CS 6474/CS 4803
Social Computing:
Computational Methods – Part 2

Munmun De Choudhury

munmund@gatech.edu

Week 6 | February 14, 2024

- Quantitative Approaches
  - Supervised learning
  - Unsupervised learning
- Psychometric Approaches
- Qualitative Approaches
- Cross-disciplinary Approaches
- Multi-method Approaches

- Quantitative Approaches
  - Supervised learning
  - Unsupervised learning
- Psychometric Approaches
- Qualitative Approaches
- Cross-disciplinary Approaches
- Multi-method Approaches

Evaluation Approaches for Supervised Learning Settings

- Quantitative metrics
  - Accuracy, precision, recall, and other performance metrics.
- Use of metrics
  - How to interpret and apply metrics for evaluating model performance.
- Beyond metrics

You Don't Know How I Feel: Insider-Outsider Perspective Gaps in Cyberbullying Risk Detection

Seunghyun Kim,[1] Afsaneh Razi,[2] Gianluca Stringhini,[3] Pamela Wisniewski,[2] Munmun De Choudhury[1]
[1]Georgia Institute of Technology, [2]University of Central Florida, [3]Boston University
seunghyun.kim@gatech.edu, afsaneh.razi@knights.ucf.edu,
gian@bu.edu, pamwis@ucf.edu, munmund@gatech.edu

Abstract

Cyberbullying is a prevalent concern within social computing research that has led to the development of several supervised machine learning (ML) algorithms for automated risk detection. A critical aspect of ML algorithm development is how to establish ground truth that is representative of the phenomenon of interest in the real world. Often, ground truth is determined by third-party annotators (i.e., "outsiders") who are removed from the situational context of the interaction; therefore, they cannot fully understand the perspective of the individuals involved (i.e., "insiders"). To understand the extent of this problem, we compare "outsider" versus "insider" perspectives when annotating 2,000 posts from an online peer-support platform. We interpolate this analysis to a corpus containing over 2.3 million posts on bullying and related topics, and reveal significant gaps in ML models that use third-party annotators to detect bullying incidents. Our results indicate that models based on the insiders' perspectives yield a significantly higher recall in identifying bullying posts and are able to capture a range of explicit and implicit references and linguistic framings, including person-specific impressions of the incidents. Our study highlights the importance of incorporating the victim's point of view in establishing effective tools for cyberbullying risk detection. As such, we advocate for the adoption of human-centered and value-sensitive approaches for algorithm development that bridge insider-outsider perspective gaps in a way that empowers the most vulnerable.

Introduction

Bullying is a prolonged, repetitive, and aggressive behavior from one individual directed to another (Smith et al. 1999). It has been shown to have a long-term negative impact on victims, especially youth. The damage inflicted on young victims of bullying can last throughout their lives, negatively influencing their health, wealth, social, and mental well-being (Zwierzynska, Wolke, and Lereya 2013). What exacerbates these negative impacts is that bullying comes in various forms and happens across multiple environments, ranging from home, school, to one's workplace. In recent years, bullying has transcended offline person-to-person circumstances to include *cyberbullying*, a form of bullying that occurs online, and has affected more than half of youth within the U.S. (Anderson 2018) and ∼65% of all youth within their lifetime (Brochado, Soares, and Fraga 2017).

Given its prevalence, experts agree that cyberbullying is a problem that must be addressed in order to protect the mental health, safety, and well-being of our youth (Thomas, Connor, and Scott 2015) and because the damage inflicted on young victims of bullying can last throughout their lives (Zwierzynska, Wolke, and Lereya 2013). Young people rely on social media as a means to make new friends, develop their social networks, as well as forming bridging and bonding social capital (Ellison, Steinfield, and Lampe 2007); however, ∼70% of youth have experienced "drama" amongst their online friends (Lenhart et al. 2015) and express concerns towards social media websites in tackling cyberbullying (Hamm et al. 2015). Meanwhile, youth and young adults also use social media to make sensitive disclosures and seek support around important personal issues, including bullying victimization and sexual abuse (De Choudhury and De 2014; Andalibi et al. 2016). However, in some cases, youth report being further traumatized due to cyberbullying that result from these sensitive online disclosures, which were meant for seeking support (Razi, Badillo-Urquiola, and Wisniewski 2020) or to gain therapeutic benefits (Ernala et al. 2017). Although cyberbullying is typically against the terms of service of most platforms, the problem still persists; it is nearly impossible for a handful of content moderators engaged in volunteer labor to manually keep up with the increasing volumes of online interactions (Van Royen et al. 2015).

Consequently, there have been many attempts over the years to build and evaluate sophisticated and robust computationally-driven systems to detect bullying, whether offline or online (see (Rosa et al. 2019) for a comprehensive review). An effective automated or a mixed-initiative system that combines machine and human efforts to detect cyberbullying could offer helpful resources to victims of cyberbullying. Such systems can augment the abilities of content moderators of online platforms so that they can intervene and mitigate behaviors that may be deemed inappropriate per the norms of a community (Van Cleemput, Vandebosch, and Pabian 2014). However, recent work (Ziems, Vigfusson, and Morstatter 2020) points out key limitations, such as the lack of publicly available training data and a robust standard for determining ground truth, that have made existing cyberbullying detection algorithms unfit for real-world use. Notably, to

TalkLife Dataset

- Selected Relationships, Family, Friends, Bullying, and Mental Health based on iterative manual inspection of post categories, excluding those that were irrelevant (e.g., Life Hacks, Poetry, etc.)

TalkLife Categories			
Relationships	**Family**	Self Harm	**Friends**
Hopes	**Bullying**	Health	Work
Music	Helpful Tips	Parenting	Education
Religion	LGBT	Pregnancy	Positive
Mental Health	My Story	Poetry	Eating Disorders
Addiction	Self-Care	I Need Help	New Parents
Gaming	Grief	Anxiety	Disabilities
Depression	Life Hacks	Politics	Ask TalkLife
Others			

- Cyberbullying detection:
  - Prevent bias in annotators for being conservative or liberal in identifying what constitutes bullying in posts
  - Try to be consistent with the mental model of the post authors when they assigned categories

Category	**Number of Posts**	**Number of Users**
Bullying	40248	32907
Mental Health	434456	112892
Family	227243	84417
Friends	676729	126799
Relationships	983752	212850

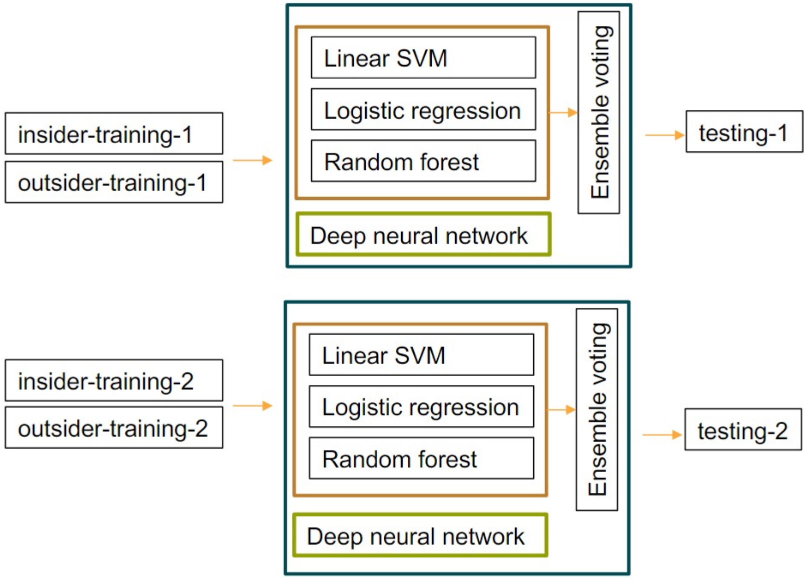
Ground Truth Annotation

Category	Example Post
Positive-Bullying	My boyfriend slapped me and hasn't apologise and is now telling his family about it. Do I have cause to be upset?
	is it normal to feel like all your friends are againt you like they are talking about you behind your back and you know its happening but you cant do anything about it....if you try to you only get bullied and put down..and they just laugh and hurt you emotionally and physically
Negative-Bullying	You don't have right to punish who apologized you from deep inside. They suffer with guiltiness and also they realized they done something wrong, but not accepting apologize and punishing them is the rude way!
	Why do people hate me?

- Five annotators
  - Three male, two female – all undergraduate students
  - had previously experienced bullying or seen their narratives online either directly or indirectly
- A total of 2,000 posts used for the annotation process
- Initial annotation resulted in 267 posts of clear "yes" and 1,381 of clear "no" (Fleiss K 0.47)
  - Remaining 352 posts were re-examined to come to consensus
- Final dataset of 535 bullying and 1,465 non-bullying posts (Fleiss' K 0.79)

Feature Engineering and Models

- Psycholinguistic Attributes (LIWC)
- Sentiment – Stanford CoreNLP's deep learning tool to get the positive, negative, or neutral scores of each post (Manning et al., 2014)
- Open Vocabulary – top 500 n-grams (n=1,2) from each post
- Hate Lexicon (Saha, Chandrasekharan, and De Choudhury, 2019)



- Linear Support Vector Machine, Random Forest, Logistic Regression, Ensemble Voting, and deep neural network model (Founta et al., 2019)
    - 10-fold cross-validation for hyperparameter tuning

		Linear SVM										Logistic Regression									
		Prec		Rec		F1		AUC		Accr		Prec		Rec		F1		AUC		Accr	
		In	Out	In	Out	In	Out	In	Out	In	Out	In	Out	In	Out	In	Out	In	Out	In	Out
Full	Pos-Bully	0.61	0.80	0.70	0.25	0.64	0.35	0.62	0.58	0.62	0.58	0.61	0.80	0.72	0.29	0.64	0.40	0.61	0.60	0.61	0.60
	Neg-Bully	0.65	0.55	0.53	0.92	0.57	0.69					0.68	0.57	0.51	0.91	0.54	0.69				
-Hate	Pos-Bully	0.60	0.80	0.72	0.29	0.64	0.39	0.61	0.60	0.61	0.60	0.61	0.80	0.72	0.29	0.64	0.40	0.62	0.60	0.61	0.60
	Neg-Bully	0.66	0.57	0.50	0.91	0.54	0.70					0.68	0.57	0.52	0.91	0.55	0.69				
-Sent	Pos-Bully	0.61	0.83	0.71	0.27	0.64	0.38	0.61	0.60	0.61	0.60	0.61	0.82	0.75	0.29	0.66	0.39	0.62	0.60	0.61	0.60
	Neg-Bully	0.65	0.56	0.51	0.92	0.55	0.70					0.66	0.57	0.48	0.91	0.52	0.69				
-n-gram	Pos-Bully	0.56	0.81	0.76	0.30	0.64	0.39	0.58	0.59	0.58	0.59	0.61	0.79	0.72	0.28	0.64	0.38	0.61	0.59	0.61	0.59
	Neg-Bully	0.64	0.56	0.40	0.88	0.47	0.68					0.68	0.56	0.50	0.90	0.53	0.69				
-LIWC	Pos-Bully	0.52	0.79	0.77	0.28	0.61	0.29	0.52	0.54	0.52	0.54	0.49	0.79	0.73	0.29	0.57	0.30	0.50	0.55	0.50	0.55
	Neg-Bully	0.58	0.54	0.27	0.80	0.32	0.63					0.62	0.54	0.27	0.80	0.27	0.63				

