Modeling Stress with Social Media Around Incidents of Gun Violence on College Campuses

KOUSTUV SAHA, Georgia Institute of Technology
MUNMUN DE CHOUDHURY, Georgia Institute of Technology

Stress constitutes a persistent wellbeing challenge to college students, impacting their personal, social, and academic life. However, violent events on campuses may aggravate student stress, due to the induced fear and trauma. In this paper, leveraging social media as a passive sensor of stress, we propose novel computational techniques to quantify and examine stress responses after gun violence on college campuses. We first present a machine learning classifier for inferring stress expression in Reddit posts, which achieves an accuracy of 82%. Next, focusing on 12 incidents of campus gun violence in the past five years, and social media data gathered from college Reddit communities, our methods reveal amplified stress levels following the violent incidents, which deviate from usual stress patterns on the campuses. Further, distinctive temporal and linguistic changes characterize the campus populations, such as reduced cognition, higher self pre-occupation and death-related conversations. We discuss the implications of our work in improving mental wellbeing and rehabilitation efforts around crisis events in college student populations.

CCS Concepts: • **Human-centered computing** → *Empirical studies in collaborative and social computing*; • **Applied computing** → *Psychology*;

Additional Key Words and Phrases: social media; Reddit; campus mental health; mental health; stress; gun violence; crisis; wellbeing; college students

ACM Reference format:
Koustuv Saha and Munmun De Choudhury. 2017. Modeling Stress with Social Media Around Incidents of Gun Violence on College Campuses. *Proc. ACM Hum.-Comput. Interact.* 1, 2, Article 92 (November 2017), 27 pages.
https://doi.org/10.1145/3134727

1 INTRODUCTION

Stress is a psychological reaction that occurs when an individual perceives that environmental demands exceed his or her adaptive capacity [73]. One of the populations particularly vulnerable to stress is college students [8]. Stress is one of the most commonly identified impediment to academic performance and student retention in colleges [92]. When stress becomes excessive, students experience physical and psychological impairment, and intensified stress can undermine resilience factors, such as hope.

However, external factors are also known to exacerbate college student stress [68]. A prominent set of such environmental attributes includes exposure to traumatic and violent events, which

The authors were partly supported through NIH grant #1R01GM11269701.
Authors' addresses: Koustuv Saha, School of Interactive Computing, College of Computing, Georgia Tech, Atlanta, GA 30332, US, koustuv.saha@gatech.edu; Munmun De Choudhury, School of Interactive Computing, College of Computing, Georgia Tech, Atlanta, GA 30332, US, munmund@gatech.edu.

2573-0142/2017/11-ART92 $15.00
https://doi.org/10.1145/3134727

can have a profound impact on college students' perceived stress and stress responses [72]. Overwhelming amounts of stress emanating from this exposure can impact students' ability to cope, and regulate their emotions. In some cases, persistent stress episodes may eventually lead to serious, long-term negative mental health outcomes, such as post-traumatic stress disorder, acute stress disorder, borderline personality disorder, or adjustment disorder [91].

Violent incidents on college campuses, ranging from mass shootings to acts of terrorism have proliferated in the past few years. A survey from Everytown for Gun Safety Support Research[1] reports that between 2013 and 2016, 76 incidents of gun violence alone have occurred on U.S. college campuses, resulting in more than 100 casualties. It is reported that many of these incidents not only affected those involved in the incidents directly, they have also left a profound negative psychological impact on the general campus population [92]. Colleges are valued institutions that help build upon a society's foundations and serve as an arena where the growth and stability of future generations begin. In that light, it is vital to understand the impacts that violent incidents have within the psyches of students on a college campus.

However, methods to assess the stress experiences of college students are plagued by challenges in access to timely information, exacerbated by the social stigma of the condition, lack of awareness of the condition, and the noted acceptance of stress in colleges as a "badge of honor" [6]. Since many of these assessment techniques rely on retrospective recall, they are susceptible to missing out real-time, fine-grained information or sudden fluctuations in psychological signals [81]. Moreover, because of the stigma associated with stress, self-reported and active solicitation techniques to identify students' stress vulnerabilities are likely to be less reliable [6]. Violent incidents on campuses may further amplify these limitations. Due to the unique circumstances presented to a community exposed to a crisis, rehabilitation efforts that utilize student contributed data on stress and well-being are likely to be difficult to implement, conduct and leverage in a timely fashion [81].

This paper addresses these gaps by inferring and analyzing the perception of stress on college campuses around gun violence, as manifested on social media. Our work is motivated from two complementary research directions. Studies in psycholinguistics and crisis informatics have found promising evidence that the linguistic characteristics of content shared on social media can help us understand and infer psychological states of individuals and communities [25, 48, 77]. Second, over 90% of young adults, or individuals of college going age use social media [2], providing a promising opportunity to study college students' mental well-being status unobtrusively and passively using social media data [6]. We focus on the following three research questions in this paper:

RQ1. How can we automatically infer the expression of stress in social media posts?
RQ2. How can we quantify *temporal* changes in stress expressions following gun violence on campuses?
RQ3. How can we quantify *linguistic* changes in stress expressions following gun violence on campuses?

We focus on 12 gun violence incidents reported over the past five years (2012-2016) on different U.S. college campuses. For each campus, we obtain historical data around the incidents, contributed in college-specific Reddit communities. Then, targeting RQ1, we first develop an inductive transfer learning approach to infer stress expressed in Reddit posts, which achieves a mean accuracy of 82%. We apply this classification framework to identify high stress posts shared in the 12 different college-specific Reddit communities. Then to address RQ2 and RQ3, that is, to assess the extent to which expressions of high stress change in the aftermath of the incidents, we develop computational techniques drawing from the time series analysis and computational linguistics literatures.

[1]everytownresearch.org Accessed: 2017-04-09
[2]pewinternet.org Accessed: 2017-04-20

We employ a causal inference based analytical approach, in conjunction with these computational techniques, to examine the outcomes of our RQs. We observe that, compared to a control (gun violence free) time period on each campus, our methods reveal a rise in the volume of posts expressing high stress following the violent incidents, including a considerable change in the patterns of stress expressed in the immediate aftermath of the incidents. Then, psycholinguistic characterization of the high stress posts indicates that campus populations exhibit reduced cognitive processing and greater self attention and social orientation, and that they participate more in death-related conversations. Additionally, a lexical analysis of high stress posts shows distinctive temporal trends in the use of incident-specific words on each campus, providing further evidence of the impact of the incidents on the stress responses of campus populations.

To our knowledge, we present the first multi-campus study of the expression of stress in social media in colleges affected by gun-violence incidents, including methods to quantify changes in these stress expressions. We situate our findings in the context of psychological theories surrounding trauma and crisis. In closing, we discuss the implications of our work in improving stress measurement around crisis events, and in the design of tools that can enable timely detection of and intervention to tackle the psychological impacts of such crises.

2 RELATED WORK

2.1 Measuring Stress on College Campuses

There is a rich body of literature that has focused on assessing stress, its causes, and its effects in college students [8]. A variety of personal and academic life factors and environmental stressors have been found to precipitate college student stress [68]. Stressful episodes, in turn, are associated with cognitive deficits in students (e.g. concentration difficulties), decreased life satisfaction, and poor health behaviors [92].

Due to these multi-faceted risk factors and consequences of the stress experience, researchers over the past several decades have employed a variety of techniques to devise global and event-specific measures of stress in college students. For example, several investigations have modified life-event scales in an attempt to measure global perceived stress [16]. However, due to reliance on a specific list of events, this approach is insensitive to stress emanating from unforeseen or unanticipated circumstances such as crises. In other work, subjective measures of response to specific stressors have also been widely used [34]. However, it has been identified that it can be difficult and time-consuming to adequately develop, and empirically and theoretically validate an individual measure every time a new stressor is identified, such as following an environmental upheaval [17]. The widely adopted Perceived Stress Scale [17] addresses many of these challenges in stress measurement and has been employed in the study of college student stress [68]. However, data obtained from such psychometric instruments are sensitive to self-report and retrospective recall biases, and may not necessarily reveal factors associated with the stigma of stress. Moreover, such measurements of stress can only be conducted periodically, posing difficulties in understanding the temporal dynamics or evolution of college student stress.

To overcome these limitations, in recent years, researchers have employed wearable sensing technologies and experience sampling methodologies that can obtain real-time information on psychological symptoms in college students [32, 40, 88]. Although these techniques capture rich and dense behavioral signals, they lack in terms of scalability, e.g., expanding to different campus populations at the same time, and require active compliance from the participants for discovering longitudinal patterns. In this work, we seek to address these gaps in the literature, by employing unobtrusively and passively gathered naturalistic social media data to measure levels of stress in multiple college student populations.

2.2 Measuring Psychological Impacts of Gun Violence

A growing body of work in behavioral sciences has attributed that violence in any form of an attack can lead to short and long term psychological effects among individuals [74, 75, 91]. Studies have characterized such effects as “invisible wounds” in terms of psychological and cognitive injuries [37, 50]. Relatedly, Cook et al. studied the propagation of shock waves following gun violence tragedies in the schools of U.S. [19]. It is well established that exposure to violence is not only correlated with mental health concerns like depression, anxiety and distress [13, 27], it also negatively impacts a student’s academic performance [9, 41, 56, 80]. In an earlier work, Asmussen and Creswell conducted an in-depth qualitative case-study, for understanding the response to gun violence on college campuses [3], and recently Schildkraut et al. studied moral panic about shootings among college students [71].

As also noted in these studies, timely and objective measurement of the impact of crisis events on a community’s well-being is a significant challenge. Many crisis events inflict emotional toll on the affected populations, leading to disruptions in day-to-day life [81]. Although surveys like the Trauma Symptom Checklist [11] provide a mechanism to assess psychological responses to crises, such surveys are often administered only after life has restored to normalcy [81]. Because respondents are likely to be displaced in space and time from the actual occurrence of the events, such methods are likely to be impacted by the respondents’ memory and recall bias.

In this paper, we address the gaps of assessing the psychological impacts of violence, by utilizing student contributed data of college student populations, shared in time periods preceding and succeeding the violent events. Since social media is recorded in the present and preserved, it minimizes the hindsight bias sometimes induced by retrospective analyses of psychological states. Moreover, due to the ability to gather longitudinal social media data, we can quantitatively assess the psychological impacts of a violent incident by comparing with the patterns manifested during a violence-free period.

2.3 Language and Crisis

Over the years, research in crisis informatics has utilized language as a tool and lens to understand how major crisis events unfold in affected populations, and how they are covered on traditional media [29, 39, 44], online media such as blogs [48], and social media [59, 77]. This body of work has consistently shown that online platforms emerge as a safe haven for people, enabling them to interact and express themselves during times of upheavals in their environment [1, 26, 48, 77].

Since language is known to signal an individual’s underlying psychological states [15], a complementary line of research has employed social media language analysis for understanding and improving crisis rescue [2, 14, 57, 65, 85, 86]. In addition, prior studies have also identified psychological and affective responses associated with violence [42, 53, 84]. Specifically, centering around gun violence in schools, such as the Sandy Hook and the Virginia Tech shooting, multiple studies have examined the reactions of affected populations as gleaned via social media [10, 33, 59, 90].

However, there is limited work which focuses on quantifying the psychological wellbeing manifested on social media during crisis events. In an early work, Cohn et al. identified language based psychological markers of 9/11 victims, based on blog posts [18]. Recently, De Choudhury et al. examined the notion of desensitization in response to persistent violence during Mexican drug war [25], Delgado Valdes et al. analyzed the psychological effects of crime and violence [82], and Lin et al., studied the evolution of emotions in the aftermath of the Boston Marathon bombings [46]. Taken together, this body of work motivates our goal of leveraging social media to examine the expression of psychological stress around gun violence related crisis events on college campuses—a hitherto under-explored community in the crisis informatics literature.

2.4 Inferring Mental Health States with Social Media

There has been a growing body of work focusing on developing quantitative approaches in social media analytics to assess mental health and wellness states [76]. Topics of investigation in this emergent area include, mood and depressive disorders [24], post-traumatic stress disorder [20], and suicidal ideation [43]. While we are inspired by this body of work, we note that stress inference from social media is particularly challenging, and existing methods of inferring other mental health states cannot be employed to quantify stress from social media. This is because, according to the DSM-V [4], stress is a different psychological construct compared to depression or suicidal ideation. Further, stress inference from social media is challenging, because of the lack of appropriate ground truth data, and also because its DSM-V defined attributes (e.g., lack of control) are difficult to measure using standard lexicon-based approaches like LIWC—a common tool adopted in prior work. In this paper, we introduce transfer learning as a way to circumvent the challenges surrounding unavailability of ground truth, and then develop a machine learning framework that is able to infer levels of stress by factoring in not only the words in social media text, but also by modeling the semantic relationship between them.

