

“I Will Not Drink With You Today”: A Topic-Guided Thematic Analysis of Addiction Recovery on Reddit

Robert P. Gauthier
University of Waterloo
Waterloo, Ontario, Canada
rpgauthier@uwaterloo.ca

Mary Jean Costello
Homewood Research Institute
Guelph, Ontario, Canada
JCostello@homewoodhealth.com

James R. Wallace
University of Waterloo
Waterloo, Ontario, Canada
james.wallace@uwaterloo.ca

ABSTRACT

Recovery from addiction is a journey that requires a lifetime of support from a strong network of peers. Many people seek out this support through online communities, like those on Reddit. However, as these communities developed outside of existing aid groups and medical practice, it is unclear how they enable recovery. Their scale also limits researchers’ ability to engage through traditional qualitative research methods. To study these groups, we performed a topic-guided thematic analysis that used machine-generated topic models to purposively sample from two recovery subreddits: *r/stopdrinking* and *r/OpiatesRecovery*. We show that these communities provide access to an experienced and accessible support group whose discussions include consequences, reflections, and celebrations, but that also play a distinct metacommunicative role in supporting formal treatment. We discuss how these communities can act as knowledge sources to improve in-person recovery support and medical practice, and how computational techniques can enable HCI researchers to study communities at scale.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in collaborative and social computing**; **User studies**; • **Computing methodologies** → **Machine learning**.

KEYWORDS

Reddit; machine learning; thematic analysis; addiction; online communities

ACM Reference Format:

Robert P. Gauthier, Mary Jean Costello, and James R. Wallace. 2022. “I Will Not Drink With You Today”: A Topic-Guided Thematic Analysis of Addiction Recovery on Reddit. In *CHI Conference on Human Factors in Computing Systems (CHI ’22)*, April 29-May 5, 2022, New Orleans, LA, USA. ACM, New York, NY, USA, 17 pages. <https://doi.org/10.1145/3491102.3502076>

1 INTRODUCTION

Recovery from substance addiction [1] can involve long and difficult journeys [99]. A key component of those journeys is having a strong network of peers who can support a person as they work

towards a healthy, productive, and meaningful life [11, 17, 48, 53]. Common sources of this support are health professionals, rehabilitation programs, and 12-step programs like Alcoholics Anonymous (AA) or Narcotics Anonymous (NA). However, several barriers can prevent people from participating in these groups, like physical distance, lack of cultural similarity to peers, the stigma surrounding addiction, and a program’s appeal [28, 32, 48, 64]. As a result, many seek out less formal support communities through social networking platforms like Reddit or Facebook [39, 67]. However, since these communities have developed outside of existing support groups and clinical practice, it is unclear whether they provide appropriate and effective support to those who seek it.

Our work investigates what is discussed in online peer-to-peer communities that have formed around addiction recovery and contributes an understanding of how these communities support their members’ recovery journeys. We focused our inquiry on the use of Reddit (www.reddit.com), a pseudonymous social networking site where communities can discuss sensitive topics that people may not feel comfortable disclosing face to face or on sites where they are personally identifiable, like Facebook [63]. We investigated discussions in two subreddits, *r/stopdrinking* and *r/OpiatesRecovery*, where people can seek out advice about recovery from addiction to two common substances, alcohol and opiates [22].

To understand how these online communities support recovery, we built on previous Human-Computer Interaction (HCI) research (e.g., [2, 3, 82, 83]) by developing computationally-supported qualitative research methods [71]. In particular, we addressed key drawbacks of existing research in this space, arising from the tension between the time intensity of qualitative research goals and the scale of online communities. That is, existing research has commonly resorted to: 1) sampling only a small set of posts from each community to enable human researchers to develop a qualitative understanding of the materials (e.g., [98]); or 2) focusing on quantitative analysis (e.g., [2]) and losing much of the ‘thick’ understanding of these communities [18].

In this work, we applied computational techniques to perform a ‘topic-guided thematic analysis’ of discourses on recovery subreddits. First, we used Latent Dirichlet Allocation (LDA) [16], an unsupervised topic modelling technique, to develop models for each subreddit from four years of posts. We used these models’ topics to generate purposive samples [29, 42] by identifying related keywords and representative threads from both subreddits. We then performed reflexive thematic analysis [18] on our purposive samples’ threads to develop and review our themes. During our analysis, we performed inductive coding and grounded our interpretations in the communities’ original contexts by looking at associated threads

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHI ’22, April 29-May 5, 2022, New Orleans, LA, USA

© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-9157-3/22/04...\$15.00

<https://doi.org/10.1145/3491102.3502076>

on Reddit’s website. This combination of unsupervised topic modelling, purposive sampling, and reflexive thematic analysis enabled us to develop qualitative understandings of these communities while sampling from more than 150,000 threads. We present the results of our analysis in terms of two research questions: 1) How are stories used for addiction recovery in these Reddit communities?; and, 2) How do community members support each other’s recovery?

Our research contributes an empirical understanding of the discussions people on recovery journeys are having online, and the information they have sought and shared on social networks at a large scale. We show that the communities comprise experienced members, are perceived as accessible, and provide a channel for sharing lived experiences such as personal stories, advice on common problems, and emotional support. These resources are leveraged by people experiencing addictions, their family, and their friends. Further, we show that Reddit enables meta-discussions that help people from under-represented groups navigate in-person programs, for example women seeking women mentors, and people struggling with references to ‘a higher power’ in 12-step programs. In discussing these findings, we describe how our themes provide a holistic understanding of addiction recovery that includes online communities, and can inform practice for both mutual aid programs and healthcare practitioners. Finally, we reflect on the effectiveness of computational techniques in supporting the development of a qualitative understanding of online communities.

2 RELATED WORK

Experts today view addiction recovery as an ongoing journey that requires a variety of supports to enable those in recovery to “...develop a healthy, productive, and meaningful life.” [99, p. 236]. While professional treatment programs vary in implementation, they often share common elements such as an emphasis on education, development of coping skills, and management of co-occurring symptoms such as post-traumatic stress disorder [1]. In many programs, emphasis is placed on support and mentorship through mutual aid groups, which may involve participation in well known 12-step groups, like AA and NA, or alternative groups, such as Self-Management And Recovery Training (SMART) and Moderation Management (MM) [28, 48]. That is, the formation of a lasting, positive, behaviour-dependent, and supportive network of peers is considered a key component of long-term success [11, 17, 48, 53].

There is a growing body of work in the HCI literature that examines how structured online health communities can provide this support network, facilitate emotional support and information exchange [36], and help their members manage health challenges (e.g., [47, 60]). Notably, these structured communities have been found to be helpful for those with chronic conditions, including addiction [12, 80, 84, 92, 100, 105, 106]. To date, this research has largely focused on structured communities, such as InTheRooms [12, 86, 87] and MedHelp [25, 58], and has found that they offer similar benefits to in-person support groups, such as AA and NA [12].

However, these structured online health communities also have drawbacks. They are often accessible only to those with registered accounts, potentially deterring people, and their friends and family,

from accessing needed support. They also may place an emphasis on certain topics, perspectives for successful treatment, or belief systems that do not work for everyone [48, 54]. For example, InTheRooms places an emphasis on the 12-step programs AA and NA [49], whereas MedHelp’s focus on connecting people with medical professionals emphasizes experts’ opinions [56]. For these reasons, there is growing interest in understanding what kind of support open-access online health communities on platforms like Reddit provide their members.

Further, the scale of these open-access online communities makes them difficult to study. Previous HCI research has begun exploring approaches to computer-supported qualitative research methods (e.g., [2, 3, 24, 35, 71]). Nevertheless, researchers investigating addiction tend to make one of two compromises in their approach: 1) they sample a small subset of posts from each community to enable human researchers to develop a qualitative understanding of the materials (e.g., [33, 93, 98]), limiting their studies from including the years of discussions that members of the communities have access to, or 2) they focus on quantitative analysis (e.g., [82, 83, 95]) limiting much of the ‘thick’ understanding of these communities that qualitative analysis could have developed [18].

In our work, we make two contributions towards understanding the benefits of online communities focused on recovery: We first demonstrate use of computational methods for qualitative analyses of online communities’ discussions, called a ‘topic-guided thematic analysis’, to overcome limitations of existing methods. We then perform an analysis on two addiction recovery communities on Reddit, describe how they support addiction recovery, and show how they help people from under-represented groups navigate in-person programs.

2.1 Analysis of Online Communities, Computational Support, and Sampling

One of the most significant challenges of studying online communities is their scale: each community is potentially comprised of hundreds of thousands of posts from tens of thousands of people over a period of years. Thematic analysis is a time-intensive method, due to the amount of reading, re-reading, and reviewing involved, that aims to explore and develop understandings of complex data [18]. The scale of large online communities both amplifies the amount of time needed for analysis and makes finding data that contain interesting aspects difficult. Research to date often overcomes this challenge through different approaches to *sampling*; such as, selecting posts from a small time period [98] or those associated with ‘hot topics’ at a given point in time [33]. For instance, Wadley et al. [98], in their investigation of r/StopSmoking chose a sample frame of the 732 posts made during April 2014, and then randomly selected 100 posts from within that sample for manual coding. While these sampling approaches enable manual coding of the data, they also have substantial limitations: for instance, they limit researchers’ opportunity to familiarize themselves with the data and develop a contextual understanding of the communities, and can exclude data from dominant or seasonal trends [62].

Computational techniques provide an opportunity to overcome these sampling limitations, and enable an in-depth, qualitative understanding of online discourse [35, 71]. For example, Latent Dirichlet Allocation (LDA) [16], an unsupervised modelling approach that can identify latent topics and associated threads within an online community [59], can be used to purposively sample [29, 42] discourses for analysis. Researchers have employed the use of computational methods to derive a variety of topics from large data corpora for some time (e.g. [8, 30, 31, 34, 68, 74]). LDA has been particularly useful to HCI researchers in identifying latent topics in Reddit communities (e.g., [2, 3, 82, 83]), and has been described by Ammari et al. [3] as part of a ‘roadmap’ for using computational techniques to better understand social relationships online.

Inspired by their work, we further develop this roadmap with a focus on a qualitative, *human* understanding of online discourse. We applied computational techniques to perform a ‘topic-guided thematic analysis’ of discourses on Reddit, where unsupervised LDA is used to sample for a thematic analysis. Our approach parallels explanatory mixed methods designs [29], and explores the use of computational techniques to augment human researchers’ abilities, as described by Muller et al. [71].

The use of LDA to purposively sample for thematic analysis has several advantages. First, our two LDA models use the breadth of each corpus of threads (144,422 threads from *r/stopdrinking* and 14,079 threads from *r/OpiatesRecovery*) which allows topics to emerge from all threads, rather than from a small sample. Second, LDA assumes that each thread comprises a mixture of topics, enabling the identification of secondary and/or latent topics [16], which aligns with how Reddit threads can involve multiple people contributing different viewpoints on both initial posts and subsequent responses. Third, purposive sampling via LDA enables identification of multiple threads for each topic, providing us opportunities to iteratively validate the models for human semantic sense [59] and to identify the samples needed for thematic analysis.

2.2 Open-Access Online Communities and Addiction Recovery

Large communities exist on open-access platforms around issues like addiction [33]. In particular, research has shown a large degree of participation on social networking platforms for topics like smoking cessation [98] and diabetes [76]. Previous work has also shown that online communities possess the same treatment mediators present for in-person group support [33]. For instance, Q&A participation in online communities has been found to be motivated by altruism and efficacy; known mediators for in-person mutual aid groups that are associated with an increased likelihood of recovery [79]. As such, these open-access online communities can be considered to be mutual aid groups.

