

# A Structured Approach to Understanding Recovery and Relapse in AA

Yue Zhang<sup>1</sup>, Arti Ramesh<sup>1</sup>, Jennifer Golbeck<sup>2</sup>, Dhanya Sridhar<sup>3</sup>, and Lise Getoor<sup>3</sup>  
SUNY Binghamton<sup>1</sup> University of Maryland, College Park<sup>2</sup> University of California, Santa Cruz<sup>3</sup>  
{yzhan202, artir}@binghamton.edu, golbeck@cs.umd.edu, {dsridhar, getoor}@ucsc.edu

## ABSTRACT

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 take a structured approach to understand recovery and relapse from AUD using social media data. To do so, we combine linguistic and psychological attributes of users with relational features that capture useful structure in the user interaction network. We evaluate our models on AA-attending users extracted from the Twitter social network and predict recovery at two different points—90 days and 1 year after the user joins AA, respectively. Our experiments reveal that our structured approach is helpful in predicting recovery in these users. We perform extensive quantitative analysis of different groups of features and dependencies among them. Our analysis sheds light on the role of each feature group and how they combine to predict recovery and relapse. Finally, we present a qualitative analysis of the different reasons behind users relapsing to AUD. Our models and analysis are helpful in making meaningful predictions in scenarios where only a subset of features are available and can potentially be helpful in identifying and preventing relapse early.

## KEYWORDS

Social Media Analysis, Probabilistic Graphical Models, Recovery from alcoholism, Alcoholics Anonymous

### ACM Reference Format:

Yue Zhang<sup>1</sup>, Arti Ramesh<sup>1</sup>, Jennifer Golbeck<sup>2</sup>, Dhanya Sridhar<sup>3</sup>, and Lise Getoor<sup>3</sup> SUNY Binghamton<sup>1</sup> University of Maryland, College Park<sup>2</sup> University of California, Santa Cruz<sup>3</sup> {yzhan202, artir}@binghamton.edu, golbeck@cs.umd.edu, {dsridhar, getoor}@ucsc.edu. 2018. A Structured Approach to Understanding Recovery and Relapse in AA. In *Proceedings of The 2018 Web Conference (WWW 2018)*. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3178876.3186019>

## 1 INTRODUCTION

Alcoholism, or alcohol 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 [30]. AUD can coexist with, 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 [19].

According to substance abuse recovery models, social support and inclusion play a pivotal role in a person’s recovery from addiction [1, 13, 24, 34]. 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 to recovery from AUD [10, 16]. 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 development of addiction and more generally, in consumption of alcohol [13]. Particularly, in the extremely sensitive recovery period, such influences have been shown to precipitate relapses in recovering alcohol addicts [19].

In this work, we present a structured approach to understand recovery from AUD. 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. 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 subpopulation, and can potentially provide insights on how to design more targeted and in-depth studies.

Specifically, 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 topic and sentiment of tweets, psycho-linguistic features from linguistic inquiry word count (LIWC) [33], and structural features from users’ interactions with friends.
- We encode dependencies among these features in a recently developed graphical modeling framework, hinge-loss Markov random fields [2], and jointly reason about the recovery/relapse of users. Our model effectively captures the relationship among the various features and serves as a template for modeling recovery from AUD.
- We perform predictions at different time periods and show that our structured 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

This paper is published under the Creative Commons Attribution 4.0 International (CC BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

WWW 2018, April 23–27, 2018, Lyons, France

© 2018 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC BY 4.0 License.

ACM ISBN 978-1-4503-5639-8/18/04.

<https://doi.org/10.1145/3178876.3186019>

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, signaling the importance of modeling social factors in recovery.