We do note that, recently, there has been some research quantifying stress with social media data as well: Lin et al. developed deep learning models [45], and Zhao et al. built machine learning classifiers [93] to quantify an individual's stress from social media data. However, these works used a sentence labeling approach to derive ground truth on stress—e.g., a tweet saying "I feel stressed" was labeled positive. Such methods may be limiting, in that they are not theoretically motivated from the psychology literature, such as alignment with DSM criteria for stress, or nuances of stress expression that may not be captured via directed self-disclosures. Flippant references to stress on Twitter may further confound classifiers built on such data. Moreover, these works have not applied these techniques in the context of college student populations, or to understand the psychological impact of crisis events.

Especially relevant to our work, social media has also been harnessed for assessing mental well-being of college students. Bagroy et al. used data from college subreddits, to build an index of collective mental well-being [6]. Manago et al. found that social networking helps in satisfying psychosocial needs of college students [47], and Moreno et al. studied mental health disclosures by college students on social media [55]. Prior research also inferred other behavioral attributes and psychological attributes of college students, using social media [49, 83]. To the best of our knowledge, our study is the first of its kind to infer the expression of stress from social media posts, and examine the evolution of stress among college populations exposed to gun violence.

3 DATA

3.1 Gathering Campus-Specific Gun Violence Data

For our study, we adopted the definition of gun-related violence on college campuses, as published by Everytown for Gun Safety Support Research[1] – *"a shooting involving discharge of a firearm inside a college building or on campus grounds and not in self-defense"*. Everytown for Gun Safety is an American nonprofit organization, which conducts gun violence research in the U.S. However, since there is no single database for gun violence incidents on college campuses, we adopted a snowball approach to curate our dataset [5, 60] – 1) First, we collected a seed list of gun-related violence incidents on US college campuses from Everytown for Gun Safety Research; prior studies have leveraged Everytown's gun violence data as well [5]. 2) Next, we augmented this seed list with additional incidents, which qualify the same definition given above—to do so we consulted different credible online sources in an iterative fashion[3].

[3] A sample of these sources are gunviolencearchive.org, time.com, motherjones.com, huffingtonpost.com, en.wikipedia.org; All Accessed: 2017-04-09

Our curated list consisted of gun-related violence incidents in and within a close proximity of a US college campus, all of which happened between 2012 and 2016. Sample incidents in our list included the 2012 Oikos University shooting, the 2013 Santa Monica College shooting, and the 2016 UCLA shooting. Aside from purely gunfire based incidents, we note that this list included attacks with the involvement of gun as well as other weapons and violent activities (e.g., car ramming, butcher knife etc.), such as the case of the 2014 Isla Vista massacre at UCSB, and the 2016 OSU attack.

Before inclusion in our ensuing analysis, for each of these incidents, we additionally assessed the veracity of their occurrence and known after-effects from various online news sources, obtained through hand-curated search engine queries.

3.2 Finding Campus-Specific Social Media Data Sources

Having gathered the above campus specific gun-related violence data, we moved on to acquiring social media data of the respective campuses. For this purpose, we used the popular social media platform Reddit as the source of data for this research. Due to its forum structure, the platform is extensively used for both content sharing, as well as for obtaining feedback and information from a variety of communities of interest. These communities are known as "subreddits", and they include many geographically localized ones dedicated to specific university campuses. This allows a large sample of posts shared by students of a university to be collected in one place. Additionally, Reddit is a popular social media platform among the college student demographic, where they have been observed to discuss a wide variety of issues related to their personal and campus life. Moreover, semi-anonymity of Reddit is known to enable candid self-disclosure around stigmatized topics like mental health [23]. Prior work has further demonstrated that Reddit subreddit data specific to college campuses may be utilized to infer the mental well-being of the students [6].

Using this research as a foundation, among the incidents involving gun-related violence on college campuses (ref. previous subsection), we shortlisted those schools which have subreddit communities with at least 500 subscribers on the day of incident on campus. For the purpose, we utilized Reddit's subreddit search functionality feature, and retrieved number of subscribers from Reddit Metrics[4]. We found 12 such colleges meeting the criteria, and the number of subscribers in these subreddits ranges between 969 (r/NAU) to 8,936 (r/OSU).

Our choice of the above threshold of subscribers was inspired by prior work which made psychological inference from campus subreddit posts [6]. Essentially, this work provides a rough estimate of the number of unique subscribers in college subreddits that were sufficiently representative of the size of the student body at the same university, as well as of its student demographic distribution. Therefore we used these estimates in our work as a threshold to select the college subreddit pages corresponding to the gun-related violence incidents.

3.3 Compiling Treatment and Control Data from Social Media

For the above identified subreddits, we now describe our data collection technique. Since our study involves examining statistical differences in the expression of stress around the gun-related violence incidents on college campuses, we needed to ensure that the measured differences in stress are attributable to the incidents, instead of another unobserved or latent variable. In the statistics literature, these concerns around quantification of an "outcome" (stress) are typically mitigated by adopting randomized experimental approaches, where, given a "treatment" (gun-related violence incident) in the target population, an equivalent population is assigned to a "control" (gun-related violence free) condition to rule out the effects on the outcome that are attributable to confounding or omitted variables [38, 63]. An experimental approach being inappropriate in our case, we adopted

[4]redditmetrics.com Accessed: 2017-04-09

Table 1. List of gun-related violence in U.S. college campuses during 2012-16 used in our work. We also include the date, number of casualties and descriptive statistics of the corresponding subreddit communities.

College	Incident	#Casualties	Subreddit	Users	#Posts
University of Southern California	2012-10-31	4	r/USC	1,143	2,676
University of Maryland	2013-02-12	3	r/UMD	2,201	9,578
University of Central Florida	2013-03-18	1	r/ucf	2,886	13,708
Massachusetts Institute of Technology	2013-04-18	3	r/mit	1,568	1,682
Purdue University	2014-01-21	1	r/Purdue	3,605	11172
University of California Santa Barbara	2014-05-23	21	r/UCSantaBarbara	3,278	17,682
Florida State University	2014-11-20	4	r/fsu	3,859	8,150
University of South Carolina	2015-02-05	2	r/Gamecocks	1,903	1,661
University of North Carolina at Chapel Hill	2015-02-10	3	r/chapelhill	2,025	1,177
North Arizona University	2015-10-09	4	r/NAU	969	1,025
University of California, Los Angeles	2016-06-01	2	r/ucla	6,301	9,454
Ohio State University	2016-11-28	14	r/OSU	8,936	35,372

a statistical matching technique, drawing from the causal inference literature, to compile our data [67]. Specifically, for each of the 12 violent incidents, we identified two separate time periods of campus subreddit data collection:

***Treatment* Period.** We identified a period of two months following and a period of two months preceding the gun-related violence in each of the campuses. Our rationale behind the choice of the duration of period of analysis stems from prior work [43], wherein it has been observed that effects of a societal upheaval persists a limited period of time. Given our focus on college campuses that tend to follow a 4 month semesterly or a 2.5 month quarterly academic system, we deduced that a four month period around each incident that closely follows the academic system will be able to glean meaningful stress changes that are attributable to the incident. Let us assume it as the *Treatment* period.

***Control* Period.** For the combined period of two months before and two months after the gun-related violence incident on each campus, we identified an equivalent period of four months during the previous year. Gathering data from exactly the same period in the past year (when no gun violence was reported) is likely to rule out confounding effects in the measurement of temporal or linguistic differences in stress attributable to academic calendar factors, or other seasonal and periodic events that impact students' experiences, lifestyle, and activities. Moreover, since we identified this period specific to each campus, we ruled out the possibility of incorporating confounding effects attributable to campus characteristics or the nature of student population and their demographics. Let us call this time period as the *Control* period.

Finally, we leveraged the archive of all the Reddit data made available online by Google BigQuery[5] to collect data for each of the college subreddits in the *Treatment* and *Control* periods, which we refer to as *Treatment* and *Control* datasets respectively. BigQuery is a cloud based managed data warehouse, that allows third parties to access large publicly available dataset through simple SQL-type queries. Our final dataset consists of 113,337 posts[6], whose descriptive statistics are summarized in Table 1. We further demarcate each of the *Treatment* and *Control* datasets into *Before* and *After* samples based on whether the date of a post included in the *Treatment* (or *Control*) dataset is prior to or following the date of the reported incident at the corresponding campus (or the same date in the previous year).

[5]bigquery.cloud.google.com Accessed: 2017-04-09
[6]In this paper, we refer to 'posts' within campus subreddits as a unified term for both posts and comments.

4 METHODS

4.1 Building a Stress Classifier

Our first research goal is to assess levels of stress manifested in the social media posts of different college campuses (RQ1). In the absence of ground truth labels on this data, we adopted a transfer learning approach, similar to [6], wherein we first built a supervised machine learning model to classify stress expressions in posts into binary labels of *High Stress* and *Low Stress*. Then we adopted this classifier to automatically label posts in the campus subreddits.

Class Definitions. Our stress class definition is based on the established psychometric measure of stress given by the Perceived Stress Scale (PSS) [17]. In the widely used 10-item version of PSS, the interpretation of scoring identifies three categories: Scores ranging from 0-13 are considered minimal stress; those ranging from 14-26 are considered moderate stress; and scores ranging from 27-40 are considered extremely stressed. However, typically, very few people score in the third category of extreme stress – except for those who suffer from chronic stress challenges. Scores around 13 in the scale are considered average that typically split respondents' scores into two classes. Moreover, factor analysis [36] is known to reveal two factors, based on this scoring. This motivated our choice of the two classes – *Low Stress* and *High Stress*.

Transfer Learning Data. For the purpose of building a stress classifier, we collected all 1402 posts from the subreddit r/stress from December 2010 to January 2017. The r/stress community allows individuals to self-report and disclose their stressful experiences and is a support community. For example, two (paraphrased) post excerpts say: "*Feel like I am burning out (again...) Help: what do I do?*"; and "*How do I calm down when I get triggered?*". The community is also heavily moderated; hence we considered these 1402 posts as ground-truth data for *High Stress* posts. Next, we utilized a second dataset of over 100,000 random posts obtained by crawling the landing page of Reddit; this dataset was used in prior work to develop a social media based mental well-being index for college campuses [6]. We employed this dataset as a source of ground-truth data for *Low Stress* posts to build the stress classifier. To approximately balance the two classes, we used a randomly sampled 2000 posts from this dataset.

Establishing Linguistic Equivalence. In our transfer learning framework, we employ a training dataset, which is obtained from non-college campus subreddits. Hence we situate our rationale behind its appropriateness to build the stress classifier. We note that the primary demographic of Reddit constitutes young adults[2], which is also the predominant demographic of college students. Therefore we do not anticipate large linguistic differences between the content of the training dataset and our college campus posts. Nevertheless, to further demonstrate linguistic equivalence between the two sources, we borrow methods from the domain adaptation literature [21]. Specifically, we employed an approach involving pairwise comparison of word vectors [7, 69]. This technique involves first constructing word vectors using the frequently occurring *n*-grams in each source of data, and then calculating a distance metric, e.g., cosine similarity, to assess their linguistic similarity. Cosine similarity of word vectors is an effective measure of quantifying the linguistic similarity between two datasets [62], and a high value would indicate that the posts in the two datasets are linguistically equivalent.

To apply this method, we first extracted the most frequent 500 *n*-grams from our training dataset, and the same from the posts of every campus subreddit. Next, using the word-vectors of these top *n*-grams (obtained from the Google News dataset of about 100 billion words [54]), we computed the cosine similarity of the two datasets in a 300-dimensional vector space. We observed that the similarity ranged between 0.94 and 0.96, with a mean value of 0.95, providing significant confidence in our ability to use of the training data in building a stress classifier for college campus posts.

