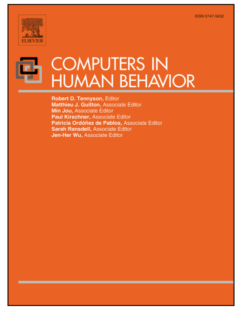
Accepted Manuscript



Examining thematic similarity, difference, and membership in three online mental health communities from reddit: A text mining and visualization approach

Albert Park, Mike Conway, Annie T. Chen

PII: S0747-5632(17)30518-6

DOI: 10.1016/j.chb.2017.09.001

Reference: CHB 5146

To appear in: *Computers in Human Behavior*

Received Date: 15 February 2017

Revised Date: 8 August 2017

Accepted Date: 1 September 2017

Please cite this article as: Park A., Conway M. & Chen A.T., Examining thematic similarity, difference, and membership in three online mental health communities from reddit: A text mining and visualization approach, *Computers in Human Behavior* (2017), doi: 10.1016/j.chb.2017.09.001.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Examining Thematic Similarity, Difference, and Membership in Three Online Mental Health Communities from Reddit: A Text Mining and Visualization Approach

Albert Park, PhD[a], Mike Conway, PhD[a], Annie T. Chen, PhD[b]

[a]Department of Biomedical Informatics, School of Medicine

University of Utah

421 Wakara Way Ste 140, Salt Lake City, UT 84108-3514, USA

[b]Department of Biomedical Informatics and Medical Education, School of Medicine

University of Washington

Box SLU-BIME 358047, 850 Republican St, Building C, Seattle, WA 98109-4714, USA

al.park@utah.edu, mike.conway@utah.edu, atchen@uw.edu

Corresponding author:

Albert Park will handle correspondence at all stages of refereeing and publication.

Phone: (206) 743-7843

Email: al.park@utah.edu or alpark1216@gmail.com

ACCEPTED MANUSCRIPT

Examining Thematic Similarity, Difference, and Membership in Three Online Mental Health Communities from Reddit: A Text Mining and Visualization Approach

Abstract

Objectives. Social media, including online health communities, have become popular platforms for individuals to discuss health challenges and exchange social support with others. These platforms can provide support for individuals who are concerned about social stigma and discrimination associated with their illness. Although mental health conditions can share similar symptoms and even co-occur, the extent to which discussion topics in online mental health communities are similar, different, or overlapping is unknown. Discovering the topical similarities and differences could potentially inform the design of related mental health communities and patient education programs. This study employs text mining, qualitative analysis, and visualization techniques to compare discussion topics in publicly accessible online mental health communities for three conditions: Anxiety, Depression and Post-Traumatic Stress Disorder.

Methods. First, online discussion content for the three conditions was collected from three Reddit communities (r/Anxiety, r/Depression, and r/PTSD). Second, content was pre-processed, and then clustered using the k-means algorithm to identify themes that were commonly discussed by members. Third, we qualitatively examined the common themes to better understand them, as well as their similarities and differences. Fourth, we employed multiple visualization

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

techniques to form a deeper understanding of the relationships among the identified themes for the three mental health conditions.

Results. The three mental health communities shared four themes: sharing of positive emotion, gratitude for receiving emotional support, and sleep- and work-related issues. Depression clusters tended to focus on self-expressed contextual aspects of depression, whereas the Anxiety Disorders and Post-Traumatic Stress Disorder clusters addressed more treatment- and medication-related issues. Visualizations showed that discussion topics from the Anxiety Disorders and Post-Traumatic Stress Disorder subreddits shared more similarities to one another than to the depression subreddit.

Conclusions. We observed that the members of the three communities shared several overlapping concerns (i.e., sleep- and work-related problems) and discussion patterns (i.e., sharing of positive emotion and showing gratitude for receiving emotional support). We also highlighted that the discussions from the r/Anxiety and r/PTSD communities were more similar to one another than to discussions from the r/Depression community. The r/Anxiety and r/PTSD subreddit members are more likely to be individuals whose experiences with a condition are long-term, and who are interested in treatments and medications. The r/Depression subreddit members may be a comparatively diffuse group, many of whom are dealing with transient issues that cause depressed mood. The findings from this study could be used to inform the design of online mental health communities and patient education programs for these conditions. Moreover, we suggest that researchers employ multiple methods to fully understand the subtle differences when comparing similar discussions from online health communities.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

keyword. Consumer Health Information; Anxiety Disorders; Depression; Stress Disorders, Post-Traumatic; Unsupervised Machine Learning; Consumer Health Information;

1. Introduction

Social media platforms, including online health communities, have become popular resources to exchange social support [1] (e.g., informational, emotional, instrumental) with others [2, 3]. This movement helps individuals to cope with and manage their illnesses while also providing the means to overcome barriers like geographical isolation, physical challenges, and stigma of disease. For example, one in four Internet users living with a chronic condition sought information from a peer with a similar condition by 2011 [4]. Peers can also offer advice on condition management [5], emotional support [6, 7, 8], and information to address everyday issues [2, 9].

Studies have consistently shown individuals can gain positive effects from interacting with other individuals in similar circumstances. Online interactions have been shown to improve depression [10, 11, 12, 13, 14], anxiety [12, 13, 15], stress [12, 13], and negative mood [16], as well as facilitate coping [17] and empowerment [10, 11, 13, 15, 17, 6]. Moreover, members of online health communities consistently emphasize the benefits of participation with respect to their treatment decisions, symptom management, clinical management, and outcomes [10, 11].

Individuals suffering from mental disorders often experience difficulty in obtaining support from their immediate social ties due to social stigma and discrimination associated with their illnesses [18, 19]. For such individuals, online health communities can be a useful medium to express their thoughts and feelings. However, extant literature has also reported that negative emotion can

ACCEPTED MANUSCRIPT

spread through interaction [20], and members of mental health communities have shown significant increases in anxiety, anger, and negative emotion following reports of celebrity suicides [21].

Although mixed results exist with respect to the effect of social media on mental health outcomes, the popularity among users is increasing. Additionally, an increasing number of researchers have employed statistical methods to study mental health in social media[1] [22]. For example, researchers have found that individuals have engaged in discussions about their mental illness on social media [18, 23], found associations between use of multiple social media platforms and symptoms of depression and anxiety [24], found chatter supporting marijuana use for Post-Traumatic Stress Disorder (PTSD) treatments [25], used social media to predict individuals at risk for depression [26], compared the longitudinal psychological changes in members of an online depression health community against members of other online health communities [27], classified social media contents with depressive symptoms [28, 23], characterized smoking and drinking problems [29, 30], tracked opioid related discussions [31], and classified substance addiction phases [32].

In this study, we examine the nature of online discussion pertaining to three mental health conditions: anxiety, depression, and PTSD. The connection between anxiety and depression, its shared symptomatology, and co-morbidity has been a subject of previous research [33]. Moreover, it has been observed that anxiety and depression often co-occur in the presence of stressful and traumatic events, and in connection to other health conditions such as chronic pain [34, 35]. Thus, the extent to which discussion topics in these communities are similar, different, or overlapping is of interest. However, at least to our knowledge, comparison of discussion topics has not been a focus of previous research on online

[1] here broadly defined to include internet discussion communities like Reddit

ACCEPTED MANUSCRIPT

mental health communities.

In addition, individuals who seek social support online often do so in order to find informational and emotional support or other types of support sources [36]. Though the types of exchanged social support with regard to depression have been studied [19, 37, 38], less is known about online social support exchanges concerning the other two conditions of interest. Discovering the most import discussion topics and understanding how members are utilizing respective communities could potentially inform the design of related mental health communities and patient education programs.

We aim to fill these gaps in the literature with this study and answer three research questions (RQ):

RQ1 : what are the main themes expressed in the communities?

RQ2 : how much thematic overlap, similarity, and difference exists among the communities?

RQ3 : what can we understand about the overlapping member base?

Our approach employs document clustering techniques, along with qualitative and visual analysis, to compare discussion topics in online health communities focusing on anxiety, depression, and PTSD. In this work, we focus on Reddit, a highly popular social gathering platform. Reddit has been shown to be a well utilized social media platform for stigmatized illnesses, including mental disorders [18]. We focus on discussion topics in the following three subcommunities: `r/Anxiety`, `r/Depression`, and `r/PTSD`[2].

ACCEPTED MANUSCRIPT

[2] To maintain clarity, we use `r/` followed by a fixed width style when anxiety, depression, and PTSD are referring to communities

ACCEPTED MANUSCRIPT

2. Material and methods

2.1. Data collection

Our corpus was based on Reddit (`http://www.reddit.com/`), a popular social networking, online gathering, and news exchanging platform. In 2015, Reddit members participated in over 88,000 subreddits (i.e., topically focused sub-communities) and generated 83 billion page views. Members of the Reddit wrote over 73 million individual posts (i.e., a submission that starts a conversation) along with over 725 million comments (i.e., a submission that replies to posts or other comments) in subreddits [39].

We used Reddit's official Application Programming Interface (API) [40], called the Python Reddit API Wrapper (PRAW) [41], to collect data from Reddit.com. Reddit allows members to organize sub-communities called 'subreddits'. We focused on three subreddits: `r/Anxiety`, `r/Depression`, and `r/PTSD`. We collected the title, author id, timestamp, post or comment id, parent id (i.e., the targeted comment or post, in which the author was replying), number of direct replies, scores (i.e., the difference between up votes and down votes), and the content. All three subreddits are public communities allowing anyone with Internet access to view the content. By December of 2015, the `r/Anxiety` subreddit had been active for 7 years with 73,501 members; the `r/Depression` subreddit had been active for 6 years with 126,348 members; the `r/PTSD` subreddit had been active for 7 years with 5,257 members.

Between the months of Oct, 2015 to Dec, 2015, we downloaded a total of 7,410 posts and 132,599 associated comments that were made by 41,967 unique members. These comments were generated between January of 2011 and December of 2015. Because the Reddit API limited downloading posts to 1,000 posts at a time, we used different methods to collect larger datasets that could cover various topics discussed in these subreddits. First, we downloaded the

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

'top' rated posts of all time. Top posts were determined by members as they vote 'up' or 'down' over the lifetime of the posts [42]. We started with the top rated posts to systematically collect the most relevant topics to the community as a whole. We relied on collective opinions of the communities as a logical starting point for understanding the most relevant topics of the communities. Second, we supplemented our dataset by collecting 'hot' posts — recent top rated posts (see [43] for an account of the Reddit voting system). We added hot posts to cover newly emerging important and informative topics; however, we noticed a large overlap with top rated posts. Third, we added 'new', the most recent posts, for up to 15 days to cover diverse topics shared in these communities. Fourth, we removed repeated posts. The `r/Depression` subreddit was far more active than the `r/Anxiety` or `r/PTSD` subreddit in terms of the total number of posts, unique members, words, and associated comments. Thus, we downloaded only 3 days of new posts from the `r/Depression` subreddit whereas 15 days from the `r/Anxiety` and `r/PTSD` subreddits, to gather comparable sized datasets. Table 1 summarizes the dataset used for this research.

