Mental Health Discourse on reddit:
Self-disclosure, Social Support, and Anonymity

Munmun De Choudhury
School of Interactive Computing
Georgia Institute of Technology Atlanta GA
munmund@gatech.edu

Sushovan De
Department of Computer Science
Arizona State University Tempe AZ
sushovan@asu.edu

Abstract

Social media is continually emerging as a platform of information exchange around health challenges. We study mental health discourse on the popular social media: reddit. Building on findings about health information seeking and sharing practices in online forums, and social media like Twitter, we address three research challenges. First, we present a characterization of self-disclosure in mental illness communities on reddit. We observe individuals discussing a variety of concerns ranging from the daily grind to specific queries about diagnosis and treatment. Second, we build a statistical model to examine the factors that drive social support on mental health reddit communities. We also develop language models to characterize mental health social support, which are observed to bear emotional, informational, instrumental, and prescriptive information. Finally, we study disinhibition in the light of the dissociative anonymity that reddit's throwaway accounts provide. Apart from promoting open conversations, such anonymity surprisingly is found to gather feedback that is more involving and emotionally engaging. Our findings reveal, for the first time, the kind of unique information needs that a social media like reddit might be fulfilling when it comes to a stigmatic illness. They also expand our understanding of the role of the social web in behavioral therapy.

Introduction

Health information seeking and sharing practices online are known to be effective in helping people cope with respective problems (Fox, 2013). Studies have shown that online fora and support groups provide a conducive environment allowing people to get connected with others who share similar difficulties, misery, pain, condition, or distress and thus act as inexpensive and convenient vehicles for obtaining help and advice around health challenges (Eysenbach et al. 2004). Moreover, honesty and self-disclosure are important therapeutic ingredients of these health communities (Johnson & Ambrose, 2006).

Recently, social media sites have begun to emerge as increasingly adopted platforms wherein health information seeking and sharing practices are apparent. Distinct from online fora, these social systems are more holistic, in the sense that millions of people use them to post about the mundane goings on of their lives. Additionally, unlike many online communities, most social media sites have a personal permanent identity associated with user profiles. Consequently, these platforms provide a rich ecosystem to study the variety of self-disclosure, social support, and disinhibition that engender health related discourse.

Specifically, we examine a highly popular social news and entertainment media: reddit (*http://www.reddit.com/*). reddit is an interesting online social system that has the attributes of a forum: it allows sharing blurbs of text and media as posts that invite votes and commentary. At the same time it is often used as a social feed of information broadcasted from people's contacts and audiences. Another distinctive aspect of reddit is its "throwaway" accounts. These are temporary identities often used as an "anonymity cloak" to discuss uninhibited feelings, sensitive information, or socially unacceptable thoughts momentarily; information otherwise considered unsuitable for the mainstream.

Given the unique characteristics of reddit, in this paper we are interested in studying the nature of discourse around the important health challenge of mental illness. Mental illness, in particular, is a kind of health concern where the value of emotional and pragmatic support as well as self-disclosure has been recognized over the years. Studies have demonstrated that both self-disclosure and social support are beneficial in improving perceived self-efficacy, and helping improve quality of life (Turner et al. 1983). Further, note that mental illness is regarded as a social stigma—there is evidence that people with mental illness tend to be guarded about what they reveal about their condition (Corrigan, 2004; De Choudhury et al. 2014). The dissociative anonymity cover that reddit allows makes study of such a stigmatic illness interesting because of the disinhibition effect induced by anonymity (Caplan & Turner, 2007).

In the light of the above discussion, the following are the research challenges we address in this paper:

- **RQ 1:** What kind of language attributes and content characterize self-disclosure in reddit communities relating to mental health?
- **RQ 2:** What factors drive social support on mental health oriented posts? What are the various forms of social support that characterize commentary on mental health postings?
- **RQ 3:** What kind of online disinhibition do we observe in the mental health information seeking and sharing practices of redditors who choose to be anonymous versus others who do not?

Based on a large dataset of several thousand reddit users, posts, and comments, our findings show that the reddit communities we study allow a high degree of information exchange around a range of issues concerning mental health. These range from self-disclosure of challenges faced in day-to-day activities, work, personal relationships, to specific queries about mental illness diagnosis and treatment. Feedback on mental health postings also ranges a wide spectrum, from emotional and instrumental commentary, to informational and prescriptive advice. In fact, lowered inhibition and self-attention focused posts receive greater support. Finally, looking through the lens of anonymity as enabled by throwaway accounts in reddit, we find that a small but notable fraction of redditors use the feature as a cover for more intimate and open conversations around their experiences of mental illness. In fact, despite the negative or caustic nature of content shared by anonymous redditors, online disinhibition of this nature garners more emotional and instrumental feedback through commentary.

Our observations demonstrate that reddit fills an interesting gap between online health forums and social media and networks like Twitter and Facebook, when it comes to mental health discourse. Moreover, it is established that any psychological consequence depends on the activities a technology enables, attributes of the user, and how the two interact. This research, hence forms a crucial point when trying to determine the effect of the social web on a grave and stigmatic concern like mental illness.

Related Work

Prior research in psychology has examined the important role of social support in combating depression (George et al. 1989). It is argued that social intimacy, social integration, nature of social networks as well as individual perception of being supported by others are important and indispensable in encouraging mental illness recovery (Turner et al. 1983).

The internet is increasingly used for seeking and sharing health information online, and such activity is known to have connections to healthcare utilization and health-related behaviors (Sillence et al. 2007; Liu et al. 2013). Literature on online support groups notes that they are popular sources of information and support for many internet users (White & Dorman, 2001). These forums tend to have sharp contrast when compared with similar offline groups; for instance, people are likely to discuss problems that they do not feel comfortable discussing face-to-face (Johnson & Ambrose,

2006). Moreover such online health communities (OHCs) are known to foster well-being, a sense of control, self-confidence, social interactions, and improved feelings. In this light, approaches to community building have also been proposed (Wicks et al. 2010).

Turning to research on social media, a growing body of work has demonstrated that social media is an increasingly adopted platform allowing users to communicate around a variety of health concerns (Paul & Dredze, 2011; Fox, 2013; De Choudhury et al. 2014). Newman et al. (2011) interviewed people with significant health concerns who participated in both OHCs and Facebook. Oh et al. (2013) examined people's use of Facebook for health purposes and showed that emotional support was a significant predictor of health self-efficacy.

