Computers in Human Behavior 165 (2025) 108544

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Full length article

Participant behavior and community response in online mental health communities: Insights from Reddit

Virginia Morini [a,b,*], Maria Sansoni [c], Giulio Rossetti [b], Dino Pedreschi [a], Carlos Castillo [d,e]

[a] Department of Computer Science, University of Pisa, Pisa, Italy
[b] Institute of Information Science and Technologies "A. Faedo" (ISTI), National Research Council (CNR), Pisa, Italy
[c] Department of Psychology, Catholic University of the Sacred Heart, Milan, Italy
[d] Department of Engineering, Universitat Pompeu Fabra, Barcelona, Spain
[e] ICREA, Barcelona, Spain

ARTICLE INFO

Keywords:
Online mental health communities
Reddit
Mental health
Self-disclosure
Posting intents
Community response

ABSTRACT

The growing presence of online mutual-help communities has significantly changed how people access and provide mental health (MH) support. While extensive research has explored self-disclosure and social support dynamics within these communities, less is known about users' distinctive behavioral patterns, posting intents, and community response. This study analyzed a large-scale, five-year Reddit dataset of 67 MH-related subreddits, comprising over 3.4 million posts and 24 million comments from approximately 2.4 million users. We categorized subreddits based on the Diagnostic and Statistical Manual of Mental Disorders and compared the behavioral patterns in these communities with Reddit non-MH ones. Leveraging Reddit's post flair feature, we defined a ground truth for post intents and applied an automated classification method to infer intents across the dataset. We then used causal inference analysis to assess the effect of community responses on subsequent user behavior. Our analysis revealed that MH-related subreddits featured unique characteristics in content length, throwaway account usage, user actions, persistence, and community response. These online behaviors mirrored those in other mutual-help Reddit communities and resonated with offline patterns while diverging from non-support-oriented subreddits. We also found that seeking support and venting are the predominant posting intents, with users tending to maintain consistent intents over time. Furthermore, we observed that receiving comments and reactions significantly influenced users' follow-up engagement, fostering increased participation. These findings highlight the supportive role of online MH communities and emphasize the need for tailored design to optimize user experience and support for individuals facing MH challenges.

1. Introduction

Mental health (MH) issues affect millions of individuals worldwide, posing significant challenges to personal well-being and societal health systems (World Health Organization, 2022). Traditionally, support for those experiencing MH challenges has been primarily offered through face-to-face interactions in clinical settings, support groups, or personal relationships (Davidson, Chinman, Sells, & Rowe, 2006). However, these conventional support systems face several limitations, including geographical constraints, time restrictions, and the persistent stigma around MH (Andrade et al., 2014). Moreover, overburdened healthcare systems in many countries have led to long waiting lists and limited availability of MH services. At the same time, the financial cost of psychotherapy creates a substantial gap between those requiring support and those able to access it (Saxena, Thornicroft, Knapp, & Whiteford, 2007).

In response to these challenges, peer support MH communities within social media platforms have emerged as valid alternatives, offering advantages that may help mitigate many of the limitations of traditional support systems (Naslund, Aschbrenner, Marsch, & Bartels, 2016). These digital spaces provide 24/7 accessibility, transcending geographical boundaries and time zones (Barak, Boniel-Nissim, & Suler, 2008). Also, they offer a cloak of anonymity that can encourage individuals to seek help without fearing social repercussions (De Choudhury & De, 2014). Most importantly, they connect people who share similar experiences, fostering a sense of community and understanding that can be difficult to find in offline settings (Naslund et al., 2016). At the same time, online mutual-help communities are not without drawbacks. For instance, the quality of shared information can be variable, with the potential for rapid spread of misinformation (Naslund et al., 2016)

* Corresponding author at: Department of Computer Science, University of Pisa, Pisa, Italy.
E-mail address: virginia.morini@phd.unipi.it (V. Morini).

https://doi.org/10.1016/j.chb.2024.108544
Received 4 August 2024; Received in revised form 12 December 2024; Accepted 17 December 2024
Available online 24 December 2024

or negative interactions, such as harassment or persistent exposure to triggering content (Saha, Ernala, Dutta, Sharma, & De Choudhury, 2020). Moreover, the lack of professional oversight raises questions about the appropriateness of the advice, particularly for individuals with severe MH conditions (Baek, Bae, & Jang, 2013).

Despite these concerns, the growing popularity of these online communities has increased, particularly in the wake of the COVID-19 pandemic (Merchant et al., 2022), which has exacerbated MH issues globally. This trend underscores their importance in the current MH landscape, as they continue to attract users seeking support, connection, and understanding (Proferes, Jones, Gilbert, Fiesler, & Zimmer, 2021). Given the complex nature of online MH communities, understanding how these digital spaces function in practice is essential. This process involves examining user behavior within these spaces, identifying the needs users seek to fulfill, and evaluating whether MH communities effectively provide the intended support (Berry et al., 2017; Bucci, Schwannauer, & Berry, 2019).

The advent of big data analysis techniques and models based on artificial intelligence enables researchers to access and analyze large-scale data from these communities while preserving user anonymity (Chen & Xu, 2021; De Choudhury & De, 2014; Hopfgartner, Ruprechter, & Helic, 2022). These tools, combined with a multidisciplinary approach incorporating psychological theories and offline evidence (Joseph, Citraro, Morini, Rossetti, & Stella, 2023), provide an unprecedented opportunity to gain insights into user interactions (D'Agostino et al., 2017; Morini et al., 2024), content patterns (Garg, 2024; Garg et al., 2022), and community dynamics (Cunha, Weber, Haddadi, & Pappa, 2016) within these online environments. The findings arising from such research are crucial for improving the design and moderation of these online spaces (Saha et al., 2020), and for informing MH professionals and policymakers about the role these communities play in the landscape of MH support (Naslund et al., 2016).

1.1. Reddit mental health communities

As discussed above, online mutual-help communities have the potential to serve as valuable spaces for individuals facing MH challenges by providing an environment where users can share experiences, offer support, and interact with others facing similar difficulties (Naslund et al., 2016). Among social media platforms, Reddit has emerged as a significant hub for MH discussions and support (Proferes et al., 2021). Reddit's structure of topic-specific communities, known as subreddits, combined with features like pseudonymity through usernames, full anonymity through throwaway accounts, and the ability to write longer posts of up to 40,000 characters, make it particularly suited for in-depth, self-disclosure and personal discussions (De Choudhury & De, 2014; De Choudhury, Kiciman, Dredze, Coppersmith, & Kumar, 2016; Park, Conway, & Chen, 2018). Another crucial feature of Reddit's communities is the presence of moderators and subreddit-specific guidelines. Moderators, often individuals with personal experience in the topic discussed, maintain community standards by enforcing rules, removing inappropriate content, and providing crisis support resources (Saha et al., 2020). This peer-led structure, especially in MH communities, could enhance interaction quality and promote adherence to community norms. Collectively, these characteristics could contribute to creating a supportive environment, which may explain why MH-focused subreddits are among the most active communities on Reddit, surpassed only by subreddits dedicated to political and news discussions (Proferes et al., 2021).

From a research perspective, Reddit's structure offers the opportunity to study – in a data-informed way – social support mechanisms and self-disclosure attempts in online MH discourse. Indeed, the availability of large-scale data has facilitated, over the last decade, analyses of interaction patterns, language use, and community dynamics across various MH subreddits (D'Agostino et al., 2017; De Choudhury & Kiciman, 2017; Garg, 2024; Hickey et al., 2023; Joseph et al., 2023). Furthermore, Reddit's diverse ecosystem enables comparative studies between MH-focused subreddits and other community types (Low et al., 2020), providing insights into the distinctive characteristics of MH discussions and support-seeking behaviors.

1.2. Posting intents in mental health communities

A crucial aspect of MH communities is self-disclosure, which involves revealing personal information to others to express feelings, build trust, and establish intimacy (Cozby, 1973). This self-disclosure primarily occurs in online settings through user-generated posts, where individuals write about their experiences, thoughts, and emotions. These posts serve as the primary medium for users to share their stories and engage with the community.

Motivations for disclosing span between an interpersonal-intrapersonal continuum (Luo & Hancock, 2020). Interpersonally, disclosures aim to foster intimacy and connection, which is particularly significant for individuals experiencing loneliness or social anxiety. For example, a user might share their struggles with social situations, seeking to connect with others who have similar experiences (Berry et al., 2017). Intrapersonally, self-disclosure serves as a mechanism for releasing emotions and reducing stress, especially for those with high stress or low self-esteem, who find a safe space for expression in online social platforms like Reddit.

Psychological literature and empirical studies on online MH groups identify various specific intents behind online self-disclosure. One of the most common is *seeking support* (Cutrona & Suhr, 1992), primarily an interpersonal motive. Social support is understood as the degree to which an individual feels assured of being loved, valued, and able to rely on others when needed. In MH communities, individuals often seek various types of support, such as practical advice or information on managing symptoms, emotional empathy, understanding, and validation of their feelings and experiences (Cutrona & Suhr, 1992). Another common intent is *offering help* to others that complements the motive of seeking support. Findings from different works (Chen & Xu, 2021; Cunha et al., 2016), indicate that Reddit users who receive social support are more likely to continue disclosing personal information, seeking help, and offering support to others in the future. This positive feedback aligns with social learning theory (Bandura & Walters, 1977), which suggests that in social settings, individuals learn through observation and emulation of others.

At the opposite end of the continuum, *venting emotions* is a distinct intent characterized by an intrapersonal nature. Venting involves expressing emotions freely, often linked to the disinhibition effect observed in online settings (Suler, 2004). This might involve posts expressing frustration, anger, or sadness without necessarily seeking specific advice or support. Disinhibition is frequently associated with increased self-disclosure (Derlega, Metts, Petronio, & Margulis, 1993), and can be benign or toxic, depending on the emotional tone and the consequences of the disclosure on the community. Lastly, users may post with the intent to *share their progress*. This intent straddles the line between interpersonal and intrapersonal motivations. Sharing success can aim to motivate others and celebrate one's own achievements. For example, a user might post about successfully completing a week of therapy, overcoming a phobia, or maintaining a consistent medication routine. According to D'Agostino et al. (2017), a significant portion (15%) of posts discussing addiction and substance abuse are focused on sharing successful or positive experiences during recovery.

Understanding the actual motivations behind users' posts is crucial for the evolution of these digital support spaces. Indeed, this knowledge can inform the design of tailored community guidelines, help moderators offer effective support, and provide MH professionals with valuable insights into online help-seeking behaviors.

1.3. Community response in mental health communities

Another key component of MH communities is the nature and quality of community response, which refers to the interactions and feedback users receive after posting content. This can be measured through various metrics, including the volume of comments and reactions, as well as the tone and quality of these interactions. According to the Social Support Behavioral Code (Cutrona & Suhr, 1992), community responses can be categorized into several types of support, such as emotional (expressions of empathy and understanding), informational (advice or information), instrumental (tangible help or services), and network (connecting users with similar experiences).

The dynamics underlying these interactions can impact the supportive atmosphere and the effectiveness of these communities, as demonstrated by recent research that has shed light on the factors influencing users' continued engagement. Among others, Chen and Xu (2021) and Cunha et al. (2016) identified the perceived empathy of community responses and the volume of comments received as key determinants of users' likelihood to return and participate. Conversely, Saha et al. (2020) observed that negative or dismissive responses might discourage participation or even exacerbate MH concerns.

These findings align with the reciprocity norm in social psychology, as described by Ferster and Skinner (1957), which posits that individuals who receive support are more inclined to reciprocate it in the future. This positive reinforcement acts as a reward, contributing to the ongoing vitality and supportive nature of online MH communities. Hence, by identifying the aspects of community response that push users to participate again in the discussion, it is possible to enhance the retention of users who seek MH support.

1.4. Research challenges and study aims

While extensive research has been conducted on Reddit's MH communities (De Choudhury & Kiciman, 2017; De Choudhury et al., 2016; Park et al., 2018; Saha et al., 2020), certain aspects remain unaddressed in the understanding of distinctive user behavioral patterns, posting intents, and community response. Drawing from psychology and computer science, this study aims to address these challenges by introducing and analyzing a large-scale dataset of 67 MH-related subreddits, categorized based on the Diagnostic and Statistical Manual of Mental Disorders- 5th Edition (DSM-5) (APA, 2013).