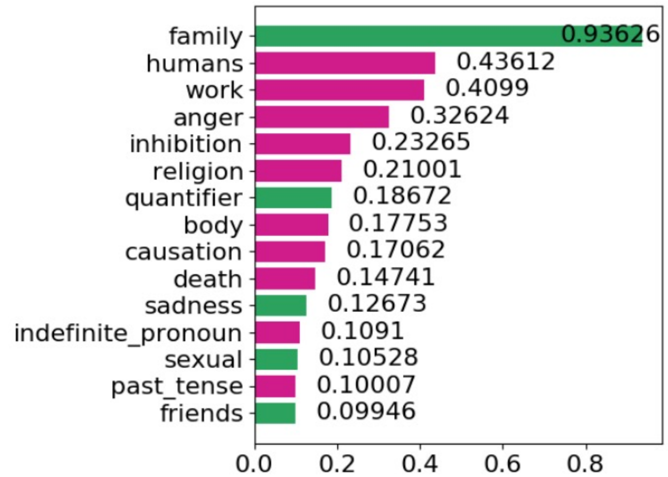
		Random Forest										Ensemble									
		Prec		Rec		F1		AUC		Accr		Prec		Rec		F1		AUC		Accr	
		In	Out	In	Out	In	Out	In	Out	In	Out	In	Out	In	Out	In	Out	In	Out	In	Out
Full	Pos-Bully	0.55	0.30	0.78	0.24	0.62	0.27	0.54	0.54	0.54	0.54	0.60	0.82	0.73	0.28	0.64	0.38	0.61	0.59	0.61	0.59
	Neg-Bully	0.29	0.54	0.31	0.85	0.30	0.65					0.67	0.56	0.49	0.91	0.53	0.69				
-Hate	Pos-Bully	0.59	0.30	0.57	0.24	0.58	0.27	0.58	0.54	0.58	0.54	0.62	0.82	0.72	0.27	0.65	0.37	0.62	0.59	0.62	0.59
	Neg-Bully	0.58	0.54	0.58	0.84	0.58	0.64					0.69	0.56	0.53	0.91	0.57	0.69				
-Sent	Pos-Bully	0.55	0.29	0.79	0.24	0.63	0.26	0.55	0.54	0.55	0.54	0.61	0.84	0.73	0.28	0.65	0.38	0.62	0.59	0.62	0.60
	Neg-Bully	0.30	0.53	0.32	0.84	0.31	0.64					0.66	0.56	0.52	0.91	0.56	0.69				
-n-gram	Pos-Bully	0.56	0.30	0.75	0.24	0.62	0.27	0.55	0.54	0.55	0.54	0.60	0.82	0.73	0.26	0.64	0.36	0.61	0.58	0.61	0.58
	Neg-Bully	0.57	0.54	0.35	0.84	0.35	0.65					0.67	0.56	0.49	0.90	0.53	0.68				
-LIWC	Pos-Bully	0.49	0.32	0.72	0.20	0.56	0.24	0.49	0.54	0.49	0.54	0.50	0.79	0.72	0.28	0.57	0.29	0.50	0.54	0.50	0.54
	Neg-Bully	0.24	0.53	0.27	0.89	0.25	0.66					0.62	0.54	0.28	0.80	0.29	0.63				

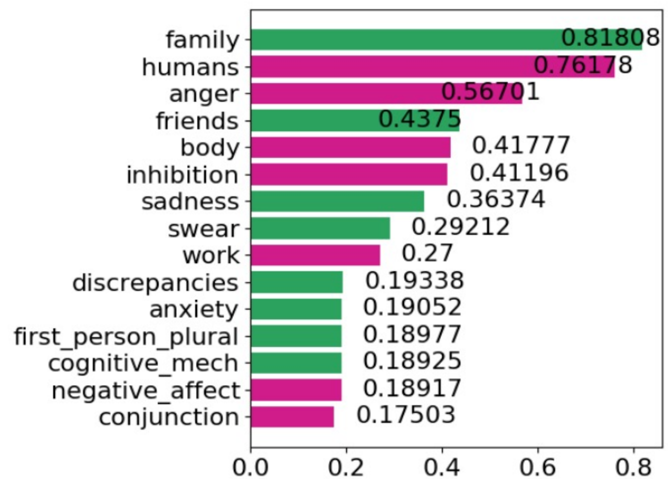
		Neural Network Model									
		Prec		Rec		F1		AUC		Accr	
		In	Out	In	Out	In	Out	In	Out	In	Out
Full	PB	0.66	0.80	0.36	0.17	0.48	0.32	0.64	0.64	0.63	0.71
	NB	0.57	0.54	0.80	0.92	0.66	0.68				
-Hate	PB	0.64	0.75	0.39	0.23	0.49	0.33	0.62	0.67	0.65	0.73
	NB	0.51	0.55	0.78	0.91	0.65	0.68				
-Sent	PB	0.71	0.77	0.36	0.21	0.47	0.31	0.65	0.66	0.64	0.71
	NB	0.57	0.54	0.85	0.92	0.68	0.68				
-n-gram	PB	0.69	0.72	0.34	0.21	0.44	0.31	0.66	0.66	0.64	0.71
	NB	0.56	0.54	0.84	0.92	0.67	0.68				
-LIWC	PB	0.67	0.58	0.03	0.03	0.06	0.06	0.57	0.57	0.61	0.67
	NB	0.51	0.51	0.98	0.98	0.67	0.67				

Top Predictive Features

- Analyze the most discriminative features for the best performing models
  - K-best univariate statistical scoring model using mutual information
  - ANOVA to establish statistical significance
- LIWC was consistently highly predictive for both models

Insider model	Outsider model
Unigrams	
know, **afraid**, *parents*, *although*, *guilty*, **bout**, **rude**, *drugs* **kids**, *asked*, **bullshit**, *anymore*, *embarassing*, **guess**, *grown*	*parents*, **rude**, *get*, *know*, **point**, **whole**, *guess*, **bullshit**, *life*, *stuff*, *kik*, **assume**, *anymore*, **admit**, **inside**
Bigrams	
much hate, *always stuff*, **people calling**, *around online*, *school boy*, **people dont**, *want live*, **im ugly**, *say things*, **seem like**, **need stop**, **fat disgusting**, **can't deal**, *sometimes feel*, **try make**	**anyone ever**, **always get**, **can't deal**, **fat disgusting**, *always online*, *school boy*, **people dont**, *want live*, *anyone want*, *sometimes feel*, **im ugly**, **real life**, **fuck fuck**, **white people**, **need stop**





Error Analysis

- Posts that outsider model "missed," but insider model labeled correctly as Bullying

- Capturing implicit references
  - *"I just wanna say that THIS IS NOT OKAY. WEARING A SHORT SKIRT ISN'T A OPEN INVITATION, YOU CANNOT TOUCH SOMEBODY LIKE THAT THINKING IT'S OKAY JUST CAUSE OF WHAT THEY'RE WEARING!!!"*

- Framing of experiences
  - *"I'm so tired of being picked on by others. I'm so tired of being alone and broken, over and over again, then I'm told that I'm nothing and will never be anything. No one cares about me and no one ever has. literally no one gives a damn about me. That's why I cut myself over and over again. I just want to die so I can stop the pain"*

- Quantitative Approaches
  - Supervised learning
  - Unsupervised learning
- Psychometric Approaches
- Qualitative Approaches
- Cross-disciplinary Approaches
- Multi-method Approaches

Evaluation Approach for Unsupervised Learning

- No ground truth available
- Examples –
  - Community detection
  - Social recommendations and social ranking
  - Personalization of content

A Social Media Study on Mental Health Status Transitions Surrounding Psychiatric Hospitalizations

SINDHU KIRANMAI ERNALA, Georgia Institute of Technology, USA
KATHAN H. KASHIPAREKH, Georgia Institute of Technology, USA
AMIR BOLOUS, Georgia Institute of Technology, USA
ASRA ALI, Zucker Hillside Hospital, Psychiatry Research, USA
JOHN M. KANE, Zucker Hillside Hospital, Psychiatry Research, USA
MICHAEL L. BIRNBAUM, Zucker Hillside Hospital, Psychiatry Research, USA
MUNMUN DE CHOUDHURY, Georgia Institute of Technology, USA

For people diagnosed with a mental illness, psychiatric hospitalization is one step in a long journey, consisting of *clinical recovery* such as removal of symptoms, and *social reintegration* involving resuming social roles and responsibilities, overcoming stigma and self-maintenance of the condition. Both clinical recovery and social reintegration need to go hand-in-hand for the overall well-being of individuals. However, research exploring social media for mental health has considered narrower, disjoint conceptualizations of people with mental illness – either as a patient or as a support-seeker. In this paper, we combine medical records with social media data of 254 consented individuals who have experienced a psychiatric hospitalization to address this gap. Adopting a theory-driven, Gaussian Mixture modeling approach, we provide a taxonomy of six heterogeneous behavioral patterns characterizing peoples' mental health status transitions around hospitalizations. Then we present an empirically derived framework, based on feedback from clinical researchers, to understand peoples' trajectories around clinical recovery and social reintegration. Finally, to demonstrate the utility of this taxonomy and the empirical framework, we assess social media signals that are indicative of individuals' reintegration trajectories post-hospitalization. We discuss the implications of combining peoples' clinical and social experiences in mental health care and the opportunities this intersection presents to post-discharge support and technology-based interventions for mental health.

CCS Concepts: • **Human-centered computing** → **Computer supported cooperative work**; **Empirical studies in collaborative and social computing**; **Social media**.

Additional Key Words and Phrases: mental health, social media, psychiatric hospitalization, health status transitions, Facebook

ACM Reference Format:
Sindhu Kiranmai Ernala, Kathan H. Kashiparekh, Amir Bolous, Asra Ali, John M. Kane, Michael L. Birnbaum, and Munmun De Choudhury. 2021. A Social Media Study on Mental Health Status Transitions Surrounding Psychiatric Hospitalizations. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 155 (April 2021), 32 pages. https://doi.org/10.1145/3449229

Authors' addresses: Sindhu Kiranmai Ernala, sernala3@gatech.edu, Georgia Institute of Technology, USA; Kathan H. Kashiparekh, kathan.kashiparekh@gatech.edu, Georgia Institute of Technology, USA; Amir Bolous, agirgis6@gatech.edu, Georgia Institute of Technology, USA; Asra Ali, AAli22@northwell.edu, Zucker Hillside Hospital, Psychiatry Research, USA; John M. Kane, jkane2@northwell.edu, Zucker Hillside Hospital, Psychiatry Research, USA; Michael L. Birnbaum, mbirnbaum@northwell.edu, Zucker Hillside Hospital, Psychiatry Research, USA; Munmun De Choudhury, munmund@gatech.edu, Georgia Institute of Technology, USA.

