Published in final edited form as:

Proc ACM Hum Comput Interact. 2019 November ; 3(CSCW): . doi:10.1145/3361108.

The Language of LGBTQ+ Minority Stress Experiences on Social Media

KOUSTUV SAHA,
Georgia Institute of Technology, Atlanta, Georgia, USA

SANG CHAN KIM,
Georgia Institute of Technology, Atlanta, Georgia, USA

MANIKANTA D. REDDY,
Georgia Institute of Technology, Atlanta, Georgia, USA

ALBERT J. CARTER,
University of Michigan, Ann Arbor, Michigan, USA

EVA SHARMA,
Georgia Institute of Technology, Atlanta, Georgia, USA

OLIVER L. HAIMSON,
University of Michigan, Ann Arbor, Michigan, USA

MUNMUN DE CHOUDHURY
Georgia Institute of Technology, Atlanta, Georgia, USA

Abstract

LGBTQ+ (lesbian, gay, bisexual, transgender, queer) individuals are at significantly higher risk for mental health challenges than the general population. Social media and online communities provide avenues for LGBTQ+ individuals to have safe, candid, semi-anonymous discussions about their struggles and experiences. We study *minority stress* through the language of disclosures and self-experiences on the r/lgbt Reddit community. Drawing on Meyer's minority stress theory, and adopting a combined qualitative and computational approach, we make three primary contributions, 1) a theoretically grounded *codebook* to identify minority stressors across three types of minority stress—*prejudice events, perceived stigma*, and *internalized LGBTphobia*, 2) a *machine learning classifier* to scalably identify social media posts describing minority stress experiences, that achieves an AUC of 0.80, and 3) a *lexicon* of linguistic markers, along with their contextualization in the minority stress theory. Our results bear implications to influence public health policy and contribute to improving knowledge relating to the mental health disparities of LGBTQ+ populations. We also discuss the potential of our approach to enable designing online tools sensitive to the needs of LGBTQ+ individuals.

Corresponding author: koustuv.saha@gatech.edu.

Keywords

LGBTQ+; mental health; minority stress; Reddit; social media; stigma

1 INTRODUCTION

Over the past three decades, there has been increasing public and scientific awareness of lesbian, gay, bisexual, transgender, and additional sexual and gender minority (LGBTQ+) lives, experiences, struggles, and other matters. LGBTQ+ is an acronym used to refer to individuals who use these sexual or gender identity labels as personally meaningful for them, acknowledging that sexual and gender identities are complex and historically situated[1].

Studies in the 1980s by Schneider et al. uncovering the concerning rates of reported suicidal behavior among gay youth was one of the first notable public research centered around LGBTQ+ issues and their mental health [96]. Thereafter, numerous studies have provided evidence of heightened risk of mental illnesses and suicide in LGBTQ+ populations [34, 89, 104]. Alarmingly, Marshal et al.'s meta-analysis reported that sexual minority individuals were three times as likely to report suicidality and attempted suicide in contrast to others who did not hold these identities [63].

Despite growing evidence, the causes and correlates of these mental health disparities are underexplored. The experiences of LGBTQ+ people must be better understood in order to develop care and intervention strategies that cater to their struggles [80]. Unfortunately, most mental health interventions rarely cater to the unique needs of LGBTQ+ individuals, and most counselors never receive training in working effectively with LGBTQ+ clients [106].

Sociologist Ilan H. Meyer in 1995 provided a theoretical framework to conceptualize the mental health of LGBTQ+ individuals [68]. Drawing upon initial conceptualization by Brooks [14], Meyer's *minority stress theory*, which has enjoyed widespread empirical support, contextualizes minority stress as "psychosocial stress derived from minority status" that the LGBTQ+ individuals experience [68]. State-of-the-art approaches of measuring minority stressors rely on self-reports, surveys, and under-representative convenience samples, limiting generalizability due to the challenges of accessing a stigmatized, hard-to-reach population [9]. These approaches may also suffer from limitations due to the discomfort of the respondents who have to recollect sensitive experiences of prejudice, stigma, discrimination, violence, and social rejection and isolation [31, 91]. Consequently, there is a gap in research around understanding and quantifying stressors that are identified by distressed LGBTQ+ individuals to bear negative impacts on their mental health.

The Internet has been identified to be a place where many LGBTQ+ individuals are coming out, finding peers, and seeking help, because of prevailing ignorance and prejudice about LGBTQ+ issues in offline contexts [65, 66]. Additionally, we note that recent social

[1] The "plus" in LGBTQ+ is inclusive of additional groups, including but not limited to asexual, intersex, queer, questioning, pansexual, non-binary, etc.

computing literature has situated semi-anonymous and anonymous social media sites like Reddit as platforms to study mental health since they enable non-judgmental and candid discourse around issues that otherwise might be stigmatizing in the society [4, 27]. This paper posits that such online self-disclosure and support seeking practices of LGBTQ+ individuals, therefore, offer a new opportunity to tackle the challenges involved in lessening LGBTQ+ mental health disparities.

In particular, we target the research question of, how *can we automatically infer minority stress expressions on social media at scale?*. By leveraging public data of online discussions of self-identifying LGBTQ+ individuals on Reddit, we develop computational and analytic approaches to understand, assess, and examine minority stress and draw meaningful insights into the unique mental health challenges and needs of LGBTQ+ individuals. We make three primary contributions:

- A codebook to identify minority stressors in social media per Meyer's minority stress theory.
- A machine learning classifier to identify social media posts on experiences of minority stress.
- A lexicon of social media language markers contextualized in the minority stress theory.

We collect 12.6K posts from *r/lgbt* subreddit, and we qualitatively annotate 350 of them regarding whether they contain expressions of minority stress. In the process, we develop a codebook based on Meyer's minority stress theory, operationalizing three types of minority stressors, *prejudice events, perceived stigma*, and *internalized LGBTphobia*. Next, we develop a machine learning classifier that uses features based on word embeddings, psycholinguistic attributes, open-vocabulary based *n*-grams, hateful keywords, and linguistic expressions of mental health symptoms to identify minority stress in the language of self-disclosing LGBTQ+ individuals. After demonstrating this classifier to provide robust and stable performance with an AUC of 0.80, we machine label our entire dataset. We proceed to study the linguistic markers associated with minority stress and build a lexicon of minority stress markers using an unsupervised language modeling technique [37]. We situate our observations of a qualitative examination of these linguistic markers in the minority stress theory, thereby establishing both face and construct validity of the lexicon. The lexicon additionally provides aggregated inferences about the nature and topics of discourse associated with the language of minority stress on social media — we find that the posts that are associated with minority stress are mostly personal and about self-life experiences, whereas the posts that do not express minority stress, are about general issues faced by the minority communities, and demonstrate a sense of community, inclusiveness, diversity, and are about raising awareness.

To our knowledge, this is the first study that utilizes the discussions in a semi-anonymous LGBTQ+ online community to develop a framework for characterizing the mental health challenges of LGBTQ+ people. Our findings have widespread implications for online community design and functioning, that can better support the mental health needs of a marginalized and stigmatized population like LGBTQ+ individuals. The research also offers new avenues for public health intervention and policy change to better address the mental health disparities of LGBTQ+ individuals.

Ethics and Disclosure.

Given the sensitivities of this work, we include a self-reflexivity statement. a) Because we use publicly accessible, historical, deidentified posts from Reddit without any interaction with the authors of these posts, our work did not qualify for approval from the relevant institutional review boards. Nevertheless, we took great care in the manner data and analyses are presented in the paper, for instance, by avoiding any personally identifiable information and paraphrasing any quote that we reference to reduce traceability. b) These paraphrased excerpts of self-experiences by Reddit posters are only used in this paper to help ground our results. However, some of this content may be sensitive, so we suggest caution to the readers. c) Multiple coauthors on the papers hold different LGBTQ+ identities, which has enabled us to incorporate sensitivity in the framing of the work and contextualize our findings in the light of our own lived experiences. d) Despite these above points, we recognize and acknowledge the limitations of our methodological approach and our position as researchers and outsiders to this particular online community. We describe our limitations and ethical considerations further in the Discussion section.

2 BACKGROUND AND RELATED WORK

2.1 Minority Stress Theory

LGBTQ+ individuals are known to suffer from widespread health disparities [40], and are also thrice more likely than others to experience a mental health condition [74]. LGBTQ+ mental health disparities and risks are associated with minority stressors such as depression, post-traumatic stress disorder, suicidal ideation, and substance abuse [56, 69]. Generally speaking, minority stress is a concept developed by sociologists and psychologists to explain the unique pressures faced by members of minority groups [14, 68]. Specific to LGBTQ+ individuals, “stigma, prejudice, and discrimination create a hostile and stressful social environment that causes mental health problems” [68]. Minority stress theory proposes that sexual and gender minority health disparities can be explained largely by stressors induced by a hostile, homophobic, and transphobic culture, which often leads to harassment, mistreatment, and victimization, and may ultimately impact access to care [32, 41, 50, 59, 69]. Minority stressors related to one’s LGBTQ+ identity can cause significant stress which could ultimately affect physical and mental health outcomes [69].

Most prior minority stress research has relied on convenience-based samples, such as surveys, to collect data on minority stress’s prevalence and impacts. A notable study concluded that the LGBTQ+ population requires public health attention beyond what can be targeted with surveys [76]. Examining the effects of minority stress factors on LGBTQ+ individuals’ mental health is complicated, as recruitment for research studies and subsequent behavioral interventions remains limited and challenging due to difficulty accessing the community [64]. These limitations ultimately result in ambiguity about the extent to which LGBTQ+ individuals experience stressors such as prejudice, rejection and discrimination, and how to quantify these stressors to support interventions.

Our research aims to leverage LGBTQ+ online discussions to identify minority stressors, situated in minority stress theory [68]. By focusing on data contributed by self-disclosing LGBTQ+ individuals on semi-anonymous social media (Reddit), our approach tackles challenges of accessing this “hard to reach” population [64] by focusing on individuals spanning geographies, and a range of sexual and gender identities.

2.2 LGBTQ+ People’s Social Media Use

Many LGBTQ+ people engage in online activities and communities, which can be affirming and positive online spaces for them [51]. A 2011 national survey of LGBTQ+ youth reported that this population spent more time online, and were more likely to have close online friends, compared to non-LGBTQ+ youth [79]. Social media platforms enable LGBTQ+ people to seek and find health information [49, 61], yet this practice can sometimes be invalidating when one’s specific identity or health concern is not represented online [61]. Tumblr has often been recognized as particularly LGBTQ+- friendly [18, 21, 36, 77]. Some of Reddit’s features, such as anonymous and pseudonymous identities, enable LGBTQ+ communities to form and thrive [26, 39].

LGBTQ+ individuals often present identities differently among separate and sometimes overlapping networks of people across the social media ecosystem [33, 44]. This multiplicity is often necessary during identity construction [44], but also involves employing different features on multiple platforms to make self-presentation decisions that fit one’s identity and allow one to avoid stigma [33]. Posting on social media can also be a way to weed out who in one’s network is and is not supportive of one’s LGBTQ+ identity [12]. Some LGBTQ+ individuals face challenges when using social media sites, due to the persistence of personal data linked to a past identity [46], and privacy related complexities [12, 21]. Yet, social media can be an important place for online LGBTQ+ presentation due to the ability to maintain boundaries between different identities and networks, thus enabling a relatively safe space for identity exploration and transition [17, 44].

Researchers have shown how social media sites and other social technologies often do not fully account for the needs of LGBTQ+ users [1, 21, 48]. Some of the difficult experiences LGBTQ+ users face involve elements of minority stress [36, 45, 61, 95]. Moreover, despite data suggesting the utility of online support for LGBTQ+ individuals, online communities have been underutilized to understand this population’s mental health. Our work aims to fill this existing gap in prior research.

2.3 Stigma, Self-Disclosure, and Mental Health Support Seeking on Social Media

When people face emotional challenges, they often wish to disclose that experience to others [81], and people tend to disclose more in online settings [102]. In social media contexts, people often face difficulty deciding if and how to disclose emotional experiences that carry stigma [2], and must manage difficult tensions between disclosure desires and impression management [75]. Thus, people often disclose selectively to particular audiences [5, 105], often after substantial time has passed [2, 98], or share less or no content [72]. Yet despite the barriers to disclosing stigmatized experiences on identified social media sites [11], when people do disclose, they often receive positive benefits such as receiving social support [6, 45, 75].

