

Gender and Cross-Cultural Differences in Social Media Disclosures of Mental Illness

Munmun De Choudhury
Georgia Tech
Atlanta, GA 30332 USA
munmund@gatech.edu

Sanket S. Sharma
Georgia Tech
Atlanta, GA 30332 USA
sanket@gatech.edu

Tomaz Logar
United Nations Global Pulse
New York, NY 10017 USA
tomaz@unglobalpulse.org

Wouter Eekhout
Leiden University
2311 EZ Leiden, Netherlands
w.t.k.eekhout@fgga.leidenuniv.nl

René Clausen Nielsen
United Nations Global Pulse
New York, NY 10017 USA
rene@unglobalpulse.org

ABSTRACT

Cultural and gender norms shape how mental illness and therapy are perceived. However, there is a paucity of adequate empirical evidence around gender and cultural dimensions of mental illness. In this paper we situate social media as a “lens” to examine these dimensions. We focus on a large dataset of individuals who self-disclose to have an underlying mental health concern on Twitter. Having identified genuine disclosures in this data via semi-supervised learning, we examine differences in their posts, as measured via linguistic attributes and topic models. Our findings reveal significant differences between the content shared by female and male users, and by users from two western and two majority world countries. Males express higher negativity and lower desire for social support, whereas majority world users demonstrate more inhibition in their expression. We discuss the implications of our work in providing insights into the relationship of gender and culture with mental health, and in the design of gender and culture-aware health interventions.

Author Keywords

social media; twitter; reddit; mental health; depression; suicide; health; gender; culture

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION

I do not know how to communicate with the experts. He told me that I have some kind of disease in my mind, but I think I am okay. He kept asking me to express my feelings toward the earthquake, but I feel embarrassed

if I tell people my own feelings. — Taiwanese natural disaster victim, quoted in [34].

The United Nations’ Millennium Development Goals established in 2000 identify large-scale cultural, social, economic, demographical and political processes underlying differential health risks in the global population¹. In fact, it is established that rarely does biology act alone to determine health inequities [49]. In the context of mental health in particular, gender and culture based differences are significant. Such differences may emanate from a biomedical, social, economic, or an epidemiological perspective [7, 54].

However, significant data gaps exist in the understanding of the gender and cultural dimensions of mental health². Gender and cross-cultural analysis is important to improve our understanding of the epidemiology of mental health problems, decisions and disclosures of these problems in different subgroups, and also to increase potential for greater public participation in health. It is noted that overlooking gender or culture based differences can have drastic consequences. This includes, misdiagnosis, misappropriation of interventions, and a “one-size-fits-all” approach to extend help to those who may have unique needs [49, 47].

In recent years, a new research direction has established social media data as a way to understand mental health challenges in people [15, 13, 24]. At the same time, other work has established social media sites as powerful platforms of self-disclosure and social support-seeking around psychological distress [14, 3, 2]. However, identifying gender based and cross-cultural context is critical in the use of such passively sensed big data for making sense of mental health and well-being.

Our motivation springs from the observation that cultural and gendered expression of different subgroups may differ markedly from the “typical”, largely western populations, on which current social media investigations of mental health are

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
CSCW '17, February 25-March 01, 2017, Portland, OR, USA
Copyright © 2017 ACM ISBN 978-1-4503-4335-0/17/03 \$15.00
DOI: <http://dx.doi.org/10.1145/2998181.2998220>

¹<http://www.un.org/millenniumgoals/>

²http://data2x.org/wp-content/uploads/2014/11/Data2X_MappingGenderDataGaps_FullReport.pdf

based [15, 13]. For instance, depression in women, in particular, is an outstanding challenge in majority world countries like India where the cultural context indicates males and females to inhabit different social worlds. Gender inequality, family violence, and restrictions to independence of young women permeates family life in India. These norms are known to increase risk of mental illness in women [35]. Consequently, rarely are analytical insights and intervention mechanisms able to be transplanted without modification from one subgroup to another, and still provide the same value [27]. It is known that digital behaviors such as disclosure and expression of emotion in social media are influenced strongly by gender and culture based behavioral and linguistic norms. As sociologist Goffman also notes in his celebrated book “Stigma” [18], individuals with a socially discredited attribute such as mental illness, tend to manage impressions of themselves in social settings in order to protect their identities. These actions might derive heavily from their gender and cultural context.

In this paper, we present a gender based and cross-cultural quantitative examination of mental health content shared on social media. Specifically, we study mental illness related self-disclosures made on Twitter, and analyze and characterize them along gender based and cross-cultural dimensions. Our contributions include:

- A machine learning approach to identify genuine self-disclosures of mental illness from noisy social media posts. We use vector-representations of content shared on external mental health support communities as weak labels to infer genuine social media disclosures in a semi-supervised manner. Based on psychology expert consultation, this method is found to yield 96% accuracy.
- Statistical comparisons between the disclosures of female and male users, and by users from four prominent English speaking countries: western countries U.S., U.K., and majority world countries India, and South Africa. Specifically, we examine different linguistic style, affective, behavioral and cognitive attributes, and content characteristics based on a topic modeling approach.

For our work, we use a dataset of half a million Twitter users and nearly 1.5 million posts. Our findings reveal significant differences in how different gender and cultural subgroups, who disclose their mental health concerns, express themselves on Twitter, compared to equivalent control subgroups. Unlike their male peers, female users in our dataset express more positivity, greater involvement in social and familial concerns, higher propensity to engage in health discourse, and show a desire to seek help on the platform. Turning to different cultural subgroups, we find that users from majority world countries demonstrate more inhibition and express fewer negative emotions, compared to their peers from the western countries. Topically, content from the former cohort reveals signs of an avoidant attitude and shame, whereas that from the latter group expresses decreased self-esteem and heightened loneliness.

We discuss how naturalistically generated and unobtrusively gathered social media data can help understand gender and

culture based differences among individuals who disclose their mental health challenge. Our work bears implications for the design of gender and culturally aware interventions, so as to bring tailored, timely help to individuals in need.

Privacy and Ethics. We leverage public data from Twitter and Reddit for our work; hence our work did not qualify for approval from our Institutional Review Board. Nevertheless, we took greater care in de-identifying and paraphrasing any content we present as examples to support our investigation. Importantly, our work does not make any diagnostic claims about mental illness experiences of the population we study.

RELATED WORK

Gaps in Gender and Cultural Dimensions of Mental Health

Gender. Prior work has identified significant differences in the health challenges experienced by gender subgroup populations [38]. Prior research has specifically identified the need for data collection efforts that can address largely unreported causes of women’s excess disease burdens, and parse out the contributions of sex and gender, and their interaction, in the etiology, onset, progression and prevention of different health conditions [37]. Sexual violence and mental health issues among women are especially known to suffer from paucity of adequate data [46]. For a long time, clinicians have reported that women receive more services for mental illness in clinical settings than men [1]. There is therefore a need to understand the expression of mental illness in gender subgroups to identify those at greater risk.

Other works exploring gender-based dimensions of mental illness include the work of Kawachi and Berkman [30]. The authors observed that the protective effects of social ties on mental health are not uniform across gender groups in society. Further, they found that social connections may paradoxically increase levels of mental illness symptoms among women with low resources. In general, Astbury found that gender differences in mental disorders extend beyond differences in the rates of various disorders or their differential time of onset or course, and include factors that can affect susceptibility, disclosure, diagnosis, and adjustment to mental disorder [1].

Culture. According to Shweder [45], culture is “to be a member of a group is to think and act in a certain way, in the light of particular goals, values, pictures of the world; and to think and act so is to belong to a group.” Culture can be defined as any characteristic of a group of people, which can affect and shape their beliefs and behaviors, including mental well-being [45].

Cultures vary in the extent to which expression of distress is socially sanctioned and reported. However, cross-cultural and cross-national studies of mental illnesses are limited [50]. Guillemin et al. [20] noted that cultural groups vary in disease expression and in their use of various health care systems. They go on to argue for the need to develop mental health and quality of life assessment measures specifically geared toward populations in non English-speaking countries. Similarly, Yeomans and Forman noted that many diagnostic

methods and models have been derived from studies of samples from industrialized countries; however their application to diverse cultural populations needs attention to reliability and validity [57].

In terms of specific cultural studies, Yeh [56] studied the particular and unique mental health concerns of Asian immigrant youths, whereas Stack [48] investigated the effect of media on imitative suicide in Japan and compared it to the American context. Further, Wang et al. [54] observed that Chinese individuals are generally reluctant to express distress and often attribute such distress to external or physical causes.

In majority world countries, it is noted that socio-political and economic issues, inequalities in the workplace, societal and cultural expectations are often known to impact individuals' risk to distress and mental illness. The 14-countrywide WHO Multinational study conducted in 1995 around the prevalence, nature and determinants of mental illnesses in general medical care settings provides some broad insights into some of these cross-national dimensions of mental health [52]. Patel et al. [39] reported that the startling finding of this study was that, despite the use of standardized methods in all country-specific centers, there were enormous variations in most variables known to be linked to mental health. Indeed, the only similarities across centers were the general observations of the ubiquity of mental illnesses, and the association of mental illness and disability after adjustment for physical disease severity. On the other hand, specific variables showed substantial variations; the prevalence rates of mental illness ranged from 7-52%, clinician recognition of mental illness varied from 5-60%, and the association of key variables such as gender, physical ill-health and education with mental illness were in opposite directions in different centers. These findings demonstrate the need for comparative cross-cultural and cross-national studies that can identify disclosure practices of mental illnesses, the local needs and thereby inform local policy and interventions [40].

We envision that leveraging social media data to understand gender and culture based differences in mental health disclosures can contribute toward closing these data gaps. More data driven capabilities of mental health inferencing and monitoring can also increase the visibility of these issues and provide an impetus for vulnerable subgroups to seek help and for clinicians to offer more treatment or intervention options.

Social Media and Health

Research in recent years has revealed that social media data, especially language and conversational patterns can be a powerful source of information toward understanding and detecting health challenges in individuals and populations. These efforts include utilizing social media to understand conditions and symptoms related to diseases [41], substance abuse [36, 33], postpartum depression [13], eating disorders [8], and other mental health disorders [15, 11, 51].

Self-disclosure, the “process of making the self known to others”, has been noted to be a basic element in the attainment of improved health [28]. Thus, another line of research has examined how social media platforms might be allowing honest and candid expression of thoughts, experiences and be-

liefs around mental health concerns [14, 3]. De Choudhury and De [14] explored how use of Reddit for mental health purposes is characterized by disinhibiting behavior. Andalibi et al. [3] qualitatively characterized the variety of self-disclosures that are shared on Instagram via the hashtag “#depression”. While these works study self-disclosures in online communities of support, we extend this body of work by developing an automated method to identify mental illness related self-disclosures in arbitrary social media content, that can have significant implications for how social media platforms are utilized to extend help to those in need.