Table 1: Characteristics of three Reddit communities studied

	Anxiety Disorders	Depression	PTSD
Dates posts were written	9/2011-12/2015	1/2011-10/2015	7/2011-12/2015
Num. of posts	3,677	1,934	1,799
Num. of comments	49,929	67,599	15,071
Num. of members	15,336	23,916	2,712
Range of num. of comments in posts	1 to 200	1 to 201	1 to 87
Mean num. of comments in posts (Stdev)	13.58 (19.61)	34.95 (41.17)	8.38 (7.56)
Median num. of comments in posts	6	18	7
Total num. of words	3,079,219	3,573,228	1,615,328

We restricted our analysis to publicly available discussion content and [Blank for blind review] Institutional Review Board (IRB) [ethics committee] exempted the study procedure and data from review [Blank for blind review] . Although publicly available, individuals from the communities of our interest are nevertheless suffering from stigmatized illness. For this reason, we followed the

ACCEPTED MANUSCRIPT

guidelines suggested by Bruckman [44] and Eysenbach [45] to modify and de-identify our example quotes to ensure members' anonymity and to protect their privacy.

2.2. RQ1: What are the main themes expressed in the communities?

Automatically identifying discussion themes involves organizing the data content into external classifications like Unified Medical Language System (UMLS) [46] or clusters. In this study, we used k-means clustering, a widely used unsupervised clustering algorithm [47]. Previous research has employed document clustering techniques to analyze discussion content in online health communities [48]. Document clustering techniques create clusters of documents that are similar to one another, but dissimilar from documents in other clusters [47]. In other words, we are creating topically similar clusters that contain high volumes of the same terms, a useful method for identifying main discussion themes in a large collection of documents. We elected to use an unsupervised algorithm because of the lack of a ground truth dataset.

We first used the Python Natural Language Toolkit (NLTK) (Version 3.1) [49] and Scikit-learn (Version 0.17) [50] to pre-process our dataset, which involved the removal of stop words, punctuation, high- and low-frequency terms, as well as tokenization. We then represented our data in vector spaces by generating term frequency matrices and weighted the terms according to their normalized term frequencies [50]. Each post and its associated comments were considered as a single document. We used Scikit-learn [50] to cluster our data using its default parameters for k-means clustering and estimated topic similarity with cosine similarity. After experimenting with varying numbers of clusters for each condition, we generated 15 clusters for each community for comparative purposes.

One limitation of the k-means clustering algorithm is that the algorithm can

ACCEPTED MANUSCRIPT

produce different clusters depending on the starting seeds [47]. Therefore, to check the validity of our clustering result, we repeated the clustering process and manual assignment of descriptive labels 10 times, each with 15 clusters, using the same parameters and procedures. Each time, we manually labeled the clusters according to its dominant theme, and created new labels if needed. In other words, if k-means clustering algorithm produces identical results in all 10 procedures, we would only need a total of 15 labels to describe all 150 clusters. However, if results were vastly different, we would need up to 150 labels to describe all the clusters. We then calculated the overlapping terms for identically labeled clusters that were generated from different clustering processes. In other words, we calculated overlapping terms of cluster 'A', if procedure 1 and 2 both produced cluster 'A'. We considered the 50 most frequently occurring terms when calculating the overlapping terms and also tracked the total number of labels required to describe all 150 clusters. We used the degree of overlap between identically labeled clusters in solutions with different starting seeds as a validity check of our cluster solution. We provide detailed information on our validity check process in the appendix.

To characterize each cluster, we qualitatively examined the most frequent terms in each cluster, as well as the titles and contents of several randomly selected example posts and their associated comments. We followed an open coding process [51], to identify and assign each cluster a descriptive label. This method was used to elicit unknown, emerging themes grounded in the data. We then visualized an overview of discussion themes as a bubble chart and a network visualization. We employed D3 [52] to construct a bubble chart, and made the cluster size (i.e., the number of documents in each cluster) proportional to the bubble size. We used Gephi (Version 0.9.1) [53], a popular network visualization tool to generate a network visualization. We applied the ForceAt-

ACCEPTED MANUSCRIPT

las2 [54] layout to gain an overview of the discussion theme network structure. In this network, each node represented a cluster. We sized the nodes in proportion to the normalized sized of the clusters, but kept the label as a fixed size for the purpose of reading labels of smaller nodes. To determine the edge weight, we employed the 20 most frequently occurring words of each cluster as a proxy for that cluster, calculated Jaccard similarity between each pair of clusters, and employed this similarity measurement as the edge weight between each pair. Jaccard similarity is a common method for comparing the similarity and diversity of sets [55].

2.3. RQ2: How much thematic overlap, similarity, and difference exists among the communities?

To examine thematic similarities and differences among identified discussion themes from RQ1, we first represented discussion themes as a Venn diagram to visualize thematic overlaps. We then qualitatively compare and contrast the common themes among the three subreddits.

We also applied the Louvain modularity algorithm [56] available in Gephi to determine the similarity among clusters in the network visualization. Modularity — a widely used method to identify community structures in a network — measures the vertices in a group of nodes and then compares to a random connection [57]. For our purposes, modularity can identify natural divisions of subgroupings of nodes (i.e., community structure) with respect to frequently occurring terms in the network representation. In other words, modularity can illustrate how clusters are topically similar or dissimilar from one another. We used the edge weight and randomize feature and set the resolution to 1. The distance between theme nodes can be also influenced by the layout in network analysis; thus we visualized Jaccard similarity scores as a heatmap to provide greater detail on topical similarities and differences.

ACCEPTED MANUSCRIPT

2.4. RQ3: What can we understand about the overlapping member base?

We investigated the characteristics of members who participated in multiple subreddits and the most commonly discussed themes by these overlapping members. To explore whether overlapping themes in the three subreddits were determined by overlapping memberships, we identified the five most common themes discussed by the overlapping members and the posting characteristics of these members.

3. Results

3.1. RQ1: What are the main themes expressed in the communities?

In this section, we present the main themes expressed in the three communities:`r/Anxiety`, `r/Depression`, and `r/PTSD` subreddits (Figure 1 on the next page).

3.1.1. Anxiety

We generated 15 clusters using the `r/Anxiety` subreddit discussion content. Many clusters including SOCIAL ANXIETY[3], MEDICATION, SCHOOL, PANIC ATTACK, and THERAPY/THERAPIST contained terms and labels which clearly differentiated the clusters from one another. However, a few clusters, such as POSITIVE EMOTION and GRATITUDE, shared terms. We distinguished these clusters from one another using the terms that they did not share, and the titles and contents of clustered posts. Although most of the labels are self-explanatory, WHAT DO YOU THINK may need a further explanation. The `r/Anxiety` subreddit contained many posts asking opinions of others on various topics from daily experiences to family situations. While the topics varied, the cluster label, WHAT DO YOU THINK, is based on this similarity in rhetorical style. Table 2 on

[3]to maintain clarity, we use small caps style when referring to cluster themes

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

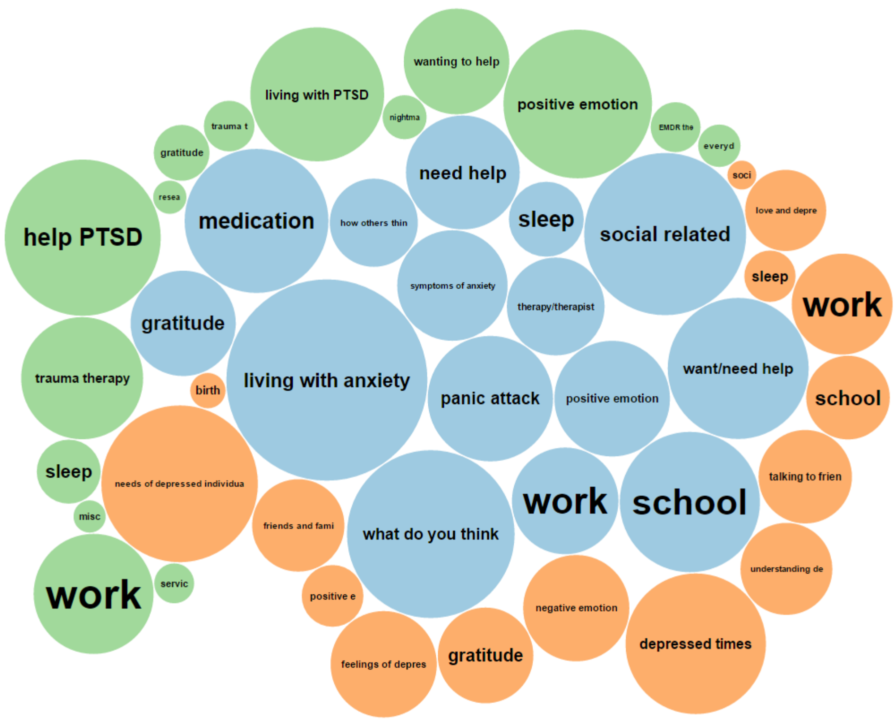


Figure 1: A bubble chart summarizing clustering results of the `r/Anxiety` (blue), `r/Depression` (orange), and `r/PTSD` (green) subreddits.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

page 13 summaries the clustering result for the `r/Anxiety` subreddit and shows the themes in descending order of cluster size.

Table 2: Characteristics of the `r/Anxiety` clusters and sample terms.

Size (%)	Cluster label	Sample terms
584 (15.88%)	living with anxiety	day, try, getting, anxious, love, want, life, years, long, days
406 (11.04%)	what do you think	think, people, help, way, better, make, need, anxious, right, bad
379 (10.31%)	social related	social, people, think, friends, anxious, talk, phone, life, bad, talking,
299 (8.13%)	medication	doctor, medication, taking, effects, help, meds, dose, lexapro, zoloft, prescribed
287 (7.81%)	want/need help	want, make, help, person, need, talk, work, understand, feeling, depression
284 (7.72%)	school	school, college, high, years, people, work, home, family, friends, anxious
227 (6.17%)	panic attack	panic, attack, heart, anxious, felt, fear, doctor, stop, calm, breathing
193 (5.25%)	positive emotion	good, great, awesome, people, hope, thank, congrats, luck, proud, happy
187 (5.09%)	need help	help, need, think, people, worry, say, talk, tell, mind, scared
176 (4.79%)	symptoms of anxiety disorder	feeling, anxious, panic, symptoms, life, people, fear, hard, times, stomach
164 (4.46%)	work	job, work, day, people, boss, working, home, jobs, stress, anxious
160 (4.35%)	gratitude	thanks, love, great, good, sharing, better, awesome, nice, glad, helped
138 (3.75%)	therapy/therapist	help, therapist, doctor, therapy, appointment, talk, medication, issues, health, insurance
112 (3.05%)	how others think of you	thoughts, think, thought, people, thinking, feeling, fear, person, negative, helped
81 (2.20%)	sleep	sleep, night, bed, asleep, day, wake, fall, sleeping, tired, morning

3.1.2. Depression

We generated 15 clusters for the `r/Depression` subreddit. Clusters including BIRTHDAY, SCHOOL, SLEEP, WORK, and GRATITUDE were clearly differentiated from one another. We distinguished clusters such as TALKING TO FRIENDS and FRIENDS AND FAMILY, which shared identical or semantically similar terms, using the procedure described in the previous subsection. Table 3 summaries the clustering result for the `r/Depression` subreddit and shows the themes in descending order of cluster size.