Social media platforms like Reddit that enable pseudonymity enable disclosure for stigmatized experiences and communities [3, 4]. As one example of the positive benefits of Reddit as a platform that enables anonymous disclosures, Andalibi et al. found that throwaway accounts on Reddit can be helpful for men to disclose past experiences of sexual abuse, despite the substantial stigma that men face around sexual abuse [4]. Anonymity or pseudonymity further facilitate people feeling safe in an online community of similar others [2, 44].

In a parallel line of research, researchers have employed computational and machine learning approaches to study stigmatizing experiences and mental health concerns on social media sites, specifically Reddit [27, 29, 30, 38, 99]. One study examined a variety of mental health support communities on Reddit, such as *r/SuicideWatch* and *r/depression* and found that self-disclosure expressions align with clinical literature on mental health [27]. Researchers also used self-experience posts on Reddit communities to build machine learning classifiers that identify expressions of mental health (e.g., depression, suicidal ideation, mental illness severity) [10, 20, 92, 93].

Together, this body of research motivates both our objective of studying self-disclosure expressions related to stigmatized experiences on Reddit, as well as our approach to scalably investigate the language associated with a specific kind of stigmatized mental health experience (minority stress) that caters to a specific population (LGBTQ+ identities).

3 DATA

This paper studies the language shared by self-disclosing LGBTQ+ individuals on social media. In particular, we use Reddit – it is a widely used semi-anonymous online social forum. The platform organizes discussions into a variety of sub-communities called *subreddits*. There are dedicated subreddits that focus on LGBTQ+ issues. Using Reddit's subreddit search feature, we surfaced subreddits such as *r/lgbt, r/LGBTPolitics, r/dixiequeer, r/lgbtaww*, and *r/ainbow* that are contextually related. For the purposes of this paper, where we focus on the language related to self-experiences and mental health, we direct our attention to the largest of all these subreddits, *r/lgbt*, which defines itself as a safe space for members of GSRM (Gender, Sexual, and Romantic Minority) to "discuss their lives, issues, interests, and passions" and welcomes all GSRM members "beyond lesbian, gay, bisexual, and transgender people." It has been active for the last nine years, has considerable traffic, and is heavily moderated by a group of 25 moderators who ensure access to high-quality content. As of March 2019, the subreddit has about 295,000 subscribers. Together, our choice of this community is driven by the rationale that it provides us a large-scale, broad, and diverse dataset with minimal or low noise (due to content moderation) towards studying the language of LGBTQ+ individuals, and distinguishing minority stress against non-minority stress content on social media.

We collect two and a half years of posts from *r/lgbt* subreddit using Google BigQuery that hosts the entire corpus of Reddit data. We filter in posts for ones that are "text-only" and are not removed or deleted by the author or the community moderators. The final dataset contains a total of 12,645 posts. The timestamps of the posts range from January 2016 to May 2018. Fig. 1 provides the top keywords in the dataset and the distribution of post lengths. We randomly sampled 350 posts from this dataset for hand annotation for signals of minority stress.

4 ANNOTATING MINORITY STRESS ON SOCIAL MEDIA

We use the Meyer minority stress model [68] to guide our work. The model describes stress processes in three primary categories: experiences of prejudice events (such as discrimination and violence), perceived stigma (such as expectations of rejection, which often involves concealing one's identity), and internalized homophobia [68].

4.1 Annotation Approach

While Meyer's [68] minority stress conceptualizations influenced our coding and we applied them in somewhat of a directed coding fashion [53], when developing the codebook (see Table 1) we also allowed concepts and meanings to emerge from posts in somewhat of an open coding fashion [101]. The differences between Meyer's categories [68] and ours stem from three main factors: 1) difference in time ([68] was published in 1995, while our data and coding is from 2018-2019); 2) difference in identities ([68] was specifically about gay men, while our data includes LGBTQ+ identities); and 3) our data is in a computer-mediated, community context, in which people disclosed information about their LGBTQ+ identities and often received support and advice in response, whereas [68] involved closed-ended scales that participants answered individually.

Two authors independently coded 20 randomly-selected posts from the larger dataset and discussed them one by one in detail. Together, we made decisions on all posts with coding discrepancies, and revised the codebook based on agreeable themes and concepts. For example, we initially had different interpretations for posts in which minority stress was discussed in the past tense. Through careful discussion, we decided to annotate posts as minority stress only in the present tense. Another example was our decision to annotate minority stress only when the post described the poster's experience, rather than posts where the minority stress was experienced by another person (e.g, a friend). Next, we coded an additional 30 randomly-selected posts and discussed our results through the same process. We found two discrepancies throughout this process and updated the codebook accordingly: minority stress related to political discrimination, and hate targeted toward members of the LGBTQ+ community from others within the community. The total coded 50 posts produced an excellent inter-rater reliability, with an overall Cohen's kappa (κ) coefficient of 0.91. With respect to the three types of minority stress, *Prejudice Events* resulted in a κ of 0.88, *Perceived Stigma* resulted in a κ of 0.86, and *Internalized LGBTphobia* resulted in a κ of 1.0.

Next, one coder independently coded 300 posts, where for any posts that they were unsure about, the two coders discussed the post and agreed on how to code it (used to annotate 16 borderline posts). After coding was completed, the other coder rated the “Minority Stress” label on a random sample of 50 posts to validate the ratings. This post-hoc coding led to a Cohen’s κ of 0.79 with an agreement of 90.4%, which is considered substantial agreement [58], therefore validating our annotation process. Out of the net 350 annotated posts (that were to be used in training the machine learning classifier), 145 were identified to express minority stress.

Fig. 2 presents word trees related to two prominent words in the expert-coded dataset, *hate*, and *phobia*. These give a sense of how these keywords are used in the *r/lgbt* subreddit. Keywords such as “god”, “sex”, “family”, “people”, homo“phob”ia, homo“phob”ic, “biphobia”, “parents”, etc, show the context of self-experiences and discourse on the subreddit. Table 1 presents the codebook that we build and subsequently use in our next two research objectives of our work. We describe the three types of minority stress in the next subsection.

4.2 Codebook Description

The first type of minority stress in the codebook is *Prejudice Events*, which consists of explicit actions of rejection toward an individual related to their LGBTQ+ identity. This includes physical and verbal violence; legal discrimination, exclusion, or lack of basic rights; and anxiety or stress about discriminatory political powers. For example, one poster wrote about verbal abuse they experienced at school: *“Recently, I came out to my closest friends as bisexual. They spread it around like wildfire and now everyone treats me different at school. Now some jerks at school students scream faggot at me when I go to my locker. I just wanna hide in my room and cry forever.”* Another discussed physical violence and rejection from family and friends: *“I was kicked out of my online video game squad just for being gay (they said gays are pedophiles). My dad was outside my room listening when this happened so he grabbed me by the throat and kicked me out.”*

Our second minority stress codebook category is *Perceived Stigma*, which involves internalized fears and expectations of rejection related to one’s LGBTQ+ identity This category includes peoples’ fears and anxieties around how they are perceived by others, which influences their actions in several ways, such as monitoring their behavior and concealing their identity. Perceived Stigma also involves thoughts about not fitting in or not being accepted by others. As an example, one individual wrote, *“At school, I have great friends and a good family at home. But I’m a closeted gay. If I ever came out, I know my friends would never talk to me again and my family would disown me. Because of this, I have zero motivation to come out.”* Perceived stigma also includes anxiety or stress about potential future discriminatory political powers, as the following quote exemplifies: *“I’m pre-hormone transgender. And I’m not out of the closet to the whole world yet. I have ambitious plans of working for the United States government and I’m currently on track for it. I’m worried that if I come out and transition, transphobia will get in the way of me receiving this job.”*

Our third minority stress codebook category is *Internalized LGBTphobia*. This consists of internalized negative societal attitudes toward LGBTQ+ people, resulting in negative feelings toward oneself from oneself. Internalized LGBTphobia includes self-hate, negative attitudes and feelings, feelings of isolation or of ending up alone, and distaste for other LGBTQ+ people. As one example of self-hate, a poster wrote, *"I hate my face, voice, legs, chest, hands, feet, clothing, hair, my mannerisms, my perceived identity by others, my inability to cope with dysphoria, and literally everything. It's so hard to see people who are the gender I am living the life I could have had. Now in my twenties, it's too late to be my true self. If only I didn't let my confusion and embarrassment waste all those years. I will never be me, I'll always just be this thing."* Another described, *"I hate everything about myself. My head is broken anyway, so I should just bash it in with a hammer."*

The minority stress categories have a substantial amount of overlap in many of the posts. This is to be expected, as the categories are not mutually exclusive, and have been found to be significantly correlated [57]. In our analysis, we focus primarily on detecting minority stress *overall* rather than detecting subtle nuances between the three categories.

While our codebook and the examples in our dataset are representative of the broader minority stress literature as reviewed in Section 2.1, we see several differences. First, because our data includes a broad set of LGBTQ+ identities, we see a wide range of minority stressors. Some, such as anxiety about not being accepted, and being victims of discriminatory actions, are unfortunately pervasive across all LGBTQ+ identities. However, we also see that some minority stressors are perpetuated by people from some subsets of the LGBTQ+ population to other subsets, such as prejudice events in which cisgender LGBTQ+ individuals rejected transgender and/or non-binary individuals. The other primary difference in our codebook and data as compared to previous literature is the online, community-based aspect of people's posts, in which they used the subreddit as an online space in which disclosures were often a means to vent and ask for advice and support from other LGBTQ+ people. These aspects of our dataset are different than survey-based studies where minority stress was determined by people's answers to validated scales, and provide rich information that enabled us to build a classifier to detect minority stress's linguistic features.

5 DEVELOPING A CLASSIFIER TO ASSESS MINORITY STRESS

Our next goal centers around scalably inferring the presence of minority stress in social media language. We draw on natural language analysis techniques to build a machine learning classifier of minority stress using the above gathered expert-labeled annotated dataset. As any other classification methodology, our approach involves tuning both the machine learning algorithm (and corresponding parameters) and the language features.

5.1 Language Features

This paper uses a variety of features that consider the linguistic, lexical, and semantic aspects of language, which are briefly described below.

Latent Semantics (Word Embeddings).—To capture the semantics of language beyond raw keywords, we use word embeddings, which are essentially vector representations of words in latent semantic dimensions. A number of studies have revealed the potential of word embeddings in improving a number of natural language analysis and classification problems [70]. In particular, we use pre-trained word embeddings (GloVe) in

50-dimensions that are trained on word-word co-occurrences in a Wikipedia corpus of 6B tokens [85].

Psycholinguistic Attributes (LIWC).—Prior literature in the space of social media and psychological wellbeing has established the potential of using psycholinguistic attributes in building predictive models [28, 92, 100] We use the Linguistic Inquiry and Word Count (LIWC) lexicon to extract a variety of psycholinguistic categories (50 in total). These categories consist of words related to *affect, cognition and perception, interpersonal focus, temporal references, lexical density and awareness, biological concerns*, and *social and personal concerns* [103].

Hate Lexicon.—As outlined in our codebook, minority stress is often associated with offensive or hateful language used against LGBTQ+ individuals. To capture these linguistic cues, we leverage the lexicon used in recent research on online hate speech and psychological wellbeing [71, 91]. This lexicon was curated through several iterations of automated classification, crowdsourcing, and expert inspection. Among the categories of hate speech, we use binary features of presence or absence of those keywords that corresponded to *gender* and *sexual orientation* related hate speech.

Open Vocabulary (*n*-grams).—Drawing on prior work where open-vocabulary based approaches have been extensively used to infer psychological attributes of individuals [94,97], we also extracted the top 500 *n*-grams ($n = 1,2,3$) from our dataset as features.

Sentiment.—An important dimension in social media language is the tone or *sentiment* of a post. Sentiment has been used in prior work to understand psychological constructs and shifts in the mood of individuals [43, 90]. We use Stanford CoreNLP's deep learning based sentiment analysis tool [62] to identify the sentiment of a post among positive, negative, and neutral sentiment label.