2573-0142/2021/4-ART155 $15.00
https://doi.org/10.1145/3449229

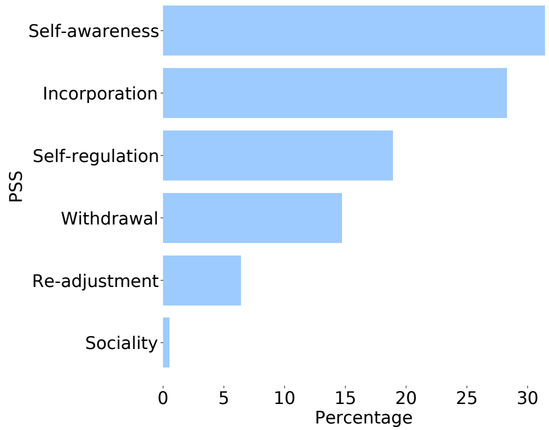

NAMI

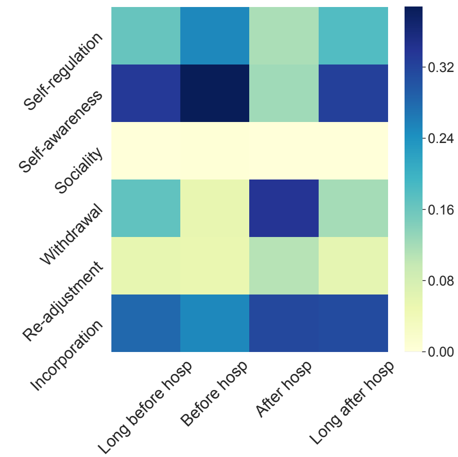
RQ1: What self-presentation and behavioral signals on social media characterize individuals' mental health statuses around psychiatric hospitalizations?
RQ2: What trajectories on social media showcase transitions between these statuses surrounding the hospitalizations?
RQ3: What social media-based signals are indicative of social reintegration trajectories of individuals following hospitalizations?

- Facebook archives of 254 consented participants who have experienced at least one hospitalization for
  - psychosis (N=142)
  - mood disorders (N=106)
  - other mental health conditions (N=6).

- Across all participants, we compile over 980 thousand Facebook posts around 372 hospitalization events.

- **Possible Selves framework:** as a theoretical lens to capture and interpret peoples' conceptions of self-knowledge and alternate versions of themselves in the future





PSS	***Increase*** **in behaviors**	***Decrease*** **in behaviors**	**Example posts and behaviors**
Self-regulation	Use of words indicative of positive emotions***. *"I just wanted to say that everyone danced so well tonight. I really enjoyed the performance."*	All other actions. E.g., use of function words, pronouns,first-person singular pronouns, show focus on the present in language, post content indicative of mental health symptoms and experiences, use of impersonal pronouns, show focus on the past in language*, third-person plural pronouns, words related to anger, body, work, sadness, death, leisure*, words indicative of negative emotions**, words about friends*. *"My roommate was watching OUAT, and I remember how quick Ruby was to condemn Regina."*	*"It is world humanitarian day. I'm doing something good, somewhere for someone else. Join me. #WHD2012 #IWASHERE"*
Self-awareness	Use of function words, personal pronouns, first-person singular pronouns, words related to cognitive processes, first-person plural pronouns, words related to anger**, words related to money*, show focus on the present in language, show focus on the future in language*, posting content indicative of mental health symptoms or experiences. *"I just can't sleep, I watched American Horror Story for the whole day. I promised myself I would wake up early and clean my room."*	Posts indicative of positive and negative emotions, sending messages to friends on FB, posting photos on FB, words related to leisure, one-on-one interactions on FB, words indicative of anxiety*, sharing feelings with status updates on FB**, shares on FB*. *Number of FB shares relatively dropped by 100%.*	*"It's just a slap in the face when you are your only sole motivation and advice giver... you have no one saying 'keep going', 'i'm proud of you"', "You work hard on your mental health to the point your new psychiatrist doesn't want you on meds anymore.", "Feeling accomplished and great.", "Hungry and bored again. blah!"*
Sociality	Overall posts and activities on FB, uploading photos and cover photos on FB, sharing feelings via posts on FB, likes on FB, one-on-one interactions on FB, sending messages on FB, co-tagged with others on FB, use of informal words. *Number of FB posts relatively increased by 341%.*	Post content related to mental health symptoms and experiences, use of second-person pronouns, first-person plural pronouns, words indicative of negative emotions, showing focus on the past in language, use of words related to leisure, adding new friends on FB. *"These winds are blowing down everything except the Trump tower."*	*"I have a lot of best friends lol so happy... national best friend day to everyone who are my best friend."*
Withdrawal	No action.	Use of function words, pronouns, posting content indicative of mental health symptoms and conditions, posting content indicative of positive emotions**, use of person pronouns, showing focus on the present in language, use of first-person singular pronouns, anger, words related to body. *"omg! You've got great hair styling skills sister."*	*"[The user] went to [a certain music festival]", "[The user] added education to his timeline", "[The user] added [a city] to his current city."*
Re-adjustment	Use of words related to leisure, sexual words, words related to work, ingestion***, pronouns, function words , informal words, co-tagging with others on FB, words related to anger, death, adding new friends on FB, use of words related to health***, showing focus on the past in language***, sending messages to friends on FB***, shares on FB***. *"I'm on the verge of a manic episode. WHAT DO I DO?"*	Only use of third-person singular pronouns. *"Woot! commemorating my 7th good hair day in a row"*	*"Friends always: fight for you, include you, respect you... stand by you. People believe your actions more than your words.", "I am not my hair. I am not this skin. I am not your expectations. I am a soul that lives within", "#oldclassmates reunion", "#fuckedup", "#amen".*
Incoporation	Showing focus on the present in language*, words indicative of negative emotions***, posting content indicative of mental health symptoms and experiences, us of informal words**, function words, pronouns, sending messages to friends on FB, use of personal pronouns, first-person singular pronouns, words related to religion**, body one-on-one interactions on FB, use of words about friends***, health**, sharing feelings with status updates on FB***. *"my girl..I need your help pls get back to me as soon as possible"*	Use of words indicating positive emotions, co-tagging with others on FB, overall posting and activities on FB, use of words related to anxiety*. *"Aww! Thanks for the feel better card!"*	*"Anyone with a TV watching the movie Avengers I might be able to join?", "Anyone coming from [a location] that might be able to give me a ride to [another location]?", "Sorry I missed your show last night. Make sure you keep me posted with everything going on."*

- Quantitative Approaches
  - Supervised learning
  - Unsupervised learning
- Psychometric Approaches
- Qualitative Approaches
- Cross-disciplinary Approaches
- Multi-method Approaches

Psychometric Approaches

- Face validity
- Construct validity
- Concurrent validity

Social Media as a Measurement Tool of Depression in Populations

Munmun De Choudhury **Scott Counts** **Eric Horvitz**
Microsoft Research, Redmond WA 98052
{munmund, counts, horvitz}@microsoft.com

ABSTRACT
Depression is a serious and widespread public health challenge. We examine the potential for leveraging social media postings as a new type of lens in understanding depression in populations. Information gleaned from social media bears potential to complement traditional survey techniques in its ability to provide finer grained measurements over time while radically expanding population sample sizes. We present work on using a crowdsourcing methodology to build a large corpus of postings on Twitter that have been shared by individuals diagnosed with clinical depression. Next, we develop a probabilistic model trained on this corpus to determine if posts could indicate depression. The model leverages signals of social activity, emotion, and language manifested on Twitter. Using the model, we introduce a social media depression index that may serve to characterize levels of depression in populations. Geographical, demographic and seasonal patterns of depression given by the measure confirm psychiatric findings and correlate highly with depression statistics reported by the Centers for Disease Control and Prevention (CDC).

Author Keywords
behavior, depression, emotion, health, language, social media, mental health, public health, Twitter, wellness

ACM Classification Keywords
H.4.3; H.5.m

INTRODUCTION
Depression affects more than 27 million Americans and is believed to be responsible for the more than 30,000 suicides each year [2,14]. Besides being directly debilitating to sufferers, depression can adversely affect chronic health conditions, such as cardiovascular disease, cancer, diabetes, and obesity. It is also known to have negative influences on individuals' family and personal relationships, work or school life, and sleeping and eating habits.

Over the coming 20 years, depression is projected to be the leading cause of disability in high-income nations such as the United States [16]. The World Health Organization (WHO) now ranks major depression as one of the most burdensome diseases in the world [2,16]. Although a number of primary care programs have been devised for its detection and treatment, the majority of the millions of Americans who meet depression criteria are untreated or undertreated [11]. Furthermore, ethnic minority groups such as Mexican Americans and African Americans are significantly less likely to receive depression therapies than are other ethnic groups [9].

As part of a national-scale effort to curb depression, every few years the Centers for Disease Control and Prevention (CDC) administers the Behavioral Risk Factor Surveillance System (BRFSS) survey via telephone to estimate the rate of depression among adults in the US [2]. However the large temporal gaps across which these measurements are made, as well as the limited number of participant responses (on the order of thousands) makes it difficult for agencies to track and identify risk factors that may be associated with mental illness, or to develop effective intervention programs.

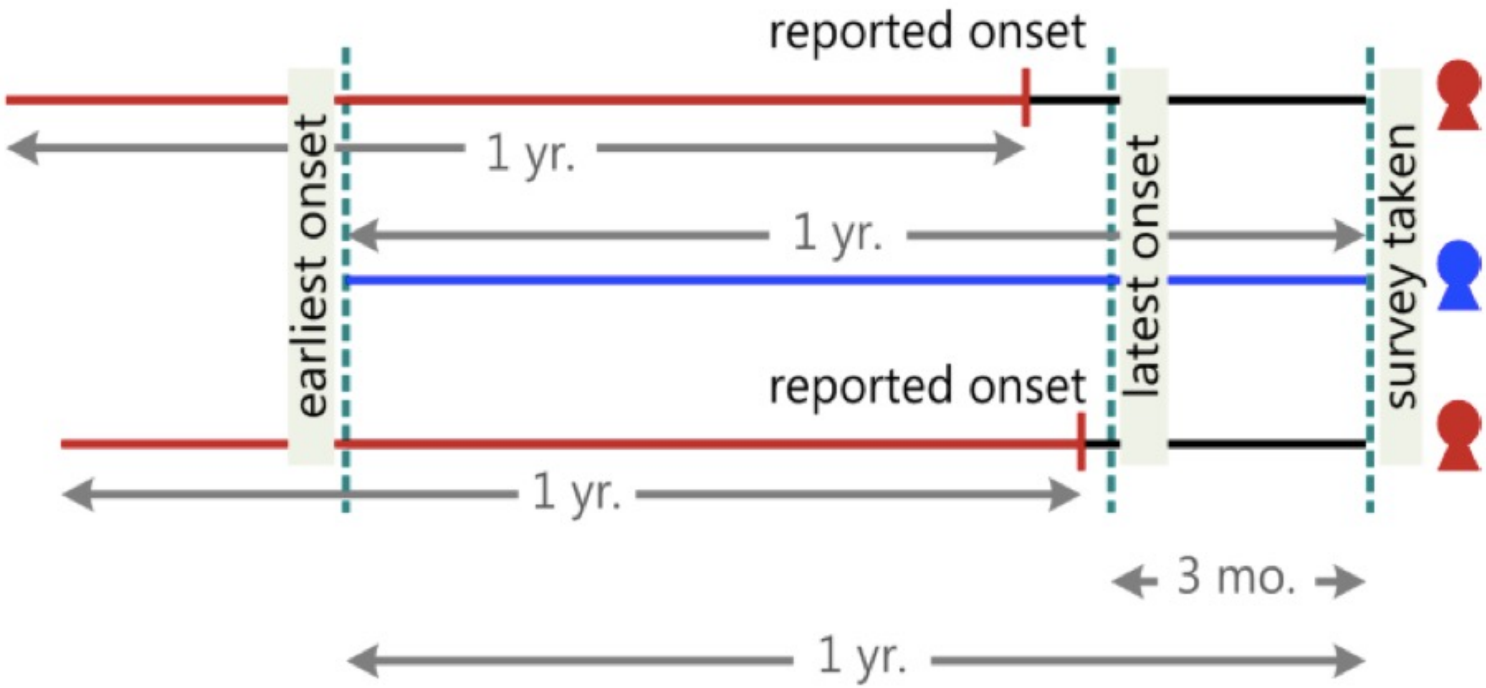
We examine the potential of social media as a new tool for mental health measurement and surveillance. Platforms such as Twitter and Facebook are increasingly gaining traction among individuals allowing them to share their thoughts and emotions around a variety of happenings in everyday life. The emotion and language used in social media postings may indicate feelings of worthlessness, guilt, helplessness, and self-hatred that characterize depression as manifested in everyday life. Additionally, depression sufferers often show withdrawal from social situations and activities—i.e., the etiology of depression typically includes social environmental factors [17]. Characterization of social media activity and changing social ties within social media can provide measurement of such withdrawal and capture the depression sufferers' social context in a manner that might help detect depression in populations.