Firstly, most existing studies have focused on a single community or a limited number of communities, lacking a comprehensive analysis of the diverse mental disorders-related subreddits. Through this contribution, we aim to face this limitation by examining a wide range of Reddit communities dedicated to MH discourse. Additionally, community-specific behavioral patterns are rarely compared to other types of online communities, making it difficult to determine whether these patterns are community or platform-dependent. Accordingly, we introduce six additional datasets encompassing a range of online discussion forums, both support-related and general-interest. Thus, our first research question seeks to uncover these aspects:

RQ1: *What are the distinguishing behavioral patterns of different Reddit mental disorder support communities? How do they compare to non-MH-related Reddit communities and offline behaviors?*

Secondly, previous studies have provided insights into users' posting intents mainly focusing on the 'seeking support' motivation. To address this gap, we leverage Reddit's post flairs – labels used by users and moderators to indicate specific post purposes – to train an automated classifier. This classifier allows us to determine posting intents across our dataset, providing insight into users' actual motivations for self-disclosure. Therefore, our second research question is:

RQ2: *What motivates users to open a discussion in MH communities? Is their posting intent consistent over time? Does community response depend on the user's posting intent?*

Lastly, while some studies have examined the impact of community responses on users' subsequent behavior, there is a need for a more thorough understanding of how these interactions shape user follow-up engagement across different MH subreddits. By capturing the volume and tone of community responses to users' posts, we investigate how these factors affect future posting and commenting frequency and the likelihood of users remaining in their community or switching to another one. This leads us to our third and final research question:

RQ3: *How does community response impact subsequent user behavior and engagement in MH communities?*

The remainder of this paper is structured as follows: Section 2 outlines our methodology for data collection and analysis. Section 3 presents our results, addressing each research question in turn. Finally, Section 4 discusses our findings, their implications, and limitations of the study, and suggests directions for future research.

2. Methods

In this section, we outline the data and methodologies used to investigate our primary research objectives: (i) understanding how Reddit MH support communities structure themselves and are different from not MH-related online groups; (ii) investigating the intents behind user postings in the platform, and the resulting community response, and (iii) studying how this influences users' subsequent behavior. Fig. 2 provides an overview of our analytical process, illustrating the three main approaches employed to address these research aims.

2.1. Data description

As previously described, Reddit is a social platform featuring user-generated content organized into thematic communities known as *subreddits*. Users can submit *posts* and engage in discussions through *comments*, which can be direct responses to posts (first-layer comments) or replies to other comments (second-layer comments and beyond), creating a hierarchical structure. Participants express approval or disapproval through *upvotes* and *downvotes*. The net result of these votes is reflected in a post or comment's *score*, which influences its visibility. Reddit offers unique features that distinguish it from other social platforms. First, users can write longer posts and comments, facilitating more detailed discussions. The platform also allows users to maintain anonymity through *throwaway accounts*, typically created for temporary use to ask sensitive questions or share personal stories without linking them to their primary account. Additionally, Reddit enables users to categorize *posts* using *flairs*, which are tags typically defined by the moderators of each subreddit. Flairs, such as 'medications' or 'hospitalization', help identify the topic of discussion, while others like 'need advice' or 'seek support' indicate the poster's intent. Moderators can assign or modify flairs after a post is published, ensuring proper categorization. However, users cannot apply multiple flairs to a single post. In Fig. 1, we provide a toy example of a Reddit thread highlighting the above-mentioned features.

The datasets used in this work were collected through the `pushshift.io` Reddit API (Baumgartner, Zannettou, Keegan, Squire, & Blackburn, 2020)[1] and consist of posts and comments shared on the platform over five years, from 01/01/2018 to 31/12/2022. Data were cleaned by removing duplicated and empty data entries, content from deleted user accounts, and contributions from subreddit moderators or identified Reddit bots.[2] Furthermore, we anonymized all data to prevent any re-identification of online users, ensuring the privacy and ethical handling of the stored information. In the following, we describe in detail the Reddit MH dataset and the Reddit comparison dataset.

[1] https://github.com/pushshift/api.

[2] https://botrank.pastimes.eu/.

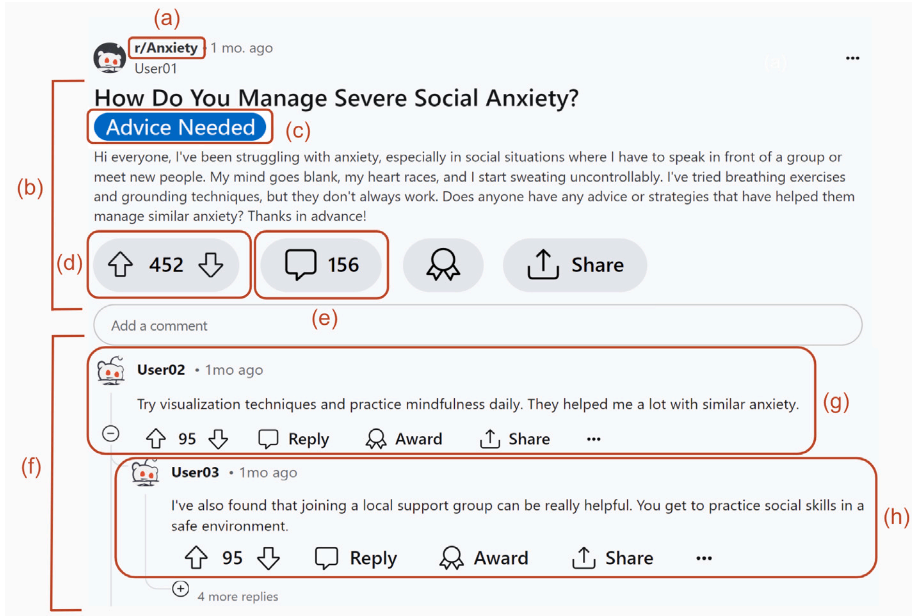

Fig. 1. Anatomy of a reddit thread. (a) the subreddit in which the post has been shared; (b) the post; (c) the post flair; (d) the score of the post that consists of a fuzzy estimate of upvote (up arrow) minus downvote (down arrow); (e) the number of comments on the post; (f) post's comments; (g) first-layer comment; (h) second-layer (nested) comment.

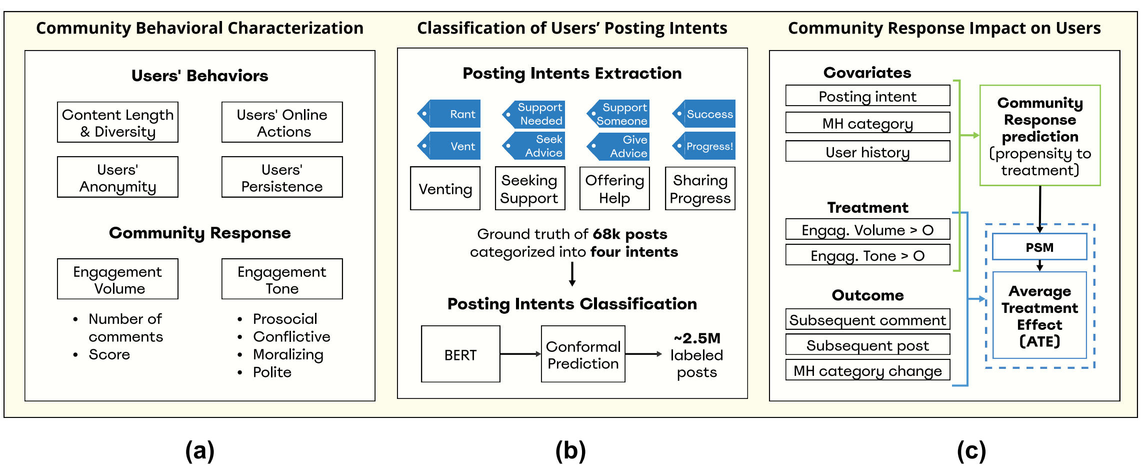

Fig. 2. Overview of the analytical process. The diagram provides an overview of the three key components of the methodology presented in this paper: (a) *Community Behavioral Characterization*, which involves analyzing user behaviors and engagement patterns; (b) *Classification of Users' Posting Intents*, using a BERT model to categorize posts into intents based on a ground truth of Reddit post flairs; and (c) *Community Response Impact Analysis*, evaluating how community engagement influences subsequent user activity through the Propensity Score Matching (PSM) method.

Reddit MH Dataset. To collect an exhaustive list of MH-related subreddits, we adopted as a reference the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition – DSM-5 (APA, 2013). Accordingly, we obtained an initial set of subreddits related to the general MH diagnostic categories (e.g., `r/eating_disorders`) or to specific mental disorders (e.g., `r/AnorexiaNervosa`). We further enriched our dataset by considering subreddits included in our seed set's 'Related Communities' section. After excluding subreddits that were inactive during the study period, had fewer than 5000 subscribers, or were not specifically focused on a mental disorder, we identified a final list of 67 subreddits. These subreddits were grouped into 13 categories according to the mental disorder diagnostic categories specified in the DSM-5. The resulting dataset comprises a total of 2,430,281 unique users, contributing 3,441,212 posts and 24,038,431 comments. A full description of mental disorders categories and specific disorders included can be found in Table 1, while a complete list of subreddits collected for each category is available in Table 1 of *SI*.

Additionally, we collected the posts' flairs as a proxy to capture the intents that drive users to write on the platform. These are not included in the `pushshift` API, hence we used the PRAW python library[3] to collect all flairs from those subreddits in our MH dataset that have the *post flairs* feature enabled in the considered five-year period. By manually assessing the quality and consistency of the extracted flairs, we identified 66 recurrent post flairs that we grouped into four distinct posting intents. *Offering Help* refers to posts in which users provide support, advice, or resources to others. *Seeking Support* encompasses posts from users looking for help, advice, or support, instead. *Venting* includes expressions of frustration, anger, or disappointment, while *Sharing Progress* involves users sharing success or progress stories. A

[3] https://praw.readthedocs.io/en/stable/index.html.

Table 1
Overview of Reddit MH communities. Description of the Reddit MH dataset, in terms of mental disorder diagnostic categories and specific disorders included. Each category is identified by its clinical name, the reference code used throughout the paper, the number of associated subreddits, the specific disorders considered, and their relative prevalence by the percentage of content collected.

Dataset (Code, # of subreddits)	Specific disorders (% of content)
Anxiety D. (ANXI, 8)	General (61%); Social anxiety d. (22%); Specific phobias (13%); Panic d. (3%); Selective mutism (0.4%).
Bipolar and Related D. (BIPO, 4)	Bipolar d. type I (60%); General (23%); Bipolar d. type II (17%).
Depressive D. (DEPR, 4)	Major depressive d. (90%); Dysthymia (0.5%); Postpartum d. (0.5%).
Dissociative D. (DISS, 2)	Dissociative identity d. (87%); Depersonalization-dereализ. d. (13%).
Feeding and Eating D. (FEED, 7)	Binge eating d. (31%); General (25%); Anorexia nervosa (17%); Bulimia nervosa (16%); Avoidant/restrictive food d. (11%).
Neurodevelopmental D. (NEUR, 13)	Attention-deficit/hyperactivity d. (53%); Autism spectrum d. (44%); Tic d. (1.7%); Specific Learning Disabilities (0.9%).
Obsessive-Compulsive and Related D. (OBSE, 6)	Obsessive-compulsive d. (66%); Body dysmorphic d. (10%); Hairpulling d. (9.3%); Compulsive hoarding(7.4%); Excoriation/skin-picking d. (7.1%).
Personality D. (PERS, 6)	Borderline personality d. (74%); Schizoid personality d. (9.5%); Avoidant personality d. (7.7%); Narcissistic personality d. (6.8%); Schizotypal personality d. (1.9%).
Schizophrenia and Other Psychotic D. (SCHI, 2)	Schizophrenia (79%); Schizoaffective d. (21%).
Sexual D. (SEXL, 3)	Vaginismus (35%); Substance-induced sexual dysf. (35%); Erectile dysf. (29%).
Sleep-Wake D. (SLWK, 5)	Insomnia (37%); Breathing-related sleep d. (28%); Narcolepsy (26%); Restless legs syndrome (4.6%); Delayed sleep phase d. (4.1%).
Substance-Related and Addictive D. (SUBS, 5)	Drugs addiction(87%); General (7.1%); Alcohol use d. (4.8%); Gambling d. (0.8%).
Trauma- and Stressor-Related D. (TRMA, 2)	Post-traumatic stress d. (100%).

complete list of categorized flairs is available in Table 2 of *SI*. Only posts tagged with a flair representing one of these four intents – totaling 67,876 posts – were selected as the ground truth to train an automated classifier aimed at identifying the intent of posts lacking a flair (see Section 2.3).