In the context of mental health in particular, Moreno et al., (2011) demonstrated that status updates on Facebook could reveal symptoms of major depressive episodes, while Park et al. (2013) found differences in the perception of Twitter use between depressed and non-depressed users—the former found value in Twitter due to the ability to garner social awareness and engage in emotional interaction. On similar lines, De Choudhury et al. (2013) examined Twitter postings of individuals suffering from depression to build models that predict the future occurrence of depression.

Weaving the above threads together, health information seeking and sharing on OHCs have been studied extensively in the light of the benefits of self-help, social support and empathy, however little investigation about the same has been done on social media. Today, many social media platforms are becoming a routine go-to place for broadcasting and discussing topics of daily life. It is therefore useful to examine how this kind of habitual disinhibition on social media impacts the manner in which people derive help on their health concerns, especially on ones like mental illness, known to be socially stigmatic (for definition see Corrigan (2004)). Focusing on reddit, note that we use a broad definition of "mental health" to span shared content around any aspect of the ailment experience.

Finally, while prior work has shown the potential of social media in bearing markers of mental health concerns, our knowledge about the nature of self-disclosure on health
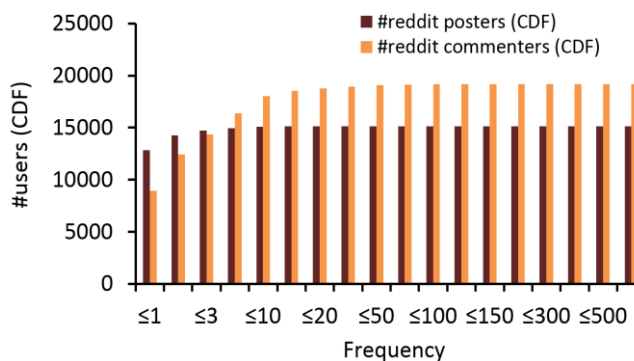
Figure 1. Cumulative density distribution (CDF) of number of redditors authoring posts and authoring comments. X axis—number of posts/comments, Y axis—number of users.

concerns is rather limited, so is the nature of emotional support and advice that social media communities provide. Through the lens of reddit, hence we are interested in investigating the language and content of people's self-disclosure, what characterizes the nature of community feedback, and the attributes that drive such social support.

Data and Methods

reddit is a social news website where registered users submit content in the form of links or text posts. Users, also known as "redditors", can then vote each submission "up" or "down" to rank the post and determine its position or prominence on the site's pages. These two attributes associated with a post are referred to as "upvotes" and "downvotes". Redditors can also comment on posts, and respond back in a conversation tree of comments. Content entries, that is the posts, are organized by areas of interest or sub-communities called "subreddits", such as politics, programming, or science.

As of 2013, reddit's official statistics included 56 billion page views, 731 million unique visitors, 40,855,032 posts, and 404,603,286 comments (*http://blog.reddit.com/2013/12/top-posts-of-2013-stats-and-snoo-years.html*).

We used of reddit's official API (*http://www.reddit.com/dev /api*) to collect posts, comments and associated metadata from several mental health subreddits: specifically using a Python wrapper PRAW (*https://praw.readthedocs.org/en/latest/index.html)*. The subreddits we crawled were: *alcoholism, anxiety, bipolarreddit, depression, mentalhealth, MMFB* (Make Me Feel Better)*, socialanxiety, SuicideWatch*. All of these subreddits host public content.

In order to arrive at a comprehensive list of subreddits to focus on, we utilized reddit's native subreddit search feature (*http://www.reddit.com/reddits*) and searched for subreddits on "mental health". Two researchers familiar with reddit employed an initial filtering step on the search results returned, so that we focus on high precision subreddits discussing mental health concerns and issues. Thereafter, we focused on a snowball approach in which starting with a
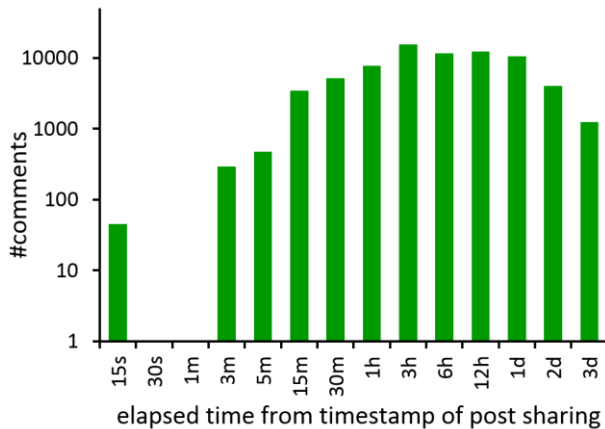
few seed subreddits (*mentalhealth, depression*), we compiled a second list of "related" or "similar" subreddits that are listed in the profile pages of the seed subreddits. Following a second filtering step, we arrived at the list of subreddits listed above.

For each of these subreddits, we obtained daily crawls of their posts in the *New* category. Corresponding to each post we collected information on the title of the post, the body or textual content, id, timestamp when the post was made, author id, and the number of upvotes and downvotes it obtained. Since posts gather comments over a period of time following the time of sharing, we crawled all of the comments per post that were shared over a three day period after the post was made. Qualitative examinations of the subreddits of interest revealed that 90% or more of the comments to any post were typically made in a three day window following the time the post is made—hence the choice. The crawl of the subreddits used in this paper spanned between Nov 8, 2013 and Dec 28, 2013.

Posts per day	247.1 (±41.3)
Average posts per user	0.75 (±2.21)
Mean, Median length of posts (words)	96.4, 64
Average upvotes per post	8.37 (±23.2)
Average downvotes per post	1.26 (±3.71)
Average comments per post	5.03 (±7.52)
Mean, Median comment length (words)	27.03, 16
Average comments per user	2.69 (±8.49)
Average upvotes per comment	2.39 (±4.16)
Average downvotes per comment	2.06 (±3.58)
Total number of posts	72,236

Table 1. Basic descriptive statistics of the reddit dataset.

