A Structured and Linguistic Approach to Understanding Recovery and Relapse in AA

SHAWN BAILEY, YUE ZHANG, and ARTI RAMESH, SUNY Binghamton
JENNIFER GOLBECK, University of Maryland, College Park
LISE GETOOR, University of California, Santa Cruz

Alcoholism, also known as Alcohol Use Disorder (AUD), is a serious problem affecting millions of people worldwide. Recovery from AUD is known to be challenging and often leads to relapse at various points after enrolling in a rehabilitation program such as Alcoholics Anonymous (AA). In this work, we present a structured and linguistic approach using hinge-loss Markov random fields (HL-MRFs) to understand recovery and relapse from AUD using social media data. We evaluate our models on AA-attending users extracted from: i) the Twitter social network and predict recovery at two different points—90 days and 1 year after the user joins AA, respectively, and ii) the Reddit AA recovery forums and predict whether the participating user is currently sober. The two datasets present two facets of the same underlying problem of understanding recovery and relapse in AUD users. We flesh out different characteristics in both these datasets: i) in the Twitter dataset, we focus on the social aspect of the users and the relationship with recovery and relapse, and ii) in the Reddit dataset, we focus on modeling the linguistic topics and dependency structure to understand users' recovery journey. We design a unified modeling framework using HL-MRFs that takes the different characteristics of both these platforms into account. Our experiments reveal that our structured and linguistic approach is helpful in predicting recovery in users in both these datasets. We perform extensive quantitative analysis of different groups of features and dependencies among them in both datasets. The interpretable and intuitive nature of our models and analysis is helpful in making meaningful predictions and can potentially be helpful in identifying and preventing relapse early.

CCS Concepts: • **Information systems** → **Web mining**; **Social networks**; • **Computing methodologies** → **Machine learning**.

Additional Key Words and Phrases: Social Media Analysis, Probabilistic Graphical Models, Modeling recovery from alcoholism, Alcoholics Anonymous

ACM Reference Format:
Shawn Bailey, Yue Zhang, Arti Ramesh, Jennifer Golbeck, and Lise Getoor. 2020. A Structured and Linguistic Approach to Understanding Recovery and Relapse in AA. *ACM Trans. Web* 1, 1, Article 1 (January 2020), 35 pages. https://doi.org/10.1145/3423208

1 INTRODUCTION

Alcoholism, or alcohol use disorder (AUD), is a serious problem affecting millions of people worldwide. Statistics from the World Health Organization estimate that there are 208 million people affected by AUD and it is the cause for 33 million deaths worldwide [44]. AUD can coexist with,

Authors' addresses: Shawn Bailey, sbailey6@binghamton.edu; Yue Zhang, yzhan202@binghamton.edu; Arti Ramesh, artir@binghamton.edu, SUNY Binghamton; Jennifer Golbeck, golbeck@cs.umd.edu, University of Maryland, College Park; Lise Getoor, getoor@ucsc.edu, University of California, Santa Cruz.

1559-1131/2020/1-ART1 $15.00
https://doi.org/10.1145/3423208

contribute to, or result from several different psychological ailments including depression, making it extremely important to study it. Further, previous work emphasize that even while enrolled in a recovery program, people are susceptible to relapse, especially in the early stages of recovery [30].

According to substance abuse recovery models, social support and inclusion play a pivotal role in a person's recovery from addiction [3, 23, 38, 49]. This is also the founding premise for social support groups such as Alcoholics Anonymous (AA). AA brings people suffering from AUD together, providing them with a platform to share their recovery experiences and has proven to be one of the highly successful approaches for recovery from AUD [17, 26]. On a similar note, social learning theory emphasizes that a significant amount of human learning is by imitation and emulation, causing social influences to play an important role in the development of addiction and more generally, in the consumption of alcohol [23]. Particularly, in the extremely sensitive recovery period, such influences have been shown to precipitate relapses in recovering alcohol addicts [30].

With the rise in online interactions, online support groups and discussion forums that function similar to an in-person meeting albeit online have gained popularity as an alternative to AA [1]. Online interaction data from these forums contain verbose posts describing recovery and relapse experiences. We observe that users use words that signify a myriad of emotional states toward alcohol having different nuances, such as some show a positive approach, motivation, and determination, while others may indicate a constant struggle. We observe that during recovery, often users may develop other psychological conditions: other obsessions, feelings of guilt, shame, rejection, that impedes their recovery and drives them toward relapse [13].

In this work, we present a unified framework using HL-MRFs and show how to model the different structural and linguistic aspects in the recovery process. In our first approach, we present a structured approach to understand recovery from AUD, where we focus on the relational interactions between AA attendees and their friends to understand the effect of these relationships on their recovery. One way to study this is using online interaction data from social media platforms. These social connections induce structure in the network of interactions among users (which we refer to as structural/relational features), helping us study the impact of social factors in users' recovery/relapse. To study the effect of social connections and their impact on recovery, we use data collected from the Twitter social network. Next, to understand recovery from the AA-attending user's perspective and the support they receive at different points in their recovery, we turn to online support forums from Reddit. Here, we adopt a socio-linguistic approach to the problem, where we develop fine-grained linguistic models for verbose and descriptive data from support forums and encode dependencies from the linguistic structure and user network to understand their journey toward recovery and relapse. While AA members on social media may not be representative of the entire population of AA, nonetheless, we believe our study helps in understanding this important and growing subpopulation that uses online platforms for recovery support and help, and can potentially provide insights on how to design more targeted and in-depth studies.

Our contributions in this paper are:

- We first show how to extract fine-grained linguistic, psychological, and structural features from users' online interactions. Our feature set includes linguistic features such as fine-grained topics, sentiment, psycho-linguistic features from linguistic inquiry word count (LIWC) [47], and structural features from users' interactions with friends and linguistic dependency structures.
- We encode dependencies among these features in a recently developed graphical modeling framework, hinge-loss Markov random fields (HL-MRFs) [4], and jointly reason about the recovery/relapse of users. We develop three different HL-MRF models capturing features and structural dependencies at different levels: user-friend relationship in the underlying

social network, interactions among users in the discussion forums, post-level and user-level linguistic features, and sentence-level linguistic expressions and structural dependencies among them in relation to recovery and relapse. Our models thus possess the capability to model a variety of online interaction scenarios related to recovery. Our models effectively capture the relationship among the various features and serve as a template for modeling recovery from AUD.

- We evaluate the effectiveness of our models on two datasets we collected and labeled from Twitter and Reddit, respectively. We perform extensive experimentation on both the datasets: i) evaluate effectiveness of predicting recovery at different time periods on the Twitter dataset, and ii) construct different meaningful variations of the models for both Twitter and Reddit datasets, and show that our structured and linguistic approach is capable of reliably predicting recovery in AA users. We also present an in-depth quantitative and qualitative analysis of the different feature groups and their dependencies and their corresponding effectiveness in predicting recovery. We observe that we can predict recovery and relapse accurately by just considering features from the structural interactions with friends in the Twitter dataset, signaling the importance of modeling social factors in recovery. Similarly, in the Reddit dataset, we find that fine-grained topic features that capture different psychological and emotional struggles faced by users during recovery are helpful in predicting recovery and relapse accurately. We also include an analysis on the number of groundings of different rules that helps us understand the target users for the different kinds of rules in our Reddit models. This endeavor helps in identifying users who follow the general trend and those who follow a specific pattern, which in turn is helpful in potentially constructing personalized interventions.

Specifically, we expand on Zhang et al. [66] by including more elaborate socio-linguistic models, additional experimental results, and more in-depth analysis of recovery and relapse on a different dataset that is more targeted toward recovery and relapse, a support forum from Reddit (*AlcoholicsAnonymous*) that functions similar to an online AA meeting. It had 5, 458 users when we scraped it and has helped 709 users stay sober out of the 1074 users we labeled. By expanding our analysis, our aim is to develop a panoramic understanding of the recovery problem from multiple angles by studying two online AA communities. The two datasets present two facets of the same underlying problem of understanding recovery and relapse in AUD users. While the Twitter dataset throws light on the social connections of the AA attendees and how that possibly affects their short-term and longer term recovery, the verbose and detailed posts on Reddit dataset by AA users on their journey helps us expand on the linguistic cues in understanding recovery. We first develop a more detailed vocabulary of words and phrases organized into fine-grained topics that are indicative of the different phases in the recovery process. There is no existing detailed vocabulary at this fine-grained level and we believe this contribution helps in understanding minute aspects of the recovery journey. We then show how to construct HL-MRF models for this data by paying specific attention to fine-grained linguistic features and their dependencies. We develop two socio-linguistic models that capture fine-grained topics and the dependency parse structure to understand recovery. We then present experimental results that elicit the different nuances in modeling required to ably study this problem by performing an in-depth analysis of features and groundings of the rules in the model.

Our models possess superior interpretability at the rule level, allowing for encoding and understanding the different factors contributing to recovery and relapse. While our models are able to achieve superior prediction performance, the interpretable and intuitive nature of the logical rules are what make our models useful. Our feature analysis helps in identifying the effect of the different features on users' recovery/relapse and helps in making meaningful predictions in scenarios where only a subset of features are available. Our models and analysis can potentially pave the way for conducting in-depth studies to understand the nuances in the recovery process and identifying early signs of relapse and preventing them before they occur.

2 RELATED WORK

Previous work in understanding recovery and relapse using online interaction data can be grouped into two broad categories: i) a social media analysis approach, which focuses on analyzing social media data and presenting aggregate statistics, and ii) a predictive modeling approach, designing models for predicting health-related attributes. We also discuss work on social theories on AUD that help us in designing the models and drawing important conclusions.

The recent explosion of social media has led to a growing interest in using social network interaction data to study issues of public health [16, 29, 33, 34, 42, 64, 67]. We discuss some of the ones related to substance abuse, mental health, and use of health support forums for coping with various illnesses specifically as they are more related to our work. Chisolm et al. [34] use a college social network, Yik Yak, to perform an exploratory analysis of the presence of substance abuse and health issues in college campuses. Balani et al. and Choudhury et al. [5, 12, 53, 58] focus on identifying mental health issues particularly focusing on self-disclosure on Reddit. Their analysis on the presence and prevalence of self-disclosure on mental health issues in social media data paves the way for our study. We also similarly employ self-disclosure for labeling data in our recovery/relapse prediction problem. Mejova et al. [41] and Paul et al. [45, 46] show how Twitter data can used a digital socioscope to study individual-level and group-level human behavior and interaction at scale. In our study, we draw upon and expand on their approach on using Twitter for studying human behavior and the importance of modeling structural interactions/influence and their impact on recovery.

There is also a growing body of literature on analyzing data from health support forums [20, 61]. Since forum data is textual in nature, these works generally focus on analyzing the linguistic cues. Choudhury et al. [7, 11] and White et al. [60] study the presence of supportive language, Xu et al. [63] and Huang et al. [25] study linguistic cues in mental health support forums, and Coulson et al. [9] and Manikonda et al. [39] analyze data from a support group for irritable bowel syndrome and weight-loss, respectively. Rey-Villamizar et al. [51] study the use of anxious words, Kramer et al. [35] and Kalman et al. [31] and Bluic et al. [6] study the linguistic and social markers, Ding et al. [15] focus on interpreting social media-based substance use prediction models, and Wang et al. [59] focus on the emotional aspects of online health forums. Ding et al. [14] develop an unsupervised learning approach to understand word usage among substance abuse users. Tamersoy et al. [55], Harikumar et al. [21], and Maclean et al. [37] analyze recovery forums on alcohol and other substance abuse. All these works collectively emphasize the importance of linguistic cues. It is important to note that these works mainly focus on word usage and their impact on prediction. In our work, we expand on this line of research by modeling fine-grained linguistic cues along with the social interaction structure using relational models. Our fine-grained linguistic cues include fine-grained topics that are designed specifically for recovery and relapse and discovery and representation of phrases and the linguistic dependency structure, which are key to identifying the right linguistic signals. Our interpretable relational approach to understanding recovery is uniquely positioned in its ability to model all the different signals in this data.

Grant et al. [19] study the importance of using online mediating measures for maintaining sobriety. Their case study indicates positive output from online mediation though they also mention the need to have more context-specific mediation techniques. Curtis et al. [10] perform case studies through questionnaires on social media and conclude that disseminating recovery support within a social media platform may be the ideal just-in-time intervention needed to decrease the rates of recurrent drug use. Their results also suggest that cross-platform solutions capable of transcending generational preferences are necessary and one-size-fits-all digital interventions should be avoided. This further ascertains the importance of adaptable, versatile, and interpretable solutions like ours for this problem.

There is also a growing body of work on predictive analysis of social media data. Perhaps the closest work to ours is Hossain et al.'s work on developing a model to predict geo-location based on whether a particular tweet signifies drinking and whether the user was drinking while writing that tweet [24]. Other approaches focus on aggregate social behavior and report analysis of these behaviors over time in specific groups of users [27, 42]. Tomkins et al. [56] present a probabilistic model for cyberbullying detection and Zhang et al. [65] present an analysis of cyberbullying on an anonymous messaging application. Other predictive analysis work surrounding social media include Chowdhury et al.'s work on studying pharmacovigilance in social media posts [8, 32], Walker et al.'s work on identifying eating disorders [57], and McIver et al.'s work on identifying sleep disorders using social network data [2, 40].

While previous work on social theories for alcohol use disorder emphasize that social influence is a contributing factor in the development of substance abuse problems including AUD [43, 54, 62], our relational approach provides a convenient way to encode these structural attributes and evaluate their effectiveness in the prediction problem. Our work differs from these existing approaches in that we explicitly encode the structural information in a probabilistic framework, allowing us to effectively reason about their contribution in recovery/relapse. We model the dependencies between linguistic and structural information, thus helping in understanding both their individual and combined contribution to predicting recovery. Further, for complex verbose descriptions of users' recovery, our models capture both fine-grained signals corresponding to recovery and structural dependencies among them (both within the post, among posts, and among users), thus bringing the ability to thoroughly understand users' journey toward recovery and the factors hindering it. Our models are encoded using first-order logical rules, thus possessing superior interpretability at the rule-level, allowing us to capture and study these intricate dependencies and also simultaneously being accessible to domain experts to specify these dependencies as opposed to other black-box machine learning prediction models that are normally used for prediction scenarios. Also, our work focuses on collecting and labeling data, extracting features, and designing models that are especially focused on users attending alcoholics anonymous (AA), while most existing work focus on the discussion of an issue (such as substance abuse, mental health) and not particularly on the recovery. Thus, our study sheds light on the recovery process and can be useful for designing targeted interventions.