Relying on social media as a behavioral health assessment tool has other advantages as well. For instance, in contrast to the self-report methodology in behavioral surveys, where responses are prompted by the experimenter and typically comprise recollection of (sometimes subjective) health facts, social media measurement of behavior captures social activity and language expression in a naturalistic setting. Such activity is real-time, and happens in the course of a

WebSci'13, May 2–4, 2013, Paris, France.

Social Media Language and Behavior Predicts Onset of Major Depressive Disorder

476 individuals (233 female) recruited via Amazon's Mechanical Turk; took CES-D/BDI and consented to accessing their Twitter data

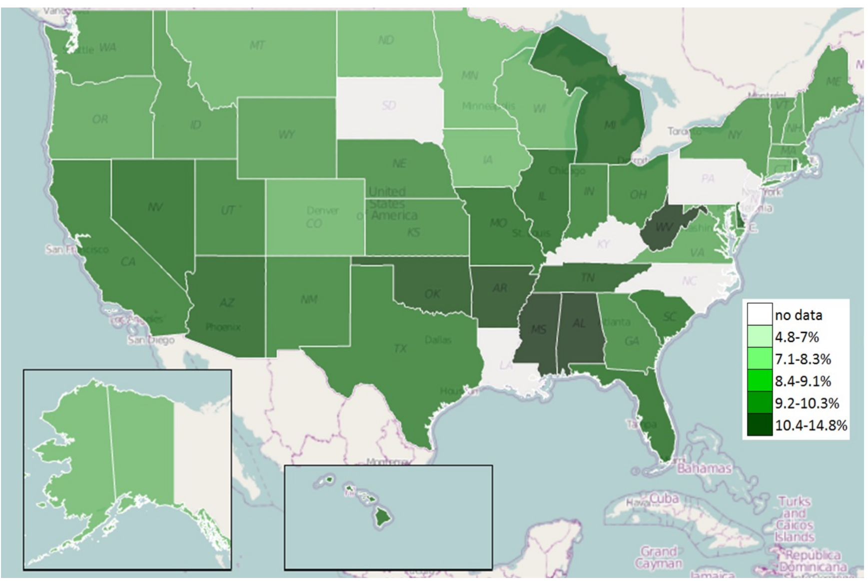


A Social Media Index

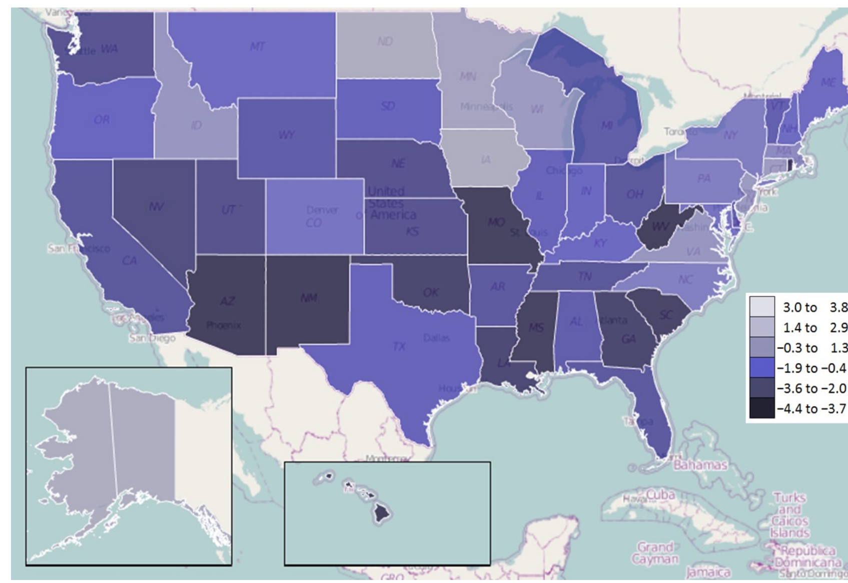
$$SMDI(t) = \frac{n_d(t) - \mu_d}{\sigma_d} - \frac{n_s(t) - \mu_s}{\sigma_s}$$

standardized difference between frequencies of depression-indicative and standard posts, compared to a period before between *k* and *t-1* ($1 \leq k \leq t-1$)

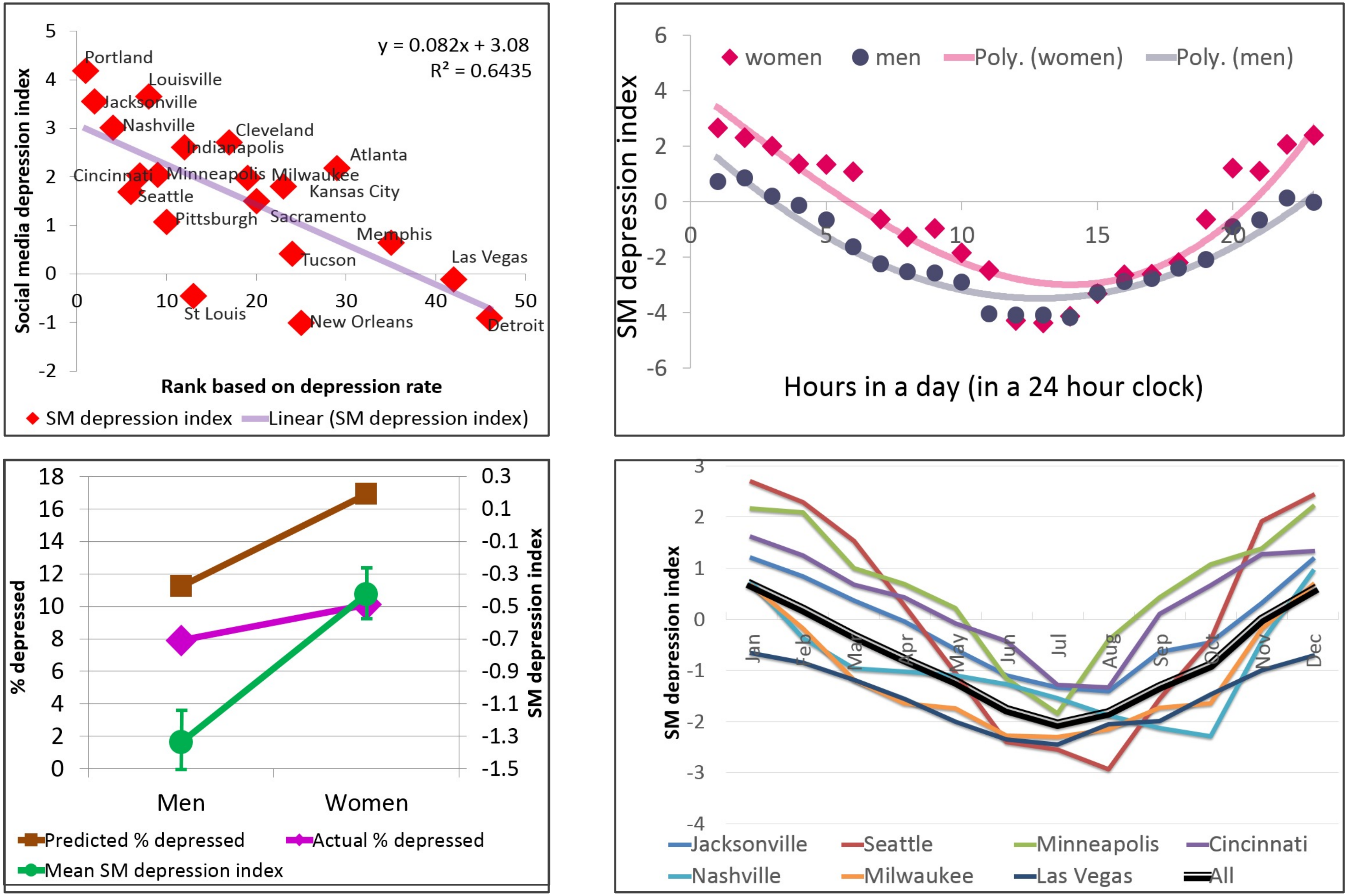
actual (BRFSS data)



predicted (SMDI)



least squares regression fit yields correlation of 0.52



y = 0.082x + 3.08
R² = 0.6435
Portland
Louisville
Jacksonville
Nashville
Cleveland
Indianapolis
Atlanta
Cincinnati
Minneapolis
Milwaukee
Kansas City
Seattle
Pittsburgh
Sacramento
Memphis
Tucson
Las Vegas
St Louis
New Orleans
Detroit
Social media depression index
Rank based on depression rate
SM depression index
Linear (SM depression index)
women
men
Poly. (women)
Poly. (men)
SM depression index
Hours in a day (in a 24 hour clock)
% depressed
SM depression index
Men
Women
Predicted % depressed
Actual % depressed
Mean SM depression index
SM depression index
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Jacksonville
Seattle
Minneapolis
Cincinnati
Nashville
Milwaukee
Las Vegas
All

- Quantitative Approaches
  - Supervised learning
  - Unsupervised learning
- Psychometric Approaches
- Qualitative Approaches
- Cross-disciplinary Approaches
- Multi-method Approaches

Qualitative Approaches

- Expert review
  - Sanity check
  - Independent assessments
- Surveys
- Interviews and Focus Groups

Marginalization and the Construction of Mental Illness Narratives Online: Foregrounding Institutions in Technology-Mediated Care

SACHIN R. PENDSE, Georgia Institute of Technology, USA
NEHA KUMAR, Georgia Institute of Technology, USA
MUNMUN DE CHOUDHURY, Georgia Institute of Technology, USA

People experiencing mental illness are often forced into a system in which their chances of finding relief are largely determined by institutions that evaluate whether their distress deserves treatment. These governing institutions can be offline, such as the American healthcare system, and can also be online, such as online social platforms. As work in Human-Computer Interaction (HCI) and Computer Supported Cooperative Work (CSCW) frames technology-mediated support as one method to fill structural gaps in care, in this study, we ask the question: how do online and offline institutions influence how people in resource-scarce areas understand and express their distress online? We situate our work in U.S. Mental Health Professional Shortage Areas (MHPSAs), or areas in which there are too few mental health professionals to meet expected needs. We use an analysis of *illness narratives* to answer this question, conducting a large scale linguistic analysis of social media posts to understand broader trends in expressions of distress online. We then build on these analyses via in-depth interviews with 18 participants with lived experience of mental illness, analyzing the role of online and offline institutions in how participants express distress online. Through our findings, we argue that a consideration of institutions is crucial in designing effective technology-mediated support, and discuss the implications of considering institutions in mental health support for platform designers.