We present some descriptive statistics of our crawled data. Our dataset contained 20,411 posts with at least one comment, and 97,661 comments in all, with 27,102 unique users who made posts, comments or both. A set of 7,823 users (28.79%) were found to write both at least one post and comment. CDF of the user distribution over posts and comments is given in Figure 1. The figure shows the expected heavy tail trend observed in several social phenomena. Also see Figure 2 for the distribution of comments over time following post share. It illustrates the quick responsivity culture in the communities we study (peak at 3 hours). Some of the additional statistics of our dataset are given in Table 1. Further, example titles of a few posts in our dataset are given in Table 2.

Figure 2. Comment distribution over time. Each bin shows the number of comments shared within a certain time interval from the timestamp of post.

Scammed out of $2600, so depressed
Just sharing... and maybe some advice will come.
How did you feel when your antidepressant started working?
Inability to concentrate destroying my grades
Has anyone else gained weight due to depression?
I can't do this anymore negative thoughts have taken over my mind these past few months
Am I depressed, I don't think I am but I keep sabotaging everything I do.
i'm afraid i might have depression that won't go away by itself and that it needs to be treated professionally...
loneliness, therapy, and just a sharing of weird feelings

Table 2. Titles of a few mental health oriented reddit posts.

RQ 1 Results: Self-disclosure

General Linguistic Attributes

To understand the nature of self-disclosure in reddit posts, we first examine the general linguistic attributes manifested in their content. In Table 3, we first present a list of the most popular (stopword eliminated) unigrams that appear in reddit postings. We intended to look at these highly shared unigrams more deeply and systematically, hence we organized these unigrams (stopword inclusive) in various semantic categories provided by the psycholinguistic lexicon LIWC (*http://www.liwc.net/*).

We find that among the unigrams in Table 3, there are words that extensively span **emotional or affective expressions** (*happy, love, bad, anxiety, good, hate*) e.g.:

"*I've been recently wondering if my **love** to numb the world around me has turned me into an alcoholic...*"
"*Has anyone else battled numbness, **loss** of feeling during recovery? Does it ever get **better**? i've been sober for about 2 years and still have **pretty severe anxiety** at times*".

We observe presence of **relationships and social life** words too (*family, friends, people, person, parents*) e.g.:

"*i get really anxious out when i go home for big events.*"
"*i do love my **family**, they're just really loud and argumentative sometimes*".

Unigram	Count	Unigram	Count	Unigram	Count
feel	24381	talk	6574	months	4809
time	17008	feeling	6013	deny	4725
life	16671	family	5949	felt	4598
people	15026	job	5783	home	4552
friends	11233	started	5341	find	4510
things	10128	long	5256	ago	4320
day	9429	love	5192	parents	4295
years	9261	bad	5169	end	4255
depression	9032	lot	5092	safe	4215
back	8979	told	5079	live	4198
school	8305	thought	5069	hate	4124
anxiety	8080	avoid	5042	hard	4104
work	7851	person	4913	night	3813
make	7599	anymore	4908	care	3727
good	7315	happy	4853	pretty	3711

Table 3. Most frequent 45 unigrams in reddit posts, along with their frequencies of usage.

Temporal indicators in reddit discourse is also visible— e.g., *time, day, years, months*:

"*hi all, i'm ten **weeks** sober **today** and while i wish i could say i'm physically and mentally in great shape, the truth is i judge my **days** by how i feel less bad, as oppose to good*".

Work and daily grind oriented words are common as well, because lifestyle irregularities are often associated with the psychopathology of mental illness (Prigerson et al. 1995)—e.g., *life, school, work, job*:

"*I am completely broke, can't afford rehab, and can't take time off **work***".

We also find a fair number of **cognitive words** in these highly used unigrams (*felt, hard, feeling, lot*), e.g.:

"*I'm new here, but having anxiety like I haven't **felt** in a long time*"
"*I find a **lot** of strength in going to a concert.I have been **understanding** of my anxiety and depression since i was about 8, and i hated it*".

These observations are supported by psychology literature, where cognitive biases as manifested through dysfunctional attitudes, depressive attributional biases, and negative automatic thoughts were found to be characteristic of mental illness (Eaves & Rush, 1984).

Further, **inhibition words** like *avoid, deny, safe* demonstrate that redditors are perhaps using the platform to broadcast their thoughts to an audience of strangers or weak ties, on issues and topics they might consider to be socially stigmatic to be discussed elsewhere:

"*i can't **escape** the feeling of fright i have at all times. i don't feel **safe** in my own home*"
"*ive been **denying** my (assumed) depression symptoms for close to two years now writing them off ...*"

Comparing across different LIWC semantic categories over all posts, we observe noticeable differences— Kruskal-Wallis one-way analysis of variance indicated the differences across categories to be significant (χ^2(39; *N*=20411)=9.24; *p*<10^{-4}). Table 4 reports the top 8 most common LIWC categories, the mean proportion of words from each category in the posts, and the corresponding standard deviation. Note that the percentages over all categories sum to greater than 100%, since a word could belong to multiple categories.

Observing closely, many of the categories whose corresponding unigrams appeared in Table 3, are also highly prominent categories globally over all posts, as given in Table 4. Not shown in Table 4, perhaps intuitively, *negative emotion, anger,* and *sadness* words were considerably more prominent than *positive emotion* words (a Wilcoxon signed rank test reveals that the differences are statistically significant (*z*=-6.08, *p*<.001)). Likely, these redditors experience several negative emotions: hence mental instability, helplessness, loneliness, restlessness manifest in their postings (Rude et al. 2004).

Semantic category	% category words	Std. dev.
Cognitive	19%	0.0579
2nd, 3rd person pronoun	17%	0.0554
Inhibition	13%	0.0526
Present	11%	0.0493
1st person singular prn.	9%	0.0423
Affect	8%	0.0474
Social	8%	0.0539
Time	7%	0.0364

Table 4. Top 8 most frequent semantic categories from LIWC, as observed in reddit posts.

Non-first person personal pronouns were the second most prominent category, implying the tendency of the authors of these posts to be interactive with their audiences perhaps to seek social support and advice. We also find the 1st person singular pronouns to be extremely common: perhaps indicating that many of the posts that redditors share on mental health are about themselves, their experiences, and

the health and social issues they are facing. In fact, high self-attentional focus is a known psychological attribute of mental illness sufferers (Chung & Pennebaker, 2007).