Reddit Comparison Dataset. To determine whether the behavioral patterns observed in MH-related subreddits were specific to the discussion topic or, instead, inherent to the Reddit platform, we extended our analysis beyond MH discourse. We strategically curated diverse datasets to enable comparisons, encompassing both Reddit communities that share similarities with our seed data and those that are distinctly different. These communities cover *General Support* (GENSP, 8 subreddits), addressing seeking advice on broad personal challenges; *Chronic Diseases* (CHRO, 13 subreddits), focusing on long-term health conditions like Diabetes and Alzheimer; *Crafts* (CRFT, 14 subreddits), centered on mutual learning and sharing on creative activities and DIY projects like origami and sewing; *Financial Advice* (FINAD, 10 subreddits), involving financial management advice; *Politics* (POLIT, 6 subreddits), discussing government and societal issues; and *Memes* (MEME, 14 subreddits), used for humor and entertainment. The first two categories could be considered adjacent to our MH dataset, as they provide support on sensitive personal matters. *Crafts* was chosen because it represents a unique niche in which users share common

Table 2
Our ground truth of posting intents. For each considered posting intent: example flairs and the number of posts with those flairs included in the ground truth.

Posting intent	Collected flairs	# P
Offering Help	'Supporting Someone', 'Offering Advice', 'Advice to Give'…	3788
Seeking Support	'Need Support', 'Advice Request', 'Seeking Reassurance'…	24,548
Venting	'rant/vent', 'vent', 'Venting'…	27,650
Sharing Progress	'Success!', 'Good News', 'Recovery Story'…	11,890

hobbies, while *Financial Advices* was selected for its tailored advice nature but on a topic unrelated to diseases. *Politics* was chosen instead for its high engagement on sociopolitical issues and *Memes* for its volatile nature, ensuring a broad spectrum for analysis.

To ensure a valid comparison, we used the same time period as our MH dataset, and we selected subreddits with a comparable number of subscribers to the MH-related ones. Further, we opted not to exclude users who appeared in both the comparison and MH datasets to present a realistic overview of users' behavioral patterns. Nevertheless, a supplementary analysis, excluding these users, yielded minor variations (less than 3%) on each observed pattern, suggesting the robustness of our results against the influence of shared users. The complete list of subreddits collected for each comparison dataset is provided in Table 3 of *SI*.

2.2. Community behavioral characterization

As shown in Fig. 2(a), to extract unique community behavioral patterns from the Reddit MH and Comparison datasets, we considered several dimensions: (i) **Content length & Lexical diversity:** We computed the average word count for posts and comments to measure the length of discussions in each of them. To assess lexical diversity in both posts and comments, we analyzed Type-Token Ratio growth curves and fitted them to Heaps' Law ($V = KN^{\beta}$, where V is vocabulary size, N is text length, and K and β are free parameters determined empirically) using the first 1M tokens from each category. (ii) **Users' online actions:** Users were categorized based on their activity as either posting only, commenting only, or engaging in both activities. (iii) **Users' anonymity:** We identified throwaway accounts by following the approach presented in De Choudhury and De (2014), which consists of using regular expression matching for the term 'throw*' within usernames. (iv) **Users' persistence:** We categorized users as either occasional contributors (participating once via a post or comment) or active members. (v) **Community response:** We analyzed the volume and tone of social interactions (i.e., comments) for each post. Specifically, we considered Posts' Engagement Volume (the number of comments and the post's score) and Posts' Engagement Tone. The tone was assessed using the 'social behavior' dimension provided by the psycholinguistic lexicon LIWC,[4] which consists of *prosocial behaviors*, *moralization*, *interpersonal conflict*, and *politeness* linguistic markers. As reported in the LIWC documentation (Boyd, Ashokkumar, Seraj, & Pennebaker, 2022), the values of these linguistic markers computed across different social media textual data never exceed 2. For each post, we computed these four indicators by aggregating their first-layer comments with a minimum length of six words. It is noteworthy that when computing LIWC scores on our MH dataset, most engagement tone dimensions frequently registered zero values – politeness (76.44%), interpersonal conflict (74.44%), and moralization (69.75%) – indicating their minimal presence in the comments analyzed. Therefore, except for prosocial behaviors, results regarding these dimensions are reported in

[4] https://www.liwc.app/.

Table 3
Behavioral metrics across Reddit communities. For *MH Reddit Datasets* (a) and *Comparison Reddit Datasets* (b) the number of posts (P), the number of comments (C), the post length (PL), the comment length (CL), the number of users (U), the percentage of users that only wrote comments (OCU), the percentage of users that only wrote posts (OPU), the percentage of users that wrote both posts and comments (BU), the percentage of users that wrote using a throwaway account (ThrwAcc), and the percentage of occasional users (OccU).

(a) MH Reddit datasets										
	# P	# C	Avg PL	Avg CL	# U	% OCU	% OPU	% BU	% ThrwAcc	% OccU
ANXI	477k	2.2M	145	50	457k	49	18	32	1.3	42
BIPO	256k	2.1M	131	49	148k	47	11	42	1.2	31
DEPR	661k	2M	166	52	565k	38	29	33	2.3	46
DISS	43k	257k	160	73	29k	41	18	41	1.5	36
FEED	100k	482k	133	54	89k	50	15	35	1.2	39
NEUR	752k	7.9M	145	57	675k	54	11	35	0.8	33
OBSE	172k	835k	149	51	155k	50	15	35	1.8	39
PERS	227k	1.4M	171	60	168k	49	13	38	1.7	34
SCHI	68k	448k	104	43	40k	44	19	37	1.1	39
SEXL	35k	217k	123	46	28k	45	18	36	2.1	39
SLWK	76k	611k	133	57	87k	52	13	35	0.6	38
SUBS	405k	4.1M	117	38	439k	54	15	30	1.4	38
TRMA	171k	1.5M	227	73	132k	57	10	34	1.7	35
MH	**3.4M**	**24M**	**149**	**53**	**2.4M**	**46**	**16**	**37**	**1.6**	**35**

(b) Comparison Reddit datasets										
	# P	# C	Avg PL	Avg CL	# U	% OCU	% OPU	% BU	% ThrwAcc	% OccU
CHRO	303k	2.8M	113	55	266k	48	15	37	0.6	35
CRFT	742k	5.5M	31	25	517k	57	7	36	0.2	34
FINAD	694k	8.4M	108	48	785k	50	16	34	1	37
GENSP	1M	7.1M	192	52	976k	33	25	42	3.3	40
MEME	1.3M	8.8M	6	22	1.6M	73	11	16	0.2	44
POLIT	879K	20.7M	25	42	829k	80	8	12	0.4	42

RQ1 – when comparing the MH dataset with Comparison datasets – but are excluded from the RQ2 and RQ3 analyses.

Descriptive statistics and basic frequency analyses were computed for each of the above-mentioned dimensions. Furthermore, in the community response analysis, we compared the distributions of the considered engagement dimensions using Jensen–Shannon divergence to identify similarities and differences in engagement patterns across different Reddit communities.

2.3. Classification of users' posting intents

To determine users' posting intents, we relied on Reddit post flairs as ground truth for labeling posts and then used an automated classification model to infer the intents of posts without an assigned flair (see Fig. 2(b)). As described in Section 2.1, our ground truth dataset consisted of 67,876 posts categorized into four intents. Table 2 provides examples of the flairs associated with each intent, and the number of posts collected for each. We then split this dataset into two stratified subsets: 90% for training and validation and 10% for testing. To classify the posting intents, we trained a machine-learning model using this labeled dataset. We employed the BERT (Bidirectional Encoder Representations from Transformers) model (Devlin, Chang, Lee, & Toutanova, 2018), specifically the `bert-base-uncased` version from Hugging Face,[5] tailored for sequence classification tasks. As displayed in Table 2, the dataset was quite unbalanced, especially for the *Offering Help* class. Also, a manual inspection of the dataset suggested difficulties in the discrimination between some of the classes, like *Venting* and *Seeking Support*. In order to mitigate possible effects related to these dataset characteristics, we implemented different strategies in our training pipeline and assessed them via a 3-fold stratified cross-validation. We selected the training procedure that obtained the best Macro F1 score. Specifically, we tried all the possible combinations of (i) setting class weights inversely proportional to class frequencies, (ii) oversampling minority classes, and (iii) concatenating posts with their first-layer comments using a special separator token within the constraints of BERT's 512-token limit for sequence length. Other hyperparameters of the training network, such as learning rate, batch size, and the number of epochs, were set via an empirical evaluation to 10^{-5}, 3, and 3, respectively. The configuration that obtained the higher Macro F1 score was the one that adopted all of the three strategies described, obtaining an average of 81% on the three validation sets of the cross-validation.

However, a notable limitation of the model is its potential inability, when applied across the entire dataset of collected MH-related posts, to capture the full spectrum of posting intents that motivate users to engage on the platform. There are, for instance, posts that do not fit into any of the four intents, such as questions about medication and/or treatments. To address this issue, we leveraged the Conformal Prediction method presented in Norinder, Carlsson, Boyer, and Eklund (2014) to quantify the degree of reliability of the obtained predictions and thus categorize posts that presented high uncertainty with a label *Other*. This method involved applying the trained model to a calibration dataset, extracting class probabilities, and ranking them for each true class by prediction probability. Test set probabilities were compared with those of the calibration set to determine their rank and calculate p-values as a measure of conformity. By setting an error rate and comparing the p-values of test samples to those of calibration samples, we were able to identify valid predictions. We determined an appropriate error rate by manually identifying an inflection point in the trade-off between data loss and model performance. Specifically, by setting an error rate of 0.26, we boosted the Macro F1-score from 82.0 to 87.0%, losing 24% of invalid data. Notice that we opted for this strategy since we were strongly interested in having a robust annotation for posting intents to give a realistic overview of Reddit mental disorder-supportive ecosystem. Table 4 in the *SI* presents representative examples of annotated posts for each posting intent.

2.4. Inference of community response impact on users behavior

This section outlines the methodology we followed to measure how community responses to posts impact user behavior within MH communities. As illustrated in Fig. 2(c), we explored whether community feedback influenced future posting, commenting, and transitions to other MH categories. To this aim, we relied on a robust causal inference

[5] https://huggingface.com/.

method that was successfully employed in empirical studies on Reddit MH communities (Chen & Xu, 2021; De Choudhury & Kiciman, 2017), namely Propensity Score Matching (PSM). PSM estimates the effect of a treatment by evaluating the difference between the observed outcomes of suitably-defined treatment and control groups. It involves: (i) selecting appropriate covariates, i.e., variables that could influence both treatment and outcomes; (ii) defining outcomes that reflect treatment effectiveness; and (iii) identifying treatment and control groups, ensuring comparability based on propensity scores, i.e., probability of receiving treatment given the covariates. For example, in evaluating a tutoring program's effectiveness, PSM would match students with similar backgrounds and test scores (covariates), where one received tutoring (treatment) and one did not (control), to assess the program's impact on future grades (outcome). This controls for variables affecting both treatment likelihood and outcomes. Following this matching principle, our implementation of PSM is described in Algorithm 1.

Algorithm 1: Propensity Score Matching

Input: Set of posts P, treatments T, covariates C, outcomes O
Output: Average Treatment Effect (ATE) and significance for each outcome in O

Definitions:
- **Covariates C:** Post intent, MH category, users' history (i.e., # of prior posts/comments).
- **Treatments T:** Engagement volume (# comments > 0 and post score > 0), engagement tone (prosocial behaviors > 0).
- **Outcomes O:** Subsequent post, Subsequent comment, MH category change (yes/no).