3 DATA

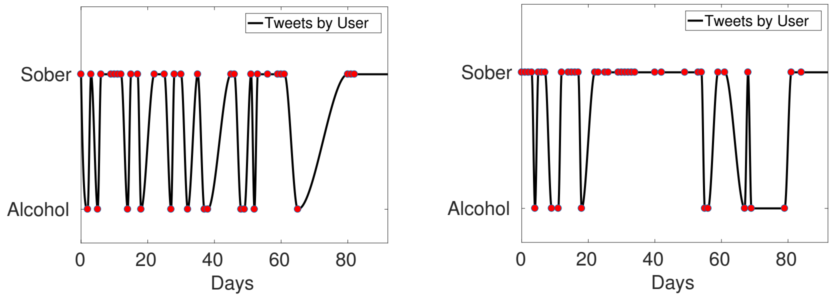
In this section, we describe how we collect data for AA-attending users from two online platforms, Twitter and Reddit, and label them as recovered/relapsed depending on whether they have been drinking after they joined a recovery program. We also present some interesting observations from the data that we use in the later sections to construct features and prediction models.

3.1 Twitter Dataset

In order to collect information about AA members on Twitter, we extract tweets using the Twitter Search API using the search term "first AA meeting", as we are only interested in AA-attending users. As our goal is to study their recovery, we look for the tweet in which they mention attending their first AA meeting. The Twitter Search API only allows access to recent tweets at each instance it is invoked. Since there aren't many users mentioning attending their first AA meeting in each search, we repeat this search in regular intervals over a four year window from January 1, 2013 to February 1, 2017 for collecting the data in this paper. From this set, we eliminate jokes and people attending to support an alcoholic friend or family member, by manually examining the tweets mentioning attending the first AA meeting. We then identify the users corresponding to the "first AA meeting" tweets. This collection and filtering process results in 691 users. We then use the Twitter API to collect the most recent 3, 200 tweets of each of the filtered users attending their first AA meeting, the maximum allowable under Twitter's Terms of Service. We separate these tweets into two groups: tweets tweeted *before* and *after* the first AA meeting tweet. After collecting data on AA-attending users, we then proceed to label these users as recovered/relapsed by looking at the tweets after joining AA.

Example Tweets
People say *sobriety* is hard work but how hard was *alcoholism*? That was the worst full time job. It didn't pay well and the benefits **sucked**.
Why do I continue to *drink* when I know how **sick** it makes me.
Sobriety **sucks**, time for a *drink*!

Table 1. Examples of tweets where users mention alcohol/sober words, but do not directly imply consuming/refraining from alcohol, respectively



(a) Recovered user with many alcohol tweets

(b) Relapsed user with many sober tweets

Fig. 1. Graphs showing recovered and relapsed users with alcohol-related and sober tweets, respectively

To design a labeling scheme for recovery/relapse, we first closely examine users' tweets after they join AA. Table 1 illustrates example tweets where presence of words related to alcohol or sobriety provide a noisy signal for labeling. The alcohol/sober words in the tweets are *italicized* and the adjectives referring to these words are typed in **bold**. For example, the first tweet mentions both alcohol and sober words but indicates user's intention to become sober. Similarly, the second and third tweets in the table contain mentions of alcohol/sober words but do not directly imply drinking alcohol or staying sober. Furthermore, we observe that both recovering and relapsing users tweet about alcohol and sobriety significantly. Figure 1(a) shows an example of a recovered user whose tweets contain significant number of alcohol-related words. The red dots refer to alcohol and sober word mentions and we can clearly see that the user oscillates between using both alcohol and sober

Category	Words
Alcohol	drunk, beer, bar, wine, alcohol, wasted, hungover, hangover, turnt, vodka, liquor, whiskey, tequila, alcoholic, champagne
Sober	recovery, sober, sobriety, #recovery, #sobriety

Table 2. Alcohol/Sober word vocabulary

Label	Tweet
Relapse	I am drunk as ever.
Relapse	Taking shots after I left work tonight was not a good idea.
Recover	I've officially been sober for 4 months
Recover	Soon, I'll be 5 months sober

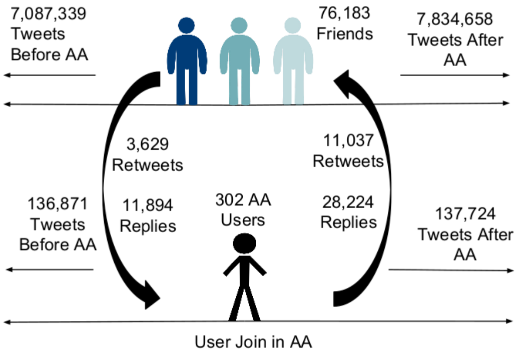
Table 3. Examples of tweets where users mention consuming or refraining from alcohol after joining AA

words, with more alcohol-related words. Similarly, Figure 1(b) gives an example of a relapsed user tweeting significantly about sobriety.

For this reason, we adopt a careful approach to labeling users with recovery/relapse labels. The users are carefully labeled by three annotators manually by looking through tweets of the user and determining if they have relapsed or are staying sober at the 90-day milestone and 1-year milestone. We rely on self-disclosure of drinking or remaining sober after joining AA, indicated by the user via tweeting as is typically done for social-media data [5, 11]. Due to the sensitive nature of this dataset and the privacy concerns that surround sharing it on online crowdsourcing platforms, researchers closely associated with the project performed the annotation. The annotation task was completed first before starting the modeling task so that the annotation does not interfere with the development of the models. To establish if a user has recovered or not at the 90-day mark, we look carefully at the tweets by the user after joining the AA meeting. We examine whether the user's tweet mentions consuming alcohol before the 90-day mark to establish that he/she did not recover. Similarly, to establish whether a user remained sober, we look at tweets where the user mentions being sober after joining AA. Table 3 gives examples of tweets from users labeled as recovered/relapsed depending on the content of their tweets after joining AA. We exclude users for whom there was no explicit mention of consuming or refraining from consuming alcohol (i.e., being sober) from our dataset. This eliminated 389 of the 691 initial users, leaving 302 AA-attending users (226 relapsed users and 76 recovered users) in our sample, with a 25.2% recovery rate. We refer to these users as AA users.

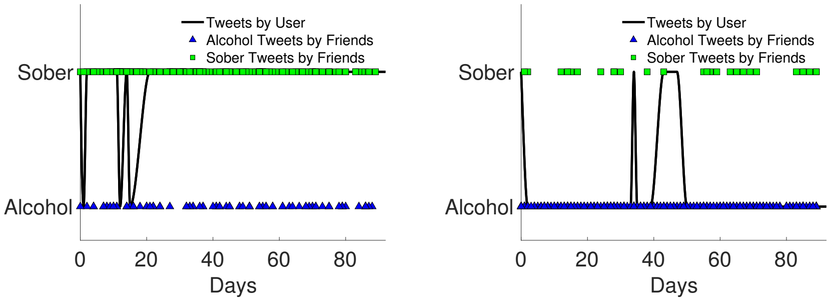
We construct an alcohol/sober vocabulary as we label the users, adding the alcohol and sober words from tweets that are significant for labeling to a vocabulary. Table 2 gives the alcohol/sober vocabulary. To also gather words that co-occur with these alcohol and sober words, we show how this vocabulary can be expanded using seeded Latent Dirichlet Allocation (seeded LDA) [28] in Section 4.2.2 by using the words in the vocabulary as seed words. This helps in extending our vocabulary to other data.

For each of these AA users, we collect friend information (we refer to bi-directional followers on Twitter as friends). We again collect the most recent 3, 200 tweets for each of the friends and separate them into *before* and *after* the corresponding AA users join AA. Figure 2(a) gives details about our dataset. We have 302 AA users in our dataset. We collect the most recent 3, 200 tweets for each of these users, giving us a total of 274, 595 AA user tweets. Of these tweets, 137, 724 tweets occur after the users join AA and 136, 871 tweets occur before the users join AA. For 302 AA users, we have a total of 76, 183 friends in our dataset. For each friend, we also collect the most recent 3, 200 tweets, giving us a total of 14, 921, 997 tweets. We again split these tweets into before/after the respective AA users join AA, giving us 7, 087, 339 and 7, 834, 658 tweets, respectively.



(a) Figure showing data statistics: number of AA users, friends, and tweets

Fig. 2. Figure showing data statistics: number of AA users, friends, and tweets in the Twitter dataset



(a) Recovered user with sober friends. Each green dot and blue dot represents the mention of a sober word and alcohol word by AA-user's friend, respectively, during the user's recovery period. The user's mention of these words is captured using the black line.

(b) Relapsed user with alcoholic friends. Each green dot and blue dot represents the mention of a sober word and alcohol word by AA-user's friend, respectively, during the user's recovery period. The user's mention of these words is captured using the black line.

Fig. 3. Graphs showing alcoholic and sober tweets by friends for a recovered and a relapsed user, respectively

The contribution of social influence from the friend/peer network in one's drinking habits is especially important during the very fragile period of recovery [43, 54, 62]. We observe that recovered users generally tend to have friends who tweet significantly about sobriety rather than about alcohol. Figure 3(a) gives an example of a recovered user in our dataset. Plotting the mention of alcohol and sober words in their friends' tweets during the AA-attending user's recovery period, we see that this user has more friends who mention sober words (each green dot denotes a mention of a sober word) as opposed to alcohol words (blue dots). The user oscillates between alcohol and sober word usage in the initial period and then transitions to only mentioning sober words in the tweets (solid black line). Similarly, we find that relapsed users tend to have more friends who tweet about alcohol. Figure 3(b) gives an example of a relapsed user whose friends tweet significantly about alcohol rather than about sobriety. Here, we can see that the AA-user oscillates again between mentions of sober/alcohol words and finally staying with alcohol mentions throughout the recovery period. While these are just mentions of alcohol or sober words without contextual information, these observations guide our feature engineering and structured prediction models that take into account these structural interactions and their corresponding effect on recovery/relapse.

3.2 Reddit Dataset

While Twitter is a suitable dataset for studying structural relationships in the data, the character limit in the tweets limits the amount of linguistic analysis possible in the data. Hence, to further understand the recovery process from a linguistic perspective, we expand our analysis to include another data source, an online discussion forum for recovery from AUD in Reddit. We collect and label data instances from the *AlcoholicsAnonymous* subreddit, that functions similar to an AA meeting albeit online, facilitating users to share their recovery experiences. We present some interesting observations from the data that we use in the later sections to construct meaningful features and prediction models.

3.2.1 Data Collection. In order to avoid Reddit's limitation of the last 1000 submissions being viewable on the site, we turn to a third party website that periodically scrapes and stores all posts on Reddit, called pushshift.io. We scrape the entire subreddit, and group the data by each user. Submissions are original posts made by a redditor, and comments are replies to either submissions or other comments found in the discussion of a particular submission. We collectively refer to submissions and comments as posts. We scrape a total of 5, 185 submissions, 48, 668 comments, from a total of 5, 458 redditors. From this initial set of data, we keep all submissions and comments from users whose posts spanned at least 290 days. This process gives a total of 591 users. When we noticed that most of these users were labeled as currently sober, we additionally scrutinized and labeled all users that mentioned the word relapse in their posts. Note that not every user who mentioned some past relapse were necessarily currently drinking, according to the contexts of their more recent posts. The additional users added some cases of relapsed users, but the class imbalance remained. The presence of class imbalance is another reason for choosing HL-MRF models for this prediction problem as they are known for performing well on datasets despite there being a class imbalance [48]. This gives us an additional 483 redditors. In all, our Reddit dataset comprises of a total of 1, 074 redditors.

3.2.2 Interesting Observations and Examples. As we explore the Reddit dataset, we observe that sometimes the difference between those that recovered and those that relapsed were subtle. Like the Twitter dataset, users frequently mention both sobriety and relapse. Here, we present some interesting types of users and some of their posts to provide insight on the recovery process.

The three examples in Table 4 represent different types of users that we find in the Reddit dataset in a general sense. For instance, consider the first post in Table 4 from a user who has recovered. We call such a user an *advise giver*, someone that advises other users about recovery. Notice that they use words related to relapsing quite a bit. The second post is from a user that we consider recovered, but is difficult to classify. We refer to such a user a *struggler*, someone who is not drinking but still thinks about drinking while being sober. Notice that context is key here, in particular the overall tense of the sentence: past, present, or future. It is not enough to solely search for negative words. One strategy we use here is to search for common phrases, rather than single words, where the

Types of Reddit Users	Example Posts
Advise-giver	This is exactly the right way to look at a **relapse** not only if it is you but also a friend. Some people feel alienated after **relapsing** and have trouble making it back due to fear of the perceived stigma against **relapsers**. Although I feel most groups are good about this there is no reason to treat a person like they are "dirty" if you know they've **relapsed** and came back in. On any given day we are one prayer or one bad decision away from going back out. We learn from our mistakes and move forward.
Struggler	The **cravings** are slowly creeping back in. I'm on day 45 of being sober and I want to **drink** so badly. My life has been feeling like it's on autopilot for so long now. I can barely remember my day and I feel like I'm either always waking up to go to work or trying to sleep. I've been trying to distract myself with games but it's not helping that much.
Relapser, but lies about recovery	I'm an alcoholic...For 3 months I've been **going to meetings**. Started once or twice a week then about 7 weeks ago I started going a minimum of 5 days a week. I've **had a sponsor** for about 6 weeks. I say this not to be proud of what I've done but because I'm not proud...I've been continuing to live a lie. I still drink most nights and tell **my group** and **my sponsor** I've been **sober**. I **don't drink** nearly the same quantity I used to but that's not the point. I'm ready truly ready to stop this madness. Thank you for letting me share.

Table 4. Examples of Reddit posts where users mention alcohol/sober words, but do not directly imply consuming/refraining from alcohol, respectively

phrases have an implied tense. For instance, "relapsed today" and "recently relapsed" strongly imply that their relapse happened recently, while "relapse" itself is talked about in general, for instance, by users giving advice to a struggling user, or by a user simply sharing their past. As an example of the future tense, sometimes a struggler will plan a relapse, saying they "may relapse", "might relapse", "will relapse", "relapse tomorrow", and so forth. The third post is an example of a relapser's post that uses a lot of sober words. This user reaches a milestone, gets a one month chip, attends meetings, and has a sponsor that they talk with. Additionally, they mention the word sober in a present tense, but admit that they are lying about it. Although there are also negative words in this post, there are many positive signs linguistically as well, which are indicated in bold (*one month chip*, *going to meetings*, *sober*). These observations help us in constructing fine-grained linguistic topic features in Section 4.2.2.

3.2.3 Data Pre-processing. Often, we could not guarantee that each redditor was consistently active on the subreddit, so the notion of splitting the data and analyzing after 90 days that we adopt for Twitter data is not relevant here. Typically, a redditor shares their personal story, saying they relapsed a certain amount of time ago, and mentions their time sober as well. During annotation, 59 redditors were skipped as they are either bots on the subreddit or users with few posts to analyze and typically only shared outside links to personal blogs. 35 users are relatives of an alcoholic seeking advice and they were skipped as well.

