Assessing the Mental Health of College Students by Leveraging Social Media Data

The mental health of college students is a growing concern and gauging the mental health needs of this group is difficult to assess in real-time and in scale. The ubiquity and widespread use of social media, particularly among young adults, provides opportunities for various stakeholders to proactively assess the mental health of college students and provide timely and tailored support.

By Koustuv Saha and Munmun De Choudhury

DOI: 10.1145/3481834

Mental health on college campuses is a matter of growing concern as an increasing number of college students show rising levels of anxiety, depression, and suicidal ideation. According to the 2019 National College Health Assessment, 16.7 percent of students felt too depressed to function in the last two weeks from when the survey was conducted, while 8.6 percent seriously considered suicide or tried to harm themselves in the past 12 months [1]. The 2018 Association for University College Counseling Center Directors (AUCCCD) survey revealed the most frequent concerns for college counseling centers around the world are anxiety (58.9 percent), depression (48.0 percent), stress (46.9 percent), relationship issues (29.5 percent), family concerns (29.0 percent), suicidal thoughts (28.4 percent), academic performance difficulties (28.2 percent), sleep disturbance (19.1 percent), social isolation or loneliness (18.5 percent), significant previous mental health treatment history (16.5 percent), and adjustment to a new environment (15.8 percent) [2]. In addition, mental health concerns are largely underreported and the actual proportion of college students' mental health concerns are likely much higher given the stigma surrounding mental health. Studies by Eisenberg and colleagues revealed the major impediments of seeking mental health care facilities on college campuses, and how these services are largely under-utilized [3].

Mental health services on college campuses, including on-campus counseling centers and psychiatric clinics, continuously struggle to address the increasing demands of mental health consultations in a timely fashion. These services often lack resources, staff, and preparedness, leading to long wait lists and selective/infrequent consultations for many [4]. This understates a need to meet the rise in demand for mental health services with adequate and accessible resources. However, currently, campus mental health services do not have adequate means to assess the evolving nature of demand or needs. While periodic surveys of students' mental health pro-

Image by MJ Graphics / Shutterstock.com

vide some barometer of mental health incidence, in terms of medication use, daily lifestyle, suicidal thoughts, depression symptoms, as well as potentially contributing academic, environmental, personal, and social factors [5], they are accurate only in snapshots and are prone to retrospective and susceptible to biases [6]. Besides, crises on college campuses can cause acute stressful experiences and can exacerbate into long-term negative consequences, such as post-traumatic stress disorder, acute stress disorder, borderline personality disorder, or adjustment disorder [7].

It is practically unsustainable to administer surveys in real-time, and when administered after the occurrences of incidents, survey data do not always capture the time-sensitive nature of students' mental health needs and demands. Surveys are also difficult to scale and are limited to smaller study groups. With an increasing gap in the supply of mental health resources and their growing demand, college campuses need to find alternative means to gauge and forecast the demand for counseling services in order to cater to everyone who needs them.

To overcome such limitations, passive sources of data have recently been explored, which provide dense and longitudinal data at scale [8]. Given

the ubiquity and widespread use of social media, especially among the college student demographic, social media data has also been leveraged as a "passive sensor" that can act as a complementary source of unobtrusive, real-time, and naturalistic data to infer wellbeing [9]. Social media data is low-cost, large-scale, non-intrusive to collect, and has the potential to comprehensively reveal naturalistic patterns of mood, behavior, cognition, psychological states, and social milieu, both in real-time and across longitudinal time for individuals and collectives [10]. Social media language consists of an individual's personal and social discourse about day-to-day concerns, and effectively reflects their health and psychosocial wellbeing in a variety of states and contexts [11, 12]. Linguistic cues and social interactions on social media platforms have therefore, enabled researchers to study psychopathologies including depression, anxiety, stress, and loneliness [13, 14, 15, 16]. Relatedly, Bagroy et al. leveraged social media data on Reddit to develop a mental health index per college campus [17].