We further note that gender and cross-cultural examinations of health states, health behaviors, and disclosures using social media are limited. UNICEF recently undertook a study of social media content in Eastern Europe to look at attitudes towards vaccination [53]. De Choudhury et al. [15] identified differences between experiences of depression between women and men as measured from Twitter. Andalibi et al. [2] found that men are significantly more likely to adopt anonymous social media identities when engaging in sexual abuse related self-disclosure. Tsugawa et al. [51] examined how individuals with non-English speaking backgrounds expressed depressive thoughts and emotions on Twitter, and thereafter built classification models to identify markers of depression among Japanese speaking users. In other work, Ramirez-Esparza et al. [43] analyzed linguistic attributes of English and Spanish posts shared in depression forums.

Our work presents an in-depth large-scale data-driven study of individuals who choose to disclose their mental illness on Twitter. Adopting this lens of self-disclosure, we then examine gender based and cross-cultural differences in attributes of disclosure in four predominantly English speaking countries.

DATA

Twitter Data

Our study uses publicly shared mental illness self-disclosure data collected from the social media Twitter. We started by obtaining a large sample of English language candidate self-disclosure posts from the full archive of public Twitter data, around a variety of mental health concerns. Specifically, we filtered the Twitter posts shared in March 2015 containing any of the keyphrases included in Table 1. These keyphrases were collated by a combination of reference to prior work [11, 12], and consultation with a trained psychiatrist practitioner. Through these keyphrases, we sought to identify who publicly state that they have been diagnosed with, or suffering from some form of mental illness. As noted by Coppersmith et al., users may make such a statement to seek support from others in their Twitter social network, to fight the stigma of mental illness, or perhaps as an explanation of some of their behavior [11]. We obtained 1,319,064 posts from 534,829 unique users at the end of this initial data collection phase.

Parallelly, we obtained a candidate control data sample from Twitter's Firehose stream, so as to allow robust statistical comparisons between Twitter users who choose to self-disclose their mental illness, and those who do not. This dataset included a random sample of 1,513,279 posts from 673,898 unique users made on Twitter during March

i want to die	i want to end my life
i want to suicide	i tried to suicide
i [*] thinking of suicide	i thought of suicide
i am depressed	i [*] diagnosed [*] depression
i [have/had] mental illness	i [*] diagnosed [*] mental illness
i attempted suicide	i [have/had] depression
killing myself	ending my life

Table 1: Keyphrases used for obtaining candidate mental health disclosure posts from Twitter.

2015, ensuring that none of these posts matched any of the keyphrases given in Table 1.

Thereafter, for both of the candidate mental health disclosure sample and the candidate control sample of posts, we utilized Twitter’s official API³ to obtain the last 3,200 posts for each of the unique users in both datasets. For the control dataset, if any of the users had any post in their crawled posts matching one of the keyphrases above, we disregard them from our analysis. Further, employing the Google Compact Language Detector⁴, we disregarded any users, if at least 75% of their posts were not written in English. Our final candidate disclosure sample contained 51,038,914 posts from 470,337 users ($\mu = 108.5$), while the control sample contained 66,214,850 posts from 480,685 users ($\mu = 137.7$).

Gender and Country Inference

In order to allow gender based and cross-cultural comparisons of the above data, we now present an automated gender and country name inference method — Twitter does not allow individuals an ability to self-report their gender, whereas location information is known to be highly noisy [22]. We also note the need for such a method given the size of our two datasets, that makes human coding of gender and country names challenging from a practical perspective.

Gender Inference. For inferring gender of a user in our two datasets, we compared account names to existing name databases. First of all, we started cleaning the account names and subsequently identifying the first and last name on the basis of a 1-gram lookup. Thereafter, we compared the names to country-specific name databases. For this, the country origin of a Twitter user was firstly retrieved using the Country Name Inference, giving the user’s specified location. If the gender could not be retrieved using the previous method it would look up the name over all the country-specific lookup tables. For analytical simplicity, we only consider binary gender (female/male) in this paper.

Country Name Inference. For inferring the country name corresponding to a user, we adopted a stepwise approach as follows: First, we cleaned the location strings reported in the location field of Twitter user profiles — including normalization of character case and removal of non-English word roots. Then we performed a location matching exercise, wherein we split the cleaned location strings into single words, iteratively created all possible 5-to-1-gram substrings, and then

matched each n-gram to a location database based on GeoNames (<http://www.geonames.org>), preferring larger n-grams over smaller ones (“New York City” to “York”). Third, we performed disambiguation by computing geographic distances between matched text-adjacent places and assigned high likelihood to those matches that are close to each other geographically. We then sorted equally likely location alternatives by population size and choose the top one. We note that, compared to geo-located Twitter posts, this location field string lookup method has been known to yield better coverage in social media data [22].

Validation. Finally, we validated both our gender and country name inference methods based on annotations obtained from two independent raters on a sample of 100 users. We found agreement between the raters’ annotations and the one given by our method for 79% of the cases for gender, and 86% of the cases for country names ($\kappa = .77$). In our candidate disclosure and candidate control datasets, we were able to infer binary gender for 325,873 candidate mental health (of which 59% were female) and 439,224 candidate control users (of which 46% were female) respectively, and country information for 131,890 and 328,468 users for the same two groups respectively.

Within the scope of this paper, we restrict our attention to four most populated countries where English is a predominant form of expression — western countries US, UK (GB), and majority world countries India (IN), and South Africa (ZA). All of these countries fare in the 25 countries with most population⁵. Focusing on these countries provides us with a lens to examine cross-cultural differences in the disclosures of mental illnesses on social media.

Handling Population and Internet Penetration Bias. We note that the populations of the four countries are widely different, along with their overall reported internet penetration rates⁶. Hence we devise a subsampling strategy to filter users belonging to one of the four countries, from the set of users in the candidate disclosure and the control datasets with inferable country information. First, for population based subsampling, we use the inverse of the population ranks of the countries⁴ as the respective rates of sampling. Then we use the internet penetration percentages of the countries⁵ to randomly sample that fraction of users from the population subsampled sets. In this manner, across both the datasets, we obtained 211,132 Twitter users from the US, 61,816 users from the UK, 10,808 from IN, and 5,769 from ZA.

METHODS

Obtaining Genuine Mental Health Disclosure Data

We note that the candidate mental health disclosure data sample is prone to significant noise. It is possible that although a user uses one of the mental health keyphrases in their Twitter post, it may not indicate a genuine disclosure (e.g., “when I have to wake up at 6am, I feel like killing myself” does not indicate a person’s real intention of taking their life).

³<https://dev.twitter.com/overview/documentation>

⁴<https://code.google.com/p/cld2/>

⁵[https://en.wikipedia.org/wiki/List_of_countries_by_population_\(United_Nations\)](https://en.wikipedia.org/wiki/List_of_countries_by_population_(United_Nations))

⁶<http://www.internetlivestats.com/internet-users-by-country/>

To eliminate such users, we adopt a semi-supervised machine learning method [58] in which we compare the language of each user in our Twitter dataset, with the language of a self-identified set of social media users suffering from a mental illness. For the purpose, we obtain a large sample of 79,833 posts from 44,262 users made on the Reddit subcommunities r/depression, r/mentalhealth, and r/SuicideWatch between February and November 2014. We use this dataset as a weak signal of the language used by individuals identifying with a form of mental illness. Prior work has also indicated that the disclosures made on these forums are genuine disclosures of mental illnesses, and have also been validated through consultation with a psychiatrist [16].

Our approach proceeds as follows:

Step 1. We create Twitter user-centric vector-representations by collating all of their posts – this would give us as many vectors as the number of users in the candidate disclosure sample. We also similarly create a single vector representation by collating all posts in our Reddit dataset.

Step 2. Next we establish comparative validity across the Twitter vectors and the Reddit vector. This is an important step because the language of Twitter and Reddit cannot be directly compared due to the unique affordances of each site, and the seeming differences in the demographics of the two⁷. For each of these vectors, we perform linguistic normalization of tokenized items in them⁸. Then we compute the average automated readability index (ARI [44]) on all of the normalized Twitter vectors, and the Reddit vector. We observe the ARI differential between the two to be $\sim 6.7\%$ ⁹, indicating that the languages are close to standardized internet speak [25], and hence comparable following normalization.

Step 3. Next we build n -gram language models ($n = 3$) for both the normalized Twitter vectors, and the single Reddit vector. We then determine cosine similarity scores between the language models of each normalized Twitter vector and the Reddit vector.

Step 4. Finally, we obtain the distribution of cosine similarities over all normalized Twitter vectors (see Figure 1(a)). We construct the “genuine disclosure dataset” of users to be those vectors (of users) for whom the cosine similarity of their language models with that of Reddit’s is *greater than or equal to* the median similarity across all vectors (median distance = .71 $\sigma = .157$). Our final dataset of genuine mental illness disclosures consisted of 231,611 users; we will refer to this set as the MID users.

⁷Twitter posts are only 140 characters long, whereas Reddit posts tend to be more detailed and longer; Twitter posts often contain highly irregular syntax and nonstandard use of English.

⁸For normalization, we perform syntactic disambiguation of the Twitter vectors based on PoS (part-of-speech) tagging [29]. This method converts non-standard forms into dictionary forms, expands abbreviations, handles misspellings, punctuation/omission errors, phonetic spelling and intentional misspelling for verbal effect: “rite” to “right”, “sooorryy” to “sorry”.

⁹ARI for the normalized Twitter posts was 15.1, whereas that for Reddit was 16.2.

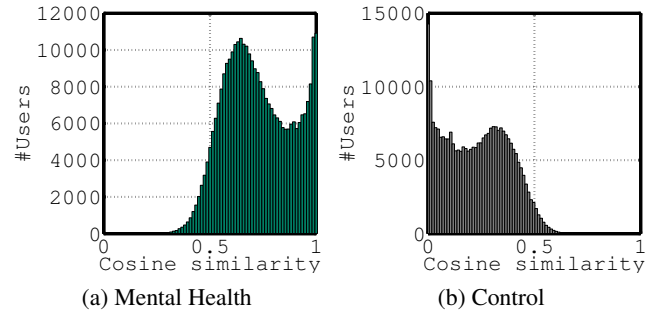


Figure 1: Distribution of the number of Twitter users over their cosine similarities with Reddit mental health content. The distributions are shown for both the candidate mental health disclosure as well as the candidate control sample.