3.2.4 Data Labeling. The dataset was labeled by two separate individuals, independently from one another. The guidelines involved looking for self-proclaimed acts of drinking while in recovery or continuing to be sober after reading all posts by each individual [5, 11]. Hence, to label the Reddit users with relapsed/recovered, we carefully read the comments and submissions of each redditor and answer two questions: "Did the user ever mention a personal relapse?" and "Is the user currently sober, based on their most recent posts?", with a *yes* or *no* answer. If there was ever a user where we felt it was unclear whether the user was currently sober, we gave them a fuzzy label. In other words, a label indicating that we were unsure. Once both individuals labeled the data set independently, we compared the labels from both individuals, keeping the labels that were agreed upon. For labels with a disagreement, the two individuals reread and discussed what the labeling should be finalized in more detail and if an agreement could not be reached and the label remained a fuzzy label, that user was excluded from the study. Since we only retained the users for whom there was no disagreement in the labels, the inter-annotator agreement is 1.0. Our goal with annotation is to eliminate as much noise as possible so that we can come up with insightful predictions from the models when trained on these labeled instances. This can be seen in the number of users for whom the data was collected and the final number of users on whom the prediction was performed. Figure 4 shows the number of users in each label for being currently sober, additionally split by the mention of a relapse at some point in their journey.

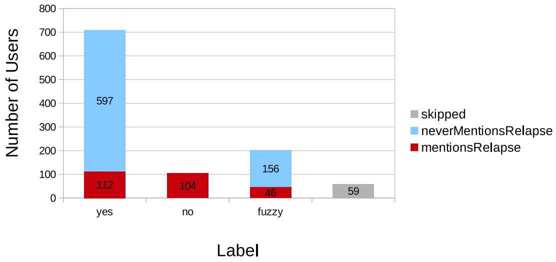


Fig. 4. Figure showing labeling for users being currently sober (indicated by yes) and not currently sober (indicated by no). Most labeled users never mention relapsing on this subreddit (597 users). Additionally, out of the 262 users that have relapsed in the past, 112 of them end up recovering.

4 AUD RECOVERY PREDICTION MODELS

In this section, we build our HL-MRF recovery prediction models for predicting relapse/recovery. We first extract features from users and then identify and encode the dependencies among them in our models of AUD recovery.

4.1 Hinge-loss Markov Random Fields

To encode the different signals from users, interactions with friends, and effectively reason about the dependencies among these signals and their effect on recovery, we develop a powerful approach using hinge-loss Markov random fields (HL-MRFs). HL-MRFs can be specified using *Probabilistic Soft Logic (PSL)* [4], a first order logical templating language. In PSL, random variables are represented as logical atoms and weighted rules define dependencies between them. An example of a PSL rule is

$$\lambda : S(a) \wedge Q(a,b) \longrightarrow R(b),$$

where S, Q, and R are predicates, a and b are variables, and λ is the weight associated with the rule. The weight of the rule indicates its importance in the HL-MRF model, which is defined as

$$\begin{aligned} P(\mathbf{Y}|\mathbf{X}) &\propto \exp\Big(-\sum_{r=1}^{M} \lambda_r \phi_r(\mathbf{Y},\mathbf{X})\Big) \\ \phi_r(\mathbf{Y},\mathbf{X}) &= (\max\{l_r(\mathbf{Y},\mathbf{X}),0\})^{\rho_r} \ , \end{aligned} \tag{1}$$

where P(Y|X) is the probability density function of a subset of logical atoms $\mathbf{Y}$ given observed logical atoms $\mathbf{X}$, $\phi_r(\mathbf{Y},\mathbf{X})$ is a *hinge-loss potential* corresponding to an instantiation of a rule r, and is specified by a linear function l_r and optional exponent $\rho_r \in \{1,2\}$. The weights are learned using maximum likelihood weight learning using ground truth labels at training time. For example, in our Twitter recovery model, U_1 is an AA attending user and U_2 is his/her friend in the Twitter network. Suppose U_2 tweets about alcohol, denoted by alcoholTweet(U_2, T) and U_1 retweets it, given by retweets(U_1, U_2, T). A PSL rule to encode the combined effect of these interactions on recovery is given by

$$\lambda : \mathit{Friends}(U_1, U_2) \wedge \mathit{alcoholTweet}(U_2, T) \wedge \mathit{retweets}(U_1, U_2, T) \rightarrow \neg \mathit{recovers}(U_1).$$

The rule captures that if U_1 retweets U_2's tweet on alcohol, then that is an indication that U_1 is drawn towards consuming alcohol and could possibly lead to U_1 not recovering from AUD. Note that this rule combines both linguistic features from tweets and structural features from interaction with friends to predict recovery.

We can also generate rules that collectively reason about two AA attending users U_1 and U_2, e.g.,

$$\lambda : \mathit{similar}(U_1, U_2) \wedge \mathit{recovers}(U_1) \rightarrow \mathit{recovers}(U_2).$$

This rule captures that if two AA users are similar, then if one recovers, there is a high possibility of the other user recovering as well. The HL-MRF model uses these rules to encode domain knowledge about dependencies among the predicates. The interpretable nature of first-order-logic rules is helpful in capturing meaningful dependencies among the different features and recovery. The continuous value representation further helps in understanding the confidence of predictions.

4.2 Feature Engineering

We first extract a suite of features from users' posts and interactions with other users. We group the features into three broad categories: i) linguistic, ii) psycho-linguistic, and iii) structural features. We describe them in detail below. Most features we extract are applicable for both the Twitter and Reddit datasets with the exception of some that are only relevant in either dataset. We identify such exceptions in the feature's description.

4.2.1 Linguistic Features. Users' posts are the strongest signal in predicting their recovery/relapse. We extract different linguistic features from AA users' and their friends' posts. Below, we explain the various linguistic features that we incorporate in our recovery prediction models.

Term Frequency. For each Twitter user, we concatenate all the tweets in the 90 days after the user joins AA. We apply standard NLP pre-processing techniques such as removing punctuation and stop words, and calculating the term frequency of remaining words for each user. To construct a succinct feature that represents the term frequency of words in the dataset, we train a logistic regression model based on the term frequency of words in the data. We choose logistic regression as it is representative of the classic machine learning models and provides a probabilistic prediction, which can be effectively integrated as a feature in HL-MRFs. The model produces scores between 0 and 1 on a test data instance, which we can incorporate as a feature in the HL-MRF model. These scores are used as a feature that we feed into our HL-MRF models. This feature abstracts away the words used and helps us expand it by capturing relationships that build on this feature. This way, we show that we can encompass simpler models using our HL-MRF models. We extract this feature similarly for our Reddit dataset, but without a 90-day timeframe.

Alcohol/Sober Word Usage. For our Twitter dataset, we use the alcohol/sober word dictionary given in Table 2 to filter tweets that use alcohol/sober words. Since AA users use alcohol and sober words significantly, we capture alcohol/sober word usage both at the tweet level and at the user level. *containsAlcohol*(U_1, T) captures if a specific tweet T by user U_1 contains at least one alcohol word, a binary feature indicated by a zero or one. *usesAlcoholWords*(U_1) is a continuous feature that captures the user's usage of alcohol/sober words: values closer to 1.0 signifying that the user uses a higher number of alcohol words and values closer to 0.0 signifying that the user uses a lower number of alcohol words. We extract this feature similarly for our Reddit dataset. Section 4.2.2 provides more details on the fine-grained topic distribution for the Reddit dataset.

Sentiment Scores. We compute sentiment scores for each Twitter users' retweet or reply from friends to users using SentiWordNet in natural language tool kit (NLTK) in Python [36]. As Goncalves et al. [18] note, sentiment analysis tools often do not work for online social networks. This is further complicated by our specific prediction problem of understanding recovery or relapse and analyzing sentiment in this domain is challenging. Goncalves et al. [18] also note that supervised classification techniques work better than Sentiwordnet for Twitter data, but in our case the prediction labels, recovery and relapse, are not directly related to sentiment. We use SentiWordNet primarily as a baseline for constructing our topic dictionaries in Section 4.2.2 as a comparison with our more specialized topic dictionaries. SentiWordNet gives the number of positive and negative words in the document. We normalize the scores and treat scores closer to 1.0 as positive and scores closer to 0.0 as negative. We conduct a similar analysis and extract sentiment scores for our Reddit dataset.

Subreddit Traffic. In general, we find that the more a user posts on the subreddit, the more likely they are to recover. More activity indicates that the user is dedicated to their sobriety. We count up how many posts each user posts and normalize these scores between 0 and 1. This feature is only relevant for the Reddit dataset.

4.2.2 Topic Features.

Coarse-grained Topic Distribution from Seeded Topic Modeling. Topic models present an easy way to understand document corpora. In our problem, we are interested in particularly isolating the tweets belonging to alcohol or sober topics. Hence, we leverage a seeded variant of topic modeling, seeded LDA [28]. Seeded LDA guides topic discovery to learn specific topics of interest to a user by allowing the user to input a set of seed words that are representative of the underlying topics in the corpus. Seeded LDA uses these seed words to improve topic-word distribution by inducing topics to obtain a high probability mass for the given seed words. Similarly, it also improves the document-topic distribution by biasing documents to select topics related to the seed words. The seed set need not be exhaustive as the model gathers related words based on co-occurrence of other words with the specified seed words in the documents. For more details, we refer the reader to [28].

For our Twitter dataset, we perform standard NLP preprocessing techniques of stop-word removal and stemming using porter stemmer on the tweets. The seed words for the alcohol and sober topics are selected from the alcohol/sober word dictionary in Table 2. We also include k un-seeded topics, to account for other topics in the document corpus. After experimenting with different values of k, we select $k = 8$, giving us a total of 10 topics. We use $\alpha = 0.0001$ and $\beta = 0.0001$ to create sparse document-topic and topic-word distributions so that fewer topics/words with high values emerge. We run the seeded LDA model on two different groups of tweets. First, we aggregate all tweets/posts for each AA user and treat that as one document. We refer to this feature as *userTopic* in our models. In the Twitter dataset, we also consider the replies and retweets exchanged between AA users and their friends and run seeded LDA on this subset of tweets (around $15,000$ tweets). We refer to this feature as *friendTweetTopic.*

Fine-grained Topic Distribution. For our Reddit dataset, the longer length of posts allows us to define fine-grained topics related to user's recovery and stage of addiction. From our observations of the data, there are 20 key topics that we focus on, 12 related to sobriety and recovery, and 8 related to relapse. It turns out that the key words/phrases that are most informative, dealing with relapse, sobriety, and drinking, all need to be handled with care. We aim to ignore these words if they are brought up in a past context. For instance, if a user is talking about a relapse that happened a long time ago, say months or years, then that occurrence has little bearing on whether they are currently sober. For the fine-grained topics, we do not stem the words and include the stop words so that the phrases in the vocabulary can be identified. Topics that are touchy typically have more phrases in the vocabulary, while other topics are smaller if the vocabulary is more clear cut. For example, to see if a user is reading the Big Book, we can search for a small set of words: big book, study, chapter, page, speaker, message. Our goal here is to create a fine-grained topic distribution of words and phrases that capture different linguistic expressions mentioned by different types of users (relapsed, recovered, struggling) during recovery, while remaining generalizable for another similar recovery dataset. The fine-grained topic distribution and analysis also provides insight on the different topics and their correlation with recovery and relapse such as development of another addiction (*drugUser*), user attending meetings regularly (*aaMeetings*), and reaching significant mile stones (*mileStones*) and their relationship with the journey toward recovery. Our choice of topics capture many positive change-related ones (such as *thriving*, *reflectChange*, *adviseGiver*, *encourager*, *mileStones*), which have been shown to be a better predictor of recovery [50].

Here is a brief description of recovery and relapse topics and some linguistic patterns they capture:

(1) **Recovery Topics**
twelveSteps - user is doing steps from the AA program
aaMeetings - user is attending AA meetings
adviceGiver - user is giving advice to others
thriving - user is doing well in AA
encourager - user is encouraging others
bigBook - user reads the big book
detoxing - user has made it through detoxing
mileStones - user has hit a mileStone, such as 1 year sober
reflectMeditateChange - user's mindset/life has changed
religionHigherPower - user talks about religion and higher power
soberWords - user talks about being sober
volunteers - user volunteers to promote/lead AA meetings

(2) **Relapse Topics**
alcoholWords - user talks about alcohol
negativeEmotions - user expresses negative emotions
drinkingWords - user talks about drinking
startingOver - user talks about relapsing, needing to start over
drugUser - user talks about drug use
medicated - user is on antidepressants, other medication
rockBottom - user hit rock bottom, e.g., divorce or bankruptcy
struggling - user is struggling to stay sober

Any given user could discuss a combination of the topics in Table 5. Based on the content of users' posts, we proceed to calculate topic distribution of users' posts across the different fine-grained topics. Figure 5 gives the topic distribution of users among the fine-grained recovery and relapse topics. To calculate this, we first count how many times a user uses a phrase from each fine-grained topic. A user is considered to be discussing a topic more than other topics if the percentage of posts in that topic is above 0.05, the score that would otherwise suggest all 20 topics being discussed equally. Since the topic distribution over 20 topics gives us values < 0.5, we scale the values by 0.45 for the topics so the topic with the highest value gets a value above 0.5 and is recognized easily as

Topic	Vocabulary (words and phrases)
twelveSteps	steps with a sponsor, working the steps, twelve steps, step # (numbers 1 through 12), my sponsor
aaMeetings	one day at a time, go to meetings, attend meetings, home group, rooms, attend aa
adviceGiver	works for you, ask for help, your step work, my advice, my suggestion, find a sponsor, helped me
thriving	great meetings, all the difference, my life has improved
encourager	keep doing what, hit meetings, don't give up, don't forget, this too shall pass, nice work, god bless, in this together, you can do it, so can you, good job, good luck,only the beginning,don't beat yourself up, it works if you work it, keep coming back,gets better
bigBook	big book, study, chapter, page, speaker, message
detoxing	hospital, prescribe, prescription, medicine, detoxed
mileStones	chip, coin, 1 year, one year, months sober, years sober
reflectChange	grow, daily, reflect, meditate, changed, journey, hope, exercise, gym, healthy, productive, saved my life, life changing, lifestyle
higherPower	higher power, religion, concept, prayer, praying
soberWords	done drinking, in sobriety, i quit, stayed sober, years sobriety, months sobriety, i quit drinking, of sobriety, months later, sober now, a huge difference, my recovery, sober living, since quitting, i'm staying sober, continuous sobriety, never drank again, stay sober, i haven't drank, when i got sober, months in, years in, months now, years now, years of sobriety, months of sobriety, been sober, abstain, abstinence, clean, clear, mo sober, my sobriety, recovered, r/stopdrinking, won't drink, stopped drinking, not drinking, don't drink
volunteers	tradition, area, conscience, conference, volunteer, speaking
alcohol	bottle, wine, consume, trigger, tequila, wasted, whiskey, shots, vodka, compulsive, beer, liquor, turnt, craving, one drink
negative	living a lie, lashed out, all alone, feel guilty, not happy, crying, depressed, shame
drinking	black out drunk, thinking about drinking, i've drunk, i've drank, decided to drink, ended up drinking, might drink, may drink, hungover, chug, drink heavily, heavy drinking, drinking heavy, drinking again, binging, binge drink, hangover, drink tomorrow, bad night, drunk again, drank again, drank tonight, start drinking, started drinking, high-functioning, blackout, bender, under the influence
startOver	confess, start over, starting over, always relapse, and relapsed, to relapse, recent relapse, of relapse, this relapse, sobriety yesterday, recently slipped, have relapsed, recently relapsed, relapses, first day back, fell again, fell off, reset, slipped, relapsed again, relapsing, relapsed <time-frame>, back from a relapse, just relapsed
drugUser	doing drugs, taking drugs, smoking, cannabis, weed, marijuana, crack, cocaine, molly, drug addict, blasted, i do drugs, na meeting, psychedelics, opiates
medicated	medicated, pills, antidepressants
rockBottom	suicidal, bankrupt, broken, divorce, rock bottom, end of my rope, jail, criminal, finances, court, disaster, disastrous
struggling	getting stuck, underlying issues, i don't know what to do, hate aa, make amends, struggled, i'm a mess, dry drunk, hung up, i need help, throwaway, mistake, struggling, been rough, wits end, institution, setback, in rehab, detoxing, withdrawal, stuck, trapped, screwed

Table 5. PSL-Reddit-1 Fine-grained Topic Dictionary

the dominant topic during inference in HL-MRFs. As the original probability value usually would be a small value even for the predominant topic(s), scaling helps the model represent this better. This way, if the user talked about a topic more than 0.05 of the time, the adjusted topic score would reflect this in terms the model can better represent. We also ensure that the adjusted score was at most 1.0. This is accomplished by taking the minimum of the scaled value and 1.0. We then extract similar topic percentages for pages and singular posts, calculating the distribution of topics across all posts in a page and the distribution of topics across posts. We refer to those as pageBased and textBased topic percentage, respectively. We also calculate a topic feature *recoveryCount* and a *relapseCount*, counting the number of words mentioned from any of the 12 recovery topics, and from the 8 relapse topics, respectively.

