



OPEN ACCESS

EDITED BY
Kirk Roberts,
University of Texas Health Science Center at
Houston, United States

REVIEWED BY
Susan McRoy,
University of Wisconsin–Milwaukee,
United States
Tavleen Singh,
University of Texas Health Science Center at
Houston, United States

*CORRESPONDENCE
Tatsawan Timakum
✉ tatsawan_tim@cmru.ac.th

RECEIVED 20 June 2023
ACCEPTED 17 October 2023
PUBLISHED 03 November 2023

CITATION
Timakum T, Xie Q and Lee S (2023) Identifying
mental health discussion topic in social media
community: subreddit of bipolar disorder
analysis. *Front. Res. Metr. Anal.* 8:1243407.
doi: 10.3389/frma.2023.1243407

COPYRIGHT
© 2023 Timakum, Xie and Lee. This is an
open-access article distributed under the terms
of the [Creative Commons Attribution License
\(CC BY\)](#). The use, distribution or reproduction
in other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted which
does not comply with these terms.

Identifying mental health discussion topic in social media community: subreddit of bipolar disorder analysis

Tatsawan Timakum ^{1*}, Qing Xie ² and Soobin Lee ³

¹Department of Information Science, Chiang Mai Rajabhat University, Chiang Mai, Thailand, ²School of Management, Shenzhen Polytechnic, Shenzhen, Guangdong, China, ³Department of Library and Information Science, Yonsei University, Seoul, Republic of Korea

Online platforms allow individuals to connect with others, share experiences, and find communities with similar interests, providing a sense of belonging and reducing feelings of isolation. Numerous previous studies examined the content of online health communities to gain insights into the sentiments surrounding mental health conditions. However, there is a noticeable gap in the research landscape, as no study has specifically concentrated on conducting an in-depth analysis or providing a comprehensive visualization of Bipolar disorder. Therefore, this study aimed to address this gap by examining the Bipolar subreddit online community, where we collected 1,460,447 posts as plain text documents for analysis. By employing LDA topic modeling and sentiment analysis, we found that the Bipolar disorder online community on Reddit discussed various aspects of the condition, including symptoms, mood swings, diagnosis, and medication. Users shared personal experiences, challenges, and coping strategies, seeking support and connection. Discussions related to therapy and medication were prevalent, emphasizing the importance of finding suitable therapists and managing medication side effects. The online community serves as a platform for seeking help, advice, and information, highlighting the role of social support in managing bipolar disorder. This study enhances our understanding of individuals living with bipolar disorder and provides valuable insights and feedback for researchers developing mental health interventions.

KEYWORDS

bipolar, mental disorder, sentiment analysis, topic modeling, network analysis, social media data

Introduction

With the rapid increase in the number of social media users, social media has become a powerful tool for sharing medical information, gathering user feedback, and establishing support networks (Moorhead et al., 2013). Individuals with severe mental health problems can challenge the stigma associated with their conditions through individual empowerment and by providing hope to others through online communities. They benefit greatly from interacting with peers online, sharing personal experiences, and strategies for living with mental illness (Naslund et al., 2016). Furthermore, social media communities provide researchers with a unique opportunity to learn about patients' health experiences, treatment preferences, and potentially discover new knowledge in the field of health science. Interactions and shared information on social media offer valuable data that can shed light on the impact of drugs, diseases, and medical treatments on patients outside controlled settings.

Previous analyses compiled information on the utilization of social media for various health-related purposes, including health interventions, health promotion initiatives, medical education, and monitoring disease outbreaks (Zhao and Zhang, 2017; Marar et al., 2019; Chen and Wang, 2021). Major social media platforms like Twitter, Facebook, and Reddit have become significant sources of online health information, including mental health-related content (Zhang et al., 2013; Pershad et al., 2018; Record et al., 2018; Foufi et al., 2019). Moreover, these platforms often provide users with natural means of self-expression, including discussions about their behavior, thoughts, and emotions, which can be indicative of their emotional wellbeing (Conway and O'Connor, 2016). This abundance of user-generated content also serves as a valuable data source for researchers. Studies in the field of data mining have employed machine learning and statistical methods to analyze online messages, focusing on mood, psycholinguistic processes, and content topics extracted from posts (Nguyen et al., 2014). Patients are increasingly turning to social media platforms to search for treatments, discuss their experiences with side effects and treatments with healthcare professionals, participate in health-related online forums, and gain further knowledge about their illnesses (Monnier et al., 2002). User-generated content from social media has proven to be a rich resource for health investigations (O'Neill et al., 2014). Users' posts reflect their thoughts and feelings about their medical experiences and often attract the attention of other patients, caregivers, and medical professionals. User-generated content on the web has become a novel source of clinical data for research purposes (Denecke, 2015).

In the era of Big Data and evidence-based research, social media data offers researchers vast opportunities to gather healthcare information, including insights into patients' health behavior and medical experiences. By leveraging interdisciplinary approaches spanning social sciences, information and computer science, and applied statistics, large datasets from social media content have been integrated into biomedical and psychological research endeavors. Collaborative efforts in Big Data science can unveil and explain patterns in psychology, human behavior, cognition, and their impact on sociocultural systems over time, ultimately generating valuable data, such as medical information (Harlow and Oswald, 2016).

Previous studies utilized computer software and social media content analysis techniques to achieve various outcomes. For example, researchers collected depression and schizophrenia-related hashtags on Twitter to identify stigmatizing attitudes and personal experiences (Reavley and Pilkington, 2014), and analyzed online community posts related to mental health issues to understand the characteristics of online depression communities in terms of language styles, topics, and sentiments (Nguyen et al., 2014). Additionally, psycholinguistics was employed as a feature for text analysis to detect emotions in online communication, recognize psychological status, and predict mental conditions. For instance, a study focused on depression analysis on Facebook using machine learning and a set of psycholinguistic features to investigate the effects of depression and provide solutions for mental health problems (Islam et al., 2018). Furthermore, researchers detected posts containing words or phrases consistent with suicidal ideation on Twitter to

observe the expression of suicidality based on linguistic patterns (O'Dea et al., 2015) and utilized linguistic metadata features, including specific words, parts of speech (POS), and lexicon-based elements from the Linguistic Inquiry and Word Count (LIWC), to detect depression in Reddit posts (Trotzek et al., 2018).

Similarly, social networks have provided insights into various aspects of bipolar disorder (BD), such as functioning and social skills. For example, linguistic and phonological features were applied to detect the early stages of BD from Twitter using a supervised machine-learning approach (Huang et al., 2017). Pathway analysis was employed to identify predictors of suicide ideation in older adults with BD, using socio-demographically targeted and social media advertising within Facebook and newsfeeds (O'Rourke et al., 2017). Furthermore, language impairment in Reddit was analyzed to study online mental health communities focusing on depression, BD, and schizophrenia, examining reports of positive emotion, exercising, and weight management (Park and Conway, 2018).

Numerous previous studies have examined the content of online health communities to gain insights into the sentiments surrounding mental health conditions. However, there is a noticeable gap in the research landscape, as no study has specifically concentrated on conducting an in-depth analysis or providing a comprehensive visualization of Bipolar disorder. Several researches demonstrated the association between language usage and mental disorders. Consequently, many researchers have concentrated on social media linguistics analysis and developed tools to detect language patterns that support clinical mental health care, such as systems for emotion detection from user-generated text. However, there remains a scarcity of research focusing on content and sentiment analysis.

Therefore, this study aims to explore the topics and sentiments related to mental health, specifically bipolar disorder (BD), by examining social media texts, in order to answer the following research questions:

- RQ1. What are the main topics and discussions related to bipolar disorder within the Bipolar subreddit?
- RQ2. What sentiments are expressed by users in the Bipolar subreddit when discussing their experiences with bipolar disorder?
- RQ3. How does the Bipolar subreddit serve as a platform for seeking help, advice, and information related to bipolar disorder?
- RQ4. What role does social support play in managing bipolar disorder within the Bipolar subreddit?