Mental Health Information Sharing

Beyond the linguistic attributes that indicate self-disclosure, we intended to examine whether redditors sought and shared information specific to their mental health concerns, e.g., symptoms or treatment related information in their posts.

For this purpose, we utilized the lexicon of "depression" terms from (De Choudhury et al., 2013). The lexicon consists of words derived from the "Mental Health" category of Yahoo! Answers. However, since we were primarily interested in keywords in this lexicon (comprising 1000 words) that would indicate symptoms or treatment aspects of mental illness, we obtained crowdsourced labels on each word on whether it is a symptom of a mental illness, a treatment related information, or otherwise. Particularly, per word, we obtained five independent judgments from crowdworkers on Amazon's Mechanical Turk. We considered English proficient crowdworkers with more than 95% approval rating. To better guide their judgment, they were instructed to refer to the following online resources: RightDiagnosis.com (*http://www.rightdiagnosis.com/m/ mental_illness/symptoms.htm)*, WebMD (*http://www. webmd.com/depression/mental-health-warning-signs),* and for treatment information, the list of antidepressant drug names from Wikipedia (*http://en.wikipedia.org/wiki/List _of_antidepressants)*.

The labeling task coded a set of 467 words which were either symptoms or treatment specific (interrater agreement Fleiss' κ=.68)—a word was considered to be a treatment or a symptom if three or more raters agreed. One researcher went through the coded words as a final step of sanity check.

Symptoms	Treatment
anxiety, mental, withdrawal, severe, insomnia, confused, fear, worried, angry, suicide, alcohol, panic, denial, swings, weight	therapy, side-effects, medication, diagnosis, 20mg, drugs, antidepressant, doctor, pills, escitalopram, psychiatrist, xanax, recovered

Table 5. Mental illness related symptom and treatment terms that are highly frequent in reddit posts.

In Table 5, we present a few common symptom keywords and common treatment keywords found in post content. We observe that symptoms related posts revolve around details about sleep, eating habits, and other forms of physical ailment—all of which are known to be associated with occurrence of a depressive episode (Posternak et al., 2006).

"*Heart/rib/chest **pains**, worse than usual… hi everyone, i have been suffering mild **panic** attacks…*"
"***Insomnia** and binge **drinking** i've had periodic insomnia for a lot of my life. lately, it seems to have gotten worse*".

Analyzing further, example post excerpts around treatment include:

"*Cognitive behavioural **therapy**? Does it work? Have you tried it? the university offers it for free where i live…*"
"*How I know the **medicine** is working. bipolar ii, currently on 1500 mg lithium, 400 mg lamictal, .5 mg risperdal, and as of one month ago, 150 mg wellbutrin.*"

Thus users also appear to discuss their therapy and treatment, even dosage levels of medication, as well as seek advice on particular treatment options.

Summarily, RQ 1 investigated disclosure in mental health subreddits: we found that the postings ranged from sharing the challenges faced in day-to-day activities, to work and personal relationships, as well as seeking detailed information on symptoms and treatment of mental illness.

RQ 2 Results: Social Support

Factors Driving Social Support

In our second research question, we first examine whether attributes of the content of posts, are predictive of the extent of social support posts receive.

Independent variables			
1st person pronoun	Conjunction	Motion	Sexual
2nd p. pronoun	Death	Negation	Social
3rd p. pronoun	Discrepancy	Neg. emotion	Space
Achievement	Exclusion	Numbers	Swear
Adverbs	Health	Perception	Tense
Assent	Home	Pos. emotion	Tentative
Bio	Inclusion	Preposition	Time
Body	Ingestion	Quantitative	Work
Cause	Inhibition	Relationships	
Certainty	Leisure	Relativity	
Cognitive	Money	Religion	
Post length	#Emoticons	#Question-centric words	

Table 6. List of semantic categories of words from LIWC and three additional post attributes that were used as independent variables in prediction task. Total variables=44.

Response variables. Since there are two different ways in which redditors can provide feedback on a post, i.e., voting and commenting, we define the following two measures of social support:

1. *"Karma"*. One of the dependent variables of social support we attempt to predict, measures the "karma" i.e., net votes that a post receives. Since every reddit post receives a certain number of upvotes (positive feedback) and downvotes (negative feedback), we define the measure of karma to be the difference between the number of upvotes and downvotes. Note that "karma" is a reddit defined measure (*http://www.reddit.com/wiki/ faq#wiki_what_is_that_number_next_to_usernames.3Fa nd_what_is_karma.3F*), and the authors do not intend to convey any philosophical or suggestive meaning.
2. *Comments*. The second dependent variable of social support is the number of comments on a reddit post.

Predictive variables. In Table 6 we present the independent variables which are the various semantic categories of words (from LIWC) computed on post content (ref. previous

section). We additionally use three different independent variables for this prediction task: the length of a post (number of whitespace delimited words), number of emoticons (*http://en.wikipedia.org/wiki/List_of_emoticons*) in post content, and the number of question-centric words (*what, where, when, which, who, whose, why, how*). The rationale to get emoticons is to observe the degree of emotional expressivity, while for the question words it is to understand to what extent posts on mental health seek explicit feedback or suggestions from the reddit community.

Statistical technique. We use negative binomial regression as our prediction method because both our dependent variables, karma and comments are counts, and negative binomial regression is typically well-suited to handle over-dispersed count outcome variables. We use a measure called deviance to evaluate goodness of fit, since this model has no direct analog of the proportion of variance explained by the predictors (R^2) in linear regression. Deviance is a measure of the lack of fit to the data in a negative binomial regression model—lower numbers are better. It is calculated by comparing a model with the saturated model—a model with a theoretically perfect fit (the intercept only model here).