CCS Concepts: • **Human-centered computing → collaborative and social computing**.

Additional Key Words and Phrases: institutions, marginalization, mental health, resource constraints, support platforms

ACM Reference Format:
Sachin R. Pendse, Neha Kumar, and Munmun De Choudhury. 2023. Marginalization and the Construction of Mental Illness Narratives Online: Foregrounding Institutions in Technology-Mediated Care. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW2, Article 346 (October 2023), 30 pages. https://doi.org/10.1145/3610195

1 INTRODUCTION

The fields of Computer Supported Cooperative Work (CSCW) and Human-Computer Interaction (HCI) have increasingly sought to understand people's everyday experiences with disempowerment [35, 103], and investigate whether technology can support both immediate needs [77] and a broader formation of counterpower [7, 51, 56]. One particular site of disempowerment for individuals in distress in the United States is the process of seeking care via the American mental health care system. An individual's ability to be treated is dependent on institutional factors largely out of

Authors' addresses: Sachin R. Pendse, Georgia Institute of Technology, Atlanta, GA, USA, sachin.r.pendse@gatech.edu; Neha Kumar, Georgia Institute of Technology, Atlanta, GA, USA, neha.kumar@gatech.edu; Munmun De Choudhury, Georgia Institute of Technology, Atlanta, GA, USA, munmund@gatech.edu.

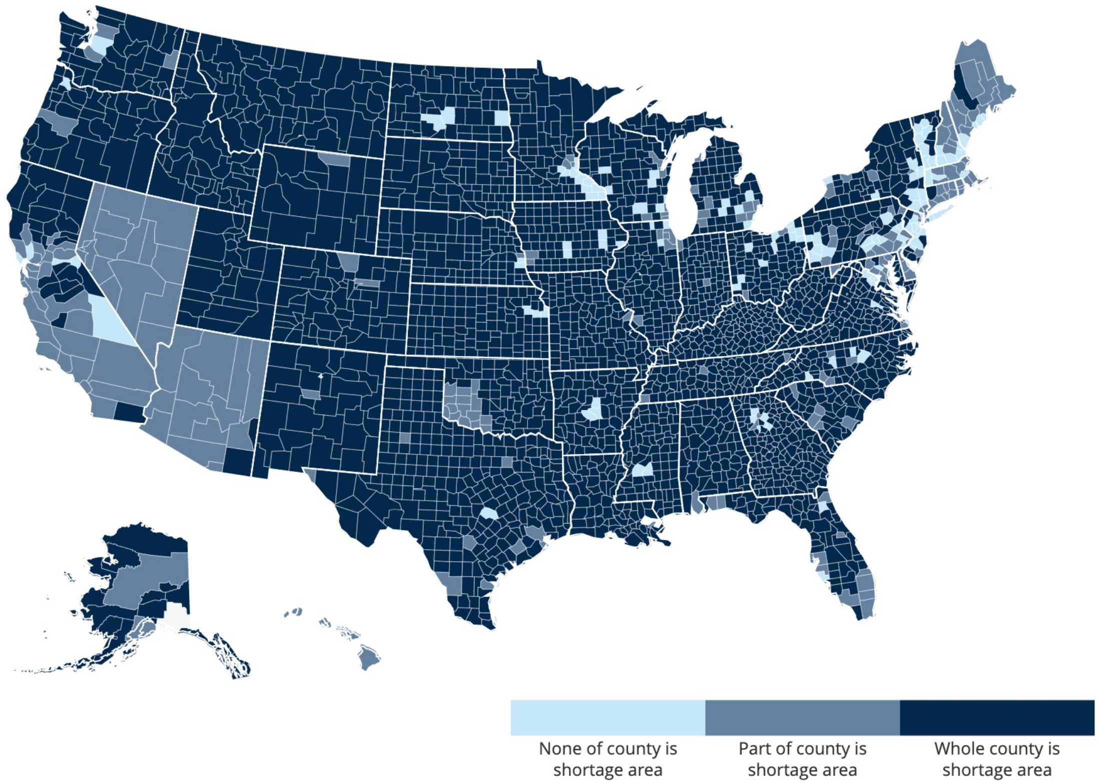
2573-0142/2023/10-ART346
https://doi.org/10.1145/3610195

Majority of individuals with mental illness do not get treatment

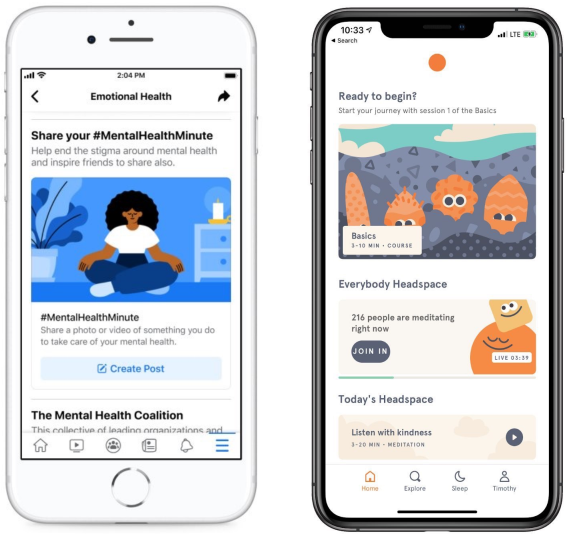
Time to treatment is more than a decade

1 in 4 forced to choose between mental health care and rent/food

Clinical care is often first accessed after suicide attempts



Statistics from Little Treatments, Big Effects by Jessica Schleider
Graphic from RuralHealthInfo.org

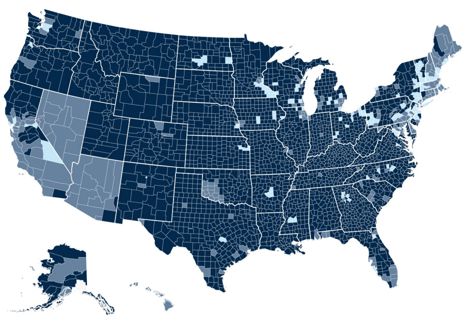


Interventions must be both effective and *acceptable*

How do online and offline institutions influence how people in resource-scarce areas understand and express their distress online?

“Illness narratives edify us about how life problems are created, controlled, made meaningful”

– Arthur Kleinman



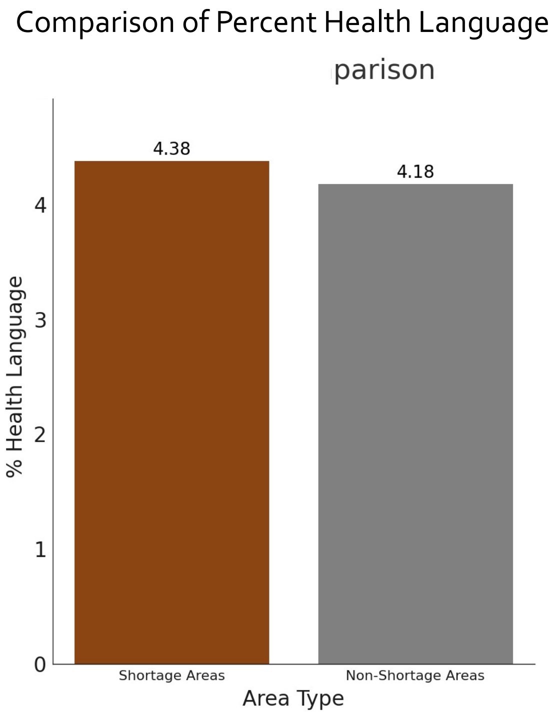
Large scale linguistic analysis:

- 1,967,582 Twitter posts
- 1,145,013 users
- 2015-2017

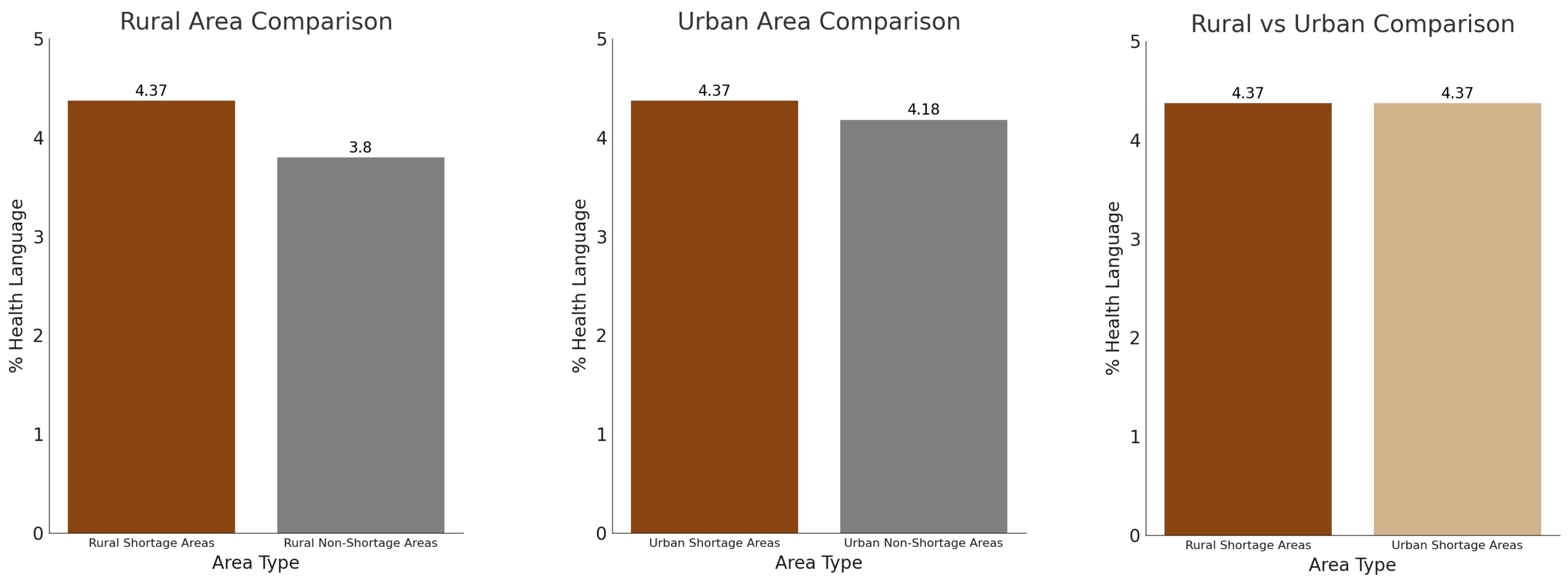
In-depth interviews with 18 participants with lived experience of mental illness

Shortage Areas
gone
tired
wish
god
hurt

Non-Shortage Areas
stab
fucking
fuck
new
worth



Significantly higher levels of somatic, religious, and community-oriented language in shortage areas



Rural Area Comparison
% Health Language
Area Type
Rural Shortage Areas
Rural Non-Shortage Areas
4.37
3.8
Urban Area Comparison
Urban Shortage Areas
Urban Non-Shortage Areas
4.37
4.18
Rural vs Urban Comparison
Rural Shortage Areas
Urban Shortage Areas
4.37
4.37
0
1
2
3
4
5

Online expressions of distress are reflective of offline resource constraints

Participants described somatic and religious symptoms

"violent diarrhea"

— Olivia (*Non-Shortage*)

"a hole in my eye"

— Abe (*Shortage*)

"I was so angry at God"

— Donna (*Shortage*)

"a queasy stomach + toothaches"

— Perla (*Shortage*)

"I largely only feel it in my body"

— Kendall (*Non-Shortage*)

All participant names are pseudonyms

Institutional rules around “valid” illness kept people from care

> “But then they screwed me over in the most classic way, they sent me to their doctor. And I didn’t have the privilege of my own psychiatrist who would vouch for me. So of course, their doctor did his job and gave me a clean bill of health, which basically screwed me for the next five years. During which time, I attempted suicide, became homeless for a long stretch, and barely survived a domestic violence relationship, one of a series of violent relationships where I was almost killed.”
>
> —Isabella

Illness narratives (influenced by online and offline institutions) had a strong influence on illness experiences.