Percentage User Recovers. For each user, we took the *recoveryCount* divided by *totalVocabSum*, the total words said in all topics.

Percentage User Relapses. Similarly, for each user, we calculate the *relapseCount* / *totalVocabSum.*

4.2.3 Psycho-linguistic Features. We extract psycho-linguistic features using LIWC [47]. We consider the LIWC categories that are most relevant to our problem – *affect* and *social* categories.

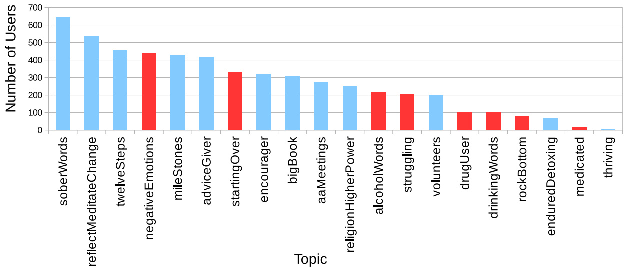


Fig. 5. Figure showing the topic distribution of users in recovery and relapse topics

Affect includes postive/negative sentiment, presence of words that signify anxiety, anger, and sadness. As discussed earlier, social support is known to be an important factor in recovery from AUD[3, 23, 38, 49]. The social category in LIWC captures the presence of words in users' or users' friends' tweets that signify family and friendship. We calculate the LIWC features on the documents created by concatenating each user's tweets. LIWC gives the number of words in each of these categories. We normalize this score across all the users, giving us a score between [0, 1] for each user, for affect and social. We refer to them as *affect* and *social* in our models. We use the psycho-linguistic features extracted using LIWC in combination with other linguistic and structural features.

4.2.4 Structural Features. We generate structural features by considering various forms of interactions between pairs of users in the Twitter and Reddit network. Note that some structural features are only relevant for the Twitter network.

Friends. We capture bi-directional followers for the AA users and refer to them as friends. We capture each pair of AA user U_1 and friend U_2 using *friends*(U_1, U_2). This feature is only relevant for the Twitter social network.

Replies. The reply network captures the tweets that are replies between the user and his/her friend in the network. Note that replies is a directed graph, with replies going from AA users to friends and from friends to AA users. We encode each pair-wise interaction in the replies network using *replies*(U_1, U_2, T), where U_1 replies to U_2 and T denotes the reply tweet.

Retweets. Similar to the replies network, the retweet network captures tweets that are retweets between the user and his/her friend in the network. Note that the retweet network is also a directed graph, containing links representing retweets from AA users to friends and from friends to AA users. We encode each pair-wise interaction in the retweet network using *retweets*(U_1, U_2, T), where U_1 retweets U_2 and T denotes the tweet that was retweeted. This feature is also only relevant for the Twitter social network.

Similarity. We also construct another derived network based on the similarity of users. We consider two ways of encoding similarity between pairs of users in the network. We first consider users' tweets 90 days before the user joins AA to 90 days after the user joins AA. We concatenate the tweets for this duration and calculate the cosine similarity between the tweets for pairs of AA users, using term frequency-inverse document frequency vectors. We refer to this similarity as tweetSimilarity(U_1, U_2). We only consider pairs of users with tweet similarity value more than the median value of 0.65 in our models. Second, we calculate the similarity between the LIWC scores for pairs of AA users calculated on concatenated user tweets before and after the user joins AA.

This value is encoded in *LIWCSimilarity*(U_1, U_2). We calculate and use *LIWCSimilarity* in both the Twitter and Reddit models.

4.3 HL-MRF Recovery Prediction Models

Here, we present our HL-MRF models that encode dependencies among the linguistic, psychological, and structural features to predict recovery. We design two main variations: i) *PSL-Twitter*, which models the structural relationships in the Twitter social network to reason about recovery, and ii) *PSL-Reddit*, which lays emphasis on the rich linguistic content in the verbose Reddit posts and structure within them to reason about recovery.

PSL-Twitter Model
Set A. Combining Linguistic Features:
Term Frequency from Logistic Regression
termFrequency(U) → recovers(U)
¬termFrequency(U) → ¬recovers(U)
Seeded LDA Topics
userTopic(U, "alcohol") → ¬recovers(U)
userTopic(U, "sober") → recovers(U)
Presence of Alcohol Words
containsAlcoholWord(U, T) ∧ attendsAA(U) → ¬recovers(U)
¬usesAlcoholWord(U) ∧ attendsAA(U) → recovers(U)
containsSoberWord(U, T) ∧ attendsAA(U)) → recovers(U)
¬usesSoberWord(U) ∧ attendsAA(U)) → ¬recovers(U)
Set B. Combining Structural and Linguistic Features:
replies(U1, U2) ∧ retweets(U2, U1, T) ∧ containsAlcoholWord(U2, T) ∧ attendsAA(U1) → ¬recovers(U1)
retweets(U1, U2) ∧ retweets(U2, U1, T) ∧ containsAlcoholWord(U2, T) ∧ attendsAA(U1) → ¬recovers(U1)
replies(U1, U2) ∧ replies(U2, U1, T) ∧ containsAlcoholWord(U2, T) ∧ attendsAA(U1) → ¬recovers(U1)
retweets(U1, U2) ∧ replies(U2, U1, T) ∧ containsAlcoholWord(U2, T) ∧ attendsAA(U1) → ¬recovers(U1)
replies(U1, U2) ∧ retweets(U2, U1, T) ∧ containsSoberWord(U2, T) ∧ attendsAA(U1) → recovers(U1)
retweets(U1, U2) ∧ retweets(U2, U1, T) ∧ containsSoberWord(U2, T) ∧ attendsAA(U1) → recovers(U1)
replies(U1, U2) ∧ replies(U2, U1, T) ∧ containsSoberWord(U2, T) ∧ attendsAA(U1) → recovers(U1)
retweets(U1, U2) ∧ replies(U2, U1, T) ∧ containsSoberWord(U2, T) ∧ attendsAA(U1) → recovers(U1)
Set C. Combining structural and topic features:
replies(U2, U1, T) ∧ friendTweetTopic(U2, T, alcohol) ∧ attendsAA(U1) → ¬ recovers(U1)
retweets(U2, U1, T) ∧ friendTweetTopic(U2, T, alcohol) ∧ attendsAA(U1) → ¬ recovers(U1)
retweets(U2, U1, T) ∧ friendTweetTopic(U2, T, sober) ∧ attendsAA(U1) → recovers(U1)
replies(U2, U1, T) ∧ friendTweetTopic(U2, T, sober) ∧ attendsAA(U1) → recovers(U1)
Set D. Combining structural, linguistic and psycho-linguistic features from LIWC on friends' tweets
affect(U1) ∧ friends(U1,U2) ∧ usesAlcoholWord(U2) → ¬recovers(U1)
affect(U1) ∧ friends(U1, U2) ∧ usesSoberWord(U2) → recovers(U1)
Set E. Collective Rules:
tweetSimilarity(U1, U2) ∧ recovers(U2) → recovers(U1)
tweetSimilarity(U1, U2) ∧ recovers(U1) → recovers(U2)
tweetSimilarity(U1, U2) ∧ ¬recovers(U2) → ¬recovers(U1)
tweetSimilarity(U1, U2) ∧ ¬recovers(U1) → ¬recovers(U2)

Table 6. HL-MRF Model combining linguistic, psycho-linguistic, and structural features for the Twitter Dataset

4.3.1 PSL-Twitter. Our complete model is presented in Table 6. We group the rules into different groups based on the features they combine. We use our HL-MRF model to: i) capture dependencies among different linguistic features, ii) capture dependencies among different linguistic and structural features, iii) capture different forms of structural interactions between AA users and friends, and

iv) reason collectively about AA user's recovery, and capture their effect on recovery. The weights of these rules are learned during training. The weights capture how important each of these rules are in predicting recovery and relapse. We explain the different rule-groups below.

Combining Linguistic Features. In the rules in set A, we capture the dependencies between linguistic features and recovery. The first two rules capture the dependency between *termFrequency* and recovery. In the second group of rules in set A, we capture the dependency between seeded LDA topic of a user and his/her recovery. *userTopic(U, "alcohol")* captures the value in the document-topic multinomial distribution for the alcohol seeded topic and *userTopic(U, "sober")* captures the value in the document-topic multinomial distribution for the sober seeded topic. In the third group of rules in set A, we capture the dependency between alcohol/sober word usage and recovery.

Nature of support	Retweets/replies containing alcohol/sober words
Supporting alcoholism	@... drink your **beer** snort your gear. RT @...: I need **vodka**. @... it's okay, we are **drunk** everyday. What are you plans for the day?!
Supporting sobriety	@... I struggled during early **sobriety**. Never skip meetings, call your sponsor or have coffee with a **sober** friend @... Do you need a **sober** companion? We're here for you. RT @...: Tips for the **sober** beginner! I contributed to @XXX's blog, which is run by the XX nonprofit

Table 7. Example Alcohol/Sober Replies/Retweets from Friends to AA Users supporting alcoholism/sobriety, respectively

Combining Linguistic and Structural Features. The rules in set B combine the linguistic features with structural features *replies* and *retweets* between pairs of users. We hypothesize that if an AA user retweets or replies to alcohol-word containing tweets by her friends, then it is more likely that she will not recover from AUD (Rules 1 – 4 in set B). Table 7 gives some examples of retweets/replies that contain alcohol words. We observe that such tweets can hurt AA user's potential to recovery as they may lead the AA user to relapse to alcohol. For example, tweets 1, 2, and 3 are tweets from friends where they mention the AA user, inviting him/her to drink or instances where the AA user retweets friends' tweets on alcohol. Similarly, interactions with friends on sobriety could potentially aid AA user's recovery from AUD. We model friends' tweets on sobriety that the AA user replies/retweets (Rules 5 – 8 in set B). Tweets 4, 5, 6 in Table 7 give examples of support sobriety. there are also friends tweets on sobriety, supporting the AA users, pointing them to necessary resources, and providing encouragement and support. We model both these signals from friends' tweets in our model.

Further, we observe that friends with whom there is reply/retweet activity have more effect on the AA user, when compared to all user's friends. Hence, we filter the user-friend network to include only pairs of users that have a significant amount of interaction in the form of retweets/replies, and model the effect of specific tweet exchanges between them that contain alcohol/sober words. For example, the first rule captures friends with whom there is a significant amount of interaction in the form of replies (*replies*(U_1, U_2)) and considers retweet exchanges between them that contain alcohol words (*retweets*(U_2, U_1, T)). Note that *replies*(U_1, U_2) does not contain specific tweet T as it considers all the replies between pairs of users. Similarly, the second rule only considers pairs of users that have a significant number of retweet exchanges (*retweets*(U_1, U_2)) and consider the specific tweets that have alcohol/sober words (*retweets*(U_2, U_1, T)).

Combining Structural and Topic Features. Here, we model the effect of friends' tweets that contain alcohol/sober words on the AA-users' recovery. *friendTweetTopic*(U_2, T, alcohol) captures if friends' replies and retweets to AA users belong to alcohol/sober topics. If they fall under alcohol/sober category, then we capture that it could affect AA users recovery negatively ($\neg$*recovers*) or positively (*recovers*), respectively.

Combining Structural and Psycho-linguistic Features. Here, we combine psychological features extracted from LIWC with network features to predict recovery. Rule 1 in Set D captures that if a user is more emotional (given by *affect*), then he/she is more likely to be affected by friends' alcoholic tweets. Similarly, more emotional users also are more likely to be affected by friends' sober tweets (rule 2 in set D).

Collective Rules. Collective rules capture that similar users tend to have similar recovery patterns. We include both similarity values *tweetSimilarity* and *LIWCSimilarity* in our model. We filter similarity values greater than the median of 0.65 and include only the propagation of recovery among AA users with similarity values more than the median. In our PSL-Twitter model we observe that using *tweetSimilarity* or *LIWCSimilarity* gives identical performance prediction. In Section 5, we perform a detailed feature-group analysis where we observe that when *LIWCSimilarity* is used instead of *tweetSimilarity*, it helps in improving prediction performance in the absence of other stronger linguistic signals from AA-users' tweets.

4.3.2 PSL-Reddit. As we explore the different sets of rules in Table 6 from the PSL-Twitter model, we find structural rules do not work as well for the Reddit dataset. While Twitter is a social networking site, the subreddit is more a forum for sharing personal experiences. Also, the majority of users on the subreddit do not know one another in real life, which means the relationship between pairs of users in Reddit is weaker when compared with Twitter users. Unlike the data we collect from Twitter where there is a character limit on the tweets, Reddit contains verbose detailed posts that describe users' recovery and relapse. The presence of verbose posts in Reddit and the absence of an underlying social network lead us to designing models focused on the linguistic content in the posts and structural relationships within/among the posts. Further, the lengthier nature of posts also increases the possibility of noise in it making HL-MRFs a good choice for modeling the uncertainty using continuous values. We explore linguistic features and structure between them at two levels: i) *PSL-Reddit-1*, which uses features and dependencies at the document and user level, and ii) *PSL-Reddit-2*, which uses features and dependencies at the sentence level.