Step 1: Calculate Propensity Scores
foreach *post* $p \in P$ **do**
  Compute propensity score using a classification model;
  Assign p to treatment or control group based on T thresholds;

Step 2: Match Treatment and Control Groups
Sort posts by propensity scores;
Stratify posts into quartiles;
Verify balance of covariates C between treatment and control groups using standardized mean differences;

Step 3: Estimate Treatment Effects
foreach *quartile* **do**
  Compute ATE as the difference in mean outcomes O between treatment and control groups;
  Perform significance testing using t-tests;
  Compute confidence intervals for ATE;

Step 4: Evaluate Results
return ATE, significance, and confidence intervals for all outcomes;

We noted that PSM required a propensity score, which measures the propensity of an item to be treated, and is computed using a classification model. Our Random Forest model for classification achieved the best performance among tested classifiers (Logistic Regression, XGBoost, Multilayer Perceptron), yielding an AUC of 74% for scores received per post, 77% for the number of comments received per post, and 70% for prosocial behaviors as identified in the first-layer comments to posts.

3. Results

In this section, we present our findings addressing the three primary research questions of our study. First, we explored the distinctive behavioral patterns observed in Reddit MH communities, comparing them with non-MH-related subreddits and offline behaviors. Next, we focused on users' intent for initiating discussions in MH communities, examining the consistency of their posting intents over time and how community responses vary based on these intents. Finally, we analyzed the impact of community responses on users' subsequent behavior and engagement within these online spaces.

3.1. Dynamics of reddit mental health communities

We next describe the distinctive behavioral patterns of various Reddit MH communities according to the different dimensions detailed in Section 2.2.

Content Length & Lexical Diversity. Analyzing content length across posts and comments, we observed distinct writing behaviors depending on the MH disorder observed, as shown in the fields post length ('PL') and comment length ('CL') in Table 3(a). For instance, participants of certain communities (e.g., *Trauma and Stressor-Related D.*) tended to have longer comments and posts ('PL' = 226.8 and 'CL' = 73.2), suggesting the complexity of trauma-related experience being shared, whereas others (e.g., *Schizophrenia and Other-Related D.*) demonstrated a predilection for relatively shorter content ('PL' = 103.8 and 'CL' = 42.7). The lexical diversity analysis yielded β values of approximately 0.50 (± 0.01) for posts and 0.51 (± 0.03) for comments, typical of English language texts.

Users Online Actions. As shown in Table 3(a), across all MH communities, the most common user profile was the only commenter ('OCU'), i.e., people who never initiated new posts but commented on other people's posts. This is particularly true for *Trauma and Stressor-Related D.* (56.9%), *Substance-Related and Addictive D.* (54.5%) and *Neurodevelopmental D.* (54.3%). In contrast, communities related to *Depressive D.* stood out, as they displayed the highest percentage of users who solely created posts ('OPU'), i.e., 28.82% and surpassed that of the second-highest community by about 10%. Instead, we did not observe strong differences in the percentage of users who both posted and commented ('BU'), i.e., from 30.3% to 42.5%, which might reflect a balance between creating new posts and reacting to other people's posts in all the considered communities.

Users Anonymity. Although the percentage of throwaway accounts was relatively low across all communities, ranging from 0.6% to 2.3% (see 'ThrwAcc' column of Table 3a), we observed some interesting patterns. Certain disorders, such as *Depressive D.* (2.3%) and *Sexual Dys.* (2.1%), showed a relatively higher percentage of throwaway account usage, suggesting a heightened desire for anonymity when discussing particularly stigmatized health conditions.

Users Persistence. We found that, on average, a consistent number of users across all MH groups (35.2%) tended to interact only once (see occasional users 'OccU' column in Table 3(a)). However, communities dealing with disorders such as *Depressive* and *Anxiety D.* presented a larger base of occasional users. Conversely, communities such as *Bipolar and Related D.* and *Neurodevelopmental D.* (31.2% and 33.3% respectively), displayed lower percentages of occasional users.

3.2. Mental health communities vs. Other subreddits

To understand whether the observed behavioral patterns across MH communities were inherent to the topic discussed or instead depended on the platform used, we repeated the analysis for the comparison datasets (i.e., *Chronic Diseases*, *Crafts*, *Financial Advices*, *General Support*, *Memes*, and *Politics*). Details are provided in Table 3(b). Also, for the sake of readability, we aggregated the 13 MH communities into a single dataset and recomputed statistics. See MH in bold in Table 3(a) for details.

Content Length & Lexical Diversity. Regarding the average length of content shared, communities oriented to providing support/advice and discussing sensitive and personal issues tended to produce fairly longer content, probably due to the necessity of providing detailed explanations and personal anecdotes typical of self-disclosure attempts (i.e., 108.6 ≤ PL ≤ 191.7 words and 47.6 ≤ CL ≤ 54.5 words). Conversely, *Crafts*, *Politics*, and particularly *Memes* ('PL' = 5.6 and 'CL' = 21.7) tended to prioritize succinct communication in posts and then fostered discussion through comment — see 'CL' and 'PL' columns in

Table 4
Community response in MH and non-MH Reddit communities. For the MH Dataset and Comparison Datasets, the volume, and tone of community responses to posts shared on the platform. We included the average and standard deviation (std) of the number of comments received and post scores. The tone of comments was described through the mean and std of LIWC 'social behavior' dimensions, extracted from the first-layer comments on posts.

Posts' engagement volume								
	Description	MH	CHRO	GENSP	FINAD	CRFT	POLIT	MEME
# Comments	# of comments received to a post.	8.4 (±25.0)	9.8 (±15.5)	6.3 (±19.7)	12.9 (±36.0)	10.2 (±20.4)	24.5 (±104.4)	11.0 (±54.2)
Post Score	The score received to a post (# of upvotes - # of downvotes).	25.9 (±150.7)	20.2 (±60.8)	9.8 (±116.5)	10.3 (±56.6)	140.0 (±512.5)	86.9 (±835.9)	562.3 (±2324)
Posts' engagement tone								
	Description	MH	CHRO	GENSP	FINAD	CRFT	POLIT	MEME
Prosocial Behaviors	Referents that signal helping or caring about others (Penner, Dovidio, Piliavin, & Schroeder, 2005).	1.0 (±1.7)	1.0 (±1.4)	1.1 (±1.7)	0.6 (±1.3)	0.5 (±1.3)	0.6 (±1.6)	0.5 (±1.8)
Moralization	Words reflecting judgment or moral evaluation about another's behavior (Brown & Levinson, 1978).	0.2 (±0.7)	0.1 (±0.5)	0.4 (±0.9)	0.2 (±0.6)	0.1 (±0.5)	0.6 (±1.6)	0.4 (±1.6)
Interpersonal Conflict	Words referring to concepts suggestive of conflict (Barki & Hartwick, 2004).	0.2 (±0.6)	0.1 (±0.5)	0.3 (±0.8)	0.1 (±0.4)	0.1 (±0.4)	0.6 (±1.4)	0.5 (±1.6)
Politeness	Referents to adherence to social norms and manners (Brady, Crockett, & Van Bavel, 2020).	0.2 (±1.0)	0.4 (±1.1)	0.2 (±0.8)	0.2 (±0.7)	0.3 (±1.1)	0.2 (±1.1)	0.3 (±1.7)

Table 3(b). Additionally, we observed that the lexical diversity in non-MH categories (0.52 ± 0.03 for postings, 0.51 ± 0.02 for comments) was similar to that of MH categories, which in turn was roughly similar to the one of regular English texts. We interpreted this result as showing that besides per-category differences in posting/content length, there were no large variations in terms of the richness of vocabulary across categories (MH and non-MH).

Users Online Actions. As shown in the 'OCU' (Only Commenter Users), 'OPU' (Only Poster Users), and 'BU' (Both poster and commenter Users) columns in Table 3(b), the most common user profile across all topics was the 'only commenters' (with the slight exception for *General Support*), thus suggesting a pattern inherent to the Reddit platform. However, when looking more in-depth at the single behaviors, we were able to unveil interesting trends. Indeed, *Politics* and *Memes* revealed a majority of users only commenting (80.4% and 72.5% respectively), suggesting again a culture of discussions on the same topic rather than opening different threads. On the contrary, communities focused on giving support or advice showed a high presence of users that both posted and commented (e.g., *General Support* 41.8%, *MH* 37.5%). Following this trend, support-oriented Reddit groups reported a higher presence of 'only posters' users with respect to other less sensitive topics (e.g., *General Support* 25.3%, *MH* 16.4%, *Financial Advice* 16.4%).

Users Anonymity. The claim that the anonymity offered by Reddit throwaway accounts fosters disclosing sensitive, often stigmatized issues was further supported by our data. Indeed, as shown in the 'ThrwAcc' column in Table 3(a), *MH* displayed a higher percentage of throwaway account usage (1.6%) compared to other topics – Table 3(b) – second only to another support-related community, i.e., *General Support* (3.3%). Further, topics like *Memes* (0.16%), *Crafts* (0.19%), and *Politics* (0.40%) showed negligible use of such accounts, reflecting less need for anonymity due to the less intimate subject matter.

Users Persistence. When examining user persistence on the platform about the discussed topic, the observed percentages of occasional users ('OccU' in Table 3) exhibited a modest variation, with a disparity of approximately 10% between the highest (*Memes* 44.4%) and lowest value (*Crafts* 34.3%). This narrow range might reflect a tendency among several Reddit users to engage only sporadically, contributing once before becoming inactive. Nevertheless, even within this trend, distinctions emerge between topics related to personal or sensitive issues and those that are not. Notably, *Memes* and *Politics* had the highest percentages of occasional users at 44.4% and 41.7%, respectively, while *Chronic Diseases* and *Crafts* had the lowest at 34.6% and 34.3%, respectively, suggesting that personal involvement in the topics discussed led to increased participation on the platform.

Community Response. Table 4 shows that the response volume within *MH* communities was marked by an average of 8.4 comments per post and an average post score of 25.9, denoting a moderate but meaningful level of interaction that was closely paralleled by the *Chronic Diseases* community. On the other end of the spectrum, the *Politics* community displayed a significantly elevated average of 24.5 comments per post and a high degree of variability (±104.4), underscoring a robust commenting culture that distinctively set it apart from other topics, as highlighted by the Jensen–Shannon divergence values in Fig. 3. In stark contrast, the *Memes* community was characterized by a fairly high average post score of 562.3, accompanied by a remarkable standard deviation of ±2324.0, likely mirroring the casual and viral dynamics of such content. Across the boards, it was evident that post scores are subject to significant fluctuations across all topics, indicating that the community's reception of content can be highly variable and potentially influenced by the nature of the content itself.

The response tone delineated more pronounced differences (see Table 4). The *MH* community displayed an average of 0.99 for prosocial behaviors with a low standard deviation (±1.7), highlighting a consistency in supportive interactions. This was closely mirrored in other mutual-help communities such as *Chronic Diseases* and *General Support*, with averages of 0.95 and 1.05, respectively. This contrasted with the lower averages and Jensen–Shannon divergence values observed in non-mutual-help topics (see Fig. 3). Furthermore, the tendency for moralization and judgment in comments was comparably higher in the *Politics* and *Memes* communities, with averages of 0.63 and 0.43, respectively. This contrasted with the moderate presence of moralization in *Mh* and *General Support* online groups. The trend for interpersonal conflict aligned similarly, with *Politics* and *Memes* showing higher average levels at 0.56 and 0.50, respectively. Noticeably, the dimension of politeness peaked in the *Chronic Diseases* community at an average of 0.36 but did not display significant differences among other topics, which all exhibit moderate values.

3.3. Posting intents & community response

In this section, we explore possible motivations behind users' decisions to initiate discussions in different Reddit MH communities.