Genuine Disclosures
I feel my mental state slowly declining. <i>I have been depressed.</i>
I'm in a dark place today. It's strange, after admitting that <i>I am depressed</i> I can recognize the slippery slope that I'm on earlier.
All I want to do is die, <i>I tried suicide</i> six times and couldn't even do it right.
I am open about most things, the latest is that <i>I have Mental Illness</i> and this has made my life extremely difficult to live and be happy
Noisy Disclosures
My headphones broke so now <i>I'm ending my life.</i> Bye
The kindergarten class I'm in charge of today doesn't know what Little Sally Walker is..... <i>I am depressed</i>
lost my headphones brb <i>killing myself</i>
ok im literally <i>killing myself</i> bc my global warming class fuck this final paper

Table 2: Example (paraphrased) genuine and noisy mental health disclosure users from Twitter, obtained using our semi-supervised learning approach.

Expert Verification of Mental Health Disclosures. Next we qualitatively verified whether posts from these users indeed were self-disclosures of mental health challenges. For the purpose, we consulted a licensed psychologist, and also included two researchers, who were familiar with mental health content shared on social media. Over a random sample of 100 posts compiled from the timeline of 100 randomly selected MID users, we obtained independent binary annotations on whether a post was likely to be related to mental health. The Fleiss’ κ for interrater agreement was found to be high, .87, along with an accuracy of 96% in distinguishing users who engage in genuine disclosures from those who do not. This establishes adequacy of our approach. Table 2 gives some paraphrased mental health disclosure posts of Twitter users who were identified to be genuine mental health disclosers by our approach.

In Figure 2 we show the pipeline of steps involved in our approach of arriving at this final dataset.

Developing an Accurate Control Dataset

Next we also note that it is possible that our candidate control dataset includes users who engage in mental illness in their posts, however did not use any of the keyphrases from Table 1. To eliminate such users, like above, we compare the

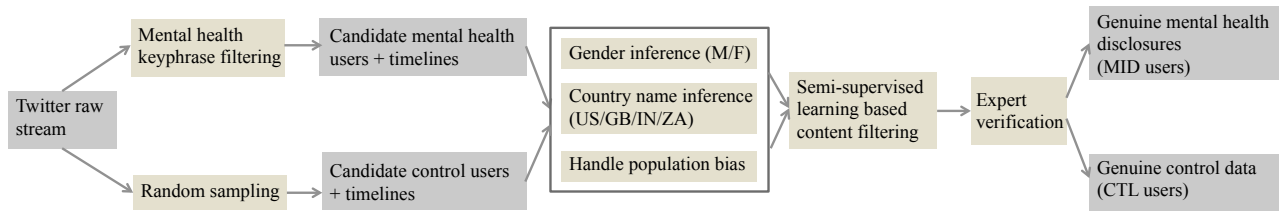


Figure 2: Flow chart of the steps involved in obtaining the MID (mental health disclosure) and CTL (control) user cohorts starting from the raw stream of Twitter data.

language used in the Twitter posts of these users with that of the Reddit mental health posts. However in this case, we are interested in the users whose language is most distinct (or least similar) from that used in the Reddit content. Hence our final control dataset is obtained by filtering for those users whose language model based similarity is *less than* the median similarity across all control user vectors (238,860 users; median distance=.38; $\sigma = .125$). See Figure 1(b) for a distribution of the cosine similarities. We will refer to this dataset as CTL users (ref. Figure 2 for the approach).

Quantifying Differences in Disclosures

In this subsection, we present methods to quantify the differences between the disclosure characteristics of females and males, and individuals reported to be from one of the four countries of interest: US, GB, IN, and ZA in the MID dataset.

Linguistic Measures

Language is a powerful source of expression [26]. A rich body of work, such as Boroditsky et al. [6] showed how the perception of objects in different languages can relate to as well as impact one’s social and psychological status. It is recognized that language, specifically one’s native language, shapes and drives one’s thoughts, actions, and social relationships [10]. Further, it is established that cross-cultural and sex differences exist in one’s underlying thought processes [21, 55]. For instance, according to Kövecses [31], cultural models are known to define one’s emotional concepts.

To quantify gender and cross-cultural dimensions in the language of individuals who engage in mental health disclosure on social media, we propose three categories of measures: (1) **affective attributes**, (2) **cognitive attributes**, and (3) **linguistic style attributes**. Measures belonging to all of these attribute categories are largely based on the psycholinguistic lexicon LIWC [42], and were motivated from prior literature that examines associations between the behavioral expression of individuals and their psychological distress, including vulnerability to mental illness [9, 14]. Specifically, with LIWC, we are able to study the psychological value of language in gender and culture subgroups—such as, parts of speech that include pronouns, articles, prepositions, conjunctives, and auxiliary verbs [9].

(1) We consider two measures of affect derived from LIWC: positive affect (PA), and negative affect (NA), and four other measures of emotional expression: *anger*, *anxiety*, *sadness*, and *swear*. Literature in mental health [49] identifies emotional expression to be key to characterizing one’s psychological vulnerability.

(2) We use LIWC to define the cognitive measures as well: (a) cognition, comprising *cognitive mech*, *discrepancies*, *inhibition*, *negation*, *death*, *causation*, *certainty*, and *tentativeness*; and (b) perception, comprising set of words in LIWC around *see*, *hear*, *feel*, *percept*, *insight*, and *relative*. Quantifying one’s cognition and perception, as manifested linguistically, can provide insights into emotional stability and cognitive complexity—these attributes are important with regard to understanding one’s mental well-being [19].

(3) Next, we consider four measures of linguistic style: (a) **Lexical Density**: consisting of words that are *verbs*, *auxiliary verbs*, *nouns*, *adjectives* (identified using NLTK’s [4] POS tagger), and *adverbs*. (b) **Temporal References**: consisting of *past*, *present*, and *future* tenses. (c) **Social/Personal Concerns**: words belonging to *family*, *friends*, *social*, *work*, *health*, *humans*, *religion*, *bio*, *body*, *money*, *achievement*, *home*, and *sexual*. (d) **Interpersonal Awareness and Focus**: words that are *1st person singular*, *1st person plural*, *2nd person*, and *3rd person* pronouns. Together, linguistic styles are known to indicate one’s underlying psychological processes (lexical density), personality (temporal references), social support and connectivity (social/personal concerns), and awareness of one’s surroundings and environment (interpersonal focus). Prior work identifies all of these cues to be valuable in understanding mental health, in both offline and online contexts, including social media [43].

Topic Modeling

Our second method for comparing mental illness disclosures uses a topic model, which have been commonly employed to analyze health data [41]. We obtain topics by running Latent Dirichlet Allocation (LDA) [5] over all posts. We pre-processed the data by removing a standard list of Twitter-specific stop words, words with very high frequency ($> 0.25 \times \text{datasize}$), and words that occur fewer than five times. Thereafter, we used Gensim’s implementation of on-line LDA from [23]. We used the default hyper-parameter settings and 100 topics, which we determined based on the value of average corpus likelihood over ten runs.

To measure topic differences in one cohort (e.g., IN MID users) over the other (e.g., UK MID users), we first compute the posterior probability of each topic separately for all posts in both cohorts. We then compute three comparison metrics: (1) the rate of change for each topic, given as the difference between the posterior topic probabilities of the cohorts, divided by the probability of the first cohort; (2) the pointwise mutual information between the posterior topic probabilities of the same cohorts; and (3) the Spearman’s rank correlation

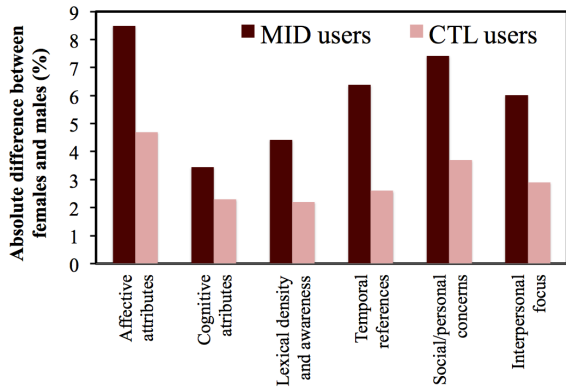


Figure 3: Mean absolute differences between female and male MID and CTL users per the various categories of linguistic measures. Difference for a specific measure is calculated as the ratio of the difference between the values of the measure for females and males, to the value of the measure among males.

between the topic distributions for the two cohorts. Additionally, we compare all gender and culture cohorts based on significance tests (e.g., Mann Whitney U test for gender, and the Kruskal Wallis test for cultural differences).

We also present a method to qualitatively examine the differences between the topics used by different MID user cohorts. For the purpose, two researchers familiar with mental health content on social media independently inspected the words associated with each of the topics given by the above topic model. They used a semi-open coding approach to develop a codebook and extracted descriptive topical themes for the topics (Cohen’s $\kappa=.74$). During the codebook development, the two annotators referred to prior literature on gender and cultural differences in mental health [46, 20, 52]. In the results section, we will present an examination of these qualitative differences.

RESULTS

Gender Differences

Linguistic Differences

Based on Table 3, we observe considerable differences, in terms of the linguistic measures, in the Twitter posts of female and male MID users.

Affective Attributes (Gender). Starting with the measures of affect, female MID users show higher sadness ($z = -39.0$; effect size Cohen’s $d = 1.1$; 15.4% higher) and anxiety ($z = -26.4$; Cohen’s $d = 1.0$; 10.7% higher). Prior literature indicates expression of these emotions to be associated with depression symptoms such as mental instability and helplessness, loneliness, and restlessness [13].