Classification Approach. Using the above training dataset, we obtained features for the stress classifier: for this, we used Stanford CoreNLP's sentiment analysis model to retrieve the sentiment class of the posts. We also obtained the top n-grams ($n = 3$) from the posts to be used as additional features. Then, using these n-gram features and the sentiment class, we developed a binary Support Vector Machine (SVM) classifier (with a linear kernel) for detecting *High Stress* and *Low Stress* in posts. Finally, we used this classifier to machine label all of the *Before* and *After* post samples shared in the *Control* and *Treatment* datasets associated with the 12 college campuses.

4.2 Quantifying Temporal Dynamics of Stress

Corresponding to RQ2, we now propose a suite of computational techniques to assess the temporal changes of stress following gun-related violence on college campuses. Drawing from the time series analysis literature, on the above classified *High Stress* posts, we develop two approaches: 1) *time domain* analysis and 2) *frequency domain* analysis.

4.2.1 Time Domain Analytic Approach. Normalized Temporal Variability of Stress. First we examined the temporal variability of *High Stress* expressed in subreddit posts around each campus specific gun-related violence (*Treatment* data), and a similar period in the previous year (*Control* data). For the purpose, we first aggregated posts shared per day, and then normalized the number of posts labeled as *High Stress* on each day. In order to assign weightage to the number of *High Stress* posts on a day as well as its proportion in this normalization, we employed a variant of the TF-IDF (term frequency-inverse document frequency) estimation technique: we multiplied the proportion of *High Stress* posts in a day with squared root of their count on the same day. We obtained this temporal variability measure of stress for both of our *Control* and *Treatment* datasets, spanning the *Before* and *After* periods. In order to reduce irregularities in time series, we smoothened the measures of the temporal variability of stress, using a non-parametric curve fitting regression method of *lowess*. Thereafter we used 0-lag cross correlation as a way to assess how the manifestations of *High Stress* around the incident differ from the same in a comparable timeframe in the past. Cross correlation measurement, that estimates the normalized cross-covariance function between two time series is an established mechanism to assess the relationship between a pair of time series signals.

Before-After Change Analysis. Continuing with time series analysis, next, to quantify to what extent stress changed, following a gun-related violence on college campus, we estimated changes in the trend of *High Stress* preceding and following the incidents. For this, we computed the z-scores of *High Stress* posts on each day for the *Before* and *After* samples in the *Treatment* dataset. z-scores quantify the standardized variation around mean value of a distribution and help estimating the relative changes in a time series data. Since z-scores don't rely on absolute values in a time series, it is a well suited metric for analyses spanning across different periods of time (like in our case), when social media activities might vary.

Next, we computed an average change in z-scores between the *Before* and *After* samples by taking difference of the mean z-scores in the two samples. Finally, we fitted linear regression models in the *Before* and *After* samples to obtain the trend manifested by *High Stress* over time.

4.2.2 Frequency Domain Analytic Approach. Crisis events are known to trigger significant disruptions in lifestyle, activities and psychological expression of affected populations [48]. In order to understand whether a gun violence incident on a college campus disrupts the general periodicity of expression of *High Stress* in social media posts, we transformed the time series of *High Stress* posts in the frequency domain. Frequency domain analyses are particularly suitable to study the rate at which the signal in a time series is varies and therefore in assessing its periodicity [87]. Taking help of Fast-Fourier Transform (FFT) algorithm [12], we obtained the distribution of frequencies (measured in terms of days) of *High Stress* posts in *Before* and *After* samples in the *Treatment* dataset. On these distributions, we quantified the disruption in periodicities of *High Stress* posts by employing two methods: 1) *spectral density analysis*; and 2) *wavelet analysis*. For the former, we compared the spectral density of the two waveforms, computed using the periodogram technique [89]. The latter analysis included, computing symmetric mean absolute percentage (SMAP) difference between the peaks at the signal waveforms of the two samples.

4.3 Quantifying Linguistic Dynamics of Stress

Per RQ3, where we are interested in assessing linguistic attributes that characterize *High Stress* posts following the campus gun violence incidents, we adopt two forms of language analysis: 1) *psycholinguistic characterization*; and 2) *incident-specific lexical analysis*.

4.3.1 Psycholinguistic Characterization. First, we seek to perform a psycholinguistic characterization of the *High Stress* posts in *Treatment* data, shared in the college campus subreddits around their respective gun-violence incidents. To do so, we employed the well-validated lexicon called Linguistic Inquiry and Word Count, or LIWC [61]. Borrowing from prior work [43], to compare the *High Stress Treatment* posts belonging to the *Before* and *After* samples, we used the following LIWC measures for understanding the expression of psychological attributes in social media: 1) *affective attributes* (categories: anger, anxiety, negative and positive affect, sadness, swear), 2) *cognitive attributes* (categories: causation, inhibition, cognitive mechanics, discrepancies, negation, tentativeness), 3) *perception* (categories: feel, hear, insight, see), 4) *interpersonal focus* (categories: first person singular, second person plural, third person plural, indefinite pronoun), 5) *temporal references* (categories: future tense, past tense, present tense), 6) *lexical density and awareness* (categories: adverbs, verbs, article, exclusive, inclusive, preposition, quantifier), 7) *biological concerns* (categories: bio, body, death, health, sexual) 8) *personal concerns* (categories: achievement, home, money, religion) and 9) *social concerns* (categories: family, friends, humans, social).

4.3.2 Incident-Specific Lexical Analysis. Second, we examined the lexical cues shared in the *High Stress* posts in *Treatment* data for each of the college campuses. Our goal here is to understand to what extent, lexical references of the specific incidents directly surface in the *High Stress* expressions shared in the college subreddits. For this, we specifically analyzed posts shared within 7 days before and after the day of incident, to focus on weekly patterns. This chosen time window of analysis was inspired from prior work [66], that demonstrates that major shifts in psychological states and emotional responses are manifested until 7 days after the date of a crisis incident. Other work in social media analytics [22, 35] further indicates that human affective and mental health patterns follow stable weekly patterns, with systematic waning and intensification through different days of the week.. We extracted 50 top occurring n-grams ($n = 1, 2, 3$) shared in the 7 days after the incident, and computed their Log Likelihood Ratio (LLR) with respect to their occurrences in posts 7 days before the incident. The LLR for an n-gram is determined by calculating the logarithm (base 2) of the ratio of its two probabilities, following add-1 smoothing. Thus, when an n-gram is comparably frequent in the two week-long periods, its LLR is close to 0; it is closer to 1, when the n-gram is more frequent in the posts after the incident, whereas, closer to -1, for the converse.

5 RESULTS

5.1 RQ 1: Inferring Stress in Social Media

5.1.1 Building a Stress Classifier. Corresponding to RQ1, we begin by presenting the results of the machine learning classifier of stress. Our binary SVM classifier used 5000 n-gram features and three boolean sentiment features of Positive, Negative and Neutral; the number of n-gram features was determined based on systematic parameter sweep. We used a k-fold (k=5) cross-validation technique to evaluate our model, and achieved a mean accuracy of 0.82. This accuracy was better than the baseline accuracy (based on a chance model) of 0.68 on this dataset. Table 2 reports the performance metrics of the stress classifier and Figure 1 shows the Receiver operating characteristic

Table 2. Performance metrics of stress classification based on k-fold cross-validation (k=5)

Metric	mean	stdev.	median	max.
Accuracy	0.82	0.11	0.78	0.90
Precision	0.83	0.14	0.77	0.92
Recall	0.82	0.09	0.78	0.88
F1-score	0.82	0.11	0.79	0.89
ROC-AUC	0.90	0.08	0.78	0.95

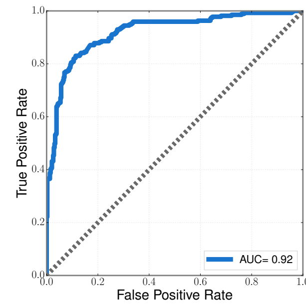


Fig. 1. ROC curve for stress classification

Table 3. Top 30 Features in stress classifier. Statistical significance reported after Bonferroni correction. (*** $p < 0.001$).

Feature	p	log(score)	Feature	p	log(score)
stress	***	9.63	thank	***	6.20
try	***	7.46	meet	***	6.17
work	***	7.20	life	***	6.07
anxiety	***	7.05	sleep	***	6.03
meditation	***	6.88	problems	***	5.98
help	***	6.81	control	***	5.95
focus	***	6.62	job	***	5.89
luck	***	6.62	good	***	5.87
breathing	***	6.44	health	***	5.87
techniques	***	6.33	week	***	5.86
feel	***	6.30	minutes	***	5.83
exercise	***	6.30	doctor	***	5.83
time	***	6.25	mental	***	5.83
play	***	6.23	relax	***	5.72
body	***	6.21	stressful	***	5.67

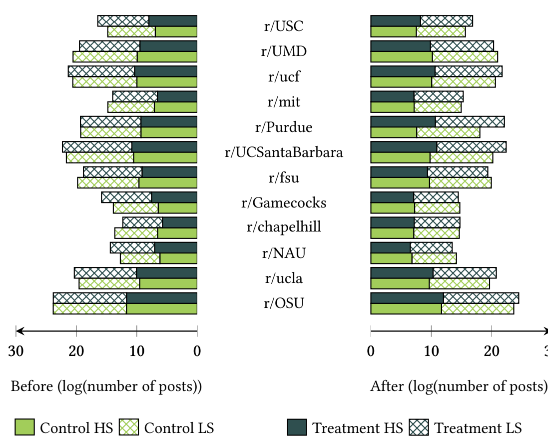
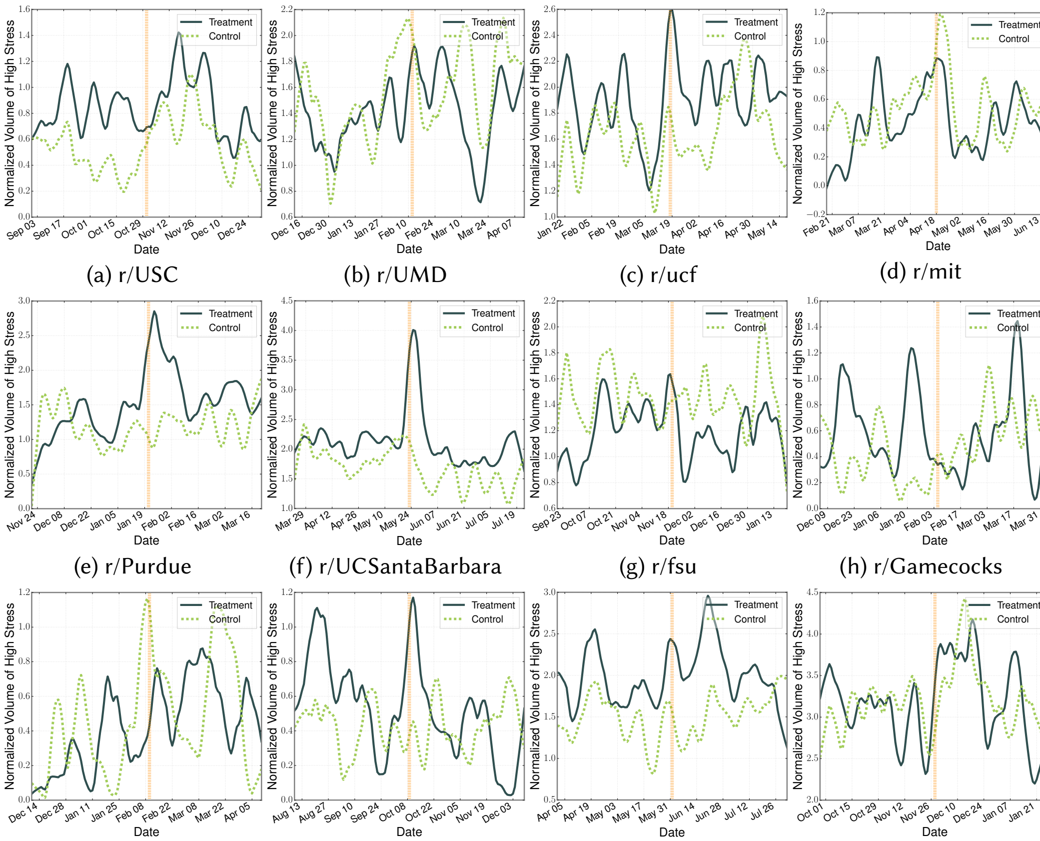


Fig. 2. *High Stress* (HS) and *Low Stress* (LS) posts in *Before* and *After* samples of *Control* and *Treatment* dataset.