Additionally, pseudonymity fosters disclosure in online forums, particularly for sensitive topics like addiction [89, 104], diabetes [76], and pregnancy loss [4]. That is, people are often less willing to discuss sensitive topics when those discussions are linked to their real identities on platforms such as Facebook or LinkedIn [76]. On the other hand, platforms that support or even encourage various degrees of anonymous participation, like Reddit, have been found to be supportive of these sensitive discussions [3, 5, 72]. The

freedom to post anonymously has been shown to enable discussions of mental health [6, 14], parenting issues [3], and sexual abuse [5].

However, the peer-to-peer nature of these open-access online mutual aid groups, and their lack of medical authority, raises questions surrounding whether the information and advice are appropriate in the context of addiction [33, 79]. That is, should these forums be officially recommended by health professionals as a place to find support? Initial work has suggested several limitations of online communities. For instance, community members may largely be new to sobriety, and advice provided by these groups may be harmful rather than helpful [27, 93]. Similar concerns arise from the perspective of members seeking help, for instance Rubya and Yarosh [87] found that community members may be less likely to self-disclose information in online meetings than when meeting face-to-face, and that geographic differences make communication less effective, or may even lead to conflict. Barrett and Murphy [7] found that online meetings were not perceived as being more accessible than face-to-face meetings, and were perceived as being less effective and of lower quality.

Further, much of the work that seeks to understand these communities is focused on the perspective of clinicians, such as looking for themes from diagnostic tools [33, 37], clinician expertise-based recommendations [47], analyzing discussions to find addiction mechanisms and treatment methods that can be assessed for clinical validity [23, 51], and categorizing users into clinical diagnoses [57, 95, 107]. While valuable, clinician perspectives may not always align with community members’ values [13].

To address this gap in our understanding, we performed analyses of two active addiction-support subreddits. We show that the topics discussed align well the support identified in the healthcare literature [11, 17, 48, 53], and are consistent with a positive, supportive network of peers. Further, we found that these open-access online communities serve a meta-communicative role in helping people to navigate difficulties with in-person groups, such as women seeking women mentors, and people struggling to accept religious aspects of AA.

3 METHOD

We chose to study two subreddits because we initially sought to establish both similarities and differences across recovery communities. However, as our analysis progressed, it evolved to focus on exploring themes that were generated from our interpretation of both subreddits’ discussions. Such an evolution is expected when performing thematic analysis using an inductive coding process [19].

We investigated two active addiction recovery subreddits focused on alcohol addiction (*r/stopdrinking*) and opiate addiction (*r/OpiatesRecovery*), retrieving corpora from *pushshift.io* [10]. These subreddits were selected because: they address recovery from use of two different classes of substances that are stereotyped as legal (alcohol) and illegal (opiates) and are current concerns of public health [81], they were the largest recovery subreddits we could find for each substance, they are publicly accessible, and they are active in terms of number of community members and posts.

We used thematic analysis to create in-depth understandings of the behaviour we observed in both subreddits. Thematic analysis enabled us to develop ‘thick’ understandings of the subreddits’

community discussions and to generate results that are accessible for both researchers and the general public, while also capturing unanticipated insights [18]. To perform the thematic analysis, we first built LDA models for each subreddit and used those models to purposively sample [29, 42] threads for each topic. Our LDA models used the full corpus of texts (144, 422 threads from *r/stopdrinking* and 14, 079 threads from *r/OpiatesRecovery*) to develop 16 topics for each subreddit. 20 discussion threads were selected from each topic in each of our two 16-topic LDA models, providing a total sample of 640 threads (composed of 640 submissions and 7828 comments). We then performed a reflexive thematic analysis [18] where we inductively coded the sampled discussions, in their living state on Reddit, and used our codes and samples to develop our themes.

Our iterative approach comprised three phases: data gathering, LDA topic modelling, and thematic analysis (Figure 1). The code used to support each activity was written in Python, and is included in the supplementary material.

3.1 Data Gathering and Ethical Considerations

As addiction recovery is a sensitive topic, we took additional steps to consider the ethical implications of our work and to protect the communities that we were interested in learning from. We reviewed Reddit's terms of service and the rules and FAQs of both of the subreddits to confirm that data was open for public use, though they did not explicitly allow use for research purposes. We included all publicly available threads from the *r/stopdrinking* and *r/OpiatesRecovery* subreddits created during 2014 to 2017 (144, 422 threads and 14, 079 threads respectively). We chose the start point of January 1, 2014 and the end point of December 31, 2017 based on our desire to allow behaviours that might be seasonal to occur multiple times and what was available when we started performing the analysis in the summer of 2018. Data for submissions and comments was downloaded from *pushshift.io* in json format [10].

We used discussion threads as our unit of analysis because we intended to examine entire threads for our thematic analysis, to preserve the context of each submission and its responses. Threads were recreated by merging `title` and `text` fields by `submission id` for both the submission and any associated comments. The downloaded data was also used to identify URLs as well as aggregated to find thread counts and distinct `user id` counts by date. Other non-aggregate information was not extracted or used from the dataset to respect both the community members' privacy (i.e., classifying and categorizing community members risks inferring private information [97]) and limitations of *pushshift.io* [10] as a snapshot of posts on Reddit (e.g., archived karma scores often do not match those currently on the website).

All published quotes are paraphrased from existing non-deleted posts to preserve pseudonymity. To respect the choice of community members who chose to delete content, we did not include deleted posts in our thematic analysis. Deleted content was identified manually when reviewing threads on Reddit for thematic analysis. We paraphrased by breaking quotes down into their thematic analysis codes, then manually constructed a new quote. We

then compared against the old quote for consistency, and Googled it to ensure anonymity.

Our study design received full approval from our institution's research ethics board, and is consistent with guidelines from the HCI community for protecting pseudonymous research participants (e.g., [20, 61]) and transparency in qualitative research [94].

3.2 LDA Topic Modelling

LDA topic modelling involved three iterative sub-phases: cleaning, building, and evaluation. These sub-phases ultimately produced two models: one for each subreddit. Each model comprised 16 topics, keywords, and a list of associated threads that we used as purposive samples for our thematic analysis. Although we cannot release the final cleaned datasets or generated models, since they contain non-paraphrased text data that could be used to identify community members, our code and a summary of outputs is available in Appendices A and B.

3.2.1 Cleaning. Our cleaning process emphasized human interpretability [9, 59], an important consideration given our goal was to create purposive samples for thematic analysis by a human researcher (e.g., [9, 71]). Our initial cleaning approach consisted of: 1) lemmatization and part of speech identification [50] using *spaCy* [44]; and 2) English stop word filtering, as well as filtering out words that were not nouns, verbs, adjectives, and adverbs [50] using *NLTK* [15] and *Mallet* [65]. This kind of 'light cleaning' improves the interpretability of models, but does not impact their stability [90, 91].

In subsequent cleaning iterations we noticed repetition of common acronyms within models' topics. To reduce this duplication and allow the LDA modelling to treat all representations as being the same word, we expanded a number of general and domain-specific acronyms (e.g., 'fyi' to 'for your information' and 'wd' to 'withdrawal') to their full form. We also observed that frequently occurring words, particularly adjectives, were causing a high level of overlap between topics. To reduce this overlap, we removed both adjectives and words that appeared in more than 25% of threads. To further improve performance, we also removed words that occurred in fewer than 20 threads using *Gensim* [85] and masked out external links (e.g., 'http://...').

3.2.2 Building. We built an LDA model for each subreddit using *Gensim* 3.8 [85]. For metadata parameters we set the number of passes to 100 along with both `alpha` and `eta` to `auto`, to allow each model to infer its own asymmetric topic distribution from the corpora [85]. We then built 10 separate LDA models for each subreddit, using the generated dictionaries and corpora, and selected the model with the maximum coherence score from the 10 generated models (Appendix A and B).

We set the number of topics in each model to 16, based on pilot runs that indicated the topic coherence had plateaued; previous research has found that topic models with higher coherence score correlate to human-interpretable topic groups [59, 75, 88]. We selected 16 topics for each subreddit because, although opiates recovery plateaued earlier at 9 topics, we wanted to gather a similar sized sample from each subreddit to allow our thematic analysis to consider each community equitably (Appendix C).

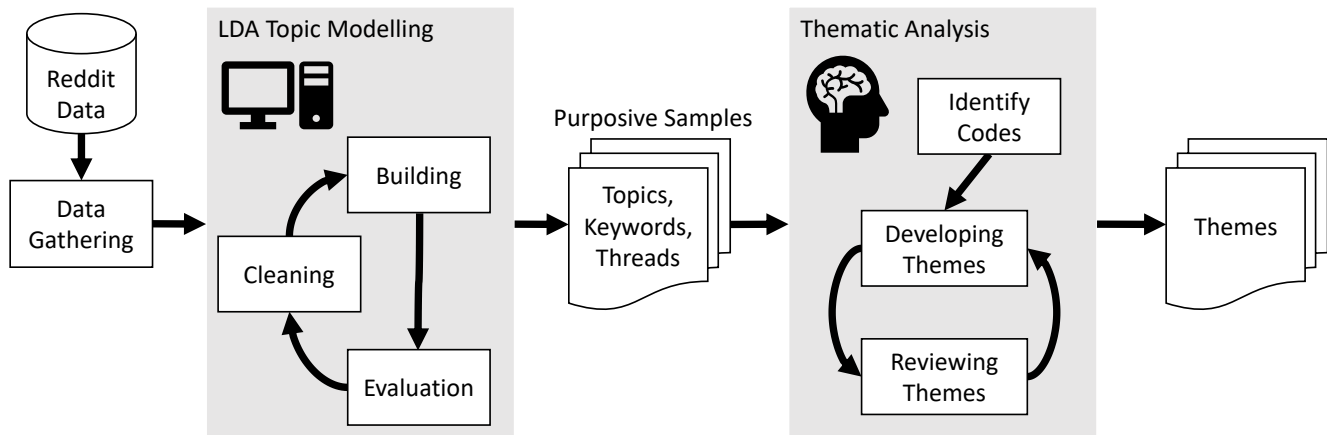


Figure 1: To perform our thematic analysis, we first collected Reddit data from 2014 to 2017 for *r/stopdrinking* and *r/OpiatesRecovery*. We then developed LDA topic models for each subreddit by iteratively cleaning, building, and evaluating the generated models, which produced purposive samples made up of keywords and threads for each topic. Finally, we performed the core of our thematic analysis by iteratively identifying codes, developing themes, and reviewing themes with the topics, keywords, and threads from our LDA model.

3.2.3 Evaluation. We then iteratively fine-tuned the LDA topic models [59]. We used the models to categorize the available threads and inspected the distribution between different topic groups. We reviewed the models, their topic terms, and the topics’ associated threads to assess their reliability [59], coherence (i.e., C_V [88]), and, most importantly, whether the topics were interpretively useful [9] to both the authors and a separate pilot group of 12 HCI researchers. During these reviews we reflected on whether additional cleaning was required. During early iterations, we also tried using Jaccard’s distance as a measure of topic similarity [3]. However, as we adjusted our cleaning process the Jaccard’s distance measures approached 1.0 for almost all identified topics, and so we did not ultimately use it to guide our topic modelling. Instead, we relied on our semantic interpretation of generated topics, consistent with our need to inform our thematic analysis.

3.2.4 Purposive Sampling. To purposively sample we used each subreddit’s LDA model to retrieve 20 representative submission ids for each of the 16 topics, giving a total of 640 threads for analysis. We identified representative threads by calculating the probability of each topic occurring in each thread, and selected the 20 threads with the highest probability. We then generated URLs for each submission id that could access the discussion threads. We accessed each thread through Google Chrome to ground our analysis in the context of the Reddit communities. The 320 *r/stopgaming* threads were composed of 3302 comments with a median of 7 comments per thread (mean = 10.32). The 320 *r/OpiatesRecovery* threads were composed of 2526 comments with a median of 6 comments per thread (mean = 7.89).