“My health has been defining me lately. #depression has invaded my peace and #anxiety has exhausted my thoughts. Pain isn’t always physical.” (↑ female)

“Why am I even here ... No one needs or wants me ... I’m useless” (↑ female)

Whereas we find that male MID users express more NA ($z = 9.8$; Cohen’s $d = .5$; 2.6% higher), anger ($z = 15.8$; Cohen’s

	μ (male)	μ (female)	z	p
Affective attributes				
PA	0.0647	0.0693	-33.928	***
NA	0.0148	0.0144	9.858	**
anger	0.0215	0.0203	15.879	***
anxiety	0.0041	0.0046	-26.451	***
sadness	0.0077	0.0089	-39.002	***
swear	0.0144	0.0130	22.562	***
Cognitive attributes				
Cognition				
cognitive_mech	0.1363	0.1338	19.832	***
inhibition	0.0407	0.0380	29.935	***
causation	0.0147	0.0144	9.553	**
certainty	0.0060	0.0059	4.763	*
tentativeness	0.0221	0.0205	42.423	***
Perception				
see	0.0140	0.0134	17.609	***
hear	0.0087	0.0085	8.042	**
feel	0.0092	0.0087	17.496	***
percept	0.0337	0.0326	14.963	***
insight	0.0229	0.0225	5.993	*
relative	0.0183	0.0180	8.797	**
Lexical Density and Awareness				
verbs	0.1553	0.1528	21.155	***
auxiliary verbs	0.0710	0.0696	13.084	**
articles	0.0320	0.0277	37.686	***
prepositions	0.0642	0.0684	-18.451	***
conjunctions	0.0362	0.0368	-4.664	*
adverbs	0.0483	0.0498	-18.163	***
inclusive	0.0218	0.0223	-7.161	**
exclusive	0.0198	0.0208	-16.967	***
Temporal references				
present tense	0.1077	0.1114	-33.391	***
future tense	0.0071	0.0064	35.784	***
Social/Personal Concerns				
family	0.0048	0.0054	-23.292	***
friends	0.0032	0.0037	-27.404	***
social	0.0937	0.0943	-4.797	*
health	0.0092	0.0097	-18.017	***
religion	0.0042	0.0039	10.615	**
bio	0.0429	0.0452	-21.303	***
body	0.0150	0.0155	-8.910	**
achievement	0.0157	0.0170	-32.293	***
home	0.0047	0.0055	-36.581	***
sexual	0.0156	0.0148	14.940	***
death	0.0037	0.0038	-5.406	*
Interpersonal focus				
1st p. singular	0.0489	0.0443	26.127	***
1st p. plural	0.0035	0.0038	-16.785	***
2nd p.	0.0185	0.0199	-8.021	**
3rd p.	0.0069	0.0071	-6.974	**

Table 3: Differences between posts from female and male MID users based on linguistic measures. Statistical significance (based on Wilcoxon rank sum tests, with degrees of freedom $df = 194, 406$) is reported following Bonferroni correction. Only significant measures are shown.

$d = .6$; 5.3% higher), and use more swear language ($z = 22.5$; Cohen’s $d = .8$; 9.5% higher) in their posts.

“The past week has been horrible. Depression is robbing me of the peace I have felt. I’m isolated in a house full of people.” (↑ female)

“Honestly fuck everyone, y’all gonna miss me when I’m gone” (↑ male)

Interestingly, female users also tend to use more PA in their shared content ($z = -33.9$; Cohen’s $d = 1.1$; 7.1% higher),

perhaps as a way to demonstrate a positive outlook publicly, despite the mental health challenge that they might be experiencing.

“There’s something about those eyes helping to wake me up everyday that makes the days brighter.” (↑ female)

We also report the extent to which the female and male CTL users differ, along all of these affective attributes. Per Figure 3, this mean difference is found to be only 4.7%, which is considerably low compared to 8.5% in the case of the MID cohort.

Cognitive Attributes (Gender). Next, female MID users show lowered cognition and perception, in other words, greater cognitive impairment in their Twitter posts compared to male MID users. For instance, inhibition ($z = 29.9$; Cohen’s $d = 1.0$) is lower in females relative to males by 6.5%. Through lower usage of certainty words, female peers tend to demonstrate heightened emotional instability ($z = 4.7$; Cohen’s $d = .5$).

“So now here I am, confused and full of questions. Am I born to lose or is this just a lesson?” (↑ male)

The female and male users in the CTL cohort, however, differ by only 2.3% across the various cognition and perception attributes, as shown in Figure 3.

Lexical Density and Awareness (Gender). Next, lexical density of the social media content of female MID users is higher compared to their male peers, as observed through the usage of prepositions ($z = -18.4$; Cohen’s $d = .7$; 6.4% higher), conjunctions ($z = -4.6$; Cohen’s $d = .5$; 1.6% higher), adverbs ($z = -18.1$; Cohen’s $d = .9$; 3.0% higher), inclusive ($z = -7.1$; Cohen’s $d = .5$; 2.3% higher), and exclusive ($z = -16.9$; Cohen’s $d = .7$; 4.9% higher). However, compared to the male MID users, they show lowered awareness of objects and their surroundings, as measured via the proportion of verbs ($z = 21.5$; Cohen’s $d = .6$; 1.6% lower), auxiliary verbs ($z = 13.1$; Cohen’s $d = .9$; 1.9% lower), articles ($z = 37.6$; Cohen’s $d = 1.1$; 13.4% lower) in Twitter posts. Note that, in the case of the female and male users in the CTL cohort, per Figure 3, the mean difference across all of these measures is observed to be only 2.2%. This indicates that the female and male MID users show differences beyond that accounted for in the control sample.

Temporal References (Gender). Female MID users tend to be more focused on the here and now, due to their greater use of present tense words ($z = -33.3$; Cohen’s $d = .8$; 3.4% higher). On the other hand, male MID users show a greater future orientation compared to the females, via the usage of more future tense in their Twitter language ($z = 35.7$; Cohen’s $d = .9$; 9.2% higher). Such differences in temporal references in language are however not observable between the female and male CTL cohort. Compared to a mean difference of 6.4% in the case of the MID cohort, it is only 2.6% for the CTL users.

Social/Personal Concerns (Gender). There are a variety of differences in the social and personal concerns that manifest in the Twitter posts of female and male MID users. First, male

MID users display lower sense of achievement ($z = -32.2$; Cohen’s $d = 1.1$; 8.1% lower)—a known sign of lowered self-esteem [8]. On the other hand, female MID users express greater concern about health ($z = -18.0$; Cohen’s $d = 1.1$; 6.0% higher) and body ($z = -8.9$; Cohen’s $d = .9$; 2.7% higher) compared to their male peers. This might indicate their greater self-awareness of their wellness status or perceptions of their physical health.

“Over the past 2 years I have been hit with physical and mental pain. The pain is real. It is still there.” (↑ female)

“My stomach sinks everytime.” (↑ female)

An interesting finding here is the observation that male MID users exhibit lower use of social ($z = -4.7$; Cohen’s $d = .1$), friends ($z = -27.4$; Cohen’s $d = 1.1$), family ($z = -23.2$; Cohen’s $d = .9$), and bio words ($z = -21.3$; Cohen’s $d = 1.0$). This may imply that these users are less socially concerned or bothered. By the same token, the female peers might be using such language more extensively in their Twitter posts in order to explicitly seek help from their social networks or to feel supported.

“Hard to really feel sick with this support group. #Family” (↑ female)

“I miss having someone, a friend to talk to all night” (↑ female)

Finally, male MID users show a greater interest in religious discussions compared to females ($z = 10.6$; Cohen’s $d = .8$; 5.1% higher).

“God is working things out for you, even when you don’t feel it. Have faith and be thankful.”

While analyzing the social and personal concerns expressed by the female and male users in the CTL cohort, based on Figure 3 we do not observe such extensive differences: the mean difference between the cohorts per these measures is only 3.7%.

Interpersonal Focus (Gender). Increased use of first person singular pronouns ($z = 26.1$; Cohen’s $d = 1.1$; 10.2% higher) in the posts of male MID users shows their self-focused behavior and disclosure of personal stories.

“I decided yesterday evening to go back to this thinking place. there was no one else there, just me. I had let my parents know where I was” (↑ male)

Additionally, lower use of second person pronouns ($z = -8.0$; Cohen’s $d = .5$; 3.0% lower) as well as that of third person pronouns ($z = -6.9$; Cohen’s $d = .8$; 3.4% lower) in the content of the male users tell us that they tend to be less interactive, and engage in lesser discourse about others. For the female and male users in the CTL cohort (Figure 3), the interpersonal focus measures account for only 2.9% of the difference.

Topical Differences

Digging deeper, between the female and male gender MID cohorts on Twitter, we find significant differences in topics of the posts. Figure 4(a) gives a distribution of the number of topics over their respective z -statistic values (based on a Mann-Whitney U test and Kruskal Wallis tests). These tests examine the differences in the likelihood of prevalence of top-

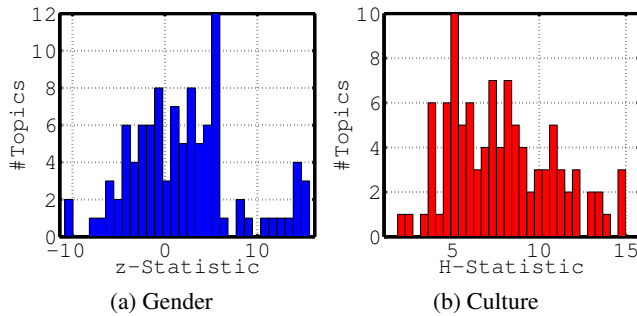


Figure 4: Distribution of the number of topics per statistical significance tests comparing topic distributions between gender and culture groups. For (a) we use the Mann-Whitney U test, while for (b) we use the Kruskal Wallis test.

ics in the posts of female and male MID users. We find that several topics are associated with highly negative or highly positive z -scores ($p < .0005$; $df = 99$), indicating significant topical differences between the gender groups.

Next, we find that, per the three topic comparison metrics, the Spearman rank correlation for topical comparison between females and males is $\rho = .31$ ($p < .01$ based on a Mann Whitney U test), the pointwise mutual information (PMI) is $.38$ ($p < .01$ based on a Mann Whitney U test), while the mean percent difference in the likelihoods of the 100 topics is 59.3% ($p < .001$ based on a Mann Whitney U test). However, the respective metrics between males and females in the CTL cohort are $\rho = .14$, $PMI = .61$, and a 33.6% difference in the topic likelihoods. This reveals significant topical differences for the MID cohort, above and beyond the differences that exist due to underlying gender based latent factors.

We now present some insights we gained by examining the posts associated with the topics that distinguish the female and male cohorts most distinctively. We first identify the two topics that are more likely to be prevalent in posts from male MID users than female. The posts associated with topic #13, for instance, express contemplative negative thoughts and hopelessness (e.g., ‘never’, ‘low’, ‘mess’, ‘lonely’, ‘drunk’):

“Sometimes I wonder if **anyone still** looks out for me. I am a **mess** that **nobody** wants to clean up. I’m a **wreck**”

“Some things are better left **unsaid**. **Lonely** nights make for **long** nights. Being **drunk** sometimes makes it **easier**.”