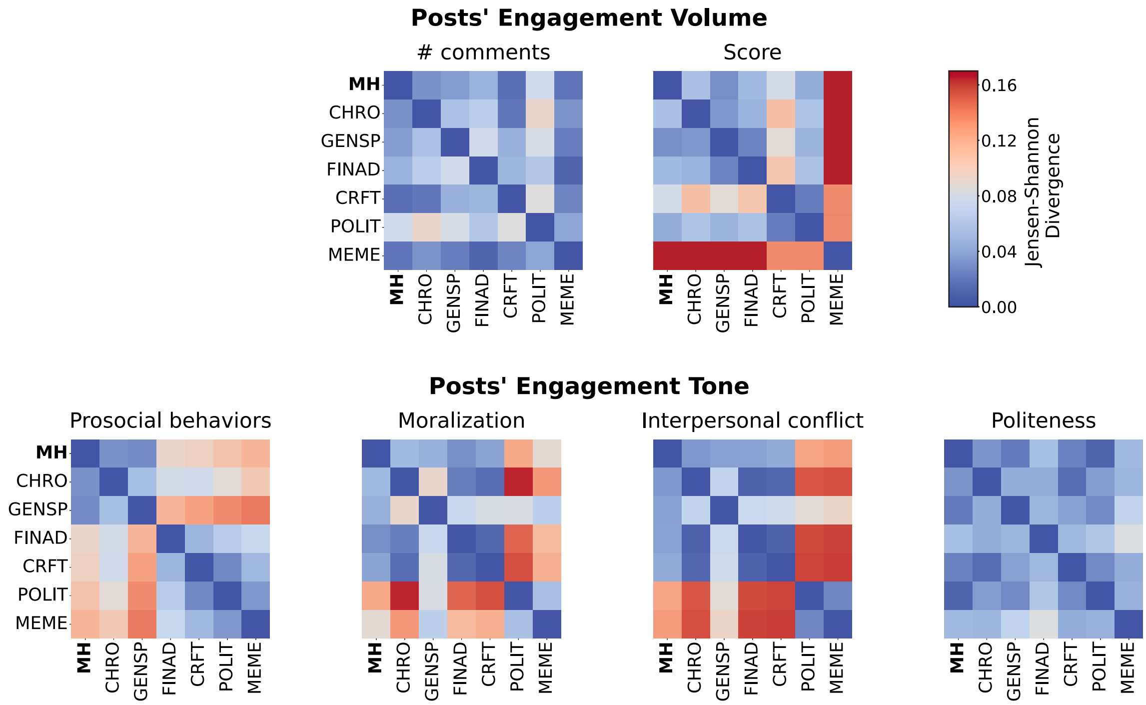

Fig. 3. Differences in community response between MH and non-MH reddit communities. For each community engagement dimension, heatmap of Jensen–Shannon divergence between distributions in the *Mh* and *Comparison* datasets.

Specifically, we examined the consistency of users' intentions, identifying if they persistently posted with the same intent (potentially suggesting users take certain 'roles'), and how these actions influenced community engagement. Notice that in the following analysis, we discarded posts labeled with the class *Other* due to uncertainty about the posting's intents, which could affect the results' robustness and reliability. Accordingly, we ended up with 2,571,479 labeled posts and 18,308,008 comments.

Posting Intents across MH Communities. At an aggregate level, the desire to seek support emerged as the most prevalent motive for sharing posts, constituting 49% of observed cases. Following this, venting comprised 36% of posts, with offering help and sharing progress representing smaller proportions at 8% and 7%, respectively.

Refining the granularity of the analysis, we identified six groups of communities characterized by similar posting intents frequencies in terms of cosine similarities (as shown in Fig. 4 and in Fig. 1 of *SI*):

1. *Dissociative D.*, *Feeding and Eating D.*, and *Bipolar and Related D.* showed, on average, a strong preference for venting posts (62%) and a lower prevalent tendency to seek support (22%).
2. *Substance-Related and Addictive D.*, *Anxiety D.*, and *Neurodevelopmental D.* predominantly seeked support or advice (59%) with a lower rate of venting (25%).
3. *Personality D.*, *Depressive D.*, and *Schizophrenia and Other Psychotic D.* communities featured a more balanced distribution of posting intents, displaying almost equal interest in both venting (41%) and seeking support (45%).
4. *Trauma- and Stressor-Related D.* and *Obsessive-Compulsive and Related D.* favored posting venting content (48%) over seeking support (39%).
5. *Sexual Dys.*, on the other hand, while showing seeking support and venting intent frequencies similar to other groups (37% and 30% respectively), stood out for a higher presence of posts for sharing progress (22%) and offering help (11%).
6. *Sleep-Wake D.* presented a unique pattern with the highest intent for sharing support-seeking posts (72%), with a significantly lower emphasis on venting (18%).

All groups made an exception for the fourth, exhibited negligible and comparable percentages of posts tailored to offer help and share progress (ranging from 5% to 10%).

Shifts in Users Posting Intents. As a second step, we explored whether users participating in MH communities maintained consistent posting intents over time, akin to adopting specific 'roles' within the community. To this aim, we modeled users' transitions between different posting intents using Markov chains, quantifying the probability of shifting intent over time. To assess the significance of these transitions, we implemented two null models: the first randomized the posting intents while preserving their overall distribution, and the second randomized the order of intents for each user, thus removing temporal dependencies. Each model was simulated 1000 times to generate expected transition probabilities, which were then compared against the observed data. Our results indicated that the transitions observed in the dataset were highly significant ($p < 0.001$) when the original intent distribution was preserved. However, when the temporal order was disrupted, some transitions – specifically from *Sharing Progress* and *Venting* to *Offering Help*, and from *Seeking Support* to *Sharing Progress* – were not statistically significant.

By looking at Fig. 5, it is evident that some users maintained a consistent posting intent, indicative of a fixed role within the community, while others frequently changed their posting intents, suggesting dynamic roles. Notably, users focused on *Seeking Support*, *Venting*, and *Offering Help* exhibited self-transition probabilities of 61%, 55%, and 42% respectively, indicating a preference to continue with their original intents rather than switch. In contrast, users who initially posted about their progress or successes were less likely (23%) to maintain this intent, often shifting to *Venting* (37%) or *Seeking Support* (33%). Similarly, those initially *Offering Help* were somewhat prone to change to *Seeking Support* or *Venting* (26% and 25%, respectively). However, individuals starting with *Seeking Support* or *Venting* were less likely (under 8%) to transition into roles like *Offering Help* or *Sharing Progress*.

Community Response to Different Posting Intents. Here, we investigate whether the level of community response in MH-related subreddits (as defined in Section 2.2) varied based on the intent of postings,

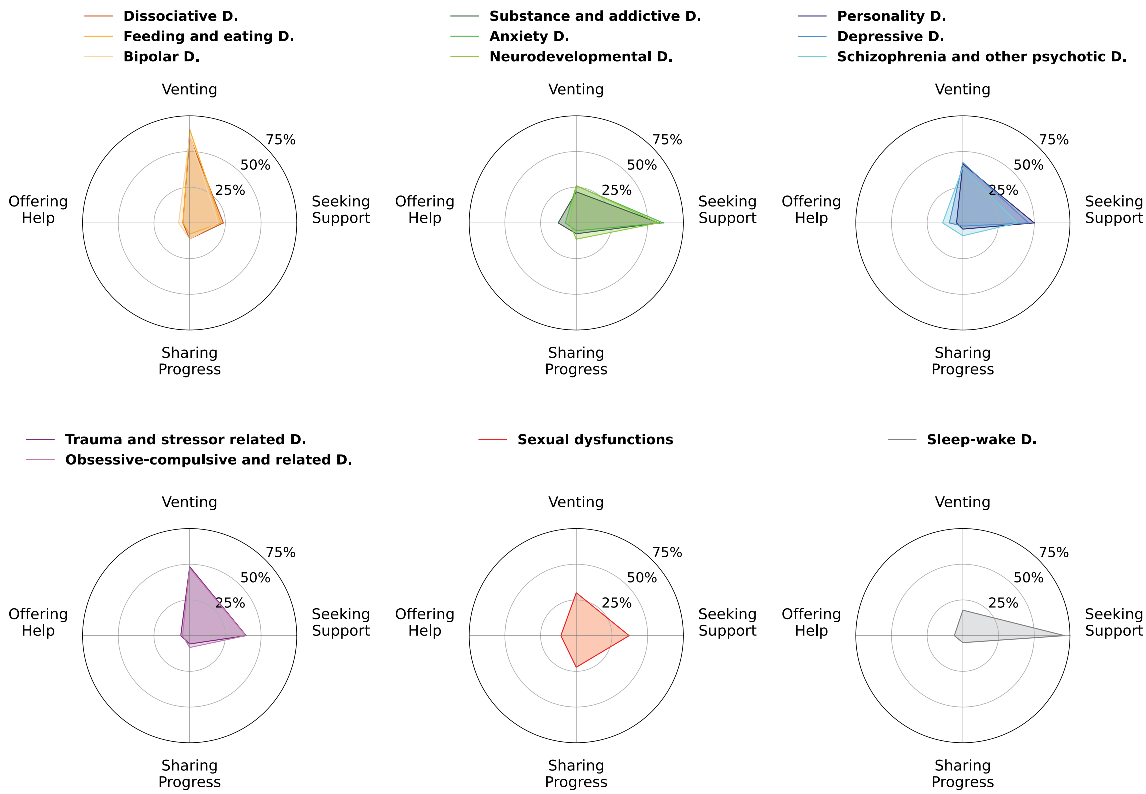

Fig. 4. Posting intent in MH communities. Radar plots display the distribution of posting intents across considered Reddit MH disorder communities. Disorders are grouped together based on similarities in posting intents, showing the relative frequencies of *seeking support*, *venting*, *offering help*, and *sharing progress* for each disorder-related community.

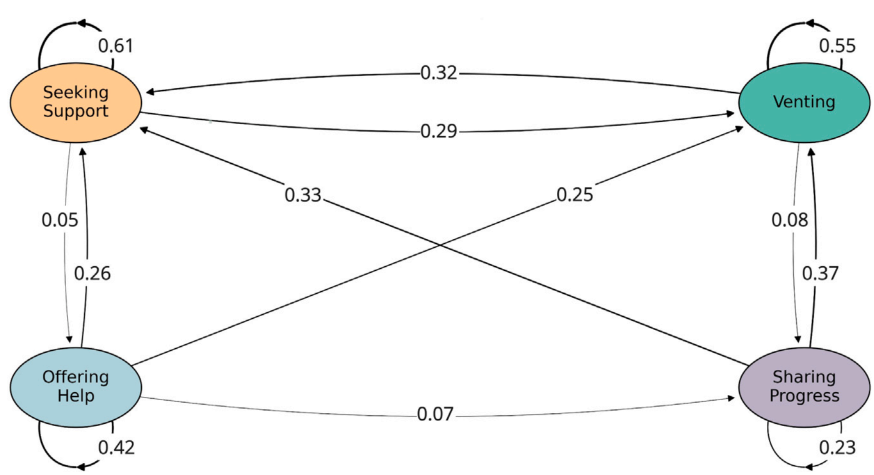

Fig. 5. Transition patterns in users posting intents. The directed graph represents users' transition probabilities between posting intents in MH subreddits. Nodes represent users' posting intents. Edges indicate statistically significant transitions, with their thickness corresponding to the magnitude of transition probabilities.

specifically exploring if certain intents elicited greater engagement. To this end, we computed the distribution of each community response metric across different posting intents for the entire dataset (see Fig. 2 of *SI*), and applied the Kolmogorov–Smirnov test to assess if the differences in distributions were statistically significant. Results confirmed significant variations across posting intents ($p < 0.001$ for all metrics). In terms of scores, posts *Sharing Progress* were the most well-received, with a median score of 9. This was followed by *Venting*, *Seeking Support*, and *Offering Help* posts, with respective median scores of 4, 2, and 1. Conversely, the number of comments tended to be similar across intents, with a median of 3, except for *Offering Help* posts, which had a median of 0 comments, thus confirming themselves as the intent with the lowest volume of engagement. Furthermore, such kinds of posts consistently showed the highest likelihood of receiving zero replies across all 13 MH categories analyzed (see Fig. 3 of *SI*), with only a 39% average probability of receiving at least one comment. This was markedly low when compared to other posting intents, such as *Venting*, which had a 77% probability, *Sharing Progress* at 78%, and *Seeking*

Table 5
Impact of community response on users follow-up behaviors. Average Treatment Effects (ATEs) for treatments *score* and *number of comments* across four quartiles and all users combined. Each quartile's treatment size is indicated in parentheses. Bold values highlight overall ATEs that are consistent across quartiles.