Table 3: Characteristics of the `r/Depression` clusters and sample terms

Size (%)	Cluster label	Sample terms
354 (18.30%)	needs of depressed individuals	need, depressed, talk, doctor, mental, understand, family, medication, anxiety, therapy
285 (14.74%)	depressed times	feeling, years, depressed, bad, hard, long, suicide, talk, days, end
163 (8.43%)	feelings of depression	feeling, bad, depressed, end, hard, sorry, worse, feels, tried, months
163 (8.43%)	negative emotion	bad, shit, trying, hard, fucking, fuck, depressed, sad
147 (7.60%)	work	work, job, hard, right, working, lot, trying, hate, money, days
132 (6.83%)	gratitude	thank, great, awesome, nice, job, proud, glad, amazing, beautiful, wow
126 (6.51%)	talking to friends	talk, friends, say, friend, person, care, tell, said, told, talking
123 (6.36%)	friends and family	friends, friend, family, best, away, love, college, great
122 (6.31%)	understanding depressed individuals	depressed, understand, feeling, bad, worse, need, hard, mental, illness, problems
101 (5.22%)	school	school, college, year, high, class, semester, grades, work, friends, parents
95 (4.91%)	love and depression	love, happy, hate, depressed, days, feeling, great, bad, guy, relationship
55 (2.84%)	positive emotion	happy, hope, luck, best, glad, great, thank, wish, awesome, love
38 (1.96%)	sleep	sleep, bed, night, wake, sleeping, dreams, asleep, morning, awake, waking
18 (0.93%)	birthday	birthday, happy, hope, friends, love, wish, great, celebrate, enjoy, facebook
12 (0.62%)	social related	living, friends, eat, social, joke, conversation, games, media, enjoy, topic

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

3.1.3. PTSD

Like the previous two subreddits, we generated 15 clusters for the `r/PTSD` subreddit. Many clusters including TRAUMA THERAPY, WORK, SLEEP, TRAUMA TRIGGER, EMDR THERAPY, NIGHTMARE, ANIMAL, and RESEARCH were clearly distinguishable. A few clusters, such as SLEEP and NIGHTMARE shared similar terms but also had distinctive and non-overlapping terms. They were distinguished from one another using the same procedure as above. Table 4 summaries the clustering result for the `r/PTSD` subreddit and shows themes in descending order of cluster size.

Table 4: Characteristics of the `r/PTSD` clusters and sample terms.

Size (%)	Cluster label	Sample terms
352 (19.57%)	help for PTSD	help, need, sorry, trying, hope, therapy, friends, family, trauma, relationship
319 (17.73%)	positive emotion	good, thanks, better, talk, hope, great, yes, love, sure, able
259 (14.40%)	living with PTSD	years, life, want, happened, help, day, year, hard, therapy, home
215 (11.95%)	trauma therapy	trauma, therapy, therapist, help, better, symptoms, abuse, brain, talk, traumatic
208 (11.56%)	work	work, anxiety, panic, job, deal, new, started, months, trying, hard
161 (8.95%)	wanting to help (by others)	help, want, talk, support, understand, care, symptoms, experience, health, situation
59 (3.28%)	sleep	sleep, night, wake, asleep, hours, sleeping, bed, nightmares, dreams, awake
45 (2.50%)	gratitude for sharing (techniques and stories)	thank, sharing, reading, posting, writing, beautiful, blog, share, powerful, appreciate
37 (2.06%)	trauma trigger	trigger, triggers, makes, trauma, triggered, flashbacks, watch, warnings, music, movie
36 (2.00%)	EMDR therapy	emdr, therapist, therapy, trauma, work, session, memories, sessions, helpful, effective
28 (1.56%)	nightmare	nightmares, prazosin, wake, doctor, anxiety, medication, fear, sorry, therapy, dose
26 (1.45%)	everyday issues	home, work, today, boyfriend, personal, hate, friends, dealing, couple, issues
23 (1.28%)	service animal	dog, service, dogs, help, animal, trained, support, better, able
16 (0.89%)	research	study, questions, survey, research, information, link, university, project, article, contact
15 (0.83%)	misc.	good, worry, guys, sorry, thanks, work, bad, getting, bit, reading

Figure 2 on the next page is an overview of the discussion theme network, in which blue, red, and green represent discussion themes from the `r/Anxiety`, `r/Depression`, and `r/PTSD` subreddits, respectively.

To check the validity of our clustering method, we repeated the clustering process and manual assignment of descriptive labels 10 times, each with 15 clusters, using the same parameters and procedures. On average k-means clustering of the `r/Anxiety` subreddit produced 80% overlapping terms for identically labeled clusters and 29 unique labels were used to describe 10 k-means clustering results. For the `r/Depression` subreddit, a total of 25 unique labels were used and on average the labels contained 75% overlapping terms for identically labeled clusters. K-means clustering for the `r/PTSD` subreddit produced

ACCEPTED MANUSCRIPT

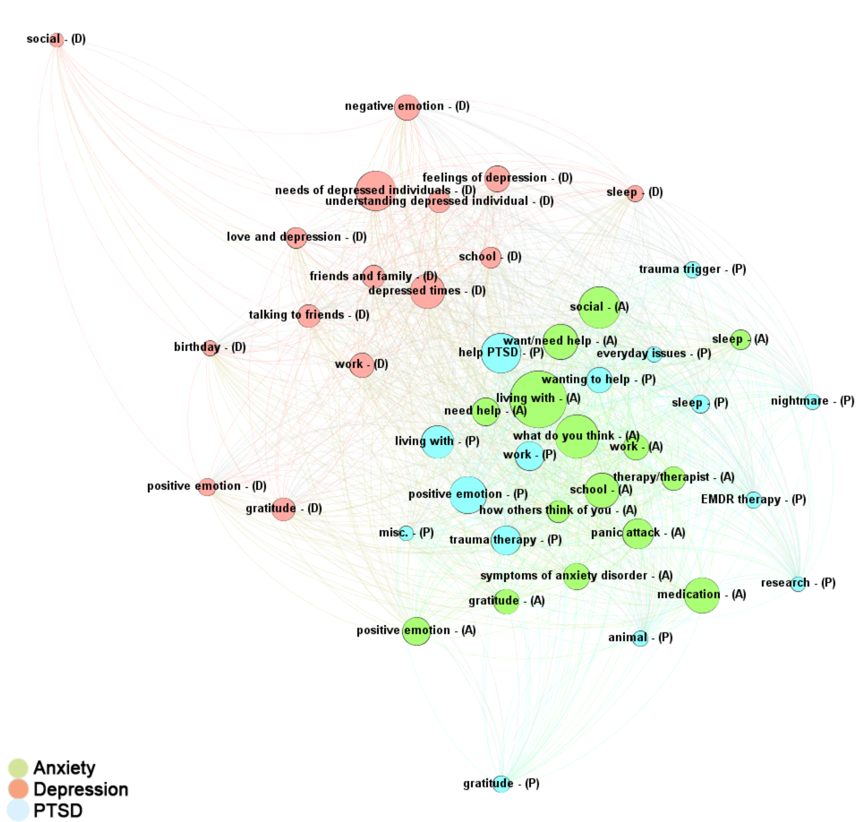


Figure 2: An overview of discussion theme network. The `r/Anxiety` denoted by (A) is in blue, the `r/Depression` denoted by (D) is in red, and the `r/PTSD` denoted by (P) is in green.

69% overlapping terms for identically labeled clusters and required a total of 28 unique labels. Detailed information on the validity check process as well as all the labels and portions of overlaps for the individual labels for each subreddit is provided in the appendix.

ACCEPTED MANUSCRIPT

3.2. RQ2: How much thematic overlap, similarity, and difference exists among the communities?

The three subreddits shared four discussion themes: sharing of POSITIVE EMOTION, GRATITUDE for received emotional support, and discussion related to SLEEP and WORK. Common themes are summarized in Figure 3 on page 16 as a Venn diagram. In the following sections, we present the results of our qualitative study of the four shared discussion themes.

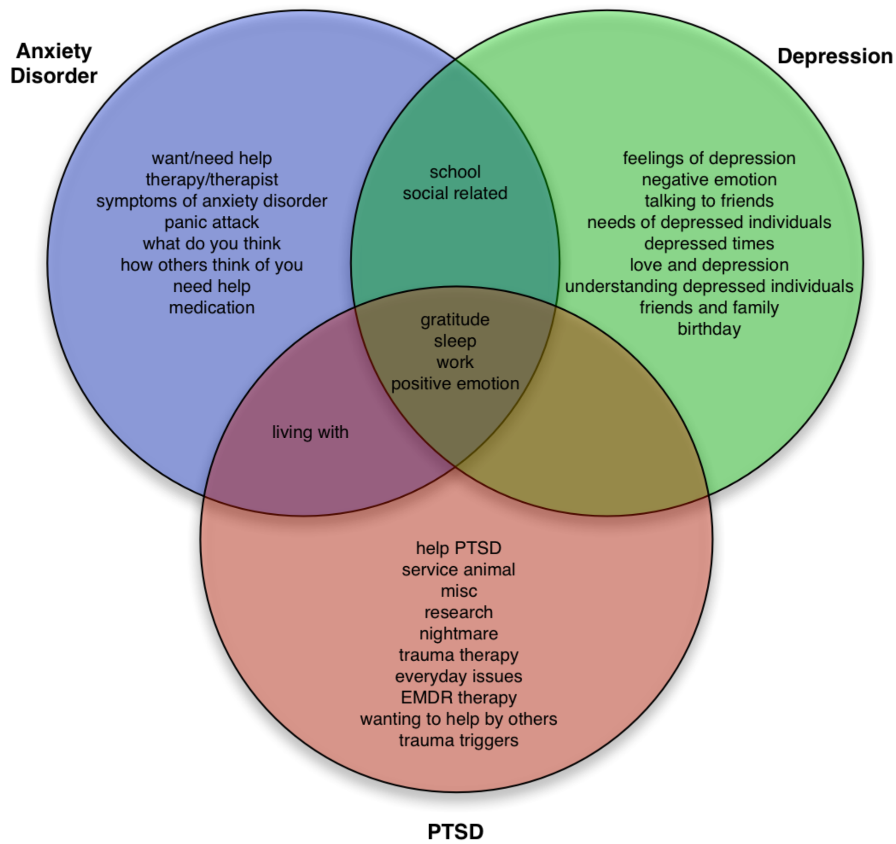


Figure 3: Summary of common themes and members as a Venn diagram

ACCEPTED MANUSCRIPT

3.2.1. Sharing of Positive Emotion

Sharing positive emotion is a well-documented practice in online health communities [7, 48, 58], including communities focusing on mental health [18]. Our dataset also consistently showed sharing positive emotion as a method of support. Below are canonical examples of Reddit members' attitude towards receiving positive emotion.

> "I had a girlfriend who abused me emotionally. I left her and I don't regret it. We need to have positive people in our lives!" - a comment from the `r/Depression` subreddit

> "Thanks! I still feel terrible, but I am more confident and I feel like I can get through this, with your support. This is an awesome community. As soon as I can [do so], I'll be sure to return the favor!" - a comment from the `r/PTSD` subreddit

It is not unexpected to find sharing positive emotion as a common theme in all three subreddits. Although the effect of positive emotion on clinical outcomes remains uncertain, previous literature suggests that exposure to positive emotions could alleviate negative emotion. For example, positive emotions could help individuals suffering from the distress of diseases or injuries [59], anxiety disorders [60], and chronic stresses [61, 62].

3.2.2. Showing Gratitude for Emotional Support

In our dataset, members frequently displayed gratitude for receiving positive posts, a common practice in online health communities [37]. The following are canonical examples of showing gratitude for received emotional support.