3.3 Thematic Analysis

For our reflexive thematic analysis [18] we took the realist stance that the continued existence of these subreddits implies a perception of value by their communities and that seeking to understand

addiction recovery is a complex process with many different contributing processes and possible outcomes. To focus on the experiential knowledge of the subreddits, we used inductive coding rather than try to force the behaviour of the communities into current understandings of addiction recovery or categorize the communities’ members.

We based our reflexive thematic analysis approach on the 6 phases described by Braun and Clarke [18]. The familiarization phase occurred as part of the LDA topic modelling. Using what was learned from familiarization, the first author then worked with the purposive samples to identify codes and to develop and review themes. The first author read each thread in its original context on Reddit to understand what data-driven codes were present. The first author iteratively compared the threads and codes to develop themes (Appendix D); for example, the common code pair ‘seeking information’ and ‘providing information’ was combined with codes for different types of seekers, such as ‘Atheists’ and ‘Female’, and support group sub-codes, such as ‘overwhelmed’ and ‘fear of stigmas’ to develop the theme ‘Navigating 12-Step programs’. Threads could contain multiple codes, codes could contribute to multiple themes, and thus threads could also contribute to multiple themes. The first author gathered supporting quotations for each theme from multiple threads that contributed supporting codes. Finally, the first and third authors reviewed the themes by looking at additional threads from each subreddit to confirm that they were present and came to agreement that the themes fit the data.

To consider the individual researcher positions inherent in qualitative research, such as our reflexive thematic analysis, we conducted group reflections with the first, second, and third authors on the identified themes [29]. During these group reflections we discussed the topics used for sampling, the themes developed, the example quotes, and the first author’s experiences and thoughts from interacting with content on a sensitive subject. We did not

seek to establish inter-rater reliability, since it is inconsistent with reflexive thematic analysis as described by Braun and Clarke [19], and because there was a single coder and “codes are the process not the product” of our work [66]. The final two phases, developing/reviewing of themes and producing of the report, occurred jointly between the first and third authors. The first author brought forward immersive experience from the previous phases and the third author provided experience communicating to the intended HCI audience.

4 RESULTS

An initial review of the subreddits identified that both were active and growing over the four-year period in terms of distinct user names and active threads (Figure 2). We defined a distinct user name as one with at least one submission or comment within the month, where over the entire four year period *r/stopdrinking* had 58,407 distinct user names compared to 10,668 for *r/Opiates-Recovery*. We defined an active thread as any with at least one new submission or comment during the month, where over the entire four year period *r/stopdrinking* had 144,422 active threads, and *r/OpiatesRecovery* had 14,079.

4.1 How are stories used for addiction recovery in these Reddit communities?

Our thematic analysis revealed that the communities engage in a range of discussions related to addiction and recovery, including stigmatized and personal areas, such as relapse, body weight, personal finances, and legal trouble. We now describe how discussions on these subreddits used stories to share experiences, provided peer encouragement, established the consequences of addiction, and exposed substance related concerns (Table 1).

4.1.1 Sharing Experiences. The most frequent theme that emerged from our analysis focused on community members sharing experiences with one another through stories. This theme is consistent with previous work (e.g., [98]), as the sharing of stories serves as a way for people to relate to each other and to learn from others’ successes and failures. These stories served as an opportunity to self-reflect or relate to others, often through comparing their lives to books or movies (Table 1, Sharing Stories). Other stories would discuss common experiences, such as dreams about drinking:

“I was so relieved when I finally realized it was in a dream. At first I thought it was impossible that I was in the middle of my 1st glass of beer. I felt that I had fallen off the wagon and thought I may as well enjoy the beer or maybe it was a dream like last time. Then I realized it really was a dream!”

These stories also often described how community members fought against internal monologues debating substance use, and reflected on what caused a relapse to both seek out help and serve as a warning to others:

“After 3 months I relapsed today. It wasn’t even on my former drug of choice. So others know it was not worth it as I feel regret for my choice in addition to how shitty I feel when on opiates. The only upside is that this destroyed any remaining belief in my mind

that opiates are amazing. I feel like what used to comfort me now disgusts and depresses me. I feel like I’m poisoning myself.

Not feeling sober sucks and I can’t wait for this to be over again”

4.1.2 Peer Support. We also observed that community members were supportive of one another — including messages of solidarity, encouragement, affirmation, and celebration. This theme highlights and reminds us that the community is participating in each others’ journeys of recovery. For example, the previous quote describing a relapse after three months provided an opportunity for another community member to encourage the original poster towards recovery:

“Remember that you were clean for all those days so they were not a waste. These things happen as we all have ups and downs. Sometimes from the downs we can see what we’re missing and improve next time. Keep smiling, keep your spirits up, and keep going.”

Another form of peer support we noted was reoccurring ‘open check-in threads’, where any community member could mark sobriety milestones in their recovery journey. These threads serve a similar purpose to the ‘chips’ or ‘sobriety coins’ that are handed out during in-person 12-step programs to mark sobriety milestones such as 24 hours or 1 month, but also could be more frequent, such as daily, weekly, monthly, or around holidays. Comments on these check-in threads were often as short as “*Day 50!*”, but also could be multiple paragraph stories describing the community member’s current situation. In addition to open check-in threads, community members created personal threads where they shared events like birthdays as personal milestones where they recommitted to sobriety and were celebrated by the community:

“Thirty Nine! I’m 39 years old today and I’m less than a week from 1 year sober. I feel like a million bucks. Quitting drinking with y’all was so worth it.”

» “Sobriety is such a great birthday present that you gave yourself. Happy Birthday!”

Throughout threads involving peer support, we observed an unusual acronym/phrase in *r/stopdrinking*: “IWNDWYT”, or “I will not drink with you today”. Several variations on this acronym and its full form were observed throughout the threads, including “IWNDWY” (i.e., I will not drink with you), “IWND” (i.e., I will not drink), and “IWND (anymore) WYT” (i.e., I will not drink anymore with you today). We saw the acronym used to respond to stories of both struggles, such as not drinking after the death of a loved one, and struggles, such as returning to the community after experiencing a relapse. We observed that members from *r/stopdrinking* used this acronym to express solidarity with each other in both good and bad times of living in recovery.

4.1.3 Consequences. In addition to community members encouraging and supporting one another, discussions often covered the consequences of substance use and of recovery, including interactions with the legal system, financial costs, academic costs, loss of social connections, and quality of life. While the two subreddits

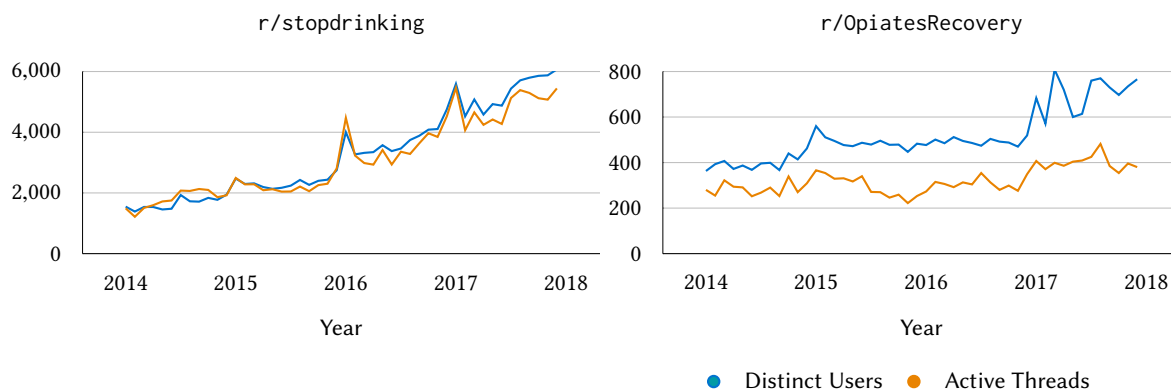


Figure 2: Distinct users and active threads for r/stopdrinking (right) and r/OpiatesRecovery (left) showing that both counts are experiencing an upwards trend over time for both subreddits.

Theme	Sub-Themes	Paraphrased Example Quote
Sharing Experiences	Self Reflections	“Reading <i>This Naked Mind</i> and thinking about my feelings and what alcohol took from me has been enlightening. I was able to establish a critical perspective that showed me how warped my thoughts had subconsciously become.”
	Sharing Failures Sharing Successes	“I remember how much I struggled at 4 months and how I couldn’t understand why it wasn’t getting easier to resist the cravings. Now that I’m at 6 months I am finally understanding why everyone means then they say the cravings don’t go away they just change. Thankfully now, despite it being a shitty week, I am not thinking about using as my first thought. Remember it does get easier, so don’t give up.”
	Waking Up	“Still experiencing the occasional vivid dream of taking pills. I guess its because it was so prominent in my life for so long. What sucks most is that after these dreams the craving is so strong. At least I am starting to feel disappointed in the high even even in the dream.”
Peer Support	Check-ins	“Day 5, Monday Night. Really wanted to drink but I resisted!”
	Encouragement	“exercise works wonders! try different activities like yoga and working out. keep up the good effort!”
	Solidarity	“It’s great how our lives don’t have to be like that anymore!”
Consequences	Benefits of Recovery Costs of Recovery	“I am trying to find rehab or detox facilities in the southern US that will take my government issued insurance. Does anyone have any suggestions?”
	Harm from Substance Use	“I saw in the newspaper that someone got picked up for their 5th dui. This made me think about my own dui from several years ago and realize how great it is to be free of both alcohol and the legal system.”
	Substance Related Concerns	Pain Management
Socializing		“It’s super bowl season and while we aren’t huge into sorts my significant other and I do like the cultural aspect. What do people suggest as bars are clearly now off the table?”

Table 1: Example themes that show diverse discussions related to addiction and recovery occurring on the subreddits. Includes themes identified and paraphrased example quotations from the subreddits.

often focused on different consequences, such as drinking and driving for *r/stopdrinking* and drug court for *r/OpiatesRecovery*, the general theme of consequences covered positive, negative, and confusing aspects of both addiction and living in recovery.

In threads that discussed negative consequences, we observed community members expressing relief that they had either already worked through the consequence or no longer had to worry about a consequence of addictive substance use now that they were living in recovery (Table 1). We also observed members receiving peer advice on how to handle the negative consequences that had occurred in the past or that they currently faced:

“Was finally arrested for a theft from a long time ago. Was back when I stole to feed by my H addiction and I knew consequences would eventually catch up to me. I was glad that I wasn’t scared of jail withdrawal thanks to being clean.”

» “It’s good that you’re clean as judges like to see that. Be sure to find a decent lawyer and don’t be tempted to relapse”

» » “Yep, I definitely plan to be honest about my past to the court and explore my options with my lawyer. In Jail I was even offered pills and I was proud that I was able to say No!”

The cost of recovery was a negative consequence frequently discussed on Reddit. These discussions covered insurance, drugs, and rehab facilities, and highlighted the financial burden to stopping use; especially when medicated withdrawal is necessary. Conversely, in threads that focused on the positive consequences of recovery, people would discuss how much money they had saved due to living in recovery, for example:

“When I see how much I saved I can’t help but think it’s awesome to not be wasting money on alcohol. I won’t drink with you today.”

We also observed the communities discussing how being in recovery had brought about confusing consequences such as impacts on body weights in *r/stopdrinking*:

“Has anyone else experienced weight loss from no longer drinking?”