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 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 emphasize that social influence is a contributing factor in the development of substance abuse problems including AUD [29, 36, 42]. The recent explosion of social media has led to a growing interest in using social network interaction data to study issues of public health [9, 21, 28]. Chisolm et al. [21] use a college social network, Yik Yak, to study substance abuse and health issues in college campuses. Balani et al. and Choudhury et al. [3, 8] focus on identifying mental health issues, Walker et al. [38] focus on identifying eating disorders, and McIver et al. [26] identify sleep disorders using social network data. Mejova et al. [27] and Paul et al. [31, 32] use Twitter data to analyze public health issues. Hossain et al. [14] develop a model to predict geo-location based on whether a particular tweet signifies drinking and whether the user was drinking while writing that tweet. Other approaches focus on aggregate social behavior and report analysis of these behaviors over time in specific groups of users [17, 28]. There is also a growing body of literature on analyzing data from health support forums [11, 41]. Choudhury et al. [5, 7] and White et al. [40] study the presence of social support, Xu et al. [43] and Huang et al. [15] study linguistic cues in mental health support forums, and Coulson et al. [6] and Manikonda et al. [25] analyze data from a support group for irritable bowel syndrome and weight-loss, respectively. Rey-Villamizar et al. [35] study the use of anxious words, Kramer et al. [22] and Kalman et al. [20] study the linguistic markers, and Wang et al. [39] focus on the emotional aspects of online health forums. Tamersoy et al. [37], Harikumar et al. [12], and Maclean et al. [23] analyze recovery forums on alcohol and other substance abuse. 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 to recovery/relapse. We model the dependencies between linguistic and structural information, thus helping in understanding both their individual and combined contribution to predicting recovery.

## 3 DATA

In this section, we describe how we collect data for alcoholic users from Twitter and label them with recovery/relapse 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 our prediction models.

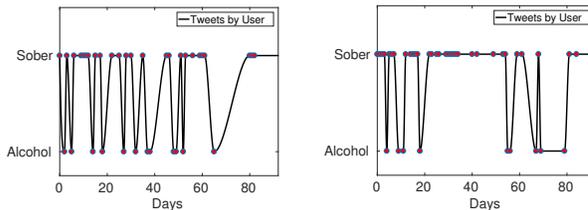
In order to collect information about AA members on Twitter, we extract tweets that mention "first AA meeting" over a four year

window from January 1, 2013 to February 1, 2017. From this set, we eliminate jokes and people attending to support an alcoholic friend or family member. This collection and filtering process results in 691 users. We then use the Twitter API to collect their most recent 3,200 tweets, the maximum allowable under Twitter’s Terms of Service. We separate these tweets into *before* and *after* the first AA meeting. After collecting data on AA-attending users, we then proceed to label these users as recovered/relapsed.

### 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/sobriety words, but do not directly imply consuming/refraining from alcohol, respectively**



(a) Recovered user with many alcohol (b) Relapsed user with many sober tweets

**Figure 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/sobriety 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/sobriety 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. 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. To establish if a user has recovered or not at the 90-day mark, we first filter tweets that mention alcohol/sobriety words after the user attends the first AA meeting. To determine users and friends’ alcohol/sobriety word usage, we construct an alcohol/sobriety vocabulary by mining tweets in our dataset. Table 2 gives the alcohol/sobriety vocabulary. We use this dictionary to filter tweets that use alcohol/sobriety words. Then, we examine whether the user’s tweet refers to consuming alcohol before the 90-day

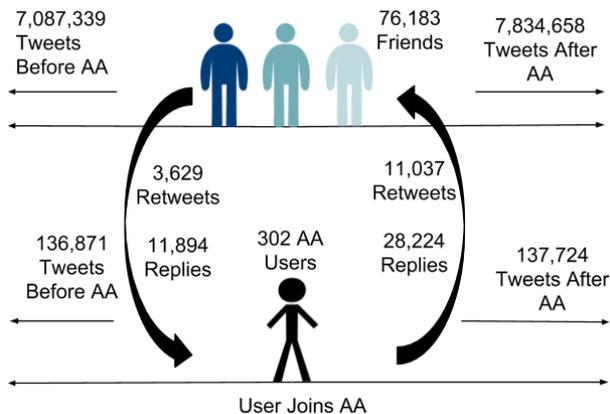
Category	Words
Alcohol	drunk, beer, bar, wine, alcohol, wasted, hungover, hangover, turnt, vodka, liquor, whiskey, tequila, alcoholic, champagne
Sober	recovery, sober, sobriety, #recovery, #sobriety

**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**

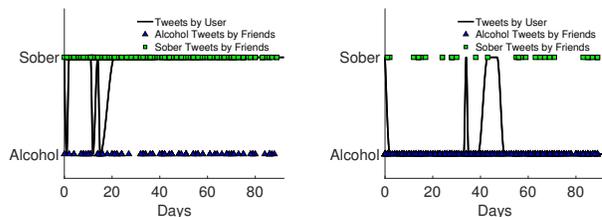
mark to establish that he/she did not recover. Similarly, to establish whether a user remained sober, we search for tweets where the user mentions being sober after joining AA. Table 3 gives examples of tweets from users labeled as recovered/relapsed depending on the content of their tweets after joining AA. We excluded subjects 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.