> "As someone with both anxiety disorder and depression, I find your comment very helpful in my times of weakness. Thank you so much!!!" - a comment from the `r/Anxiety` subreddit

ACCEPTED MANUSCRIPT

> "This is awesome, thank you so much for sharing your stories. Your positive reaction is overwhelming and I am thankful." - a comment from the `r/Depression` subreddit

> "Thank you for this insightful post! I hope that someday this will be available to those with other types of PTSD. Positive stories like this are encouraging, thanks!" - a comment from the `r/PTSD` subreddit

Similar to sharing positive emotion, showing gratitude has been suggested to be beneficial for individuals suffering from mental health conditions. For example, gratitude has led to higher levels of perceived emotional support, and lower stress and depression [63]. Moreover, making a conscious effort to acknowledge gratitude has been suggested to have emotional and interpersonal benefits [64], although further investigation regarding clinical outcomes is warranted.

3.2.3. Sleep

Sleep-related discussions were salient in all three subreddits. Subreddit members spoke of their issues generally, and to inquire whether others had similar experiences:

> "[...] I have problems sleeping as it is [...]" - a comment from the `r/Anxiety` subreddit

> "Who else is experiencing sleep trouble??" - a comment from the `r/PTSD` subreddit

The nature of sleep-related discussions was different in the three subreddits. Awakening, being tired throughout the day, and having trouble sleeping were commonly discussed in the SLEEP clusters for the `r/Anxiety` and `r/PTSD` subreddits. Panic was more salient in the `r/Anxiety` and `r/PTSD` subreddits, and nightmares, in the `r/PTSD` subreddit. For the `r/PTSD` subreddit, the k-means

ACCEPTED MANUSCRIPT

clustering algorithm yielded two sleep related clusters: SLEEP and NIGHTMARE. The SLEEP cluster contained general sleep related problems as well as nightmare issues as shown below. This result is consistent with the DSM-5 classification [65] and extant literature on PTSD and sleep [66].

> "[...] Whenever I see a fire or a related image, I get horrible nightmares. I wish it would stop! [...]" - a comment from the `r/PTSD` subreddit

In the `r/Depression` subreddit, discussions relating to sleep were much different. Topics that were frequently discussed by the `r/Depression` community included feeling unrefreshed after sleeping, regardless of how long they slept.

> "[...] man, I slept for 12 hours last night but I feel like I only got a couple of hours of sleep [...]" - a comment from the `r/Depression` subreddit

Another major theme in the `r/Depression`'s SLEEP cluster was members' desire to die in their sleep.

> "Yeah I feel you. I have a nice car, house, and a nice life, yet every night when I go to bed, I hope I never wake up." - a comment from the `r/Depression` subreddit

Although the results of the cluster analysis showed that all three subreddits had discussions related to sleep, our qualitative analysis showed differences in context. In the `r/Anxiety` and `r/PTSD` subreddits, members talked about issues related to sleep troubles such as nightmares, whereas the `r/Depression` subreddit, discussions were more about unrefreshing sleep and/or feelings of wanting to die, which may be exacerbated at bedtime.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

3.2.4. Work

As was the case with sleep, work-related discussions were prevalent in all three subreddits. In the `r/Anxiety` and `r/PTSD` subreddits, conversations were about the difficulty of keeping, performing, or getting a job due to their symptoms.

> "So I had to quit my job again because of my anxiety. This has been happening for the past several years. [...]" - a comment from the `r/Anxiety` subreddit

> "I've been jobless for the last couple years, so I was happy to get a new job. But now, I am always tired and worn out. I realize it's part of adjusting to a new life. However, now I've started to get flashbacks and nightmares again." - a comment from the `r/PTSD` subreddit

However, in the `r/Depression` subreddit, work-related content was about working too much, quitting/time off from work due to depression, and venting members' dislike of their work.

> "[...] I have no interest in working, but I have no choice but to make a living [...]" - a comment from the `r/Depression` subreddit

> "I had to take a leave of absence from work. So I can sort out my life stressors [...]" - a comment from the `r/Depression` subreddit

3.2.5. Shared themes in the `r/Anxiety` and `r/Depression` subreddits

The `r/Anxiety` and `r/Depression` subreddits shared two discussion themes: discussion of SCHOOL and SOCIAL related issues.

> "I am forced to take some time off from school due to anxiety and depression [...]" - a comment from the `r/Anxiety` subreddit

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

> "[...] I'm 16 years old. Till now, I was always homeschooled. I just now started to attend public school. [...] I am socially awkward and I get anxious around students my age. But it's getting worse since I started public school. [...]" - a comment from the `r/Depression` subreddit

As shown in the example, many members explicitly mentioned anxiety and depression together, especially with a topic regarding school. This demonstrated the importance of school-related issues for both subreddits, though the foci were slight different. For the `r/Anxiety` subreddit, social anxiety was the main topic of the SOCIAL cluster, whereas for the `r/Depression` subreddit, the topics ranged widely from reminiscing about one's past social life, to using social media to cope with depression.

3.2.6. Shared themes in the `r/Anxiety` and r/PTSD subreddits

The `r/Anxiety` and `r/PTSD` subreddits shared one common theme, LIVING WITH their respective conditions. Both clusters showed issues relating to daily struggles or mundane bad experiences, however. Another similar topic between the two subreddits was related to help. The `r/PTSD` subreddit had two help related clusters, WANTING TO HELP and HELP FOR PTSD. Many members of the `r/PTSD` subreddit were individuals who did not have PTSD but wanted to help other individuals who are suffering from PTSD. We labeled this cluster as WANTING TO HELP. The other cluster was labeled HELP FOR PTSD in which members were explicitly asking for help. The `r/Anxiety` subreddit had both types of discussions (i.e., WANT/NEED HELP and NEED HELP); however, in our qualitative analysis, the `r/Anxiety` subreddit had far fewer discussions where friends and family members were asking for advice on how to help individuals with anxiety disorder. Moreover, those discussions on 'wanting to help' were typically clustered together with discussions regarding 'asking for help', thus

ACCEPTED MANUSCRIPT

we called this cluster WANT/NEED HELP. Another help related cluster from the `r/Anxiety` subreddit was NEED HELP, in which members were expressing their needs and explicitly asking for help. Though these topics are somewhat related, we have treated these clusters as different in Figure 3 on page 16 to preserve the contextual distinction.

We found that the four common themes — POSITIVE EMOTION, GRATITUDE, SLEEP, and WORK — were not necessarily collocated in the network (Figure 4 on the following page). For instance, POSITIVE EMOTION node from the `r/Depression` cluster (red) is not adjacent to POSITIVE EMOTION node from the `r/Anxiety` cluster (blue). However, common theme nodes from the `r/Anxiety` and `r/PTSD` (green) subreddits are always closer to each other than to the corresponding theme from the `r/Depression` subreddit. For instance, SLEEP nodes from the `r/Anxiety` and `r/PTSD` subreddits are closer to each other than to SLEEP node from the `r/Depression` subreddit. Moreover, the nodes for LIVING WITH — a theme found only in the `r/Anxiety` and `r/PTSD` subreddits — are located relatively close, whereas SCHOOL and SOCIAL — themes found only in the `r/Anxiety` and `r/Depression` subreddits — have more distance between nodes.

From distance differences among common nodes, we observed that the `r/Anxiety` and `r/PTSD` subreddits shared more common terms with themselves than with the `r/Depression` subreddit. To validate this observation, we applied the Louvain modularity algorithm and color coded the communities according to the modularity result (Figure 4 on the next page).

We observed that the algorithm clearly divided the network into two main communities of nodes. All the discussion themes from the `r/Depression` subreddit were grouped as one community and all the discussion themes from the `r/Anxiety` and `r/PTSD` subreddits were grouped as another community. A

ACCEPTED MANUSCRIPT

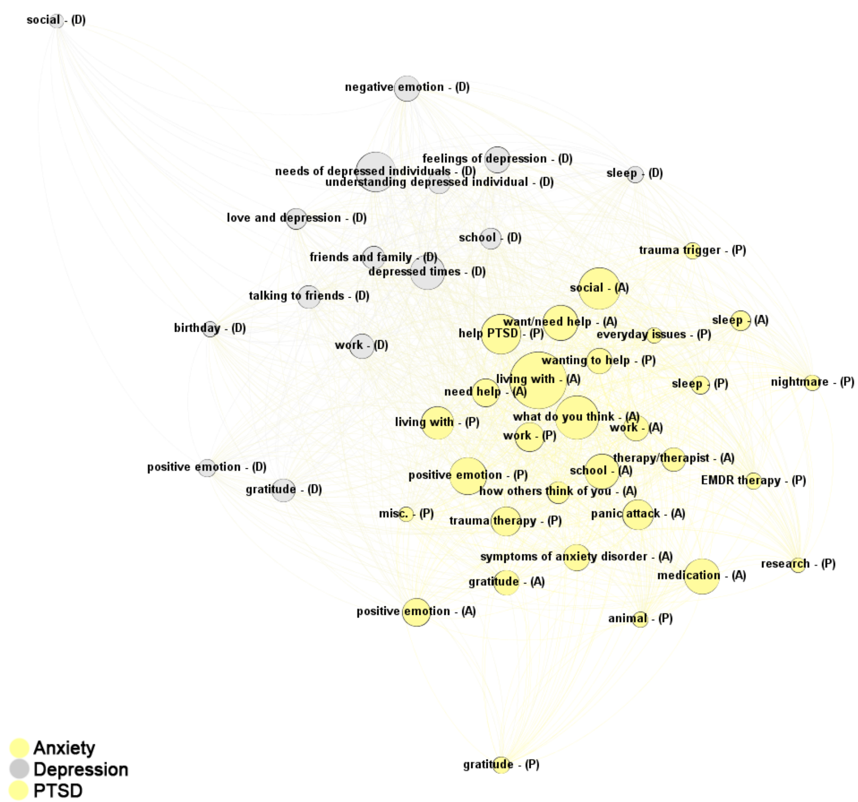


Figure 4: An overview of modularity structure in discussion themes. The `r/Anxiety` denoted by (A) and the `r/PTSD` denoted by (P) are in the same group (yellow), whereas the `r/Depression` denoted by (D) is in another group (gray).

heatmap (Figure 5 on the following page) also shows that `r/Anxiety` and `r/PTSD` subreddit are strongly linked compared to each other than to the `r/Depression` subreddit. Similar to the Figure 2 on page 15, SOCIAL from the `r/Depression` and GRATITUDE from the `r/PTSD` show less commonly shared words with any other nodes. The Louvain modularity algorithm and a heatmap support the assertion that concerns of the `r/Anxiety` and `r/PTSD` subreddit members are semantically more similar to one another than to the `r/Depression` subreddit.