Treatment	Outcome					
		Q1 (47.7%)	**Q2 (66.5%)**	**Q3 (76.3%)**	**Q4 (91.9%)**	**All (70.5%)**
Post Score	Subsequent comment	2.0%	3.0%	8.0%	13.0%	**13.0%**
	Comment in a different MH category	−2.0%	−5.0%	−6.0%	−7.0%	**−6.0%**
	Subsequent post	−5.0%	4.0%	7.0%	16.0%	7.0%
	Post in a different MH category	0.0%	−1.0%	−2.0%	−2.0%	−2.0%
		Q1 (51.1%)	**Q2 (76.7%)**	**Q3 (84.2%)**	**Q4 (94.2%)**	**All (76.5%)**
# Comments	Subsequent comment	4.0%	8.0%	9.0%	14.0%	**17.0%**
	Comment in a different MH category	−4.0%	−8.0%	−9.0%	−9.0%	**−9.0%**
	Subsequent post	−3.0%	8.0%	8.0%	17.0%	8.0%
	Post in a different MH category	−2.0%	−2.0%	−2.0%	−0.0%	**−4.0%**

Support at 82%. Regarding the tone of responses, particularly prosocial behavior markers as expressed in first-layer comments to post, *Offering Help* posts showed the highest median score at 0.62, followed by *Seeking Support* at 0.56, *Venting* at 0.47, and *Sharing Progress* at 0.41.

3.4. Community response & follow-up behaviors

Finally, we explored the impact of community response to users' posts on users' subsequent behavior on the platform. Specifically – as described in Section 2.4 – we relied on PSM causal inference analysis to examine whether users post and comment again or switch the MH category depending on whether they receive engagement to their posts or not. As displayed in Table 5, the analysis revealed several consistent effects across all quartiles. Receiving a reaction (i.e., a score on the post) significantly increased the likelihood of making a subsequent comment, with an average treatment effect (ATE) of 13.0%. Additionally, this engagement reduced the likelihood of making a subsequent comment in a different MH category, with an ATE of −6.0%. Similarly, the number of comments received had a huge impact: receiving at least one comment significantly increased the likelihood of making another comment (ATE of 17.0%) and decreased the likelihood of commenting in a different MH category (ATE of −9.0%). There was also a smaller reduction in the propensity to make a subsequent post in a different MH category (ATE of −4%). These effects were all statistically significant ($p < 0.0001$) with very narrow confidence intervals. However, the effect of engagement volume on subsequent posting behaviors revealed some heterogeneity, particularly in the first quartile. Indeed, the likelihood of making a subsequent post after receiving a reaction generally had a positive ATE of 7%, but notably, the first quartile displayed a negative effect. A similar pattern was observed for posting in a different MH category after receiving a reaction, with an overall slightly negative ATE of −2%, except for the first quartile, which showed a different trend. Receiving comments also had a heterogeneous effect on the likelihood of making a subsequent post, with an overall ATE of 8%, yet again, the first quartile presented differing results. Lastly, the engagement tone indicator considered in this analysis (linguistic markers of prosocial behavior) led to both heterogeneous and minimal treatment effects across quartiles. For this reason, we decided to omit it in Table 5.

4. Discussion

In the following section, we advance answers to our research questions on MH Reddit support communities based on the results presented in the previous section.

4.1. RQ1: What are the distinguishing behavioral patterns of different reddit mental disorder support communities? how do they compare to non-MH-related reddit communities and offline behaviors?

In summary

Similarities in content length, user actions, user persistence, use of throwaway accounts, and community responses were observed between different types of support and advice-oriented communities on Reddit, distinguishing them from non-mutual-help ones. Users in Reddit mental disorder communities often exhibited behavioral patterns reflective of their offline counterparts.

Behavior in online and offline communities. Our study suggests that users' behaviors in Reddit MH communities are, in various ways, consistent with offline communities. For instance, our analysis revealed that people in the *Trauma and Stressor-Related D.* group, which comprises subreddits dedicated to Post-Traumatic Stress Disorder (PTSD), tend to write the longest comments and posts. Previous research (Crespo & Fernández-Lansac, 2016) suggests that longer narratives from individuals affected by PTSD may reflect a need to elaborate and emotionally process traumatic events rather than avoid them, which would typically result in shorter narratives. In these communities, users may process their traumatic experiences by sharing them with others facing similar challenges, incorporating repetitions, adding details, and providing extensive information as a coping mechanism.

Conversely, other MH communities (e.g., *Schizophrenia and Other-Related Disorders*) tend to favor shorter posts and comments, potentially mirroring some offline characteristics of the condition as well. Anomalies in thinking, language, and communication are prominent features of *Schizophrenia and Other-Related Disorders*, particularly the so-called 'negative symptoms', such as poverty of speech, alogia, and a lack of content in speech (Hartopo & Kalalo, 2022).

MH vs. non-MH communities. By comparing such observed patterns with a wide range of Reddit communities (both mutual-help and non-mutual-help), we identify similarities among communities dedicated to support and advice-seeking. Indeed, *MH*, *Chronic Diseases*, *General Support*, and *Financial Advices* were characterized by notably lengthy content, particularly in posts. This extensive self-disclosure is typical of supportive discourse, where personal and sensitive topics often require detailed explanations, personal anecdotes, and nuanced discussions, resulting in longer contributions (Barak et al., 2008; Cozby, 1973; Mohan et al., 2017; Shi & Khoo, 2023).

Posting and commenting behavior. Concerning users' online actions across MH communities, the most common user profile was the only commenter. However, specific patterns may be observed: individuals within the *Substance and Addictive D.* group, for example, exhibited a consistent activity primarily in the form of short comments, ranking as the second most engaged community in terms of commenting frequency across all. This feature may be attributed to shared characteristics inherent in the disorder. Phenomena such as compulsivity, craving, and continued use despite adverse consequences are generally consistent across various types of addictions (APA, 2013). Consequently, the commonality in experiences among users in the community may lead individuals to comment on existing posts rather than create new ones due to the overlap with content that has already been shared.

However, when extending our analyses to the broader Reddit communities, we observe that while users across most communities primarily engaged in commenting – highlighting a common behavior on Reddit – mutual-help communities exhibited a relatively higher proportion of users who both posted and commented. This dual role of support provider and seeker, as corroborated by previous studies (De Choudhury & Kiciman, 2017; Valdez & Patterson, 2022), suggests a collective commitment to building a supportive network, indicating a stronger sense of community and mutual engagement in these forums. Equity theories (Fisher, Nadler, & Whitcher-Alagna, 1983) offer a framework for understanding the balanced exchange of posts and comments in these communities. Within this context of aid relationships, individuals evaluate their contributions (e.g., the support they provide) against the benefits they receive (e.g., the support they get in return). When both parties perceive a balance – such as asking for help through a post and receiving it via comments – the relationship is considered equitable, fostering reciprocity and mutual support.

Engagement over time. Despite the general trend of sporadic participation on Reddit, the relatively higher user retention rates in mutual-help communities support the idea of reciprocal support engagement again. The sustained involvement we witnessed in these communities suggests that the reciprocity of giving and receiving support creates a more engaging and rewarding experience, fostering a sense of belonging and continuous participation (Fisher et al., 1983).

It is important to state, however, that uniform engagement levels were not evident across all MH groups. The *Depressive D.* community, for instance, exhibited the lowest percentage of users actively commenting on existing posts, the highest rate of posts lacking comments, and the greatest prevalence of occasional users. These characteristics portrayed it as a community with relatively low engagement and support. The limited participation within the *Depressive D.* group, potentially associated with the features of the disorder that diminishes interest or pleasure in activities or social interactions (APA, 2013), may contribute to the tendency of users to disengage from the group after a single interaction. According to equity theories (Fisher et al., 1983), the observed lack of reciprocity due to the characteristics of this condition may lead to a perceived imbalance between seeking and receiving help, leading to diminished engagement in the community.

Users seeking to contribute anonymously. Our findings on the distribution of throwaway accounts in MH communities indicated that the observed patterns aligned with the challenges individuals face in offline settings. The highest rates of throwaway accounts were indeed observed in the *Sexual Dys.* and *Depressive D.* groups. This trend is coherent with the offline experience of individuals with such conditions: for instance, people with *Sexual Dys.* often encounter significant social stigma, leading to feelings of shame and embarrassment about their condition (Foster et al., 2022). This, in turn, often discourages them from seeking help and treatment (Moreira et al., 2005). Similarly, narratives about *Depressive D.* frequently remain unshared, as individuals perceive these thoughts as too negative or 'untellable' due to their content, such as suicidal ideation and self-harming behaviors (Yeo, 2021). Consequently, individuals often find a diminished sense of vulnerability when sharing their experiences within mediated environments that offer anonymity. This anonymity facilitates more extensive, spontaneous, and intimate self-disclosure regarding taboo and stigmatized topics, such as sexual impairments or depressive thoughts (Yeo, 2021).

The use of anonymity via throwaway accounts was indeed another distinguishing feature of mutual-help communities on Reddit, particularly in *MH* and *General Support* topics. In contrast, it was scarcely used in topics not related to personal and intimate issues, such as *Memes*, *Politics*, and *Crafts*. Anonymity fosters a more open and supportive environment by reducing the barriers to self-disclosure, which is essential for building trust and facilitating the exchange of support, especially in often stigmatized topics (De Choudhury & De, 2014; Foster et al., 2022; Zent, 2023). These findings align with social support theories, which emphasize the importance of self-disclosure and anonymity in supportive online environments (Berry et al., 2017; De Choudhury & De, 2014; Zulkarnain & Jan, 2019).

Community response. Behavioral patterns were also shaped by community responses. Analyzing responses in mutual-help communities (i.e., *MH*, *Chronic Diseases*, and *General Support*) revealed moderate but significant interactions characterized by consistent prosocial behaviors. Overall, this tendency of mutual-help communities was also depicted by relatively low rates of moralization and interpersonal conflicts. This suggests that users in mutual-help communities engage in behaviors that are cooperative, helpful, and empathetic (Penner et al., 2005), a crucial feature for creating a supportive atmosphere (Soós, Coulson, & Davies, 2022; Tsvetkova & Macy, 2015). Cognitive empathy (i.e., perspective-taking), affective empathy (i.e., shared emotions), and associative empathy (i.e., identification with the target) can, in fact, be elicited by messages revealing personal adversities or challenges, frequently resulting in helping behaviors (Wei & Liu, 2020).

In contrast, non-mutual-help communities, especially *Politics* and *Memes*, exhibited high variability in community response to posts. These communities showed a robust and larger commenting culture and high post scores, but also relatively higher levels of moralization/judgment (Brown & Levinson, 1978) and interpersonal conflict (Barki & Hartwick, 2004). These differences likely reflect the distinct needs and goals of each community. For example, craft and humor-focused communities prioritize creativity and entertainment, fostering a sense of affiliation through shared interests. Political communities, however, often aim to provoke debate and discussion, leading to a large number of replies and potentially more negative tones. Interestingly, politeness in comments, while not high, appeared to be a consistent feature across Reddit, regardless of the topic. This suggests that politeness is more platform-dependent than topic-dependent, possibly reflecting the efforts of Reddit moderators in maintaining civil discourse across diverse communities (Saha et al., 2020).

4.2. RQ2: What motivates users to open a discussion in MH communities? Is their posting intent consistent over time? Does community response depend on the user's posting intent?

In summary

The primary posting intents in MH communities were seeking support/advice and venting, with users generally maintaining consistent roles over time. Notable transitions from sharing progress and offering help to venting and seeking support highlighted the ups and downs of users' journeys. Posts sharing progress received the highest scores/reactions, possibly due to community rewards.

Posting intent. Our findings indicated that the most common posting intents across all MH communities were seeking support/advice and venting. This aligned with core behaviors in mutual-help communities, both offline (in therapy and support groups) and online. Venting emotions could be indeed seen as a specific modality of seeking help: it may function as a form of self-help through emotional regulation or as a way to solicit support from others by communicating one's feelings to the external world (Koole, 2010; Nils & Rimé, 2012). These findings support existing literature, which indicates that users are more inclined to post on platforms when they have specific needs, such as venting emotions, seeking support, or requesting advice, rather than sharing positive news or offering assistance (Berry et al., 2017). This trend highlights the community's overarching preference for empathetic listening and support over advice-giving or celebrating achievements.