**Figure 2: Figure showing data statistics: number of AA users, friends, and tweets**

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 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) Recovered user with sober friends (b) Relapsed user with alcoholic friends

**Figure 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 [29, 36, 42]. 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, we see that this user has more friends who mention sober words as opposed to alcohol words. 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. We use these observations to develop our structured prediction models that take into account these structural interactions and their corresponding effect on recovery/relapse.

## 4 AUD RECOVERY PREDICTION MODELS

In this section, we build our HL-MRF recovery prediction models for predicting relapse/recovery<sup>1</sup>. We first extract features from users’ interactions 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, their tweets, and interactions with friends and effectively reason about the dependencies among these signals and their effect on recovery, we propose a powerful approach using hinge-loss Markov random fields (HL-MRFs). HL-MRFs can be specified using *Probabilistic Soft Logic (PSL)* [2], 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

$$\lambda : P(a) \wedge Q(a, b) \rightarrow R(b),$$

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

$$P(\mathbf{Y}|\mathbf{X}) \propto \exp \left( - \sum_{r=1}^M \lambda_r \phi_r(\mathbf{Y}, \mathbf{X}) \right)$$

$$\phi_r(\mathbf{Y}, \mathbf{X}) = (\max\{l_r(\mathbf{Y}, \mathbf{X}), 0\})^{\rho_r}, \quad (1)$$

<sup>1</sup>Our models are available at <https://github.com/yzhan202/zhang-www18-experiments>

where  $P(Y|X)$  is the probability density function of a subset of logical atoms  $Y$  given observed logical atoms  $X$ ,  $\phi_r(Y, X)$  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 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  $\text{alcoholTweet}(U_2, T)$  and  $U_1$  retweets it, given by  $\text{retweets}(U_1, U_2, T)$ . A PSL rule to encode the combined effect of these interactions on recovery is given by

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

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.,

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

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' tweets and interactions with their friends. We group the features into three broad categories: i) linguistic, ii) psycho-linguistic, and iii) structural features. We describe them in detail below.

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

**Term Frequency.** For each 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 effectively combine the term frequency features, we train a logistic regression classifier with term frequency of the words as features and use the prediction scores as a feature in our model. We refer to this feature as *termFrequency* in our model.

**Alcohol/Sober Word Usage.** 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.  $\text{containsAlcohol}(U_1, T)$  captures if a specific tweet  $T$  by

user  $U_1$  contains alcohol/sober words.  $\text{usesAlcoholWords}(U_1)$  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.

**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 [18]. 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 [18].

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  $k$  unseeded 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 for each AA user and treat that as one document. We have 302 AA users, giving us 302 documents in the first group. We refer to this feature as *userTopic* in our models. Next, we 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*.

**Sentiment Scores.** We compute sentiment scores for each retweet or reply from friends to users using SentiWordNet in natural language tool kit (NLTK) in Python [4]. SentiWordNet gives the number of positive and negative words in the document. We normalize the scores and treat scores closer to 1.0 as having positive sentiment and scores closer to 0.0 as negative sentiment.

**4.2.2 Psycho-linguistic Features .** We extract psycho-linguistic features using LIWC [33]. We consider the LIWC categories that are most relevant to our problem – *affect* and *social* categories. Affect includes positive/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 [1, 13, 24, 34]. 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.

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

**Table 4: Example Alcohol/Sober Replies/Retweets from Friends to AA Users supporting alcoholism/sobriety, respectively**

**4.2.3 Structural Features.** We generate structural features by considering various forms of interactions between pairs of users in 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  $friends(U_1, U_2)$ .

*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.

*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. 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  $LIWCsimilarity(U_1, U_2)$ .

### 4.3 HL-MRF Recovery Prediction Models

Here, we present our structured HL-MRF model that encodes dependencies among the linguistic, psychological, and structural features to predict recovery. Our complete model is presented in Table 5. 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.