The findings of this study can contribute to a better understanding of individuals living with BD and offer valuable insights and feedback for researchers developing health interventions. Social media analytics can be used to track the impact and reach of health interventions in real time, allowing for the refinement and adaptation of interventions to maximize their effectiveness. Furthermore, the research framework of this study can be applied to the study of other health conditions as well.

Materials and methods

This study aimed to examine social media discussions within Bipolar subreddit communities. The experiment utilized topic modeling and sentiment analysis to identify hidden topics and observe polarity in social media data, respectively.

Dataset

The data used for this study was collected from the Reddit community (www.reddit.com). A finite set of two subreddits related to bipolar disorder, including <https://www.reddit.com/r/bipolar> and <https://www.reddit.com/r/bipolar2>, was selected for analysis. The subreddit “r/bipolar” community, dedicated to bipolar-related issues, has over 85.5 k members, while the subreddit “r/bipolar2” has 13.5 k members who live with bipolar disorder type 2 and encompass the entire bipolar spectrum.

The posts and comments were collected from November 26th, 2018, and ended on December 28th, 2018. During this time period, a total of 1,460,447 posts were collected as plain text documents by scraping Reddit using Python Reddit API Wrapper (PRAW). Posts consist of the initial textual statements that initiate communication with other users, while comments are replies to these posts, organized in a tree-like structure (Gkotsis et al., 2016) as shown in Figure 1.

Reddit is one of the largest and most active online communities, where users can engage in discussions and share their experiences on a wide range of topics, including mental health. It boasts a large community of members, and many of them have an extensive history of previous submissions. The platform also hosts substantial content related to various diseases and medical conditions. The language of Reddit text posts is more structured than that on other major social media platforms, such as Twitter. It provides a unique environment for studying online communities due to its structure of subreddits, which are individual, topic-specific communities within the larger platform. While there may be multiple subreddits related to bipolar disorder on Reddit. We selected the “r/bipolar” and “r/bipolar2” subreddits based on factors such as their popularity, activity level, and relevance to the research objectives. Both subreddits are dedicated to in-depth discussions specifically related to bipolar disorder, making it a suitable and focused community to investigate the topics and sentiments related to this mental health condition. Furthermore, Reddit’s terms and conditions allow the use of its content for research purposes, which constitutes a major advantage for researchers. The Reddit forum also tends to be distinct from similar offline groups. Users are more likely to discuss problems that they do not feel comfortable talking about in face-to-face interactions (Johnson and Ambrose, 2006).

The data used in this study was collected from Reddit, a social media platform, in accordance with the platform’s terms of service and community guidelines. The data collection process strictly adhered to the ethical standards outlined by Reddit, ensuring respect for user privacy, anonymity, and the responsible use of the platform’s content. Consequently, we confirm that the collection, analysis, and reuse of social media data were conducted in strict

accordance with Reddit’s policies and terms of use, as well as all relevant institutional regulations.

Data processing

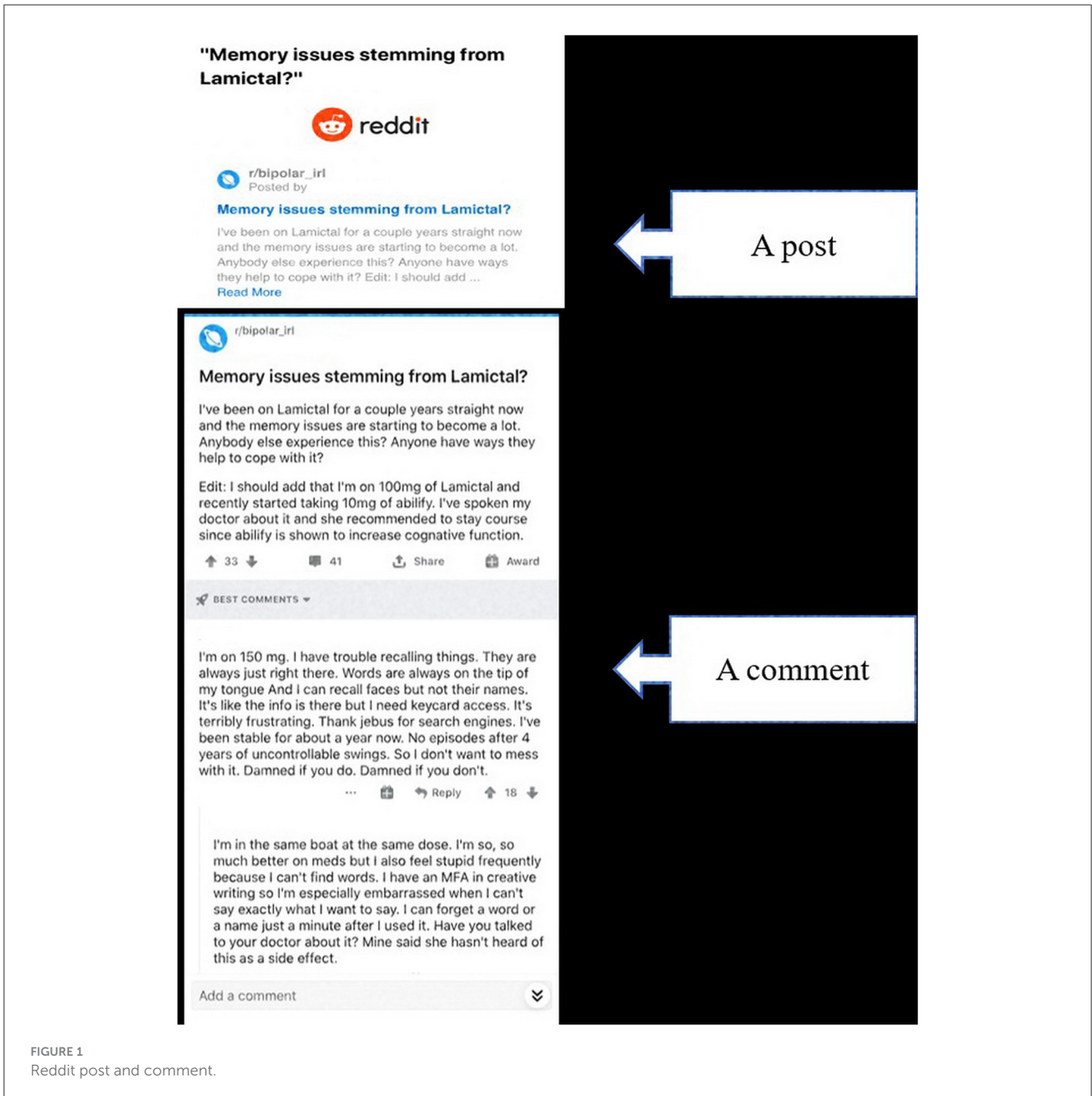
Pre-processing

Before analyzing the dataset, it underwent a pre-processing process to ensure the quality of the data prior to conducting specific analyses. This process included sentence segmentation, which involved splitting the sentences based on full-stop delimiters using the OpenNLP Sentence Detector (Apache OpenNLP Development Community, 2011). Additionally, stop word removal eliminated words that do not convey meaningful information, such as “a,” “is,” and “are.” Furthermore, Stanford CoreNLP (Manning et al., 2014) was utilized for part-of-speech tagging (POS) to identify the word types (noun, verb, adverb, and adjective) in the dataset based on tokens and context. This technique proved valuable for examining complex words and accurately determining the meaning of words within sentences. Furthermore, lemmatization was employed to analyze the lemma of each term. This step reduces inflection and derivationally related forms of words, returning them to their standard form (Manning et al., 2009; Song et al., 2015).

Topic analysis

The topic modeling approach was employed for content analysis to summarize the topics of bipolar disorder (BD) content in Reddit. Topic modeling is a probabilistic generative model based on the assumption that documents contain a mixture of topics consisting of the highest probability for each word. This statistical model helps identify abstract topics that occur in collections of documents. A topic represents a cluster of words that frequently appear together. In this study, we utilized Latent Dirichlet Allocation (LDA) for topic analysis, an unsupervised model that automatically clusters words into topics and associates documents with those topics. The underlying assumption is that a text document has a probability distribution over a mixture of “topics,” where each topic is associated with a distribution over words, and each word is drawn from the mixture (Blei et al., 2003).