Posting intent over time and user roles. By examining users' transitions across different intents over time, we observed that users tended to maintain consistent posting roles within the community, with the notable exception of those who shared progress. This consistency, particularly among users who primarily sought support and those who vented, underscored the role that these communities play in expressing and potentially fulfilling emotional needs. Users may find a stable, supportive environment to seek emotional validation and empathy repeatedly. In addition, our findings presented a significant transition in roles between seeking support and venting, indicating a feedback loop where users seeking help often ended up venting their emotions and vice versa. This could suggest an intertwined relationship between them where the emotional catharsis from venting prompts further support-seeking behavior. This cyclical pattern can be seen as a coping mechanism to manage ongoing stress and emotional challenges (Brown, Westbrook, & Challagalla, 2005; Gloria & Steinhardt, 2016).

Notably, our results exhibited a significant role shift from offering help and sharing progress to venting and seeking support, with the transition from sharing progress to venting exceeding even the likelihood of maintaining the same intent. This pattern aligns with the non-linear nature of MH journeys, characterized by fluctuations in well-being and support needs (Morini et al., 2024; Prochaska & Velicer, 1997). The pronounced transition from sharing progress to venting or seeking support reflects the complex emotional landscape of recovery. While positive experiences can build psychological resources (Fredrickson, 2001), they may also heighten awareness of ongoing challenges. This shift could represent healthy emotional engagement, reflecting increased self-awareness and proactive help-seeking behaviors (Gross, 1998). Recalling and sharing positive experiences can also reactivate memories of past challenges (Kensinger & Ford, 2020), prompting emotional processing through venting or seeking support.

Community response by posting intent. When analyzing community responses to different posting intents, we found no significant differences in general user engagement across intents, indicating that users tended to engage similarly with various types of posts. However, there were two notable exceptions. Posts that shared progress received higher engagement scores. This may reflect a form of social reinforcement, where positive milestones are celebrated and encouraged, fostering a sense of community achievement and belonging. This aligns with Social Exchange Theory, where positive contributions are reciprocated with increased social rewards (Blau, 2017; Homans, 1958). Conversely, posts offering help received the lowest engagement scores and had the highest tendency not to receive replies. Low engagement with help-offering posts could stem from a variety of factors, including the bystander effect, where individuals may assume others will respond, reducing their personal sense of responsibility (Darley & Latané, 1968). Additionally, the nature of the help offered might not resonate with immediate community needs, leading to lower interaction rates. The reciprocity norm also suggests that people are more likely to respond when they receive something in return, which may not be immediately apparent in help-offering posts (Gouldner, 1960).

4.3. RQ3: How does community response impact subsequent user behavior and engagement in MH communities?

In summary

Community responses significantly impacted user behavior in MH communities, with engagement (comments or scores) positively reinforcing continued participation. Users who received reactions were more likely to remain active and stay within a MH community. Comments had a stronger effect on sustaining user engagement than scores, likely due to their qualitative, empathetic nature.

Community response and subsequent behavior. Our findings underlined a significant positive reinforcement effect on user behavior due to receiving reactions or replies to their posts from the community. Indeed, results obtained from causal inference analysis indicated that users who received engagement, whether in the form of scores or comments, were more likely to remain active within the subreddit by commenting and posting again. This result aligns with findings from various mutual-help communities on Reddit. For instance, in gambling self-help communities (Cunha et al., 2016), MH-related subreddits (Chen & Xu, 2021), and weight loss communities (Hopfgartner et al., 2022), receiving comments or upvotes has been shown to increase the likelihood of users making additional posts and comments. Conversely, negative interactions in hateful subreddits have been demonstrated to reduce user retention and engagement (Hickey et al., 2023). Moreover, we found that receiving reactions or comments on posts tended to anchor users within their initial MH category, reducing the likelihood of them posting or commenting in different MH communities. Again, this finding supports previous research on Reddit (Chen & Xu, 2021; Hamilton, Zhang, Danescu-Niculescu-Mizil, Jurafsky, & Leskovec, 2017; Zhang, Hamilton, Danescu-Niculescu-Mizil, Jurafsky, & Leskovec, 2017), which has displayed that sustained engagement and interactions increase the likelihood of continued activity within the same community rather than branching out to others. Additionally, the Social Identity Theory (Tajfel, Turner, Austin, & Worchel, 1979) supports this notion as well, as individuals derive a sense of identity and self-esteem from their group memberships, reinforcing their participation in their initial community and making them less inclined to seek assistance elsewhere. Lastly, we observed that while both comments and scores influenced subsequent user behavior, comments had a stronger effect compared to scores. This difference may be due to the qualitative nature of comments, which provide more empathetic and personalized feedback compared to the quantitative score metric, in turn reinforcing their continued participation (Green et al., 2020; Saha, Gakhreja, Das, Chakraborty, & Saha, 2022).

4.4. Methodological contributions and practical implications

Our study offered both methodological advancements and practical insights for the analysis and development of online MH communities.

From a methodological perspective, we introduced a large-scale dataset mapping Reddit's MH support environment over a five-year period. By categorizing subreddits according to DSM-5 diagnostic categories, we enabled an in-depth analysis of various MH conditions, moving beyond single-subreddit approaches. Additionally, the inclusion of comparative datasets – consisting of both support-related and general-interest subreddits – allowed for the identification of patterns specific to MH communities. Moreover, our automated intent classification model, based on posts' flairs, may provide a tool for large-scale analysis of user motivations in these communities. This model, trained on a ground truth dataset based on user-assigned flairs, demonstrated high accuracy in classifying posts into distinct intent categories i.e., seeking support, venting, offering help, and sharing progress.

The findings of this work have also practical implications for researchers, platform developers, public health agencies, and MH professionals. Our results highlighted that users derived significant value from interactions within these online forums. Many engaged repeatedly and meaningfully through both commenting and posting, often disclosing personal information in long posts and returning to the community multiple times. Also, the general tone of community responses, marked by prosocial linguistic markers, underscored the supportive environment these spaces foster. Public health agencies can build upon these results by providing guidance to developers and moderators of these forums to enhance their supportive nature further. Moreover, our analysis revealed that MH communities on Reddit shared characteristics with other mutual-help groups while differing from non-support-oriented communities. This suggests that online forums for MH support could benefit from tailored designs that facilitate common user needs, such as seeking support and venting, potentially enhancing user experience and engagement. The study also emphasized the crucial role of community response in user retention and participation. Since timely responses to support-seeking posts appeared particularly important for maintaining user engagement, platform moderators might consider implementing systems to flag unanswered posts, potentially improving community responsiveness. Interestingly, while posts offering help aligned with community goals, they tended to receive less engagement. This represents an opportunity for platform designers to investigate and potentially develop strategies to encourage interaction with these supportive contributions.

4.5. Limitations and future directions

As with most data-driven research, this study presents several limitations that on the other hand suggest potential areas for future research. Firstly, it is important to clarify that our study cannot determine the psychological states or actual diagnoses of Reddit users. Online self-help groups are indeed frequented by a variety of individuals, including those without diagnoses, their friends and relatives, and those merely considering their MH. However, our findings aimed to offer a comprehensive view of how these communities operate and engage in discussions rather than highlighting the severity of user conditions or predicting signs of disease.

Secondly, this study was limited to a single platform, Reddit, in a single language, primarily involving users from the US.[6] However, Reddit is one of the largest online communities today and hosts several subreddits dedicated to MH issues (Proferes et al., 2021). Although Reddit users may not represent the global population of individuals with MH issues, by including all subreddits related to specific disorders as classified in the DSM-5 (APA, 2013), we provided a broad picture of the Reddit MH environment.

Another limitation was our use of automated classification of intents, as it was inherently imperfect. While this method allowed us to categorize millions of posts efficiently based on user-assigned ground truth (post flairs), unlike previous studies that focused on small-scale collections, the categorization remained fairly coarse-grained with only four intents. This high-level categorization facilitated broad comparisons across different MH categories but did not capture the rich semantic variations within each intent category. For instance, within *seeking support* posts, there may be a meaningful distinction between users seeking advice about treatment options versus those looking for emotional validation, or between those discussing pharmacological approaches versus lifestyle changes. Future research could improve semantic analysis of these communities in several ways: (i) by including additional intents and specific discussion topics, potentially retrieved from user-assigned post flairs, (ii) by employing embedding-based approaches to analyze the semantic relationships between posts and identify common themes or patterns within each intent category, and (iii) by conducting phrase-level analysis to better understand what specific language patterns characterize different types of posting intents. Furthermore, by adopting more advanced Natural Language Processing techniques, researchers could better capture the tone of community responses, extracting richer dimensions from user-generated content.

While acknowledging these constraints, the ecosystem of human experiences captured in these online spaces continues to offer valuable insights that can inform and improve MH support strategies both online and offline.

CRediT authorship contribution statement

Virginia Morini: Writing – original draft, Visualization, Resources, Methodology, Investigation, Data curation, Conceptualization. **Maria Sansoni:** Writing – original draft, Validation, Resources, Investigation, Conceptualization. **Giulio Rossetti:** Writing – original draft, Validation, Investigation, Conceptualization. **Dino Pedreschi:** Writing – original draft, Validation, Investigation, Conceptualization. **Carlos Castillo:** Writing – original draft, Visualization, Supervision, Resources, Methodology, Investigation, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported by: the EU NextGenerationEU programme under the funding schemes PNRR-PE-AI FAIR (Future Artificial Intelligence Research); the EU – Horizon 2020 Program under the scheme "INFRAIA-01-2018-2019 – Integrating Activities for Advanced Communities" (G.A. n.871042) "SoBigData++: European Integrated Infrastructure for Social Mining and Big Data Analytics" (http://www.sobigdata.eu); PNRR-SoBigData.it "Strengthening the Italian RI for Social Mining and Big Data Analytics" - Prot. IR0000013; the Department of Research and Universities of the Government of Catalonia (SGR 00930); EU-funded project "SoBigData++" (grant agreement 871042); MCIN/AEI/10.13039/501100011033 under the Maria de Maeztu Units of Excellence Programme (CEX2021-001195-M).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.chb.2024.108544.

Data availability

Data will be made available on request.

References

Andrade, L. H., Alonso, J., Mneimneh, Z., Wells, J., Al-Hamzawi, A., Borges, G., et al. (2014). Barriers to mental health treatment: results from the who world mental health surveys. *Psychological Medicine*, *44*(6), 1303–1317.

APA, A. P. A. (2013). *Diagnostic and statistical manual of mental disorders*. The American Psychiatric Association.

Baek, Y. M., Bae, Y., & Jang, H. (2013). Social and parasocial relationships on social network sites and their differential relationships with users' psychological well-being. *Cyberpsychology, Behavior, and Social Networking*, *16*, 512–517.

Bandura, A., & Walters, R. H. (1977). *vol. 1*, *Social learning theory*. NJ: Prentice hall Englewood Cliffs.

[6] https://www.statista.com/statistics/325144/reddit-global-active-user-distribution/.

Barak, A., Boniel-Nissim, M., & Suler, J. (2008). Fostering empowerment in online support groups. *Computers in Human Behavior*, *24*(5), 1867–1883.

Barki, H., & Hartwick, J. (2004). Conceptualizing the construct of interpersonal conflict. *International Journal of Conflict Management*, *15*(3), 216–244.

Baumgartner, J., Zannettou, S., Keegan, B., Squire, M., & Blackburn, J. (2020). The pushshift reddit dataset. *vol. 14*, In *Proceedings of the international AAAI conference on web and social media* (pp. 830–839).

Berry, N., Lobban, F., Belousov, M., Emsley, R., Nenadic, G., & Bucci, S. (2017). # whywetweetmh: understanding why people use twitter to discuss mental health problems. *Journal of Medical Internet Research*, *19*(4), Article e107.

Blau, P. (2017). *Exchange and power in social life*. Routledge.

Boyd, R. L., Ashokkumar, A., Seraj, S., & Pennebaker, J. W. (2022). *The development and psychometric properties of liwc-22* (p. 10). Austin, TX: University of Texas at Austin.

Brady, W. J., Crockett, M. J., & Van Bavel, J. J. (2020). The mad model of moral contagion: The role of motivation, attention, and design in the spread of moralized content online. *Perspectives on Psychological Science*, *15*(4), 978–1010.

Brown, P., & Levinson, S. C. (1978). Universals in language usage: Politeness phenomena. In *Questions and politeness: Strategies in social interaction* (pp. 56–311). Cambridge University Press.