(ROC) curve of the same. We find that our classifier yields low number of false positives (average precision 0.82), as well as low false negatives (average recall 0.82), indicating robust performance on test data. We conclude that our classifier is able to successfully classify Reddit posts to be expressions of *High Stress* or *Low Stress*.

What are the top predictive features of this classifier? In Table 3, we report the top 30 features of our stress classifier. We observe that a notable number of verbs or action-based nouns occur in this list, such as, *try*, *work*, *help*, *focus*. Additionally, we observe the presence of words which are contextually related to the expression of stress, like *stress*, *anxiety*, *stressful*, and *relax*. Aligning with prior work that has examined the correlates or factors precipitating stress [70], other notable words which occur in the top features include – 1) work-related: *work* and *job*; and 2) health-related: *health*, *body* and *sleep*.

5.1.2 Expert Validation of Stress Classifier. In order to understand the temporal and linguistic dynamics of *High Stress*, specific to our problem (RQ2 and RQ3), we apply the stress classifier to machine label the posts in both the *Treatment* and *Control* dataset. With the help of three human



(a) r/USC (b) r/UMD (c) r/ucf (d) r/mit

(e) r/Purdue (f) r/UCSantaBarbara (g) r/fsu (h) r/Gamecocks

(i) r/chapelhill (j) r/NAU (k) r/ucla (l) r/OSU

Fig. 3. Temporal variation in the expression of *High Stress*. The reference line represents the date of gun-violence incident.

raters, expert in social media analytics and the study of affect dynamics, we validated a random sample of 151 of the classifier labeled posts (79 *High Stress* and 72 *Low Stress* posts). Our experts adopted the Perceived Stress Scale [17] for examining how the specific concerns measured in the scale (e.g., feelings of nervousness, anger, lack of control) were expressed in each post they rated. Presence of these concerns meant a *High Stress* label, while their absence indicated *Low Stress*. Our raters reached high agreement in this task (Fleiss' $\kappa = 0.84$), and we obtained an accuracy of 82%[7] for the stress classification.

5.2 RQ 2: Temporal Dynamics of Stress

For RQ2, we begin by summarizing the results of class-wise stress distribution on each of the campuses in Figure 2. Comparing the *Treatment* and *Control* datasets spanning the *Before* and *After* periods, we find that: 1) For the *Treatment* dataset, the proportion of *High Stress* posts in the *Before* sample ranges between 35% and 45%, averaging at 40% (10,043 out of 24,737), whereas, the same for posts in the preceding year, ranges between 33% and 43%, averaging at 40% (9,415 out of 23,430); 2) the proportion of *High Stress* posts in the *After* sample of *Treatment* dataset, ranges

[7]This should not be confused with cross-validation accuracy of stress classification. Coincidentally, we obtained same value in both cases.

between 33% and 47%, with a mean value of 41% (12,834 out of 31,370), while a similar period in the preceding year, reveals a mean proportion of 42% (9,528 out of 22,816). These numbers convey that the proportion of posts expressing *High Stress* in the college subreddits remains comparably similar over an extended period of time, despite a gun violence incident on the campus. However, as our ensuing time series analysis will show, we do observe significant changes in the *patterns of expression* of *High Stress* posts in the aftermath of gun violence.

5.2.1 Time Domain Analysis of High Stress Posts. To understand how stress varies following incidents of gun related violence on college campuses, we first focus on time domain analysis of the expression of *High Stress* campus subreddits.

Temporal Variability of Stress. In Figure 3, we show the normalized volume of *High Stress* content in the *Treatment* and *Control* datasets. We observe that *High Stress* posts are shared in the college subreddits all throughout the period spanning both the *Treatment* and *Control* datasets, in varying degrees. These posts consist of content ranging across varieties of academic and college-life specific topics including, admission, examination, or assignments: "*I really should be doing homework right now...*"; and "*I applied to the PhD program. I have emailed them twice in the past few weeks, but they keep saying they aren't done reviewing applications. [..] What should I do?*". This observation aligns with prior literature that situates various college-life specific factors to be attributable to student stress [68], and that stress is a persistent psychological observation among college students [8].

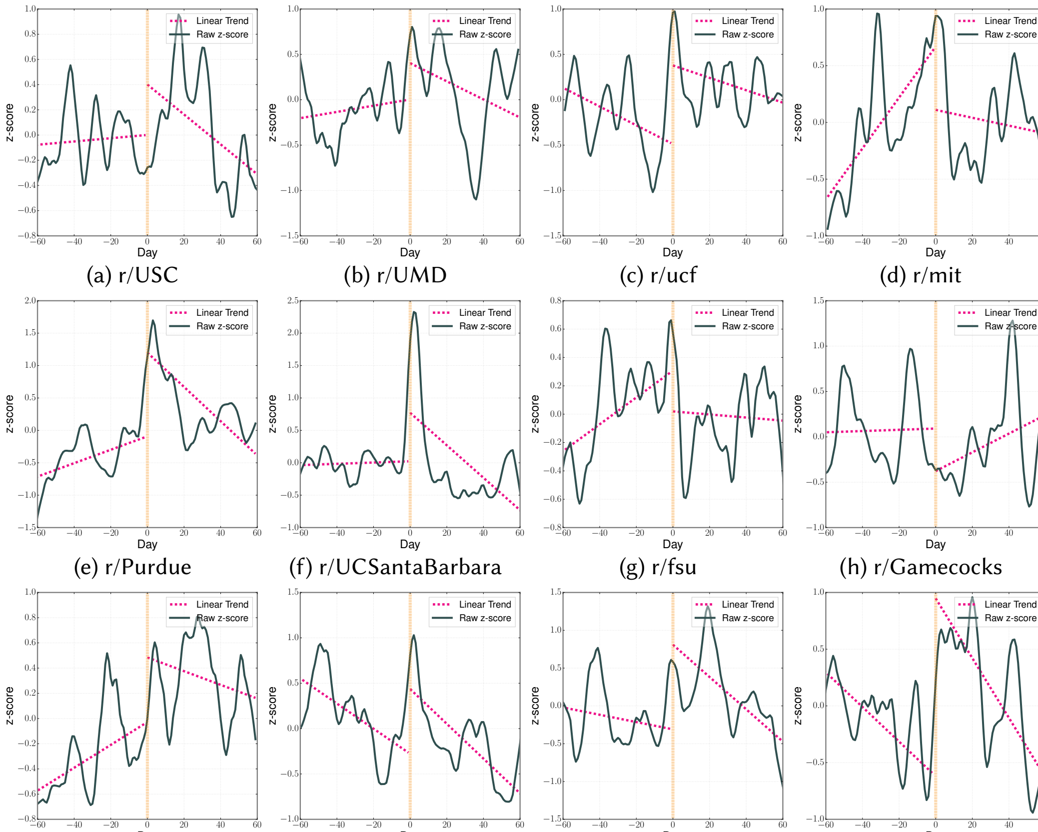
When we specifically examine the day of the gun violence incident and its vicinity, we observe a peak of the normalized volume of *High Stress* posts in a majority of the subreddits, considerably distinct in r/ucf, r/Purdue, r/UCSantaBarbara, r/NAU, r/OSU. The peak in stress in the *Treatment* year, as compared to the *Control* year supports a weak causal claim: that the campus gun violence contributes to an increased stress immediately following the incident. We also observe that the mean normalized stress in the *Treatment* year is higher than the same for *Control* across all campuses (1.35 vs. 1.19), with the maximum difference observed in r/ucla (0.49) and r/UCSantaBarbara (0.45).

To assess whether the above reported differences are statistically significant, we report the results of a cross-correlation analysis of the temporal occurrences of *High Stress* posts in the *Control* and *Treatment* datasets in Figure 6(a). We find negligible values of 0-lag cross correlations between the two time series, ranging between -0.002 and 0.003, with a mean value of 0.000 . This indicates that the differences between the pattern of *High Stress* between year of incident and a similar period in the year prior to it are indeed significant.

Before-After Change Analysis. However, how does the expression of *High Stress* in the college subreddits change in the aftermath of the gun violence incidents, compared to that before? To answer this question, we report the findings of our proposed before-after analysis (ref. Methods).

Within the *Treatment* dataset, first we conduct a cross-correlation analysis between the temporal occurrences of *High Stress* posts, which were shared *Before* and *After* the date of incident. A 0-lag cross-correlation for the *Before* and *After* samples (Figure 6(a)) ranges between -0.016 and 0.019, with a mean value of 0.005, revealing negligible correlation in the pattern of *High Stress* following the incident, as compared to before it. Next, we compute the z-scores of *High Stress* expressed on each day (Figure 4). At a glance, we observe the mean change in z-score between the *Before* and *After* samples ranges from -0.30 (r/NAU) to 0.83 (r/Purdue), with 9 out of 12 subreddits exhibiting a positive change in expression of *High Stress* (Figure 6(b)). We conducted Mann-Whitney U tests of *Before* and *After* day-wise z-scores, for which we converted the dates to ordinals by counting the number of days on the either ends of the date of incident. These tests revealed statistical significance for each of the subreddits, with U ranging between 13,585 and 18,562,124, with a mean of 2,695,608.

More careful examination of Figure 4 indicates that the z-scores of *High Stress* in the days following the incident in most of the subreddits have a trend line (based on fitting a linear model) yielding a negative slope. Specifically, we observe the most negative slopes in the cases of r/Purdue



(a) r/USC (b) r/UMD (c) r/ucf (d) r/mit

(e) r/Purdue (f) r/UCSantaBarbara (g) r/fsu (h) r/Gamecocks

(i) r/chapelhill (j) r/NAU (k) r/ucla (l) r/OSU

Fig. 4. Variation of z-scores of stress expressed in the *Before* and *After* samples in the *Treatment* dataset.

(-0.03) and r/OSU (-0.03). However, the trend line fits for *High Stress* z-scores in the *Before* period do not show such a trend– the mean slope during the period preceding the gun violence incidents is 0.001, revealing approximately a stable pattern.

Overall, our results suggest that, the expression of *High Stress* in the aftermath of gun violence shows an abrupt shift in their temporal pattern, peaking significantly around the day of the incidents, and thereafter showing a downward trend.

5.2.2 Frequency Domain Analysis of High Stress Posts. Recall our final analysis for RQ2 centers around understanding how the various gun violence incidents on campuses disrupt the periodicity of sharing *High Stress* posts. For this, working within the frequency domain, we apply Fast-Fourier Transform (FFT) on the distribution of *High Stress* posts in *Treatment* data (ref. Methods). In Figure 5, for each college subreddit, we show the distribution of frequencies $F(t)$ during the *Before* and *After* periods respectively, in a heatmap format. The color intensity of a cell in a specific heatmap indicates the probability of a certain frequency, $P(F(t))$ (measured in terms of days). Discussing our main observations from the heatmaps, in case of r/USC (Figure 5(a)), we find that the *High Stress* posts in the *Before* period showed high periodicity (i.e., exhibit peaks in expression) around every 4 and 13 days, whereas the same in the *After* period occurred at every 5, 7 and 11 days.

In order to quantify if and to what extent periodicities of *High Stress* expression in the *Before* and *After* samples, as given by the FFT approach, are disrupted around the gun violence incidents, we

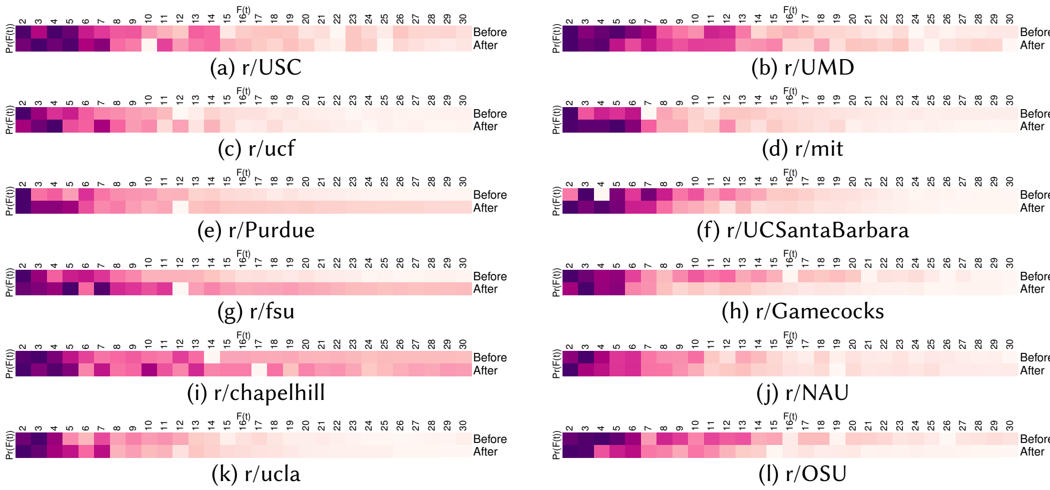


Fig. 5. Frequency distribution heatmaps of stress in *Treatment* dataset. The x-axis, $F(t)$ represents frequency where t is in terms of days, and the density of color, $Pr(F(t))$, represents the probability of *High Stress* at $F(t)$.