PSL-Reddit-1. In the first model for Reddit, we include post-level and user-level linguistic features: relapse and recovery topics, positive and negative sentiment, and psycho-linguistic features from LIWC. Table 8 gives the rules in *PSL-Reddit-1.* Set A in Table 8 gives the topic rules combining relapse and recovery topics, users' participation in the subreddit, and mentions of relapse by the user. Set B gives the rules on sentiment, where they are used separately as well as together with the discussion forum structure given by *replies.* In Set C, we construct collective rules using similarity based on LIWC scores.

PSL-Reddit-2. In the second model for Reddit, we consider each sentence in users' posts and extract key features using a dependency parser and combine them with topic features in *PSL-Reddit-1.* The topic features are re-extracted using a simplified topic dictionary. We use spaCy [22] for constructing the dependency parses. A breakdown of the different components of the dependency

***PSL-Reddit-1:* Post and User-level Features and Dependencies**
U, U_1, U_2: AA-attending user or friend; T_1, T_2: sober topics; T_3, T_4: alcohol topics; P_1, P_2: posts, Pg: Reddit Forum Page
Set A: Based on topics discussed among each user and subredditTraffic
$\mathit{subredditTraffic}(U) \longrightarrow \mathit{currentlySober}(U)$
$\mathit{soberTopic}(U, T_1) \land \mathit{soberTopic}(U, T_2) \longrightarrow \mathit{currentlySober}(U)$
$\mathit{percentageUserRecovers}(U) \longrightarrow \mathit{currentlySober}(U)$
$\mathit{alcoholTopic}(U, T_3) \land \mathit{alcoholTopic}(U, T_4) \longrightarrow \neg\, \mathit{currentlySober}(U)$
$\mathit{percentageUserRelapses}(U) \longrightarrow \neg\, \mathit{currentlySober}(U)$
Set B: Based on Sentiment Rules
$\mathit{posSentiment}(U) \longrightarrow \mathit{currentlySober}(U)$
$\mathit{posSentiment}(Pg) \land \mathit{pageMap}(Pg, U) \longrightarrow \mathit{currentlySober}(U)$
$\mathit{posSentiment}(U, P) \land \mathit{replies}(U, U_1, P_1) \land \mathit{replies}(U_1, U, P_2) \longrightarrow \mathit{currentlySober}(U)$
$\mathit{negSentiment}(U) \longrightarrow \neg\, \mathit{currentlySober}(U)$
$\mathit{negSentiment}(Pg) \land \mathit{pageMap}(Pg, U) \longrightarrow \neg\, \mathit{currentlySober}(U)$
$\mathit{negSentiment}(U, P_1) \land \mathit{replies}(U, U_1, P_1) \land \mathit{replies}(U_1, U, P_2) \longrightarrow \neg\, \mathit{currentlySober}(U)$
Set C: Based on LIWC Rules
$\mathit{similar}(U_1, U_2) \land \mathit{currentlySober}(U_1) \longrightarrow \mathit{currentlySober}(U_2)$
$\mathit{similar}(U_1, U_2) \land \mathit{currentlySober}(U_2) \longrightarrow \mathit{currentlySober}(U_1)$
$\mathit{similar}(U_1, U_2) \land\neg\, \mathit{currentlySober}(U_1) \longrightarrow \neg\, \mathit{currentlySober}(U_2)$
$\mathit{similar}(U_1, U_2) \land\neg\, \mathit{currentlySober}(U_2) \longrightarrow \neg\, \mathit{currentlySober}(U_1)$

Table 8. Table showing the rules in *PSL-Reddit-1* model

parse is given in Figure 6. For each sentence, we focus on the subject, action, and context labels in the dependency parse and extract the features described below.

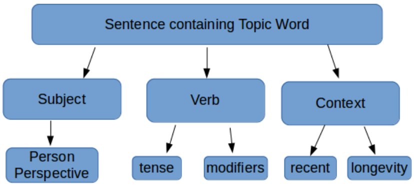


Fig. 6. Breakdown of Sentence Analysis using dependency parser.

Person perspective. For each sentence, we look for pronouns indicating if the sentence uses first, second, or third person perspective.

containsTopicWord. For each sentence, we look for a word in any given topic. The topic words are actual words, rather than phrases. This is not as elaborate as the topic phrases in our PSL-Reddit-1 model as we are integrating the topic dictionary with structure information from the dependency parser.

partOfSpeech. For each extracted topic word in a sentence, we record the part of speech of the word. We concentrate on topic words identified in *PSL-Reddit-1* such as *drink* and *sober*. Additionally for a verb, the part of speech contains information on the tense and person perspective. For instance, VBP means Verb, non-3rd person singular present.

posModifiers. For each sentence, we extract words that qualify the verb of the sentence and assert it semantically. For instance "*still* drinking" emphasizes drinking and qualifies it in a positive manner.

negModifiers. For each sentence, we extract words that qualify the verb of the sentence and negate/reduce the semantic contribution of the word. For instance "*not* drinking" and "*stopped* drinking" negate/reduce the contribution of the word "drinking".

longevityContext. Here, we extract words that refer to long periods of time, for instance, "*three years* sober".

recentContext. We also extract words that refer to recent periods of time, for instance, "drank *yesterday*", "*just* drank", "drinking *today*".

inPost and postMap. *inPost* identifies the post corresponding to the sentence being analyzed and *postUser* maps the post to the corresponding Reddit user.

Sentence	Context	Perspective	Topic Word	Tense	Neg Modifier	Prediction
Personally I have not relapsed in almost 34 years and don't plan for it to happen today.	34 years	I (1st person)	relapsed	VBN	not	Recovered
It feels like a punishment because you have only been sober a few weeks.	few weeks	you (2nd person)	sober	N/A	N/A	Irrelevant
I woke up the beast, I'm drinking again.	N/A	I'm (1st person)	drinking	VBG	N/A	Relapsed

Table 9. Dependency Parser Examples

Our main goal here is to simplify the topic dictionary shown in Table 5 and present a more general model for recovery and relapse prediction that is capable of capturing the linguistic patterns at the sentence level. In *PSL-Reddit-1*, we use phrases to get meaningful context. Here, we use the dependency parser to extract the verb tense of topic words that involve some action, such as drinking (the verb form of drink) or relapsing. This allows us to use a simpler topic dictionary than what is shown in Table 5, only containing the words and not exact phrases or linguistic patterns. This paves the way for the model to be generalizable to other datasets. The trade off is lower prediction values than *PSL-Reddit-1*, but for a more general and automated approach that still performs relatively well and is easily extensible to future datasets. We believe that the combination of a comprehensive linguistic analysis both using a vocabulary (*PSL-Reddit-1*) and showing how to capture sentence-level patterns using HL-MRFs (*PSL-Reddit-2*), makes our models and analysis helpful for understanding recovery and relapse better.

Rules for predicting currentlySober. The PSL rules in Table 10 show how to combine the dependencies between different features extracted from the dependency parser for each sentence to reason about recovery and relapse. We focus on the fine-grained topics and words associated with them and first identify the sentences by the user that contain these words. Then, we incorporate contextual information surrounding the topic word such as the parts of speech tag of the topic word, positive and negative modifiers, first, second, or third person, temporal context, and connect them to the user through the post in which the sentence is present. For example, the first rule captures that if a user's most posts belong to the fine-grained topic "sober" and it is mentioned in a first person context along with a longevity context, then there is a higher chance of the user being currently sober. Below, we outline the sequential process of constructing the rule templates for predicting *currentlySober* in Table 10.

***PSL-Reddit-2:* Linguistic Structural Dependencies at the Sentence-level using Dependency Parser**
S: Sentence, W: Word, POS: Part of Speech, P: Post, PER: first, second, or third person, CON: Context, MOD: Modifier, U: User
1. *userTopic(U, "Sober") ∧ containsSoberWord(S, W, POS) ∧ firstPerson(S, PER) ∧ longevityContext(S, CON) ∧ inPost(S, P) ∧ postMap(P, U) ⟶ currentlySober(U)*
2. *userTopic(U, "Encourager") ∧ containsEncouragingWord(S, W, POS) ∧ secondPerson(S, PER) inPost(S, P) ∧ postMap(P, U) ⟶ currentlySober(U)*
3. *userTopic(U, "TwelveSteps") ∧ containsTwelveStepsWord(S, W, POS) ∧ firstPerson(S, PER) ∧ longevityContext(S, CON) ∧ inPost(S, P) ∧ postMap(P, U) ⟶ currentlySober(U)*
4. *userTopic(U, "AAMeetings") ∧ containsAAMeetingWord(S, W, POS) ∧ firstPerson(S, PER) ∧ longevityContext(S, CON) ∧ inPost(S, P) ∧ postMap(P, U) ⟶ currentlySober(U)*
5. *userTopic(U, "AdviceGiver") ∧ containsAdviceWord(S, W, POS) ∧ secondPerson(S, PER) ∧ inPost(S, P) ∧ postMap(P, U) ⟶ currentlySober(U)*
6. *userTopic(U, "BigBook") ∧ containsBigBookWord(S, W, POS) ∧ firstPerson(S, PER) inPost(S, P) ∧ postMap(P, U) ⟶ currentlySober(U)*
7. *userTopic(U, "MileStones") ∧ containsMileStoneWord(S, W, POS) ∧ firstPerson(S, PER) ∧ inPost(S, P) ∧ postMap(P, U) ⟶ currentlySober(U)*
8. *userTopic(U, "ReflectChange") ∧ containsReflectChangeWord(S, W, POS) ∧ firstPerson(S, PER) ∧ inPost(S, P) ∧ postMap(P, U) ⟶ currentlySober(U)*
9. *userTopic(U, "HigherPower") ∧ containsHigherPowerWord(S, W, POS) ∧ firstPerson(S, PER) ∧ inPost(S, P) ∧ postMap(P, U) ⟶ currentlySober(U)*
10. *userTopic(U, "Volunteers") ∧ containsVolunteerWord(S, W, POS) ∧ firstPerson(S, PER) ∧ inPost(S, P) ∧ postMap(P, U) ⟶ currentlySober(U)*
11. *userTopic(U, "Drinking") ∧ containsDrinkWord(S, W, POS) ∧ firstPerson(S, PER) ∧ negModifier(S, MOD) ∧ inPost(S, P) ∧ postMap(P, U) ⟶ currentlySober(U)*
12. *userTopic(U, "StartingOver") ∧ containsStratingOverWord(S, W, POS) ∧ firstPerson(S, PER) ∧ recentContext(S, CON) ∧ inPost(S, P) ∧ postMap(P, U) ⟶ ¬ currentlySober(U)*
13. *userTopic(U, "Struggling") ∧ containsStruggleWord(S, W, POS) ∧ firstPerson(S, PER) ∧ recentContext(S, CON) ∧ inPost(S, P) ∧ postMap(P, U) ⟶ ¬ currentlySober(U)*
14. *userTopic(U, "Drinking") ∧ containsDrinkWord(S, W, POS) ∧ firstPerson(S, PER) ∧ recentContext(S, CON) ∧ posModifier(S, MOD) ∧ inPost(S, P) ∧ postMap(P, U) ⟶ ¬ currentlySober(U)*
15. *userTopic(U, "Alcohol") ∧ containsAlcoholWord(S, W, POS) ∧ firstPerson(S, PER) ∧ recentContext(S, CON) ∧ inPost(S, P) ∧ postMap(P, U) ⟶ ¬ currentlySober(U)*
16. *userAlcoholWords(U) ∧ containsAlcoholWord(S, W, POS) ∧ firstPerson(S, PER) ∧ posModifier(S, MOD) ∧ inPost(S, P) ∧ postMap(P, U) ⟶ ¬ currentlySober(U)*
17. *userTopic(U, "DrugUse") ∧ containsDrugWord(S, W, POS) ∧ firstPerson(S, PER) ∧ recentContext(S, CON) ∧ inPost(S, P) ∧ postMap(P, U) ⟶ ¬ currentlySober(U)*
18. *userTopic(U, "DrugUse") ∧ containsDrugWord(S, W, POS) ∧ firstPerson(S, PER) ∧ posModifier(S, MOD) ∧ inPost(S, P) ∧ postMap(P, U) ⟶ ¬ currentlySober(U)*
19. *userTopic(U, "RockBottom") ∧ containsRockBottomWord(S, W, POS) ∧ firstPerson(S, PER) ∧ recentContext(S, CON) ∧ inPost(S, P) ∧ postMap(P, U) ⟶ ¬ currentlySober(U)*
20. *userTopic(U, "RockBottom") ∧ containsRockBottomWord(S, W, POS) ∧ firstPerson(S, PER) ∧ posModifier(S, MOD) ∧ inPost(S, P) ∧ postMap(P, U) ⟶ ¬ currentlySober(U)*
21. *userTopic(U, "NegEmotions") ∧ containsNegEmotionsWord(S, W, POS) ∧ firstPerson(S, PER) ∧ recentContext(S, CON) ∧ inPost(S, P) ∧ postMap(P, U) ⟶ ¬ currentlySober(U)*
22. *userTopic(U, "NegEmotions") ∧ containsNegEmotionsWord(S, W, POS) ∧ firstPerson(S, PER) ∧ posModifier(S, MOD) ∧ inPost(S, P) ∧ postMap(P, U) ⟶ ¬ currentlySober(U)*

Table 10. Table showing rules in *PSL-Reddit-2* model

To contextualize the presence of alcohol/sober words in posts, we first select users that talk about a specific topic given by the first predicate *userTopic(U, "topic")*, where topic refers to the different fine-grained topics. Next, we select sentence(s) in the post which contain topic-related words, which is given by the second predicate, such as *containsSoberWord.* Now, we examine if the sentence mentions the topic words in the first person, given by *firstPerson.* Then, we contextualize the mention of a topic word in a temporal context, which captures if the user has been sober for quite some time, given by *longevityContext.* We then map this back to the user through *inPost*, which captures the post which contains the sentence *S* and *postMap*, which captures the user who wrote the post.

We incorporate a couple of modifications to the above template to capture some scenarios. To capture users who provide advice to others and encourage others in the forum, we capture if the mention of the topic words is in second person. Rules 2 and 5 in Table 10 capture this phenomena. Since the presence of drinking words is prevalent across all users in the dataset, we further use the negative modifiers to capture if the user mentions drinking in the first person in a negative way, which would indicate they feel negatively toward drinking and indicate being currently sober.

Rules for predicting ¬currentlySober. We follow a similar process for constructing the ¬*currentlySober* rule templates with a few exceptions. Contrary to the currently sober rules, if a user uses positive modifiers along with drinking words, this would indicate that they feel positively toward drinking and possibly indicate that they are not currently sober. For users who specify words in the StartingOver, DrugUse, RockBottom topics, we capture if they mention that in a recent context (given by *recentContext*), which would possibly indicate a recent relapse. Additionally, we analyze the number of groundings of the rules to make them more general versus more specific. In rules 17 – 22, we do not include the predicates *posModifier* and *recentContext* in the same rule as they lead to specific rules creating too few groundings. We discuss this more in our grounding analysis in Section 5.2.4.