Brown, S. P., Westbrook, R. A., & Challagalla, G. (2005). Good cope, bad cope: adaptive and maladaptive coping strategies following a critical negative work event. *Journal of Applied Psychology*, *90*(4), 792.

Bucci, S., Schwannauer, M., & Berry, N. (2019). The digital revolution and its impact on mental health care. *Psychology and Psychotherapy: Theory, Research and Practice*, *92*(2), 277–297.

Chen, Y., & Xu, Y. (2021). Exploring the effect of social support and empathy on user engagement in online mental health communities. *International Journal of Environmental Research and Public Health*, *18*(13), 6855.

Cozby, P. C. (1973). Self-disclosure: a literature review. *Psychological Bulletin*, *79*(2), 73.

Crespo, M., & Fernández-Lansac, V. (2016). Memory and narrative of traumatic events: A literature review. *Psychological Trauma: Theory, Research, Practice, and Policy*, *8*(2), 149.

Cunha, T. O., Weber, I., Haddadi, H., & Pappa, G. L. (2016). The effect of social feedback in a reddit weight loss community. In *Proceedings of the 6th international conference on digital health conference* (pp. 99–103).

Cutrona, C. E., & Suhr, J. A. (1992). Controllability of stressful events and satisfaction with spouse support behaviors. *Communication Research*, *19*(2), 154–174.

D'Agostino, A. R., Optican, A. R., Sowles, S. J., Krauss, M. J., Lee, K. E., & Cavazos-Rehg, P. A. (2017). Social networking online to recover from opioid use disorder: A study of community interactions. *Drug and Alcohol Dependence*, *181*, 5–10.

Darley, J. M., & Latané, B. (1968). Bystander intervention in emergencies: diffusion of responsibility. *Journal of Personality and Social Psychology*, *8*(4p1), 377.

Davidson, L., Chinman, M., Sells, D., & Rowe, M. (2006). Peer support among adults with serious mental illness: a report from the field. *Schizophrenia Bulletin*, *32*(3), 443–450.

De Choudhury, M., & De, S. (2014). Mental health discourse on reddit: Self-disclosure, social support, and anonymity. *vol. 8*, In *Proceedings of the international AAAI conference on web and social media* (pp. 71–80).

De Choudhury, M., & Kiciman, E. (2017). The language of social support in social media and its effect on suicidal ideation risk. *vol. 11*, In *Proceedings of the international AAAI conference on web and social media* (pp. 32–41).

De Choudhury, M., Kiciman, E., Dredze, M., Coppersmith, G., & Kumar, M. (2016). Discovering shifts to suicidal ideation from mental health content in social media. In *Proceedings of the 2016 CHI conference on human factors in computing systems* (pp. 2098–2110).

Derlega, V. J., Metts, S., Petronio, S., & Margulis, S. T. (1993). *Self-disclosure*. Sage Publications, Inc.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Ferster, C. B., & Skinner, B. F. (1957). Schedules of reinforcement.

Fisher, J. D., Nadler, A., & Whitcher-Alagna, S. (1983). Four conceptualizations of reactions to aid. *New Directions in Helping*, *1*, 51–84.

Foster, S., Pomerantz, A., Bell, K., Carvallo, M., Lee, J., & Lee, J. (2022). Victims of virility: Honor endorsement, stigma, and men's use of erectile dysfunction medication. *Psychology of Men & Masculinities*, *23*(1), 47.

Fredrickson, B. L. (2001). The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. *American Psychologist*, *56*(3), 218.

Garg, M. (2024). Mental disturbance impacting wellness dimensions: Resources and open research directions. *Asian Journal of Psychiatry*, *92*, Article 103876.

Garg, M., Saxena, C., Krishnan, V., Joshi, R., Saha, S., Mago, V., et al. (2022). Cams: An annotated corpus for causal analysis of mental health issues in social media posts. arXiv preprint arXiv:2207.04674.

Gloria, C. T., & Steinhardt, M. A. (2016). Relationships among positive emotions, coping, resilience and mental health. *Stress and Health*, *32*(2), 145–156.

Gouldner, A. W. (1960). The norm of reciprocity: A preliminary statement. *American Sociological Review*, 161–178.

Green, B. M., Van Horn, K. T., Gupte, K., Evans, M., Hayes, S., & Bhowmick, A. (2020). Assessment of adaptive engagement and support model for people with chronic health conditions in online health communities: combined content analysis. *Journal of Medical Internet Research*, *22*(7), Article e17338.

Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology*, *2*(3), 271–299.

Hamilton, W., Zhang, J., Danescu-Niculescu-Mizil, C., Jurafsky, D., & Leskovec, J. (2017). Loyalty in online communities. *vol. 11*, In *Proceedings of the international AAAI conference on web and social media* (pp. 540–543).

Hartopo, D., & Kalalo, R. T. (2022). Language disorder as a marker for schizophrenia. *Asia-Pacific Psychiatry*, *14*(3), Article e12485.

Hickey, D., Schmitz, M., Fessler, D., Smaldino, P., Muric, G., & Burghardt, K. (2023). No love among haters: Negative interactions reduce hate community engagement. arXiv preprint arXiv:2303.13641.

Homans, G. C. (1958). Social behavior as exchange. *American Journal of Sociology*, *63*(6), 597–606.

Hopfgartner, N., Ruprechter, T., & Helic, D. (2022). Retention and relapse in gambling self-help communities on reddit. In *International conference on social informatics* (pp. 305–319). Springer.

Joseph, S. M., Citraro, S., Morini, V., Rossetti, G., & Stella, M. (2023). Cognitive network neighborhoods quantify feelings expressed in suicide notes and reddit mental health communities. *Physica A. Statistical Mechanics and its Applications*, *610*, Article 128336.

Kensinger, E. A., & Ford, J. H. (2020). Retrieval of emotional events from memory. *Annual Review of Psychology*, *71*(1), 251–272.

Koole, S. L. (2010). The psychology of emotion regulation: An integrative review. *Cognition and Emotion*, *13*, 8–177.

Low, D. M., Rumker, L., Talkar, T., Torous, J., Cecchi, G., & Ghosh, S. S. (2020). Natural language processing reveals vulnerable mental health support groups and heightened health anxiety on reddit during covid-19: Observational study. *Journal of Medical Internet Research*, *22*(10), Article e22635.

Luo, M., & Hancock, J. T. (2020). Self-disclosure and social media: motivations, mechanisms and psychological well-being. *Current Opinion in Psychology*, *31*, 110–115.

Merchant, R., Goldin, A., Manjanatha, D., Harter, C., Chandler, J., Lipp, A., et al. (2022). Opportunities to expand access to mental health services: A case for the role of online peer support communities. *Psychiatric Quarterly*, *93*(2), 613–625.

Mohan, S., Guha, A., Harris, M., Popowich, F., Schuster, A., & Priebe, C. (2017). The impact of toxic language on the health of reddit communities. In *Advances in artificial intelligence: 30th Canadian conference on artificial intelligence, Canadian AI 2017, edmonton, AB, Canada, May 16-19, 2017, proceedings 30* (pp. 51–56). Springer.

Moreira, E., Jr., Brock, G., Glasser, D. B., Nicolosi, A., Laumann, E. O., Paik, A., et al. (2005). Help-seeking behaviour for sexual problems: The global study of sexual attitudes and behaviors. *International Journal of Clinical Practice*, *59*(1), 6–16.

Morini, V., Citraro, S., Sajno, E., Sansoni, M., Riva, G., Stella, M., et al. (2024). Online posting effects: Unveiling the non-linear journeys of users in depression communities on Reddit. *Computers in Human Behavior Reports*, Article 100542.

Naslund, J. A., Aschbrenner, K. A., Marsch, L. A., & Bartels, S. J. (2016). The future of mental health care: peer-to-peer support and social media. *Epidemiology and Psychiatric Sciences*, *25*(2), 113–122.

Nils, F., & Rimé, B. (2012). Beyond the myth of venting: Social sharing modes determine the benefits of emotional disclosure. *European Journal of Social Psychology*, *42*(6), 672–681.

Norinder, U., Carlsson, L., Boyer, S., & Eklund, M. (2014). Introducing conformal prediction in predictive modeling. a transparent and flexible alternative to applicability domain determination. *Journal of Chemical Information and Modeling*, *54*(6), 1596–1603.

Park, A., Conway, M., & Chen, A. T. (2018). Examining thematic similarity, difference, and membership in three online mental health communities from reddit: a text mining and visualization approach. *Computers in Human Behavior*, *78*, 98–112.

Penner, L. A., Dovidio, J. F., Piliavin, J. A., & Schroeder, D. A. (2005). Prosocial behavior: Multilevel perspectives. *Annual Review of Psychology*, *56*, 365–392.

Prochaska, J. O., & Velicer, W. F. (1997). The transtheoretical model of health behavior change. *American Journal of Health Promotion*, *12*(1), 38–48.

Proferes, N., Jones, N., Gilbert, S., Fiesler, C., & Zimmer, M. (2021). Studying reddit: A systematic overview of disciplines, approaches, methods, and ethics. *Social Media+ Society*, *7*(2), Article 20563051211019004.

Saha, K., Ernala, S. K., Dutta, S., Sharma, E., & De Choudhury, M. (2020). Understanding moderation in online mental health communities. In *Social computing and social media. participation, user experience, consumer experience, and applications of social computing: 12th international conference, SCSM 2020, held as part of the 22nd HCI international conference, HCII 2020, copenha gen, Denmark, July 19–24, 2020, proceedings, part II 22* (pp. 87–107). Springer.

Saha, T., Gakhreja, V., Das, A. S., Chakraborty, S., & Saha, S. (2022). Towards motivational and empathetic response generation in online mental health support. In *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval* (pp. 2650–2656).

Saxena, S., Thornicroft, G., Knapp, M., & Whiteford, H. (2007). Resources for mental health: scarcity, inequity, and inefficiency. *The Lancet*, *370*(9590), 878–889.

Shi, J., & Khoo, Z. (2023). Online health community for change: Analysis of self-disclosure and social networks of users with depression. *Frontiers in Psychology*, *14*, Article 1092884.

Soós, M. J., Coulson, N. S., & Davies, E. B. (2022). Exploring social support in an online support community for tourette syndrome and tic disorders: analysis of postings. *Journal of Medical Internet Research*, *24*(10), Article e34403.

Suler, J. (2004). The online disinhibition effect. *Cyberpsychology & Behavior*, *7*(3), 321–326.

Tajfel, H., Turner, J. C., Austin, W. G., & Worchel, S. (1979). An integrative theory of intergroup conflict. *Organizational Identity: A Reader*, *56*(65), 9780203505984–16.

Tsvetkova, M., & Macy, M. (2015). The contagion of prosocial behavior and the emergence of voluntary-contribution communities. In *Social phenomena: From data analysis to models* (pp. 117–134). Springer.

Valdez, D., & Patterson, M. S. (2022). Computational analyses identify addiction help-seeking behaviors on the social networking website reddit: Insights into online social interactions and addiction support communities. *PLOS Digital Health*, *1*(11), Article e0000143.

Wei, L., & Liu, B. (2020). Reactions to others' misfortune on social media: Effects of homophily and publicness on schadenfreude, empathy, and perceived deservingness. *Computers in Human Behavior*, *102*, 1–13.

World Health Organization (2022). Mental health. (Accessed 30 July 2024).

Yeo, T. E. D. (2021). Do you know how much i suffer?: How young people negotiate the tellability of their mental health disruption in anonymous distress narratives on social media. *Health Communication*, *36*(13), 1606–1615.

Zent, M. (2023). Anonymous online support: Investigations of identity and heterogeneous groups in online recovery support. In *Extended abstracts of the 2023 CHI conference on human factors in computing systems* (pp. 1–8).

Zhang, J., Hamilton, W., Danescu-Niculescu-Mizil, C., Jurafsky, D., & Leskovec, J. (2017). Community identity and user engagement in a multi-community landscape. *vol. 11*, In *Proceedings of the international AAAI conference on web and social media* (pp. 377–386).

Zulkarnain, N. A., & Jan, J. M. (2019). Solicitation of online social support: Self-disclosure of eating disorder patients. *Journal of Nusantara Studies (JONUS)*, *4*(1), 244–364.