» “Yeah, I found it weird how even though I was eating more chocolate bars I still lost weight during my first few months. I also found that tracking my diet was easier when I wasn’t binge eating in the middle of the night while blackout drunk”

While the question and response suggested a loss of weight, we saw many different perspectives throughout the thread, including this response where a member discussed weight gain and raised changing exercise routines:

» “Believe it or not I have gained weight. Also, I’ve been having better workouts and am pumped! Has anyone else noticed they can increase the weights they lift more rapidly? I know it keeps me motivated to keep checking in.”

We came to understand that Reddit enabled community members to develop a better understanding of their own experiences by comparing them to the consequences experienced by others. Overall,

this theme illustrated how the discussion of consequences took on a variety of forms, were largely personal in nature, could provide direction to other members in navigating their own path to recovery.

4.1.4 Substance Related Concerns. In both communities we observed discussions that were focused on concepts that were tightly connected to the substance’s context. Members of *r/OpiatesRecovery* were often concerned about managing acute and chronic pain, frequently associated with a recent injury or planned surgery. For example:

“I was in the same situation as you early in my recovery. When I went to the doctor I told them that I was allergic to codeine so that they wouldn’t consider giving me any. Instead they gave me this awful syrup I had never heard of before. It didn’t work for me but luckily someone else I know in recovery is a nurse who directed me to tesselon perles. It doesn’t work very well but at least it was something. I think there’s other over the counter options as well as long as you don’t abuse them.”

If your doctor knows you’re struggling with opiates and they still prescribed this to you, then just take it as recommended and be careful. If they don’t know then I’d say try to find a way to ask them for something that is non-narcotic.”

The urgent and fearful tone of the concerns highlighted to us that it is important to consider the beliefs of the community when treating the individual, such as the perception that effective pain management drugs are opiates, and therefore pain treatment introduces opportunities to relapse; either through a person taking the drugs in good faith to treat pain, or through the temptation to obtain a legal prescription to feed an addiction. As these concerns were raised in *r/OpiatesRecovery*, we saw that the community would respond with sharing stories and support, and sometimes provide advice on alternative drugs that could be explored with the help of a physician.

Discussions with similar fearful tones were observed in *r/stopdrinking*. However, rather than pain management, these discussions were focused on concerns around socializing. For instance, we observed a fear of choosing between risking relapse and maintaining social commitments :

“So I’m part of a group trip planned for Ireland, However now that I’m in recovery I’m growing increasingly nervous about the planned pub visits where the rest of the group will be drinking. I know I can avoid drinking normally but I’m scared it will be different when I’m in a different place. Any suggestions you might have would be appreciated. Thanks.”

Another concern was that recovery stigma was driving a wedge into existing friendships:

“I found out that some friends went out last Saturday and didn’t invite me because they thought that I don’t like to drink beer. I miss going out with them and I feel excluded just because I’ve stopped drinking.”

Reading multiple threads with similar concerns highlighted to us that the social context of alcohol consumption complicates the struggle that can occur when community members seek to balance their existing social networks with their recovery. The community would try to be supportive and offer suggestions when social concerns were raised. The suggestions commonly included taking time to develop a personal plan for the situation and establishing a trusted in-person social network in order to be able to provide support. These suggestions show the community seeking to help members learn skills that can both help them handle the current situation as well as be reused during their recovery journeys. Another type of suggestion was to consider alternative activities that the entire social group could enjoy. This type of suggestion highlighted to us how the *r/stopdrinking* community tries to encourage members to look past their fears and develop and maintain healthy connections to their friends rather than fixating on the stigma that addiction creates a wall between you and your friends.

4.2 How do community members support each other's recovery?

During our thematic analysis, we saw indications that Reddit made it possible to have discussions that may be difficult to address in person, and that the ability to share information easily through URLs fostered sharing both within the recovery communities and from outside groups. The sharing nature of Reddit was observed to be particularly supportive in discussions on social relationships and activities, knowledge seeking, and in supporting formal treatment of addiction (Table 2).

4.2.1 Social Relationships and Activities. Community members sought out suggestions for new activities, taking advantage of the subreddit's accessibility and the community's collective experience. This theme is particularly important because, for many living in recovery, previous social relationships and activities were formed around their drug of choice, leaving a void to be filled and many opportunities for relapse. It also mirrors recovery programs which seek to maintain balance, structure, and routine in a person's daily life [21, 73]. Responses frequently included links to related content, such as music videos on YouTube. The discussions would often be initiated by a request, such as:

“The local methadone clinic, while pricey, helped me stay off heroin for 6 days and is saving me money. In other news I wanted to see if anyone had any new music suggestions for when I'm feeling down. I've been wanting to find new music to help with my boredom. Thanks for any suggestions you can give. I'm really liking this site and am glad that I was shown it by a friend.”

While many sought out advice for off-line activities, we also observed that community events that provided members with online activities that helped to fill voids in their social lives such as Book and Movie clubs (Table 2). These events were often synchronous, such as in the case of movie nights, and provided members a sense of being together. Other events were less tightly-coupled, such as book or reading clubs, where members would post about a shared reading assignment over a 1-week period.

4.2.2 Sharing Knowledge and Lived Experience. Many people came to the subreddits as a step towards gaining a better understanding of addiction and recovery. This desire for knowledge came from community members living in recovery, as well as those more generally impacted by addiction in some way, such as family or friends, or those actively using (Table 2). These discussions highlighted to us that the Reddit platform enables knowledge to be sought, distributed, and discussed by the community. Responses to questions often included formal sources of information, like scientific research results, therapy approaches, new discoveries in the media, and the experiences of others, and was supported by the ability to directly link to content online via URLs. Additionally, some members disclosed that they felt the sense of anonymity provided by Reddit helped make them feel more open to sharing their personal experiences:

“I hear you. While I like NA I find Reddit's real anonymity provides me something which makes me feel free to share more personally.”

For family and friends, Reddit also provides a venue for advice, and access to a community with lived experience, that may not otherwise be accessible. Many used the platform as a way to reach out for advice on how to approach a loved one's recovery:

“My older brother lost his job a year ago. I figured out he has been binge drinking all the time and I want to help. Can someone who can relate to this give me some advice?”

On both subreddits we observed community members responding to questions like these with their lived experiences. We also saw the original posters respond with thanks, showing that these two Reddit communities were considered a valuable knowledge source by outsiders who had been impacted by addiction. We also noted responses that referred people to other subreddits or external organizations such as:

“I suggest you check out *r/AlAnon* as they have resources for family and friends of people with alcohol problems. We're the people with those problems here.”

These responses showed us that the subreddits could act as a knowledge bridge that connected community members living in recovery with others impacted by addiction. These responses also showed us that the subreddits exist within an ecosystem of both online and in-person groups that are interconnected with one another through a multitude of addiction recovery journeys.

Throughout the discussions we also saw indications that the communities wanted their members and any readers to be aware that their lived experiences supplemented professional expertise rather than replaced it. For instance, when discussing treating withdrawal and any symptoms, community members frequently mentioned that they where not medical experts and that that the experiences they described should be discussed with a care provider:

“That medicine can help with sleeping but it is actually an anti-psychotic, To use it you seriously need to talk with your doctor more. I took it for years but I was always under supervision. While I was glad to get off it a few months ago how you use it and stop using it is

Theme	Sub-Themes	Paraphrased Example Quote
Social Relationships and Activities	Filling a Void	“Since I stopped drinking my life has felt empty. What do you suggest to handle the boredom?” » “You got to fill that emptiness! You can get started with anything you wanted to try in the past. It might be hard starting but motivation will come if you try.”
	Group Activities	“During the movie club we chat while watching the linked stream in sync as best as we can.”
	Healthy Activities	“I picked up cross-training, biking, and running. I also found myself enjoying golf.”
	Leisure Activities	“If you’re bored trying reading Game of Thrones. With five currently out and thousands of pages it will keep you occupied a while.”
Sharing Knowledge and Lived Experience	Family and Friends	“tl;dr my older sibling is addicted to opiates and I have no idea how to help them.”
	Managing Addiction	“I know exercise sucks to get started during withdrawal but believe me when I say you will feel better after you finish. I read this is because of runner’s high where the body releases endorphins which helps the opiate receptors feel less starved.”
	Managing Consequences	
	Managing Recovery	
	Managing Withdrawal	
Understanding Addiction	“I can’t understand how any amount of alcohol can be good for us as since I stopped drinking recently and I feel way healthier. Does anyone know if the claim the moderated drinking extends your life has been refuted by science? I can’t believe this claim.”	
Supporting Formal Treatment of Addiction	Female Support	“I want to find a group that has members who have common ground with me as a young female if possible.”
	Higher Power Concerns	“After going back to NA I remember why i stopped. I just cant figure out how to reconcile not believing win god with the 12-steps. I know if supposed to be spiritual not religious but i just have such a hard time with how that works in my head.”
	Newcomer Support	“As long as you go to an AA meeting you’re at the best one. Any meeting should welcome you.”

Table 2: Example themes that show what the community members are discussing to support each other’s recovery. Includes themes identified and paraphrased example quotations from the subreddits.

individual and you really need to talk to your doctor about what is right for you.”

Similarly, when discussing experiences with the justice system they would emphasize that they were not lawyers and contacting proper representation was important. From these discussions we learned that the communities acknowledged limitations of their experiential knowledge and sought to cooperatively contribute to their members’ addiction recovery journeys rather than replace other contributors.

4.2.3 Supporting Formal Treatment of Addiction. We observed widespread efforts to support formal treatment, such as detox clinics, and 12-step mutual aid groups, such as AA and NA, in both subreddits. Community members encouraged one another to seek out and make use of these supports as part of living in recovery. Also, both subreddits served as a platform for members with concerns about these groups to seek out advice from the rest of the community. In these cases, we postulate that Reddit’s pseudonymity helped members to seek out this advice without risk to their in-person relationships. For instance, we observed a member raise concerns about the in-person nature of AA and NA violating their anonymity when seeking advice about how to handle a relapse:

“While I am involved in twelve step programs for both alcohol and opiates many of the people there are connected to my halfway which makes me worried

my family will be called or that I might get kicked out”

Since the subreddit is pseudonymous and separate from the AA and NA groups, it allows community members to discuss their questions and concerns. In these cases, the subreddits offer access, through the Reddit platform, to a large supplementary group of peers that can offer assistance in navigating 12-step groups.

In other cases, the subreddit helped members struggling with aspects of the 12-step programs, such as how to seek out advice on how to participate. In particular, we noted examples where women (e.g., Table 2, Supporting Formal Treatment of Addiction) and atheists/agnostics sought advice on how to participate in 12-step mutual aid groups that are predominantly male focused, and founded by people with a belief in a higher power:

“While I respect that everyone has different beliefs that may involve religion it just isn’t for me. When I see higher power I have a hard time accepting that it’s not a reference to God. Any suggestions?”

The discussion about the above quote included responses such as the following that the original poster responded to with thanks:

» “When I was in a similar situation I found the following quote by Jung helpful:

God is the name by which I designate all things which cross my path violently and recklessly, all things which alter my plans and intentions, and change the course of my life, for better or for worse.”

We also observed discussions that emphasized difficulties when discussing such concerns at AA meetings due to perceptions of judgment:

“Someone at my AA meeting was explaining how your higher power doesn’t have to be god. However when I looked around there was a lot of head shaking and silent judgment. What was going on with this? Is there something I’m not understanding?”

» “Some meetings involve a lot of God talk to the point it makes some people think AA is religious. When someone explains how to take the steps less literally that is when the silent judgment happens (which is unfortunate).”

» » “Thank you!”