The second topic more prevalent among male MID users is topic #57, and it shows detachment from the social realm and hesitation to seek help (e.g., ‘invisible’, ‘help’, ‘ask’, ‘relationship’):

“If I were going to **kill** myself, I wouldn’t **tell** anyone. If I’m already **invisible**, why see me to favor your own self righteousness?”

“I love my **relationship** but it’s fun to wonder who was just too shy or too **afraid** to ask me out. I have **suspicious** but I’m self centered.”

Contrastively, we examine the top two topics that manifest more extensively in posts from female MID users, compared to the males. Posts associated with topic #86 indicate the

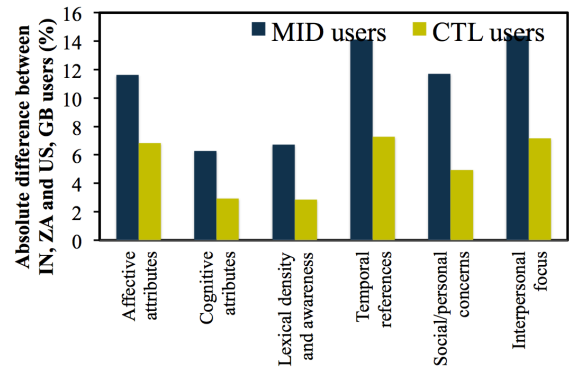


Figure 5: Mean absolute differences between IN, ZA and US, GB MID and CTL users per the various categories of linguistic measures. Difference for a specific measure is calculated as the ratio of the difference between the aggregated values of the measure for IN, ZA and US, GB users, to the value of the measure among US, GB users.

presence of a positive outlook, motivational spirit to cope with mental health challenges, and a desire for disclosure and help seeking (e.g., ‘fear’, ‘rejection’, ‘bury’, ‘lose’, ‘stay’, ‘say’, ‘okay’):

“you’re afraid to tell **people** how you feel because you **fear rejection**, so you **bury** it deep inside yourself where it only **destroys** you more.”

“Sadly, you can’t **lose** what you never had, you can’t keep what’s not yours, and you can’t **hold** on to something that doesn’t want to **stay**.”

Finally, through topic #45, the topic with the second highest likelihood of prevalence in female MID users’ posts over males, the female MID users share personal experiences around mental illness, including self-assessments and self-realization (e.g., ‘pain’, ‘control’, ‘realization’, ‘reminder’):

“I used to hurt **myself**, because it was the only **pain** I could **control**.”

“Daily **reminder** that life has **taught** me: don’t waste your time **worrying** about the **people** who don’t like you, you don’t live to please anyone.”

Cross-Cultural Differences

Linguistic Differences

The second part of our empirical investigation focused on characterizing the usage of different linguistic measures across the four cultures: western countries US, GB, and majority world countries IN and ZA. In Table 4, we report the mean values of each statistically significant linguistic measure in each of the four cultures, as well as the results of Kruskal-Wallis significance tests comparing them. For the purposes of the ensuing discussion, we performed post hoc multiple comparison Wilcoxon tests to report which specific pairs of cultures differed most in terms of the different linguistic measures. Further, we discuss our results by comparing the majority world cultures IN and ZA, with the western cultures US and GB. The linguistic differences between these two cohorts in the CTL group are shown in Figure 5.

	μ (US)	μ (GB)	μ (IN)	μ (ZA)	H	p
Affective attributes						
PA	0.062	0.063	0.070	0.066	104.3	**
NA	0.014	0.014	0.014	0.013	17.5	*
anger	0.020	0.019	0.018	0.016	127.9	***
anxiety	0.004	0.005	0.004	0.004	87.9	**
sadness	0.008	0.009	0.007	0.007	79.6	**
swear	0.014	0.012	0.010	0.009	239.7	***
Cognitive attributes						
Cognition						
cognitive_mech	0.135	0.133	0.142	0.139	150.8	***
negation	0.023	0.023	0.025	0.026	87.2	**
discrepancies	0.018	0.018	0.018	0.018	25.5	*
causation	0.015	0.014	0.016	0.015	37.0	*
certainty	0.016	0.016	0.017	0.017	41.2	*
tentativeness	0.021	0.021	0.023	0.021	142.0	***
Perception						
see	0.013	0.012	0.014	0.014	69.9	**
hear	0.009	0.008	0.008	0.008	314.3	***
feel	0.008	0.009	0.009	0.009	52.7	**
percept	0.030	0.032	0.032	0.034	120.1	***
insight	0.023	0.023	0.025	0.024	40.7	*
Lexical Density and Awareness						
verbs	0.149	0.149	0.152	0.152	26.7	*
auxiliary verbs	0.069	0.070	0.072	0.074	215.2	***
articles	0.030	0.030	0.033	0.034	186.7	***
prepositions	0.067	0.066	0.072	0.072	210.6	***
adverbs	0.048	0.051	0.046	0.049	208.8	***
exclusive	0.020	0.019	0.021	0.021	38.4	*
Temporal references						
past tense	0.030	0.031	0.025	0.025	369.9	***
future tense	0.007	0.007	0.008	0.008	103.8	**
Social/Personal Concerns						
family	0.005	0.005	0.004	0.004	154.1	***
social	0.101	0.101	0.093	0.082	1613.5	***
health	0.010	0.010	0.009	0.009	124.5	***
religion	0.004	0.005	0.004	0.004	162.3	***
bio	0.044	0.044	0.043	0.039	27.8	*
body	0.015	0.015	0.015	0.013	56.6	**
work	0.007	0.008	0.006	0.007	281.9	***
home	0.006	0.006	0.004	0.005	158.2	***
sexual	0.014	0.015	0.013	0.012	131.0	***
death	0.004	0.004	0.003	0.003	38.4	*
Interpersonal focus						
1st p. singular	0.038	0.041	0.046	0.041	128.0	***
2nd p.	0.023	0.023	0.018	0.015	290.1	***
3rd p.	0.007	0.007	0.007	0.006	49.9	**

Table 4: Differences between posts of MID users from the four cultures: US, GB, IN and ZA, per different linguistic measures. Statistical significance (based on Kruskal Wallis tests, with $df = 147, 187$) is reported following Bonferroni correction. Only significant measures are shown.

Affective Attributes (Culture). First, MID users from IN and ZA express relatively higher PA ($H = 104.3$; eta-squared effect size $E_r^2 = .45$) and lower NA ($H = 17.5$; eta-squared effect size $E_r^2 = .19$), anger ($H = 127.9$; eta-squared effect size $E_r^2 = .53$), anxiety ($H = 87.9$; eta-squared effect size $E_r^2 = .42$) and sadness ($H = 79.6$; eta-squared effect size $E_r^2 = .44$) compared to those from US and GB. In other words, users from the latter two cultures, US and GB, tend to express emotional content that is more emotionally outspoken in comparison with the majority world cultures (11.6% higher). We conjecture those in the majority world cohort might be expressing more positivity as a coping mechanism or as a face-saving mechanism [14]. This difference is also

considerably larger than the affective differences observed for the cultures in the CTL cohort (6.8%); see Figure 5.

“I hate myself with a burning passion right now, I should be out on the sled” (↑ US, GB)

“I’ve been crying all afternoon I feel like a bag of shit” (↑ US, GB)

“Being unhappy is the only thing I do right.” (↑ US, GB)

Cognitive Attributes (Culture). We also observe greater manifestation of cognitive impairment and emotional instability among MID users from US and GB compared to those from IN and ZA. Mean difference between the groups is 6.3%, while the same in the case of the CTL cohort is only 2.9%. This is observable especially from the lower use of measures of cognitive_mech ($H = 150.8$; eta-squared effect size $E_r^2 = .64$), certainty ($H = 41.2$; eta-squared effect size $E_r^2 = .33$), discrepancies ($H = 25.5$; eta-squared effect size $E_r^2 = .31$), and percept ($H = 120.1$; eta-squared effect size $E_r^2 = .55$).

“you think youre doing okay, then suddenly its a nighttime and you feel alone and youre not sure how to distract yourself anymore” (↑ IN, ZA)

Lexical Density and Awareness (Culture). The lexical density measures show lower values in the posts of US and GB MID users, relative to those in the case of IN and ZA users. That is, the former group uses fewer prepositions ($H = 210.6$; eta-squared effect size $E_r^2 = .63$), adverbs ($H = 208.8$; eta-squared effect size $E_r^2 = .69$), and exclusive words ($H = 38.4$; eta-squared effect size $E_r^2 = .37$). Additionally, they also show reduced awareness of their social and environmental context, as measured via the usage of articles ($H = 186.7$; eta-squared effect size $E_r^2 = .49$). We note this to be more than twice relative to the differences in terms of lexical density and awareness observed in the two groups within the CTL cohort (ref. Figure 5).

Temporal References (Culture). Next, the MID users from IN and ZA show an increased future orientation in their Twitter posts, compared to those from US and GB, as indicated in the usage of future tense words ($H = 103.8$; eta-squared effect size $E_r^2 = .41$). On the other hand, the latter group discussed more of their past experiences and events, via the use of past tense words ($H = 369.9$; eta-squared effect size $E_r^2 = .65$). This difference is also much larger than what is observed in the case of the CTL cohort (7.2%, compared to 14.1% in the former).

“I felt so lonely and I started to cry but nobody understood.. Nobody saw just how broken I really was.” (↑ US, GB)

Social/Personal Concerns (Culture). Increased levels of social concerns, measured via family ($H = 154.1$; eta-squared effect size $E_r^2 = .58$), and social words ($H = 1613.5$; eta-squared effect size $E_r^2 = .87$), also tend to be observable in the posts shared by MID users from western cultures US and GB, compared to IN and ZA. These users also discuss more able health ($H = 124.5$; eta-squared effect size $E_r^2 = .56$), body ($H = 56.6$; eta-squared effect size $E_r^2 = .32$), and bio ($H = 27.8$; eta-squared effect size $E_r^2 = .27$) compared

to the IN and ZA users. We conjecture the latter (majority world) group to be more self-conscious in utilizing a public platform like Twitter to discuss about the health aspects relating to a stigmatized illness.

“I miss friendships with everyone that has ever held me close. Please be patient with me and I promise you that I’ll be back!” (↑ IN, ZA)

“I go through these spouts of depression and I could lay in bed for weeks. I just dont care about a single thing about this health condition” (↑ US, GB)

We further observe that the IN and ZA MID users converse less about sensitive or ‘taboo’ topics like religion ($H = 162.2$; eta-squared effect size $E_r^2 = .68$), death ($H = 38.4$; eta-squared effect size $E_r^2 = .28$), and sexual ($H = 131.0$; eta-squared effect size $E_r^2 = .47$) in their Twitter posts, compared to their peers from US and GB. This might indicate a latent social norm in the former cohort to avoid sensitive discourse on a publicly accessible platform like Twitter. Overall, the differences in these two sets of MID culture groups is much higher (11.7%) compared to the difference observable in the CTL cohort (4.9%).