5 EXPERIMENTAL EVALUATION

In this section, we present quantitative and qualitative results of our model on the annotated Twitter and Reddit AA datasets. We conduct a suite of experiments to evaluate the predictive ability of our model in predicting recovery in both the datasets. For the Twitter dataset, we consider two time periods: i) 90 days after the user joins AA, and ii) 1 year after the user joins AA. Next, we consider different combinations of features and their dependencies for both datasets and evaluate the contribution of different feature groups in predicting recovery. Our experimental analysis sheds light on the importance of different feature groups and dependencies among them in predicting recovery and relapse.

5.1 PSL-Twitter Experiments

5.1.1 Recovery Prediction at 90 days. In our first set of experiments, we predict recovery of AA users at 90 days after the user joins AA. Table 11 gives the results of our PSL-Twitter model. We compare our PSL-Twitter model to a logistic regression model that uses all the linguistic and psycho-linguistic features derived from users' tweets. The only additional features we include in our PSL models are structural features that capture pair-wise interactions between AA users and their friends in the network. Aggregate features such as number of friends, number of friends mentioning alcohol in their tweets, number of friends retweeting alcohol/sober tweets unfortunately do not add value to the prediction problem as there is not much difference in these values across users, especially relapsing and recovering users. Figure 3(a) gives an example of this phenomena in the dataset, where a recovering user has friends that mention both alcohol and sober words. While we capture the strength and contribution of each such interaction differently using our HL-MRF models, we can see that the actual number of mentions is not a helpful feature. This is the reason we build our modeling approach with particular emphasis on pair-wise interactions that consider interactions that add value to the prediction problem and do not include the noisier aggregate signals in the logistic regression baseline.

For all PSL and logistic regression experiments, we perform 5-fold cross-validation. Table 11 shows the comparison results between logistic regression and our PSL-Twitter model. We report area under the precision recall curve for the positive class (AUC-PR Pos., recovery), the negative class (AUC-PR Neg., relapse) and area under the receiver operating characteristics curve (AUC-ROC). AUC-PR and AUC-ROC are combination metrics analogous to F1 but more suitable for continuous-valued predictions. AUC-PR is more informative than AUC-ROC for datasets with class imbalance [52], so we include both in our evaluation. We observe that our PSL-Twitter model performs better at both recovery and relapse prediction. Statistically significant values with a rejection threshold of $p = 0.05$ are typed in **bold**.

Model	AUC-PR Pos.	AUC-PR Neg.	AUC-ROC
Logistic Regression	0.580	0.890	0.711
PSL-Twitter	**0.755**	**0.940**	**0.903**

Table 11. Area under precision-recall curve and ROC values for recovery and relapse prediction at 90 days for PSL-Twitter and Logistic Regression models.

5.1.2 Recovery Prediction at 1 year. Next, we predict recovery at 1 year after the user joins AA. 26 people were removed from the dataset as they could not be labeled with recovery/relapse after 90 days. 13 users moved from recovery to relapse, decreasing the number of recovered users in 1 year. Hence, there are 239 relapsed and 37 recovered users in our 1-year data.

We perform two experiments here: first, we train the model on data from 90 days and use the trained model to predict recovery at the 1-year mark. Second, we train on data from 1 year after the user joins AA and use the trained model to predict recovery at the 1-year mark. Table 12 gives the results from training on 90-day data and 1-year data. We observe that training on data from 90 days gives us a superior prediction performance for recovery and relapse, when compared to training on 1-year data. In AA, 90 days is regarded as a crucial milestone for recovery. In this experiment, we aim to study that empirically in online data. First, we observe that the data becomes noisier and the signals become more diluted in the time period between 90 days and 1 year. The most relevant information corresponding to recovery and relapse occur before 90 days in our data as well, confirming the need to choose this time period as a critical milestone in recovery. The data is noisier and less relevant for the prediction problem after the 90-day mark. Hence, our results confirm that the first 90 days after the user joins AA are crucial in modeling the user's path to recovery.

We also notice that AUC-PR Pos. scores, i.e., recovery prediction scores are lower than the corresponding scores in the 90-day prediction model. We explain this as follows. There is a reduction in the number of recovered users in the dataset with 13 users moving from recovery to relapse and 26 people who could not be labeled as they do not mention being sober after 90 days. This brings down the number of recovered users from 76 to 37. Thus, for predicting at the 1-year mark, there is an even more pronounced class imbalance, which also contributes to a reduction in the prediction performance.

Training Period	AUC-PR Pos.	AUC-PR Neg.	AUC-ROC
90-day	0.5849	0.9793	0.9351
1-year	0.4509	0.9652	0.8733

Table 12. Area under precision-recall curve and ROC values for recovery and relapse prediction at 1 year for PSL-Twitter trained on 90-day data and 1-year data.

5.1.3 *Analysis of Linguistic, Psycho-linguistic, and Structural Feature Groups in* PSL-Twitter.

In this section, we present analysis of different features groups in our model and their respective contribution to predicting recovery. We perform two different analyses of the features. First, we construct different versions of our models by leaving out different sets of features/rules and analyze the corresponding effect on prediction performance. We also analyze the weights learned by our model for the different rule sets in Table 6.

To empirically evaluate the predictive ability of various feature groups in predicting recovery, we construct variations of the *PSL-Twitter* model by leaving out features, groups of features, and dependencies among them. To distinguish the *PSL-Twitter* model from other variants, we refer to this model as *PSL-Twitter-All* in below experiments. Figure 7 shows the AUC-ROC, AUC-PR Pos. (recovery), and AUC-PR Neg. (relapse) results for the different PSL models. We explain the different variations of our *PSL-Twitter* model below.

PSL-Twitter-All. This model uses all the rules shown in Table 6. The performance scores from this model are captured in the far left in Figure 7.

PSL-Linguistic. PSL-Linguistic model uses linguistic features drawn from AA users' tweets and dependencies among them. The rules for this model are captured in set A in Table 6. Note that the PSL-Linguistic model does not use any linguistic features on friends' tweets, relational, or psycho-linguistic features.

Variations of PSL-Relational Twitter Model

PSL-Relational
Set D. Combining structural and linguistic on friends' tweets (excluding psycho-linguistic features from LIWC):
friends(U1, U2) ∧ usesAlcoholWord(U2, aw) → ¬ recovers(U1)
friends(U1, U2) ∧ usesSoberWord(U2, sw) → recovers(U1)

PSL-Topic
Set C. Combining structural and topic features without sentiment:
replies(U2, U1, T) ∧ friendTweetTopic(U2, T, alcohol) ∧ attendsAA(U1) → ¬ recovers(U1)
retweets(U2, U1, T) ∧ friendTweetTopic(U2, T, alcohol) ∧ attendsAA(U1) → ¬ recovers(U1)
retweets(U2, U1,T) ∧ friendTweetTopic(U2, T, sober) ∧ attendsAA(U1) → recovers(U1)
replies(U2, U1, T) ∧ friendTweetTopic(U2, T, sober) ∧ attendsAA(U1) → recovers(U1)
Set D. Combining structural and linguistic on friends' tweets (excluding psycho-linguistic features from LIWC):
friends(U1, U2) ∧ usesAlcoholWord(U2, aw) → ¬ recovers(U1)
friends(U1, U2) ∧ usesSoberWord(U2, sw) → recovers(U1)

PSL-Psychological (Affect)
Set D. Combining structural features and linguistic on friends' tweets with *affect*:
affect(U1) ∧ friends(U1, U2) ∧ usesAlcoholWord(U2, aw) → ¬ recovers(U1)
affect(U1) ∧ friends(U1, U2) ∧ usesSoberWord(U2, sw) → recovers(U1)

PSL-Psychological (Social)
Set D. Combining structural features and linguistic on friends' tweets with *social*:
social(U1) ∧ friends(U1, U2) ∧ usesAlcoholWord(U2, aw) → ¬ recovers(U1)
social(U1) ∧ friends(U1, U2) ∧ usesSoberWord(U2, sw) → recovers(U1)

PSL-Sentiment
Set C. Combining structural and topic features with sentiment:
replies(U2, U1, T) ∧ friendTweetTopic(U2, T, alcohol) ∧ postiveSentiment(U2, T) ∧ attendsAA(U1) → ¬ recovers(U1)
retweets(U2, U1, T) ∧ friendTweetTopic(U2, T, alcohol) ∧ postiveSentiment(U2, T) ∧ attendsAA(U1) → ¬ recovers(U1)
retweets(U2, U1, T) ∧ friendTweetTopic(U2, T, sober) ∧ postiveSentiment(U2, T) ∧ attendsAA(U1) → recovers(U1)
replies(U2,TU1, T) ∧ friendTweetTopic(U2, T, sober) ∧ postiveSentiment(U2, T) ∧ attendsAA(U1) → recovers(U1)

PSL-LIWCSimilarity
Set E. Collective Rules with LIWCSimilarity:
LIWCsimilarity(U1, U2) ∧ recovers(U2) → recovers(U1)
LIWCsimilarity(U1, U2) ∧ recovers(U1) → recovers(U2)
LIWCsimilarity(U1, U2) ∧ ¬ recovers(U2) → ¬ recovers(U1)
LIWCsimilarity(U1, U2) ∧ ¬ recovers(U1) → ¬ recovers(U2)

Table 13. Table showing the additional rules in different variations of PSL-Relational Twitter model.

PSL-Relational. Next, we consider the model with structural features from interactions with friends in the network and dependencies among them, as captured in rules in sets B, D, and E. The rules in set D are modified to exclude the psycho-linguistic features from LIWC, making it a model which relies only on the presence of alcohol/sober words in the structural interactions between AA users and friends (retweets/replies). The performance scores for this model is captured immediately after *PSL-Linguistic* in Figure 7. We notice that even without linguistic features from AA users' tweets and including only structural interactions with friends in the network, we can achieve reasonably high prediction scores. This demonstrates the importance of modeling structural interactions for understanding recovery and relapse.

We conduct experiments on variations of *PSL-Relational* by incrementally adding other linguistic/ psycho-linguistic features extracted from friends' tweets and combining them with structural features. These models do not use any linguistic features on user's tweets and solely rely on linguistic analysis of friends' tweets and structural features to predict AA user's recovery. Some of

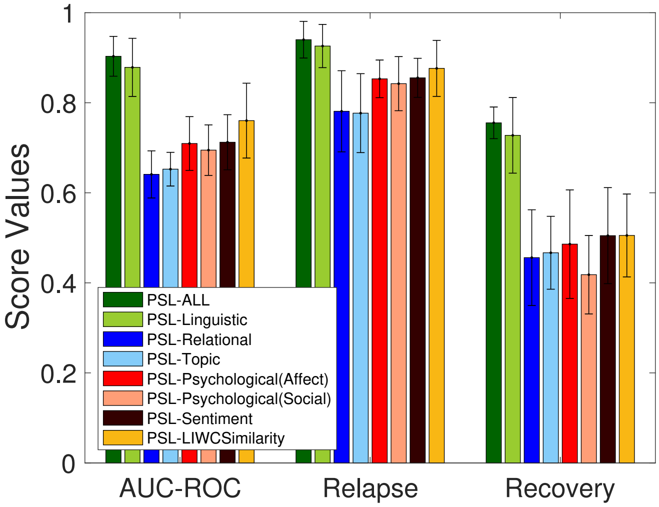


Fig. 7. Graph showing area under precision-recall curve and ROC values for recovery and relapse prediction for the different variations of *PSL-Twitter* model

these features do not help in improving the prediction performance when other stronger linguistic features from AA users' tweets are present, but when added to the relational model, they help in achieving a significant performance improvement. This exercise is helpful in understanding how to effectively combine linguistic and psycho-linguistic signals with relational features and how the different signals from structural interactions affect user's recovery. Table 13 gives the different variations and the changes to the different rule sets in the *PSL-Twitter* model in Table 6. We explain each of these models in detail below.

PSL-Topic. PSL-Topic model adds the *friendTweetTopic* feature, adding rules in set C to the *PSL-Relational* model. Notice that adding rules in C gives a slight performance improvement in AUC-ROC and AUC-PR. Pos. (Recovery) scores. This demonstrates that topic of friends' tweets on retweets/replies is a helpful signal in predicting recovery/relapse.

PSL-Psychological (Affect). PSL-Psychological (Affect) model uses all the rules in sets *B*, *C*, and *E* as given in Table 6 that are part of *PSL-Topic* model. The rules in set *D* are enhanced to consider the *affect* of users and their interaction with structural features as given by rules under *PSL-Psychological (Affect)* in Table 13.

PSL-Psychological (Social). PSL-Psychological (Social) model uses almost the same rules to *PSL-Psychological (Affect)* model, except instead of using *affect* feature, it uses the *social* feature in set *D*. We notice that between *affect* and *social*, including *affect* gives a better performance improvement.

PSL-Sentiment. PSL-Sentiment considers all the rules in *PSL-Topic* sets *B*, *D*, and *E*. The rules in set *C* are updated to include a combination of topic and sentiment as shown in Table 13 inside *PSL-Sentiment.* Notice that here the topic of friends' tweets together with the sentiment of the tweet accounts for their sentiment towards alcohol/sobriety and helps in accurately modeling the effect of the friends' tweet on user's recovery. The high performance scores of this model reaffirms that social interactions play a crucial role in user's recovery. We notice that including sentiment with topic gives us the most improvement when compared to adding other psycho-linguistic features to *PSL-Relational.*

PSL-LIWCSimilarity. PSL-LIWCSimilarity model uses all the rules in sets B, D, and E. The collective rules in set E are replaced with *LIWCSimilarity* instead of *tweetSimilarity.* We notice that this gives us a significant performance improvement over using tweet similarity. In fact, the scores for *PSL-LIWCSimilarity* are greater than *PSL-Sentiment* which includes other linguistic features such as topics, sentiment, and affect. While including *LIWCSimilarity* is not significantly helpful when other linguistic features from users' tweets are present, in absence of these features, we find that it achieves a significant improvement. This helps in understanding the importance of collective rules in our model and how to construct efficient models to predict recovery using combinations of available features.

Analysis of Learned Weights. The weights of the PSL-Twitter model are learned at training time. The weights capture the predictive capability of the model in predicting recovery. Analyzing the weights, we find that rules containing linguistic features get the highest weights after training, ascertaining that the textual content in the tweets is the strongest signal in predicting recovery. Following that, the next highest weights are observed in the rules that combine network and linguistic features, with alcohol-related topics/words getting higher weights than sober-related topics. Since the interactions with friends emerge as a strong feature in the learned weights, we conduct experiments on relational features and their dependence with psychological features to understand their role in predicting recovery.