In addition, members expressed concern over being stigmatized in-person by users of other substances. For example, one community member who wanted to attend in-person 12-step meetings felt stigmatized because they were an opiate user instead of an alcoholic:

“Why are good meetings so hard to find? I’m new to abstinence from opiates and the first NA meeting I tried to go to landed me in the middle of the ghetto. When I tried to broaden my search in a meeting finder app to include AA meetings but then I got rejected because I identified as a heroin addict instead of an alcoholic. So I tried only looking at NA, but I once again ended up at a meeting location that was in a tiny room, in the ghetto, and had no parking.

Why are these meetings so hard to find?? I can’t understand this since I live in a large city. Does anyone else find they have to bend over backwards to find a god damn meeting to attend??”

Multiple community members responded with advice, such as:

» “You could try identifying yourself as an alcoholic and just not mentioning that you are a drug addict until you feel out the meeting for mentions of drugs.”

Seeing concerns and responses like these showed us that the Reddit platform was enabling community members to have pseudonymous meta discussions about AA and NA that might be viewed as inappropriate at the meetings themselves. In all of these cases the community was supportive of its members, and tried to work through their concerns with a focus on finding what could best support their recovery.

5 DISCUSSION & IMPLICATIONS

Our results show a considerable depth of support and richness of lived experience present in online communities that largely mirrors best practice from the medical community. These findings contrast existing work by showing diversity in community members and their lived experiences. For instance, previous work has raised

concerns that members may be new to sobriety or may give inappropriate or harmful advice (e.g., [87, 93]). Indeed, we identified that discussions on Reddit enable communication between different groups impacted by addiction, including those with years of recovery experience, those just starting out on their own journeys, and family members and friends. Community members often raised concerns about sensitive, addiction-related issues such as withdrawal, body weight, legal trouble, and personal finances. We also observed discussions covering concepts such as stories, seeking advice, and informational support [58, 86, 87, 98, 102].

Our results also show the important meta-communicative role that these communities play in supporting formal treatment of addiction. Community members provided encouragement to seek out help, advice on navigating 12-step programs like AA and NA, and helped others as they struggled with differences in norms and values. These discussions were particularly valuable for under-represented groups like women struggling with a lack of female mentors, and for atheists/agnostics with concerns about references to a higher power. These examples provide a valuable counterpoint to past work which has found that online communities may be perceived as being less effective and of lower quality than face-to-face meetings [7], and that individuals may be less willing to disclose sensitive information online [87]. This contrast suggests the communities play a distinct and important complementing role by supporting queries that may be difficult to address in person.

These observations add to a growing body of research (e.g., [3, 5, 76, 103]) that shows how Reddit provides a distinct environment where members feel comfortable seeking out a network of support, and where they can share personal or stigmatized experiences that they wouldn’t necessarily disclose with in-person groups. We now discuss implications of this work for both research and practice moving forward, and reflect on the usefulness of topic-guided thematic analysis in conducting the analysis.

5.1 Implications for Addiction Recovery Programs Online and Offline

Our analysis shows how communities on Reddit provide mediators for positive recovery outcomes identified in the literature, like access to recovery role models, abstinent contingent social support, and an environment that increases commitment to recovery [11, 48, 52, 70]. For example, we observed that check-in threads provide a useful parallel to 12-step meetings and enable community members to mark milestones together and celebrate successes. Members explore new recreational activities and strategies to handle difficulties such as social changes, and experienced members act as mentors, matching the role that they would play for in-person 12-step programs. Similarly, communities help members develop new friendships through social activities, such as movie nights, reading circles, and song sharing groups. These elements correspond to emphases on in-person mutual aid groups like education, development of coping skills, and management of co-occurring symptoms [1].

The similarity between online and in-person support groups, and the substantial number of active members and discussions on Reddit, raises the question of whether in-person groups, such as AA and NA, might benefit from the online communities’ lived experiences.

In-person groups are typically much smaller than the 6000 people we observed engaged online in *r/stopdrinking*, with about 17 members on average [78]. These groups may also benefit from the many different perspectives online, including access to under-represented groups, that may not always be present in-person. Our results show how the large community and body of experiential knowledge work as a supplementary resource for in-person mutual aid groups, and highlights opportunities for HCI to help people leverage these resources moving forward.

The potential benefits of online communities to people who attend in-person aid groups are myriad. For instance, more regular peer support like daily check-in threads may be impractical in person, but may benefit some members. Shared activities, like synchronous online movie watching, may be valuable to those who typically attend in-person meetings but have difficulty finding in-person alternatives to substance use. Mentors may not have the resources they need to advise other members, but may find those resources online through shared stories, frequently asked questions about consequences, or more niche concerns such as those around pain management.

Thus, there is an opportunity for HCI researchers to make the valuable lived experience from online groups more readily available to in-person groups. For example, addiction treatment specialists could use Reddit's experiential knowledge to create and share a curated selection of frequently asked questions, advice, and activities for in-person addiction treatment programs. Computational techniques, like those used in our analysis, may be particularly effective in identifying concepts of interest and potential alternative treatments (e.g., [23]).

5.2 Expert Use of Online Experiential Knowledge

Our analysis also shows how communities like *r/stopdrinking* and *r/OpiatesRecovery* exist alongside of the medical profession and traditional mutual aid groups. These communities are dedicated to responding to questions and discussing sensitive issues, and are an important source of patient expertise and lived experience [96]. For example, we observed discussions on consequences of recovery, which included concerns about addiction recovery treatment costs and risks involving insurance, drugs, and rehab facilities. The separation from medical expertise is both a strength and weakness. We saw that these communities help people on their journeys of recovery by discussing sensitive and highly personal issues such as pain management alternatives, different consequences of addiction and recovery, and getting the right support from formal treatment. Although the communities make efforts to be clear they are not sources of medical expertise, providing advice may also mean that inappropriate, out-of-date, or even incorrect information is being shared, and that there is potential for harm [33, 36, 45] — particularly when discussing medication, e.g., incorrectly changing dosages could lead to overdosing. By understanding that patient expertise and medical expertise are complimentary [40, 46] they can be used together to gain a more holistic understanding of addiction recovery journeys.

Bridging treatment specialists into these communities is a tempting way for HCI research to contribute clinical knowledge to the

discussions. For instance, previous work has focused on how to weave specialists' expertise into online communities [47] and the potential to supplement posts with clinician validation [23]. However, when learning from and interacting with these communities, it is important for both researchers and specialists to be respectful and non-judgmental as not dismiss patient expertise or disrupt community norms and rules; they are valuable safe spaces for members that serve a collaborative role in recovery journeys.

An alternative approach informed by our research is to consider what support these communities want from specialists, and what clinicians can gain from an understanding of the issues discussed. For example, we identified that pain management is a common concern in *r/OpiatesRecovery*. People discussed being afraid to see a doctor about an illness out of a fear that they begin using an addictive substance again. They also discussed how to interact with medical providers in ways that could avoid common prescriptions, and to seek out alternative treatments. The communities' online discussions can be additional sources of patient experiences for medical training approaches, such as narrative medicine [69] and situated learning [55] which enhance awareness and empathy, to help practitioners better understand and respond to the experiences and needs of patients impacted by substance use disorders. Similarly the discussions of social isolation in *r/stopdrinking* could serve as an additional source of patient experiences for therapists being trained on how to emphasize with and assist patients attending recovery programs. HCI practitioners working in the social computing space can play a valuable role by developing new tools that can help experts access this experiential knowledge in ways that respect the anonymity of online communities.

5.3 Topic-Guided Thematic Analysis

Finally, in reflecting on our topic-guided process, we found that the opportunity to 'prime' our thematic analysis with a qualitative understanding of topics was invaluable. The LDA topics helped us sample in a way that allowed us to develop both general themes, like sharing stories [98], but also novel and nuanced themes that are specific to the context of addiction recovery, such as difficulties with pain management and supporting formal treatment of addiction. This process aligned with our goal of obtaining a 'thick' understanding [18] of the communities' experiential knowledge of addiction recovery, while mitigating challenges of scalability when studying very large online communities. We argue that use of LDA to guide thematic analysis provided two key benefits: 1) the work performed as part of the LDA modelling process was informative in itself, and 2) purposive sampling based on generated topics improved our thematic analysis.

First, we found that the *iterative* nature of the process, and human involvement in developing and interpreting the model was invaluable to our thematic analysis. While our topic modelling activities leveraged computational techniques (i.e., left side of Figure 1), a human researcher played a substantial role in iteratively cleaning data, building models, and evaluating their utility. For instance, we first became aware of the 'IWNDWYT' acronym when cleaning data, which prompted us to pay more attention to it later on. When evaluating models, we noticed that optimizing around quantitative measures like coherence [88] and Jaccard's distance did not yield a

single solution, but did give us an opportunity to engage with and reflect on topics from multiple perspectives.

Second, we believe that purposively sampling from each topic improved our thematic analysis (i.e., right side of Figure 1). Sampling from each topic ensured that a range of data was represented in our thematic analysis, rather than simply those that were ‘hot’ on a given day (e.g., [33, 62, 98]). Sampling an equal number from each topic also ensured that a range of data was represented, compared to, for instance, a random sample (e.g., [98]). Topics often provided invaluable ‘hints’ for thematic analysis, and ultimately some translated directly to themes (e.g., Sharing Experiences), some were combined into broader themes (e.g., Activities, Consequences), and even topics that were less coherent provided data to code and helped us develop and review our themes.

To realize such benefits we needed to consider whether LDA modelling was an appropriate technique for our research questions and intended methods. We determined that because our unit of analysis, discussions, is long-form text with multiple authors and multiple co-occurring topics, LDA was an appropriate method for purposive sampling. However, for other analyses and research contexts, a different choice of topic model may be more suitable. For instance, for discussions on Twitter data is more sparse and biterm modelling [101] or LDA with author aggregation [43] are likely to better capture latent topics. Similarly, if the topic models were intended to provide a baseline for deductive theme and code identification, then techniques with enhanced within-topic aspect identification, such as Attention-based Aspect Extraction [41], should be considered.

Our work fits within a larger body of computational social science research that explores how technology can help qualitative understanding by human researchers (e.g., [35, 71]). We demonstrated a specific application of thematic analysis where computing *supports*, not replaces, a human researcher. Indeed, we believe there are opportunities to more deeply integrate computational techniques into thematic analysis phases like identifying codes, developing themes, and reviewing themes (Figure 1). In the interest of transparency [94], and to enable others to expand on our work, we have shared our full source code in the paper’s supplementary materials.

6 LIMITATIONS

As qualitative research, our work develops an understanding of addiction recovery communities on Reddit and the topics that they discuss. These methods have the advantage of allowing us to engage with materials from the ‘wild’, derive themes from our observations, and to validate our findings within their original context. However, the choice to carefully study two subreddits focused on substance use also has limitations, particularly when generalizing to the broader addiction recovery community, or when comparing to communities centred on, for example, weight loss, fitness, or mental health. For instance, substance-specific issues like stigma and local legality for alcohol or opiate use differ from those for smoking or foods. In this work, we intentionally focused on understanding the under-explored supports for substance use on Reddit, and while we identified themes like stories and seeking advice that have also been found in contexts like smoking cessation (e.g., [98]), additional research is required to examine similarities and differences across

communities for the many distinct, but related, medical diagnoses grounded in addiction.

We also needed to manage some limitations of our topic modelling approach, and to balance model optimization with our ultimate goal of performing a *qualitative* analysis of discussions in these communities. As Baumer et al. [8] notes, the models we generated needed to provide “scaffolding for human interpretation.” We came to understand the importance of *good enough* models which provide new perspectives of the data, and the trap of searching for the *best* model. In short, perfect was the enemy of useful. To develop the models we used best practices for LDA analysis [59], including practices for data cleaning, and selecting a number of topics for each model based on maximizing its reliability and coherence scores. We also discussed pilot models as a group and identified features of the data that were making topics both interpretable and uninterpretable. These discussions informed our tokenization, lemmatization, and filtering of stop words, frequent terms, and infrequent terms. However, the interpretive nature of both model development and thematic analysis are simultaneously a strength and a limitation of our methods [8, 71].