PREDICTING THE NEED FOR ON-CAMPUS MENTAL HEALTH SERVICES

This article briefly discusses our research in leveraging social media in understanding the mental health of college students. In particular, we discuss two studies, which are in the context of crises on college campuses: one around gun violence incidents on college campuses, and another around public service announcements (PSAs) of counseling recommendations sent out after student deaths on college campuses. Then, we discuss establishing the construct validity of social media as a passive source to understand the mental wellbeing of college students.

These studies used Reddit as a source of college student social media data. Reddit is one of the most popular social media platforms, which caters to the age group between 18–29 years, as 65 percent of Reddit users are young adults [18]. This age demographic aligns well with the typical college student population, making Reddit a suitable choice for studying college communities. Reddit is a social discussion website consisting of diverse communities known as "subreddits" that offer demographic, topical, or interest-specific discussion boards. Many colleges have a dedicated subreddit community, which provides a common portal for the students on campus to share and discuss a variety of issues related to their personal, social, and academic life. The college subreddits name themselves after the college communities, which they rep-

Table 1. Top 16 (U.S.) college subreddits by member count (June 2018). These subreddits customize their icons and name their members with college nicknames and mascots.

Subreddit	Self-Description	#Members	Icon
r/UIUC	This subreddit is for anyone/anything related to UIUC	18,900 Illini	
r/berkeley	GO BEARS!	14,280 bears	
r/aggies	Anything Texas A&M community related!	13,477 Ags	
r/gatech	A subreddit for my dear Georgia Tech Yellow Jackets.	13,295 readers	
r/UTAustin	Welcome to The University of Texas at Austin	13,101 Longhorns	
r/OSU	The Ohio State University	13,348 4-string QBs	
r/ucf	Reddit for University of Central Florida	13,348 Knights	
r/UCSD	Members associated with the UC San Diego.	9,680 Tritons	
r/rutgers	For news relevant to Rutgers University.	9,654 Scarlet Knights	
r/VirginiaTech	A reddit for Hokies	9,605Hokies	
r/Purdue	Purdue University's subreddit.	9,602 Boilermakers	
r/rit	Rochester Institute of Technology official subreddit.	9,304 RITedditors	
r/UMD	The official subreddit of the University of Maryland.	9,172 ReddiTerps	
r/uofm	University of Michigan subreddit	8,756 wolverines	
r/ucla	A place for UCLA students, faculty, and fans! Go Bruins!	8,912 Bruins	
r/ASU	Subreddit for Arizona State University	8,316 Sun Devils	

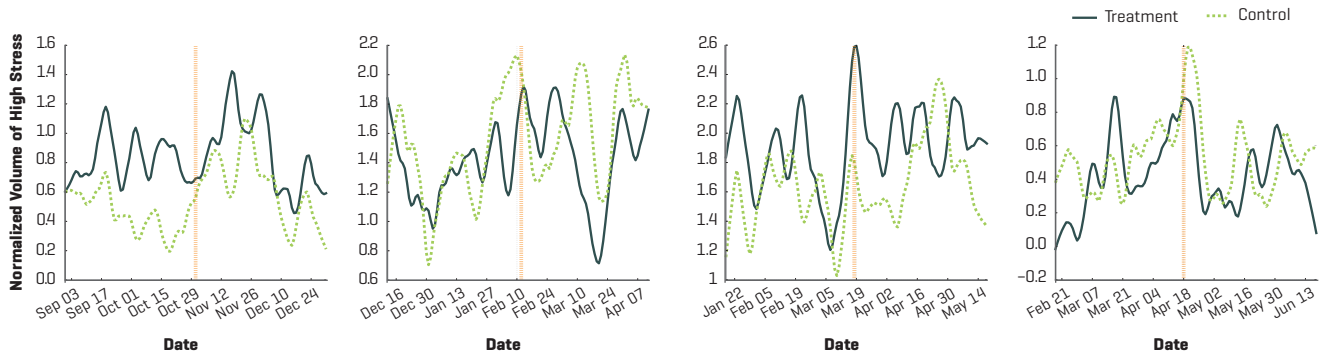
Table 2. Comparing the mean percentage difference between "before" and "after" periods around counseling recommendations after student deaths on college campuses in the Treatment (Tr) and Control (Ct) groups.