Interpersonal Focus (Culture). Finally, we observe the MID users from the western cultures US and GB to demonstrate a greater tendency of social interaction (use of second person pronoun words ($H = 290.1$)), as well as attention to people around them (use of third person pronouns ($H = 49.9$; eta-squared effect size $E_r^2 = .36$)). However, the users from IN and ZA express greater self pre-occupation and self-attentional focus, as measured in their use of more first person singular pronouns ($H = 128.0$; eta-squared effect size $E_r^2 = .54$). We also note that the aggregated interpersonal focus in these two sets of MID users (14.4%) is much higher in contrast to the CTL users (7.2%).

“I suffered and was embarrassed to talk. When I spoke up my suffering lifted. There are people who will listen.” (↑ IN, ZA)

Topical Differences

We now move over to examining the differences in the topics prevalent in the four cultures based on the LDA topic model. We first report on the statistical differences between the likelihoods of prevalence of topics in the four culture groups. In Figure 4(b) we show a distribution of the number of topics over their respective H -statistic values, based on Kruskal Wallis tests of comparing the topics of the four cultures. The figure indicates that a large number of topics have large H -statistic values (e.g., over 10; $p < .0005$; $df = 96$), in turn revealing topical differences between the culture groups.

Next, investigating these topical differences further, we present the differences per the three different metrics: (1) Mean Percentage Difference in Topic Likelihood; (2) Spearman’s Rank Correlation; and (3) Pointwise Mutual Information (PMI). The results are given in Table 5. Across all pairwise differences for the three metrics, we observe that US and GB MID users aggregatively differ less topically compared to US-IN, US-ZA, GB-IN or GB-ZA. For instance, the Spearman’s rank correlation of topics used by US and IN MID users is only .17, however it is much higher for US-GB (.43). This shows that US-GB topics are more similar to each other

in their usage, compared to the topics prevalent in US and IN posts. Another example is the measure of pointwise mutual information (PMI) for US and ZA. It is only .18, indicating their there are considerable differences in the likelihoods of various topics across these two cultures.

Examining these numbers computed on the CTL users, we observe less significant differences in topic likelihoods. Together, these indicate topical differences in the posts of MID users from the four cultures to be higher than that observed in their CTL cohort peers.

	US	GB	IN	ZA
<i>Mean Percentage Difference in Topic Likelihood (%)</i>				
US	0 (0)	22.1 (24.5)	84.4 (69.6) ***	71.8 (64.2) ***
GB		0 (0)	58.6 (50.2) **	49.2 (37.5) **
IN			0 (0)	44.6 (34.9) *
ZA				0 (0)
<i>Spearman’s Rank Correlation</i>				
US	1 (1)	0.436 (.519) **	0.174 (.295) ***	0.209 (.352) ***
GB		1 (1)	0.430 (.502) **	0.535 (.683) *
IN			1 (1)	0.321 (.504) **
ZA				1 (1)
<i>Pointwise Mutual Information (PMI)</i>				
US	1 (1)	0.491 (.638) *	0.123 (.224) ***	0.183 (.327) ***
GB		1 (1)	0.285 (.391) **	0.391 (.482) *
IN			1 (1)	0.313 (.542) **
ZA				1 (1)

Table 5: Topical differences between the posts of MID users from the four cultures, based on three metrics. Corresponding numbers in gray indicate the same for the CTL users.

We now discuss the context of usage of the top two topics that have a greater likelihood of occurrence in western cultures (US and GB posts) versus that in the majority world cultures (IN and SA posts shared by the MID users). Topic #29 centers around words like ‘kill’, ‘stop’, ‘hate’, ‘pain’, ‘life’ all of which generally revolve around self-deprecating, self-critical thoughts and self-destruction.

“my god **life** makes me want to **kill** myself”

“**everything** good is taken from me. **EVERYTHING**”

“I’d rather my **heart stop** beating and my lungs **fail** me that continue living through this constant **pain**”

Next, topic #54 includes words like ‘alone’, ‘lonely’, ‘people’, indicative of loneliness and lack of social support. US and GB MID user posts associated with high likelihood of this topic tend to bear a tone of decreased self-esteem, and greater self-loathing.

“**lonely** even when I’m not **alone**”

“no one will ever be able to **understand**”

“my **family** is supposed to be loving and accepting and they are the ones who **trigger** me the most”

On the other hand, the two top topics that are more prevalent in MID users from majority world cultures, IN and ZA, compared to those from western cultures, US and GB, revolve around significantly different content. For instance, topic (#17) manifests confessions and regrets of individuals (e.g., ‘faith’, ‘regret’, ‘strong’).

“I know I need to be **strong** but its just **hard** to now a days.
Regret”

“Will I ever **beat** this? I was doing so **good** but what happened to me”

“I put so much **faith** in you being the one, I was so **excited**. It’s just **depression, paranoia** and **panic** got the better of me.”

Through topic #86, the MID users in IN and ZA express be-reavement and marginalization due to the stigma associated with mental illness (e.g., words ‘pretend’, ‘ashamed’, ‘embarrassed’, ‘struggle’). We observe that, these MID users may be trying to resolve their problems through posts associated this topic, believing that mental health can be maintained by avoiding bad thoughts and exercising will power. In fact, Wang et al. [54] have suggested that talking to another individual or group (e.g., a mental health worker) about psychological problems may be viewed by non-American sufferers as bringing disgrace on the family.

“Beyond tired. I want this to end. I can’t **pretend** anymore, need to be **strong**”

“I’m still so **embarrassed**, but I’m **hurt** more by him not being there for me. I don’t want to **struggle** alone.”

“**Fear** is the biggest reason I don’t want to **admit** that I’m depressed. **Need** to get over this. I can **do** it.”

Interaction Between Gender and Culture

Our final empirical investigation considers the interaction between gender and culture, as measured linguistically. For the sake of simplicity, we do not include topical analysis for this last investigation. In Table 6 we present mean values of the different linguistic measures, grouped by categories, for male and female MID users in each of the four cultures. Since not all linguistic measures were significant based on Kruskal-Wallis significance tests, we computed the means based on those that were indeed significant. We additionally report in the table the mean relative difference between male and female MID users grouped by the Western (US, GB) and Majority world (IN, ZA) cultures.

We derive two major observations from this investigation. First, for the linguistic measure categories, females and males show systematic differences across the four cultures, that align with our findings from the gender analysis. For instance, as we had observed before, females for all of the cultures tend to express more positive affect and less negative affect compared to males from the corresponding cultures. They also show lower cognitive processing, express more social/personal concerns on Twitter, and tend to be more interpersonally focused — all of which we had observed before. This indicates that gender differences are persistent over cultures, at least in the context of the four cultures we study here.

Our second observation revolves around assessing the extent of linguistic difference between females and males spanning the four cultures. Even in the light of the above findings around systematic differences in the linguistic expression of the two gender groups across cultures, not all cultures show the same extent of difference. Interestingly, the relative difference between females and males are greater for the two majority world cultures (IN, ZA), in contrast to the west-

	μ (US)	μ (GB)	μ (IN)	μ (ZA)	Diff (W)	Diff (M)
Affect – positive						
male	0.0351	0.0367	0.0195	0.0291		
female	0.1056	0.1040	0.1036	0.0970	-192.2%	-332.6%
Affect – negative						
male	0.0151	0.0175	0.0207	0.0183		
female	0.0065	0.0040	0.0035	0.0052	67.14%	77.25%
Cognition						
male	0.0572	0.0556	0.0720	0.0687		
female	0.0190	0.0192	0.0082	0.0115	66.10%	85.98%
Perception						
male	0.0268	0.0200	0.0315	0.0222		
female	0.0062	0.0131	0.0038	0.0131	55.57%	64.42%
Lexical Density						
male	0.0339	0.0315	0.0309	0.0297		
female	0.0939	0.0971	0.1015	0.1027	-192.4%	-237.6%
Temporal references – past						
male	0.0119	0.0114	0.0061	0.0075		
female	0.0478	0.0496	0.0438	0.0423	-317.4%	-540.2%
Temporal references – future						
male	0.0093	0.0098	0.0132	0.0110		
female	0.0040	0.0041	0.0019	0.0041	57.78%	74.16%
Social/personal concerns						
male	0.0048	0.0127	0.0042	0.0125		
female	0.0339	0.0261	0.0377	0.0296	-352.4%	-467.2%
Interpersonal focus – self						
male	0.0508	0.0591	0.0788	0.0681		
female	0.0259	0.0235	0.0122	0.0229	54.66%	75.45%
Interpersonal focus – others						
male	0.0105	0.0075	0.0058	0.0067		
female	0.0192	0.0228	0.0190	0.0181	-143.7%	-197.8%

Table 6: Mean values of different measures across male and female MID users in each of the four cultures. Only statistically significant measures were considered in this calculation. For each measure category, we also show relative percentage differences between male and female groups in the western (Diff (W)) and majority world (Diff (M)) cultures.

ern cultures (US, GB). For instance, positive affect difference between females and males in the former culture group is -332.6%, whereas it is only -192.2% in the case of the latter.

Causal links, e.g., whether culture indeed impacts gendered expression of mental health on social media cannot be easily derived. However, putting it together, these observations indicate that, with regard to social media disclosures of mental health, culture and gender are closely intertwined.

DISCUSSION

Theoretical and Practical Implications

Our work has provided some of the first detailed insights into how gender and cultural attributes relate to the content shared by individuals who disclose about their mental health challenges on these platforms.

Reflecting on Gender Differences. In the sample of mental health disclosing Twitter users we studied in this paper, there were considerable differences in the manner in which females and males appropriated the platform. Specifically, based on our different linguistic measures, females expressed higher sadness and anxiety, but lower anger and NA. These observations align with prior work in social psychology. Lieberman and Goldstein [32] also used the LIWC program to find that women in online support groups who experienced depression used high volume of anxiety words. Female mental health disclosers in our data also expressed greater social and familial concerns, compared to males. Literature has indicated that women tend to rely more on the social network of their family and the community, rather than the individual; whereas men exhibit a relative orientation toward stoicism [20]. These observations provide credence to our findings.