**4.3.1 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.

**4.3.2 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 4 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 4 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

---

**PSL-Recovery Model**

---

**Set A. Combining Linguistic Features:****Term Frequency from Logistic Regression**

$$\begin{aligned} & \text{termFrequency}(U) \rightarrow \text{recovers}(U) \\ & \neg \text{termFrequency}(U) \rightarrow \neg \text{recovers}(U) \end{aligned}$$

**Seeded LDA Topics**

$$\begin{aligned} & \text{userTopic}(U, \text{"alcohol"}) \rightarrow \neg \text{recovers}(U) \\ & \text{userTopic}(U, \text{"sober"}) \rightarrow \text{recovers}(U) \end{aligned}$$

**Presence of Alcohol Words**

$$\begin{aligned} & \text{containsAlcoholWord}(U, T) \wedge \text{attendsAA}(U) \rightarrow \neg \text{recovers}(U) \\ & \neg \text{usesAlcoholWord}(U) \wedge \text{attendsAA}(U) \rightarrow \text{recovers}(U) \\ & \text{containsSoberWord}(U, T) \wedge \text{attendsAA}(U) \rightarrow \text{recovers}(U) \\ & \neg \text{usesSoberWord}(U) \wedge \text{attendsAA}(U) \rightarrow \neg \text{recovers}(U) \end{aligned}$$

**Set B. Combining Structural and Linguistic Features:**

$$\begin{aligned} & \text{replies}(U1, U2) \wedge \text{retweets}(U2, U1, T) \wedge \text{containsAlcoholWord}(U2, T) \wedge \text{attendsAA}(U1) \rightarrow \neg \text{recovers}(U1) \\ & \text{retweets}(U1, U2) \wedge \text{retweets}(U2, U1, T) \wedge \text{containsAlcoholWord}(U2, T) \wedge \text{attendsAA}(U1) \rightarrow \neg \text{recovers}(U1) \\ & \text{replies}(U1, U2) \wedge \text{replies}(U2, U1, T) \wedge \text{containsAlcoholWord}(U2, T) \wedge \text{attendsAA}(U1) \rightarrow \neg \text{recovers}(U1) \\ & \text{retweets}(U1, U2) \wedge \text{replies}(U2, U1, T) \wedge \text{containsAlcoholWord}(U2, T) \wedge \text{attendsAA}(U1) \rightarrow \neg \text{recovers}(U1) \\ & \text{replies}(U1, U2) \wedge \text{retweets}(U2, U1, T) \wedge \text{containsSoberWord}(U2, T) \wedge \text{attendsAA}(U1) \rightarrow \text{recovers}(U1) \\ & \text{retweets}(U1, U2) \wedge \text{retweets}(U2, U1, T) \wedge \text{containsSoberWord}(U2, T) \wedge \text{attendsAA}(U1) \rightarrow \text{recovers}(U1) \\ & \text{replies}(U1, U2) \wedge \text{replies}(U2, U1, T) \wedge \text{containsSoberWord}(U2, T) \wedge \text{attendsAA}(U1) \rightarrow \text{recovers}(U1) \\ & \text{retweets}(U1, U2) \wedge \text{replies}(U2, U1, T) \wedge \text{containsSoberWord}(U2, T) \wedge \text{attendsAA}(U1) \rightarrow \text{recovers}(U1) \end{aligned}$$

**Set C. Combining structural and topic features:**

$$\begin{aligned} & \text{replies}(U2, U1, T) \wedge \text{friendTweetTopic}(U2, T, \text{alcohol}) \wedge \text{attendsAA}(U1) \rightarrow \neg \text{recovers}(U1) \\ & \text{retweets}(U2, U1, T) \wedge \text{friendTweetTopic}(U2, T, \text{alcohol}) \wedge \text{attendsAA}(U1) \rightarrow \neg \text{recovers}(U1) \\ & \text{retweets}(U2, U1, T) \wedge \text{friendTweetTopic}(U2, T, \text{sober}) \wedge \text{attendsAA}(U1) \rightarrow \text{recovers}(U1) \\ & \text{replies}(U2, U1, T) \wedge \text{friendTweetTopic}(U2, T, \text{sober}) \wedge \text{attendsAA}(U1) \rightarrow \text{recovers}(U1) \end{aligned}$$