LDA topic modeling has been a well-established and widely used technique in natural language processing and topic modeling for many years. Consequently, the availability of numerous open-source implementations, libraries, and resources for LDA makes it relatively easier to integrate into existing workflows. LDA is a relatively lightweight algorithm that can be trained efficiently on smaller datasets. LDA might be a more practical choice for tasks with limited training time and computational resources. In contrast, another approach, such as BERT Topic, being a newer approach, has fewer implementations and resources available, which could affect its adoption in certain contexts. BERT Topic is based on BERT, a large-scale language model that demands substantial computational resources and training time. Compared to BERT Topic, LDA provides a more interpretable representation of topics. LDA models assign probabilities to words in each topic, allowing users to understand and label the topics based on the most representative words. This level of interpretability is crucial



in domains where human understanding and explainability are important, such as social sciences and content analysis.

It is important to note that both LDA and the summarization algorithms assume the documents to be a “bag of words” and do not take grammar into account. This purely statistical approach relies on the meanings of documents being conveyed through words (Arora and Ravindran, 2008). Regarding the LDA topic model (shown in Figure 2), the process of generating topics for a document is as follows:

- The dirichlet distribution (α) generates the topic distribution (θ_d) for document d .
- The multinomial distribution of the topic (θ_d) assigns the topic ($z_{d,n}$) for word n of document d .

- The corresponding topics from the dirichlet distribution (β_k) generate the topic distribution ($\varphi_{z_{d,n}}$).
- The multinomial distribution of topics ($\varphi_{z_{d,n}}$) generates the words ($w_{d,n}$).

N represents the different variables demonstrating the observed words in the various documents. K represents the total number of topics, and D represents the total number of documents.

Subsequently, LDA was employed for topic modeling to uncover hidden topics and terms in Reddit posts. This process analyzes the extensive amount of social media text to determine the connections between topics and summarizes the dataset on a scale that is impossible to achieve through manual annotation. In this study, we configured the algorithm to

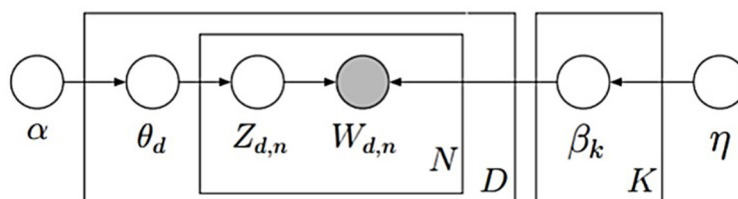


FIGURE 2
LDA topic model.

identify 30 topics per dataset, with each topic displaying its top 30 words.

The LDA algorithm was employed to perform clustering, aiming to identify coherent topics within the dataset. It was observed that a higher coherence score yielded better clustering results. Coherence served as a metric for determining the optimal number of clusters. The coherence scores were calculated using the scikit-learn metric package. We conducted experiments by varying the number of cluster parameters from 5 to 50, with intervals of 5. Among these values, the coherence score was highest when the number of topics (K) was set to 10. However, we conducted a manual inspection of the data. Upon manual inspection, it became evident that the results were most distinct and interpretable when the K -value was set to 30. Smaller K -values, such as 5 and 10, rendered the characteristics of the topics unclear and incomplete. Conversely, the largest K -value (50) resulted in an overlap of topic words across most topics. Consequently, we determined the optimal number of topics to be 30 for the overview topics and to be 20 for the sentiment topics.

Finally, we considered the list of top words in the probability distribution for each topic to assign a name to that specific topic. In other words, this interpretative process of labeling LDA topic outputs involves selecting a set of words from each topic's probability distribution that best represents the theme or subject of that topic. These words essentially serve as labels or names for the topics. Therefore, this process entails selecting words from each topic's probability distribution to provide a meaningful name or label that captures its content or theme. Afterward, we measured the agreement among three annotators regarding the labeling of the topics.

Sentiment analysis

An essential task in sentiment analysis is to classify the polarity of a given text at the sentence or document level, determining whether the expressed opinion is positive, negative, or neutral. Advanced sentiment classification can also identify emotions such as happiness, sadness, and anger, providing insights into the psychological patterns in text based on the analysis of human verbal behavior (Gottschalk and Bechtel, 2008). Therefore, sentiment analysis examines sentiment and detects individual words and phrases in the text to reflect the emotional scales.

To investigate the polarity properties of social media posts from Reddit, the SentiStrength software (<http://sentistrength.wlv.ac.uk/download.html>) was utilized for automatic sentiment analysis.

SentiStrength is specifically designed for analyzing social web texts and predicts the strength of positive and negative sentiment in short texts, even for informal language, by employing a lexical approach. The sentiment analysis captures reactions to events through free texts, offering new insights into human behavior and people's positive or negative attitudes toward an event (Thelwall and Buckley, 2013). The software operates in two modes: supervised and unsupervised. In the unsupervised mode, it employs predefined sentiment strength weights for the lexicon, while in the supervised mode, it utilizes a training set of data to automatically adjust the lexicon term weights and produce more accurate results.

At the core of SentiStrength is a list of 2,608 words and word stems, each associated with a typical positive or negative sentiment strength for social web texts. The software assigns a positive sentiment strength ranging from 1 (no positive sentiment) to 5 (very strong positive sentiment) and a negative sentiment strength ranging from -1 (no negative sentiment) to -5 (very strong negative sentiment) to each text (zeros are not used). In the presence of multiple sentiment words, the strongest sentiment word is chosen, while in the absence of sentiment words, a neutral sentiment is assumed. Additionally, the software incorporates special rules for handling negations, questions, booster words (e.g., "very"), emoticons, and various other special cases (Thelwall et al., 2010, 2012).

Results

Bipolar disorder topic modeling

To analyze the content of Reddit data, we assume that the number of 30 topics in LDA corresponds to the number of issues or events that describe each dataset. Therefore, we fixed $K = 30$ throughout this study. For each topic, we present only 30 words with higher probabilities under that topic.

In this paper, we display the top 10 ranked topics based on the topic probability distribution scores. Each topic consists of 30 words, with stronger influences indicating higher relevance to the topic. Using this information, we manually labeled each topic with a corresponding word. Additionally, we removed groups of words that appeared as stand-alone words or special characters, as they did not provide any relevant information about the topic. We replaced them with the next word in the rank. The main idea of each dataset's content is provided below.

The top 10 topics related to bipolar disorder (BD) and their top 30 words are as follows:

Topic 1. Help and advice (0.161): Social media users shared their conditions, asked for help and information, and expressed appreciation.

Topic 2. Bipolar disorder (0.158): Discussions about symptoms, mood swings, diagnosis, and medication.

Topic 3. Social process (0.139): Patients expressed the need for mental support from their social environment, such as family and friends.

Topic 4. Therapy (0.120): Experiences shared regarding medicine, doctors, and psychiatrists.

Topic 5. Drugs (0.116): Discussions about medications used for BD and related mental health issues on Reddit, including Lithium, Lamictal, and Wellbutrin.

Topic 6. Episode (0.084): Personal accounts of specific periods or conditions experienced by individuals living with BD.

Topic 7. Emotion expression (0.059): Patients with BD expressed their feelings about living with the illness, including emotions such as hate, sadness, and crying.

Topic 8. Friend (0.054): Frequent mentions of individuals who interacted with patients in social settings.

Topic 9. Relaxation (0.047): Techniques shared for relaxation during times of stress and anxiety, such as smoking, drinking, taking drugs, sleeping, eating, and meditation.

Topic 10. Drug-side effect (0.045): Discussions about the side effects of drugs used for BD, with weight gain being the most frequently mentioned side effect (Table 1).