Mental Health Concerns (DASS: Depression, Anxiety, Stress, and Suicidal Ideation).—Since minority stress concerns the mental health state of LGBTQ+ individuals, we use linguistic features that identify commonly associated mental health symptomatic expressions of depression, anxiety, stress, and suicidal ideation (DASS). To estimate this, we replicate the transfer learning classifiers built in recent research [92, 93]. These classifiers are trained on those subreddits that are most closely associated with each of these mental health conditions. That is, their positive examples comprise the posts shared on *r/depression* for depression, *r/anxiety* for anxiety, *r/stress* for stress, and *r/SuicideWatch* for suicidal ideation, and on the other hand the negative examples are extracted from a collated sample of 20M posts, gathered from 20 subreddits from Reddit's landing page, such as *r/AskReddit, r/aww*, and *r/movies*. These classifiers are SVM models with linear kernels, use 5000 *n*-grams (n=1,2,3) as features, and are trained on balanced datasets of positive and negative class data. They show a mean cross-validation accuracy ranging between 0.79 and 0.88, and mean test accuracy ranging between 0.81 and 0.91. When expert-assessed with the DSM-5 clinical framework, these classifiers were found to have 87% agreement with manual ratings [93]. We use these classifiers to machine label our dataset with the presence (or absence) of expressions corresponding to the above mental health attributes, and the labels are incorporated as features in the minority stress classifier.

5.2 Machine Learning Model

We use the 350 manually-coded posts from the previous section to build a machine learning classifier with a total of 659 features, as described above. We consider and evaluate multiple classifiers, including Naive Bayes, Logistic Regression, Random Forest, Support Vector Machine, and Multilayer Perceptron (MLP) algorithm. We use stratified κ-fold cross-validation ($k = 5$) to parameter tune our classifiers. Table 2 summarizes the performance metrics of these models. All of these classifiers perform better than the baseline accuracy of 58% on our dataset (based on a chance model). We find that the Neural Network based MLP classifier outperforms all with a median AUC of 0.80, median precision of 0.75, and median recall of 0.74. Table 3 summarizes the performance metrics of this classifier, where we find that the classifier is reasonably stable (stdev. = 0.03) across the five folds, and Table 4 summarizes the step-wise improvement with the addition of each kind of feature in the MLP model. For the rest of the paper, we use the MLP as the minority stress classifier.

We use K-best univariate statistical scoring model using mutual information to obtain the relative importance among features, and establish their statistical significance using ANOVA to obtain the top features of the minority stress classifier, which are reported in Table 5. Note that this table only includes the “interpretable features”, and excludes word-embedding dimensions. We find that the features obtained by the DASS classifier contributed relatively the most among the features. Many of the psycholinguistic (LIWC) categories are significant, which also aligns with the literature on a number of topics related to mental health, self-disclosure, and stress [38, 83, 92]. For instance, the affective categories, relating to *affective attributes*, such as *anxiety, negative affect, positive affect, and anxiety* show high relative importance. We also find *cognitive attributes*, such as *tentativeness, causation, inhibition, certainty, negation* also show high relative importance — these keywords are related to an individual’s cognitive functioning [82], and are known to be associated with first-hand accounts of the real world happenings, events, and experiences [13,15] — Boals and Klein found that individuals describing painful relationship breakup use more cognitive mechanisms, particularly causal words. These align with the dataset we are working on, e.g., *“[..]I told her that sometimes I want to kill myself to make those memories go away, and I “think” I am unable to “because” I’m too scared to do it, and I am scared one day I won’t be scared anymore. I cried so much that I could not breathe and almost passed out. She told me that she “understood” [..].”*

6 THEORY-DRIVEN POST-HOC ANALYSES ON MINORITY STRESS LANGUAGE

We now use the minority stress classifier to label all our 12,645 posts in our dataset. We find that 35% of these posts (4,419) are predicted to contain minority stress expression. This section first analyzes the linguistic markers associated with minority stress, and then situates them in the Minority Stress Theory [68]. Doing so, we essentially establish construct and face validity of the linguistic markers in the context of LGBTQ+ minority stress experiences. Finally, conducting an error-analysis we relook at the intricacies of this classifier.

6.1 Analyzing Language Cues Associated with Minority Stress Theory

6.1.1 Finding discriminating language cues in the language of minority stress.—We first examine the language markers associated with minority stressors. We employ an unsupervised language modeling technique known as the Sparse Additive Generative Model (SAGE) [37]. Given any two documents, SAGE selects discriminating keywords by comparing the parameters of two logistically parameterized multinomial models, using a self-tuned regularization parameter to control the tradeoff between frequent and rare terms. We use SAGE to identify discriminating *n*-grams (*n*=2,3,4) between the posts of present and absent minority stress. The magnitude of SAGE value of a linguistic token signals the degree of its "uniqueness", and in our case a positive SAGE (more than 0) indicates that the *n*-gram is more representative for the presence of minority stress, whereas a negative SAGE denotes greater representativeness for its absence.

6.1.2 What do the discriminating keywords say?—Table 6 reports the top 30 discriminating keywords that occur in present and absent minority stress posts. We find that keywords such as *tell mom, telling people, started talking*, and *say things* occur more frequently in the posts that express Minority Stress. These keywords are related to sharing and disclosing, and sometimes to individuals who are closely related to the discloser, such as "mom", "dad", or "friends", such as in *"I'm a coward to be unable "tell my mom" to her face that I'm gay,"* These resonate strongly with prior work that people who faced parental rejection when disclosing their LGBTQ+ identity suffered greater psychological distress [86], on the difficulties around disclosing LGBTQ+ identity [45], and the LGBTQ+ individuals' need to weigh costs vs. benefits when deciding to reveal or conceal their identity, primarily due to stigma [16, 22]. We also observe keywords that express self-experiences and life events, such as "started feeling", "days later", *didn't feel*, primarily occur in the posts that have minority stress expressions. For example, a post says, *""started feeling" these issues when I was 16 or 17, and because of my parents, I was like, 'No I won't do this' and shoved them back down again."* These keywords relate to Meyer's finding that concealing one's LGBTQ+ identity for long periods of time, an important coping strategy for many, can lead to increased stress [69].

In contrast, keywords relating to support and gratefulness, such as *thanks (in) advance, thanks (for) reading*, and *greatly appreciated* occur in abundance in posts that do not express minority stress. Such posts are mostly about seeking advice such as on localities in towns, or relationship, *"any tips on flirting and just befriending her in general would be "greatly appreciated""*. We also notice many generic keywords or the group names, such as *gay lesbian, lgbtq community*, and *lgbt friendly* are more frequent in the posts that do not contain minority stress. This could be associated with the fact that these individuals who are more openly LGBTQ+ are receiving the positive benefits of disclosure and thus are likely to be less distressed [73]. Additionally, these posts are plausibly by those who want to disseminate the sense of community, diversity, and inclusiveness, for example, *" I thank the*

LGBTQ+ community for being together, through friends and social media," and *"What is your resolution or hope for the LGBTQ community[..]"*.

Together, one contrasting theme observed in the posts that express minority stress compared to the ones that do not, is that while the former is mostly "personal" and "self-experiences", the latter is to raise awareness, and talk about less-personal and more general issues related to the gender and sexual minorities. These posts are less likely to relate directly to the individual's mental state, and relate instead to the community broadly.

6.1.3 Aligning the language of minority stressors with the minority stress theory.—

We focus on the top discriminating keywords associated with minority stress (as obtained via SAGE analysis) because these are most likely to be the linguistic markers of minority stressors and minority stress. For every post in our dataset, we obtain the cosine similarity of word embedding representations [70, 85, 94] with the descriptions of each category in our codebook— *Prejudice Events, Perceived Stigma*, and *Internalized LGBTphobia*. We use 300-dimensional lexico-semantic latent space of word vectors (pre-trained on the Wikipedia corpus of 6B tokens [85]). We label each post with a high propensity of belonging to those minority stress categories, where their similarity is above a certain threshold (0.80) [88] (see Fig. 4b for the distribution of minority stress categories and their overlap). Then, on the basis of frequency distribution with respect to minority stress categories per post, we obtain those tokens that stand out in minority stress language. Fig. 4a plots this distribution, where the radial bar plots reveal the probabilistic likelihood percentage per keyword in each of the categories.

We find that many keywords show very similar frequency distribution across the three categories. This could be because most posts on *r/lgbt* are long and explain multiple issues related to individuals' self-experiences, which is also why multiple categories of minority stress are co-morbid on the posts (see Section 4). Now per category, we look at the most frequent keywords, to understand the language associated with different types of minority stress.

Prejudice Events.: Keywords like *didnt want, didnt feel*, and *didnt say*, occur with greater than 20% probability in this category. All of these contain a negation followed by an action word. We conjecture that these are related to describing life events where the individual experienced unpleasant, violent, or nonconsensual activities resulting from societal prejudice, eg., *"I tried to explain that it wasn't really consensual, and I didn't want it"*. We find that *gay people*, and *gay person* occur heavily in posts expressing Prejudice Events: *"whatever that religious people have done and said about women, and specifically "gay people" is extremely sad. Too hurtful. Too stupid!'*.

Perceived Stigma.: Just like in the case of *prejudice events, perceived stigma* category also includes negated action verbs (*didnt want, didnt feel*, and *didnt think*). For instance, *"I didn't feel very comfortable around my coworkers despite their friendliness."* Literature in psycholinguistics and expressive writing found that negation has a high correlate with inhibition [23, 47]. Inhibition is related to much of the Perceived Stigma section of the codebook (see Table 1), which involves shifting one's behavior and concealing one's identity in anticipation of potentially being rejected by others. Keywords that highlight temporal events, such as *started talking, months after, started feel, thought gay* are also prominent in this category. Temporal keywords are indicators of discourse on self-disclosure on mental health [27, 103]: *"I started to feel tense when I asked that [..]."*

Internalized LGBTPhobia.: Keywords such as *want live* and *feel bad* that express the feelings are also prominent in this type of minority stress, for example, *"I "want to live" and be free as boys and girls that are allowed to express themselves."* Internalized LGBTphobia has been discussed as an internalization of the prejudice experienced by LGBTQ+ people, and may be an antecedent of psychological distress [107]. The keywords in this category about wanting to live and feeling bad may signal this internalization of prejudice in which one becomes hyper-focused about their own feelings and emotions. In addition, the presence of keywords such as *im gay, thought gay*, and *didn't feel* could be indicative of the fact that this category is more about *self-focused* behavior and distress, for example *"My biggest issue with this is that it paints a bad picture of the LGBT community and that my crush might avoid me because "im gay" and not interested in girls."*

6.2 Error Analysis on the Minority Stress Classifier

This section revisits our classification task, and drills deeper into the feature-level nuances to understand how and what linguistic markers help improve the accuracy, or alternatively what factors contribute towards misclassifications. Our analyses are inspired by error analysis techniques in social media language analysis research [19, 25]. We quantitatively identify posts with very similar lexical and semantic characteristics, but contrasting outcomes on minority stress expressions, and then qualitatively examine the differences and similarities in social media language of LGBTQ+ individuals that contribute in (mis)classifying the minority stress expressions.

As observed previously, the top features in our classifiers correspond to psycholinguistic attributes and word-embedding dimensions. For every post in our expert-labeled dataset, we repurpose their vector representation across the psycholinguistic and word-embedding dimensions to obtain its pair-wise similarity with other posts. We refer to the confusion matrix (Fig. 3c), and study cases of False Positives (FPs) and False Negatives (FN), against cases of True Positives (TPs) and True Negatives (TNs) in our pooled κ-fold cross-validation ($k = 5$) classification task.

6.2.1 Analyzing the False Positives.—First, we conduct a precision-centric analysis, which investigates the FP cases. We obtain the pair-wise similarity of all the posts in TP and FP, and those that are in TN and FP. We particularly look at those posts that occur in the top 10-percentile of both the similarities. These are essentially those posts which are extremely similar in their language, and are all identified to have minority stress, however, 16 of these posts do not actually express minority stress. We find the following commonly occurring themes when we look deeper into these examples.