Per our topic model analysis, we further observed differences between female and male mental health disclosers. Male users appropriated Twitter to show greater detachment from the social realm and hesitation to seek help. Examining literature on mental health trauma, we find evidence supporting this observation. Norris et al. [37] found that the differences in PTSD severity were greater women compared to men. The authors speculated this to be attributable to adherence to more traditional gender roles, in which male “machismo” inhibits disclosure of distress.

Reflecting on Cultural Differences. Turning to cultural comparison of social media disclosers of mental health concerns, one of our findings revealed that individuals from the majority world countries, IN and ZA were less likely to express negative emotions in their Twitter posts compared to those from western countries, US and GB. The former group also expressed relatively less cognitive impairment and lowered tendency of social interaction as measured via pronoun use. As Marsella and Christopher [34] have argued, intrusive symptoms of many mental health concerns may be universal, however, avoidance or numbing may be more culturally based. Yeomans et al. [57] also discovered cultural factors in how people handle traumatic stress: for example, Asians are more reluctant to express distress in public, which aligns with our findings.

Moreover, based on our topic analysis, the IN and ZA cohorts were less likely to be disinhibiting or candid in their social media discourse, compared to their western peers. For instance, they used fewer words relating to self-critical or self-loathing thoughts, but shared evidence of bereavement and shame. This finding supports observations in prior cross-cultural investigations of mental health in eastern and western cultures; where individuals in the former culture were found to feel more socially stigmatized while sharing these experiences socially [17]. These practices have been attributed to social, political, economic, and communication barriers. Broadly, these observations indicate notable cultural differences in the Twitter activities of IN and ZA based mental health disclosers. They support the view that a conflict exists between traditional cultural values for these groups and the way in which psychological and social support are sought in western cultures.

Gender and Culture. Finally, we observed notable patterns in the interaction between gender and culture groups, as measured through language. Kövecses [31] posited that “Although we find a great deal of commonality in emotion concepts and metaphors both across languages/cultures and through time, we can see a great deal of variation as well”. Our findings of this interaction study align with this position of Kövecses — the female and male mental health disclosers showed similar linguistic characteristics across the four cultures, however, the extent to which they differed in each culture, was markedly distinct.

Summarily, as pointed out earlier, gender and culture based dimensions of mental health are less understood today, primarily because of the challenges in obtaining reliable and normalized data on the same [49, 47]. In this paper, we showed the potential of social media as a way to quantify and assess these differences. To do so, we proposed, developed, and evaluated a rigorous and principled methodology that automatically and accurately identified disclosures of mental health challenges on social media, specifically Twitter. Obtaining adequate access to gold standard data (aka ground truth) is of prime importance in quantitative studies of mental health and social media. The most reliable data is gathered via self-reported means like surveys and user studies, however are difficult to scale. Gathering public data around mental health topics from social media incurs a much lower overhead, but the false positive rates in such data acquisition approaches can be alarmingly high. Our technique balances this through a semi-supervised approach, in which we leverage a much more reliable source of mental health data to glean genuine disclosures in another large but noisy source. Domain expert validation revealed the effectiveness of our approach. In short, beyond mental health, we believe our method provides a generalized template for gathering reliable data in problem contexts where either data acquisition is difficult, or noise, social and psychological considerations pose significant challenge in obtaining and curating quality data.

Besides, while our findings largely focus on validating known attributes of gender and culture in the context of mental health, our work brings to the fore novel mechanisms to do so. It proposes a rigorous quantitative framework through which the findings can be studied in large populations, tested for generalization, or adapted to multiple online, gender and cultural contexts. Finally, we were also able to discover nuances in mental health expression across the gender and culture subgroups, that may be challenging to quantify through traditional means. These primarily include recognizing the role of specific linguistic constructs and topics in mental health disclosures. Thus, a gender and cross-cultural approach to mental health, examined from naturalistic, unobtrusive data collected from social media, can provide guidance to the identification of appropriate responses from the mental healthcare system, as well as for healthcare policy globally.

Design Implications and Ethics

Social media data can be an unobtrusive mechanism to gather information about one’s emotional and psychological state, and our findings indicate that this rich information source may be employed to understand and characterize mental

health disclosures. Hence our methods can influence the design and development of timely and personalized intervention tools and provisions that bring help to mental health disclosers on social media. Such interventions can include applications and services that can serve as early warning systems, and tools that can provide personalized alerts and psychosocial support resources to individuals, based on their linguistic expression, topic usage, and emotional distress. By allowing individuals monitor their linguistic and topical trends, interventions can also include technologies that serve as a personal diary-type narrative for self-reflection and self-awareness of mental well-being over time.

While design considerations along these lines have been proposed in prior work on western populations [13], for these interventions to be viewed as credible and effective more broadly, they would need to be modified for those assuming different gender role or belonging to different cultures. In fact, the relationship between gender and cultural attributes and technology use, such as social media have begun to constitute a major thrust in CSCW and social computing research. Below we describe some concrete design considerations that can help make mental health intervention technologies more sensitive and adaptive toward one's gender and culture.

Making Interventions Culture-Aware. Virtually all cultures have therapeutic systems that reflect their particular world view and values [57]. Design considerations of mental health interventions delivered via online or social media channels could leverage these cultural constructs. For instance, many eastern cultures often take more of a “socio-centric” than an “egocentric” view of society [57]—this observation was notable in our comparisons of majority world and western cultures as well. Hence interventions can prioritize provisioning social support on social media to individuals from majority world cultures, such as through recommending supportive links, communities and groups, as well as by providing a communication channel for candid discourse. On the other hand, interventions could focus on alternative schemes, such as providing subtle nudges, promoting self-awareness and self-reflection, and triggering personalized alerts to enable those in western cultures better cope with and manage their mental health challenges.

It is also known that ethnic groups tend to show lower compliance to mental health interventions, possibly due to the implications of social stigma and shame involved in seeking mental health help, or perhaps due to deep-seated beliefs about mental illness [47]. In fact, our results showed that perceptions of stigma and emotional inhibition were higher in the cultural groups from the majority world countries. Hence appropriate efforts need to be made in both designing and deploying interventions to these cultural groups in particular. Techniques such as gamification, or external incentive or reward structures could be incorporated in the above outlined interventions to increase motivation and compliance with systems geared towards improving one's mental well-being.

Making Interventions Gender-Aware. Heightened levels of self-introspection and hesitation to seek help are known to foster a sense of uncertainty and emotional vulnerability [30].

These tendencies are more pronounced in men [1]. In fact, our findings reveal that male mental health disclosers showed a detachment from their social realms, as well as were less willing to acknowledge their psychological problems. Therefore, social support based interventions can be adequately and specifically directed to male mental health disclosers on social media platforms. For instance, they could be connected with other help seekers in an online community, so that they could confide in each other, share experiences and challenges, and find emotional support through an informal channel.

Certainly, none of these interventions would be successful without appropriate and sensible privacy and ethical considerations in place, due to the sensitivities that exist around a stigmatized condition like mental illness. Although our methods are based on public data collected from social media, we suggest caution in how our findings are interpreted in gendered and cultural contexts. Importantly, our method or findings should not be used for making diagnostic claims about mental illness, or as standalone detectors of mental illness in gender based or cultural subgroups. We also do not intend our work and findings to be leveraged in ways that amplify perceptions of gender or culture based stereotypes on social media, or use these attributes for discriminatory purposes around mental health topics.

Limitations

There are some limitations to our work that one should consider. First, we acknowledge that our findings are limited by our data acquisition capabilities. We relied on a set of hand curated keyphrases to seed our Twitter data collection. Although these were validated via consultation with a psychologist, they do not possibly include all of the ways in which Twitter users self disclose their mental health concerns. We were also limited by our ability to accurately infer gender and country information for the users in our dataset. While the methods we used have been employed in prior work and yield high precision content, we cannot rule out that our method suffers from a low recall problem.

Further, our method of obtaining mental health self-disclosure information only captures a subpopulation of Twitter users with a form of mental illness (i.e., those who are speaking publicly about what is usually a very private matter). This may not truly represent all aspects of the population as a whole. Moreover, this method in no way verifies whether this diagnosis is genuine (i.e., people are not always truthful in self-reports on Twitter), or whether they are clinically validated. However, given the stigma often associated with mental illness, it seems unlikely users would self-disclose that they are diagnosed with a condition they do not have.

We also note caveats in our semi-supervised machine learning method of identifying genuine mental illness disclosures, and eliminating noisy control data samples. We used Reddit as a way to obtain weakly labeled mental health disclosure information. However, we do note that Reddit demographics and platform usage practices might be considerably different from that of Twitter. We believe our normalization step for comparative validity to be helpful in curbing the effects of

these differences, but future work could explore other means to obtain labeled mental health content.

Our results, while suggestive, are correlational by nature. Therefore, it is impossible to say whether the expression of specific linguistic or topical constructs was the driving factor behind different gender and cultural subgroups' mental health disclosures on social media.

Finally, we comment on the cultural grouping adopted in this work—US and GB which were defined as western cultures, and IN and ZA which were included as majority world countries. Our choice was driven by an attempt to coarsely identify broad cultural dimensions that may simplify our analytic approaches. However, we acknowledge that the four countries chosen are not culturally distinct—for instance, IN and ZA have a rich colonial history, which might explain some of the observed similarities between them and GB, in contrast to US. Certainly, other groupings are possible. For instance, several organizations use metrics of economic prosperity to identify developed and developing cultures¹. How our findings generalize over alternative cultural groupings is an interesting direction toward future research.

CONCLUSION

Psychology research recognizes considerable data gaps in gender based and cultural dimensions of global mental health. In this paper, we provided some of the first quantitative insights into gender and cultural differences associated with individuals who self-disclose to suffer from a mental health concern on Twitter. To identify genuine disclosures, we developed a semi-supervised learning based framework, which yielded high accuracy (96%) based on expert feedback. Thereafter, we explored the range of differences in the content shared by female and male users, and users who report on Twitter to be from US, UK, India or South Africa. We observed male users to express higher negativity and lower desire for social support, whereas majority world users (India and South Africa) demonstrated more inhibition in their expression. Our findings help validate, via use of social media data, a number of known characteristic differences in mental health experience of gender and cultural subgroups. We believe our work can encourage re-thinking of privacy-honoring health interventions to be more gender and culture aware, so that they could bring appropriate and personalized help and support to vulnerable individuals on social media.

ACKNOWLEDGEMENTS

We would like to thank Bapu Vaitla and Rebecca Furst-Nichols from the United Nations Foundation for helping frame the problem statement and for fruitful discussions. For gender inferences we would like to thank Thomas Baar from Leiden University's Centre for Innovation, Parisa Zahedi from Risa-IT, Pim van de Pavoordt and Gerard Simons from Qualogy, and Maral Dadvar who helped through UN Volunteers. This work was supported by a United Nations Foundations grant to De Choudhury.

