of the Boulder floods. *Health Communication*, *30*, 135–143. doi: 10.1080/10410236.2014.974124

Sutton, J., Spiro, E. S., Fitzhugh, S., Johnson, B., Gibson, B., & Butts, C. T. (2014, May). Terse message amplification in the boston bombing response. In *Proceedings of the 11th International ISCRAM Conference* (pp. 612–621). doi: 10.1080/10410236.2014.974124

Sylwester, K., & Purver, M. (2015). Twitter language use reflects psychological differences between democrats and republicans. *PloS ONE*, *10*, e0137422. doi: 10.1371/journal.pone.0137422

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*, 24–54. doi: 10.1177/0261927X09351676

Thistlethwaite, D. L., & Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology*, *51*, 309. doi: 10.1037/h0044319

UCLA Statistical Consulting Group. (2015). *Regression with stata: Beyond ols.* Retrieved from http://www.ats.ucla.edu/stat/stata/webbooks/reg/chapter4/statareg4.htm

Volkova, S., Bachrach, Y., Armstrong, M., & Sharma, V. (2015, april). Inferring latent user properties from texts published in social media. In *29th AAAI Conference on Artificial Intelligence* (pp. 4296–4297). Palo Alto, CA. Retrieved from http://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9358/9858

Wicke, T., & Silver, R. C. (2009). A community responds to collective trauma: An ecological analysis of the James Byrd murder in Jasper, Texas. *American Journal of Community Psychology*, *44*, 233–248. doi: 10.1007/s10464-009-9262-8

Wojcik, S. P., Hovasapian, A., Graham, J., Motyl, M., & Ditto, P. H. (2015). Conservatives report, but liberals display, greater happiness. *Science*, *347*, 1243–1246. doi: 10.1126/science.1260817

Doc creation date: 11.8.2016