Non Self-Experiences.: We find that the FP posts include *posts that are not self-experiences of minority stress*. Although our psycholinguistic features include all kinds of pronouns (1st, 2nd, and 3rd person), our classifier is unable to discriminate such occurrences of self-experiences of minority stress. Co-reference and semantic role labeling techniques can help in predicting such instances correctly [87]. An example excerpt, *“..my best friend is a trans-man who has been constantly abused [..] is now taking anxiety medication and have managed to survive thus far. Does someone know of any safe spaces for trans-men who are recovering from abuse or from suicide?.”*

Past Experiences.: We find the presence of FP posts where the individuals share about their past experiences of minority stress, however, they have “recovered” from the same in the present (see example below). Such instances are classified as the presence of minority stress experience in our classifier, as it is unable to incorporate the temporal discourse of events. NLP techniques such as temporal discourse parsing may help the classifier in identifying such examples [52], such as in, *..had closeted feelings for a long time and obviously couldn’t do anything about them since I grew up in a conservative household. I felt I was doomed trying to make it in a world where the more masculine and assertive man dominates while the rest of us have to sit by on the sidelines..*

Seeking Relationship Advice.: We find a few FP instances that are similar in lexico-semantics with the TP, especially because they disclose about their feelings or problems associated with belonging to the minority communities, but these posts do not explicitly express about minority stress per se, and rather seek advice or support regarding some aspect of their lives, such as relationship: *I mostly kept my sexuality to myself until now, and I became lonely and depressed [..]. I don’t think I’m straight [..] I don’t like the idea of sex though I fantasize about it at times, but I can’t see myself doing it. It’s only been a week since I came to this realization, but the evidence is very one-sided. I like women a lot. I enjoy being around them and sometimes being intimate. I just know that I also find men attractive… I’m wondering if I should tell this to a person I know who seems cool and discrete or discuss it with my therapist first?*

6.2.2 Analyzing the False Negatives.—

Next, we conduct a recall-centric analysis, which investigates the FN cases. We obtain the pair-wise similarity of all the posts in TPs and FNs (Fig. 4c). As above, we look at those posts that occur in the top 10-percentile of the similarities — these include 52 TP and 22 FN posts. These are essentially those posts that are extremely similar in their lexico-semantics, however, the FNs were (incorrectly) not caught by our classifier as expressing minority stress.

Asking question(s).: We find a few FN examples that are very similar to the TPs, however, these posts do not explicitly state their minority stress experiences, rather ask question(s), such as on perceived stigma. For example, *“Being bisexual, I feel hate in every group of people, even the LGBTQ+ community. The LGBTQ+ community is so judgy towards me for being attracted to both men and women. It’s really hard feeling like I’m not accepted anywhere except when with other bisexual people. Is the LGBTQ+ community is really accepting?.”*

Lack of context and non-explicit expression of self-experience.: FNs may occur in posts that do not make explicit statements of minority stress experiences. Although these posts are very similar to the TPs in terms of the keyword use, the classifier is unable to understand their underlying context, e.g., *" Guys! I'm in need of a chest binder. I'm in extreme dysphoria that badly affects my daily life. I want someone to donate one and be my lifesaver.."*

Typoes and variants of slangs.: We also find FNs when certain words are misspelled. NLP preprocessing techniques of normalizing keywords or spell correctors may help in overcoming misclassifications, e.g., *"I'm a young closeted lesbian in college. I just moved to a new town and know nobody here. I finally made a new friend, who seemed awesome! Today, mid conversation while discussing about outfit, she turns to me and says, "At least I don't look like a* ***dke*** *though I guess!" Haha! :)I'm done. Giving up. I'll never find my people. Sorry I just need to vent somewhere."*

7 DISCUSSION

This paper provides a novel theoretically-grounded approach to assess minority stress in the discussions shared on social media by LGBTQ+ minorities. Notably, our minority stress codebook, a key contribution of this work, provides a novel theoretically-grounded approach to characterize minority stressors in the discussions shared on social media by self-identifying LGBTQ+ minorities. It builds from Meyer's work in three ways: includes a broad range of LGBTQ+ identities, includes stressors perpetuated by some subsets of the LGBTQ+ population to other subsets, and involves an online, community-based aspect. These characteristics apply to many LGBTQ+ communities across the Internet on sites such as Instagram, Tumblr, Twitter, Facebook groups, YouTube, and TrevorSpace. While they are yet to be validated (in future research), we believe that the results from our work extend far beyond Reddit and would be of great use in designing interventions to help LGBTQ+ people on a variety of online platforms.

Technically, the generalizability of our classifier (across online communities and social media platforms) is motivated by the success of transfer learning methodologies used in a number of recent work [10, 92, 93]. These studies trained supervised machine learning classifiers on one (domain-specific) dataset, and applied them on another unlabeled dataset (including other platforms). One reason that these classifiers work is because language across social media platforms is not very different, and if linguistic equivalence (with methods proposed in the above prior work) is established between the training and unlabeled datasets, then the machine learning classification works reasonably well with minimal dataset-specific customization.

7.1 Social Media Interventions for LGBTQ+ Minorities

Currently, most of existing LGBTQ+ online communities only serve as a safe networking place but are missing or equipped with the very limited capacity to proactively identify individuals' risk to or experience of different minority stressors. Even on the subreddit considered here or other prominent online social networking sites for LGBTQ+ youth, such as TrevorSpace, proactive intervention, i.e., referring vulnerable individuals to a hotline or instant messaging service, is provided based on forum administrators' or moderators' observations on the community discussions with little decision support. Risk assessment as a manual process is labor intensive and costly given the increasingly large online communities, and the severity and urgency that might underscore many of the calls for help or support. Our methods and findings may be utilized to close these gaps and expand existing efforts as described below:

(1) Moderation and Support Matching Efforts.—LGBTQ+ individuals whose content contain phrases and other linguistic constructs relating to minority stressors (as also revealed in the linguistic markers in Section 6), as revealed by our methods, may be flagged in the interfaces of moderators and other clinical experts for help and support. Community moderators may also be allowed to maintain a “risk list” in their interfaces that would include individuals forecasted by our methods to exhibit signs of minority stress. This would allow improved preparedness to bring timely and tailored help to those in need. Further, on being informed that an individual in the community could be experiencing minority stress, moderators and experts may make provisions to connect them with appropriate mental health resources including The Trevor Project, which provides a national, 24-hour confidential suicide hotline for LGBTQ+ youth, online chat and confidential text messaging applications, and Trans Lifeline, a crisis hotline for transgender people. Additionally, trusted peers in the communities could be “matched” to such content who are knowledgeable about the specific cultural considerations and issues faced by LGBTQ+ individuals with mental health challenges, and platform affordances to field private messages with relevant information on help-seeking or therapy can be incorporated in the moderation and support matching efforts.

(2) Therapeutic Efforts.—Self-disclosures and expressive writing on online platforms are associated with positive therapeutic outcomes [38], because these platforms allow people to vent and give structure to stigmatized experiences [6]. Our results show that these observations are particularly valid for the community we study in this paper; individuals who find the open and feel safe enough to express experiences of minority stress. Accordingly, we envision that journaling tools that support positive therapeutic experiences may be built and integrated directly or indirectly with social media platforms, wherein posts that document one’s minority stress experiences could be logged voluntarily, serving as a timestamped archive of one’s thoughts, feelings, and experiences around holding a gender/sexual minority identity. The psychotherapy literature has identified the unique benefits of such archival writing. For instance, it can help an individual develop a logical narrative of events, experiences, and mental health challenges, and thereby enable them to meet self-care and coping goals [84]. Tumblr used to be an important social media site where this type of journaling took place, but with recent policy changes of the platform, some LGBTQ+ people feel less welcome on Tumblr [2]. Therefore, new online spaces and applications that integrate with them are needed for networked journaling so that LGBTQ+ people can receive therapeutic benefits.

Journaling tools integrated with social media can further allow LGBTQ+ individuals to be self-reflective and more empowered: reframing minority stress experiences have been known

[2] washingtonpost.com/technology/2018/12/04/before-tumblr-banned-adult-content-it-was-safe-space-exploring-identity/

to partially mitigate the negative impact of prejudicial and discriminatory environments and support healthy identity development [54]. The archives logged through this journaling tool can also complement counseling efforts tailored to the sexual/gender minority experience. When shared with a therapist, the archives can aid them "by entering the client's mental constructs via the written word" [7], or by understanding those thoughts and feelings which the client might be "unable to vocalise" [24].

7.2 Public Health Implications

Health Disparities of a "Hard to Reach" Population.—The rise and pervasiveness of new channels of communication such as social media have brought both new opportunities and challenges for public health professionals. Many scholars have noted LGBTQ+ individuals to constitute a "hard to reach" population [67]. Consequently, existing public health research suffers from issues of over-sampling of the "visible" sections of this hidden population and an inability to capture diverse, complex emotions [64], as noted in Sections 1 and 2.

We believe that the approach and results presented in this paper will provide important data to inform evidence-based decision support systems for optimized public health understanding and decision-making, as well as for designing mental health intervention tailored for LGBTQ+ people, especially those who use online communities of the type considered here. Our research can further support documenting, understanding, and addressing the environmental factors that contribute to health disparities in the LGBTQ+ community, by generating diverse samples in terms of sexuality and gender identities and a different type of data, immediate and unmediated by researchers.

Clinical and Therapeutic Practice.—The American Counseling Association's Code of Ethics states, "Counselors gain knowledge, personal awareness, sensitivity, and skills pertinent to working with a diverse population" [8], still, some have suggested that training programs that focus on LGBTQ+ affirmative counseling lack a focus on empirically informed treatments [55]. Our findings can be used to guide positive public health policy change relating to gender and sexual minorities, including existing clinical practices and training programs targeted at this population. Our data can also provide a new source of information that can be appropriated to make these practices more (LGBTQ+) client-centered, such that the clinicians are cognizant of the insidious challenges these individuals face, can be more aware in the examination of biases and values underlying LGBTQ+ experiences, as well as can implement appropriate intervention strategies.

Minority Stress Research.—In addition, the results can guide future research in the area of LGBTQ+ mental health by identifying variables that have received little attention in previous studies or formulation of minority stressors in existing research [35]. More concretely, understanding how LGBTQ+ people communicate online about minority stressors provides first steps for indicators of a population-level understanding of the types and amounts of minority stress experienced by LGBTQ+ people more broadly. Future work could fine-tune these analyses and consider how prevalent minority stress categories may be for LGBTQ+ people at population-levels.

7.3 Ethics

We recognize this work concerns a sensitive subject area presenting ethical, methodological and epistemological challenges (as in [67]), because it focuses on self-reports of stigmatizing experiences of a marginalized population (LGBTQ+ individuals' minority stress).

We note that an automated way to gauge minority stressors in shared content might call upon negative impacts such as social discrimination and rejection, or even reinforce some of the very minority stressors like prejudice and stigma, that are detrimental to well-being. These issues can amplify if the minority stress classifier is misused in situations or by bad actors where the individual may not desire to have their minority stressors quantified despite discussing their distresses online, or may prefer to keep their gender or sexual identities private. We also acknowledge an apparent tension – an LGBTQ+ individual, who has not come out yet, may not wish to have their minority stressors identified on social media, but at the same time, may still wish to receive help, advice, and support to deal with their everyday struggles. Our classifier will not be able to support the needs of such individuals, or more broadly, those who are not present in the types of online communities we focus on in this work.

Additionally, there are ethical complexities associated with employing automated classification-based monitoring of minority stressors on an online platform. First, employing algorithmic methods may silence the speech of those stigmatized individuals who adopt these online platforms as safe spaces for LGBTQ+ conversations, but resist their data being computationally analyzed. Beyond Reddit, there are ethical complexities associated with using the classifier on other social media platforms or any non-anonymous user feeds (such as on Twitter or Facebook), where individuals might prefer not to disclose their sexual identity. Further, by its nature as a classifier of textual content, our approach, if used to allocate resources such as support, may not allocate these resources correctly. It can only detect minority stress based on the text that people choose to share, it likely is more accurate on the text that directly rather than indirectly conveys minority stress, and it involves false negatives (people who may benefit from support but are not identified correctly by the classifier). Additionally, people who face minority stress cope differently, and require different amounts of support and resources; our automated approach would not know how to best allocate these resources. Taken together, these ethical considerations lead us to ask: How do we ensure that a minority stress classifier is used in a way that is ethically compliant with, and does not cause harm to, the individuals and communities whose data it analyzes, and also be appropriately leveraged to extend timely help and support to the same individuals and communities?

We believe that the potential tools discussed under design implications, and their intended efficacy need to be cognizant of the preferences, goals, and values of the concerning LGBTQ+ individuals, so as to lessen unintended consequences. One way to ensure this is an opt-in approach, in which members of an online community are given the choice to explicitly opt in or out of a mental health intervention that might make use of their data and offer them support in times of distress. Similarly, for the journaling tool, we indicated the possibility of sharing of archives with a therapist. These design approaches need to factor in boundary regulation considerations in the client-therapist interpersonal relationship, and need to develop adequate data and informational abstractions to manage LGBTQ+ clients' privacy expectations.

Additionally, as developers of this classifier, we maintain ethical responsibility to limit its use to applications that will benefit rather than harm LGBTQ+ individuals and communities. Of course, it would be naive to assume that this fully mitigates the risks described above, but this approach does reduce unintended consequences. Nevertheless, any decision and interventions using our classification approach require careful and in-depth supplemental ethical analysis, beyond the empirical analysis we present in this paper. We acknowledge that this classifier cannot be used as-is, or for direct intervention; instead, it is meant to complement and assist human intervention, where content that is computationally classified as containing minority stress will be inspected by human experts before any action is taken.