ACCEPTED MANUSCRIPT

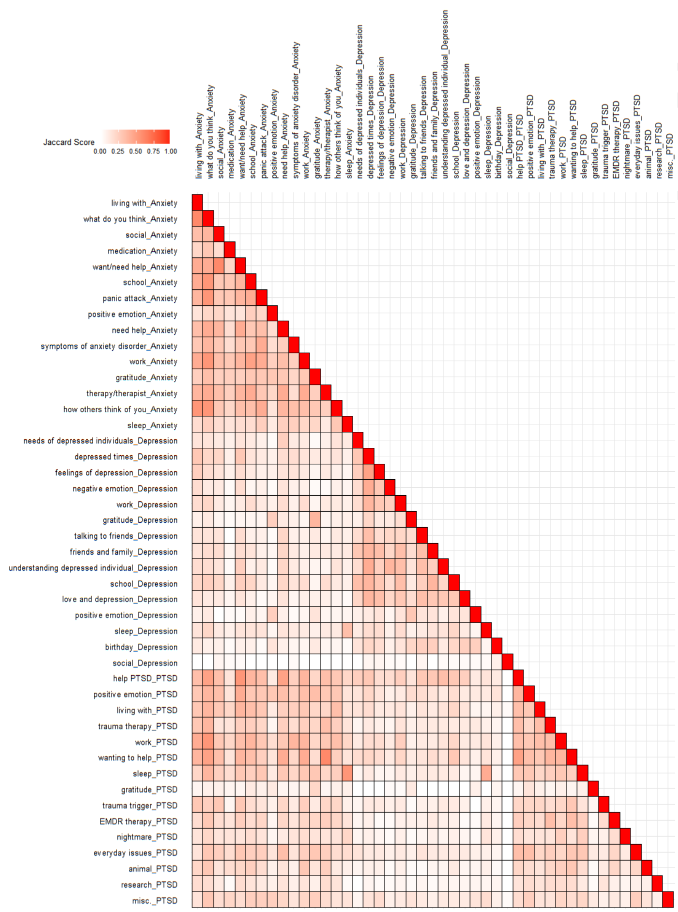


Figure 5: Heatmap representation of discussion themes.

ACCEPTED MANUSCRIPT

3.3. RQ3: What can we understand about the overlapping member base?

We investigated the extent to which members participated in multiple subreddits to verify that common themes were not mainly due to overlapping memberships. 5.96% (2,357 out of 39,541 members) participated in more than one of the three subreddits; these members participated in multiple topics and their discussions were not especially concentrated in overlapping themes. Only a small number of individuals participated in all three subreddits (n=65); some of these individuals explicitly mentioned the comorbidity and their thoughts and experiences concerning the conditions.

> "In my opinion, Anxiety is your mind speeding up – having more thoughts and worries while depression is your body slowing down – having less energy and sleeping more. When you add slowed body to sped up mind, everything becomes out of balance. You feel like you are constantly fighting youself, and nothing gets done except that you tire yourself out." - a comment from the `r/Anxiety` subreddit

> "I have Anxiety, Depression and PTSD. I often think it's scary what's inside my head." - a comment from the `r/Depression` subreddit

The `r/Anxiety` and `r/Depression` subreddits had a substantial number overlapping members (n=2,037), however, the two subreddits were also much bigger than PTSD 1 on page 7. The most commonly discussed themes by the overlapping members differed in the two subreddits. *Anxiety Disorders* and *PTSD* shared 217 overlapping members and also showed differences in what overlapping members most commonly discussed in the subreddits. `r/Depression` and *r/PTSD* did not have a common theme, but shared 233 members between the two subreddits. The main discussions made by these 233 members were also

ACCEPTED MANUSCRIPT

different, suggesting differential uses of these subreddits (Table 5 on the next page).

Table 5: Characteristics of members who participated in multiple subreddits and most commonly discussed themes by these overlapping members

	in Anxiety Disorder	in Depression
Num. of Overlapping Members	2,037	2,037
Num. of Total Posts	8,124	8,547
Mean (Stdev)of Num. posts	3.99(8.44)	4.20(7.41)
Five Most Occurring Themes (%)	1. what do you think (24.57%) 2. want/need help (17.23%) 3. social anxiety(9.05%) 4. gratitude (9.02%) 5. positive emotion (8.12%)	1. depressed times (32.76%) 2. gratitude (12.05%) 3. understanding depressed individual (10.82%) 4. negative emotion (10.04%) 5. talking to friends (9.86%)
	in Anxiety Disorder	in PTSD
Num. of Overlapping Members	217	217
Num. of Total Posts	962	1,507
Mean (Stdev)of Num. posts	4.43(5.42)	6.94(10.58)
Five Most Occurring Themes (%)	1. what do you think (22.97%) 2. want/need help (18.61%) 3. gratitude (9.36%) 4. panic attack (9.25%) 5. social anxiety (8.52%)	1. help PTSD (26.74%) 2. work (13.87%) 3. trauma therapy (13.07%) 4. living with PTSD (9.42%) 5. positive emotion (9.09%)
	in Depression	in PTSD
Num. of Overlapping Members	233	233
Num. of Total Posts	1,214	1,799
Mean (Stdev)of Num. posts	5.21(11.50)	7.72(12.21)
Five Most Occurring Themes (%)	1. depressed times (40.53%) 2. negative emotion (13.10%) 3. understanding depressed individual (12.85%) 4. gratitude (9.31%) 5. talking to friends (7.41%)	1. help PTSD (28.40%) 2. work (14.34%) 3. trauma therapy (12.62%) 4. living with PTSD (11.17%) 5. wanting to help (9.34%)

4. Discussion

Understanding the nature of online discussion from similar online health communities can be challenging, especially if the members share similar symptoms and co-morbidity. In this study, we not only to compare the overall discussion themes and the contextual variations among the same themes, but also to identify differences in participation and discussion styles using content from Reddit. It has been reported that anxiety and depression often co-occur in the presence of stressful and traumatic events [34, 35]. Thus, we analyze `r/Anxiety`, `r/Depression`, and `r/PTSD` subreddits.

ACCEPTED MANUSCRIPT

We first employed cluster analysis to examine the 15 main themes that were discussed in the `r/Anxiety`, `r/Depression` and `r/PTSD` subreddits. As expected, there were common topics that appeared in multiple subreddits. In particular, the three subreddits shared four discussion themes: POSITIVE EMOTION, GRATITUDE, SLEEP, and WORK. To gain a better insight, we then qualitatively analyze the four common themes.

Sharing of POSITIVE EMOTION and showing of GRATITUDE are themes that have also been reported in past research on health-related online communities [18, 7, 48, 58], and it was not unexpected to see these expressed here. SLEEP- and WORK-related problems were salient in all three subreddits, though they were discussed in a slightly different manner. Through manual examination, we discovered that members of the `r/Anxiety` and `r/PTSD` subreddits described their issues differently from members of the `r/Depression` subreddit (see Results).

This result was corroborated by our theme network analysis, in which the Louvain modularity algorithm separated the `r/Depression` subreddit's discussion themes from the `r/Anxiety` and `r/PTSD` subreddits' discussion themes. A heatmap also shows darker representations between `r/Anxiety` and `r/PTSD` subreddits compared to topics in the `r/Depression` subreddit. Although the topics of the discussions were the same, our approach underline the need to focus on different issues pertaining to SLEEP and WORK with these conditions.

The prevalence of mentions of specific medications, treatments, and support resources also highlighted the differences in subreddits. The names of common medications were present among the top cluster keywords for the `r/Anxiety` and `r/PTSD` subreddits, and the `r/PTSD` clusters also included specific therapies and support resources, such as Eye Movement Desensitization and Reprocessing (EMDR) therapy and service animals. In contrast, cluster topics

ACCEPTED MANUSCRIPT

in the r/Depression subreddit focused more on contextual aspects of depressive episodes such as affect (negative emotion, positive emotion), interpersonal interactions (friends and family, talking to friends, understanding depressed individuals), and situations in which depressive symptoms may occur (birthdays, love and depression).

A number of overlapping members exist among the three subreddits. Only a small fraction of members participated all three subreddits (n=65), but a larger sum of members (n=2,037) participated in both the r/Anxiety and r/Depression subreddits. However, these members were discussing a variety of topics, not just the commonly shared topics. Given the small number of shared members between subreddits and the variety of topical interests expressed, we concluded that these members were not the main reason for the commonly shared discussion themes.

Taken together, these results suggest that the r/Anxiety and r/PTSD subreddit members are more likely to be individuals whose experiences with a condition are more long-term, and who are interested in treatments and medications. The r/Depression subreddit members may be a more diffuse group, some who may be dealing with long-term issues, but perhaps who are dealing with transient issues that cause depressed mood. This may also account for the larger size of the r/Depression subreddit. The word 'depression'perhaps has a larger set of connoted meanings, some clinical and others not; and thus, those who participate in this subreddit may be a more diffuse and transient group.

The contribution of this work is twofold: first, we illustrated the differences in the nature of online discussion from communities sharing similar symptoms and co-morbidity. Our findings inform more nuanced discussion themes and suggest researchers to employ multiple methods to fully understand the subtle differences. Second, from a practical perspective, understanding these subtle

ACCEPTED MANUSCRIPT

differences in the nature of online discussion could used to inform the design of online mental health communities and patient education programs for these conditions.

5. Limitations and Future Directions

This study has various limitations. First, this study employed data from one social networking site. As mentioned in the Introduction, Reddit is a widely used platform, but it is more frequently used by certain demographic segments, particularly by younger males [67, 68]. This bias toward a younger audience enabled us to identify particular areas of interest of members, such as school and work. However, in future studies it would also be useful to examine other online health communities that address these conditions to better characterize the needs of people who experience the conditions, but may not be represented in the Reddit community.

Second, the topic of online discussions is prone to change as the discussion progresses [69]. We expect many of the longer discussions (i.e., higher number of comments. note Table 1 on page 7) to have multiple topics, however, our method of analysis would only identify single topic for each of those discussions. Thus, different machine learning algorithms, such as latent Dirichlet allocation that can produce multiple topics for a single document, could produce different results. Moreover, misspellings, abbreviations, contractions, and community-specific nomenclatures are common in online health communities [70]. A high prevalence of these cases could alter the clustering result by changing the overall counts of important terms. However, we did not encounter these cases during our manual examination of the most frequently occurring terms for each cluster.

Third, the number of words that were used in to calculate thematic similarity could have influenced the rendering our visualizations. In our method, we pre-

ACCEPTED MANUSCRIPT

specified the use of the 20 most frequently occurring words of each cluster to determine the edge weight between nodes for the network visualization and to determine the proximity of discussion themes in the heatmap. If we had considered a larger number of words in the visualization processes, the overall visualizations could look different. For instance, the two outlier themes (e.g., SOCIAL from the `r/Depression` and GRATITUDE from the `r/PTSD`) could have more common words to other themes.

This difference in group composition provokes some interesting questions. First, does the difference in content suggest different usage intents on the part of the subreddit members? Second, if so, do subreddits fulfill the needs of these different types of members equally well? Short-lived participation is generally viewed as a challenge of managing online health communities, due to issues like lurking [71] and dropping out [72, 73]. If subreddit members visit to obtain a solution to a transient issue or simply to vent and move on, their needs might have been fulfilled but without much contribution to the community. Similarly, the informational and emotional support needs for those who are looking for more long-term solutions are different. How to design online health communities that can support both types of needs and members while sustaining the overall activities of the communities is an unanswered question.

Other than group composition, one might consider what features online health communities might provide to help users find content [74] or members [75, 76] that are important to them. Based on the results presented in this paper, it could be useful to provide interactive functionality for members to locate posts and other members [77] that discuss particular types of medications and treatments, but also to identify content based on contextual elements of experience, such as social occasions, the need for understanding, and so on. Also, considering the temporality of participants' experiences (e.g. long-term,

ACCEPTED MANUSCRIPT

transient, etc.) is paramount.

Though beyond the scope of this study, it would be interesting to consider how the informational and emotional support content of these communities compares to the content that is delivered in Internet-based interventions for anxiety, depression and PTSD (e.g. [78, 79]). Although potential overlaps of users between online health communities and Internet-based interventions may exist, it is unlikely that any of these avenues could reach the entire population who are suffering from these mental health conditions [80]. Thus, understanding what each of these different avenues can and cannot offer is important.