- Quantitative Approaches
    - Supervised learning
    - Unsupervised learning
- Psychometric Approaches
- Qualitative Approaches
- Cross-disciplinary Approaches
- Multi-method Approaches

Cross-disciplinary Approaches

- Incorporating Insights:
  - Integrating knowledge from sociology, psychology, and related disciplines.

A Social Media Study on Demographic Differences in Perceived Job Satisfaction

KOUSTUV SAHA, Georgia Institute of Technology, USA
ASRA YOUSUF, Georgia Institute of Technology, USA
LOUIS HICKMAN, Purdue University, USA
PRANSHU GUPTA, Georgia Institute of Technology, USA
LOUIS TAY, Purdue University, USA
MUNMUN DE CHOUDHURY, Georgia Institute of Technology, USA

Effective ways to measure employee job satisfaction are fraught with problems of scale, misrepresentation, and timeliness. Current methodologies are limited in capturing subjective differences in expectations, needs, and values at work, and they do not lay emphasis on demographic differences, which may impact people's perceptions of job satisfaction. This study proposes an approach to assess job satisfaction by leveraging large-scale social media data. Starting with an initial Twitter dataset of 1.5M posts, we examine two facets of job satisfaction, *pay* and *supervision*. By adopting a theory-driven approach, we first build machine learning classifiers to assess perceived job satisfaction with an average AUC of 0.84. We then study demographic differences in perceived job satisfaction by geography, sex, and race in the U.S. For geography, we find that job satisfaction on Twitter exhibits insightful relationships with macroeconomic indicators such as financial wellbeing and unemployment rates. For sex and race, we find that females express greater pay satisfaction but lower supervision satisfaction than males, whereas Whites express the least pay and supervision satisfaction. Unpacking linguistic differences, we find contrasts in different groups' underlying priorities and concerns, e.g., under-represented groups saliently express about basic livelihood, whereas the majority groups saliently express about self-actualization. We discuss the role of frame of reference and the "job satisfaction paradox", conceptualized by organizational psychologists, in explaining our observed differences. We conclude with theoretical and sociotechnical implications of our work for understanding and improving worker wellbeing.

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

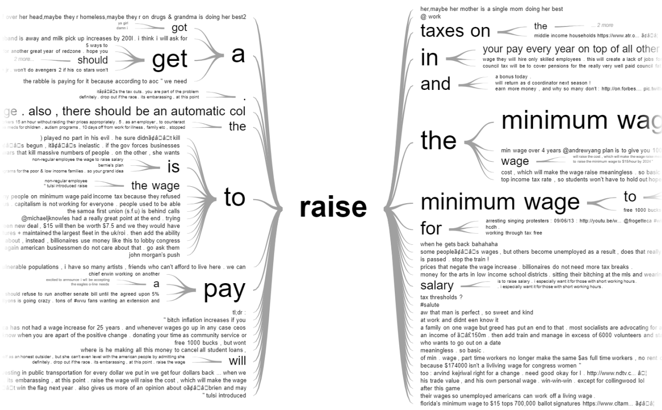
Additional Key Words and Phrases: job satisfaction, demographic differences, sex, gender, race, geography, Twitter, social media, macroeconomic constructs, workplace

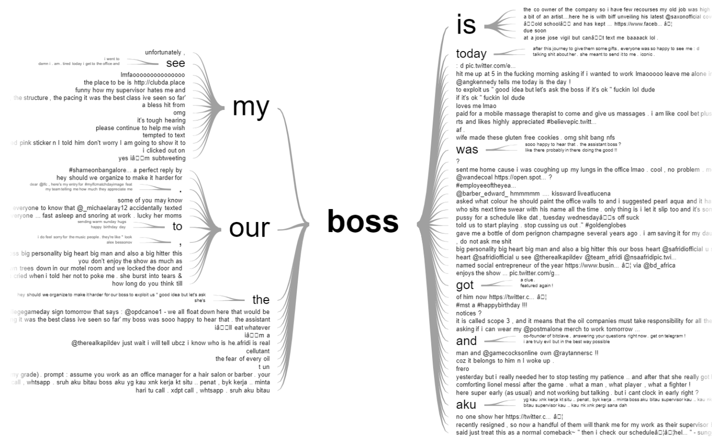
ACM Reference Format:
Koustuv Saha, Asra Yousuf, Louis Hickman, Pranshu Gupta, Louis Tay, and Munmun De Choudhury. 2021. A Social Media Study on Demographic Differences in Perceived Job Satisfaction. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 167 (April 2021), 29 pages. https://doi.org/10.1145/3449241

Authors' addresses: Koustuv Saha, Georgia Institute of Technology, Atlanta, GA, USA, koustuv.saha@gatech.edu; Asra Yousuf, Georgia Institute of Technology, Atlanta, GA, USA, ayousuf@gatech.edu; Louis Hickman, Purdue University, West Lafayette, IN, USA, lchickma@purdue.edu; Pranshu Gupta, Georgia Institute of Technology, Atlanta, GA, USA, prangupt@gatech.edu; Louis Tay, Purdue University, West Lafayette, IN, USA, stay@perdue.edu; Munmun De Choudhury, Georgia Institute of Technology, Atlanta, GA, USA, munmund@gatech.edu.

2573-0142/2021/4-ART167 $15.00
https://doi.org/10.1145/3449241

Pay and Supervision Satisfaction





Example Post	Relevance	Valence
Pay-Related		
how the fuck have I spent 90% of my pay in 2 days	Not Relevant	NA
I just remembered my bonus funds are reset for this semester	Not Relevant	NA
I got overtime on this paycheck...We get a bonus on next paycheck.#WinWin	Relevant	Positive
it may be years until the day, my dreams will match up with my pay	Relevant	Negative
Supervision-Related		
I jut said I forged my mom's signature In front of my supervisor	Not Relevant	NA
God Bless all Small Businesses all work hard because Jesus is our Boss IJN Amen	Not Relevant	NA
I LOVE my supervisor!	Relevant	Positive
I cannot decide who do I hate the most among my father, my supervisor, and my ex.	Relevant	Negative

Table 3. Median metrics of **pay satisfaction** classifiers in k-fold (k=5) cross-validation.

Model	Relevance			Valence		
	Precision	Recall	AUC	Precision	Recall	AUC
Logistic Reg.	0.73	0.72	0.80	0.80	0.80	0.89
KNN	0.81	0.63	0.70	0.72	0.72	0.81
SVM (Linear)	0.73	0.73	0.81	0.80	0.80	0.89
Random Forest	0.83	0.65	0.80	0.72	0.74	0.82
AdaBoost	0.83	0.66	0.80	0.79	0.79	0.87
MLP	0.81	0.67	0.77	0.76	0.76	0.86

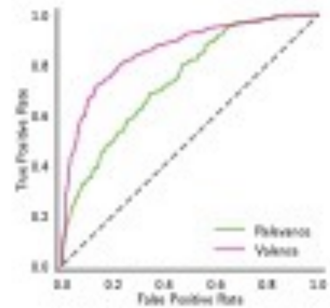


Fig. 2. ROC curves of **pay satisfaction** classifiers.

Table 4. Top features of **pay** satisfaction classifier.

Relevance	Valence
bless, bonus, check, constantly, get bonus, get raise, got raise today, income, income tax, pay, pay check, pay raise, raise today, LIWC: (1st P., Work), reduce, tax, today, week	food, increasing, need, pay, reduce, salary, spend, taxes, wage, LIWC: (Anger, Causation, 1st P. Pl., 2nd P., Fu. Tense, Aux. Verb, Relative, Achievement), Sentiment: (Pos., Neg.)

Table 5. Median metrics of **supervision satisfaction** classifiers in k-fold (k=5) cross-validation.

Model	Relevance			Valence		
	Precision	Recall	AUC	Precision	Recall	AUC
Logistic Reg.	0.71	0.70	0.78	0.81	0.81	0.88
KNN	0.88	0.52	0.69	0.75	0.75	0.82
SVM (Linear)	0.75	0.76	0.80	0.81	0.82	0.87
Random Forest	0.88	0.53	0.75	0.81	0.81	0.83
AdaBoost	0.88	0.57	0.78	0.77	0.77	0.85
MLP	0.87	0.60	0.77	0.78	0.78	0.85

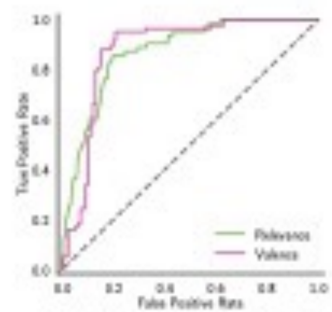
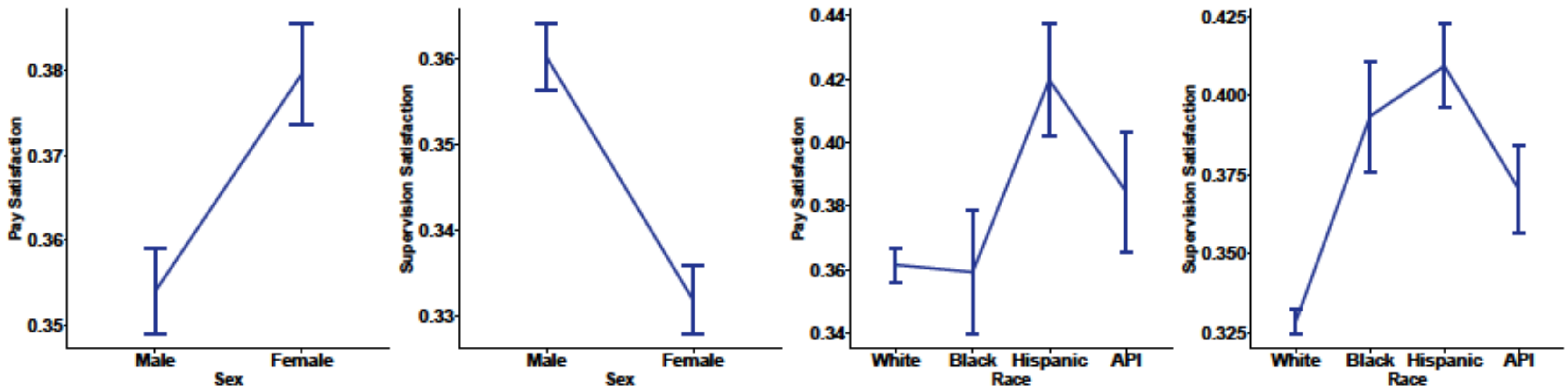