7.4 Limitations and Future Directions

We acknowledge that our work has limitations, many of which suggest interesting directions for future research. We do not make any population-centric assessments because the subreddit considered in our work cannot be considered wholesome of online discussions of LGBTQ+ individuals. Rather, our work should be seen as a proof-of-concept study to examine minority stress language on social media. Future work that studies community dynamics and makes population-centric assessments associated with LGBTQ+ individuals, should also consider the caveats concerning missingness and quality of social media datasets, including [42].

Our work inherently suffers from self-selection biases, that it only works on the language of the individuals who self-select to express themselves on online communities, particularly those that are LGBTQ+ friendly. Relatedly, we only study the language in minority stress expressions on social media. Incorporating other behavioral and communicative signals like frequency of posting, topic of interest, and support-seeking or support-giving nature of posting, can help us to comprehensively understand minority stress on social media. Future work can investigate these disclosures on minority self-experiences across other online communities and social media platforms.

As we briefly discussed in our posthoc and error analyses, the classifiers can be further improved with more sophisticated models of machine learning and natural language processing. This can be tuned with respect to the objective of the problem, where our objective was to balance between predictability and interpretability — i.e., to not only build a stable model that reveals the potential in machine learning to scalably infer the language of minority stress, but also to help us understand the linguistic nuances in expressing minority stress on social media.

Building intersectionality and diversity into strategies to conduct LGBTQ+ mental health research is an important goal [78]. A limitation of this work is that we have not considered how minority stressors might manifest themselves differently among gender and sexual minorities with intersecting identity facets. Further, use of the queer theory to understand LGBTQ+ experiences is advocated, because it builds upon the idea that gender is part of the essential self and emphasizes the socially constructed nature of sexual acts and identities [60]. Situating our minority stressor identification approach in these theoretical lenses constitute an important area for future work.

8 CONCLUSION

This paper studied the language of minority stress experiences of LGBTQ+ identities on social media. Drawing on Meyer's minority stress theory, and adopting a combined qualitative and computational approach, this paper examined the language on a LGBTQ+ online community on Reddit (*r/lgbt* subreddit), and makes three primary contributions. First, a theoretically grounded *codebook* to identify minority stressors across three types of minority stress— *prejudice events, perceived stigma*, and *internalized LGBTphobia*. Second, a *machine learning classifier* to identify social media posts expressing minority stress experiences at scale. The classifier used a variety of features, ranging across word embeddings, psycholinguistic attributes, hateful keywords, sentiment, and open-vocabulary based *n*-grams, and achieved a mean AUC of 0.80. Finally, we conducted deeper post-hoc analysis on minority stress language to obtain *lexicons* of linguistic markers, along with their contextualization in the minority stress theory. We believe our work bears the potential to help understand the prevalence of minority stress in online discussions, and support tailored interventions sensitive to the needs of LGBTQ+ individuals and communities.

ACKNOWLEDGEMENT

This project is partly supported by research grants from Mozilla (RK677) and NIH (R01GM112697).

REFERENCES

[1]. Ahmed Alex A.. 2018 Trans Competent Interaction Design: A Qualitative Study on Voice, Identity, and Technology. Interacting with Computers 30 (2018).

[2]. Andalibi Nazanin and Forte Andrea. 2018 Announcing Pregnancy Loss on Facebook: A Decision-Making Framework for Stigmatized Disclosures on Identified Social Network Sites In Proc. CHI (CHI '18). New York, NY, USA.

[3]. Andalibi Nazanin, Haimson Oliver L., De Choudhury Munmun, and Forte Andrea. 2018 Social Support, Reciprocity, and Anonymity in Responses to Sexual Abuse Disclosures on Social Media. ACM TOCHI 25 (2018).

[4]. Andalibi Nazanin, Haimson Oliver L, De Choudhury Munmun, and Forte Andrea. 2016 Understanding social media disclosures of sexual abuse through the lenses of support seeking and anonymity. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. ACM, 3906–3918.

[5]. Andalibi Nazanin, Morris Margaret E., and Forte Andrea. 2018 Testing Waters, Sending Clues: Indirect Disclosures of Socially Stigmatized Experiences on Social Media. PACM HCI 2, CSCW (11 2018), 1–23.

[6]. Andalibi Nazanin, Ozturk Pinar, and Forte Andrea. 2017 Sensitive Self-disclosures, Responses, and Social Support on Instagram: The Case of #Depression In Proc. CSCW New York, NY, USA.

[7]. Anthony K. 2000 Information Technology. Counselling in cyberspace. COUNSELLING-RUGBY-, 11 10 (2000), 625–627.

[8]. American Counseling Association. 2005 ACA code of ethics.

[9]. Auerswald Colette L, Greene Karen, Minnis Alexandra, Doherty Irene, Ellen Jonathan, and Padian Nancy. 2004 Qualitative assessment of venues for purposive sampling of hard-to-reach youth: an illustration in a Latino community. Sexually transmitted diseases 31, 2 (2004), 133–138.

[10]. Bagroy Shrey, Kumaraguru Ponnurangam, and De Choudhury Munmun. 2017 A Social Media Based Index of Mental Well-Being in College Campuses. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems.

[11]. Bazarova Natalya N., Choi Yoon Hyung, Whitlock Janis, Cosley Dan, and Sosik Victoria. 2017 Psychological Distress and Emotional Expression on Facebook. Cyberpsychology, Behavior, and Social Networking 20, 3 (2017).

[12]. Blackwell Lindsay, Hardy Jean, Ammari Tawfiq, Veinot Tiffany, Lampe Cliff, and Schoenebeck Sarita. 2016 LGBT Parents and Social Media: Advocacy, Privacy, and Disclosure During Shifting Social Movements In Proc. CHI (CHI '16). New York, NY, USA.

[13]. Boals Adriel and Klein Kitty. 2005 Word use in emotional narratives about failed romantic relationships and subsequent mental health. Journal of Language and Social Psychology 24, 3 (2005), 252–268.

[14]. Brooks Virginia R.. 1981 Minority Stress and Lesbian Women. Lexington Books, Lexington, Mass.

[15]. Brubaker Jed R, Kivran-Swaine Funda, Taber Lee, and Hayes Gillian R. 2012 Grief-Stricken in a Crowd: The Language of Bereavement and Distress in Social Media. In ICWSM.

[16]. Cain Roy. 1991 Stigma Management and Gay Identity Development. Social Work 36, 1 (1991), 67–73.

[17]. Cannon Yuliya, Speedlin Stacy, Avera Joe, Robertson Derek, Ingram Mercedes, and Prado Ashely. 2017 Transition, Connection, Disconnection, and Social Media: Examining the Digital Lived Experiences of Transgender Individuals. Journal of LGBT Issues in Counseling 11 (2017). 10.1080/15538605.2017.1310006

[18]. Cavalcante Andre. 2018 Tumbling Into Queer Utopias and Vortexes: Experiences of LGBTQ Social Media Users on Tumblr. Journal of Homosexuality 0 (2018).

[19]. Chancellor Stevie, Kalantidis Yannis, Pater Jessica A, De Choudhury Munmun, and Shamma David A. 2017 Multimodal Classification of Moderated Online Pro-Eating Disorder Content. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems ACM, 3213–3226.

[20]. Chancellor Stevie, Lin Zhiyuan, Goodman Erica L, Zerwas Stephanie, and De Choudhury Munmun. 2016 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 ACM, 1171–1184.

[21]. Cho Alexander. 2017 Default publicness: Queer youth of color, social media, and being outed by the machine. New Media & Society 20, 9 (2017).

[22]. Corrigan Patrick and Matthews Alicia. 2003 Stigma and disclosure: Implications for coming out of the closet. Journal of Mental Health 12, 3 (1 2003), 235–248. 10.1080/0963823031000118221

[23]. Creswell J David, Lam Suman, Stanton Annette L, Taylor Shelley E, Bower Julienne E, and Sherman David K. 2007 Does self-affirmation, cognitive processing, or discovery of meaning explain cancer-related health benefits of expressive writing? Personality and Social Psychology Bulletin (2007).

[24]. Cullen D. 2000 A byte size study of online counselling: who is doing it and what is lit like. Ph.D. Dissertation. Ph. D. Dissertation. MSc dissertation, University of Bristol.

[25]. Culotta Aron. 2014 Estimating county health statistics with Twitter. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems ACM, 1335–1344.

[26]. Darwin Helana. 2017 Doing Gender Beyond the Binary: A Virtual Ethnography. Symbolic Interaction 40, 3 (2017).

[27]. De Choudhury Munmun and De Sushovan. 2014 Mental Health Discourse on reddit: Self-Disclosure, Social Support, and Anonymity. In ICWSM.

[28]. Munmun De Choudhury Michael Gamon, Counts Scott, and Horvitz Eric. 2013 Predicting depression via social media. In International AAAI Conference on Web and Social Media.

[29]. De Choudhury Munmun and Kiciman Emre. 2017 The language of social support in social media and its effect on suicidal ideation risk. In Eleventh International AAAI Conference on Web and Social Media.

[30]. De Choudhury Munmun, Kiciman Emre, Dredze Mark, Coppersmith Glen, and Kumar Mrinal. 2016 Discovering shifts to suicidal ideation from mental health content in social media In Proc. CHI

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

[32]. Denton F Nicholas, Scales Rostosky Sharon, and Danner Fred. 2014 Stigma-related stressors, coping self-efficacy, and physical health in lesbian, gay, and bisexual individuals. Journal of Counseling Psychology 61, 3 (2014), 383.

[33]. Devito Michael A, Marie Walker Ashley, and Birnholtz Jeremy. 2018 “Too Gay for Facebook”: Presenting LGBTQ+ Identity Throughout the Personal Social Media Ecosystem. PACM HCI CSCW (2018).

[34]. Diaz Rafael M, Ayala George, Bein Edward, Henne Jeff, and Marin Barbara V. 2001 The impact of homophobia, poverty, and racism on the mental health of gay and bisexual Latino men: findings from 3 US cities. American journal of public health 91, 6 (2001), 927.

[35]. Dillon Frank R, Worthington Roger L, Bielstein Savoy Holly, Craig Rooney S, Becker-Schutte Ann, and Guerra Rachael M. 2004 On becoming allies: A qualitative study of lesbian-, gay-, and bisexual-affirmative counselor training. Counselor Education and Supervision 43, 3 (2004), 162–178.

[36]. Duguay Stefanie. 2014 “He has a way gayer Facebook than I do”: Investigating sexual identity disclosure and context collapse on a social networking site. New Media & Society (9 2014), 1461444814549930.

[37]. Eisenstein Jacob, Ahmed Amr, and Xing Eric P. 2011 Sparse additive generative models of text. (2011).

[38]. Kiranmai Ernala Sindhu, Rizvi Asra F., Birnbaum Michael L., Kane John M., and De Choudhury Munmun. 2017 Linguistic Markers Indicating Therapeutic Outcomes of Social Media Disclosures of Schizophrenia. PACM HCI 1, CSCW (2017).

[39]. Farber Rebecca. 2017 ‘Transing’ fitness and remapping transgender male masculinity in online message boards. Journal of Gender Studies 26, 3 (5 2017), 254–268.

[40]. Fisher Celia B and Mustanski Brian. 2014 Reducing health disparities and enhancing the responsible conduct of research involving LGBT youth. Hastings Center Report (2014).

[41]. Frost David M, Lehavot Keren, and Meyer Ilan H. 2015 Minority stress and physical health among sexual minority individuals. Journal of behavioral medicine 38, 1 (2015), 1–8. [PubMed: 23864353]

[42]. Gaffney Devin and Nathan Matias J. 2018 Caveat emptor, computational social science: Large-scale missing data in a widely-published Reddit corpus. PloS one (2018).

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

[44]. Haimson Oliver L.. 2018 Social Media as Social Transition Machinery. Proc. ACM Hum.-Comput. Interact. 2, CSCW (2018).

[45]. Haimson Oliver L., Brubaker Jed R., Dombrowski Lynn, and Hayes Gillian R.. 2015 Disclosure, Stress, and Support During Gender Transition on Facebook In Proc. CSCW (CSCW ‘15). New York, NY, USA.

[46]. Haimson Oliver L., Brubaker Jed R., Dombrowski Lynn, and Hayes Gillian R.. 2016 Digital Footprints and Changing Networks During Online Identity Transitions In Proc. CHI (CHI ‘16). New York, NY, USA, 2895–2907.