Predictors of Karma	β		Std. Err.	z
Intercept	20.22	***	0.052	150.79
1st person pronoun	11.21	***	0.047	86.48
Negative emotion (NA)	-10.8	***	0.016	-65.70
Relationships	8.62	***	0.011	61.96
Cognitive	8.36	**	0.054	59.49
Emoticons	7.30	***	0.051	58.69
Positive emotion (PA)	6.42	**	0.067	67.26
Social	5.85	**	0.054	43.68
Inhibition	-5.58	**	0.051	-43.49
Health	4.30	***	0.068	41.89
Certainty	1.35	*	0.034	23.83
Home	1.34	*	0.065	27.85
Tentative	-1.26	**	0.060	-19.52
Post length	-1.12	***	0.028	-10.68
Swear	-0.87	**	0.041	-18.24
Religion	0.79	*	0.024	17.09

*** $p<.0001$; ** $p<.001$; * $p<.01$

Null deviance	328.351
Residual deviance	69.246
Log likelihood	-52.489
Error degrees of freedom	20,367

Table 7. Results of predicting the number of karma on a post using the independent variables from Table 6. The prediction model is a negative binomial regression model. The top 15 predictive variables (in terms of their β coefficients are shown).

Prediction results. Now we analyze the results of our two prediction tasks—predicting the measure of karma of posts, and the number of comments. Table 7 presents the results of negative binomial regression in predicting karma. We observe that compared to a baseline null model (intercept only model), the LIWC semantic categories as well as post attributes (ref. Table 6) provide considerable explanatory power, with an improvement in deviance (328.351-69.246=259.105). This difference in deviances follows a χ^2

distribution, hence using a χ^2 test we show significant statistical power of our model over baseline in explaining our data: $\chi^2(44; N=20411)=328.35-69.24=259.11; p<10^{-5}$.

Next, Table 8 presents the results of another negative binomial regression, with comment count of posts as the dependent variable. The overall model explains a considerable amount of deviance compared to the null model, $\chi^2(44; N=20411)= 294.896-78.252=216.644; p<10^{-5}$.

The contribution of the different independent variables in the two prediction tasks is notable. In both, greater use of *1st person pronouns* in posts garners more karma and comments. We conjecture that personal accounts of mental health concerns engages more redditors. Interestingly, while both *negative* and *positive emotion* are significant predictors of karma and comment count, greater NA hinders karma but drives more comments. At the same time, greater PA leads to more karma though with fewer comments.

Next, posts with lowered sense of *inhibition* receive greater karma and comments, likely because the post authors are discussing their mental health concerns in a candid and unrestrained way. Posts about *relationships*, *social* aspects, and *health* also receive greater social support in the form of karma and comments. Shorter posts and with lesser *swear* content receive more karma, however posts which are longer or more elaborate in nature garner greater commentary. Finally, presence of more *emoticons* in mental health posts is associated with higher karma, while posts that are more *question-centric* receive more comments, likely because the latter seek explicit feedback and advice from the greater reddit audience.

Predictors of Comments	β		Std. Err.	z
Intercept	17.35	***	0.014	128.41
Health	11.7	***	0.110	106.36
Inhibition	-8.63	***	0.083	-103.9
1st person pronoun	5.36	***	0.105	74.71
Post length	3.46	***	0.037	53.51
Body	3.35	***	0.044	61.14
2nd person pronoun	3.29	**	0.019	48.21
Relationships	2.22	***	0.026	45.38
Question-centric words	2.06	**	0.032	44.38
Death	1.99	**	0.039	41.03
Cognitive	1.59	***	0.066	44.09
Social	1.45	**	0.038	28.16
Time	1.34	**	0.021	33.81
3rd person pronoun	-1.06	**	0.059	-17.97
Negative emotion (NA)	0.63	**	0.141	13.4
Positive emotion (PA)	-0.32	*	0.044	-9.71

*** $p<.0001$; ** $p<.001$; * $p<.01$

Null deviance	294.896
Residual deviance	78.252
Log likelihood	-63.856
Error degrees of freedom	20,367

Table 8. Results of predicting the number of comments on a post using the independent variables from Table 6. The prediction model is a negative binomial regression model. Top 15 predictive variables (in terms of β coefficients are shown).

Characterizing Social Support

Our next question of investigation is, what is the *type* of social support that reddit users provide on postings in mental health subreddits? Our observation stems from prior health literature, where social support concerning illnesses has been characterized (Turner et al. 1983; George et al. 1989). From the literature, we identified a set of four categories of social support around health concerns— *prescriptive, informational, instrumental,* and *emotional*.

Identifying types of social support in reddit commentary involved characterizing the content of comments into thematic clusters. Therefore we built a language model, that makes use of the unigrams and bigrams present in comment content to automatically infer clusters of support types. Note that unsupervised learning is more appropriate here because of the lack of ground truth. Our model is based on Latent Dirichlet Allocation (LDA) (Blei et al. 2003), a widely popular probabilistic topic model that yields clusters of word tokens (unigrams and bigrams here). For the purposes of this paper, we made use of the LDA implementation provided by Stanford Topic Modeling Toolbox (*http://nlp.stanford.edu/software/tmt/ tmt-0.2/*). The model yielded a clustering of the comments into 20 different topics.

Over the LDA output, two researchers performed qualitative labeling of the clusters, so as to identify *those* clusters whose comment content revolved around social support and advice, and if so, what was the type of support that the community was providing in them. That is, we intended to identify the clusters in which the comments bore *prescriptive, informational, instrumental,* and *emotional* information. Following the labeling task to identify which four clusters reflected the four types of social support above, the Fleiss' κ for interrater agreement was found to be .77, and disagreements were resolved through discussion.

We observe that there is considerable variation in the presence of each type of social support: there are more emotional (36%) and prescriptive comments (32%), than those which are informational (13%) or instrumental (19%). A Kruskal-Wallis one-way analysis of variance indicated the differences to be significant (χ^2(3; N=97661)=5.88; $p<10^{-3}$). To demonstrate differences across the four social support themes, we show context of use of one popular unigram/bigram from each category in Table 9.