Bipolar disorder conversation sentiment analysis

By employing topic-based sentiment analysis on social media data, it becomes possible to identify people's moods, related issues, and emotional responses to specific situations. For instance, when discussing a health topic, the mention of a drug name often evokes strong positive or negative sentiments within the conversation. To gauge the emotional intensity within the bipolar disorder (BD) community, we utilized SentiStrength 2.3 for sentiment analysis. This tool adopts a lexical approach that begins with a collection of terms known to have sentiment associations. It then applies a set of rules to predict the sentiment of texts based on the occurrence of these words. After obtaining the results of positive and negative text classification, we integrated these findings into the topic modeling approach to determine the common topics associated with each polarity.

Polarity classification results

The system employed sentiment analysis to determine the sentiment strength in Reddit data, categorizing each text as neutral, positive, or negative. However, for the purposes of this study, we focused solely on reporting the positive and negative sentiment text to observe the experiences related to bipolar disorder. Each text was assigned two numerical values by SentiStrength: a score between 1 and 5 for positive sentiment strength, and a score between -1 and -5 for negative sentiment strength. Additionally, a score of 1 or -1 indicated no sentiment, while a score of 5 or -5 represented

a strong sentiment of that type. For example, a social web text with a score of 1-4 would convey weak positive sentiment and strong negative sentiment. The system utilized two scales because a conviction can encompass both positivity and negativity, and the objective was to detect the expressed viewpoint rather than the overall polarity (Thelwall et al., 2010). Below are some examples of the short text analysis results:

“Cried at work Cried at work”

(Negative) 1 -4 Cried [-3] at [0] work [0] [[Sentence=-4,1=word max,1-5]] [[[1, -4 max of sentences]]]

“Wow, fantastic response. Thank you so much for sharing, you encouraged me!”

(Positive) 4 -1 Wow[2] fantastic[2][+1 Multiple Positive Words] response[0] [[Sentence=-1,4=word max, 1-5]]

Thank [1] you [0] so [0] much [0] For [0] sharing [0] you [0] encouraged [1] me [0] [[Sentence= -1,2=word max, 1-5]]

[[[4, -1 max of sentences]]]

The software assigned a score to each word in the sentence, and a total weighted score was calculated to predict the text's polarity. Each word received a numerical value based on its strength. For instance, the word “cried” was assigned a strongly negative score of -3, while both “wow” and “fantastic” were rated as moderately positive with a score of 2.

The results of the sentiment analysis indicate that there are more negative sentiment (641,264) than positive sentiment (392,352) within the discussions of Reddit-related bipolar disorder communities. Examples of the experimental results are presented in Table 2.

Polarity topics results

In order to investigate the polarity of emotions surrounding BD in social media discourse, we identified common issues that were publicly discussed within the Reddit-related bipolar community. The texts from the sentiment analysis results were divided into two datasets: positive and negative. These datasets were then subjected to LDA topic modeling to uncover hidden topics within each polarity. We set the number of topics to 20 and specified that each topic should consist of 20 words with the highest probabilities.

The LDA results present the content of documents with 20 different topic mixtures and 20 words per topic. For the purpose of this paper, we selected the top five topics to report. We manually assigned labels to these top five topics based on the words associated with each topic.

Table 3 illustrates the positively-related topics of BD within the Reddit community. These topics include “Medication”, where users express their feelings and opinions regarding drugs used in bipolar treatment, such as lamictal, abilify, and lithium, and discuss their effectiveness. The next topic is “Social wellbeing”, which explores the sense of social inclusion, individuals' lifestyles, and overall quality of life. The third topic is “Personal concerns”, which delves into social factors with a greater focus on personal experiences, such as career, relationships, and children. The fourth topic is “Social support”, which highlights the importance of having friends, family, and others to rely on during times of need and

TABLE 1 Top-10 topics of bipolar disorder discussion in the Reddit communities.

Topic label	Topic 1		Topic 2		Topic 3		Topic 4		Topic 5	
	Help and advice		Bipolar disorder		Social process		Therapy		Drugs	
Probability scores	0.16115		0.1582		0.13998		0.12064		0.1164	
Frequency of words	Thank	63,864	Bipolar	82,296	Bipolar	33,047	Med	74,865	Lithium	42,276
	Thanks	30,261	Depression	38,808	Help	29,800	Help	30,637	Lamictal	40,493
	Hear	22,372	Mood	33,029	Feel	29,300	Doctor	23,993	Effect	39,410
	Glad	22,207	Diagnose	32,433	Time	25,765	Try	22,148	Dose	24,659
	Help	20,575	Hypomanic	23,530	Tell	24,328	Feel	19,591	Doctor	23,847
	I-hope	17,930	Mixed	21,250	Friend	24,168	Time	19,086	Med	19,571
	Sorry	16,814	Hypomania	20,783	Love	22,309	Medication	17,821	Medication	19,000
	FEEL	15,686	Feel	20,517	Life	21,434	Psychiatrist	15,681	Start	17,773
	Share	14,531	Manic	20,183	Try	20,360	Life	14,047	Mood	16,725
	Post	12,577	Experience	18,888	Relationship	19,564	Bipolar	10,744	Try	14,625
	Try	11,852	Diagnosis	17,897	yourself	19,170	Therapist	9,408	Help	14,529
	I-feel	11,619	Depressive	17,731	Support	19,095	Ago	8,997	Time	13,106
	Look	11,398	Mania	17,601	Understand	18,129	Luck	8,426	Wellbutrin	12,825
	Hope	10,149	Time	17,131	Hard	16,953	Talk	8,400	Doc	12,065
	Reply	8,317	Mixed-episode	16,853	Mental	15,527	Tell	8,054	Experience	11,876
	Definitely	8,170	Doctor	15,534	Talk	15,490	Start	7,940	Dosage	11,590
	Hard	7,448	Psychiatrist	15,424	Care	14,930	Diagnosis	7,911	Feel	10,685
	Comment	7,353	Symptom	15,310	Person	14,107	Help-i	7,240	Psychiatrist	10,085
	Read	7,349	Episode	15,059	Family	13,691	Symptom	7,113	Low	9,934
	Nice	7,274	Medication	14,658	I-feel	12,972	Mood	6,975	Stop	9,426
	Happy	7,233	I-feel	14,625	Mean	12,851	Pdoc	6,642	Prescribe	9,250
	Time	7,107	Depressive-episode	13,299	Deal	10,732	Stop	6,599	Lamotrigine	9,159
	Sound	7,066	Start	10,787	I-hope	10,461	Diagnose	6,233	Pdoc	8,941
	Advice	6,974	Bipolar-disorder	10,571	Learn	9,668	Doc	6,188	Depression	8,885
	Luck	6,405	Manic-episode	10,052	Diagnosis	8,958	Effect	5,950	Increase	8,304
	Word	5,795	Anxiety	9,670	Sound	8,542	Depression	5,486	Mood-stabilizer	7,867
	Start	5,719	Rapid	9,606	Illness	8,348	Prescribe	5,360	Notice	7,810
	Helpful	5,633	Rapid-cycling	9,491	Feeling	8,027	Pill	5,348	Anxiety	7,704
	Sub	5,507	Happen	9,208	Hurt	7,983	Appointment	5,323	Add	7,563
	Appreciate	5,480	Med	8,363	Mental-illness	7,829	Therapy	4,927	Drug	7,260

(Continued)

TABLE 1 (Continued)