**Set D. Combining structural, linguistic and psycho-linguistic features from LIWC on friends' tweets**

$$\begin{aligned} & \text{affect}(U1) \wedge \text{friends}(U1, U2) \wedge \text{usesAlcoholWord}(U2) \rightarrow \neg \text{recovers}(U1) \\ & \text{affect}(U1) \wedge \text{friends}(U1, U2) \wedge \text{usesSoberWord}(U2) \rightarrow \text{recovers}(U1) \end{aligned}$$

**Set E. Collective Rules:**

$$\begin{aligned} & \text{tweetSimilarity}(U1, U2) \wedge \text{recovers}(U2) \rightarrow \text{recovers}(U1) \\ & \text{tweetSimilarity}(U1, U2) \wedge \text{recovers}(U1) \rightarrow \text{recovers}(U2) \\ & \text{tweetSimilarity}(U1, U2) \wedge \neg \text{recovers}(U2) \rightarrow \neg \text{recovers}(U1) \\ & \text{tweetSimilarity}(U1, U2) \wedge \neg \text{recovers}(U1) \rightarrow \neg \text{recovers}(U2) \end{aligned}$$

---

**Table 5: HL-MRF Model combining linguistic, psycho-linguistic, and structural features**

of users. Similarly, the second rule only considers pairs of users that have a significant number of retweet exchanges ( $\text{retweets}(U_1, U_2)$ ) and consider the specific tweets that have alcohol/sober words ( $\text{retweets}(U_2, U_1, T)$ ).

**4.3.3 Combining Structural and Topic Features.** Here, we model the effect of friends' tweets that contain alcohol/sober words on the AA-users' recovery.  $\text{friendTweetTopic}(U_2, T, \text{alcohol})$  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 ( $\neg \text{recovers}$ ) or positively ( $\text{recovers}$ ), respectively.

**4.3.4 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  $\text{affect}$ ), 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).

**4.3.5 Collective Rules.** Collective rules capture that similar users tend to have similar recovery patterns. We include both similarity values  $\text{tweetSimilarity}$  and  $\text{LIWCSimilarity}$  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-Recovery model we observe that using  $\text{tweetSimilarity}$  or  $\text{LIWCSimilarity}$  gives identical performance prediction. In Section 5, we perform a detailed feature-group analysis where we observe that when  $\text{LIWCSimilarity}$  is used instead of  $\text{tweetSimilarity}$ , it helps in improving prediction performance in the absence of other stronger linguistic signals from AA-users' tweets.

## 5 EXPERIMENTAL EVALUATION

In this section, we present quantitative and qualitative results of our model on the annotated Twitter AA user dataset. We conduct a suite of experiments to evaluate the predictive ability of our model in predicting recovery. 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 and evaluate the contribution of different feature groups in predicting recovery. Then, we present qualitative analysis of the different reasons corresponding to user’s relapse to alcoholism.

### 5.1 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 6 gives the results of our PSL-Recovery model. We compare our PSL-Recovery 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 pairwise interactions between AA users and their friends in the network, which the logistic regression model cannot encode. For all PSL and

Model	AUC-PR Pos.	AUC-PR Neg.	AUC-ROC
Logistic Regression	0.580	0.890	0.711
PSL-Recovery	<b>0.755</b>	<b>0.940</b>	<b>0.903</b>

**Table 6: Area under precision-recall curve and ROC values for recovery and relapse prediction at 90 days for PSL-Recovery and Logistic Regression models.**

logistic regression experiments, we perform 5-fold cross-validation. Table 6 shows the comparison results between logistic regression and our PSL-recovery 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). We observe that our PSL-recovery model performs better at both recovery and relapse prediction. Statistically significant values with a rejection threshold of  $p = 0.05$  are typed in **bold**.

### 5.2 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 7 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. 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 scores in the 90-day prediction model. This is because of the 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, hence contributing to a reduction in the 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 7: Area under precision-recall curve and ROC values for recovery and relapse prediction at 1 year for PSL-Recovery trained on 90-day data and 1-year data.**

### 5.3 Analysis of Linguistic, Psycho-linguistic, and Structural Feature Groups

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 analyze the weights learned by our model for the different rule sets in Table 5. Then, we construct different versions of our models by leaving out different sets of features/rules and analyze the corresponding effect on prediction performance.