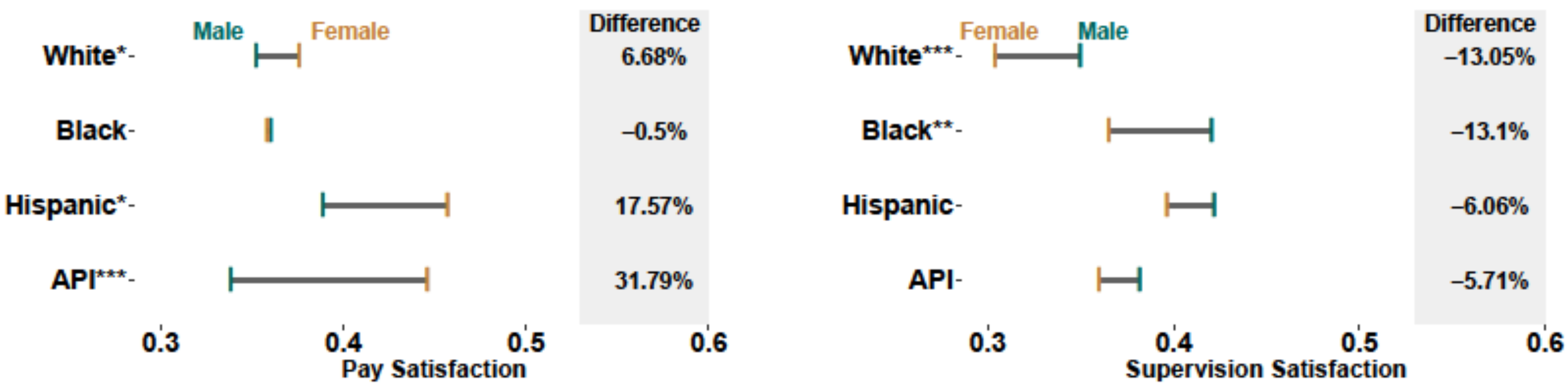
Fig. 3. ROC curves of **supervision satisfaction** classifiers.

Table 6. Top features of **supervision** satisfaction classifier.

Relevance	Valence
ass, awesome, bitch, boss, congratulations, ceo, early, gave, go home, hiring, job like, leave, love, lunch, proud, seriously, shift, shit, supervisor gave, team, team lead, tied team, told	amazing, best, bitch, boss, cool, congrats, fired, fuck, happy, hate, like, love, proud, shit, LIWC: (Anger, N. Affect, Swear, Negation, Quantifier, Achievement), Sentiment: (Pos., Neg.)



(a) Sex: Pay (b) Sex: Supervision (c) Race: Pay (d) Race: Supervision



(e) Sex & Race: Pay (f) Sex & Race: Supervision

Males			
n-gram	SAGE	*n*-gram	SAGE
wealth tax	2.21	average wage	0.80
civil	2.05	pay grade	0.79
rights	1.84	income tax	0.77
homelessness	1.33	federal government	0.76
sex	1.33	healthcare	0.76
technician	1.30	work late	0.75
wealth	1.29	competitive	0.74
baseball	1.17	wage growth	0.73
returns	1.15	tax cuts	0.73
center applicant	1.13	get fired	0.70
gun	1.10	unions	0.68
teams	1.08	sacked	0.65
facilities	1.08	progressive	0.61
registered nurse	1.06	growth	0.58
pay cut	1.24	wage laws	0.58
2nd shift	1.07	hr manager	0.53
football	1.06	applicant	0.53
tax rate	0.98	progressive	0.50
taxes paid	0.93	capital gains	0.48
wage growth	0.88	wage war	0.47

Females			
n-gram	SAGE	*n*-gram	SAGE
toxic workplace	-4.92	prayer	-1.16
survival guide toxic	-4.87	maternity leave	-1.15
adaptive	-4.12	medical bills	-1.05
lawful ethical	-3.82	husband	-1.04
feedback criticalthinking	-3.51	work tonight	-1.03
sexually	-1.82	hourly shift	-1.00
scare	-1.79	extra income	-0.96
bra	-1.68	supportive	-0.95
harassed	-1.68	anxiety	-0.93
manager transportation	-1.67	ass manager	-0.93
girl got raise	-1.64	emotional	-0.92
security shift	-1.63	stressful	-0.90
land like asst	-1.55	family income	-0.89
survival	-1.51	like hospitality	-0.86
touched	-1.43	bonus check week	-0.82
outfit	-1.33	drama	-0.82
home early	-1.32	screamed	-0.80
barista	-1.26	boyfriend	-0.80
clerical	-1.21	clothes	-0.79
abused	-1.20	wear	-0.76

Whites			
n-gram	SAGE	*n*-gram	SAGE
management trainee	4.35	interested	1.76
starbucks apply	3.07	innovative	1.69
way meeting	2.84	nervous	1.66
hr check	2.84	coffee shop	1.66
shy score	2.84	like sales	1.65
referrals great way	2.84	working month	1.64
supervisor panera	2.84	hiring anchorage	1.64
competitive salary	2.83	career rise	1.61
asking referrals	2.80	explore opportunities	1.61
health tracking	2.71	best options	1.60
ulta beauty	2.47	doubled pay	1.59
retail opportunity	2.48	midtown	1.51
looking score	2.38	read latest	1.49
bio	2.35	sympathy	1.38
great fit	2.34	mr president	1.37
information apply	2.29	technology	1.06
cosmetology	2.23	home health	1.01
underestimating value	2.03	hair stylist	0.58
lasvegas	1.99	look specific	0.83
minnesota	1.92	aramark	0.72

Racial Minorities			
n-gram	SAGE	*n*-gram	SAGE
reproductive	-2.70	bruh	-1.21
impeach	-2.44	faith	-1.19
facilitiesmgmt	-2.34	lord bless	-1.17
voting	-2.24	supervisor wanna	-1.16
strike	-2.17	work ethic	-1.15
civil	-2.10	annoying	-1.06
minimum wage	-2.02	medicare	-0.96
homelessness	-1.92	chinese	-0.92
extra income	-1.76	af	-0.88
rights	-1.54	patience	-0.87
reduce stress	-1.51	fucking hate	-0.84
sex	-1.47	really hate	-0.84
green	-1.37	getting fired	-0.83
finna	-1.35	pay shit	-0.82
scare	-1.34	bitch ass	-0.80
wealth tax	-1.33	work environment	-0.80
niggas	-1.32	mississippi	-0.78
ghetto	-1.32	safety	-0.76
southwest	-1.27	stress	-0.74
need extra income	-1.21	household income	-0.70

- Quantitative Approaches
  - Supervised learning
  - Unsupervised learning
- Psychometric Approaches
- Qualitative Approaches
- Cross-disciplinary Approaches
- Multi-method Approaches

everytown.org

EXPLORE

SEARCH

EVERYTOWN FOR GUN SAFETY

HAPPENING NOW

Humanities & Social Sciences Communications

ARTICLE

https://doi.org/10.1057/s41599-021-00993-6 OPEN

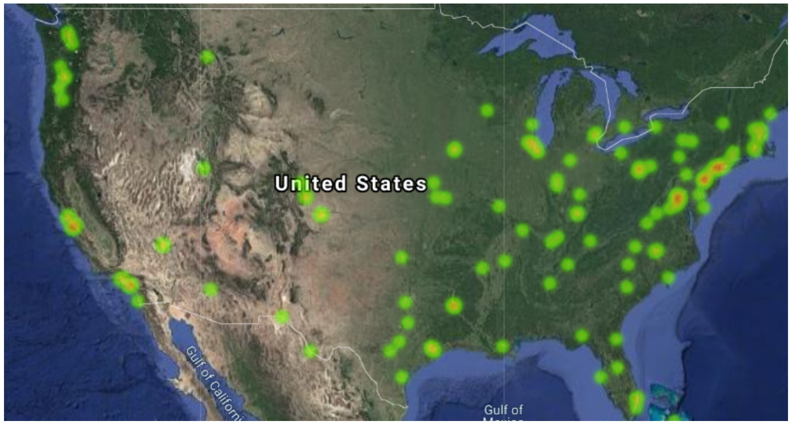
Check for updates

Impacts of school shooter drills on the psychological well-being of American K-12 school communities: a social media study

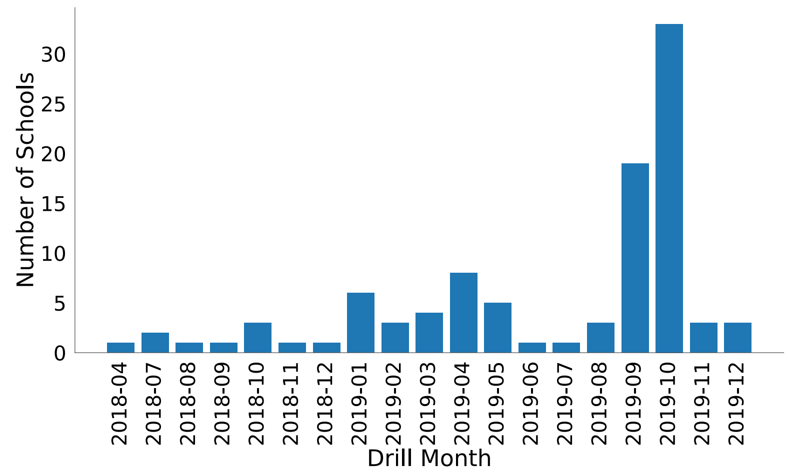
Mai ElSherief[1,5], Koustuv Saha[2,5], Pranshu Gupta[3], Shrija Mishra[3], Jordyn Seybolt[3], Jiajia Xie[3], Megan O'Toole[4], Sarah Burd-Sharps[4] & Munmun De Choudhury[3]✉

The toll from gun violence in American K-12 schools has escalated over the past 20 years. School administrators face pressure to prepare for possible active shootings, and often do so through drills, which can range from general lockdowns to simulations, involving masked "shooters" and simulated gunfire, and many variations in between. However, the broad and lasting impact of these drills on the well-being of school communities is poorly understood. To that end, this article applies machine learning and interrupted time series analysis to 54 million social media posts, both pre- and post-drills in 114 schools spanning 33 states. Drill dates and locations were identified via a survey, then posts were captured by geo-location, school social media following, and/or school social media group membership. Results indicate that anxiety, stress, and depression increased by 39–42% following the drills, but this was accompanied by increases in civic engagement (10–106%). This research, paired with the lack of strong evidence that drills save lives, suggests that proactive school safety strategies may be both more effective, and less detrimental to mental health, than drills.

[1]University of California, San Diego, CA, USA. [2] Microsoft Research, Montreal, QC, Canada. [3] Georgia Institute of Technology, Atlanta, GA, USA. [4] Everytown for Gun Safety, New York, NY, USA. [5]Authors conducted this research while affiliated to Georgia Institute of Technology. ✉email: munmund@gatech.edu



(a) Geographical distribution of K-12 schools with active shooter drills considered in this research.