Finally, another limitation is that we have little information about the people who posted online, or their motivations for doing so. Since Reddit is a pseudonymous community, we cannot infer any direct relationships between accounts, or between accounts and people. We therefore are limited in our ability to accurately generate descriptive statistics, like the number of posts or threads created by any individual, or to more fully explore interactions between individuals in these communities. We also can only consider those who posted in the forums, and have no means of understanding how others use this information (i.e., lurkers [77]), or whether individuals sought support through other means like in-person 12-step groups or other online communities [26].

7 CONCLUSIONS & FUTURE WORK

Our work shows how online communities such as Reddit support addiction recovery. It provides an understanding of the issues discussed in these communities, how they take advantage of pseudonymity and an online format to support one another, family, and friends, and the stories, advice, and emotional support that they share. We also identified that these online communities play a role in helping people navigate personal concerns about 12-step programs. Our results can be used by the HCI community, addiction recovery programs, and healthcare practitioners to develop a more holistic understanding of how online peer-to-peer communities are leveraging the Reddit platform around sensitive and health-related issues such as addiction recovery.

Our work also further explores the application of computational techniques to support qualitative research [35, 71]. We built on the ‘roadmap’ set out by Ammari et al. [3] to gain insights into the needs of stigmatized groups, like those in addiction recovery, via social networks. We show how the complementary use of computational and qualitative techniques, in this case the use of topic modelling to purposively sample for thematic analysis, can yield insights into the types of discussions occurring online at scale, and allow human researchers to more deeply engage with those discussions. We expect to see tighter integration of qualitative research and

computing moving forward, and that our work can serve as one model of partnership between human- and machine-guided analysis techniques. As a next step towards making these models more accessible to qualitative researchers, we are currently developing an open-source graphical toolkit for topic-guided analysis of online discussions, called the Computational Thematic Analysis Toolkit (e.g., [38]).

ACKNOWLEDGMENTS

We thank the University of Waterloo Games Institute and its members for their constructive feedback. We would also like to thank all of our anonymous reviewers for their valuable feedback, which particularly helped us think through the ethical considerations of our work. This work was made possible by NSERC Discovery Grant 2015-06585 and the Ontario Graduate Scholarship.

REFERENCES

- [1] American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. American Psychiatric Pub, 800 Maine Avenue, S.W., Suite 900, Washington, DC 20024.
- [2] Tawfiq Ammari, Sarita Schoenebeck, and Daniel Romero. 2019. Self-declared Throwaway Accounts on Reddit: How Platform Affordances and Shared Norms Enable Parenting Disclosure and Support. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW (Nov 2019), 135:1–135:30. <https://doi.org/10.1145/3359237>
- [3] Tawfiq Ammari, Sarita Schoenebeck, and Daniel M. Romero. 2018. Pseudonymous Parents: Comparing Parenting Roles and Identities on the Mommit and Daddit Subreddits. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 489:1–489:13. <https://doi.org/10.1145/3173574.3174063>
- [4] Nazanin Andalibi and Andrea Forte. 2018. Announcing Pregnancy Loss on Facebook: A Decision-Making Framework for Stigmatized Disclosures on Identified Social Network Sites. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (Montreal QC, Canada) (CHI '18)*. ACM, New York, NY, USA, Article 158, 14 pages. <https://doi.org/10.1145/3173574.3173732>
- [5] Nazanin Andalibi, Oliver L. Haimson, Munmun De Choudhury, and Andrea Forte. 2016. Understanding Social Media Disclosures of Sexual Abuse Through the Lenses of Support Seeking and Anonymity. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 3906–3918. <https://doi.org/10.1145/2858036.2858096> event-place: San Jose, California, USA.
- [6] Nazanin Andalibi, Pinar Ozturk, and Andrea Forte. 2017. Sensitive Self-disclosures, Responses, and Social Support on Instagram: The Case of #Depression. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (Portland, Oregon, USA) (CSCW '17)*. ACM, New York, NY, USA, 1485–1500. <https://doi.org/10.1145/2998181.2998243>
- [7] Ashley K. Barrett and Melissa M. Murphy. 2020. Feeling Supported in Addiction Recovery: Comparing Face-to-Face and Videoconferencing 12-Step Meetings. *Western Journal of Communication* 0, 0 (2020), 1–24. <https://doi.org/10.1080/10570314.2020.1786598> arXiv:https://doi.org/10.1080/10570314.2020.1786598
- [8] Eric P. S. Baumer, David Mimno, Shion Guha, Emily Quan, and Geri K. Gay. 2017. Comparing grounded theory and topic modeling: Extreme divergence or unlikely convergence? *Journal of the Association for Information Science and Technology* 68, 6 (2017), 1397–1410. <https://doi.org/10.1002/asi.23786> arXiv:https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/asi.23786
- [9] Eric P. S. Baumer, Drew Siedel, Lena McDonnell, Jiayun Zhong, Patricia Sittikul, and Micki McGee. 2020. Topicalizer: reframing core concepts in machine learning visualization by co-designing for interpretivist scholarship. *Human-Computer Interaction* 35, 5 (2020), 452–480. <https://doi.org/10.1080/07370024.2020.1734460> Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/07370024.2020.1734460>
- [10] Jason Baumgartner, Savvas Zannettou, Brian Keegan, Megan Squire, and Jeremy Blackburn. 2020. The Pushshift Reddit Dataset. *Proceedings of the International AAAI Conference on Web and Social Media* 14 (May 2020), 830–839.
- [11] Martha C. Beattie and Richard Longabaugh. 1997. Interpersonal factors and post-treatment drinking and subjective wellbeing. *Addiction* 92, 11 (1997), 1507–1521. <https://doi.org/10.1111/j.1360-0443.1997.tb02871.x>
- [12] Brandon G. Bergman, Nathaniel W. Kelly, Bettina B. Hoepfner, Corrie L. Vilsaint, and John F. Kelly. 2017. Digital recovery management: Characterizing recovery-specific social network site participation and perceived benefit. *Psychology of Addictive Behaviors* 31, 4 (2017), 506–512. <https://doi.org/10.1037/adb0000255>
- [13] Andrew BL Berry, Catherine Y. Lim, Andrea L. Hartzler, Tad Hirsch, Evette Ludman, Edward H. Wagner, and James D. Ralston. 2017. "It's Good to Know You're Not a Stranger Every Time": Communication About Values Between Patients with Multiple Chronic Conditions and Healthcare Providers. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW (Dec. 2017), 23:1–23:20. <https://doi.org/10.1145/3134658>
- [14] Natalie Berry, Fiona Lobban, Maksim Belousov, Richard Emsley, Goran Nenadic, and Sandra Bucci. 2017. #WhyWeTweetMH: Understanding Why People Use Twitter to Discuss Mental Health Problems. *Journal of Medical Internet Research* 19, 4 (2017), e107. <https://doi.org/10.2196/jmir.6173>
- [15] Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural Language Processing with Python* (1st ed.). O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.
- [16] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research* 3, Jan (2003), 993–1022. <http://www.jmlr.org/papers/v3/blei03a.html>
- [17] Rosemary A. Boisvert, Linda M. Martin, Maria Grosek, and Anna J. Clarie. 2008. Effectiveness of a peer-support community in addiction recovery: participation as intervention. *Occupational Therapy International* 15, 4 (12 2008), 205–220. <https://doi.org/10.1002/oti.257>
- [18] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. <https://doi.org/10.1191/1478088706qp0630a> arXiv:https://www.tandfonline.com/doi/pdf/10.1191/1478088706qp0630a
- [19] Virginia Braun and Victoria Clarke. 2020. One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology* 0, 0 (Aug 2020), 1–25. <https://doi.org/10.1080/14780887.2020.1769238>
- [20] Amy Bruckman. 2002. Studying the amateur artist: A perspective on disguising data collected in human subjects research on the Internet. *Ethics and Information Technology* 4, 3 (2002), 217–231. <https://doi.org/10.1023/A:1021316409277>
- [21] Anja Busse, G. Gerra, Igor Koutsenok, and Elizabeth Saenz. 2015. International standards on the treatment of drug use disorders. *Drug and Alcohol Dependence* 156 (Nov. 2015), e32. <https://doi.org/10.1016/j.drugalcdep.2015.07.1006>
- [22] Canadian Substance Use Costs and Harms Scientific Working Group. 2018. *Canadian substance use costs and harms (2007–2014)*. Technical Report. Institute for Substance Use Research and the Canadian Centre on Substance Use and Addiction, Ottawa, Ont.: Canadian Centre on Substance Use and Addiction. <http://www.ccdus.ca/ResourceLibrary/CSUCH-Canadian-Substance-Use-Costs-Harms-Report-2018-en.pdf>
- [23] Stevie Chancellor, George Nitzburg, Andrea Hu, Francisco Zampieri, and Munmun De Choudhury. 2019. Discovering Alternative Treatments for Opioid Use Recovery Using Social Media. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, 124:1–124:15. <https://doi.org/10.1145/3290605.3300354> event-place: Glasgow, Scotland Uk.
- [24] Nan-Chen Chen, Margaret Drouhard, Rafal Kocielnik, Jina Suh, and Cecilia R. Aragon. 2018. Using Machine Learning to Support Qualitative Coding in Social Science: Shifting the Focus to Ambiguity. *ACM Trans. Interact. Intell. Syst.* 8, 2 (Jun 2018), 9:1–9:20. <https://doi.org/10.1145/3185515>
- [25] Katherine Y. Chuang and Christopher C. Yang. 2012. A Study of Informal Support Exchanges in MedHelp Alcoholism Community. In *Social Computing, Behavioral - Cultural Modeling and Prediction*, Shanchieh Jay Yang, Ariel M. Greenberg, and Mica Endsley (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 9–17.
- [26] Amy M Cohn, Michael S Amato, Kang Zhao, Xi Wang, Sarah Cha, Jennifer L Pearson, George D Papandonatos, and Amanda L Graham. 2019. Discussions of alcohol use in an online social network for smoking cessation: analysis of topics, sentiment, and social network centrality. *Alcoholism: Clinical and Experimental Research* 43, 1 (2019), 108–114.
- [27] Kaitlin Light Costello. 2019. Quality of medical advice for chronic pain on social platforms. *Proceedings of the Association for Information Science and Technology* 56, 1 (2019), 381–385.
- [28] Mary Jean Costello, Yao Li, Shannon Remers, James MacKillop, Sarah Sousa, Courtney Ropp, Don Roth, Mark Weiss, and Brian Rush. 2019. Effects of 12-step mutual support and professional outpatient services on short-term substance use outcomes among adults who received inpatient treatment. *Addictive Behaviors* 98 (Nov 2019), 106055. <https://doi.org/10.1016/j.addbeh.2019.106055>
- [29] John W. Creswell and Vicki L. Plano Clark. 2018. *Designing and Conducting Mixed Methods Research* (4th ed.). SAGE, 2455 Teller Road, Thousand Oaks, California.
- [30] Paul DiMaggio, Manish Nag, and David Blei. 2013. Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding. *Poetics* 41, 6 (2013), 570–606. <https://doi.org/10.1016/j.poetic.2013.08.004> Topic Models and the Cultural Sciences.
- [31] Karthik Dinakar, Birago Jones, Henry Lieberman, Rosalind Picard, Carolyn Rose, Matthew Thoman, and Roi Reichart. 2012. You too?! mixed-initiative lda story matching to help teens in distress. In *Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media (ICWSM 2012)*. Association for the Advancement of Artificial Intelligence, Palo Alto, California, 74–81.