[47]. Hancock Jeffrey T, Landrigan Christopher, and Silver Courtney. 2007 Expressing emotion in text-based communication. In Proceedings of the SIGCHI conference on Human factors in computing systems ACM, 929–932.

[48]. Hardy Jean and Lindtner Silvia. 2017 Constructing a Desiring User: Discourse, Rurality, and Design in Location-Based Social Networks In Proc. CSCW (CSCW ‘17). New York, NY, USA.

[49]. Hawkins Blake and Giesking Jack. 2017 Seeking ways to our transgender bodies, by ourselves: Rationalizing transgender-specific health information behaviors. Proc. Association for Information Science and Technology 1 (2017).

[50]. Hendricks Michael L. and Testa Rylan J.. 2012 A conceptual framework for clinical work with transgender and gender nonconforming clients: An adaptation of the Minority Stress Model. Professional Psychology: Research and Practice 43, 5 (10 2012).

[51]. Homan Christopher M, Lu Naiji, Tu Xin, Lytle Megan C, and Silenzio Vincent. 2014 Social structure and depression in TrevorSpace In Proc. CSCW

[52]. Hovy Eduard, Mitamura Teruko, Verdejo Felisa, Araki Jun, and Philpot Andrew. 2013 Events are not simple: Identity, non-identity, and quasi-identity. In Workshop on events: Definition, detection, coreference, and representation 21–28.

[53]. Hsieh Hsiu-Fang and Shannon Sarah E.. 2005 Three Approaches to Qualitative Content Analysis. Qual. Health Res (2005).

[54]. Israel Tania, Gorcheva Raya, Burnes Theodore R, and Walther William A. 2008 Helpful and unhelpful therapy experiences of LGBT clients. Psychotherapy Research (2008).

[55]. Israel Tania, Gorcheva Raia, Walther William A, Sulzner Joselyne M, and Cohen Jessye. 2008 Therapists' helpful and unhelpful situations with LGBT clients: An exploratory study. Professional Psychology: Research and Practice (2008).

[56]. Kelleher Cathy. 2009 Minority stress and health: Implications for lesbian, gay, bisexual, transgender, and questioning (LGBTQ) young people. Counselling psychology quarterly 22, 4 (2009), 373–379.

[57]. Kelleher Cathy. 2009 Minority stress and health: Implications for lesbian, gay, bisexual, transgender, and questioning (LGBTQ) young people. Counselling psychology quarterly 22, 4 (2009), 373–379.

[58]. Richard Landis J and Koch Gary G. 1977 The measurement of observer agreement for categorical data. Biometrics (1977).

[59]. Lick David J, Durso Laura E, and Johnson Kerri L. 2013 Minority stress and physical health among sexual minorities. Perspectives on Psychological Science 8, 5 (2013), 521–548.

[60]. Lovaas Karen. 2013 LGBT studies and queer theory: New conflicts, collaborations, and contested terrain. Routledge.

[61]. Magee Joshua C., Bigelow Louisa, DeHaan Samantha, and Mustanski Brian S.. 2012 Sexual Health Information Seeking Online A Mixed-Methods Study Among Lesbian, Gay, Bisexual, and Transgender Young People. Health Education & Behavior 39, 3 (6 2012), 276–289.

[62]. Manning Christopher, Surdeanu Mihai, Bauer John, Finkel Jenny, Bethard Steven, and McClosky David. 2014 The Stanford CoreNLP natural language processing toolkit. In Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations 55–60.

[63]. Marshal Michael P, Dietz Laura J, Friedman Mark S, Stall Ron, Smith Helen A, McGinley James, Thoma Brian C, Murray Pamela J, D'Augelli Anthony R, and Brent David A. 2011 Suicidality and depression disparities between sexual minority and heterosexual youth: A meta-analytic review. Journal of adolescent health (2011).

[64]. Martinez Omar, Wu Elwin, Shultz Andrew Z, Capote Jonathan, López Rios Javier, Sandfort Theo, Manusov Justin, Ovejero Hugo, Carballo-Dieguez Alex, Chavez Baray Silvia, et al. 2014 Still a hard-to-reach population? Using social media to recruit Latino gay couples for an HIV intervention adaptation study. J. Med. Internet Res (2014).

[65]. McDermott Elizabeth. 2015 Asking for help online: Lesbian, gay, bisexual and trans youth, self-harm and articulating the 'failed'self. Health: (2015).

[66]. McDermott Elizabeth and Roen Katrina. 2012 Youth on the virtual edge: Researching marginalized sexualities and genders online. Qualitative Health Research 22, 4 (2012), 560–570.

[67]. McDermott Elizabeth, Roen Katrina, and Piela Anna. 2013 Hard-to-reach youth online: Methodological advances in self-harm research. Sexuality Research and Social Policy (2013).

[68]. Meyer Ilan H. 1995 Minority stress and mental health in gay men. Journal of health and social behavior (1995), 38–56. [PubMed: 7738327]

[69]. Meyer Ilan H. 2003 Prejudice, social stress, and mental health in lesbian, gay, and bisexual populations: conceptual issues and research evidence. Psychological bulletin 129, 5 (2003), 674.

[70]. Mikolov Tomas, Sutskever Ilya, Chen Kai, Corrado Greg S, and Dean Jeff. 2013 Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems. 3111–3119.

[71]. Mondal Mainack, Araújo Silva Leandro, and Benevenuto Fabrício. 2017 A Measurement Study of Hate Speech in Social Media. In Proceedings of the 28th ACM Conference on Hypertext and Social Media (HT '17) ACM, 85–94.

[72]. Morioka Tsubasa, Ellison Nicole B., and Brown Michael. 2016 Identity Work on Social Media Sites: Disadvantaged College Students' First Year College Transition In Proc. CSCW

[73]. Morris Jessica F., Waldo Craig R., and Rothblum Esther D.. 2001 A Model of Predictors and Outcomes of Outness Among Lesbian and Bisexual Women. American Journal of Orthopsychiatry (1 2001).

[74]. Mustanski Brian S, Garofalo Robert, and Emerson Erin M. 2010 Mental health disorders, psychological distress, and suicidality in a diverse sample of lesbian, gay, bisexual, and transgender youths. Am. J. Public Health (2010).

[75]. Newman MarkW, Lauterbach Debra, Munson Sean A, Resnick Paul, and Morris Margaret E. 2011 It's not that I don't have problems, I'm just not putting them on Facebook: challenges and opportunities in using online social networks for health. In Proceedings of the ACM 2011 conference on Computer supported cooperative work ACM, 341–350.

[76]. Quynh Nguyen Trang, Bandeen-Roche Karen, German Danielle, Nguyen Nam TT, Bass Judith K, and Knowlton Amy R. 2016 Negative treatment by family as a predictor of depressive symptoms, life satisfaction, suicidality, and tobacco/alcohol use in Vietnamese sexual minority women. LGBT health 3, 5 (2016), 357–365.

[77]. Oakley Abigail. 2016 Disturbing Hegemonic Discourse: Nonbinary Gender and Sexual Orientation Labeling on Tumblr. Social Media + Society 2, 3 (7 2016), 2056305116664217.

[78]. Paisley Varina and Tayar Mark. 2016 Lesbian, gay, bisexual and transgender (LGBT) expatriates: An intersectionality perspective. The International Journal of Human Resource Management (2016).

[79]. Pascoe Cheri J. 2011 Resource and risk: Youth sexuality and new media use. Sex. Res. Soc. Policy (2011).

[80]. Pearson Quinn M. 2003 Breaking the silence in the counselor education classroom: A training seminar on counseling sexual minority clients. Journal of Counseling & Development 81, 3 (2003), 292–300.

[81]. Pennebaker James W.. 1995 Emotion, Disclosure & Health. American Psychological Association.

[82]. Pennebaker James W and Chung Cindy K. 2007 Expressive writing, emotional upheavals, and health. Handbook of health psychology (2007), 263–284.

[83]. Pennebaker James W, Mehl Matthias R, and Niederhoffer Kate G. 2003 Psychological aspects of natural language use: Our words, our selves. Annual review of psychology 54, 1 (2003), 547–577.

[84]. Pennebaker James W and Seagal Janel D. 1999 Forming a story: The health benefits of narrative. Journal of clinical psychology 55, 10 (1999), 1243–1254.

[85]. Pennington Jeffrey, Socher Richard, and Manning Christopher D. 2014 Glove: Global Vectors forWord Representation.In EMNLP, Vol. 14 1532–1543.

[86]. Puckett Julia A., Woodward Eva N., Mereish Ethan H., and Pantalone David W.. 2014 Parental Rejection Following Sexual Orientation Disclosure: Impact on Internalized Homophobia, Social Support, and Mental Health. LGBT Health (2014).

[87]. Punyakanok Vasin, Roth Dan, and Yih Wen-tau. 2008 The importance of syntactic parsing and inference in semantic role labeling. Computational Linguistics 34, 2 (2008), 257–287.

[88]. Rekabsaz Navid, Lupu Mihai, and Hanbury Allan. 2017 Exploration of a threshold for similarity based on uncertainty in word embedding In European Conference on Information Retrieval. Springer, 396–409.

[89]. Remafedi Gary, French Simone, Story Mary, Resnick Michael D, and Blum Robert. 1998 The relationship between suicide risk and sexual orientation: results of a population-based study. American journal of public health (1998).

[90]. Saha Koustuv, Chan Larry, De Barbaro Kaya, Abowd Gregory D, and De Choudhury Munmun. 2017 Inferring Mood Instability on Social Media by Leveraging Ecological Momentary Assessments. Proc. ACM IMWUT 1, 3 (2017), 95.

[91]. Saha Koustuv, Chandrasekharan Eshwar, and De Choudhury Munmun. 2019 Prevalence and Psychological Effects of Hateful Speech in Online College Communities In WebSci.

[92]. Saha Koustuv and De Choudhury Munmun. 2017 Modeling Stress with Social Media Around Incidents of Gun Violence on College Campuses. Proc. ACM Hum.-Comput. Interact. 1, CSCW, Article 92 (12 2017), 27 pages.

[93]. Saha Koustuv, Sugar Benjamin, Torous John, Abrahao Bruno, Kıcıman Emre, and De Choudhury Munmun. 2019 A Social Media Study on The Effects of Psychiatric Medication Use In ICWSM.

[94]. Saha Koustuv, Weber Ingmar, and De Choudhury Munmun. 2018 A Social Media Based Examination of the Effects of Counseling Recommendations After Student Deaths on College Campuses In ICWSM.

[95]. Klaus Scheuerman Morgan, Branham Stacy M., and Hamidi Foad. 2018 Safe Spaces and Safe Places: Unpacking Technology-Mediated Experiences of Safety and Harm with Transgender People. PACM HCI 2, CSCW (2018).

[96]. Schneider Stephen G, Farberow Norman L, and Kruks Gabriel N. 1989 Suicidal behavior in adolescent and young adult gay men. Suicide and Life-Threatening Behavior 19, 4 (1989), 381–394.

[97]. Andrew Schwartz H, Eichstaedt Johannes C, Kern Margaret L, Dziurzynski Lukasz, Ramones Stephanie M, Agrawal Megha, Shah Achal, Kosinski Michal, Stillwell David, Seligman Martin EP, et al. 2013 Personality, gender, and age in the language of social media: The open-vocabulary approach. PloS one 8, 9 (2013), e73791.

[98]. Semaan Bryan, Britton Lauren M., and Dosono Bryan. 2017 Military Masculinity and the Travails of Transitioning: Disclosure in Social Media In Proc. CSCW (CSCW ‘17). ACM, New York, NY, USA.

[99]. Sharma Eva and De Choudhury Munmun. 2018 Mental Health Support and its Relationship to Linguistic Accommodation in Online Communities In Proc. CHI

[100]. Sharma Eva, Saha Koustuv, Kiranmai Ernala Sindhu, Ghoshal Sucheta, and De Choudhury Munmun. 2017 Analyzing ideological discourse on social media: A case study of the abortion debate In Proc. CSS

[101]. Strauss Anselm and Corbin Juliet M.. 1998 Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory SAGE Publications. Google-Books-ID: tBcEjwEACAAJ.

[102]. Suler John. 2004 The online disinhibition effect. Cyberpsychology & behavior 7, 3 (2004), 321–326.

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

[104]. Tebbe Elliot A and Moradi Bonnie. 2016 Suicide risk in trans populations: An application of minority stress theory. (2016).