Topic label	Topic 6		Topic 7		Topic 8		Topic 9		Topic 10	
	Episode		Emotion expression		Friend		Relaxation		Drug-side effect	
Probability scores	0.08456		0.05908		0.05406		0.04717		0.04573	
Frequency of words	Sleep	59,445	Feel	17,039	I-feel	16,101	Help	16,821	Weight	17,854
	Night	38,903	I-feel	16,756	Feel	14,878	Smoke	12,094	Abilify	12,770
	Bed	17,855	Time	12,838	Time	12,789	Time	11,977	Weight-gain	12,089
	Morning	15,444	I-hate	10,471	Friend	7,550	Anxiety	11,081	Eat	11,338
	Time	15,071	Try	9,214	Help	6,715	Weed	10,436	Med	10,922
	Feel	14,059	Life	7,246	Life	5,837	Try	7,876	Gain	10,757
	I-feel	12,268	Wanna	6,948	Feeling	5,134	Help-i	7,398	Lamictal	9,343
	Hour	11,709	Fuck	6,236	Love	4,977	Drink	7,210	Latuda	9,266
	Help	10,571	Shit	5,854	Tell	4,445	I-feel	6,880	Effect	8,781
	Wake	10,461	Bad	5,510	Try	4,103	Feel	6,769	Lose	7,043
	I-start	8,395	Wan	5,509	Happen	3,814	Life	5,635	Time	6,588
	I-wake	7,428	Tell	4,627	Person	3,697	Stop	4,985	Start	5,866
	Fall	6,829	Deal	4,527	Little	3,692	Alcohol	4,739	Seroquel	5,708
	Try	6,786	Happen	4,378	Talk	3,632	Panic	4,729	Gain_weight	5,687
	Seroquel	6,740	Med	4,257	Bring	3,627	Start	4,724	I-gain	5,435
	I-sleep	6,511	Help	3,955	Sometimes-i	3,598	Sleep	4,453	Help	4,586
	Happen	6,472	Die	3,864	Hard	3,417	Drug	4,432	Body	4,223
	Else	5,945	Cry	3,573	Bad	3,183	Medication	3,895	Stop	4,122
	Fall-asleep	5,737	Stop	3,489	Voice	3,088	Med	3,840	Feel	3,732
	Energy	5,643	Sometimes	3,437	Idea	3,064	I-smoke	3,806	Food	3,570
	Start	5,523	Kill	3,428	Else	3,058	Eat	3,805	I-feel	3,501
	Tired	5,261	Stay	3,376	Maybe	3,054	Bit	3,740	Depression	3,485
	Stay	5,145	Friend	3,278	Leave	2,925	Smoking	3,693	Try	3,406
	Little	5,027	Brain	3,247	Help-i	2,822	hard	3,509	Switch	3,302
	Super	4,623	Totally	3,175	College	2,763	Panic-attack	3,230	Appetite	3,153
	Couple	4,420	Feeling	3,151	Look	2,760	Job	3,207	Medication	3,152
	Dream	4,138	Care	2,984	Start	2,730	Bad	3,133	I-eat	3,130
	Tell	4,094	Else	2,938	Head	2,612	Worse	3,099	Cause	3,059
	Med	3,978	Live	2,838	Depression	2,525	Manic	3,013	Pretty	2,969
	Body	3,850	Depression	2,811	Word	2,508	Quit	2,904	Little	2,692

TABLE 2 Examples of polarity categorization.

Reddit text post	Positive score	Negative score	Polarity	Weighted scores
Glad it helped somehow somehow Even if originally bio-chemical, the “real” effects of bipolar affect us in deep psychological ways too.	2	-1	Positive	Glad[1] it[0] helped[0] somehow[0][+1 Emoticon] Even[0] if[0] originally[0] bio[0] chemical[0] the[0] real[0] effects[0] of[0] bipolar[0] affect[0] us[0] in[0] deep[0] psychological[0] ways[0] too[0] [[Sentence=-1,2=word max, 1-5]][[2,-1 max of sentences]]
Do you have a therapist you work with? Can you get an appointment soon? 10 h days consistently is a lot for anyone, maybe give yourself a break from that many hours. You need to be your best friend and advocate even though that is one of the hardest things to do. Do you have a friend or family member you can talk to? Sending you hugs.	3	-1	Positive	Do[0] you[0] have[0] a[0] therapist[0] you[0] work[0] with[0] [[Sentence=-1,1=word max, 1-5]] Can[0] you[0] get[0] an[0] appointment[0] soon[0] [[Sentence=-1,1=word max, 1-5]] 10hr[0] days[0] consistently[0] is[0] a[0] lot[0] for[0] anyone[0] maybe[0] give[0] yourself[0] a[0] break[0] from[0] that[0] many[0] hours[0] [[Sentence=-1,1=word max, 1-5]] You[0] need[0] to[0] be[0] your[0] best[1] friend[1] and[0] advocate[0] even[0] though[0] that[0] is[0] one[0] of[0] the[0] hardest[0] things[0] to[0] do[0] [[Sentence=-1,2=word max, 1-5]] Do[0] you[0] have[0] a[0] friend[1] or[0] family[0] member[0] you[0] can[0] talk[0] to[0] [[Sentence=-1,2=word max, 1-5]] Sending[0] you[0] hugs[2] [[Sentence=-1,3=word max, 1-5]][[3,-1 max of sentences]]
I actually just purchased a Bipolar II specific workbook that goes through CBT, DBT, and mindfulness, I'm really excited to get it. I'm glad they've helped you!	4	-1	Positive	I[0] actually[0] just[0] purchased[0] a[0] Bipolar[0] II/I[0] specific[0] workbook[0] that[0] goes[0] through[0] CBT[0] DBT[0] and[0] mindfulness[0] I'm[0] really[0] excited[2][1 LastWordBoosterStrength] to[0] get[0] it[0] [[Sentence=-1,4=word max, 1-5]] I'm[0] glad[1] they've[0] helped[0] you[0] [[Sentence=-1,2=word max, 1-5]][[4,-1 max of sentences]]
That's adorable!! Also try sudoku, that's supposed to be a good memory builder. My therapist also says to start journaling during the day because writing things down can leave a thicker mark on the memory. Good luck :)	5	-1	Positive	That's[0] adorable[3][+0.6 EmphasisInPunctuation] [[Sentence=-1,5=word max, 1-5]] Also[0] try[0] sudoku[0] that's[0] supposed[0] to[0] be[0] a[0] good[1] memory[0] builder[0] [[Sentence=-1,2=word max, 1-5]] My[0] therapist[0] also[0] says[0] to[0] start[0] journaling[0] during[0] the[0] day[0] because[0] writing[0] things[0] down[0] can[0] leave[0] a[0] thicker[0] mark[0] on[0] the[0] memory[0] [[Sentence=-1,1=word max, 1-5]] Good[1] luck[2][+1 Emoticon] [[Sentence=-1,4=word max, 1-5]][[5,-1 max of sentences]]
Wondering about life w/o meds	1	-1	Neutral	Wondering[0] about[0] life[0] w/o[0] meds[0] [[Sentence=-1,1=word max, 1-5]][[1,-1 max of sentences]]
I forgot to add, that the lower the dosage of the Seroquel, the drowsier you are. My effective dose was 300 mg, but I was still so tired. We bumped me up to 400 mg so I could sleep 10 h instead of 12. In addition, make sure you're taking the pill 3-4 h prior to bedtime so you concentrate the drowsiness (as much as possible) during your sleeping hours. I take mine at 5 p.m. for this reason. Just keep backing it off until the timing is right for you.	1	-2	Negative	I[0] forgot[0] to[0] add[0] that[0] the[0] lower[0] the[0] dosage[0] of[0] the[0] Seroquel[0] the[0] drowsier[0] you[0] are[0] [[Sentence=-1,1=word max, 1-5]] My[0] effective[0] dose[0] was[0] 300 mg[0] but[0] I[0] was[0] still[0] so[0] tired[-1] [[Sentence=-2,1=word max, 1-5]] We[0] bumped[0] me[0] up[0] to[0] 400 mg[0] so[0] I[0] could[0] sleep[0] 10[0] hours[0] instead[0] of[0] 12[0] [[Sentence=-1,1=word max, 1-5]] And[0] make[0] sure[0] you're[0] taking[0] the[0] pill[0] 3[0] -4[0] hour[0] prior[0] to[0] bedtime[0] so[0] you[0] concentrate[0] the[0] drowsiness[-1] as[0] much[0] as[0] possible[0] during[0] your[0] sleeping[0] hours[0] [[Sentence=-2,1=word max, 1-5]] I[0] take[0] mine[0] at[0] 5pm[0] for[0] this[0] reason[0] [[Sentence=-1,1=word max, 1-5]] Just[0] keep[0] backing[0] it[0] off[0] until[0] the[0] timing[0] is[0] right[0] for[0] you[0] [[Sentence=-1,1=word max, 1-5]][[1,-2 max of sentences]]
Lack of appetite and weight loss with Lamictal	1	-3	Negative	Lack[-1] of[0] appetite[0] and[0] weight[0] loss[-2] with[0] Lamictal[0] [[Sentence=-3,1=word max, 1-5]][[1,-3 max of sentences]]
Does anyone else feel the difference between suicidal depressed and regular depressed	1	-4	Negative	Does[0] anyone[0] else[0] feel[0] the[0] difference[0] between[0] suicidal[0] depressed[-3] and[0] regular[0] depressed[-3] [[Sentence=-4,1=word max, 1-5]][[1,-4 max of sentences]]
For those of you out there feeling sad or depressed, hopeless or helpless, anxious or suicidal, self-harming or self-medicating, or suffering from mental illness	1	-5	Negative	For[0] those[0] of[0] you[0] out[0] there[0] feeling[0] sad[-3] or[0] depressed[-3] hopeless[-3][+1 MultiplePositiveWords] or[0] helpless[-1] anxious[-2] or[0] suicidal[0] self[0] harming[-2] or[0] self[0] medicating[0] or[0] suffering[-3] from[0] mental[0] illness[-1] [[Sentence=-5,1=word max, 1-5]][[1,-5 max of sentences]]