In short, in characterizing social support in RQ 2, we observed that lowered inhibition postings that are more self-attention focused and discuss relationship and health issues seemed to gather greater community support through votes and comments. Somewhat surprisingly greater negative affect was associated with more commentary. Not all social support were of the same type either—our findings demonstrated comments to bear emotional and instrumental, to information and prescriptive advice.

| **Emotional** |
| *Good luck*! I'm afraid I don't have any advice |
| *Good luck* please talk honestly to your doctor |
| *Good luck*, this shit is tough |
| *Good luck* in the same boat my friend. |
| **Informational** |
| *Therapy*, including cognitive behavioral therapy is very helpful |
| … *therapy* to deal with your anxiety in the long term |
| … *therapy* [..] they just meet in groups to support each other |
| … *therapy* [..] with a therapist you are able to get very customized understanding |
| **Instrumental** |
| *Feel free* to pm me if you ever need |
| *Feel free* to look up what you've been prescribed |
| … *feel free*. I'm here |
| *Feel free* to pm or comment if you have questions. |
| **Prescriptive** |
| *Medication* prevents seizures, lessens withdrawal symptoms… |
| *Medication* [..] I got prescribed a benzodiazepine to calm me… |
| … *medication* asap so that you can see what it does to you |
| … *medication*, drug and alcohol counseling [..] you should be able to do that without judgment |

Table 9. Excerpts from comments along the four types of social support. Excerpts contain a widely mentioned unigram/bigram per type (shown in italics).

RQ 3 Results: Anonymity and Disinhibition

Our third and final research question revolves around examining how the ability to be anonymous on reddit impacts sharing of information around stigmatic health concerns like mental illness. An interesting feature of reddit, is the 'throwaway account'. By way of this feature, reddit enables an individual create accounts in a matter of minutes without giving out an email address. They can thus make the choice to dissociate from their reddit identity by simply using an alternate pseudonym and then leaving it behind. Prior literature demonstrates that researchers recognize dissociative anonymity of this nature (a resistance to attach to offline identity or to their actual reddit account/online persona) as the foundation of online disinhibition (Lapidot-Lefler & Barak 2012). Online disinhibition leads people to act differently than they would in identifiable online settings (Chester & O'Hara, 2007). This led us to examine the behavior of throwaway accounts in our data.

First, based on our qualitative observations, we used a simple technique to identify such anonymous throwaway accounts. We matched the regex "**throw**" to all usernames to construct a high precision set of anonymous posters. Note that, there could be users our technique cannot detect, e.g., ones who may not have this regex in their temporary identities; however our technique gives us a reliable and precise user set to work with.

Category	Non-anon.	Anon.	Diff.	z	p
Inhibition	0.0165	0.0130	-21.2%	1.712	$<10^{-3}$
Friends	0.0043	0.0050	16.4%	2.476	$<10^{-5}$
Family	0.0046	0.0052	14.9%	2.181	$<10^{-5}$
Death	0.0042	0.0048	14.8%	1.712	$<10^{-4}$
Health	0.0040	0.0044	10.6%	2.476	$<10^{-6}$
1st person prn.	0.0030	0.0033	6.8%	2.181	$<10^{-6}$
2nd person prn.	0.0189	0.0182	-4.1%	1.712	$<10^{-4}$
Cognitive	0.1939	0.2009	3.6%	2.476	$<10^{-6}$
Negative emo.	0.0424	0.0431	1.8%	2.181	$<10^{-3}$
Anger	0.0109	0.0116	6.7%	1.712	$<10^{-3}$
Anxiety	0.0088	0.0107	17.6%	2.476	$<10^{-6}$
Sadness	0.0127	0.0130	1.8%	2.181	$<10^{-5}$

Positive emo.		0.0364	0.0360	-1.2%	1.712	$<10^{-3}$

Table 10. Differences in mean use of linguistic attributes between the anonymous (the throwaway accounts) group and the rest. The statistical significance tests are based on Mann-Whitney U-tests. We adopt Bonferroni correction to correct for multiple pairwise comparisons (α=.001).

To our surprise, mental illness despite being stigmatic, our statistics reveal that a rather small percentage of users in our dataset used throwaway accounts (1,209 users; 4.46%). Nevertheless, we do notice that in this small percentage of throwaway account owners, a greater fraction (92% or 1114 users) are authors of one or more posts, while about 42% (508 users) are ones who have authored one or more comments. This shows that *sharing* seemingly personal information on mental health issues through posts is considered relatively more of a sensitive activity than *providing* support through commentary. Further, 733 users (61%) we found to post *exactly one item* (i.e., post or comment). Such "one-time" usage of throwaway accounts by a large majority shows that redditors use it as a mechanism to discuss on topics they feel guarded about, and a one-time use gives them the opportunity to not leave any trails behind, and walk away from further discussion.

Differences in Language Use

Is there any observable difference in behavior, e.g., in the light of the linguistic attributes (from LIWC) between the set of redditors who post with throwaway accounts and those who do not? Table 10 presents the mean use of top 10 LIWC semantic categories along which the two cohorts differ the most—note that for negative emotion (NA) we show a breakdown of three types of emotion, *anger, anxiety*, and *sadness*. We also present statistical hypothesis testing between the cohorts using Wilcoxon signed rank tests.

We observe notable differences—anonymous redditors tend to be more disinhibited in the content they share in their postings (21.2% less *inhibition* words); share more about personal relationships e.g., *friends, family* (14-16% more); converse more on *death* and *health* related issues (10-14% more); are more self-attention focused i.e., greater use of *1st person pronouns* (6.8% more); are less interactive with others i.e., lower use of *2nd person pronouns* (4.1% less); present more *cognitive* biases in their content (3.6% more *cognitive* words); and finally, are more negative and less positive (1.8% higher and 1.2% lower respectively)—also note greater use of *anger, anxiety, sadness* words. To summarize, these observations tell us that anonymity through these throwaway reddit accounts results in decreased feelings of vulnerability and increased self-disclosure when it comes to discourse on mental health.

### Differences in Social Support

Does the nature of feedback or social support from the greater reddit community also differ in the case of posts from anonymous accounts? We observe from Figure 3 that throwaway accounts are endorsed to a lesser extent (lower karma), perhaps due to their negative or caustic content.
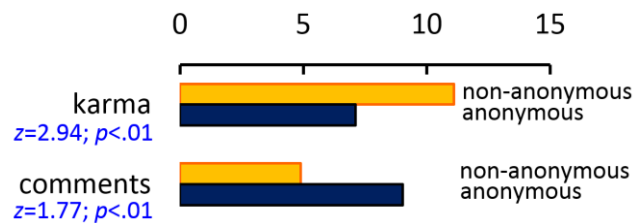
Figure 3. Comparison of extent of social support through karma and comments between throwaway and other accounts. Significance results use Mann-Whitney U-tests.