6. Conclusion

In this study, we compared online discussion content from three online mental health communities concerning conditions that similar symptoms and can potentially be co-morbid. More specifically, we collected data from Reddit, a highly popular social media platform, and analyze content from three subreddits focusing on anxiety, depression and PTSD. First, we employed cluster analysis to identify the top 15 discussion themes for each subreddit. Second, we combined text mining, visualization and qualitative analysis methods to identify thematic similarities and differences between the three subreddits. Through qualitative analysis, we observed that members of the three communities shared overlapping concerns (i.e., sleep- and work related problems) and discussion patterns (i.e., sharing of positive emotion and showing gratitude for receiving emotional support), but also exposed contextual variations in these themes among the three communities. By rendering a network visualization of the topics discussed and employing a community detection algorithm on this network, we illustrated discussions from the `r/Anxiety` and `r/PTSD` subreddits shared greater similarities to one another than to discussions from the `r/Depression` subreddit, and

ACCEPTED MANUSCRIPT

employed a heatmap to support closer examination of these similarities and differences. We also supported this finding by examining the shared members' participation and discussion. The findings from this study could be used to inform the design of online mental health communities and patient education programs for these conditions. Moreover, we suggest that researchers employ multiple methods to fully understand the subtle differences when comparing similar discussions from online health communities.

7. Acknowledgments

[Blank for blind review]

References

[1] C. E. Cutrona, J. A. Suhr, Social support communication in the context of marriage: An analysis of couples' supportive interactions., 1994.

[2] S. Fox, M. Duggan, Health online 2013: 35% of U.S. adults have gone online to figure out a medical condition, Pew Internet & American Life Project (2013) 1–55.
URL `http://www.pewinternet.org/2013/01/15/health-online-2013/;Archived_at:http://www.webcitation.org/6Y5K9a3JH`

[3] R. Gray, M. Fitch, C. Davis, C. Phillips, A qualitative study of breast cancer self-help groups, Psycho-Oncology 6 (4) (1997) 279–289.
URL `http://onlinelibrary.wiley.com/doi/10.1002/(SICI)1099-1611(199712)6:4<279::AID-PON280>3.0.CO;2-0/abstract`

[4] S. Fox, Peer-to-peer healthcare: Many people — espeically those living with chronic or rare diseases — use online connections to supplement professional medical advice, Pew Internet & American Life Project.

ACCEPTED MANUSCRIPT

URL http://www.pewinternet.org/2011/02/28/peer-to-peer-health-care-2/;Archived_at:http://www.webcitation.org/6Y5JjmduI

[5] M. Swan, Emerging patient-driven health care models: An examination of health social networks, consumer personalized medicine and quantified self-tracking, International Journal of Environmental Research and Public Health 6 (2) (2009) 492–525. doi:10.3390/ijerph6020492.

[6] H. S. Wentzer, A. Bygholm, Narratives of empowerment and compliance: Studies of communication in online patient support groups, International Journal of Medical Informatics 82 (12) (2013) e386–e394. doi:10.1016/j.ijmedinf.2013.01.008.
URL http://dx.doi.org/10.1016/j.ijmedinf.2013.01.008http://linkinghub.elsevier.com/retrieve/pii/S1386505613000191

[7] K. O. Hwang, A. J. Ottenbacher, A. P. Green, M. R. Cannon-Diehl, O. Richardson, E. V. Bernstam, E. J. Thomas, Social support in an Internet weight loss community, International Journal of Medical Informatics 79 (1) (2010) 5–13. doi:10.1016/j.ijmedinf.2009.10.003.
URL http://linkinghub.elsevier.com/retrieve/pii/S1386505609001725

[8] J. Finkelstein, O. Lapshin, Reducing depression stigma using a web-based program, International Journal of Medical Informatics 76 (10) (2007) 726–734. doi:10.1016/j.ijmedinf.2006.07.004.
URL http://linkinghub.elsevier.com/retrieve/pii/S1386505606001778

[9] A. Hartzler, W. Pratt, Managing the personal side of health: how patient expertise differs from the expertise of clinicians, Journal of Medical Internet

ACCEPTED MANUSCRIPT

research 13 (3) (2011) e62. doi:10.2196/jmir.1728.
URL http://www.ncbi.nlm.nih.gov/pubmed/21846635

[10] C. F. van Uden-Kraan, C. H. C. Drossaert, E. Taal, B. R. Shaw, E. R. Seydel, M. a. F. J. van de Laar, Empowering processes and outcomes of participation in online support groups for patients with breast cancer, arthritis, or fibromyalgia, Qualitative Health Research 18 (3) (2008) 405–417. doi:10.1177/1049732307313429.
URL http://www.ncbi.nlm.nih.gov/pubmed/18235163

[11] C. F. van Uden-Kraan, C. H. C. Drossaert, E. Taal, E. R. Seydel, M. a. F. J. van de Laar, Participation in online patient support groups endorses patients' empowerment, Patient Education and Counseling 74 (1) (2009) 61–69. doi:10.1016/j.pec.2008.07.044.
URL http://www.ncbi.nlm.nih.gov/pubmed/18778909

[12] Y. Setoyama, Y. Yamazaki, K. Namayama, Benefits of peer support in online Japanese breast cancer communities: differences between lurkers and posters, Journal of medical Internet research 13 (4) (2011) e122. doi:10.2196/jmir.1696.
URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3278108{&}tool=pmcentrez{&}rendertype=abstract

[13] Y. K. Bartlett, N. S. Coulson, An investigation into the empowerment effects of using online support groups and how this affects health professional/patient communication, Patient Education and Counseling 83 (1) (2011) 113–119. doi:10.1016/j.pec.2010.05.029.
URL http://www.ncbi.nlm.nih.gov/pubmed/20599338

[14] K. M. Griffiths, A. L. Calear, M. Banfield, Systematic review on Internet Support Groups (ISGs) and depression (1): Do ISGs reduce depressive

ACCEPTED MANUSCRIPT

symptoms?, Journal of Medical Internet research 11 (3) (2009) e40. doi:10.2196/jmir.1270.
URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2802256{&}tool=pmcentrez{&}rendertype=abstract

[15] M. T. Høybye, S. O. Dalton, I. Deltour, P. E. Bidstrup, K. Frederiksen, C. Johansen, Effect of Internet peer-support groups on psychosocial adjustment to cancer: a randomised study, British Journal of Cancer 102 (9) (2010) 1348–1354. doi:10.1038/sj.bjc.6605646.
URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2865756{&}tool=pmcentrez{&}rendertype=abstract

[16] B. Shaw, R. Hawkins, F. McTavish, Effects of insightful disclosure within computer mediated support groups on women with breast cancer, Health Communication 19 (2) (2006) 133–142. doi:10.1207/s15327027hc1902.
URL http://www.tandfonline.com/doi/abs/10.1207/s15327027hc1902{_}5

[17] P. K. H. Mo, N. S. Coulson, Developing a model for online support group use, empowering processes and psychosocial outcomes for individuals living with HIV/AIDS, Psychology & Health 27 (4) (2012) 445–459. doi:10.1080/08870446.2011.592981.
URL http://www.ncbi.nlm.nih.gov/pubmed/21854088

[18] M. De Choudhury, S. De, Mental health discourse on Reddit: self-disclosure, social support, and anonymity, in: Proceedings of ICWSM, AAAI, 2014, pp. 71–80.

[19] L. J. Barney, K. M. Griffiths, M. A. Banfield, Explicit and implicit information needs of people with depression: a qualitative investigation of problems reported on an online depression support forum, BMC Psychiatry

ACCEPTED MANUSCRIPT

11 (1) (2011) 88. doi:10.1186/1471-244X-11-88.
URL http://www.ncbi.nlm.nih.gov/pubmed/21592335$\delimiter"026E30F$nhttp://www.biomedcentral.com/content/pdf/1471-244X-11-88.pdfhttp://www.biomedcentral.com/1471-244X/11/88

[20] E. Hatfield, J. T. Cacioppo, R. L. Rapson, Emotional contagion, Current Directions in Psychological Science 2 (3) (1993) 96–99. doi:10.1111/1467-8721.ep10770953.
URL http://cdp.sagepub.com/lookup/doi/10.1111/1467-8721.ep10770953

[21] M. Kumar, M. Dredze, G. Coppersmith, M. D. Choudhury, Detecting changes in suicide content manifested in social media following celebrity suicides, in: Proceedings of the 26th ACM Conference on Hypertext & Social Media, ACM, 2015, pp. 85–94. doi:http://dx.doi.org/10.1145/2700171.2791026.

[22] M. Conway, D. O'Connor, Social media, big data, and mental health: current advances and ethical implications, Current Opinion in Psychology 9 (2016) 77–82. doi:10.1016/j.copsyc.2016.01.004.
URL http://www.sciencedirect.com/science/article/pii/S2352250X16000063

[23] P. A. Cavazos-Rehg, M. J. Krauss, S. Sowles, S. Connolly, C. Rosas, M. Bharadwaj, L. J. Bierut, A content analysis of depression-related tweets, Computers in Human Behavior 54 (2016) 351–357. doi:10.1016/j.chb.2015.08.023.
URL http://dx.doi.org/10.1016/j.chb.2015.08.023

[24] B. A. Primack, A. Shensa, C. G. Escobar-Viera, E. L. Barrett, J. E. Sidani,

ACCEPTED MANUSCRIPT

J. B. Colditz, A. E. James, Use of multiple social media platforms and symptoms of depression and anxiety: A nationally-representative study among U.S. young adults, Computers in Human Behavior 69 (2017) 1–9. doi:10.1016/j.chb.2016.11.013.
URL http://linkinghub.elsevier.com/retrieve/pii/S0747563216307543

[25] H. Dai, J. Hao, Mining social media data on marijuana use for Post Traumatic Stress Disorder, Computers in Human Behaviordoi:10.1016/j.chb.2016.12.064.
URL http://linkinghub.elsevier.com/retrieve/pii/S0747563216308949

[26] M. De Choudhury, M. Gamon, S. Counts, E. Horvitz, Predicting depression via social media, in: Proceedings of the 7th International AAAI Conference on Weblogs and Social Media (ICWSM), 2013.
URL http://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/viewFile/6124/6351

[27] A. Park, M. Conway, Longitudinal Changes in Psychological States in Online Health Community Members: Understanding the Long-Term Effects of Participating in an Online Depression Community., Journal of medical Internet research 19 (3) (2017) e71. doi:10.2196/jmir.6826.
URL http://www.ncbi.nlm.nih.gov/pubmed/28320692http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5379019

[28] D. Mowery, A. Park, M. Conway, C. Bryan, Towards automatically classifying depressive symptoms from Twitter data for population health., in: Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media, Osaka, Japan, 2016, pp. 182–191.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

[29] A. Tamersoy, M. D. Choudhury, D. H. Chau, Characterizing smoking and drinking abstinence from social media, in: Proceedings of the 26th ACM Conference on Hypertext & Social Media, ACM, 2015, pp. 139–148. doi:http://dx.doi.org/10.1145/2700171.2791247.

[30] M. Myslín, S.-H. Zhu, W. Chapman, M. Conway, Using Twitter to examine smoking behavior and perceptions of emerging tobacco products, Journal of Medical Internet Research 15 (8) (2013) e174. doi:10.2196/jmir.2534.
URL http://www.jmir.org/2013/8/e174/http://www.ncbi.nlm.nih.gov/pubmed/23989137http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3758063

[31] A. Park, M. Conway, Towards Tracking Opium Related Discussions in Social Media, Online Journal of Public Health Informatics 9 (1) (2017) e73. doi:10.5210/ojphi.v9i1.7652.
URL http://ojphi.org/ojs/index.php/ojphi/article/view/7652http://journals.uic.edu/ojs/index.php/ojphi/article/view/7652https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5461054

[32] D. Maclean, S. Gupta, A. Lembke, C. Manning, J. Heer, Forum77 : an analysis of an online health forum dedicated to addiction recovery, in: Proceedings of the companion publication of the 18th ACM conference on Computer supported cooperative work & social computing (CSCW), 2015.