5.2 PSL-Reddit Experiments

5.2.1 Recovery Prediction. We first present results of *PSL-Reddit-1* and *PSL-Reddit-2* models on the Reddit dataset compared with a Logistic Regression baseline in Table 14. Note that this is a more difficult prediction problem, as here we are considering the problem of predicting whether the user is currently sober. There are more cases of currently sober users in the Reddit dataset, which makes it hard to predict the negative class. Additionally, from the observations in our dataset, we find that 15.7% percent of recovered users mention a past relapse, making it challenging to differentiate users. The dependencies captured in our HL-MRF rules help in capturing the complex relationship of the user with alcohol, sobriety, and recovery, thus helping in predicting recovery and relapse accurately despite the significant class imbalance in the dataset.

The real challenge in this dataset is to get the model to perform well for the negative (¬currentlySober) class. With the presence of the class imbalance, obviously rules for the positive class learn higher weights, since predicting positive class would produce good overall results by the model. Our first model, PSL-Reddit-1 focuses on the presence of linguistic signals while PSL-Reddit-2 captures important contextual cues and relevance at the sentence level. For instance, the PSL-Reddit-2 model not only looks for the mention of relapse, but also captures if the user is giving a personal account of a relapse (talking in the first person perspective), and in a recent context to be classified as relapsed. A similar instance for classifying a user as currently sober would be that the user mentioned being sober in a longevity context (e.g., "I've been sober for 5 years"). PSL-Reddit-2 provides us the ability to look for how words were related to one another at a sentence level. This presents us with the opportunity to develop rules that better capture how humans would read a sentence and make a judgement on the relevance of the sentence as a whole. This explains the better AUC-PR scores for both the positive (currentlySober) and the negative (¬currentlySober) class for PSL-Reddit-2.

5.2.2 Analysis of Linguistic, Psycho-linguistic, and Sentiment Feature Groups in PSL-Reddit-1. Here, we consider variations of *PSL-Reddit-1* by considering various combinations of the different sets of

Model	AUC-PR Pos.	AUC-PR Neg.	AUC-ROC
Logistic Regression	0.931	0.537	0.711
PSL-Reddit-1	**0.953**	0.448	**0.793**
PSL-Reddit-2	**0.957**	**0.570**	**0.836**

Table 14. Area under precision-recall curve and ROC values for two PSL-Reddit models and Logistic Regression.

rules in Table 8 and examine the effect on the prediction. Figure 8 shows the different combinations. This feature/rule analysis exercise is helpful in understanding how the different features/feature groups combine to form the rules. We exclude all other rules except the set of rules in question and measure how they perform. Our analysis also draws insights on the predictive importance of different rules/features in the dataset, which is helpful in understanding why the model performs the way it does on this dataset. Our analysis and insights from this effort are helpful when extending the model to another similar dataset/domain, where only a subset of features are relevant.

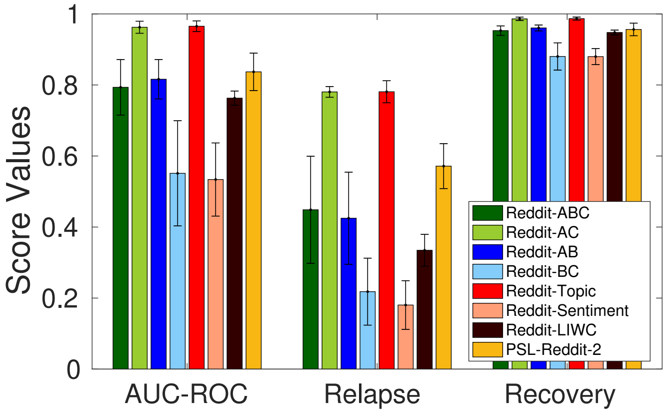


Fig. 8. Graph showing area under precision-recall curve and ROC values for recovery and relapse prediction for the different variations of *PSL-Reddit* model

PSL-Reddit-Topic. In the first variation, we isolate the fine-grained topic rules in Set A in Table 8, which capture the alcohol and sober topics and other user-level features. We find that this model achieves the best prediction performance for both recovery and relapse, indicating that the fine-grained topic analysis is helpful for understanding users' journey toward recovery.

PSL-Reddit-Sentiment. In this model, we use the sentiment rules in Set B alone and we find that this model performs the worst, indicating that sentiment in isolation can be a confusing signal. In *PSL-Reddit-2*, we use the sentiment indicators along with the alcohol/topic words they qualify, which helps in accurately modeling the sentiment in the posts.

PSL-Reddit-LIWC. In this model, we only consider the psycho-linguistic signals from LIWC in Set C and find that they outperform *PSL-Reddit-Sentiment*, indicating that they are a stronger signal when compared to sentiment. Also, note that the rules in Set C are collective in nature (i.e., jointly predicting recovery/relapse for pairs of users based on the LIWC similarity), indicating that psychological similarity is a good signal for predicting recovery and relapse and that structural signals are helpful across both datasets. This further ascertains that social and psychological signals are important features to consider when modeling this problem.

We also consider combinations of *PSL-Reddit-Topic* (A), *PSL-Reddit-Sentiment* (B), and *PSL-Reddit-LIWC* (C). We find that all combinations are able to achieve performance comparable to the best model *PSL-Reddit-Topic* for predicting recovery while fine-grained linguistic features and dependencies among them remain crucial for predicting relapse, as the dataset is skewed toward recovery. We find that *PSL-Reddit-AC*, which combines topic and LIWC features performs the best.

5.2.3 Analysis of Learned Weights. For each model of the Reddit dataset, we initialize the weights with 5.0 for each rule. From the different variations of *PSL-Reddit-1* summarized in Figure 8, it is clear that the fine-grained topic rules are the best predicting rules for this dataset. After training, we observe that these topic rules generally attain higher values than other rules, such as sentiment/collective LIWC rules, when they are combined in our model.

In the *PSL-Reddit-2* model, we observe that recovery rules generally obtain higher weights after training, which can be attributed to the class imbalance in the dataset. However, because of the additional contextual information encoded in the rules, the mere presence of topic words no longer implies recovery/relapse. For instance, for the relapse topic, words such as “hangover”, “whiskey”, or “binge” did not score high for predicting relapse. On the other hand, some relapse rules do attain higher weights. For instance, rules that get grounded for words that mention “slipped” in a recent context, “bender” in a recent context while using a positive modifier, and mentioning the drug “molly” in the first person and recent context end up with higher weights after training when compared with other rules for relapse.

The LIWC collective rules are our third best set of rules. Although the scores are not much higher for the negative class, the rules do not hurt the model when we combine rules from sets A (topic) and C (LIWC) together, and the weights stay around the 4.8 - 4.9 range after training.

Although the sentiment rules prove to not help in the prediction for the Reddit dataset, there is a clear pattern of relevance to be noted. The weights suggest that sentiment scores are most relevant for a user overall, or a given page on the subreddit. Sentiment scores on individual posts are not that useful. This makes intuitive sense, since scoring sentiment at a higher level gives a better idea of the user’s overall sentiment throughout their posts. The sentiment at page level also makes sense, because understanding the overall sentiment of a page gives some indication of the topic brought up by the original user. For instance, collective positive sentiment on a page could suggest that the original poster is celebrating a milestone.

5.2.4 Grounding Coverage Analysis. In this section, we perform an analysis of the number of groundings of rules across the PSL Reddit models to understand user behavioral patterns in our dataset. We perform this analysis on the Reddit models as the rules here focus on the AA-attending users and linguistic cues from their posts. Our analysis sheds light on the generality and specificity of the different rules across the models. To effectively classify different kinds of users, a combination of both these kinds of rules are necessary.

We show how to measure the relevance of the different rules across both models by using following three main metrics:

1) *Average Coverage*: *How many users get grounded to at least 1 rule?* Users that are not grounded in any rule in the model are not represented by the model. The coverage for each model was 100%, i.e., all users are represented by at least one rule in the model. In other words, every user was represented and recognized by our PSL Reddit models.

2) *Average groundings*: *What is the average number of instances grounded to each rule?* This metric captures the average recall of the rules across the instances. The higher number of instances

grounded to a rule indicates that the rule has greater recall in the dataset and is more general. We also need specific rules that are able to distinguish between instances correctly and effectively capture corner cases in the dataset, which rules with lesser number of grounded instances accomplish.

3) *Average overlap*: *For each user, how many rules get grounded on average?* This metric gives the degree of overlap across rules in users. When a user gets grounded to multiple rules in the model, many different features, their dependencies, and associated uncertainty is captured, paving the way for a better prediction performance.

Rule Set	Groundings	Rules	Avg Groundings	Avg Overlap	Overlap %
A (Fine-grained Topics)	33,549	97	345.87	41.27	42.55 %
B (Sentiment)	64,983	6	10,830.50	4.28	71.33 %
C (LIWC)	1,319,528	4	329,882.0	4.00	100.00 %
D (Dependency Parse Structure)	79,217	250	316.87	14.11	5.64 %

Table 15. Rule analysis of Reddit models.

Table 15 gives the grounding metrics for the different sets of rules in the PSL Reddit model. Since the average coverage is 100%, we omit this metric from the table. We categorize the rules into four groups: i) fine-grained topic rules (Set A), ii) sentiment rules (Set B), iii) LIWC rules (Set C), and iv) rules combining fine-grained topics and dependency parse structure (PSL-Reddit-2, Set D). From the table, we observe that rules in Set A have the highest degree of overlap (41.27 of the 97 rules grounded on average for each user), making them more general than other rules in our model. This means that our *PSL-Reddit-1* model is considering a user in multiple contexts simultaneously to make a more informed prediction. However, note that general rules can also prove to be almost too encompassing and not helpful in differentiating between data points. If everyone is grounded to a majority of contradictory rules, the model loses its ability to make distinctions between users being recovered or relapsed. For instance, considering the sentiment rules, 4.28 out of 6 rules are grounded to each user. This is 71.33% of all sentiment rules, which may explain why the sentiment rules do not contribute much to the prediction problem (PSL-Reddit-Sentiment in Figure 8). Using the grounding analysis, we are able to understand that users are displaying occurrences of both positive and negative sentiment within their posts, which makes the signal noisy.

Note that the total number of groundings is highest for the LIWC rules. This is due to the fact that the LIWC rules are collective rules, i.e., we are comparing two users pairwise to each other in these rules. This means that the expected number of instances grounded to these rules will be around (813 choose 2), or $330,078$, covering all possible pairwise user combinations, thus leading to a high number of groundings.

Of course, the more specific a rule becomes, a fewer number of instances will be grounded to satisfy that rule. For instance, consider the *PSL-Reddit-2* rule set (Set D). For any given user, on average only 14.11 rules were associated with that user to make a prediction by the model. However, this low overlap is expected and helpful for the prediction problem because the *PSL-Reddit-2* rule set was meant to make rules that were more specific, capturing a greater distinction for each user considered. Hence, these rules really help in modeling the distinct aspects of a user's journey toward recovery.

To highlight some factors worth considering, we present some examples of rules in the *PSL-Reddit-2* model that correspond to the most groundings for recovery and relapse. The highest grounded recovery rule comes from the sober topic, where the user mentions the word 'sober' in a first person perspective (first rule in Table 16). There are $7,268$ groundings to this rule, and the final learned weight of the rule is 3.9717. Notice that this rule combines both the presence of sober words as calculated by the sober topic feature and the presence of first person perspective in the dependency parse structure, thus only retaining users who are mentioning sobriety from a first-person perspective. Similarly, analyzing the groundings of rules for relapse, the highest grounded relapse rule is from the *starting over* topic, where the user mentions the word 'relapsed' in a first person perspective (first rule in Table 17). There are 477 groundings to this rule, and the final learned weight of the rule is 2.5455. Again, we can see that intuitively a person starting over is likely to have relapsed in the recent past and the inclusion of first-person perspective helps in disambiguating users who may be showing support and encouragement to the relapsed users.

Recall that the primary motivating reason behind the development of the *PSL-Reddit-2* model was to create more specific rules that tend to have fewer number of groundings. To prove this, we consider the rules in the sober and relapse topics which have the highest learned weights for the *PSL-Reddit-2* model. The rule with the highest weight for recovery prediction considers users mentioning the word 'abstinence' in the first person (second rule in Table 16). There are 79 grounded instances and the rule has an ending weight of 4.4406. Likewise, consider the highest scoring rule for relapse prediction, where the user mentions 'relapsed' in a recent context (second rule in Table 17). There are 16 grounded instances, and the rules has a final learned weight of 4.4257. This shows considering recent contexts is important for categorizing mentions of relapse, and that rules with less groundings have more significance in the *PSL-Reddit-2* model as they can distinguish between recoverers and relapsers better.

Recovery Rule	Groundings	Learned Weight
SoberWords(User) ∧ ContainsSoberWord(Sentence, 'soberword', Tag)) ∧ InPost(Sentence, PostId) ∧ PostedBy(PostId, User)) ∧ Perspective(Sentence, "FirstPerson") → CurrentlySober(User)	7268	3.97
SoberWords(User) ∧ ContainsSoberWord(Sentence, 'soberword', Tag)) ∧ InPost(Sentence, PostId) ∧ PostedBy(PostId, User)) ∧ Perspective(Sentence, "FirstPerson") → CurrentlySober(User)	78	4.44

Table 16. Recovery rules with most number of groundings and highest weight after training, respectively.

Relapse Rule	Groundings	Learned Weight
StartingOver(User) ∧ ContainsRelapseWord(Sentence, 'relapsed', Tag) ∧ InPost(Sentence, PostId) ∧ PostedBy(PostId, User)) ∧ Perspective(Sentence, "FirstPerson") → CurrentlySober(User)	477	2.54
StartingOver(User) ∧ ContainsRelapseWord(Sentence, 'relapsed', Tag) ∧ InPost(Sentence, PostId) ∧ PostedBy(PostId, User)) ∧ RecentContext(Sentence, Context) → CurrentlySober(User)	16	4.43

Table 17. Relapse rules with most number of groundings and highest weight after training, respectively.

6 DISCUSSION AND FUTURE DIRECTIONS

In this paper, we presented a myriad of different structured prediction models and demontrated how to incorporate various fine-grained linguistic and psycholinguistic features from users' and friends' posts and structural interactions with friends and effectively encode dependencies among them to model and understand recovery and relapse from/into AUD across two popular scenarios in social media data. We perform a thorough quantitative and qualitative analysis on the prediction problem involving features and grounding, which unearths different user behavioral patterns, throws light on the respective ability of the feature groups in predicting recovery, and helps in potentially extending our models to similar prediction problems, thus serving as a generic template for studying similar computational social science problems.

There are many exciting directions to go from here. Our structured prediction approach demonstrates how to combine dependency parsing and fine-grained topics, which can be extended further to model even more intricate relationships in the data. For example, the dependency parsing and the HL-MRF model can be improved to account for double negatives when extracting negative modifiers. For instance, "haven't stopped drinking" has two negative modifiers, "stopped" and "haven't", but they cancel each other in this phrase. Our approach can also be extended to differentiate between multiple temporal contexts in a single sentence such as, "I recently thought about my last relapse two years ago". Here, there are a few time contexts, "recently" and "two years ago". The "recently" links with "thought" and "two years ago" is associated with "relapse". Similarly, extracting context from a verb's base form may also be an extension, where we could capture the dependency in sentences such as "I wish I could stay sober". Here, the user wishes they *could* stay sober, which implies they cannot. By contrast, the sentence "I knew I could stay sober" refers to sobriety on a positive note. The sentence's context changes completely by the change of a single verb "wish" to "knew". Our approach provides the ideal foundation for modeling these complex linguistic expressions in future works.