However, they garner more comments, likely due to their greater disinhibition in expressing mental health issues.

Additionally, from Figure 4 we find that throwaway accounts receive 27% and 12% more *emotional* and
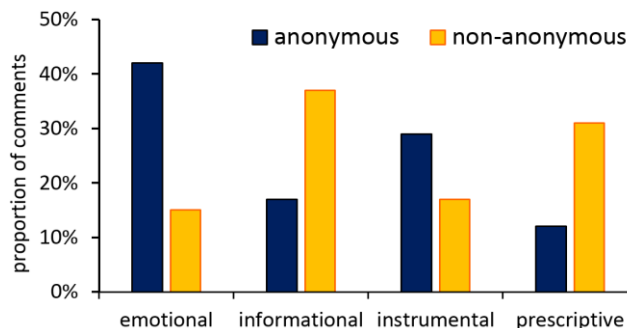
Figure 4. Posts from anonymous and non-anonymous accounts in the light of social support types.

instrumental support through comments, in comparison to non-anonymous ones. Perhaps the reddit audience tends to sympathize more with the posters, and provide more helpful and contributory feedback, suggestions, and opinions to the throwaway posters because of their honest confessions. However, throwaway accounts garner fewer comments that are *informational* or *prescriptive* in nature. Perhaps when individuals share information through their primary (non-anonymous) accounts, they are more objective in nature and seek information and help on concrete challenges around mental illness. Consequently, the audience responds through support that is laden with information about e.g., treatment, coping etc. as well as that provide propositions around ways to deal with the challenges.

In essence, our findings of RQ 3 show that anonymity by way of throwaway accounts allows greater self-disclosure around the stigmatic topic of mental health without worrying about being identified; at the same time they let other redditors to communicate prudently, or give honest advice on sensitive issues around mental health.

Discussion

Theoretical Implications

Our findings show that reddit fills an interesting gap between online health forums, and social media and social networks like Twitter and Facebook, when it comes to mental health related discourse. We find that reddit users in

these communities share, quite explicitly, information about mental health issues providing evidence of selecting this non-conventional tool as a medium that fulfils certain needs.

Self-disclosure. The clinical literature reports that a variety of factors or states are associated with mental health concerns in people; these include mood disturbances, self-derogatory thoughts, cognitive impairments, attention, communication, and judgment (Rabkin & Struening, 1976). Our observations from self-disclosure in reddit postings align with these prior findings—mood disturbances indicated by higher use of negative emotion words, self-derogatory thoughts and self-care indicated by *work, time*, and *relativity* words, cognitive impairments and judgment issues indicated by the use of *cognitive* words, attention to self as indicated by greater use of 1st person singular pronoun, and finally communication attributes reflected in the use of *social* words and increased non-first person personal pronoun use. In essence, our results demonstrate that redditors sharing posts in mental health subreddits often use it as a venue of self-expression of their experiences around their illness challenges, as well as the impact of those experiences on their work, life, and relationships. Aside from that, seeking concrete diagnosis or treatment information is also not uncommon in these communities.

Social support. Our findings show that certain types of disclosure, e.g., posts with lowered inhibition, invite greater social support than others. In fact, even though redditors are not compensated for their actions, we observe the feedback manifested in the comments to be of surprisingly high quality, and ranges from emotional and instrumental, to information and prescriptive advice. This is an important contrast to social media like Twitter, where sharing health information is most times a broadcast or an emotional outburst (Paul & Dredze, 2011; De Choudhury et al., 2013), and not necessarily around seeking specific or quality information around treatment and diagnosis.

We also note the contrast of reddit use for mental health with the social network Facebook. While prior research has shown that Facebook status updates bear some degree of health-oriented information, privacy concerns around revealing "too much to people you know" often preclude one from seeking out for actionable advice around stigmatic health concerns (Newman et al. 2011; Young & Quan-Haase, 2009). In that light, reddit acts as a convenient medium because of its lack of personally identifiable information in user profiles. In comparison to online health forums too, our findings on reddit are distinct and complementary. Health forums often focus on specific health needs of individuals, and provide tools for easy sharing of health and illness information, ranging from current treatments, to symptoms and outcomes (Eysenbach et al. 2004). We observe that users do use reddit to fulfil some of these types of needs, however they often also use it simply for emotional support—an aspect considered very valuable in mental illness therapy (Cohen & Wills, 1985).

Anonymity. Finally, the small but distinctive use of throwaway accounts in the subreddits we study indicates that this feature of anonymity allows individuals to express views and thoughts relatively freely—something that may be considered to be extreme or unacceptable to the mainstream. That more than half of these throwaway account owners (61%) have used a particular throwaway account to post *exactly once* shows their propensity to make sure they do not leave any activity trails behind. However, we observe that such anonymity does not hinder the quality of social support redditors receive—in fact they garner more comments on such postings, and as we observe, tend to provide greater emotional sustenance, and are generally more involving and helpful in their suggestions and feedback. In essence, our findings align with prior literature (Bernstein et al. 2011; Schoenebeck et al. 2013), where anonymity has been found to be a positive feature of some online communities, because when it comes to mental illness, disinhibition in the form of journaling and discourse can be an effective healing process (Rude et al. 2004).

Nevertheless, as popular a social site as reddit is, further research is needed to understand these anonymity practices around mental health concerns. Compared to analogous anonymous sites like the discussion board /b/ on 4chan on one end of the continuum, and social networks like Facebook which harvest physical world contacts (e.g., family, friends, coworkers) and identities on the other, it would be interesting to examine what factors affect disinhibition and self-perception of anonymity of users in these communities discussing stigmatic health concerns, and could be valuable insights for social media design.

Limitations

Our work is of course not free from limitations. We acknowledge that there are likely other subreddits beyond the ones we study where stigmatic information and topics as mental health are discussed. This work is a preliminary exploration, focusing on a set of high precision reddit communities, however expanding to other subreddits is a ripe area of future research. Also, understanding the extent to which the greater reddit population engaged in mental illness discourse embodies the observed behavior, is also valuable from a generalization perspective. Further note that by no means the goal of this research is to claim all of the individuals posting in the subreddits of interest actually suffered from mental illness: we can only make a weak inference about it from their interest in the forums we study. Future research will benefit from a mixed methods approach of backing up our quantitative analysis with qualitative investigations about the health status of the communities.