[105]. Vitak Jessica. 2012 The Impact of Context Collapse and Privacy on Social Network Site Disclosures. Journal of Broadcasting & Electronic Media 56 (2012).

[106]. Walker Jennifer A and Prince Trayci. 2010 Training considerations and suggested counseling interventions for LGBT individuals. Journal of LGBT Issues in Counseling (2010).

[107]. Williamson Iain R.. 2000 Internalized homophobia and health issues affecting lesbians and gay men. Health Education Research (2000).

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

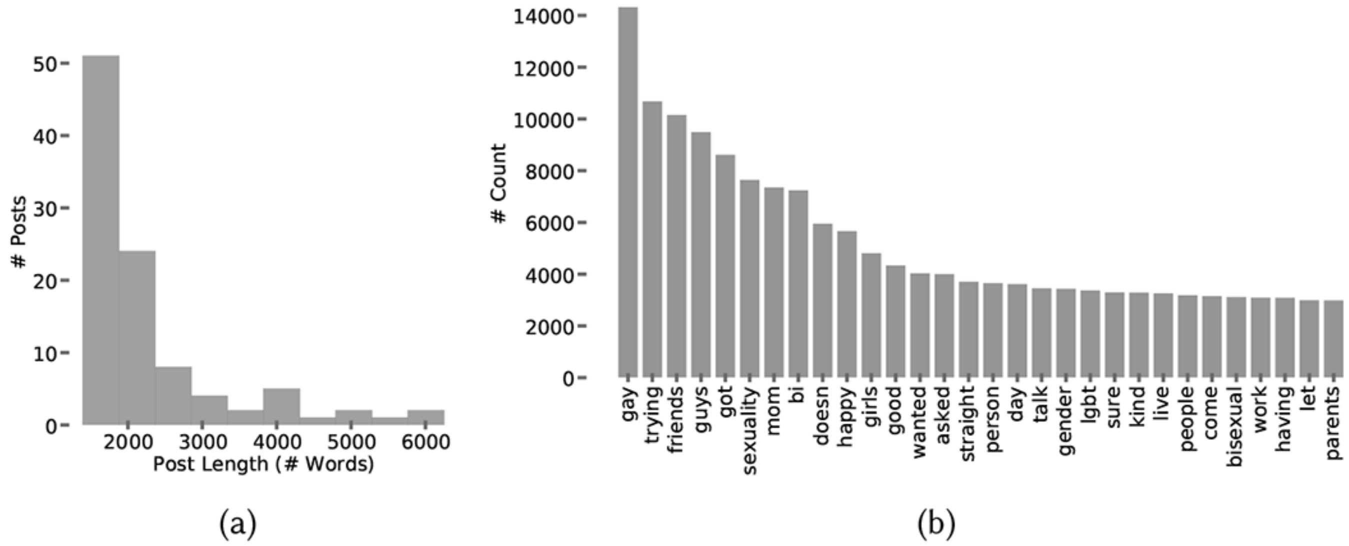


Fig. 1.

Description of data in the subreddit *r/lgbt*: (a) Histogram of the distribution of the length of posts against the number of posts, (b) Top 30 keywords used in posts.

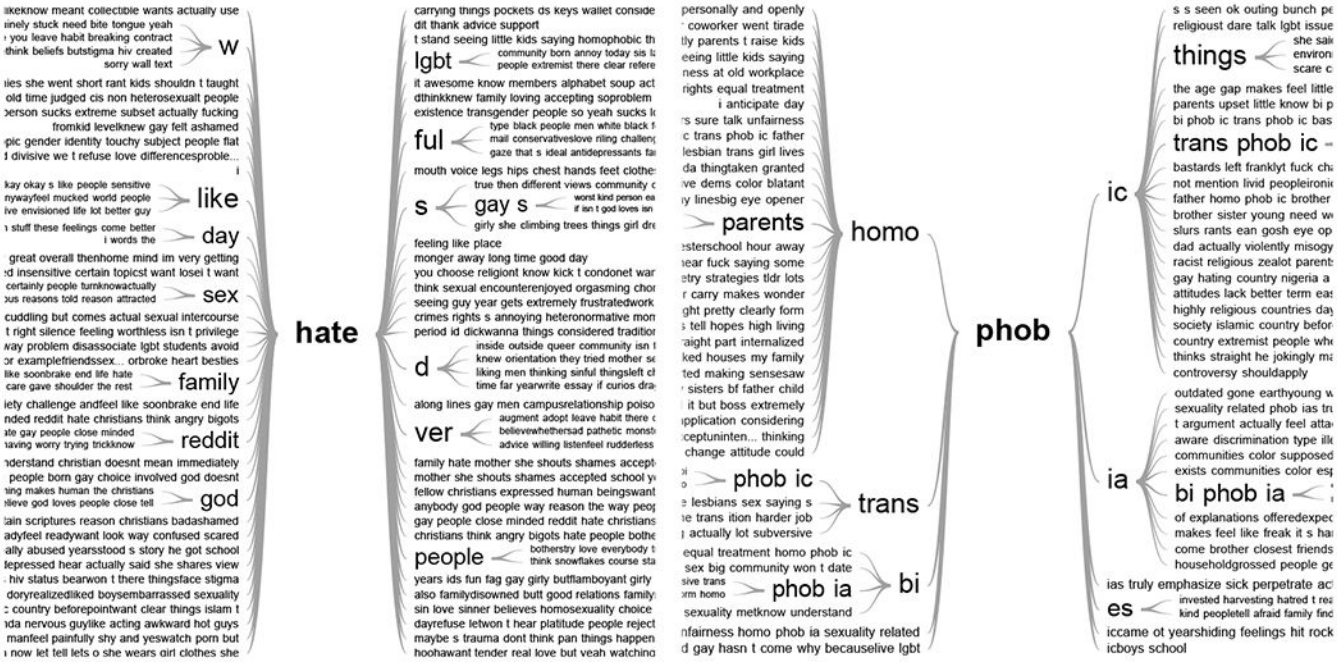
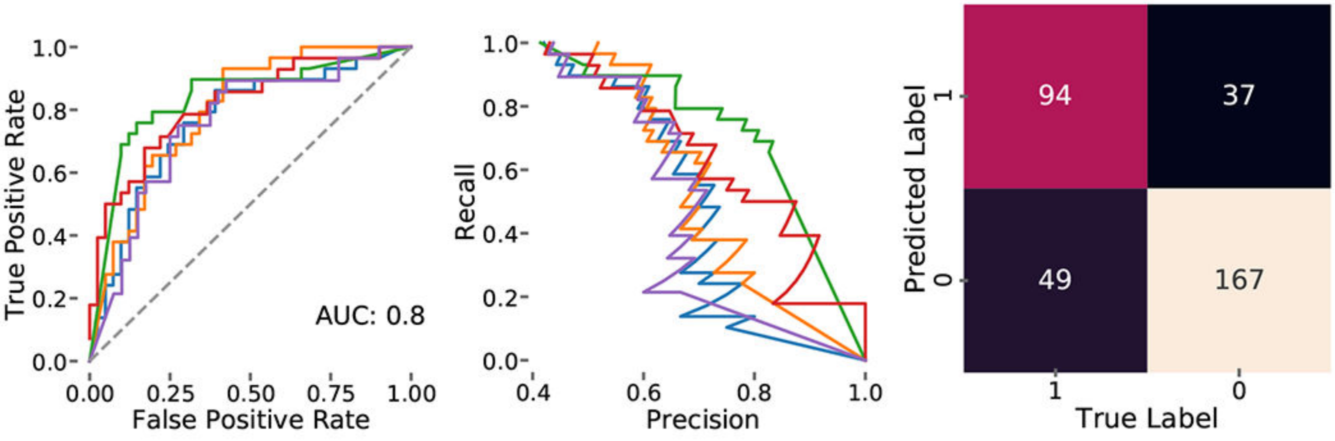


Fig. 2.

Word trees built on two prominent tokens in our expert-coded dataset: “hate” and “phob”. The visualizations represent the content in the form of co-occurrences of keywords in the dataset. The font size of keywords are proportional to their occurrence along with surrounding co-occurring keywords. For example, *phob* occurs as trans(phob)ia and trans(phob)ic, bi(phob)ic, bi(phob)ia, etc.



Predicted Label \ True Label	1	0
1	94	37
0	49	167

(a) (b) (c)

Fig. 3.

Minority Stress Classifier accuracy metrics on κ-fold cross validation: (a) ROC curve, (b) Precision-Recall Curve, (c) Pooled Confusion Matrix

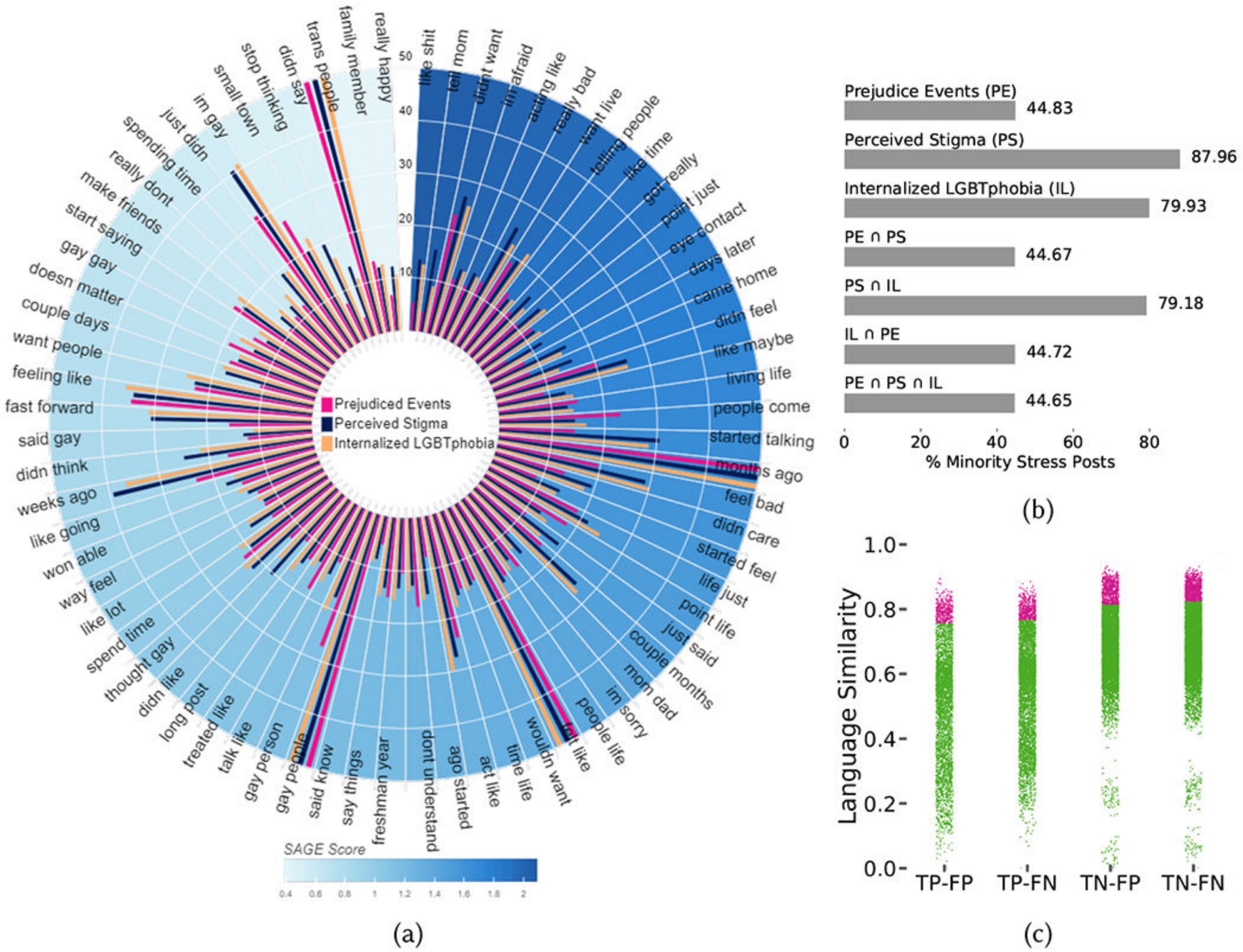


Fig. 4.

(a) Top keywords (per SAGE) with their frequency distribution in the three categories of minority stress, (b) Overlap of Minority Stress categories in among 4,419 posts predicted to contain minority stress, (c) Strip plot of pair-wise similarities of TPs and FPs, TPs and FNs, TNs and FPs, and TNs and FNs.