TABLE 3 Top-five positive topics.

Topics	Top words
Medication (0.00278)	Lamictal weight side medication abilify seroquel lithium latuda wellbutrin luck gain doctor dose night mood hope food body sleep doc
Social wellbeing (0.00266)	Awesome life happy glad advice feeling social word sense hard love step healthy high hope issue night long friend heart
Personal concerns (0.00194)	List idea hug daylio emood app kid mood love work free vibes life easy feeling relationship luck mind minute
Social support (0.00189)	Episode glad mixed real mania parent life friend manic quick psychiatrist hypomanic luck person family idea normal partner support story
Mental therapy (0.00187)	Therapy life therapist mental helpful disorder person health idea interesting luck diagnosis hard job question feeling dbt super problem group

TABLE 4 Top-five negative topics.

Topics	Top words
Drugs (0.00915)	Lamictal dose side depression medication anxiety lithium doctor weight abilify latuda wellbutrin mood seroquel psychiatrist night gain low drug stabilizer
Mood swings (0.00912)	Episode depression mood symptom mania manic depressive hypomanic hypomania diagnosis mixed anxiety psychiatrist medication disorder depressed normal experience
Symptoms (0.00428)	Night sleep anxiety depression shit work feeling friend tired episode normal body morning panic attack bed happy brain full life
Treatments (0.00409)	Life thought night feeling friend bed medication episode depression anxiety therapist depressed work doctor psychiatrist morning hard manic therapy dose
Mental disorder (0.00369)	Disorder therapist mental diagnosis therapy mood life depression anxiety medication issue illness psychiatrist doctor symptom bpd reason health episode relationship

who provide encouragement when experiencing bipolar-related symptoms, such as mood swings. Lastly, the fifth topic is “Mental therapy”, which revolves around discussions within the community regarding various treatments for the disease.

Table 4 presents the negative topics of BD in the Reddit community. The first topic is “Drugs”, where social media users share their experiences related to medications and their side effects. The following topic is “Mood swings”, where individuals living with bipolar disorder describe their experiences with mixed episodes, including depression, hypomania, and mania. The next topic is “Bipolar symptoms”, which encompasses the various symptoms associated with the condition, such as anxiety, depression, panic, and happiness, typically occurring during the night and morning and significantly impacting their lives. The fourth topic is “Treatments”, where individuals discuss their attempts to address negative feelings through strategies like meeting friends, consulting with a psychiatrist, and seeking medical assistance. Lastly, the fifth topic discussed within the bipolar sub-Reddit community is “Mental disorder”, which takes a broader perspective on the mental challenges they face in both personal and social aspects of their lives.

To illustrate the association between terms and document topics, we constructed a co-word map. The positive and negative text post were processed into graphML format to generate word co-occurrences and create a network of co-occurring words. Subsequently, we visualized the graph using Gephi software, adjusting the nodes to highlight those with high betweenness centrality for improved visualization. This allowed us to observe the graphs with key terms (nodes) representing the core concepts within each polarity. The results are presented below:

Figure 3 presents the co-word network of the positive post text. The terms that influenced other nodes the most

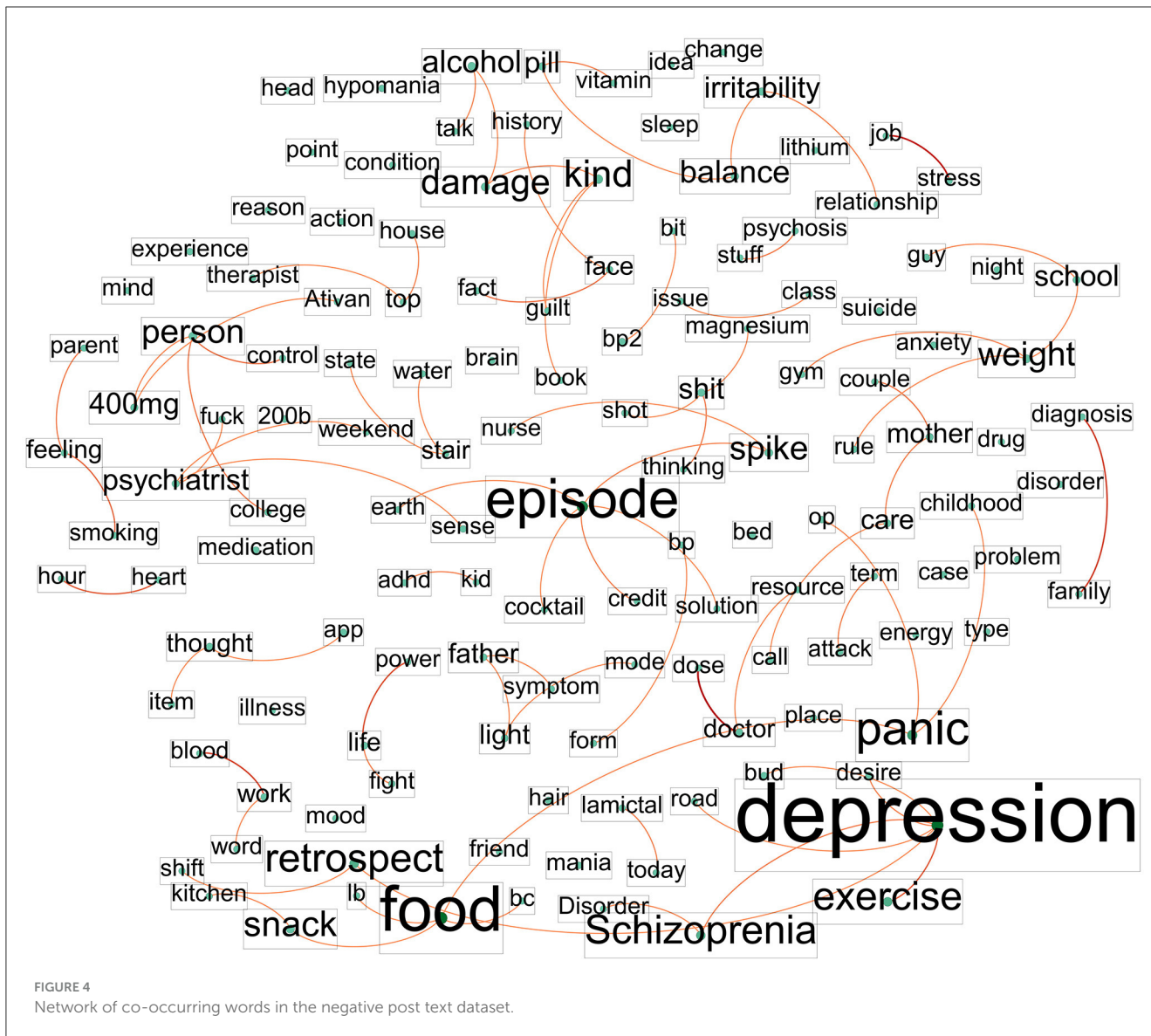
were “life” (0.00016), followed by “mind” (0.000149), “doctor” (0.000142), “episode” (0.000118), and “medication” (0.000114). These terms exhibited a high betweenness centrality score. The co-word network confirms a core concept of positive conversations, where Reddit users shared experiences about their personal lives, including medication, mind therapy, activities, and social environments. Additionally, the network reveals connected terms such as life-peace, doctor-mind, work-love, medication-change, and song-summer.