- [32] Nick Doukas. 2011. Perceived barriers to identity transformation for people who are prescribed methadone. *Addiction Research & Theory* 19, 5 (Oct 2011), 408–415. <https://doi.org/10.3109/16066359.2010.530715>
- [33] Alexandra R. D'Agostino, Allison R. Optican, Shaina J. Sowles, Melissa J. Krauss, Kiriam Escobar Lee, and Patricia A. Cavazos-Rehg. 2017. Social networking online to recover from opioid use disorder: A study of community interactions. *Drug and Alcohol Dependence* 181 (Dec. 2017), 5–10. <https://doi.org/10.1016/j.drugalcdep.2017.09.010>
- [34] M. Eickhoff and Runhild Wieneke. 2018. Understanding Topic Models in Context: A Mixed-Methods Approach to the Meaningful Analysis of Large Document Collections. In *Proceedings of the 51st Hawaii International Conference on System Sciences*. University of Hawaii at Manoa, Honolulu, Hawaii, 903–912.
- [35] James A. Evans and Pedro Aceves. 2016. Machine Translation: Mining Text for Social Theory. *Annual Review of Sociology* 42, 1 (2016), 21–50. <https://doi.org/10.1146/annurev-soc-081715-074206>
- [36] Gunther Eysenbach. 2005. Patient-to-Patient Communication: Support Groups and Virtual Communities. In *Consumer Health Informatics: Informing Consumers and Improving Health Care*, Deborah Lewis, Gunther Eysenbach, Rita Kukafka, P. Zoë Stavri, and Holly B. Jimison (Eds.). Springer New York, New York, NY, 97–106. https://doi.org/10.1007/0-387-27652-1_8
- [37] Manas Gaur, Ugur Kursuncu, Amanuel Alambo, Amit Sheth, Raminta Daniulaityte, Krishnaprasad Thirunarayan, and Jyotishman Pathak. 2018. "Let Me Tell You About Your Mental Health!": Contextualized Classification of Reddit Posts to DSM-5 for Web-Based Intervention. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management (Torino, Italy) (CIKM '18)*. Association for Computing Machinery, New York, NY, USA, 753–762. <https://doi.org/10.1145/3269206.3271732>
- [38] Robert P. Gauthier and James R. Wallace. 2022. The Computational Thematic Analysis Toolkit. In *Proceedings of the 22nd International Conference on Supporting Group Work*. *Proc. ACM Hum.-Comput. Interact.* 6, GROUP, Article 25 (January 2022), 15 pages. <https://doi.org/10.1145/3492844>
- [39] Simon Graham, Jamie Irving, Ivan Cano, and Edwards Michael. 2018. Participation with online recovery specific groups - findings from the UK Life in Recovery survey 2015. *Alcoholism Treatment Quarterly* 0, 0 (2018), 1–23. <https://doi.org/10.1080/07347324.2018.1500873>
- [40] Andrea Hartzler and Wanda Pratt. 2011. Managing the Personal Side of Health: How Patient Expertise Differs from the Expertise of Clinicians. *Journal of Medical Internet Research* 13, 3 (2011), e1728. <https://doi.org/10.2196/jmir.1728>
- Company: Journal of Medical Internet Research Distributor: Journal of Medical Internet Research Institution: Journal of Medical Internet Research Label: Journal of Medical Internet Research Publisher: JMIR Publications Inc., Toronto, Canada.
- [41] Ruidan He, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier. 2017. An Unsupervised Neural Attention Model for Aspect Extraction. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (2017-07)*. Association for Computational Linguistics, Vancouver, Canada, 388–397. <https://doi.org/10.18653/v1/P17-1036>
- [42] Orland Hoerber, Larena Hoerber, Ryan Snelgrove, and Laura Wood. 2017. Interactively Producing Purposive Samples for Qualitative Research using Exploratory Search. In *SCST@ CHIIR*. CEUR-WS, Oslo, Norway, 18–20.
- [43] Liangjie Hong and Brian D. Davison. 2010. Empirical study of topic modeling in Twitter. In *Proceedings of the First Workshop on Social Media Analytics (2010-07-25) (SOMA '10)*. Association for Computing Machinery, New York, NY, USA, 80–88. <https://doi.org/10.1145/1964858.1964870>
- [44] Matthew Honnibal and Ines Montani. 2018. spaCy · Industrial-strength Natural Language Processing in Python. <https://spacy.io/>
- [45] Jina Huh, David W. McDonald, Andrea Hartzler, and Wanda Pratt. 2013. Patient Moderator Interaction in Online Health Communities. *AMIA Annual Symposium Proceedings* 2013 (Nov. 2013), 627–636. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900205/>
- [46] Jina Huh, Rupa Patel, and Wanda Pratt. 2012. Tackling dilemmas in supporting 'the whole person' in online patient communities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (2012-05-05) (CHI '12)*. Association for Computing Machinery, New York, NY, USA, 923–926. <https://doi.org/10.1145/2207676.2208535>
- [47] Jina Huh and Wanda Pratt. 2014. Weaving Clinical Expertise in Online Health Communities. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 1355–1364. <https://doi.org/10.1145/2556288.2557293> event-place: Toronto, Ontario, Canada.
- [48] Keith Humphreys. 2003. *Circles of Recovery: Self-Help Organizations for Addictions*. Cambridge University Press, Cambridge, UK. <https://doi.org/10.1017/CBO9780511543883>
- [49] InTheRooms. 2019. In The Rooms. <https://www.intherooms.com/>
- [50] Carina Jacobi, Wouter van Atteveldt, and Kasper Welbers. 2016. Quantitative analysis of large amounts of journalistic texts using topic modelling. *Digital Journalism* 4, 1 (2016), 89–106. <https://doi.org/10.1080/21670811.2015.1093271> Publisher: Routledge_eprint: <https://doi.org/10.1080/21670811.2015.1093271>
- [51] Deepanshu Jha and Rahul Singh. 2020. Analysis of associations between emotions and activities of drug users and their addiction recovery tendencies from social media posts using structural equation modeling. *BMC Bioinformatics* 21, 18 (2020), 554. <https://doi.org/10.1186/s12859-020-03893-9>
- [52] John F. Kelly. 2017. Is Alcoholics Anonymous religious, spiritual, neither? Findings from 25 years of mechanisms of behavior change research. *Addiction* 112, 6 (2017), 929–936. <https://doi.org/10.1111/add.13590>
- [53] John Francis Kelly, Molly Magill, and Robert Lauren Stout. 2009. How do people recover from alcohol dependence? A systematic review of the research on mechanisms of behavior change in Alcoholics Anonymous. *Addiction Research & Theory* 17, 3 (Jan. 2009), 236–259. <https://doi.org/10.1080/16066350902770458>
- [54] John F. Kelly and William L. White. 2012. Broadening the Base of Addiction Mutual-Help Organizations. *Journal of Groups in Addiction & Recovery* 7, 2-4 (2012), 82–101. <https://doi.org/10.1080/1556035X.2012.705646>
- [55] Kwo-Chen Lee, Chin-Ching Yu, Pei-Ling Hsieh, Chin-ching Li, and Yann-Fen C. Chao. 2018. Situated teaching improves empathy learning of the students in a BSN program: A quasi-experimental study. *Nurse Education Today* 64 (May 2018), 138–143. <https://doi.org/10.1016/j.nedt.2018.02.013>
- [56] Vitals Consumer Services LLC. 2019. MedHelp - Health community, health information, medical questions, and medical apps. <https://www.medhelp.org/>
- [57] John Lu, Sumati Sridhar, Ritika Pandey, Mohammad Al Hasan, and Georege Mohler. 2019. Investigate Transitions into Drug Addiction through Text Mining of Reddit Data. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (Anchorage, AK, USA) (KDD '19)*. Association for Computing Machinery, New York, NY, USA, 2367–2375. <https://doi.org/10.1145/3292500.3330737>
- [58] Diana MacLean, Sonal Gupta, Anna Lembke, Christopher Manning, and Jeffrey Heer. 2015. Forum77: An Analysis of an Online Health Forum Dedicated to Addiction Recovery. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (Vancouver, BC, Canada) (CSCW '15)*. ACM, New York, NY, USA, 1511–1526. <https://doi.org/10.1145/2675133.2675146>
- [59] Daniel Maier, A. Waldherr, P. Miltner, G. Wiedemann, A. Niekler, A. Keinert, B. Pfetsch, G. Heyer, U. Reber, T. Häussler, H. Schmid-Petri, and S. Adam. 2018. Applying LDA Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology. *Communication Methods and Measures* 12, 2-3 (2018), 93–118. <https://doi.org/10.1080/19312458.2018.1430754> arXiv:https://doi.org/10.1080/19312458.2018.1430754
- [60] Lena Mamykina, Andrew D. Miller, Elizabeth D. Mynatt, and Daniel Greenblatt. 2010. Constructing Identities Through Storytelling in Diabetes Management. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Atlanta, Georgia, USA) (CHI '10)*. ACM, New York, NY, USA, 1203–1212. <https://doi.org/10.1145/1753326.1753507>
- [61] Annette Markham. 2012. Fabrication as ethical practice: Qualitative inquiry in ambiguous internet contexts. *Information, Communication & Society* 15, 3 (2012), 334–353. <https://doi.org/10.1080/1369118X.2011.641993>
- [62] Martin N. Marshall. 1996. Sampling for qualitative research. *Family Practice* 13, 6 (Jan 1996), 522–526. <https://doi.org/10.1093/fampra/13.6.522>
- [63] Alice E. Marwick and Danah Boyd. 2011. I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society* 13, 1 (Feb. 2011), 114–133. <https://doi.org/10.1177/1461444810365313>
- [64] Carmen L. Masson, Michael S. Shopshire, Soma Sen, Kim Hoffman, Nicholas Hengli, John Bartolome, Dennis McCarty, James L. Sorensen, and Martin Iguchi. 2013. Possible Barriers to Enrollment in Substance Abuse Treatment among a Diverse Sample of Asian Americans and Pacific Islanders: Opinions of Treatment Clients. *Journal of substance abuse treatment* 44, 3 (Apr 2013), 309–315. <https://doi.org/10.1016/j.jsat.2012.08.005>
- [65] Andrew Kachites McCallum. 2002. MALLET: A Machine Learning for Language Toolkit. <http://mallet.cs.umass.edu/>
- [66] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and Inter-rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. In *Proc. ACM Hum.-Comput. Interact.* 39, ACM, New York, NY, USA, 23. <https://doi.org/10.1145/3359174>
- [67] R.J. McQuaid, A. Malik, K. Moussouni, N. Baydack, M. Stargardt, and M. Morrissey. 2017. *Life in Recovery from Addiction in Canada*. Technical Report. Canadian Centre on Substance Use and Addiction, Ottawa, Ont. <http://www.ccdus.ca/ResourceLibrary/CCSA-Life-in-Recovery-from-Addiction-Report-2017-en.pdf>
- [68] Yelena Mejova, Youcef Benkhedda, and Khairani Khairani. 2017. #Halal Culture on Instagram. *Frontiers in Digital Humanities* 4 (2017), 21. <https://doi.org/10.3389/fdigh.2017.00021>
- [69] M. M. Milota, G. J. M. W. van Thiel, and J. J. M. van Delden. 2019. Narrative medicine as a medical education tool: A systematic review. *Medical Teacher* 41, 7 (Jul 2019), 802–810. <https://doi.org/10.1080/0142159X.2019.1584274>
- [70] Rudolf H. Moos. 2007. Theory-based active ingredients of effective treatments for substance use disorders. *Drug and Alcohol Dependence* 88, 2 (May 2007), 109–121. <https://doi.org/10.1016/j.drugalcdep.2006.10.010>
- [71] Michael Muller, Shion Guha, Eric P.S. Baumer, David Mimmo, and N. Sadat Shami. 2016. Machine Learning and Grounded Theory Method: Convergence, Divergence and Combination. In *Proceedings of the 19th International Conference*