Table 1.

Minority stress codebook on social media language, drawn on Meyer, 1995.

Prejudice Events

Actions of rejection towards an individual

1. Violence towards and individual
   - Verbal Violence
     - Directing slurs towards minority individuals (before or after disclosure)
     - Rejecting sexuality
   - Physical Violence
2. Legal discrimination in housing, employment, entitlements, and basic civil rights
   - Excluded based on identity
   - Lack of sexuality education in schools
3. Anxiety/stress about discriminatory political powers
4. Includes prejudice events from those in other subsets of the LGBTQ+ population (e.g., violence perpetuated from cisgender LGBQ people to trans people; violence perpetuated from gay people to bisexual people)

Perceived Stigma

Internalized fears and anxiety of expected rejection from others

Fears and anxiety that come from self about how others perceive them:

1. Individuals constantly monitoring their behavior: how one dresses, speaks, walks, and talks in contrast to expected social norms.
2. Lying to cover up identity
3. Thinking others do not or will not “accept” them
4. Feelings of not “fitting in”
5. Anxiety/stress about potential future discriminatory political powers
6. Includes perceived stigma from other subsets of the LGBTQ+ population (e.g., stigma felt by trans people from cisgender LGBQ people; stigma felt by bisexual people from gay people)

Internalized LGBTphobia

Internalization of negative societal attitudes, and negative feelings that come from self about self

1. Self-hate for being a part of the LGBTQ+ community
2. Applying negative attitudes/dislike to themselves related to LGBTQ+ identity
3. Feelings of negativity rooted from LGBTQ+ identity
4. Feelings of isolation or of ending up alone
5. Hate/distaste for others within LGBTQ+ community

Table 2.

Median metrics in *k*-fold (*k*=5) cross-validation.

Model	**Pr.**	**Rc.**	**F1**	**AUC**
Naive Bayes	0.70	0.49	0.53	0.54
Logistic Reg.	0.73	0.72	0.72	0.76
SVM (Linear)	0.74	0.74	0.74	0.77
Random F.	0.76	0.67	0.70	0.75
AdaBoost	0.73	0.72	0.72	0.77
MLP	0.75	0.74	0.75	0.80

Table 3.

Detailed accuracy metrics in *k*-fold (*k*=5) cross-validation in the Minority Stress Classifier (MLP).

Metric	Min.	Max.	Mean.	Stdev.
Precision	0.73	0.82	0.76	0.03
Recall	0.71	0.81	0.75	0.03
F1	0.72	0.82	0.75	0.04
Acc.	0.71	0.81	0.75	0.04
AUC	0.76	0.84	0.80	0.03

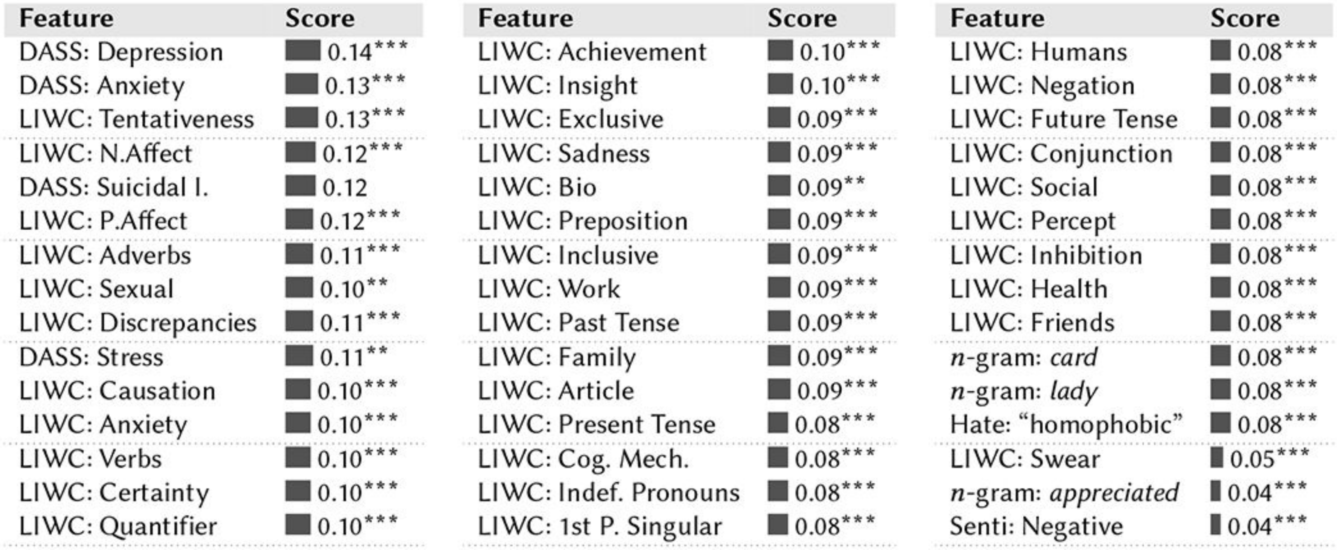
Table 4.

Incremental accuracy metrics of adding features in the Minority Stress Classifier (MLP).

Model	F1	Acc.	AUC
W2V	0.67	0.72	0.72
.+LIWC	0.71	0.71	0.75
.+.+HateLex	0.72	0.71	0.75
.+.+.+Ngrams	0.73	0.72	0.78
.+.+.+.+Senti	0.74	0.74	0.79
.+.+.+.+.+DASS	0.75	0.75	0.80

Table 5.

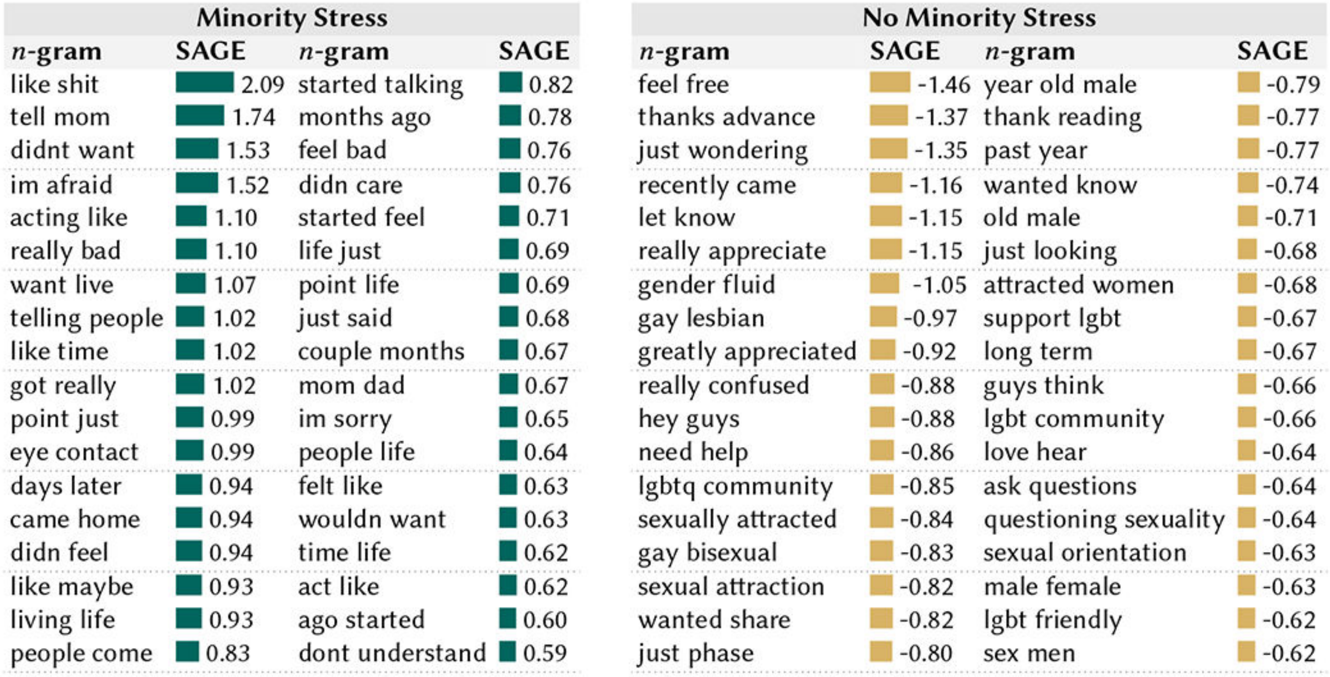
Top 45 Features in the minority stress classifier. *p*-values reported after Bonferroni correction following ANOVA (*** *p*<0.0001, ** *p*<0.001, * *p*<0.01).



Feature	Score	Feature	Score	Feature	Score
DASS: Depression	■ 0.14***	LIWC: Achievement	■ 0.10***	LIWC: Humans	■ 0.08***
DASS: Anxiety	■ 0.13***	LIWC: Insight	■ 0.10***	LIWC: Negation	■ 0.08***
LIWC: Tentativeness	■ 0.13***	LIWC: Exclusive	■ 0.09***	LIWC: Future Tense	■ 0.08***
LIWC: N.Affect	■ 0.12***	LIWC: Sadness	■ 0.09***	LIWC: Conjunction	■ 0.08***
DASS: Suicidal I.	■ 0.12	LIWC: Bio	■ 0.09**	LIWC: Social	■ 0.08***
LIWC: P.Affect	■ 0.12***	LIWC: Preposition	■ 0.09***	LIWC: Percept	■ 0.08***
LIWC: Adverbs	■ 0.11***	LIWC: Inclusive	■ 0.09***	LIWC: Inhibition	■ 0.08***
LIWC: Sexual	■ 0.10**	LIWC: Work	■ 0.09***	LIWC: Health	■ 0.08***
LIWC: Discrepancies	■ 0.11***	LIWC: Past Tense	■ 0.09***	LIWC: Friends	■ 0.08***
DASS: Stress	■ 0.11**	LIWC: Family	■ 0.09***	*n*-gram: *card*	■ 0.08***
LIWC: Causation	■ 0.10***	LIWC: Article	■ 0.09***	*n*-gram: *lady*	■ 0.08***
LIWC: Anxiety	■ 0.10***	LIWC: Present Tense	■ 0.08***	Hate: “homophobic”	■ 0.08***
LIWC: Verbs	■ 0.10***	LIWC: Cog. Mech.	■ 0.08***	LIWC: Swear	▮ 0.05***
LIWC: Certainty	■ 0.10***	LIWC: Indef. Pronouns	■ 0.08***	*n*-gram: *appreciated*	▮ 0.04***
LIWC: Quantifier	■ 0.10***	LIWC: 1st P. Singular	■ 0.08***	Senti: Negative	▮ 0.04***

Table 6.

Top discriminating *n*-grams (*n*=2,3) in posts with and without Minority Stress (SAGE Analysis [37]).



Minority Stress				No Minority Stress			
n-gram	SAGE	*n*-gram	SAGE	*n*-gram	SAGE	*n*-gram	SAGE
like shit	2.09	started talking	0.82	feel free	-1.46	year old male	-0.79
tell mom	1.74	months ago	0.78	thanks advance	-1.37	thank reading	-0.77
didnt want	1.53	feel bad	0.76	just wondering	-1.35	past year	-0.77
im afraid	1.52	didn care	0.76	recently came	-1.16	wanted know	-0.74
acting like	1.10	started feel	0.71	let know	-1.15	old male	-0.71
really bad	1.10	life just	0.69	really appreciate	-1.15	just looking	-0.68
want live	1.07	point life	0.69	gender fluid	-1.05	attracted women	-0.68
telling people	1.02	just said	0.68	gay lesbian	-0.97	support lgbt	-0.67
like time	1.02	couple months	0.67	greatly appreciated	-0.92	long term	-0.67
got really	1.02	mom dad	0.67	really confused	-0.88	guys think	-0.66
point just	0.99	im sorry	0.65	hey guys	-0.88	lgbt community	-0.66
eye contact	0.99	people life	0.64	need help	-0.86	love hear	-0.64
days later	0.94	felt like	0.63	lgbtq community	-0.85	ask questions	-0.64
came home	0.94	wouldn want	0.63	sexually attracted	-0.84	questioning sexuality	-0.64
didn feel	0.94	time life	0.62	gay bisexual	-0.83	sexual orientation	-0.63
like maybe	0.93	act like	0.62	sexual attraction	-0.82	male female	-0.63
living life	0.93	ago started	0.60	wanted share	-0.82	lgbt friendly	-0.62
people come	0.83	dont understand	0.59	just phase	-0.80	sex men	-0.62