[33] D. Watson, K. Weber, J. S. Assenheimer, L. A. Clark, E. al, Testing a tripartite model: I. Evaluating the convergent and discriminant validity of anxiety and depression symptom scales, Journal of Abnormal Psychology 104 (1) (1995) 3–14. doi:10.1037/0021-843X.104.1.3.
URL http://www.ncbi.nlm.nih.gov/pubmed/7897050http://doi.apa.org/getdoi.cfm?doi=10.1037/0021-843X.104.1.3

ACCEPTED MANUSCRIPT

[34] K. Ginzburg, T. Ein-Dor, Z. Solomon, Comorbidity of posttraumatic stress disorder, anxiety and depression: A 20-year longitudinal study of war veterans, Journal of Affective Disorders 123 (1-3) (2010) 249–257. doi:10.1016/j.jad.2009.08.006.
URL http://linkinghub.elsevier.com/retrieve/pii/S0165032709003723

[35] K. Scott, R. Bruffaerts, A. Tsang, J. Ormel, J. Alonso, M. Angermeyer, C. Benjet, E. Bromet, G. de Girolamo, R. de Graaf, I. Gasquet, O. Gureje, J. Haro, Y. He, R. Kessler, D. Levinson, Z. Mneimneh, M. Oakley Browne, J. Posada-Villa, D. Stein, T. Takeshima, M. Von Korff, Depressionanxiety relationships with chronic physical conditions: Results from the World Mental Health surveys, Journal of Affective Disorders 103 (1-3) (2007) 113–120. doi:10.1016/j.jad.2007.01.015.
URL http://linkinghub.elsevier.com/retrieve/pii/S0165032707000171

[36] A. T. Chen, What's in a virtual hug? A transdisciplinary review of methods in online health discussion forum research, Library & Information Science Research 36 (2) (2014) 120–130. doi:10.1016/j.lisr.2013.10.007.
URL http://linkinghub.elsevier.com/retrieve/pii/S0740818814000309

[37] P. R. Morrow, Telling about problems and giving advice in an Internet discussion forum: some discourse features, Discourse Studies 8 (4) (2006) 531–548. doi:10.1177/1461445606061876.
URL http://dis.sagepub.com/cgi/doi/10.1177/1461445606061876

[38] R. Savolainen, Requesting and providing information in blogs and Internet discussion forums, Journal of Documentation 67 (5) (2011) 863–886.

ACCEPTED MANUSCRIPT

doi:10.1108/00220411111164718.
URL http://www.emeraldinsight.com/doi/abs/10.1108/00220411111164718

[39] Reddit, Reddit in 2015 (2015).
URL http://www.redditblog.com/2015/12/reddit-in-2015.html;Archived_at:http://www.webcitation.org/6eTHN0TFD

[40] Reddit, Reddit API Documentation (2015).
URL https://www.reddit.com/dev/api;Archived_at:http://www.webcitation.org/6dGU2ksOW

[41] Reddit, PRAW: The Python Reddit Api Wrapper (2014).
URL https://praw.readthedocs.org/en/latest/index.html;Archived_at:http://www.webcitation.org/6dGUF2mNR

[42] Reddit, Frequently Asked Questions.
URL https://www.reddit.com/wiki/faq;Archived_at:http://www.webcitation.org/6dGUyQNkU

[43] Reddit_Members, ELI5: difference between "best", "hot", and "top" on Reddit (2013).
URL https://www.reddit.com/r/explainlikeimfive/comments/1u0q4s/eli5{_}difference{_}between{_}best{_}hot{_}and{_}top{_}on{_}reddit

[44] A. Bruckman, Teaching students to study online communities ethically, Journal of Information Ethics 15 (2) (2006) 82–98.

[45] G. Eysenbach, J. E. Till, Ethical issues in qualitative research on internet communities, BMJ 323 (7321) (2001) 1103–1105. doi:10.1136/bmj.323.7321.1103.
URL http://www.bmj.com/cgi/doi/10.1136/bmj.323.7321.1103

ACCEPTED MANUSCRIPT

[46] B. Humphreys, D. Lindberg, H. Schoolman, The Unified Medical Language System: an informatics research collaboration, Journal of the American Medical Informatics Association 5 (1).
URL http://www.ncbi.nlm.nih.gov/pubmed/9452981

[47] C. D. Manning, P. Raghavan, H. Schütze, Introduction to information retrieval, Cambridge: Cambridge university press, 2008.

[48] A. T. Chen, Exploring online support spaces: Using cluster analysis to examine breast cancer, diabetes and fibromyalgia support groups, Patient Education and Counseling 87 (2) (2012) 250–257. doi:10.1016/j.pec.2011.08.017.
URL http://linkinghub.elsevier.com/retrieve/pii/S073839911100468X

[49] S. Bird, NLTK: the natural language toolkit., in: Proceedings of the COLING/ACL on Interactive presentation sessions,)Association for Computational Linguistics., 2006, pp. 69–72.

[50] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research 12 (2011) 2825–2830. arXiv:1201.0490.
URL http://arxiv.org/abs/1201.0490

[51] A. L. Strauss, J. M. Corbin, Basics of qualitative research, Sage Publications, Newbury Park, CA, 1990.

[52] M. Bostock, V. Ogievetsky, J. Heer, D: Data-Driven Documents., IEEE transactions on visualization and computer graphics 17 (12) (2011) 2301–

ACCEPTED MANUSCRIPT

9. arXiv:1408.4626, doi:10.1109/TVCG.2011.185.
URL http://www.ncbi.nlm.nih.gov/pubmed/22034350

[53] M. Bastian, S. Heymann, M. Jacomy, Gephi: An open source software for exploring and manipulating networks.
URL http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154

[54] M. Jacomy, T. Venturini, S. Heymann, M. Bastian, ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software, PLOS ONE 9 (6) (2014) e98679. doi:10.1371/journal.pone.0098679.
URL http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0098679

[55] A. Huang, Similarity measures for text document clustering, in: Proceedings of the sixth new zealand computer science research student conference (NZCSRSC2008), no. April, 2008, pp. 49–56.
URL http://nzcsrsc08.canterbury.ac.nz/site/proceedings/Individual{_}Papers/pg049{_}Similarity{_}Measures{_}for{_}Text{_}Document{_}Clustering.pdf

[56] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks, Journal of Statistical Mechanics: Theory and Experiment 2008 (10) (2008) P10008. arXiv:0803.0476, doi:10.1088/1742-5468/2008/10/P10008.
URL http://arxiv.org/abs/0803.0476http://stacks.iop.org/1742-5468/2008/i=10/a=P10008?key=crossref.46968f6ec61eb8f907a760be1c5ace52

[57] M. E. J. Newman, M. Girvan, Finding and evaluating community structure

ACCEPTED MANUSCRIPT

in networks, Physical Review E 69 (2) (2004) 026113. arXiv:0308217, doi:10.1103/PhysRevE.69.026113.
URL http://link.aps.org/doi/10.1103/PhysRevE.69.026113

[58] B. Qiu, K. Zhao, P. Mitra, D. Wu, C. Caragea, J. Yen, G. E. Greer, K. Portier, Get online support, feel better – sentiment analysis and dynamics in an online cancer survivor community, in: 2011 IEEE Third Int'l Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third Int'l Conference on Social Computing, IEEE, 2011, pp. 274–281. doi:10.1109/PASSAT/SocialCom.2011.127.
URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6113125

[59] B. L. Fredrickson, R. W. Levenson, Positive emotions speed recovery from the cardiovascular sequelae of negative emotions, Cognition and Emotion 12 (2) (1998) 191–220. doi:10.1080/026999398379718.
URL http://www.tandfonline.com/doi/abs/10.1080/026999398379718

[60] J. Wolpe, Psychotherapy by reciprocal inhibition, A Pavlovian Journal of Research & Therapy 3 (4) (1968) 234–240. arXiv:PrintISBN:1053-881X;OnlineISSN:1936-3567, doi:10.1007/BF03000093.

[61] S. Folkman, Positive psychological states and coping with severe stress, Social Science & Medicine 45 (8) (1997) 1207–1221. doi:10.1016/S0277-9536(97)00040-3.
URL http://linkinghub.elsevier.com/retrieve/pii/S0277953697000403

[62] S. Folkman, J. T. Moskowitz, Positive affect and the other side of coping., American Psychologist 55 (6) (2000) 647–654. doi:10.1037/0003-066X.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

55.6.647.
URL http://doi.apa.org/getdoi.cfm?doi=10.1037/0003-066X.55.6.647

[63] A. M. Wood, J. Maltby, R. Gillett, P. A. Linley, S. Joseph, The role of gratitude in the development of social support, stress, and depression: two longitudinal studies, Journal of Research in Personality 42 (4) (2008) 854–871. doi:10.1016/j.jrp.2007.11.003.
URL http://linkinghub.elsevier.com/retrieve/pii/S0092656607001286

[64] R. A. Emmons, M. E. McCullough, Counting blessings versus burdens: An experimental investigation of gratitude and subjective well-being in daily life., Journal of Personality and Social Psychology 84 (2) (2003) 377–389. doi:10.1037/0022-3514.84.2.377.
URL http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.84.2.377

[65] American Psychiatric Association, American Psychiatric Association. DSM-5 Task Force, Diagnostic and statistical manual of mental disorders: DSM-5., 5th Edition, American Psychiatric Association, Washington, DC, 2013.