Acknowledgements. We thank Dhanya Sridhar for her help with initial experiments in the conference version of this paper. This work was supported by SUNY Binghamton, NSF grants CCF-1740850 and IIS-1703331. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of NSF, or the U.S. Government.

REFERENCES

[1] 2017. Research finds Reddit helping 40,000 recovering addicts find sobriety. http://www.dailymail.co.uk/femail/article-4093086/Research-finds-Reddit-helping-40-000-recovering-alcoholics-sobriety.html.

[2] Tim Althoff, Eric Horvitz, Ryen W White, and Jamie Zeitzer. 2017. Harnessing the web for population-scale physiological sensing: A case study of sleep and performance. In *Proceedings of the Conference on World Wide Web (WWW)*.

[3] American Psychological Association. 2012. Recovery principles. http://www.apa.org/monitor/2012/01/recovery-principles.aspx.

[4] Stephen H. Bach, Matthias Broecheler, Bert Huang, and Lise Getoor. 2017. Hinge-Loss Markov Random Fields and Probabilistic Soft Logic. *Journal of Machine Learning Research (JMLR)* 18 (2017), 1–67.

[5] Sairam Balani and Munmun De Choudhury. 2015. Detecting and Characterizing Mental Health Related Self-Disclosure in Social Media. In *Proceedings of the Conference on Human Factors in Computing Systems*.

[6] Ana-Maria Bliuc, Muhammad Iqbal, and David Best. 2019. Integrating computerized linguistic and social network analyses to capture addiction recovery capital in an online community. *Journal of Visualized Experiments* 147 (2019), e58851.

[7] Munmun De Choudhury and Emre Kiciman. 2017. The Language of Social Support in Social Media and its Effect on Suicidal Ideation Risk. In *Proceedings of the International Conference on Web and Social Media (ICWSM)*.

[8] Shaika Chowdhury, Chenwei Zhang, and Philip S Yu. 2018. Multi-task pharmacovigilance mining from social media posts. In *Proceedings of the Conference on World Wide Web (WWW)*.

[9] Neil S Coulson. 2005. Receiving social support online: an analysis of a computer-mediated support group for individuals living with irritable bowel syndrome. *Journal of CyberPsychology & Behavior* (2005), 580–584.

[10] Brenda L Curtis, Robert D Ashford, Katherine I Magnuson, and Stacy R Ryan-Pettes. 2019. Comparison of Smartphone Ownership, Social Media Use, and Willingness to Use Digital Interventions Between Generation Z and Millennials in the Treatment of Substance Use: Cross-Sectional Questionnaire Study. *Journal of Medical Internet Research* 21, 4 (2019), e13050.

[11] Munmun De Choudhury and Sushovan De. 2014. Mental Health Discourse on reddit: Self-Disclosure, Social Support, and Anonymity.. In *Proceedings of the International Conference on Web and Social Media (ICWSM)*.

[12] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. Predicting depression via social media.. In *Proceedings of the International Conference on Web and Social Media (ICWSM)*.

[13] Ronda L Dearing, Jeffrey Stuewig, and June Price Tangney. 2005. On the importance of distinguishing shame from guilt: Relations to problematic alcohol and drug use. *Addictive Behaviors* 30, 7 (2005), 1392–1404.

[14] Tao Ding, Warren K Bickel, and Shimei Pan. 2017. Multi-view unsupervised user feature embedding for social media-based substance use prediction. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

[15] Tao Ding, Fatema Hasan, Warren K Bickel, and Shimei Pan. 2018. Interpreting social media-based substance use prediction models with knowledge distillation. In *Proceedings of the International Conference on Tools with Artificial Intelligence (ICTAI)*.

[16] M. Dredze. 2012. How Social Media Will Change Public Health. *Journal of IEEE Intelligent Systems* 27 (2012), 81–84.

[17] Marica Ferri, Laura Amato, and Marina Davoli. 2006. Alcoholics Anonymous and other 12-step programmes for alcohol dependence. *The Cochrane Library* (2006).

[18] Pollyanna Gonçalves, Matheus Araújo, Fabrício Benevenuto, and Meeyoung Cha. 2013. Comparing and combining sentiment analysis methods. In *Proceedings of the Conference on Online social networks*.

[19] Donald S Grant and Karen E Dill-Shackleford. 2017. Using social media for sobriety recovery: Beliefs, behaviors, and surprises from users of face-to-face and social media sobriety support. *Psychology of Popular Media Culture* 6, 1 (2017), 2.

[20] Christian Greiner, Anne Chatton, and Yasser Khazaal. 2017. Online self-help forums on cannabis: A content assessment. *Journal of Patient Education and Counseling* (2017).

[21] Haripriya Harikumar, Thin Nguyen, Sunil Gupta, Santu Rana, Ramachandra Kaimal, and Svetha Venkatesh. 2016. Understanding Behavioral Differences Between Short and Long-Term Drinking Abstainers from Social Media. In *Proceedings of the International Conference on Advanced Data Mining and Applications (ADMA)*.

[22] Matthew Honnibal and Mark Johnson. 2015. An Improved Non-monotonic Transition System for Dependency Parsing. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

[23] Misra K. Epner A. Cooper G. M. Horvath, T. 2010. Social Learning Theory of Addiction and Recovery Implications. http://www.amhc.org/1408-addictions/article/48347-social-learning-theory-of-addiction-and-recovery-implications.

[24] Nabil Hossain, Tianran Hu, Roghayeh Feizi, Ann Marie White, Jiebo Luo, and Henry A. Kautz. 2016. Precise Localization of Homes and Activities: Detecting Drinking-While-Tweeting Patterns in Communities. In *Proceedings of the International Conference on Web and Social Media (ICWSM)*.

[25] Hsiao-Ying Huang and Masooda Bashir. 2016. Online Community and Suicide Prevention: Investigating the Linguistic Cues and Reply Bias. In *Proceedings of the Conference on Human Factors in Computing Systems*.

[26] George M Hunt and Nathan H Azrin. 1973. A community-reinforcement approach to alcoholism. *Journal of Behaviour Research and Therapy* 11 (1973), 91–104.

[27] Rosenquist J, Murabito J, Fowler JH, and Christakis NA. 2010. The spread of alcohol consumption behavior in a large social network. *Journal of Annals of Internal Medicine* 152 (2010), 426–433.

[28] Jagadeesh Jagarlamudi, Hal Daumé, III, and Raghavendra Udupa. 2012. Incorporating Lexical Priors into Topic Models. In *Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics (EACL)*.

[29] Deeptanshu Jha and Rahul Singh. 2019. SMARTS: the social media-based addiction recovery and intervention targeting server. *Bioinformatics* (2019).

[30] Hua Jin, Sean B Rourke, Thomas L Patterson, Michael J Taylor, and Igor Grant. 1998. Predictors of relapse in long-term abstinent alcoholics. *Journal of Studies on Alcohol* 59 (1998), 640–646.

[31] Yoram M. Kalman, Kathleen Geraghty, Cynthia K. Thompson, and Darren Gergle. 2012. Detecting Linguistic HCI Markers in an Online Aphasia Support Group. In *Proceedings of the Conference on Computers and Accessibility (SIGACCESS)*.

[32] Maulik R Kamdar and Mark A Musen. 2017. PhLeGrA: Graph analytics in pharmacology over the web of life sciences linked open data. In *Proceedings of the Conference on World Wide Web (WWW)*.

[33] Payam Karisani and Eugene Agichtein. 2018. Did You Really Just Have a Heart Attack?: Towards Robust Detection of Personal Health Mentions in Social Media. In *Proceedings of the Conference on World Wide Web (WWW)*.

[34] Animesh Koratana, Mark Dredze, Margaret S Chisolm, Matthew W Johnson, and Michael J Paul. 2016. Studying Anonymous Health Issues and Substance Use on College Campuses with Yik Yak. In *AAAI Workshop on WWW and Population Health Intelligence*.

[35] Adam D. I. Kramer, Susan R. Fussell, and Leslie D. Setlock. 2004. Text Analysis As a Tool for Analyzing Conversation in Online Support Groups. In *Proceedings of the Conference on Human Factors in Computing Systems*.

[36] Edward Loper and Steven Bird. 2002. NLTK: The Natural Language Toolkit. In *Proceedings of the ACL Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics*.

[37] 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 Conference on Computer Supported Cooperative Work & Social Computing (CSCW)*.

[38] Stephen A Maisto, Kate B Carey, and Clara M Bradizza. 1999. Social learning theory. *Journal of Psychological Theories of Drinking and Alcoholism* 2 (1999), 106–163.

[39] Lydia Manikonda, Heather Pon-Barry, Subbarao Kambhampati, Eric Hekler, and David W McDonald. 2014. Discourse analysis of user forums in an online weight loss application. In *Proceedings of the Joint workshop on social dynamics and personal attributes in social media*.

[40] David J McIver, Jared B Hawkins, Rumi Chunara, Arnaub K Chatterjee, Aman Bhandari, Timothy P Fitzgerald, Sachin H Jain, and John S Brownstein. 2015. Characterizing sleep issues using Twitter. *Journal of Medical Internet Research* 17 (2015).

[41] Yelena Mejova, Ingmar Weber, and Michael W Macy. 2015. *Twitter: a digital socioscope*. Cambridge University Press.

[42] Liesbeth Mercken, Christian Steglich, Ronald Knibbe, and Hein de Vries. 2012. Dynamics of Friendship Networks and Alcohol Use in Early and Mid-Adolescence. *Journal of Studies on Alcohol and Drugs* 73 (2012), 99–110.

[43] Megan A Moreno and Jennifer M Whitehill. 2014. Influence of social media on alcohol use in adolescents and young adults. *Journal of Alcohol Research: Current Reviews* 36 (2014), 91.

[44] World Health Organization and World Health Organization. Management of Substance Abuse Unit. 2014. *Global status report on alcohol and health, 2014*. World Health Organization.

[45] Michael J Paul and Mark Dredze. 2011. You are what you Tweet: Analyzing Twitter for public health.. In *Proceedings of the International Conference on Web and Social Media (ICWSM)*.

[46] Michael J. Paul and Mark Dredze. 2014. Discovering Health Topics in Social Media Using Topic Models. *PLOS ONE* 9 (2014), e103408.

[47] James W Pennebaker, Martha E Francis, and Roger J Booth. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology* 29(1) (2010), 24–54.

[48] Arti Ramesh, Dan Goldwasser, Bert Huang, Hal Daumé III, and Lise Getoor. 2014. Learning latent engagement patterns of students in online courses. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

[49] Julie Repper and Rachel Perkins. 2003. *Social inclusion and recovery: A model for mental health practice*. Elsevier Health Sciences.

[50] Hannah C Rettie, Lee M Hogan, and W Miles Cox. 2018. Negative attentional bias for positive recovery-related words as a predictor of treatment success among individuals with an alcohol use disorder. *Addictive Behaviors* 84 (2018), 86–91.

[51] Nicolas Rey-Villamizar, Prasha Shrestha, Farig Sadeque, Steven Bethard, Ted Pedersen, Arjun Mukherjee, and Thamar Solorio. 2016. Analysis of anxious word usage on online health forums. In *Proceedings of the International Workshop on Health Text Mining and Information Analysis*.

[52] Takaya Saito and Marc Rehmsmeier. 2015. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One* 10, 3 (2015).

[53] Sandra Servia-Rodríguez, Kiran K Rachuri, Cecilia Mascolo, Peter J Rentfrow, Neal Lathia, and Gillian M Sandstrom. 2017. Mobile sensing at the service of mental well-being: a large-scale longitudinal study. In *Proceedings of the Conference on World Wide Web (WWW)*.

[54] Bruce Simons-Morton. 2007. Social influences on adolescent substance use. *American Journal of Health Behavior* 31 (2007), 672–684.

[55] Acar Tamersoy, Duen Horng Chau, and Munmun De Choudhury. 2017. Analysis of Smoking and Drinking Relapse in an Online Community. In *Proceedings of the International Conference on Digital Health*.

[56] Sabina Tomkins, Lise Getoor, Yunfei Chen, and Yi Zhang. 2018. A socio-linguistic model for cyberbullying detection. In *Proceedings of the Conference on Advances in Social Networks Analysis and Mining (ASONAM)*.

[57] Morgan Walker, Laura Thornton, Munmun De Choudhury, Jaime Teevan, Cynthia M. Bulik, Cheri A. Levinson, and Stephanie Zerwas. 2015. Facebook Use and Disordered Eating in College-Aged Women. *Journal of Adolescent Health* 57 (2015), 157–163.

[58] Yilin Wang, Jiliang Tang, Jundong Li, Baoxin Li, Yali Wan, Clayton Mellina, Neil O'Hare, and Yi Chang. 2017. Understanding and discovering deliberate self-harm content in social media. In *Proceedings of the Conference on World Wide Web (WWW)*.

[59] Yi-Chia Wang, Robert Kraut, and John M Levine. 2012. To stay or leave?: the relationship of emotional and informational support to commitment in online health support groups. In *Proceedings of the Conference on Computer Supported Cooperative Work (CSCW)*.

[60] Marsha White and Steve M Dorman. 2001. Receiving social support online: implications for health education. *Journal of Health Education Research* 16 (2001), 693–707.

[61] 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. *Journal of Cancer* 97 (2003), 1164–1173.

[62] Mark D Wood, Jennifer P Read, Roger E Mitchell, and Nancy H Brand. 2004. Do parents still matter? Parent and peer influences on alcohol involvement among recent high school graduates. *Journal of Psychology of Addictive Behaviors* 18 (2004), 19.

[63] Ronghua Xu and Qingpeng Zhang. 2016. Understanding online health groups for depression: social network and linguistic perspectives. *Journal of Medical Internet Research* 18, 3 (2016).

[64] Shaodian Zhang, Tian Kang, Lin Qiu, Weinan Zhang, Yong Yu, and Noémie Elhadad. 2017. Cataloguing treatments discussed and used in online autism communities. In *Proceedings of the Conference on World Wide Web (WWW)*.

[65] Yue Zhang and Arti Ramesh. 2018. Fine-Grained Analysis of Cyberbullying Using Weakly-Supervised Topic Models. In *Proceedings of the International Conference on Data Science and Advanced Analytics (DSAA)*.

[66] Yue Zhang, Arti Ramesh, Jennifer Golbeck, Dhanya Sridhar, and Lise Getoor. 2018. A Structured Approach to Understanding Recovery and Relapse in AA. In *Proceedings of the Web Conference (WWW)*.

[67] Bin Zou, Vasileios Lampos, and Ingemar Cox. 2018. Multi-task learning improves disease models from web search. In *Proceedings of the Conference on World Wide Web (WWW)*.