Conclusion

We have presented, to the best of our knowledge, the first comprehensive study of mental health discourse on the social media reddit. Our results showed evidence of considerable self-disclosure around mental health issues. In characterizing the nature of social support in comments on the mental health subreddits we studied, we observed that it can span several nuanced categories, providing emotional to

prescriptive feedback. Lastly, use of dissociative anonymity as featured via throwaway accounts, although limited, was found to be adopted as an information sharing practice for open conversations relating to mental illness. Interestingly, despite the negative or caustic nature of content shared by anonymous redditors, disinhibition led to more emotional and instrumental feedback.

This research reveals how social media like reddit are fulfilling unique information and social needs of a cohort challenged with a stigmatic health concern looking through the lenses of disclosure, social support, and disinhibition. Potentially, our work may provide a wealth of resources to clinicians, health practitioners, caregivers, and policy makers to identify communities at risk.

References

Bernstein, M. S., Monroy-Hernández, A., Harry, D., André, P., Panovich, K., & Vargas, G. G. (2011). 4chan and/b: An Analysis of Anonymity and Ephemerality in a Large Online Community. In *ICWSM 2011*.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *The Journal of machine Learning research*, *3*, 993-1022.

Caplan, S., & Turner, J. (2007). Bringing theory to research on computer-mediated comforting communication. *Comp in human behavior*, *23*(2), 985-998.

Cohen, S., & Wills, T. A. (1985). Stress, social support, and the buffering hypothesis. *Psychological bulletin*, *98*(2).

Chester, Andrea, and O'Hara, Agi. (2007). Image, Identity and Pseudonymity In Online Discussions. *International Journal of Learning* 13.12 (2007): 193-203.

Chung, C., & Pennebaker, J. W. (2007). The psychological functions of function words. *Social comm.*, 343-359.

Corrigan, P. (2004). How stigma interferes with mental health care. *American Psychologist*, *59*(7), 614.

De Choudhury, M., Gamon, M., Counts, S., & Horvitz, E. (2013). Predicting Depression via Social Media. In *ICWSM*

De Choudhury, M., Morris, M. R., & White, R. W. (2014). Seeking and Sharing Health Information Online: Comparing Search Engines and Social Media. In *Proc. CHI 2014*, to appear.

Fox, S. (2013). Health. *Pew Internet*. 2013-01.

Eaves, G., & Rush, A. J. (1984). Cognitive patterns in symptomatic and remitted unipolar major depression. *Journal of Abnormal Psychology*, *93*(1), 31.

Eysenbach, G., Powell, J., Englesakis, M., Rizo, C., & Stern, A. (2004). Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. *BMJ*, *328*(7449), 1166.

George, L. K., Blazer, D. G., Hughes, D. C., & Fowler, N. (1989). Social support and the outcome of major depression. *The British J. of Psychiatry*,*154*(4), 478-485.

Johnson, G.J. and Ambrose, P.J. (2006). Neo-tribes: the power and potential of online communities in health care. *CACM*, 49(1): 107-113.

Lapidot-Lefler, N., & Barak, A. (2012). Effects of Anonymity, Invisibility, and Lack of Eye-Contact on Toxic Online Disinhibition. *Computers in Human Behavior* 28(2): 434-43.

Liu, L.S., Huh, J., Neogi, T., Inkpen, K. and Pratt, W. (2013). Health vlogger-viewer interaction in chronic illness management. In *Proc. CHI*, 49-58.

Moreno MA, Jelenchick LA, & Egan KG, et al. (2011). Feeling bad on Facebook: Depression disclosures by college students on a social networking site. *Depressions and Anxiety*. 28(6):447–455.

Newman, M. W., Lauterbach, D., Munson, S. A., Resnick, P., & Morris, M. 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 *Proc. CSCW*. 341-350.

Oh, H.J., Lauckner, C., Boehmer, J., Fewins-Bliss, R. and Li, K. (2013). Facebooking for health: an examination into the solicitation and effects of health-related social support on social networking sites. *Comp. in Human Behavior*, 29(5): 2072-2080.

Park, M., McDonald, D. W., & Cha, M. (2013). Perception Differences between the Depressed and Non-depressed Users in Twitter. In *Proc. ICWSM 2013*.

Paul, M. and Dredze, M. (2011). You are what you tweet: analyzing Twitter for public health. In *Proc. ICWSM*.

Posternak M, Solomon D, & Leon A. (2006). The naturalistic course of unipolar major depression in the absence of somatic therapy. *J. Nerv & Mental Disease* 194(5):324–29.

Prigerson, H. G., Monk, T. H., Reynolds, C. F., Begley, A., Houck, P. R., Bierhals, A. J., & Kupfer, D. J. (1995). Lifestyle regularity and activity level as protective factors against bereavement-related depression in later life. *Depression*, 3(6), 297-302.

Rabkin, J. G.; & Struening, E. L. (1976). Life events, stress, and illness. *Science*, 194(4268), 1013-1020.

Rude, S.; Gortner, E.; & Pennebaker, J. (2004). Language use of depressed and depression-vulnerable college students. *Cognition and Emotion*, 1121-1133.

Schoenebeck, S. Y. (2013). The Secret Life of Online Moms: Anonymity and Disinhibition on YouBeMom.com. In *Proc. ICWSM 2013*.

Sillence, E., Briggs, P., Harris, P.R. and Fishwick, L. (2007). How do patients evaluate and make use of online health information? *Soc. Sci. Med.*, 64(9): 1853-1862.

Turner, R. J., Frankel, B. G., & Levin, D. M. (1983). Social support: Conceptualization, measurement, and implications for mental health. *Research in Community & Mental Health*.

White, M. and Dorman, S.M. (2001). Receiving social support online: implications for health education. *Health Education Research,* 16 No. 6 (2001), 693--707.

Young, A.L. and Quan-Haase, A. (2009). Information revelation and internet privacy concerns on social network sites: a case study of Facebook. In *Proc. C&T*, 265-274.