REFERENCES

1. Mustafa Afifi. 2007. Gender differences in mental health. *Singapore medical journal* 48, 5 (2007), 385.

2. Nazanin Andalibi, Oliver L Haimson, Munmun De Choudhury, and Andrea Forte. 2016. Understanding Social Media Disclosures of Sexual Abuse Through the Lenses of Support Seeking and Anonymity. In *Proceedings of the SIGCHI conference on human factors in computing systems (CHI)*.
3. Nazanin Andalibi, Pinar Ozturk, and Andrea Forte. 2015. Depression-related Imagery on Instagram. In *Proc. CSCW'15 Companion*. 231–234.
4. Steven Bird. 2006. NLTK: the natural language toolkit. In *COLING/ACL Interactive presentation sessions*.
5. David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *the Journal of machine Learning research* 3 (2003), 993–1022.
6. Lera Boroditsky, Lauren A Schmidt, and Webb Phillips. 2003. Sex, syntax, and semantics. *Language in mind: Advances in the study of language and thought* (2003), 61–79.
7. Ana Mari Cauce, Melanie Domenech-Rodríguez, Matthew Paradise, Bryan N Cochran, Jennifer Munyi Shea, Debra Srebnik, and Nazli Baydar. 2002. Cultural and contextual influences in mental health help seeking: a focus on ethnic minority youth. *Journal of consulting and clinical psychology* 70, 1 (2002), 44.
8. Stevie Chancellor, Zhiyuan (Jerry) Lin, Erica Goodman, Stephanie Zerwas, and Munmun De Choudhury. Quantifying and Predicting Mental Illness Severity in Online Pro-Eating Disorder Communities. In *Proceedings of the 19th ACM conference on Computer supported cooperative work & social computing*.
9. Cindy Chung and James W Pennebaker. 2007. The psychological functions of function words. *Social communication* (2007), 343–359.
10. Aaron V Cicourel. 1974. Cognitive sociology: Language and meaning in social interaction. (1974).
11. Glen Coppersmith, Mark Dredze, and Craig Harman. 2014a. Quantifying mental health signals in twitter. In *ACL Workshop on Computational Linguistics and Clinical Psychology*.
12. Glen Coppersmith, Craig Harman, and Mark Dredze. 2014b. Measuring post traumatic stress disorder in Twitter. In *International Conference on Weblogs and Social Media (ICWSM)*.
13. Munmun De Choudhury, Scott Counts, Eric Horvitz, and Aaron Hoff. 2014. Characterizing and Predicting Postpartum Depression from Facebook Data. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM.
14. 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)*.

15. Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. Predicting depression via social media. In *AAAI Conference on Weblogs and Social Media*.
16. Munmun De Choudhury, Emre Kiciman, Mark Dredze, Glen Coppersmith, and Mrinal Kumar. 2016. Discovering Shifts to Suicidal Ideation from Mental Health Content in Social Media. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM.
17. Simone Freitag, Anna Grimm, and Silke Schmidt. 2011. Talking about traumatic events: A cross-cultural investigation. *Europes Journal of Psychology* 7, 1 (2011), 40–61.
18. Erving Goffman. 1963. *Stigma: Notes on the management of spoiled identity*. Prentice-Hall.
19. James J Gross and Ricardo F Muñoz. 1995. Emotion regulation and mental health. *Clinical psychology: Science and practice* 2, 2 (1995), 151–164.
20. Francis Guillemin, Claire Bombardier, and Dorcas Beaton. 1993. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *Journal of clinical epidemiology* 46, 12 (1993), 1417–1432.
21. Kira Hall and Mary Bucholtz. 2012. *Gender articulated: Language and the socially constructed self*. Routledge.
22. Brent Hecht, Lichan Hong, Bongwon Suh, and Ed H Chi. 2011. Tweets from Justin Bieber’s heart: the dynamics of the location field in user profiles. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 237–246.
23. Matthew Hoffman, Francis R Bach, and David M Blei. 2010. Online learning for latent dirichlet allocation. In *Neural Information Processing Systems (NIPS)*.
24. Christopher M Homan, Naiji Lu, Xin Tu, Megan C Lytle, and Vincent Silenzio. 2014. Social structure and depression in TrevorSpace. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. ACM, 615–625.
25. Yuheng Hu, Kartik Talamadupula, Subbarao Kambhampati, and others. 2013. Dude, srsly?: The Surprisingly Formal Nature of Twitter’s Language.. In *ICWSM*.
26. Mutsumi Imai and Dedre Gentner. 1997. A cross-linguistic study of early word meaning: Universal ontology and linguistic influence. *Cognition* 62, 2 (1997), 169–200.
27. Lilly Irani, Janet Vertesi, Paul Dourish, Kavita Philip, and Rebecca E Grinter. 2010. Postcolonial computing: a lens on design and development. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 1311–1320.
28. Adam N Joinson. 2001. Self-disclosure in computer-mediated communication: The role of self-awareness and visual anonymity. *European Journal of Social Psychology* 31, 2 (2001), 177–192.
29. Max Kaufmann and Jugal Kalita. 2010. Syntactic normalization of twitter messages. In *International conference on natural language processing, Kharagpur, India*.
30. Ichiro Kawachi and Lisa F Berkman. 2001. Social ties and mental health. *Journal of Urban health* 78, 3 (2001), 458–467.
31. Zoltán Kövecses. 2003. *Metaphor and emotion: Language, culture, and body in human feeling*. Cambridge University Press.
32. Morton A Lieberman and Benjamin A Goldstein. 2006. Not all negative emotions are equal: The role of emotional expression in online support groups for women with breast cancer. *Psycho-Oncology* 15, 2 (2006), 160–168.
33. Diana MacLean, Sonal Gupta, Anna Lembke, Christopher Manning, and Jeffrey Heer. 2015. Forum77: An Analysis of an Online Health Forum Dedicated to Addiction Recovery. In *Computer-Supported Cooperative Work and Social Computing (CSCW)*.
34. Anthony J Marsella and Michael A Christopher. 2004. Ethnocultural considerations in disasters: An overview of research, issues, and directions. *Psychiatric Clinics of North America* 27, 3 (2004), 521–539.
35. Joanna Maselko and Vikram Patel. 2008. Why women attempt suicide: the role of mental illness and social disadvantage in a community cohort study in India. *Journal of epidemiology and community health* 62, 9 (2008), 817–822.
36. Elizabeth L Murnane and Scott Counts. 2014. Unraveling abstinence and relapse: smoking cessation reflected in social media. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. ACM, 1345–1354.
37. Fran H Norris, Julia L Perilla, Gladys E Ibañez, and Arthur D Murphy. 2001. Sex differences in symptoms of posttraumatic stress: Does culture play a role? *Journal of Traumatic Stress* 14, 1 (2001), 7–28.
38. Michael Paolisso and Joanne Leslie. 1995. Meeting the changing health needs of women in developing countries. *Social Science & Medicine* 40, 1 (1995), 55–65.
39. Vikram Patel, Ricardo Araya, Mauricio de Lima, Ana Ludermitz, and Charles Todd. 1999. Women, poverty and common mental disorders in four restructuring societies. *Social science & medicine* 49, 11 (1999), 1461–1471.
40. Vikram Patel and Mark Winston. 1994. ” Universality of mental illness” revisited: Assumptions, artefacts and new directions. *The British Journal of Psychiatry* (1994).

41. Michael J Paul and Mark Dredze. 2011. You Are What You Tweet: Analyzing Twitter for Public Health.. In *ICWSM*.
42. James W Pennebaker, Martha E Francis, and Roger J Booth. 2001. Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates* 71 (2001), 2001.
43. Nairan Ramirez-Esparza, Cindy K Chung, Ewa Kacewicz, and James W Pennebaker. 2008. The Psychology of Word Use in Depression Forums in English and in Spanish: Texting Two Text Analytic Approaches.. In *ICWSM*.
44. RJ Senter and EA Smith. 1967. *Automated readability index*. Technical Report. DTIC Document.
45. Richard A Shweder. 1991. *Thinking through cultures: Expeditions in cultural psychology*. Harvard University Press.
46. Robin W Simon. 1995. Gender, multiple roles, role meaning, and mental health. *Journal of Health and Social behavior* (1995), 182–194.
47. Rachel E Spector. 2002. Cultural diversity in health and illness. *Journal of Transcultural Nursing* 13, 3 (2002), 197–199.
48. Steven Stack. 1996. The effect of the media on suicide: evidence from Japan, 1955–1985. *Suicide and Life-Threatening Behavior* 26, 2 (1996), 132–142.
49. Shelley E Taylor and Jonathon D Brown. 1988. Illusion and well-being: a social psychological perspective on mental health. *Psychological bulletin* 103, 2 (1988), 193.
50. Harry C Triandis. 1989. The self and social behavior in differing cultural contexts. *Psychological review* 96, 3 (1989), 506.
51. Sho Tsugawa, Yusuke Kikuchi, Fumio Kishino, Kosuke Nakajima, Yuichi Itoh, and Hiroyuki Ohsaki. 2015. Recognizing Depression from Twitter Activity. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 3187–3196.
52. T Bedirhan Üstün and Norman Sartorius. 1995. *Mental illness in general health care: an international study*. John Wiley & Sons.
53. Bapu Vaitla. 2014. The Landscape of Big Data for Development. *Data2x* (2014).
54. Xiangdong Wang, Lan Gao, Naotaka Shinfuku, Huabiao Zhang, Chengzhi Zhao, and Yucun Shen. 2000. Longitudinal study of earthquake-related PTSD in a randomly selected community sample in north China. *American Journal of Psychiatry* (2000).
55. Benjamin Lee Whorf, John B Carroll, Stephen C Levinson, and Penny Lee. 2012. *Language, thought, and reality: Selected writings of Benjamin Lee Whorf*. Mit Press.
56. Christine J Yeh. 2003. Age, acculturation, cultural adjustment, and mental health symptoms of Chinese, Korean, and Japanese immigrant youths. *Cultural Diversity and Ethnic Minority Psychology* 9, 1 (2003), 34.
57. Peter D Yeomans and Evan M Forman. 2009. Cultural factors in traumatic stress. *Sociocultural influences on mental health*. Boston: Blackwell (2009), 221–244.
58. Xiaojin Zhu, Zoubin Ghahramani, John Lafferty, and others. 2003. Semi-supervised learning using gaussian fields and harmonic functions. In *ICML*, Vol. 3. 912–919.