- on Supporting Group Work (GROUP '16). ACM, New York, NY, USA, 3–8. <https://doi.org/10.1145/2957276.2957280>
- [72] Emily van der Nagel and Jordan Frith. 2015. Anonymity, pseudonymity, and the agency of online identity: Examining the social practices of r/Gonewild. *First Monday* 20, 3 (Feb. 2015), 1 pages. <https://doi.org/10.5210/fm.v20i3.5615>
- [73] National Institute on Drug Abuse. 2018. Principles of Drug Addiction Treatment: A Research-Based Guide (Third Edition). <https://www.drugabuse.gov/publications/>
- [74] Laura K. Nelson. 2020. Computational Grounded Theory: A Methodological Framework. *Sociological Methods & Research* 49, 1 (2020), 3–42. <https://doi.org/10.1177/0049124117729703> arXiv:<https://doi.org/10.1177/0049124117729703>
- [75] David Newman, Jey Han Lau, Karl Grieser, and Timothy Baldwin. 2010. Automatic Evaluation of Topic Coherence. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics (HLT '10)*. Association for Computational Linguistics, USA, 100–108. <http://dl.acm.org/citation.cfm?id=1857999.1858011> event-place: Los Angeles, California.
- [76] Mark W. Newman, Debra Lauterbach, Sean A. Munson, Paul Resnick, and Margaret E. Morris. 2011. It's Not That I Don't Have Problems, I'm Just Not Putting Them on Facebook: Challenges and Opportunities in Using Online Social Networks for Health. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work (Hangzhou, China) (CSCW '11)*. ACM, New York, NY, USA, 341–350. <https://doi.org/10.1145/1958824.1958876>
- [77] Blair Nonnecke and Jenny Preece. 2000. Lurker Demographics: Counting the Silent. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (The Hague, The Netherlands) (CHI '00)*. ACM, New York, NY, USA, 73–80. <https://doi.org/10.1145/332040.332409>
- [78] General Service Office. 2019. Estimates of A.A. Groups and Memberships. https://www.aa.org/assets/en_US/smf-53_en.pdf
- [79] Sanghee Oh. 2012. The characteristics and motivations of health answerers for sharing information, knowledge, and experiences in online environments. *Journal of the American Society for Information Science and Technology* 63, 3 (2012), 543–557. <https://doi.org/10.1002/asi.21676>
- [80] Kathleen O'Leary, Stephen M. Schueller, Jacob O. Wobbrock, and Wanda Pratt. 2018. "Suddenly, We Got to Become Therapists for Each Other": Designing Peer Support Chats for Mental Health. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (Montreal QC, Canada) (CHI '18)*. ACM, New York, NY, USA, Article 331, 14 pages. <https://doi.org/10.1145/3173574.3173905>
- [81] National Institute on Drug Abuse. 2019. Drugs of Abuse. <https://www.drugabuse.gov/drugs-abuse>
- [82] Gisele Lobo Pappa, Tiago Oliveira Cunha, Paulo Viana Bicalho, Antonio Ribeiro, Ana Paula Couto Silva, Wagner Meira Jr, and Alline Maria Rezende Beleigoli. 2017. Factors Associated With Weight Change in Online Weight Management Communities: A Case Study in the LoseIt Reddit Community. *Journal of Medical Internet Research* 19, 1 (2017), e17. <https://doi.org/10.2196/jmir.5816>
- [83] Albert Park and Mike Conway. 2018. Tracking Health Related Discussions on Reddit for Public Health Applications. *AMIA Annual Symposium Proceedings* 2017 (Apr 2018), 1362–1371.
- [84] Farzana Rahman, Ivor D. Addo, and Sheikh I. Ahamed. 2014. PriSN: A Privacy Protection Framework for Healthcare Social Networking Sites. In *Proceedings of the 2014 Conference on Research in Adaptive and Convergent Systems (Towson, Maryland) (RACS '14)*. ACM, New York, NY, USA, 66–71. <https://doi.org/10.1145/2663761.2664199>
- [85] Radim Rehurek and Petr Sojka. 2010. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. ELRA, Valletta, Malta, 45–50.
- [86] Sabirat Rubya and Svetlana Yarosh. 2017. Interpretations of Online Anonymity in Alcoholics Anonymous and Narcotics Anonymous. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 91 (Dec. 2017), 22 pages. <https://doi.org/10.1145/3134726>
- [87] Sabirat Rubya and Svetlana Yarosh. 2017. Video-Mediated Peer Support in an Online Community for Recovery from Substance Use Disorders. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (Portland, Oregon, USA) (CSCW '17)*. ACM, New York, NY, USA, 1454–1469. <https://doi.org/10.1145/2998181.2998246>
- [88] Michael Röder, Andreas Both, and Alexander Hinneburg. 2015. Exploring the Space of Topic Coherence Measures. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining (WSDM '15)*. ACM, New York, NY, USA, 399–408. <https://doi.org/10.1145/2684822.2685324>
- [89] Zachary Schmitt and Svetlana Yarosh. 2018. Participatory Design of Technologies to Support Recovery from Substance Use Disorders. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW (Nov. 2018), 156:1–156:27. <https://doi.org/10.1145/3274425>
- [90] Alexandra Schofield, Måns Magnusson, and David Mimno. 2017. Pulling out the stops: Rethinking stopword removal for topic models. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*. Association for Computational Linguistics, Stroudsburg, Pennsylvania, United States, 432–436.
- [91] Alexandra Schofield and David Mimno. 2016. Comparing Apples to Apples: The Effects of Stemmers on Topic Models. *Transactions of the Association for Computational Linguistics* 4 (2016), 287–300. https://doi.org/10.1162/tacl_a_00099
- [92] Chih-Ya Shen, Hong-Han Shuai, De-Nian Yang, Yi-Feng Lan, Wang-Chien Lee, Philip S. Yu, and Ming-Syan Chen. 2015. Forming Online Support Groups for Internet and Behavior Related Addictions. In *Proceedings of the 24th ACM International Conference on Information and Knowledge Management (Melbourne, Australia) (CIKM '15)*. ACM, New York, NY, USA, 163–172. <https://doi.org/10.1145/2806416.2806423>
- [93] Shaina J. Sowles, Melissa J. Krauss, Lewam Gebremedhn, and Patricia A. Cavazos-Rehg. 2017. "I feel like I've hit the bottom and have no idea what to do": Supportive social networking on Reddit for individuals with a desire to quit cannabis use. *Substance Abuse* 38, 4 (2017), 477–482. <https://doi.org/10.1080/08897077.2017.1354956> arXiv:<https://doi.org/10.1080/08897077.2017.1354956> PMID: 28704167.
- [94] Poorna Talkad Sukumar, Ignacio Avellino, Christian Remy, Michael A. DeVito, Tawanna R. Dillahunt, Joanna McGrenere, and Max L. Wilson. 2020. Transparency in Qualitative Research: Increasing Fairness in the CHI Review Process. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–6. <https://doi.org/10.1145/3334480.3381066>
- [95] Acar Tamersoy, Munmun De Choudhury, and Duen Horng Chau. 2015. Characterizing Smoking and Drinking Abstinence from Social Media. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media (Guzelyurt, Northern Cyprus) (HT '15)*. ACM, New York, NY, USA, 139–148. <https://doi.org/10.1145/2700171.2791247>
- [96] Kenton T. Unruh and Wanda Pratt. 2008. The Invisible Work of Being a Patient and Implications for Health Care: "[the doctor is] my business partner in the most important business in my life, staying alive.". *Ethnographic Praxis in Industry Conference Proceedings* 2008, 1 (2008), 40–50. <https://doi.org/10.1111/j.1559-8918.2008.tb00093.x> eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1559-8918.2008.tb00093.x>
- [97] Sandra Wachter and Brent Mittelstadt. 2019. A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI. *Columbia Business Law Review* 2019, 2 (2019), 494–620–494–620. <https://doi.org/10.7916/cblr.v2019i2.3424> Number: 2.
- [98] Greg Wadley, Wally Smith, Bernd Ploederer, Jon Pearce, Sarah Webber, Mark Whooley, and Ron Borland. 2014. What People Talk About when They Talk About Quitting. In *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: The Future of Design (Sydney, New South Wales, Australia) (OzCHI '14)*. ACM, New York, NY, USA, 388–391. <https://doi.org/10.1145/2686612.2686671>
- [99] William L. White. 2007. Addiction recovery: Its definition and conceptual boundaries. *Journal of Substance Abuse Treatment* 33, 3 (Oct. 2007), 229–241. <https://doi.org/10.1016/j.jsat.2007.04.015>
- [100] Andrew J. Winzelberg, Catherine Classen, Georg W. Alpers, Heidi Roberts, Cheryl Koopman, Robert E. Adams, Heidemarie Ernst, Parvati Dev, and C. Barr Taylor. 2003. Evaluation of an internet support group for women with primary breast cancer. *Cancer* 97, 5 (2003), 1164–1173. <https://doi.org/10.1002/cncr.11174>
- [101] Xiaohui Yan, Jiafeng Guo, Yanyan Lan, and Xueqi Cheng. 2013. A biterm topic model for short texts. In *Proceedings of the 22nd international conference on World Wide Web (2013-05-13) (WWW '13)*. Association for Computing Machinery, New York, NY, USA, 1445–1456. <https://doi.org/10.1145/2488388.2488514>
- [102] Diyi Yang, Robert E. Kraut, Tenbroeck Smith, Elijah Mayfield, and Dan Jurafsky. 2019. Seekers, Providers, Welcomers, and Storytellers: Modeling Social Roles in Online Health Communities. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (Glasgow, Scotland UK) (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. <https://doi.org/10.1145/3290605.3300574>
- [103] Diyi Yang, Zheng Yao, Joseph Seering, and Robert Kraut. 2019. The Channel Matters: Self-Disclosure, Reciprocity and Social Support in Online Cancer Support Groups. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (Glasgow, Scotland UK) (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–15. <https://doi.org/10.1145/3290605.3300261>
- [104] Svetlana Yarosh. 2013. Shifting Dynamics or Breaking Sacred Traditions?: The Role of Technology in Twelve-step Fellowships. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 3413–3422. <https://doi.org/10.1145/2470654.2466468> event-place: Paris, France.
- [105] Chuang-Wen You, Ya-Fang Lin, Cheng-Yuan Li, Yu-Lun Tsai, Ming-Chyi Huang, Chao-Hui Lee, Hao-Chuan Wang, and Hao-Hua Chu. 2016. KeDiary: Using Mobile Phones to Assist Patients in Recovering from Drug Addiction. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (San Jose, California, USA) (CHI '16)*. ACM, New York, NY, USA, 5704–5709. <https://doi.org/10.1145/2858036.2858185>
- [106] Chuang-wen You, Kuo-Cheng Wang, Ming-Chyi Huang, Yen-Chang Chen, Cheng-Lin Lin, Po-Shiun Ho, Hao-Chuan Wang, Polly Huang, and Hao-Hua

Chu. 2015. SoberDiary: A Phone-based Support System for Assisting Recovery from Alcohol Dependence. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) (CHI '15). ACM, New York, NY, USA, 3839–3848. <https://doi.org/10.1145/2702123.2702289>

[107] Francisco Alejandro Zampieri. 2018. Analysis of Deviant Opioid Addiction Treatment Communities on Reddit. <https://smartech.gatech.edu/handle/1853/60370> Accepted: 2018-08-20T19:10:58Z Publisher: Georgia Institute of Technology.