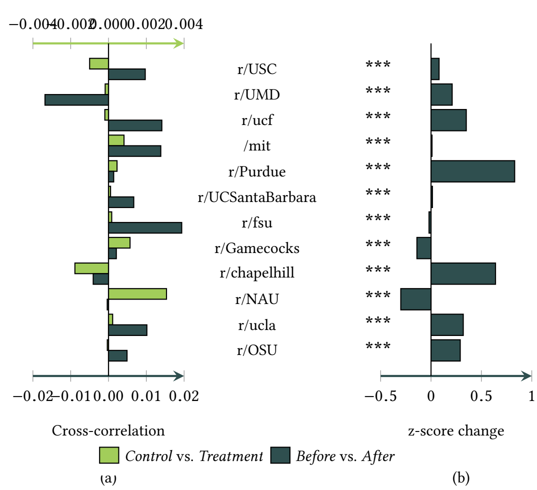


Fig. 6. (a) 0-lag cross-correlation between *Control-Treatment* and *Before-After* (*Treatment*) samples. (b) z-score changes in *After* sample compared to *Before*. p-values computed using Mann-Whitney U-test (*** $p < 0.001$).

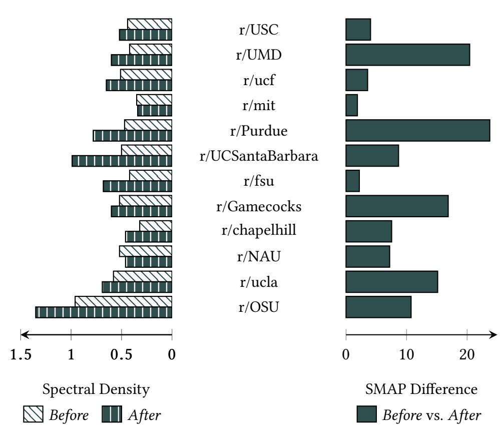


Fig. 7. Spectral and wavelet analysis of frequency waveforms of *Before* and *After* samples in *Treatment* dataset.

present the results of the spectral density and the wavelet analysis in Figure 7. For the former, the changes in mean spectral densities between the *Before* and *After* samples range from -2% (r/mit) to 96% (r/UCSantaBarbara), with a mean absolute difference of 37%. For the latter, we find that the symmetric mean absolute percentage (SMAP) differences between the frequency waveforms of the *Before* and *After* samples averages at 10, ranging between 2 (r/fsu) and 24 (r/Purdue). These results suggest that the periodicity of expression of *High Stress* was disrupted considerably following the incidents of gun violence on the 12 campuses.

Table 4. Welch's t-test comparing the psycholinguistic attributes of *High Stress Treatment* posts shared *Before* and *After* gun violence incidents. Statistical significance reported after Benjamini-Hochberg-Yekutieli False Discovery Rate correction (*** $p < .001$, ** $.001 < p < .01$, * $.01 < p < .05$).

Category	*Before*	*After*	Δ%	t-stat.	p
Affective Attributes					
Anger	0.008	0.010	23.34	3.558	***
Anxiety	0.007	0.003	-61.81	-11.499	***
Negative Affect	0.007	0.009	20.55	3.376	***
Positive Affect	0.072	0.036	-50.56	-27.978	***
Sadness	0.002	0.002	14.42	1.554	*
Swear	0.006	0.007	12.46	1.5765	*
Cognitive Attributes					
Causation	0.027	0.013	-51.95	-23.312	***
Inhibition	0.008	0.005	-36.48	-7.824	***
Negation	0.029	0.041	41.90	13.334	***
Perception					
Feel	0.004	0.006	34.59	3.225	**
Hear	0.014	0.009	-35.49	-7.518	***
Insight	0.041	0.020	-50.24	-26.544	***
Percept	0.017	0.018	4.22	1.137	*
See	0.019	0.018	-7.11	-1.896	*
Interpersonal Focus					
1st P. Plural	0.013	0.010	-24.94	-5.011	***
1st P. Singular	0.061	0.080	32.47	15.864	***
3rd P.	0.015	0.012	-18.78	-3.740	**

Category	*Before*	*After*	Δ%	t-stat.	p
Temporal References					
Future Tense	0.037	0.035	-6.15	-2.146	*
Past Tense	0.056	0.061	8.58	3.787	***
Present Tense	0.116	0.113	-2.20	-1.787	*
Lexical Density and Awareness					
Article	0.117	0.144	22.93	16.720	***
Exclusive	0.032	0.064	99.31	33.659	***
Preposition	0.219	0.181	-17.38	-21.991	***
Quantifier	0.023	0.043	86.01	25.474	***
Biological Concerns					
Bio	0.012	0.014	10.62	2.478	**
Body	0.004	0.005	16.30	2.067	***
Health	0.003	0.007	97.39	8.837	***
Death	0.001	0.003	155.22	6.407	***
Personal and Social Concerns					
Achievement	0.037	0.016	-55.66	-25.509	***
Home	0.005	0.009	93.94	10.148	***
Money	0.022	0.011	-48.75	-16.661	***
Religion	0.003	0.004	43.82	2.872	**
Family	0.002	0.003	41.67	2.606	**
Friends	0.004	0.006	65.55	5.352	***

We note that for r/Gamecocks, which we found to show aberrant pattern compared to other subreddits in the time domain analysis, according to its frequency domain analysis distribution heatmap (Figure 5(h)), there is a significant change in the periodicity of expression of high stress following the gun violence incident in the University of Southern Carolina (14% change in spectral density and an SMAP difference of 17).

5.3 RQ 3: Linguistic Dynamics of Stress

Finally, we present the results of our RQ3: the linguistic dynamics of *High Stress* expression in *Treatment* posts around the 12 campus gun violence incidents.

5.3.1 Psycholinguistic Characterization. In order to characterize the psycholinguistic cues of high stress content shared in the campus subreddits, we extracted the normalized occurrences of the LIWC attribute categories from the *High Stress Treatment* posts shared during the *Before* and *After* periods (ref. Methods). For each psycholinguistic measure, to assess whether the differences between the *Before* and *After* samples are statistically significant, we perform Welch's t-test, followed by Benjamini-Hochberg-Yekutieli False Discovery Rate (FDR) correction. These results are presented in Table 4.

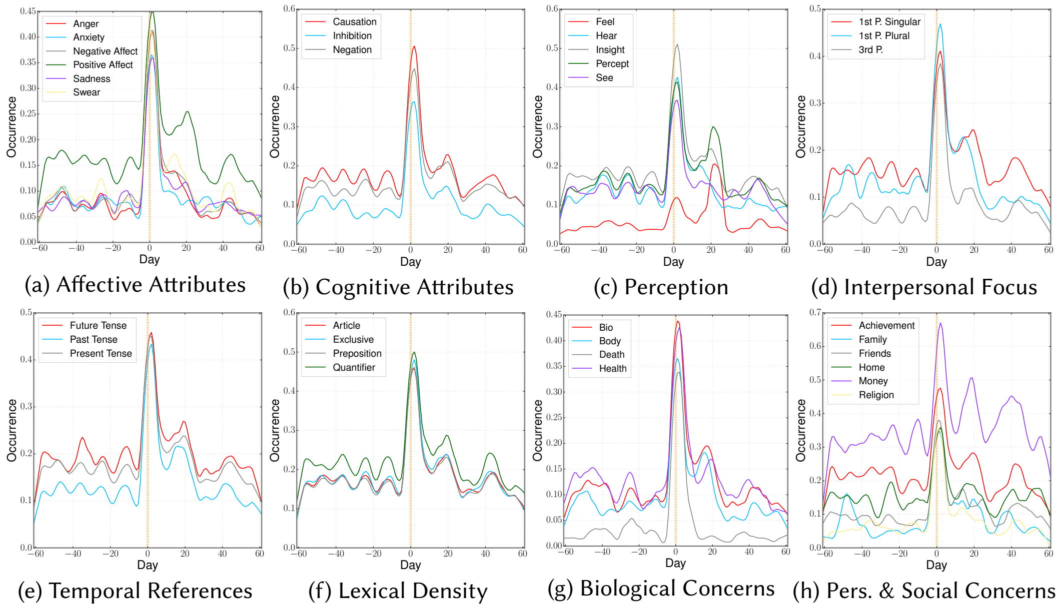
Affective Attributes. Starting with the measures under *Affective Attributes*, we observe that *High Stress* posts in the *After* dataset show higher occurrences of *anger*, *negative affect* and *swear* words. Some example post snippets include, "*why the hell do they have a giant assault rifle?*" and " *I guess since campus is a gun free zone we're all fucked*". At the same time, *High Stress* posts in *After* period show significantly lowered levels of *positive affect* words. This indicates that the students may be engaging over Reddit to express their relatively higher negative perceptions, reactions and thoughts apropos the gun violence incidents.

Cognition and Perception. Next, for the measures grouped under *Cognitive Attributes* and *Perception*, we observe that, words related to *causation*, *inhibition* and *insight* are used significantly lesser in the *After* period. Prior work has related this psycholinguistic expression to lowered cognitive functioning [6], which is a symptom of high stress. However, *negation* words occur more frequently in the *After* period, as well as words related to *feel*. Per prior work [61], this kind of greater perceptual expressiveness is known to be associated with language that depicts personal and first-hand accounts of real world happenings, events and experiences. Likewise, in our case, they indicate that the subreddit users are more expressive of their feelings in the aftermath of the campus gun violence incidents, e.g., "*I'm already home, but can't explain how am I feeling. No idea how to deal with this. Anything normal does feel way out of place at this point.*"

Linguistic Style Attributes. Corresponding to the different linguistic style attributes, the *Before* and *After High Stress* posts show distinctive *Interpersonal Focus*—we find that the use of *1st person singular* pronouns increases by 32% after gun violence, however that of *1st person plural* and *third person* pronoun words decreases. These patterns are known to indicate heightened self-attentional focus and greater detachment from the social realm [61]. We conjecture that the users posting in the college subreddits may be resorting to social media to share their personal experiences and opinions about the incident. In the case of *Temporal References*, words referring to *future* and *present* tense show reduced usage in *After* sample, whereas words belonging to *past* tense demonstrate higher occurrence in the *After* period. Higher use of past tense indicates tendency to recollect prior experiences and events [79], which in our case, might be an orientation to discuss the gun violence incident on the campus. With the exception of *preposition* words, all other function words (*exclusives*, *articles* and *quantifiers*) show significant increase in the *After* period, which are known to be related to a personal narrative writing style, often characteristic of crisis-inflicted populations [18].

Biological Concerns. Considering the measures grouped under *Biological Concerns*, our results show that words referring all of *bio*, *body*, *health* and *death* occur in a higher frequency in the *After* period. We conjecture that the *High Stress* posts shared following the gun violence incidents tend to relate to the after-effects, casualties, and implications of the incident for students' safety, well-being, and life (example post excerpt: "*I hope that no one is seriously injured or killed*").

Personal and Social Concerns. Finally, looking at the measures grouped under *Personal* and *Social* Concerns, we observe some revealing patterns. First, words related to *achievement* occur significantly lower (55%) in the *High Stress* posts spanning the period *After* the gun violence incidents. We note that this category consists of words like 'confidence', 'pride', 'progress', 'determined'. The college subreddits which are generally a platform for college-life discussions among students, lower usage of *achievement* words indicates a decline in tendency of engagements in career and academic topic related discussions in the aftermath of the campus gun related violence. Next, we find that the usage of words relating to social life and relationships: *family*, *friends* and *home* increase *After* the incidents. Through these social orientation words, we conjecture the subreddit users to share social well-being impressions as well as perceptions of solidarity in the context of the incidents. Words in the category *money*, which occur significantly less frequently in the *After* samples, show that although money is a generic contributor of stress [78], it does not remain so in the *High Stress* posts following gun related violence on the campuses. In addition, some of these incidents, such as the UNC Chapel Hill or the OSU attack were violence attributed to religious radicalism or religious hate-crime, which we conjecture to contribute to the higher occurrence of *religion* words in the *After* period.