Figure 4 displays the co-occurrence network of words in the negative post text. The top-ranking terms, which play a crucial role as bridges in the network, are “depression” (0.000256), followed by “life” (0.000164), “episode” (0.000157), “shit” (0.000101), “psychiatrist” (0.000093), and “food” (0.000091). This graph reveals that individuals living with BD expressed concerns about recurrent disorders such as schizophrenia and obesity, along with symptoms of mood swings. Furthermore, they experienced irritability in their relationships with others and struggled with negative thoughts, including suicidal ideation. Notable word pairs identified in the network include depression-schizophrenia, life-fight, episode-spike, shit-thinking, food-panic, and irritability-relationship.

Discussion

Topics discussed in the bipolar disorder online community of Reddit users

The online community of Reddit users discussing Bipolar disorder covers a variety of topics related to the condition. The top



regarding the condition, seeking understanding, empathy, and establishing a support system. Friends are regarded as essential in the lives of individuals with bipolar disorder, and conversations often revolve around the impact of friendships on mental health, maintaining relationships, and the importance of supportive friends. This finding is consistent with a study on the use of social media for social support among adolescents, including emotional, appraisal, and informational support (Selkie et al., 2020). Discussions also cover various relaxation techniques aimed at managing stress and anxiety associated with bipolar disorder. Methods such as smoking, drinking, medication use, sleeping, eating, and meditation practices are mentioned. It is important to approach coping strategies, including substance use, with caution and consult healthcare professionals for guidance. Moreover, the online community serves as a platform for seeking help and information. Users actively seek advice on managing symptoms, finding effective treatment options, and expressing

gratitude for the support they receive. Social media enables people to share information on treatments and research to improve care (De Choudhury and De, 2014; Pershad et al., 2018).

Overall, the Bipolar disorder online community on Reddit encompasses a wide range of topics related to the condition, fostering support, understanding, and the sharing of experiences among its members.

Polarity classification

In the analysis of people’s opinions and feelings toward mental health issues, specifically related to Bipolar Disorder (BD), the sentiment was assessed using the SentiStrength

software. The software analyzed user-generated text from social media, which described situations related to health problems. The results of the sentiment analysis on the Reddit dataset, under the topic of BD indicated a higher frequency of negative polarity compared to positive polarity. This finding suggests that discussions and expressions within the Bipolar Disorder online community on Reddit tended to lean toward a more negative sentiment. It indicates that individuals sharing their experiences, challenges, and emotions related to BD often conveyed negative feelings or perspectives.

It is important to note that sentiment analysis provides an automated assessment of sentiment based on textual content, and it may not fully capture the nuances and complexities of individual experiences. Additionally, the analysis is limited to the specific dataset and platform analyzed, and sentiments expressed may vary across different social media platforms and online communities. For instance, using Twitter content analysis to measure attitudes toward mental illness and analyzing mental health based on the positive language of Facebook users provided different insights into online discussions about mental health (Reavley and Pilkington, 2014; Bogolyubova et al., 2020).

Further qualitative analysis and a deeper understanding of the specific context and content of the negative sentiment would be necessary to gain insights into the concerns, challenges, and negative experiences expressed by individuals living with or relating to Bipolar Disorder in the Reddit community.

Polarity topics

The study findings revealed that individuals expressed their thoughts and emotions related to both the biological and mental aspects of their conditions. Distinct topics emerged when analyzing positive and negative sentiments. Regarding positive sentiment, Reddit users discussed receiving encouragement and support from various sources, including friends, partners, family members, doctors, and psychiatrists. They expressed satisfaction with their social wellbeing while living with a mental illness. Some users also reported positive experiences with bipolar medication and psychological therapy, although others mentioned negative effects such as weight gain from the drugs.

On the other hand, negative sentiment topics focused on the negative aspects of drugs and treatments, including instances of incorrect diagnosis by physicians when experiencing mental symptoms. Users also shared their struggles with bipolar disorder symptoms, particularly mood swings, which caused both biological, and psychological stress.

The polarity findings emphasize the results showing that social media users seek information to address their mental health problems and to connect with individuals who can understand and encourage positive emotions. These findings underscore the importance of addressing patients' wellbeing and providing appropriate treatments for their mental health.

Limitation

It is essential to acknowledge certain limitations inherent in our study. One notable limitation is the nature of bipolar disorder (BD) as a clinical diagnosis. Within the context of our research, participants on social media platforms may encompass a diverse range of individuals, including those who have received a professional diagnosis of BD, those who self-identify as bipolar, and individuals who are family members of those with BD. Consequently, this gap introduces a potential risk of misattribution when analyzing sentiments and topics related to BD within the context of discussions about BD. It is important to recognize and consider this limitation.

Furthermore, in future work, we should consider the potential value of predicting user post-behavior during the presence of mental symptoms and conducting a sentiment comparison between Bipolar Disorder (BD) and other online communities.

Conclusion

This study aimed to explore the topics and sentiments discussed in the Bipolar disorder online community on Reddit. The findings revealed that users engaged in discussions related to various aspects of bipolar disorder, including symptoms, mood swings, diagnosis, medication, therapy, social support, and relaxation techniques. They shared their personal experiences, sought advice, and expressed their emotions about living with bipolar disorder. The analysis showed a predominance of negative sentiment expressed in the community, indicating the presence of challenges and struggles associated with the condition. These findings emphasize the importance of addressing the wellbeing of individuals living with bipolar disorder and highlight the need for appropriate treatment and support. The research framework employed in this study can also be applied to gain insights into other health conditions through social media analysis.

In summary, the primary objective of the present study is to enhance the understanding of how individuals with BD communicate on social media platforms. While not providing definitive medical insights, our research offers opportunities for further exploration. We intend to investigate unmet information needs and aspects requiring professional attention in future research, guided by the input of clinicians and individuals with BD, ensuring practical relevance.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

The data used in this study was collected from Reddit, a social media platform, in accordance with the platform's terms of service and community guidelines. The data collection process strictly

adhered to the ethical standards outlined by Reddit, ensuring respect for user privacy, anonymity, and the responsible use of the platform's content. Consequently, we confirm that the collection, analysis, and reuse of social media data were conducted in strict accordance with Reddit's policies and terms of use, as well as all relevant institutional regulations.

Author contributions

TT conceived and co-designed the analysis, collected the data, and co-wrote the paper. QX co-designed the analysis and co-wrote the paper. SL co-wrote the paper. All authors contributed to the article and approved the submitted version.