Bar lengths represent relative and numbers denote absolute magnitudes. Blank entries convey no statistical significance.

Measure	% Changes in Treatment	% Changes in Control
Affective Changes		
Grief: Activation	15	–1
Grief: Valence	9	–1
Behavioral Changes		
Activity	–	–
Interaction Diversity	9	8
Interactivity	29	–1
Cognitive Changes		
Readability	14	11
Complexity	1.3	0.7
Repeatability	–3	9
Linguistic Style	481	92
Cognition and Perception	457	70
Social Context	382	49

Figure 1. Temporal variation of stress as measured on social media data of college campuses where gun violence events happened.

The orange line in the center shows the date of the on-campus gun violence event (Treatment: year of gun violence event, Control: the same period from the previous year).



resent, and often customize their pages with college logos and campus images to signal their identity. The subreddit pages use personalized Reddit icons and member names based on college nicknames and mascots. Table 1 shows examples of college subreddits, along with self-descriptions, number of members (and moderators), and personalized titles and icons.

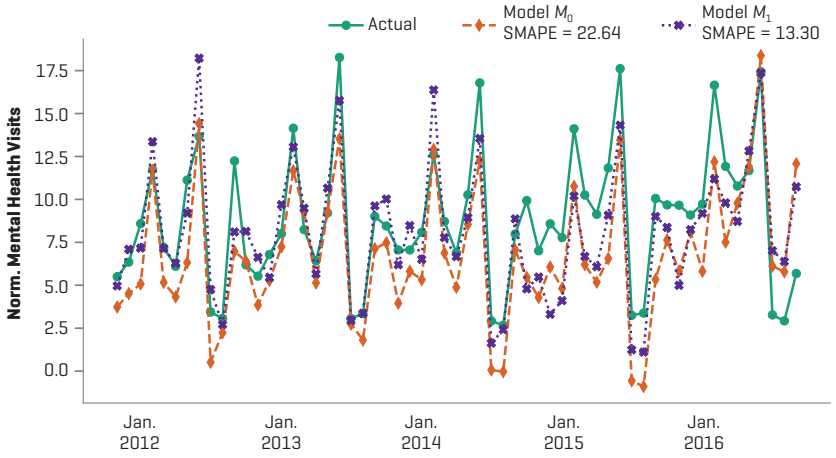
In 2017 we examined how college student stress evolves around gun violence incidents on college campuses [19]. We built a machine learning classifier of stressful expressions in social media language. Next, focusing on 12 incidents of campus gun violence between 2012 and 2016, and social media data gathered from college subreddits, this study reveals amplified stress levels following these incidents, which deviate from usual stress patterns on college campuses (see Figure 1). We examined the linguistic changes around these gun violence incidents to find distinctive characteristics such as decreased cognition and academic career-related conversations but increased self-attention, social orientation, death and family-related conversations, and the emergence of collective identity and solidarity.

The following year, we examined the effectiveness of post-crisis intervention measures in the form of counseling recommendations after student deaths on college campuses by adopting a causal inference framework on social media data [20]. We employ statistical modeling and natural language analysis to measure the psychosocial shifts in behavioral, cognitive, and affective expression of grief in individuals who are "exposed" to the counseling recommendations, compared to that in a matched control cohort (see Table 2). Drawing on crisis and psychology research, the findings suggest individuals exposed to counseling recommendations show greater grief, psycholinguistic, and social expressiveness, providing evidence of healing response to the crisis and thereby positive psychological effects of the counseling recommendations.