(a) Affective Attributes (b) Cognitive Attributes (c) Perception (d) Interpersonal Focus

(e) Temporal References (f) Lexical Density (g) Biological Concerns (h) Pers. & Social Concerns

Fig. 8. Temporal variation of statistically significant psycholinguistic attributes of *High Stress Treatment* posts shared *Before* and *After* gun related violence.

Temporal Trends of Psycholinguistic Measures. Finally, we examine the temporal trends of occurrences of the significant psycholinguistic measures in *High Stress* posts: ref. Figure 8. We observe that most of the measures belonging to the different attribute categories consistently attain a peak in their occurrences in the immediate vicinity of the gun violence incident (day: 0). Interestingly, this peak is also the maximum value attained by all of the measures, with an exception of *feel* (Figure 8c).

Comparing among the measures belonging to *Affective Attributes*, we notice that *positive affect* is expressed consistently higher than any other measure in this group. We observe a revealing shift in trends of the usage of *Interpersonal Focus* (Figure 8d) – 1) We observe a sharp increase in the usage of 1st personal plural pronouns, overtaking the usage of 1st person singular ones just immediately following the day of incident, which aligns with the emergence of collective identity as observed in prior work [48]. 2) But with days to follow, we observe the occurrence of 1st person singular takes over, indicating a rise in usage of words referring to self attention. In addition, it is interesting to note the occurrence of *death*, agrees with prior work [33] – where we find that although "death" has a minimal occurrence consistently in the *Before* period, it achieves substantial concentration in *High Stress* posts just following the day of incident for a few days.

5.3.2 Incident-Specific Lexical Analysis. Our final set of results includes an analysis of the linguistic markers as manifested in the subreddits immediately after gun violence on a college campus. For this, within the *High Stress Treatment* posts, we first extract the 30 most occurring n-grams ($n = 1, 2, 3$) on the day of incident. The 30-day *Before* and *After* temporal trends of usage of n-grams is shown in Figure 9 in a heatmap format. We observe some contrasting patterns– for instance, 'class' occurs consistently in *High Stress* posts until the incident dates, but its usage declines considerably in the week following the gun related violence. On the other hand, subreddit users converse about 'people', 'friend', 'hope' and 'feel' a lot more in the immediate aftermath of the event, as compared to their overall occurrences—aligning with the observations drawn from

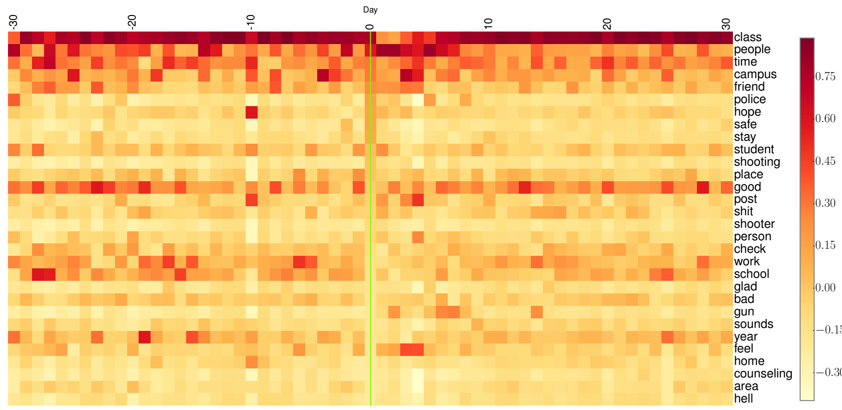


Fig. 9. Top 30 keywords used in *High Stress* posts on the day of gun violence incident ($day = 0$) across all the subreddits.

Table 5. Lexicon of selected n-grams ($n = 1, 2, 3$) occurring considerably higher in posts shared 7 days after the day of gun related violence, as compared to 7 days before.

Subreddit	*After>Before* (*LLR* $\geq$ (0.75))
r/USC	problems, night, security, shooting, party, events, fingerprint, entrances, email, dps, campus center, event, trojan, defense, safe
r/UMD	athletics, gun, supercar, cars, shoot, department, school, fire, community, sports, college, park
r/ucf	assault, assault rifle, weapon, tower 1, rifle, gun, police
r/mit	state, lincoln, stay safe, watertown, officer, officers, police, scanner, second, shots, shots fired, house, bpd, unknown, clear, confirmed, custody, dexter, fired, fuck, spruce, suspect, black, boston
r/Purdue	shooter, police, shooting, news, place, building, ee, campus, guy, heard, day, gt, know, student, people, today
r/UCSanta-Barbara	videos, victims, gun, mental, isla vista, guy, news, community, post, police, person, help, feel, love, iv, life, point, friends
r/fsu	mental, safe, shooting, strozier, ok, news, library, shooter, friends, victims, hope, stay, post, time, people, information, good
r/Game-cocks	alert, murdersuicide, public health, public health research, research center, shooter, shooting, students, support, lockdown, faculty staff, counseling center, building, health research center, cancelled
r/chapelhill	pretty, muslims, writing, religion, high, hicks, help, pound, students, execution style, execution, universal, world, abusalha, 30 serv, support, parking, unc
r/NAU	astronomy, jones, kill, kill people, meth, problem, harder, professors, self, self defense, shooter, shot, tour, guns kill people, year, guns, fight, class, defense, gun, asu, shooting
r/ucla	safe, confirmed, police, klug, shooter, gun, guns, health, mental, saying, professor, situation
r/OSU	safe, police, muslims, gun, removed, parking, post, stay, wrong

the psycholinguistic analysis above, involving the emergence of a social orientation and greater perceptual expression following the incidents. The n-grams describing the nature, manifestation, and implications of the specific campus incidents, e.g., 'police', 'shooting', 'safe' and 'gun', have dense and increased concentration of usage following the day of the incident.

Next, we drill down further for each of the 12 colleges and, employing the Log Likelihood Ratio (LLR) measure (ref. methods), extract a lexicon of the top 50 n-grams ($n = 3$) from the *High Stress* posts within 7 days following the day of gun violence incident, and then compare their occurrence in *High Stress* posts in the 7 days preceding the day of incident. Table 5 reports the n-grams, for which we obtained an LLR of over 0.75– these n-grams occur predominantly in posts after the incident, and are distinctive characteristic of what is discussed in the subreddits specifically on these days.

Taking a close look at Table 5, we observe that this lexicon includes words which embed information specific to the incidents that occurred on the different college campuses under our consideration. For instance, we notice the lexicon to encompass words related to the geographical site of the incident such as – '*campus center*' in r/USC, '*tower 1*' in r/ucf, '*isla vista*' in r/UCSantaBarbara, '*library*' in r/fsu, '*public health*' in r/Gamecocks, and '*parking* in r/OSU. Next, we note the presence of the word '*videos*' in r/UCSantaBarbara and '*murdersuicide*' in r/NAU, which are coherent with how the incidents unfolded at these campuses. Additionally, agreeing with our findings from the psycholinguistic characterization presented above, we observe the presence of '*muslims*' in r/chapelhill and r/OSU, where the incidents were attributed to be religious hate-crimes or radicalism. Finally, we find usage of words relating to the victim or the perpetrator's name and occupation in some of the subreddits, such as r/mit, r/Gamecocks, r/chapelhill and r/ucla. Summarily, this analysis shows that high stress expressed in posts of the college subreddits in the immediate aftermath of the gun related violence may be a consequence of the incidents in the respective campuses.

6 DISCUSSION

From our findings, we derive two major observations: 1) Psychological stress may be automatically inferred from social media content by employing supervised learning approaches; and 2) Inferred stress levels in a college campus may indicate the responses of individuals exposed to the reported gun-related violence incident. To arrive at these findings, we make a methodological contribution in the paper as to how stress changes, temporal and linguistic, can be measured following a violent incident on campus, drawing from machine learning and time series analysis techniques. Thus our work bears implications for researchers intending to study the socio-psychological responses of a population exposed to a crisis, and those interested in developing technologies to assist vulnerable populations following traumatic events. We discuss these implications in the following subsections.

6.1 Theoretical and Psychological Implications

Freud's psychoanalytic theory [31] argued that external reality, for example, traumatic events, can have profound effects on an individual's pysche, and can be considered to be the cause of emotional upheaval, stress, and traumatic neurosis. He suggested that the personal impact of the trauma, the inability to find conscious expressions for it, and the unpreparedness of the individual can cause a breach to the stimulus barrier and overwhelm the defense mechanisms [30]. We find that the methods we employed in this paper allow us to examine these theoretical constructs in a quantitative, data-driven manner. For instance, our linguistic analytical methods suggest distinctive psycholinguistic cues in high stress posts shared after the gun violence incident compared to before. As an example, the usage of words related to biological concerns increase remarkably following the incident. In contrast, more general topics closely related with stress in a college population, such as financial and career-related concerns [68] exhibit significant reductions in usage following the incidents.

Further, a notable finding of our work comes from our incident-specific lexical analysis— the content shared on social media immediately following the violent incidents appears to be largely topically related to the events themselves. McCann and Pearlman [51], within the framework of cognitive theory, proposed seven fundamental psychological need areas following experience of a crisis event: frame of reference, safety, dependency/trust of self and others, power, esteem, intimacy, and independence. Trauma, they argue, may cause disruptions in any of these need areas and thereby lead to troublesome emotions and thoughts such as stress. Words such as "stay safe", "support", "hope", "help", "self", that increase in usage in high stress social media posts following the incidents, indicate the expression of many of these needs.

Finally, our methods allow us to draw a variety of nuances of the patterns of acute stress responses on college campuses following the violent incidents, that tend to offset more persistent chronic stress expressions. For instance, our findings suggest that although students undergo stress all throughout the year because of academic and personal reasons [68], stress expression of a campus changes considerably after an incident of gun related violence. In essence, our findings show that, as revealed by campus social media posts, stress as a construct, is prevalent (possibly chronic in nature) across time, yet the nature of this construct changes drastically (possibly turning more acute) around a critical crisis incident. Further, we demonstrate the temporal and linguistic "signatures" of expression of such acute stress, such as altered periodicities or increase/descrease in specific psycholinguistic words, can be gleaned with our proposed machine learning and time series analysis approaches. Our findings support similar observations made with respect to the manifestation of affect and psychological states in a response to chronic violence [25], wars [48] and terrorist attacks [46]. Moreover, closely aligning with prior work [1], we also observe that the post violence acute stress levels subside within days to follow, and approach baseline levels of generally persistent chronic stress. This could be because life goes back in order, with other aspects of campus life taking priority. This interpretation is consistent with prior work in crisis informatics [48], as well as Foa and colleagues' emotional processing theory [28]. They noted that emotional experiences, such as anxiety and stress, are often relived well after the original traumatic events have occurred, although the frequency and the intensity of emotional reliving usually decreases over time.

6.2 Practical Implications

Our computational techniques provide a robust mechanism to quantify the impacts and severity of a crisis, as well as the corresponding community responses. While our findings may seem unsurprising—that stress exacerbates after violent events, and shows altered periodicities and expression patterns, that the outcomes of our proposed methodologies align with expected patterns of stress following crises, in a way, establishes the validity of our methods. Therefore, we believe our techniques will find use as unobtrusive sensors of stress and its linguistic and temporal changes during crises. We also believe that these methodologies may be leveraged in future situations where causes of stress may not be so apparent or known, as was the case in our study, e.g., assessing stress and associated student responses in everyday (crisis free) contexts, where a variety of day-to-day but unanticipated academic, personal or social concerns may contribute to stress.

Further, since our techniques leverage social media of the specific affected communities (college campuses), they can help identify their unique "signatures" or idiosyncratic patterns in stress expressions. As we observed, the temporal trajectory of high stress following the incident at the University of Southern California was distinct from the others; so were the kinds of linguistic cues that surfaced in social media content of the different campuses immediately following the incidents.

In turn, our approaches may help discover the presence or absence of protective factors surrounding stress in specific communities/campuses, including how a campus's stress expression deviates from an expected pattern of stress on any campus affected by a similar crisis. This information can be immensely valuable to crisis rehabilitation efforts, including how specific campuses may adopt policies or strategies to enhance the idiosyncratic aspects relating to the community, that exacerbate or protect against stress.