References

- Apache OpenNLP Development Community (2011). *Apache OpenNLP Developer Documentation.pdf*. OpenNLP. Available online at: <https://opennlp.apache.org/docs/1.9.0/manual/opennlp.html> (accessed July 23, 2023).
- Arora, R., and Ravindran, B. (2008). "Latent dirichlet allocation based multi-document summarization," in *Proceedings of SIGIR 2008 Workshop on Analytics for Noisy Unstructured Text Data, AND'08*. Available online at: <http://www.cse.ittm.ac.in/~ravi/papers/Rachit-AND08.pdf> (accessed July 23, 2023).
- Blei, D. M., Edu, B. B., Ng, A. Y., Edu, A. S., Jordan, M. I., and Edu, J. B. (2003). Latent dirichlet allocation. *J. Mach. Learn. Res.* 3, 993–1022. Available online at: <https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf> (accessed July 23, 2023).
- Bogolyubova, O., Panicheva, P., Ledovaya, Y., Tikhonov, R., and Yaminov, B. (2020). The language of positive mental health: findings from a sample of Russian Facebook users. *SAGE Open* 10, 2158244020924370. doi: 10.1177/2158244020924370
- Chen, J., and Wang, Y. (2021). Social media use for health purposes: systematic review. *J. Med. Inter. Res.* 23, e17917. doi: 10.2196/17917
- Conway, M., and O'Connor, D. (2016). Social media, big data, and mental health: current advances and ethical implications. *Curr. Opin. Psychol.* 9, 77–82. doi: 10.1016/j.copsyc.2016.01.004
- De Choudhury, M., and De, S. (2014). Mental health discourse on reddit: self-disclosure, social support, and anonymity. *Proc. Int. AAAI Conf. Web Soc. Media* 8, 71–80. doi: 10.1609/icwsm.v8i1.14526
- Denecke, K. (2015). *Health Web Science: Social Media Data for Healthcare*, eds Y. Zhang. Cham: Springer.
- Foufi, V., Timakum, T., Gaudet-Blavignac, C., Lovis, C., and Song, M. (2019). Mining of textual health information from Reddit: analysis of chronic diseases with extracted entities and their relations. *J. Med. Inter. Res.* 21, e12876. doi: 10.2196/12876
- Gkotsis, G., Oellrich, A., Hubbard, T., Dobson, R., Liakata, M., Velupillai, S., et al. (2016). "The language of mental health problems in social media," in *Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology*, p. 63–73. Available online at: <https://aclanthology.org/W16-0307.pdf> (accessed July 23, 2023).
- Gottschalk, L. A., and Bechtel, R. J. (eds.). (2008). *Computerized Content Analysis of Speech and Verbal Texts and Its Many Applications*. New York, NY: Nova Science.
- Harlow, L. L., and Oswald, F. L. (2016). Big data in psychology: introduction to the special issue. *Psychol. Methods.* 21, 447–457. doi: 10.1037/met0000120
- Huang, Y.-H., Wei, L.-H., and Chen, Y.-S. (2017). *Detection of the Prodromal Phase of Bipolar Disorder from Psychological and Phonological Aspects in Social Media*. Available online at: <http://arxiv.org/abs/1712.09183> (accessed June 15, 2023).
- Islam, R., Kabir, M. A., Ahmed, A., Kamal, A. R. M., and Wang, H. (2018). Depression detection from social network data using machine learning techniques. *Health Inform. Sci. Syst.* 6, 8. doi: 10.1007/s13755-018-0046-0
- Johnson, G., and Ambrose, P. (2006). Neo-tribes: the power and potential of online communities in health care. *Commun ACM* 49, 107–113. doi: 10.1145/1107458.1107463
- Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., and McClosky, D. (2014). "The stanford CoreNLP natural language processing toolkit" in *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, p. 55–60. Available online at: <https://aclanthology.org/P14-5010.pdf> (accessed July 23, 2023).
- Manning, C. D., Ragahvan, P., and Schutze, H. (2009). *An Introduction to Information Retrieval*. Information Retrieval. Available online at: <https://ds.amu.edu.et/xmlui/bitstream/handle/123456789/14697/Book%20558%20pages.pdf?sequence=1&isAllowed=y> (accessed July 30, 2023).
- Marar, S. D., Al-Madaney, M. M., and Almousawi, F. H. (2019). Health information on social media: perceptions, attitudes, and practices of patients and their companions. *Saudi Med. J.* 40, 1294. doi: 10.15537/smj.2019.12.24682
- Monnier, J., Laken, M., and Carter, C. L. (2002). Patient and caregiver interest in Internet-based cancer services. *Cancer Pract.* 10, 305–310. doi: 10.1046/j.1523-5394.2002.106005.x
- Moorhead, S. A., Hazlett, D. E., Harrison, L., Carroll, J. K., Irwin, A., and Hoving, C. (2013). A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. *J. Med. Inter. Res.* 15, 1–16. doi: 10.2196/jmir.1933
- Naslund, J. A., Aschbrenner, K. A., Marsch, L. A., and Bartels, S. J. (2016). The future of mental health care: peer-to-peer support and social media. *Epidemiol. Psychiatr. Sci.* 25, 113–122. doi: 10.1017/S2045796015001067
- Nguyen, T., Phung, D., Dao, B., Venkatesh, S., and Berk, M. (2014). Affective and content analysis of online depression communities. *IEEE Trans. Affect. Comput.* 5, 217–226. doi: 10.1109/TAFFC.2014.2315623
- O'Dea, B., Wan, S., Batterham, P. J., Calear, A. L., Paris, C., and Christensen, H. (2015). Detecting suicidality on twitter. *Intern. Intervent.* 2, 183–188. doi: 10.1016/j.invent.2015.03.005
- O'Neill, B., Ziebland, S., Valderas, J., and Lupiáñez-Villanueva, F. (2014). User-generated online health content: a survey of internet users in the United Kingdom. *J. Med. Intern. Res.* 16, 3187. doi: 10.2196/jmir.3187
- O'Rourke, N., Heisel, M. J., Canham, S. L., and Sixsmith, A. (2017). Predictors of suicide ideation among older adults with bipolar disorder. *PLoS ONE.* 16, e0187632. doi: 10.1371/journal.pone.0187632
- Park, A., and Conway, M. (2018). Harnessing reddit to understand the written-communication challenges experienced by individuals with mental health disorders: analysis of texts from mental health communities. *J. Med. Intern. Res.* 20, e121. doi: 10.2196/jmir.8219
- Pershad, Y., Hangege, P. T., Albadawi, H., and Oklu, R. (2018). Social medicine: Twitter in healthcare. *J. Clin. Med.* 7, 121. doi: 10.3390/jcm7060121
- Reavley, N. J., and Pilkington, P. D. (2014). Use of Twitter to monitor attitudes toward depression and schizophrenia: an exploratory study. *PeerJ.* 2, e647. doi: 10.7717/peerj.647
- Record, R. A., Silberman, W. R., Santiago, J. E., and Ham, T. (2018). I sought it, I Reddit: examining health information engagement behaviors among Reddit users. *J. Health Commun.* 23, 470–476. doi: 10.1080/10810730.2018.1465493
- Selkie, E., Adkins, V., Masters, E., Bajpai, A., and Shumer, D. (2020). Transgender adolescents' uses of social media for social support. *J. Adoles. Health* 66, 275–280. doi: 10.1016/j.jadohealth.2019.08.011
- Song, M., Kim, W. C., Lee, D., Heo, G. E., and Kang, K. Y. (2015). PKDE4J: entity and relation extraction for public knowledge discovery. *J. Biomed. Inform.* 57, 320–332. doi: 10.1016/j.jbi.2015.08.008

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Thelwall, M., and Buckley, K. (2013). Topic-based sentiment analysis for the social web: the role of mood and issue-related words. *J. Am. Soc. Inform. Sci. Technol.* 64, 22872. doi: 10.1002/asi.22872

Thelwall, M., Buckley, K., and Paltoglou, G. (2012). Sentiment strength detection for the social web. *J. Am. Soc. Inform. Sci. Technol.* 63, 163–173. doi: 10.1002/asi.21662

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., and Kappas, A. (2010). Sentiment in short strength detection informal text. *J. Am. Soc. Inform. Sci. Technol.* 61, 2544–2558. doi: 10.1002/asi.21416

Trotzek, M., Koitka, S., and Friedrich, C. M. (2018). Utilizing neural networks and linguistic metadata for early detection of depression indications in text sequences. *IEEE Trans. Knowl. Data Eng.* 32, 588–601. doi: 10.1109/TKDE.2018.2885515

Zhang, Y., He, D., and Sang, Y. (2013). Facebook as a platform for health information and communication: a case study of a diabetes group. *J. Med. Syst.* 37, 1–12. doi: 10.1007/s10916-013-9942-7

Zhao, Y., and Zhang, J. (2017). Consumer health information seeking in social media: a literature review. *Health Inform. Libraries J.* 34, 268–283. doi: 10.1111/hir.12192