Finally, in our most recent work, we conducted an observational study on the data of a large U.S. public university [21]. We collected more than 66,000 posts by approximately 18,000 users from the university's subreddit, and we examined if social media-based mental health assessments corresponded with the ground-truth data of mental health consultations as obtained from the university's health center over a period of over five years between 2011 and 2016. We adopted machine learning and natural language analysis methodologies to measure symptomatic mental health

Figure 2. SARIMA models to predict mental health consultations on college campuses.

Model M_0 only uses historical consultations data, and Model M_1 also additionally incorporates symptomatic mental health expressions measured on social media discussions on college subreddits.



College campuses need to find alternative means to gauge and forecast the demand for counseling services in order to cater to everyone who needs them.

expressions of depression, anxiety, stress, suicidal ideation, and psychosis on the social media data [16]. We then compared the prevalence of these expressions with the ground truth using seasonal autoregressive integrated moving average (SARIMA) models of forecasting on-campus mental health consultations (See Figure 2). We found mental health expressions on social media show a statistically significant relationship with ground-truth college mental health data. Further, incorporating social media data leads to time series predictions with r=0.86 and $SMAPE$=13.30, which outperforms forecasting models without social media data by 41 percent. Our language analysis revealed social media discussions during high mental health utilization months consisted of discussions related to academics, career, and other stressful events on campus, whereas months of low mental health consultations saliently corresponded to expressions of positive affect, collective identity, socialization, and better mental wellbeing.

Together, these studies show that social media interactions of college students can help predict ground-truth data of on-campus mental health consultations. Such contextual and timely information about the community pulse can help college stakeholders including administrators, policymakers, and wellbeing councils to gauge the needs of the students and accordingly ensure that adequate resources are available to meet the demands of mental health-related services. Additionally, campus stakeholders can hold awareness and support initiatives as a part of their campus mental health response to help students during their college stay. These works bear implications in designing proactive and tailored support and crisis rehabilitation efforts on college campuses and making resource allocation more evidence based.

References

[1] American College Health Association. *American College Health Association-National College Health Assessment II: Reference Group Executive Summary Spring 2019*. American College Health Association, Silver Spring, MD, 2019.

[2] LeViness, P. et al. *The Association for University and College Counseling Center Directors Annual Survey 2018*. Association for University and College Counseling Center Directors, Indianapolis, 2018; https://www.aucccd.org/assets/documents/Survey/2018%20aucccd%20survey-public-revised.pdf

[3] Eisenberg, D., Downs, M. F., Golberstein, E., and Zivin, K. Stigma and help seeking for mental health among college students. *Medical Care Research and Review* 66, 5 (2009), 522–541.

[4] Gallagher, R. P. *National Survey of College Counseling Centers 2014*. Project Report. The International Association of Counseling Services (IACS), Alexandria, VA, 2015; http://d-scholarship.pitt.edu/28178/1/survey_2014.pdf

[5] Bayram, N. and Bilgel, N. 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.

[6] Tourangeau, R., Rips, L. J., and Rasinski, K. *The Psychology of Survey Response*. Cambridge University Press, Cambridge, 2000.

[7] Wood, J., Foy, D. W., Layne, C., Pynoos, R., and James, C. B. 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.

[8] Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S., Harari, G., Tignor, S., Zhou, X., Ben-Zeev, D., and Campbell, A. T. StudentLife: Assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, New York, 2014, 3–14.

[9] Saha, K., Chan, L., De Barbaro, K., Abowd, G. D., and De Choudhury, M. Inferring mood instability on social media by leveraging ecological momentary assessments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1, 3 (2017), 1–27.

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

[11] Jaidka, K., Giorgi, S., Schwartz, H. A., Kern, M. L., Ungar, L. H., and Eichstaedt, J. C. Estimating geographic subjective well-being from Twitter: A comparison of dictionary and data-driven language methods. *Proceedings of the National Academy of Sciences* 117, 19(2020), 10165–10171.