Finally, we also note that, it is recognized that the impact of a violent incident transcends observed casualties, and its perception can be very subjective at an individual level. Our work provides a way to account for the "invisible wounds" [37] or "hidden casualties" [64] in a crisis, which tend not to get reported or measured adequately. In essence, we observe that in the aftermath of campus gun related violence, campus-specific social media like Reddit, acts as a unique platform allowing campus populations to express their emotion and stress about their circumstances, (semi) anonymously, amid feelings or perception of fear and trauma. Our techniques enable capturing a "quantitative narrative" of these self-disclosed stress experiences of campus populations exposed to crisis events, which, we believe, can eventually inform historical accounts about campus life.

6.3 Design Implications

Our research bears implications for designing technology that can support improving mental health provisions for campus populations during times of upheavals.

A significant challenge for college administrators is providing adequate mental health services, such as around aggravated stress levels, in a proactive, real-time manner. These efforts become even more difficult in the face of violent events on campus, due to the disruption in everyday life and activities on campus. Our work shows promise in enabling technology-assisted means to tackle these challenges.

Since sudden bursts of stress can be detrimental in the long-term [52], with our work, population-centric stress tracking tools can be built. These tools can significantly advance current practices in terms of how college authorities engage with the student community following crisis incidents. Typically these practices include broad campus-wide communication of the context and outcomes of the incident, followed by specialized programs to direct psychological counseling and rehabilitation support to students who may need help. Our work can complement existing techniques and tools for assessing stress among individuals [58]. With tools that leverage our methods, college authorities can learn about the pervasiveness of stress following a crisis event and the extent to which its normal pattern has been disrupted. This can enable them in making more informed decisions about the nature of crisis communication that should take place on campus, such as balancing informational alerts with adequately sensitive and focused assurance. Additionally, administrators will be able to reach a better understanding of students' counseling or rehabilitation resource needs. They will also be able to identify specific stress induced temporal or linguistic responses that negatively impact specific student groups. This can allow them to take adequate action in a timely manner, e.g., conducting campus-wide awareness and mitigation campaigns on mental wellbeing, or making tailored provisions to improve mental resilience and morale of the student body.

Next, often following campus violence, administrators need to make policy decisions for maintaining the safety and morale of the student body. For example, in the wake of 2012 shooting in University of Southern California (which we considered in our analysis), the administration incorporated heightened security measures and visitor restrictions in the campus[8]. The patterns and observations gleaned from tools that leverage our methods, such as the incident-specific linguistic markers, could provide evidence-based, student-contributed insights to administrators so as to make informed policy decisions to scaffold campus life succeeding crises. Over time, this ability to

[8]latimesblogs.latimes.com Accessed: 2017-04-15

identify markers of student stress and their dynamics can also contribute to improved preparedness in campus around future crisis events.

6.4 Limitations and Future Directions

Our study includes some limitations. Although our results build on expert validations, we cannot be certain about the clinical nature of stress inferred in this data, and we caution against making clinical inferences with our classifier and approach. Additionally, our study employs adequate statistical techniques such as consideration of control data to identify stress dynamics subject to the gun violence events specifically, we cannot make absolute causal claims. An interesting future extension to address this concern can be to collect quantitative evidence of stress in the campus population following a violent incident, such as through psychometric instruments, and examine if our findings predict these measurements.

Relatedly, there is a possibility that our study has not considered unknown confounding variables, which could have affected our results. In fact, it may seem that casualty statistics of the specific gun violence event may have direct impact on the observed stress patterns, and therefore should be factored into our stress assessment model. However, as noted before, our proposed approaches enable us to examine to what extent social media may serve as a source of information capturing those patterns of stress that surface from the "invisible wounds" [37] and "hidden casualties" [64], rather than directly from the reported casualties. Future work could investigate to what extent observed social media stress patterns directly relate to reported casualties. Further, our results only account for a specific nature of crisis event with a stipulated period of temporal variation. We cannot be certain if additional campus related factors that coincided with these events, e.g., policy changes or other unreported crises, could have impacted the observed analytical findings. Considering longer temporal analysis can also provide insights into the resilience of different campuses in the light of violent incidents.

We also acknowledge limitations in the use of campus-specific social media data in modeling stress and in identifying the aftereffects of violent incidents. First and importantly, we recognize that although college students constitute a demographic in which social media penetration is among the highest [6], possibly not every student on a campus uses them. Our approach is therefore not able to account for social media non-use or those who use platforms whose data cannot be collected for research purposes in an ethical manner. We also identify that there could be self-presentation bias in the Reddit expression of some users; this is because stress is a stigmatized experience, and some individuals may not feel comfortable sharing these experiences on a public online platform like Reddit. Augmenting our data with other social media sources (e.g., Twitter) including students' self-reported (qualitative and quantitative) data can circumvent some of these limitations, and constitute promising directions for future work.

6.5 Privacy and Ethics

Privacy challenges within social media research is well-recognized. In our study, we used publicly available social media data and no direct interaction was made with the users who posted on the subreddits. In the case of social media population, it is impracticable to gain informed consent from thousands of people, and therefore individuals may be unaware of the implications of social media content, with regard to their ability to signal underlying psychological risk. More importantly, we ensured that our assessment of the expression of stress is conducted at post-level with anonymized usernames, to protect the identity and safeguard detecting the mental well-being state of user, which would necessitate an individual consent and an approval from institutional review board.

7 CONCLUSION

In this paper, we proposed computational techniques to assess how the psychological stress of a campus population changes following an incident of gun violence. Our data consisted of posts shared on the Reddit communities of 12 US colleges affected by gun related violence between 2012 and 2016. Using social media as a lens to glean students' stress expression, we first built a machine learning classifier for inferring the expression of stress, which achieved a mean accuracy of 82%. Next, through time series analyses we observed that the usual periodicities of stress experienced on campuses get disrupted, and high stress levels are more prevalent following the incidents; although their effects subside a few days later. The linguistic dynamics associated with high stress posts indicated a correlation between incident-specific topics and stress expression within the campuses. We believe our work advances research in psycholinguistics and crisis informatics, through a methodology to measure the expression of stress via social media, and a framework to assess the psychological impacts of a community exposed to a crisis.

REFERENCES

[1] Ban Al-Ani, Gloria Mark, and Bryan Semaan. 2010. Blogging in a region of conflict: supporting transition to recovery. In *CHI*. ACM, 1069–1078.

[2] Jennings Anderson, Marina Kogan, Melissa Bica, Leysia Palen, Kenneth Anderson, Kevin Stowe, Rebecca Morss, Julie Demuth, Heather Lazrus, Olga Wilhelmi, et al. 2016. Far Far Away in Far Rockaway: Responses to Risks and Impacts during Hurricane Sandy through First-Person Social Media Narratives. In *Information Systems for Crisis Response and Management*.

[3] Kelly J Asmussen and John W Creswell. 1995. Campus response to a student gunman. *The Journal of Higher Education* 66, 5 (1995), 575–591.

[4] American Psychiatric Association et al. 2013. *Diagnostic and statistical manual of mental disorders, (DSM-5)*. American Psychiatric Pub.

[5] John W Ayers, Benjamin M Althouse, Eric C Leas, Ted Alcorn, and Mark Dredze. 2016. Can Big Media Data Revolutionarize Gun Violence Prevention? *arXiv preprint arXiv:1611.01148* (2016).

[6] Shrey Bagroy, Ponnurangam Kumaraguru, and Munmun De Choudhury. 2017. A Social Media Based Index of Mental Well-Being in College Campuses. (2017).

[7] Timothy Baldwin, Paul Cook, Marco Lui, Andrew MacKinlay, and Li Wang. 2013. How noisy social media text, how diffrnt social media sources?. In *IJCNLP*. 356–364.

[8] Nuran Bayram and Nazan Bilgel. 2008. The prevalence and socio-demographic correlations of depression, anxiety and stress among a group of university students. *Social psychiatry and psychiatric epidemiology* 43, 8 (2008), 667–672.

[9] Louis-Philippe Beland and Dongwoo Kim. 2016. The effect of high school shootings on schools and student performance. *Educational Evaluation and Policy Analysis* 38, 1 (2016), 113–126.

[10] Adrian Benton, Braden Hancock, Glen Coppersmith, John W Ayers, and Mark Dredze. 2016. After Sandy Hook Elementary: A year in the gun control debate on Twitter. *arXiv preprint arXiv:1610.02060* (2016).

[11] John Briere and Marsha Runtz. 1989. The Trauma Symptom Checklist (TSC-33) early data on a new scale. *Journal of interpersonal violence* 4, 2 (1989), 151–163.

[12] E Oran Brigham, E Oran Brigham, JulioJ Rey Pastor, Rey Pastor, Tom M Tom M Apostol, MargaritaMartínez Rodríguez, Miguel RamónMargarita Rodríguez, Miguel Ramón Martínez, C HenryPENNEY Edwards, DAVID EC Henry Edwards, et al. 1988. *The fast Fourier transform and its applications*. Number 517.443. Prentice Hall,.

[13] Ronald Burns and Charles Crawford. 1999. School shootings, the media, and public fear: Ingredientsfor a moral panic. *Crime, Law and Social Change* 32, 2 (1999), 147–168.

[14] Cornelia Caragea, Nathan McNeese, Anuj Jaiswal, Greg Traylor, Hyun-Woo Kim, Prasenjit Mitra, Dinghao Wu, Andrea H Tapia, Lee Giles, Bernard J Jansen, et al. 2011. Classifying text messages for the haiti earthquake. In *ISCRAM2011*. Citeseer.

[15] Cindy Chung and James W Pennebaker. 2007. The psychological functions of function words. *Social communication* (2007), 343–359.

[16] Sidney Cobb, Stanislav V Kasl, et al. 1977. Termination; the consequences of job loss. In *DHEW (NIOSH) Publication*. Vol. 77. NIOSH.

[17] Sheldon Cohen, Tom Kamarck, and Robin Mermelstein. 1983. A global measure of perceived stress. *Journal of health and social behavior* (1983), 385–396.

[18] Michael A Cohn, Matthias R Mehl, and James W Pennebaker. 2004. Linguistic markers of psychological change surrounding September 11, 2001. *Psychological science* 15, 10 (2004), 687–693.

[19] Philip J Cook and Jens Ludwig. 2000. *Gun violence: The real costs*. Oxford University Press on Demand.

[20] Glen Coppersmith, Craig Harman, and Mark Dredze. 2014. Measuring post traumatic stress disorder in Twitter. In *International Conference on Weblogs and Social Media (ICWSM).*

[21] Hal Daumé III. 2009. Frustratingly easy domain adaptation. *arXiv preprint arXiv:0907.1815* (2009).

[22] Munmun De Choudhury, Scott Counts, and Eric Horvitz. 2013. Social media as a measurement tool of depression in populations. In *Proceedings of the 5th Annual ACM Web Science Conference.* ACM, 47–56.

[23] Munmun De Choudhury and Sushovan De. 2014. Mental Health Discourse on reddit: Self-disclosure, Social Support, and Anonymity. In *International Conference on Weblogs and Social Media (ICWSM).*

[24] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. Predicting depression via social media. In *ICWSM.*

[25] Munmun De Choudhury, Andres Monroy-Hernandez, and Gloria Mark. 2014. Narco emotions: affect and desensitization in social media during the mexican drug war. In *CHI.* ACM, 3563–3572.

[26] Sean M Fitzhugh, C Ben Gibson, Emma S Spiro, and Carter T Butts. 2016. Spatio-temporal filtering techniques for the detection of disaster-related communication. *Social Science Research* 59 (2016), 137–154.

[27] Daniel J Flannery, Kelly L Wester, and Mark I Singer. 2004. Impact of exposure to violence in school on child and adolescent mental health and behavior. *Journal of community psychology* 32, 5 (2004), 559–573.

[28] Edna B Foa, Gail Steketee, and Barbara Olasov Rothbaum. 1989. Behavioral/cognitive conceptualizations of post-traumatic stress disorder. *Behavior therapy* 20, 2 (1989), 155–176.

[29] Kirsten A Foot and Steven M Schneider. 2004. Online structure for civic engagement in the September 11 web sphere. *Electronic J of Communication* 14 (2004), 3–4.

[30] Sigmund Freud. 1977. *Inhibitions, symptoms and anxiety.* WW Norton & Company.