[66] W. R. Pigeon, A. M. Gallegos, Posttraumatic stress disorder and sleep, Sleep Medicine Clinics 10 (1) (2015) 41–48. doi:10.1016/j.jsmc.2014.11.010.
URL http://linkinghub.elsevier.com/retrieve/pii/S088985291300039Xhttp://linkinghub.elsevier.com/retrieve/pii/S1556407X14001271

[67] T. Bogers, R. Wernersen, How 'social' are social news sites? exploring

ACCEPTED MANUSCRIPT

the motivations for using Reddit.com, in: iConference 2014 Proceedings, iSchools, 2014, pp. 329–344. doi:10.9776/14108.
URL https://www.ideals.illinois.edu/handle/2142/47295

[68] M. Duggan, A. Smith, 6% of online adults are Reddit users, Pew Internet & American Life Project 3.
URL pewinternet.org/Reports/2013/reddit.aspxhttp://www.webcitation.org/6fEEqzeJp

[69] A. Park, A. L. Hartzler, J. Huh, G. Hsieh, D. W. McDonald, W. Pratt, "How Did We Get Here?": Topic Drift in Online Health Discussions., Journal of medical Internet research 18 (11) (2016) e284. doi:10.2196/jmir.6297.
URL http://www.jmir.org/2016/11/e284/http://www.ncbi.nlm.nih.gov/pubmed/27806924http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5112365

[70] A. Park, A. L. Hartzler, J. Huh, D. W. McDonald, W. Pratt, Automatically Detecting Failures in Natural Language Processing Tools for Online Community Text., Journal of medical Internet research 17 (8) (2015) e212. doi:10.2196/jmir.4612.
URL http://www.jmir.org/2015/8/e212/http://www.ncbi.nlm.nih.gov/pubmed/26323337http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4642409

[71] N. Sun, P. P.-L. Rau, L. Ma, Understanding lurkers in online communities: A literature review, Computers in Human Behavior 38 (2014) 110–117. doi:10.1016/j.chb.2014.05.022.
URL http://dx.doi.org/10.1016/j.chb.2014.05.022http://linkinghub.elsevier.com/retrieve/pii/S0747563214003008

ACCEPTED MANUSCRIPT

[72] A. Park, A. L. Hartzler, J. Huh, D. W. McDonald, W. Pratt, Homophily of Vocabulary Usage: Beneficial Effects of Vocabulary Similarity on Online Health Communities Participation., AMIA ... Annual Symposium proceedings. AMIA Symposium 2015 (2015) 1024–33.
URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4765708/http://www.ncbi.nlm.nih.gov/pubmed/26958240http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4765708

[73] M. Balestra, C. Cheshire, O. Arazy, O. Nov, Investigating the Motivational Paths of Peer Production Newcomers, in: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17, no. May, ACM Press, New York, New York, USA, 2017, pp. 6381–6385. doi:10.1145/3025453.3026057.
URL http://delivery.acm.org/10.1145/3030000/3026057/p6381-balestra.pdf?ip=131.181.158.7{&}id=3026057{&}acc=OPEN{&}key=65D80644F295BC0D.CE8691788DF0BE02.4D4702B0C3E38B35.6D218144511F3437{&}CFID=761663822{&}CFTOKEN=94546269{&}{_}{_}acm{_}{_}=1494560435{_}3174bfd3ce6767e4b86cbahttp://dl.acm.org/citation.cfm?doid=3025453.3026057

[74] M. S. Ackerman, T. W. Malone, Answer Garden: a tool for growing organizational memory, in: Proceedings of the conference on Office information systems -, Vol. 11, ACM Press, New York, New York, USA, 1990, pp. 31–39. doi:10.1145/91474.91485.
URL http://portal.acm.org/citation.cfm?doid=91474.91485

[75] A. L. Hartzler, M. N. Taylor, A. Park, T. Griffiths, U. Backonja, D. W. McDonald, S. Wahbeh, C. Brown, W. Pratt, Leveraging cues from person-generated health data for peer matching in online communities.,

ACCEPTED MANUSCRIPT

Journal of the American Medical Informatics Association : JAMIA 23 (3) (2016) 496–507. doi:10.1093/jamia/ocv175.
URL http://jamia.oxfordjournals.org/lookup/doi/10.1093/jamia/ocv175http://www.ncbi.nlm.nih.gov/pubmed/26911825http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4901377

[76] A. L. Hartzler, D. W. McDonald, A. Park, J. Huh, C. Weaver, W. Pratt, Evaluating health interest profiles extracted from patient-generated data., in: AMIA ... Annual Symposium proceedings. AMIA Symposium, Vol. 2014, 2014, pp. 626–35.
URL http://www.ncbi.nlm.nih.gov/pubmed/25954368http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4419913

[77] A. L. Hartzler, B. Weis, C. Cahill, W. Pratt, A. Park, U. Backonja, D. W. McDonald, Design and Usability of Interactive User Profiles for Online Health Communities, ACM Transactions on Computer-Human Interaction 23 (3) (2016) 1–33. doi:10.1145/2903718.
URL http://dl.acm.org/citation.cfm?doid=2952594.2903718

[78] V. Spek, P. Cuijpers, I. Nyklicek, H. Riper, J. Keyzer, V. Pop, Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis, Psychol Med 37 (3) (2007) 319–328. doi:10.1017/S0033291706008944.
URL http://www.ncbi.nlm.nih.gov/pubmed/17112400

[79] C. Knaevelsrud, A. Maercker, Internet-based treatment for PTSD reduces distress and facilitates the development of a strong therapeutic alliance: a randomized controlled clinical trial., BMC psychiatry 7 (2007) 13. doi:10.1186/1471-244X-7-13.
URL http://www.ncbi.nlm.nih.gov/pubmed/17442125

ACCEPTED MANUSCRIPT

[80] P. Corrigan, How stigma interferes with mental health care, American Psychologist 59 (7) (2004) 614–625. doi:10.1037/0003-066X.59.7.614. URL http://doi.apa.org/getdoi.cfm?doi=10.1037/0003-066X.59.7.614

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Appendices

A. All labels used to describe 10 k-means clustering results for Anxiety Disorder subreddit

Cluster label	Occurrence of labels in 10 k-means	Overlapping vocabularies (in %)
medication	10	89.42
misc.	10	41.40
panic attack	10	94.18
sleep	10	88.13
therapy/therapist	10	80.83
work	10	90.93
living with anxiety	9	74.94
positive emotion	9	71.39
social anxiety	9	94.28
how others think of you	8	62.76
want/need help	8	65.64
what do you think	8	77.79
congratulation	7	84.57
school	7	70.48
anxious feeling	6	68.13
heart attack and panic attack	3	82.67
gratitude	2	90.00
medication and symptoms	2	66.00
symptoms of anxiety disorder	2	80.00
animal	1	
anxiety and relationship	1	
anxiety symptoms	1	
depression	1	
different medication experiences	1	
do I have anxiety	1	
driving	1	
mental disorder	1	
need help	1	
school and driving	1	

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

B. All labels used to describe 10 k-means clustering results for the `r/Depression` *subreddit*

Cluster label	Occurrence of labels in 10 kmeans	Overlapping vocabularies (in %)
misc.	10	18.48
birthday	10	78.93
need for depression	10	65.29
school	10	88.13
sleep	10	99.60
talking to friends	10	78.58
depressed times	9	64.61
gratitude	9	80.61
feelings of depression	8	60.86
love and depression	8	78.14
work	8	84.86
positive emotion	7	52.10
understanding depressed individual	7	80.86
congratulation	5	80.60
games	5	72.80
friends and family	4	68.00
loss and depression	4	83.67
negative emotion	4	62.67
suicide	4	89.00
medication	2	48.00
reddit	2	92.00
music	1	
talking	1	
social	1	
weather	1	

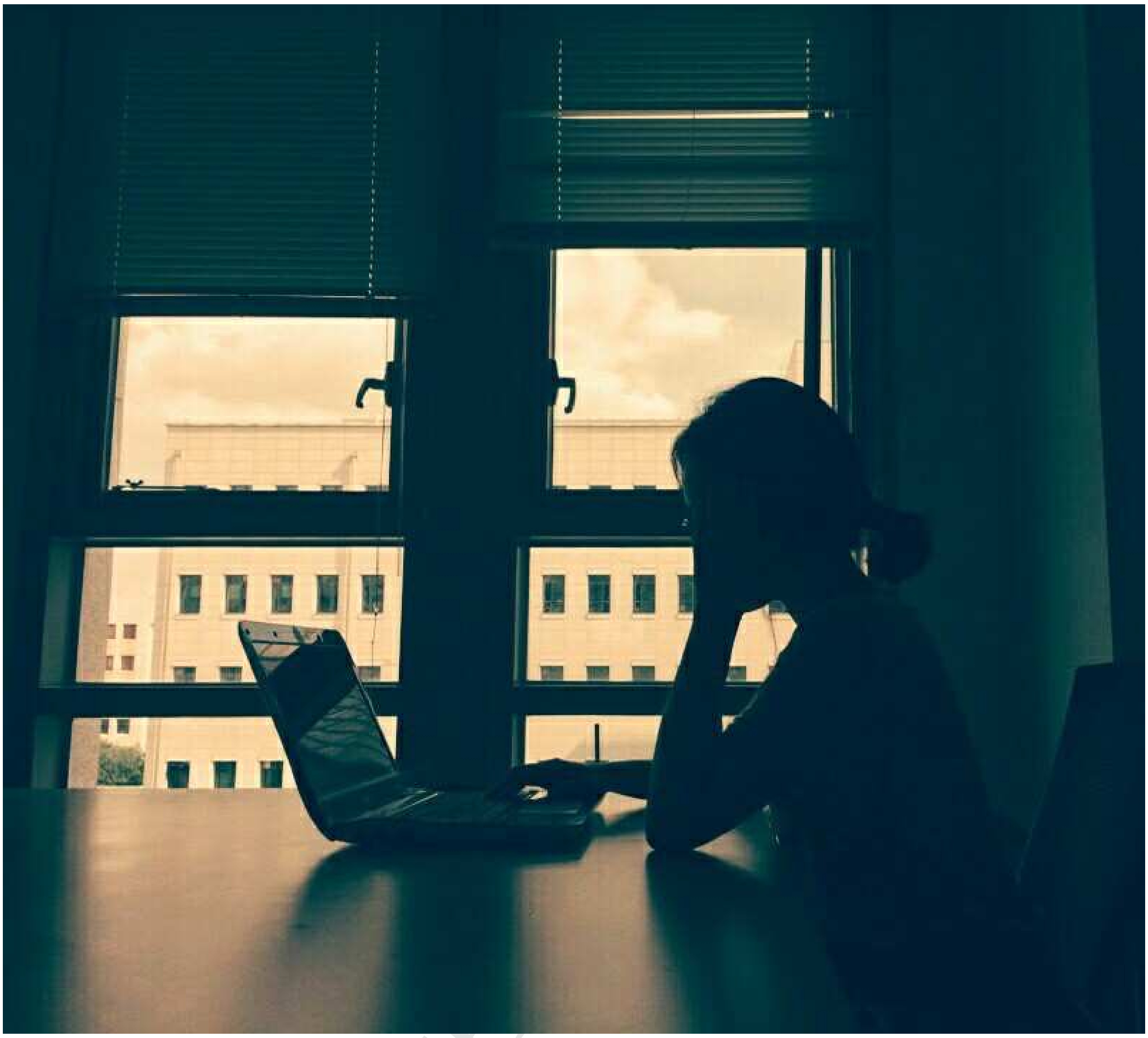
ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

1020 *C. All labels used to describe 10 k-means clustering results for the* `r/PTSD` *sub-*
1021 *reddit*

Cluster label	Occurrence of labels in 10 kmeans	Overlapping vocabularies (in %)
misc.	21	20.31
nightmare	10	86.80
animal	10	77.82
EMDR therapy	10	78.40
wanting to help	10	73.96
work	9	72.06
gratitude for sharing (techniques and stories)	9	52.89
trauma therapy	8	68.86
living with PTSD	8	73.43
help for PTSD	8	84.71
positive emotion	7	69.71
memory	7	65.62
wanting to talk	6	68.00
trauma trigger	5	62.00
therapy	4	71.33
sleep	2	80.00
symptoms and treatment	2	42.00
getting better	2	72.00
anxiety	2	66.00
diagnosis	2	38.00
research	1	
anger	1	
wanting to understand	1	
everyday issues	1	
military	1	
driving anxiety	1	
sexual	1	
doctor	1	

ACCEPTED MANUSCRIPT



ACCEPTED MANUSCRIPT

AP's contribution to this research was supported by the National Library of Medicine of the National Institutes of Health under award number T15 LM007124.

MC's contribution to this research was supported by the National Library of Medicine of the National Institutes of Health under award numbers R00LM011393 & K99LM011393.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

- Compares the nature of online discussion from online mental health communities.
- Identifies the common themes as well as the contextual variations in common themes.
- Highlights the differences in participations and discussion styles.

ACCEPTED MANUSCRIPT