(b) Distribution of active shooter drills between April 2018 and December 2019

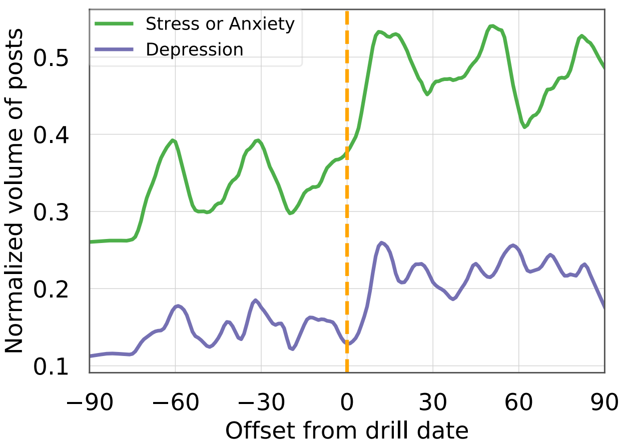
114 K-12 schools in 33 states in the 2018-19 school year

6 months of analysis (3 months pre- and post-incidents)

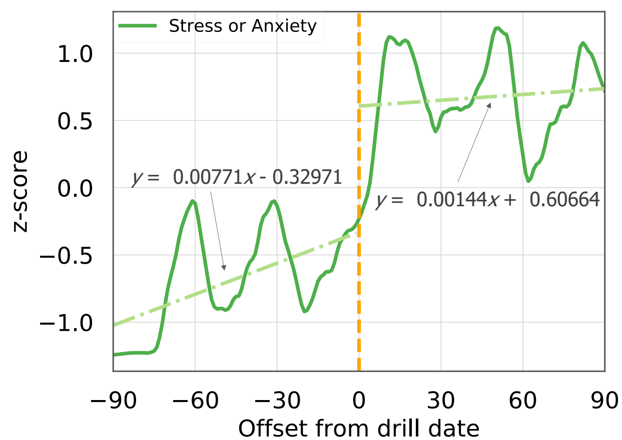
54 million social media posts

Focus group interviews with 34 stakeholders

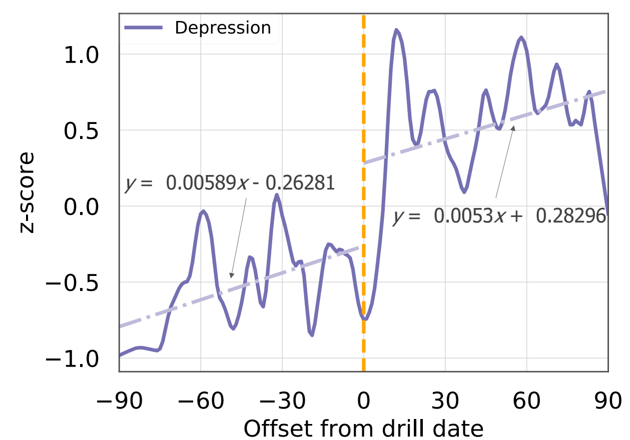
Well-being concerns increased by 38-42%



(a) Stress/Anxiety and Depression levels

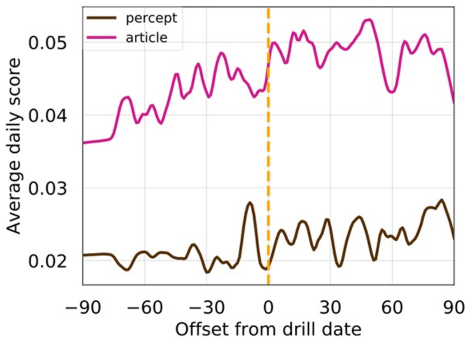


(b) Stress/Anxiety standardized trend

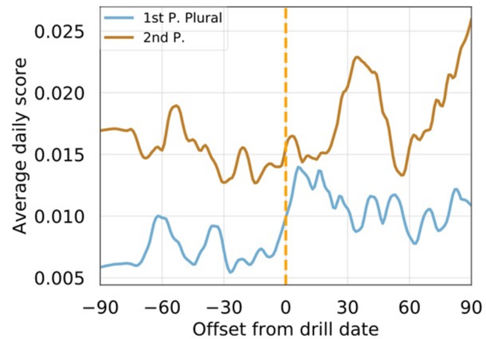


(c) Depression standardized trend

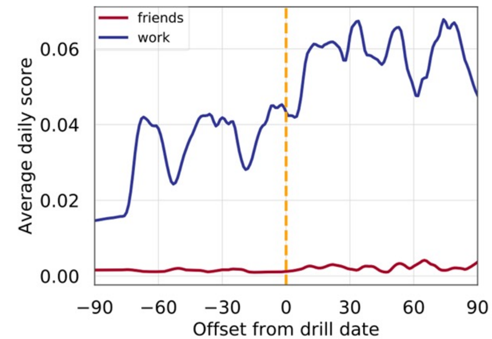
Well-being concerns increased by 38-42%



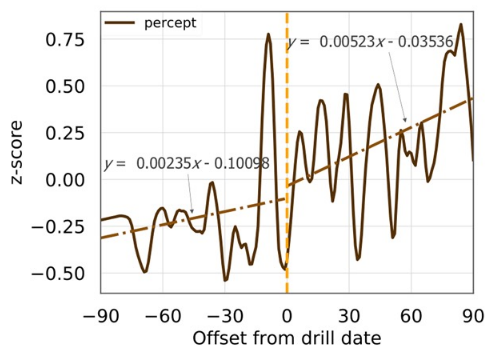
(a) Levels of use of percept and article words.



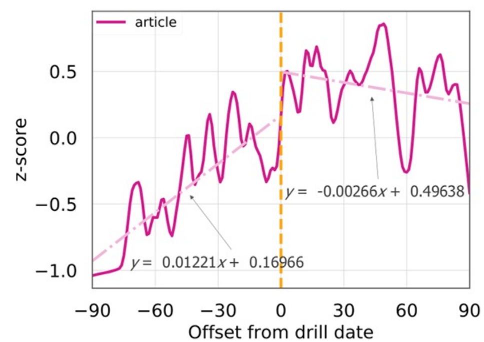
(b) Levels of 1st person plural and 2nd person pronouns.



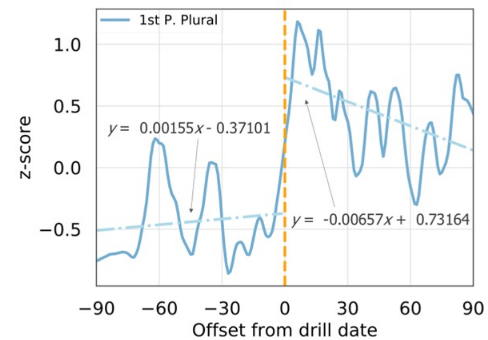
(c) Levels of friends and work words.



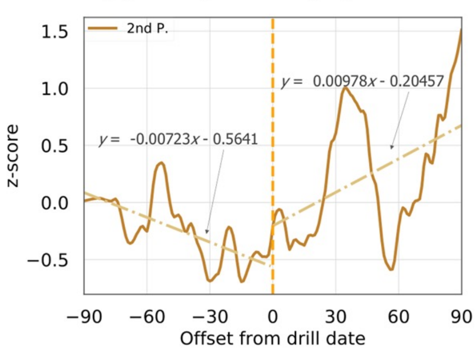
(d) Perception category trend.



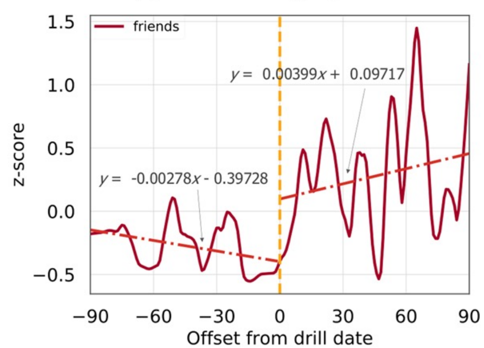
(e) Article category trend.



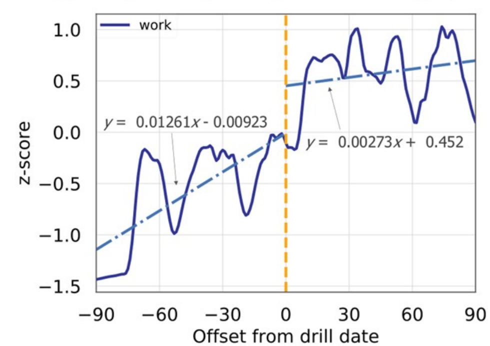
(f) 1st person plural pronoun category trend.



(g) 2nd person pronoun category trend.



(h) Friends category trend.

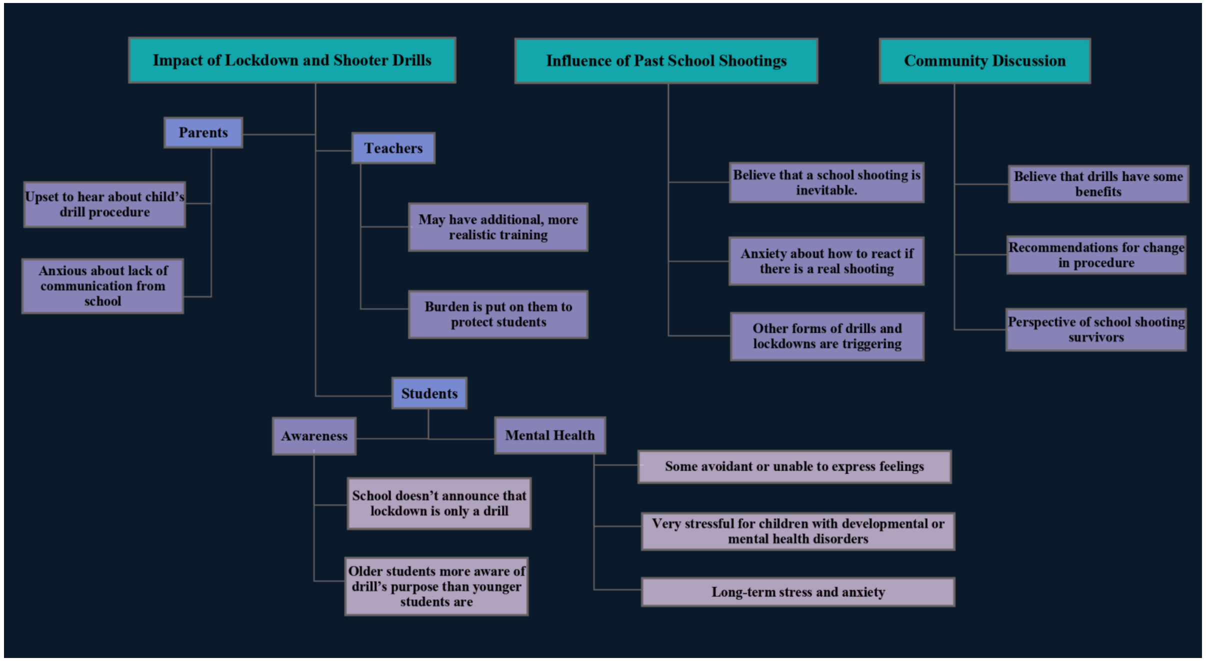


(i) Work category trend.

Focus Group Triangulation

boyd and Crawford's provocation, "in this computational turn, it is increasingly important to recognize the value of 'small data"'

6 1-h long focus group interviews with 21 parents, 11 teachers, and 2 students



- Quantitative Approaches
  - Supervised learning
  - Unsupervised learning
- Psychometric Approaches
- Qualitative Approaches
- Cross-disciplinary Approaches
- Multi-method Approaches