[31] Sigmund Freud. 2003. *Beyond the pleasure principle.* Penguin UK.

[32] Martin Gjoreski, Hristijan Gjoreski, Mitja Lutrek, and Matja Gams. 2015. Automatic detection of perceived stress in campus students using smartphones. In *Intelligent Environments (IE), 2015 International Conference on.* IEEE, 132–135.

[33] Kimberly Glasgow, Clayton Fink, and Jordan L Boyd-Graber. 2014. " Our Grief is Unspeakable": Automatically Measuring the Community Impact of a Tragedy.. In *ICWSM.*

[34] IR Gochman. 1979. Arousal, attribution, and environmental stress. *Stress and anxiety* 6 (1979), 67–92.

[35] Scott A Golder and Michael W Macy. 2011. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science* 333, 6051 (2011), 1878–1881.

[36] Paul L Hewitt, Gordon L Flett, and Shawn W Mosher. 1992. The Perceived Stress Scale: Factor structure and relation to depression symptoms in a psychiatric sample. *Journal of Psychopathology and Behavioral Assessment* 14, 3 (1992), 247–257.

[37] Terra C Holdeman. 2009. Invisible wounds of war: Psychological and cognitive injuries, their consequences, and services to assist recovery. *Psychiatric Services* 60, 2 (2009), 273–273.

[38] Paul W Holland. 1986. Statistics and causal inference. *Journal of the American statistical Association* 81, 396 (1986), 945–960.

[39] Andrew Hoskins and Ben O'loughlin. 2007. *Television and terror: Conflicting times and the crisis of news discourse.* Springer.

[40] Yun Huang, Ying Tang, and Yang Wang. 2015. Emotion map: A location-based mobile social system for improving emotion awareness and regulation. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing.* ACM, 130–142.

[41] Hallam Hurt, Elsa Malmud, Nancy L Brodsky, and Joan Giannetta. 2001. Exposure to violence: Psychological and academic correlates in child witnesses. *Archives of pediatrics & adolescent medicine* 155, 12 (2001), 1351–1356.

[42] Muhammad Imran, Shady Mamoon Elbassuoni, Carlos Castillo, Fernando Diaz, and Patrick Meier. 2013. Extracting information nuggets from disaster-related messages in social media. *ISCRAM* (2013).

[43] Mrinal Kumar, Mark Dredze, Glen Coppersmith, and Munmun De Choudhury. 2015. Detecting changes in suicide content manifested in social media following celebrity suicides. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media.* ACM, 85–94.

[44] Jeff Lewis. 2005. *Language wars: The role of media and culture in global terror and political violence.* Pluto Press.

[45] Huijie Lin, Jia Jia, Quan Guo, Yuanyuan Xue, Qi Li, Jie Huang, Lianhong Cai, and Ling Feng. 2014. User-level psychological stress detection from social media using deep neural network. In *Proceedings of the 22nd ACM international conference on Multimedia.* ACM, 507–516.

[46] Yu-Ru Lin and Drew Margolin. 2014. The ripple of fear, sympathy and solidarity during the Boston bombings. *EPJ Data Science* 3, 1 (2014), 31.

[47] Adriana M Manago, Tamara Taylor, and Patricia M Greenfield. 2012. Me and my 400 friends: the anatomy of college students' Facebook networks, their communication patterns, and well-being. *Developmental psychology* 48, 2 (2012), 369.

[48] Gloria Mark, Mossaab Bagdouri, Leysia Palen, James Martin, Ban Al-Ani, and Kenneth Anderson. 2012. Blogs as a collective war diary. In *CSCW.* ACM, 37–46.

[49] Gloria Mark, Melissa Niiya, Stephanie Reich, et al. 2016. Sleep Debt in Student Life: Online Attention Focus, Facebook, and Mood. In *CHI*. ACM, 5517–5528.
[50] Pedro Martinez and John E Richters. 1993. The NIMH community violence project: II. Children's distress symptoms associated with violence exposure. *Psychiatry* 56, 1 (1993), 22–35.
[51] I Lisa McCann and Laurie Anne Pearlman. 1990. *Psychological trauma and the adult survivor: Theory, therapy, and transformation.* Number 21. Psychology Press.
[52] Alexander C McFARLANE. 2010. The long-term costs of traumatic stress: intertwined physical and psychological consequences. *World Psychiatry* 9, 1 (2010), 3–10.
[53] Matthias R Mehl and James W Pennebaker. 2003. The social dynamics of a cultural upheaval social interactions surrounding september 11, 2001. *Psychological Science* 14, 6 (2003), 579–585.
[54] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* (2013).
[55] Megan A Moreno, Lauren A Jelenchick, Katie G Egan, Elizabeth Cox, Henry Young, Kerry E Gannon, and Tara Becker. 2011. Feeling bad on Facebook: depression disclosures by college students on a social networking site. *Depression and anxiety* 28, 6 (2011), 447–455.
[56] Ray Munni and Prahbhjot Malhi. 2006. Adolescent violence exposure, gender issues and impact. *Indian pediatrics* 43, 7 (2006), 607.
[57] Alexandra Olteanu, Sarah Vieweg, and Carlos Castillo. 2015. What to expect when the unexpected happens: Social media communications across crises. In *CSCW*. ACM, 994–1009.
[58] Helga Þórarinsdóttir, Lars Vedel Kessing, and Maria Faurholt-Jepsen. 2017. Smartphone-Based Self-Assessment of Stress in Healthy Adult Individuals: A Systematic Review. *Journal of medical Internet research* 19, 2 (2017).
[59] Leysia Palen. 2008. Online social media in crisis events. *Educause Quarterly* 31, 3 (2008), 76–78.
[60] Ellie Pavlick and Chris Callison-Burch. 2016. The gun violence database. *arXiv preprint arXiv:1610.01670* (2016).
[61] James W Pennebaker, Matthias R Mehl, and Kate G Niederhoffer. 2003. Psychological aspects of natural language use: Our words, our selves. *Annual review of psychology* 54, 1 (2003), 547–577.
[62] Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global Vectors for Word Representation.. In *EMNLP*, Vol. 14. 1532–1543.
[63] Mark Petticrew, Steven Cummins, Catherine Ferrell, Anne Findlay, Cassie Higgins, Caroline Hoy, Adrian Kearns, and Leigh Sparks. 2005. Natural experiments: an underused tool for public health? *Public health* 119, 9 (2005), 751–757.
[64] Deborah Prothrow-Stith and Sher Quaday. 1995. Hidden Casualties: The Relationship between Violence and Learning.. In *Streamlined Seminar*, Vol. 14. ERIC, n2.
[65] Hemant Purohit, Andrew Hampton, Shreyansh Bhatt, Valerie L Shalin, Amit P Sheth, and John M Flach. 2014. Identifying seekers and suppliers in social media communities to support crisis coordination. *CSCW* 23, 4-6 (2014), 513–545.
[66] Yan Qu, Chen Huang, Pengyi Zhang, and Jun Zhang. 2011. Microblogging after a major disaster in China: a case study of the 2010 Yushu earthquake. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*. ACM, 25–34.
[67] Paul R Rosenbaum and Donald B Rubin. 1985. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician* 39, 1 (1985), 33–38.
[68] Shannon E Ross, Bradley C Niebling, and Teresa M Heckert. 1999. Sources of stress among college students. *Social psychology* 61, 5 (1999), 841–846.
[69] Koustuv Saha, Larry Chan, Kaya de Barbaro, Gregory D Abowd, and Munmun De Choudhury. 2017. Inferring Mood Instability on Social Media by Leveraging Ecological Momentary Assessments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (Ubicomp)* 1, 3 (95) (2017), 27.
[70] Wilmar B Schaufeli and Maria CW Peeters. 2000. Job stress and burnout among correctional officers: A literature review. *International Journal of Stress Management* 7, 1 (2000), 19–48.
[71] Jaclyn Schildkraut, H Jaymi Elsass, and Mark C Stafford. 2015. Could it happen here? Moral panic, school shootings, and fear of crime among college students. *Crime, Law and Social Change* 63, 1-2 (2015), 91–110.
[72] Neil Schneiderman, Gail Ironson, and Scott D Siegel. 2005. Stress and health: psychological, behavioral, and biological determinants. *Annu. Rev. Clin. Psychol.* 1 (2005), 607–628.
[73] Hans Selye. 1956. The stress of life. (1956).
[74] Mark I Singer, Trina Menden Anglin, Li yu Song, and Lisa Lunghofer. 1995. Adolescents' exposure to violence and associated symptoms of psychological trauma. *Jama* 273, 6 (1995), 477–482.
[75] Karen Slovak. 2002. Gun violence and children: Factors related to exposure and trauma. *Health & Social Work* 27, 2 (2002), 104–112.
[76] Emma S Spiro. 2016. Research opportunities at the intersection of social media and survey data. *Current Opinion in Psychology* 9 (2016), 67–71.
[77] Kate Starbird, Leysia Palen, Amanda L Hughes, and Sarah Vieweg. 2010. Chatter on the red: what hazards threat reveals about the social life of microblogged information. In *CSCW*. ACM, 241–250.

[78] Thomas Li-Ping Tang and Pamela R Gilbert. 1995. Attitudes toward money as related to intrinsic and extrinsic job satisfaction, stress and work-related attitudes. *Personality and Individual Differences* 19, 3 (1995), 327–332.

[79] Yla R Tausczik and James W Pennebaker. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology* 29, 1 (2010), 24–54.

[80] Theodre Thompson and Carol Rippey Massat. 2005. Experiences of violence, post-traumatic stress, academic achievement and behavior problems of urban African-American children. *Child and Adolescent Social Work Journal* 22, 5-6 (2005), 367–393.

[81] Roger Tourangeau, Lance J Rips, and Kenneth Rasinski. 2000. *The psychology of survey response.* Cambridge University Press.

[82] Jose Manuel Delgado Valdes, Jacob Eisenstein, and Munmun De Choudhury. 2015. Psychological effects of urban crime gleaned from social media. In *Ninth International AAAI Conference on Web and Social Media.*

[83] Patti M Valkenburg, Jochen Peter, and Alexander P Schouten. 2006. Friend networking sites and their relationship to adolescents' well-being and social self-esteem. *CyberPsychology & Behavior* 9, 5 (2006), 584–590.

[84] Shari R Veil, Tara Buehner, and Michael J Palenchar. 2011. A work-in-process literature review: Incorporating social media in risk and crisis communication. *Journal of contingencies and crisis management* 19, 2 (2011), 110–122.

[85] Sudha Verma, Sarah Vieweg, William J Corvey, Leysia Palen, James H Martin, Martha Palmer, Aaron Schram, and Kenneth Mark Anderson. 2011. Natural Language Processing to the Rescue? Extracting" Situational Awareness" Tweets During Mass Emergency.. In *ICWSM.* Citeseer.

[86] Sarah Vieweg, Amanda L Hughes, Kate Starbird, and Leysia Palen. 2010. Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In *CHI.* ACM, 1079–1088.

[87] Michail Vlachos, Philip Yu, and Vittorio Castelli. 2005. On periodicity detection and structural periodic similarity. In *Proceedings of the 2005 SIAM International Conference on Data Mining.* SIAM, 449–460.

[88] Rui Wang, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T Campbell. 2014. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Ubicomp.* ACM, 3–14.

[89] Peter Welch. 1967. The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on audio and electroacoustics* 15, 2 (1967), 70–73.

[90] Shelley Wigley and Maria Fontenot. 2010. Crisis managers losing control of the message: A pilot study of the Virginia Tech shooting. *Public Relations Review* 36, 2 (2010), 187–189.

[91] Jenifer Wood, David W Foy, Christopher Layne, Robert Pynoos, and C Boyd James. 2002. An examination of the relationships between violence exposure, posttraumatic stress symptomatology, and delinquent activity: An "ecopathological" model of delinquent behavior among incarcerated adolescents. *Journal of Aggression, Maltreatment & Trauma* 6, 1 (2002), 127–147.

[92] Anna Zajacova, Scott M Lynch, and Thomas J Espenshade. 2005. Self-efficacy, stress, and academic success in college. *Research in higher education* 46, 6 (2005), 677–706.

[93] Liang Zhao, Jia Jia, and Ling Feng. 2015. Teenagers' stress detection based on time-sensitive micro-blog comment/response actions. In *IFIP International Conference on Artificial Intelligence in Theory and Practice.* Springer, 26–36.

Received June 2017; revised August 2017; accepted November 2017