[12] Boyd, R. L. and Pennebaker, J. W. A way with words: Using language for psychological science in the modern era. In *Consumer Psychology in a Social Media World*. Routledge, 2015, 250-264.

[13] De Choudhury, M., Gamon, M., Counts, S., and Horvitz, E. Predicting depression via social media. *Proceedings of the International AAAI Conference on Web and Social Media* 7, 1 (2021), 128–137.

[14] Coppersmith, G., Dredze, M., and Harman, C. Quantifying mental health signals in Twitter. In *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*. Association for Computational Linguistics, Baltimore, 2014, 51–60.

[15] Guntuku, S. C., Buffone, A., Jaidka, K., Eichstaedt, J. C., and Ungar, L. H. Understanding and measuring psychological stress using social media. *Proceedings of the International AAAI Conference on Web and Social Media* 13, 1 (2019), 214–225.

[16] Saha, K., Sugar, B., Torous, J., Abrahao, B., Kıcıman, E., and De Choudhury, M. A social media study on the effects of psychiatric medication use. *Proceedings of the International AAAI Conference on Web and Social Media* 13, 1 (2019), 440–451.

[17] Bagroy, S., Kumaraguru, P., and De Choudhury, M. A social media based index of mental well-being in college campuses. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, New York, 2017,1634–1646.

[18] Pew Research Center. Social Media Fact Sheet. April 7, 2017; https://www.pewresearch.org/internet/fact-sheet/social-media/

[19] Saha, K. and De Choudhury, M. Modeling stress with social media around incidents of gun violence on college campuses. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* 1, CSCW (2017), 1–27.

[20] Saha, K., Weber, I., and De Choudhury, M. A social media based examination of the effects of counseling recommendations after student deaths on college campuses. *Proceedings of the International AAAI Conference on Web and Social Media* 12, 1 (2018).

[21] Saha, K., Yousuf, A., Boyd, R., Pennebaker, J. W., and De Choudhury, M. Mental health consultations on college campuses: Examining the predictive ability of social media. Available at SSRN 3774189. 2021; http://dx.doi.org/10.2139/ssrn.3774189

Biographies

Koustuv Saha recently completed his Ph.D. in computer science from Georgia Tech. His research interest is in social computing and computational social science. In his research, he adopts machine learning, natural language, and causal inference analysis to examine human behavior and wellbeing using social media and online data, along with complementary multimodal sensing data. He is joining as a senior researcher at Microsoft Research, Montreal. He has been recognized as Foley Scholar, a recipient of the Foley Scholarship Award, GVU Center's highest recognition for student excellence in research contributions to computing. He is a recipient of the Snap Research Fellowship, a finalist of the Symantec Graduate Fellowship, and his research has won the Outstanding Study Design Award at ICWSM 2019. His research has been covered at prestigious media outlets, including the *New York Times*, CBC Radio, NBC, 11Alive, the Hill, and the *Commonwealth Times*.

Munmun De Choudhury is an associate professor of interactive computing at Georgia Tech. Dr. De Choudhury is best known for laying the foundation of a line of research that develops computational techniques to responsibly and ethically employ social media in understanding and improving our mental health. To do this work, she adopts a highly interdisciplinary approach, combining social computing, machine learning, and natural language analysis with insights and theories from the social, behavioral, and health sciences. Dr. De Choudhury has been recognized with the 2021 ACM-W Rising Star Award, 2019 Complex Systems Society–Junior Scientific Award, over a dozen best paper and honorable mention awards from the ACM and AAAI, and extensive coverage in popular press like the *New York Times*, the NPR, and the BBC. Earlier, Dr. De Choudhury was a faculty associate with the Berkman Klein Center for Internet and Society at Harvard, a postdoc at Microsoft Research, and obtained her Ph.D. in computer science from Arizona State University.