*5.3.1 Analysis of Learned Weights.* The weights of the PSL-Recovery 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 in Section 5.3.2.

*5.3.2 Variations of PSL-Recovery Model.* To empirically evaluate the predictive ability of various feature groups in predicting recovery, we construct variations of the *PSL-Recovery* model by leaving out features, groups of features, and dependencies among them. To distinguish the *PSL-Recovery* model from other variants, we refer to this model as *PSL-Recovery-All* in below experiments. Figure 4 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-Recovery* model below.

*PSL-Recovery-All.* This model uses all the rules shown in the Table 5. The performance scores from this model are captured in the far left in Figure 4.

*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 5. Note that the PSL-Linguistic model does not use any linguistic features on friends’ tweets, relational, or psycho-linguistic features.

---

## Variations of PSL-Relational Model

---

### PSL-Relational

**Set D. Combining structural and linguistic on friends' tweets (excluding psycho-linguistic features from LIWC):**

$$\begin{aligned} & \text{friends}(U1, U2) \wedge \text{usesAlcoholWord}(U2, aw) \rightarrow \neg \text{recovers}(U1) \\ & \text{friends}(U1, U2) \wedge \text{usesSoberWord}(U2, sw) \rightarrow \text{recovers}(U1) \end{aligned}$$

---

### PSL-Topic

**Set C. Combining structural and topic features without sentiment:**

$$\begin{aligned} & \text{replies}(U2, U1, T) \wedge \text{friendTweetTopic}(U2, T, alcohol) \wedge \text{attendsAA}(U1) \rightarrow \neg \text{recovers}(U1) \\ & \text{retweets}(U2, U1, T) \wedge \text{friendTweetTopic}(U2, T, alcohol) \wedge \text{attendsAA}(U1) \rightarrow \neg \text{recovers}(U1) \\ & \text{retweets}(U2, U1, T) \wedge \text{friendTweetTopic}(U2, T, sober) \wedge \text{attendsAA}(U1) \rightarrow \text{recovers}(U1) \\ & \text{replies}(U2, U1, T) \wedge \text{friendTweetTopic}(U2, T, sober) \wedge \text{attendsAA}(U1) \rightarrow \text{recovers}(U1) \end{aligned}$$

**Set D. Combining structural and linguistic on friends' tweets (excluding psycho-linguistic features from LIWC):**

$$\begin{aligned} & \text{friends}(U1, U2) \wedge \text{usesAlcoholWord}(U2, aw) \rightarrow \neg \text{recovers}(U1) \\ & \text{friends}(U1, U2) \wedge \text{usesSoberWord}(U2, sw) \rightarrow \text{recovers}(U1) \end{aligned}$$

---

### PSL-Psychological (Affect)

**Set D. Combining structural features and linguistic on friends' tweets with affect:**

$$\begin{aligned} & \text{affect}(U1) \wedge \text{friends}(U1, U2) \wedge \text{usesAlcoholWord}(U2, aw) \rightarrow \neg \text{recovers}(U1) \\ & \text{affect}(U1) \wedge \text{friends}(U1, U2) \wedge \text{usesSoberWord}(U2, sw) \rightarrow \text{recovers}(U1) \end{aligned}$$

---

### PSL-Psychological (Social)

**Set D. Combining structural features and linguistic on friends' tweets with social:**

$$\begin{aligned} & \text{social}(U1) \wedge \text{friends}(U1, U2) \wedge \text{usesAlcoholWord}(U2, aw) \rightarrow \neg \text{recovers}(U1) \\ & \text{social}(U1) \wedge \text{friends}(U1, U2) \wedge \text{usesSoberWord}(U2, sw) \rightarrow \text{recovers}(U1) \end{aligned}$$

---

### PSL-Sentiment

**Set C. Combining structural and topic features with sentiment:**

$$\begin{aligned} & \text{replies}(U2, U1, T) \wedge \text{friendTweetTopic}(U2, T, alcohol) \wedge \text{positiveSentiment}(U2, T) \wedge \text{attendsAA}(U1) \rightarrow \neg \text{recovers}(U1) \\ & \text{retweets}(U2, U1, T) \wedge \text{friendTweetTopic}(U2, T, alcohol) \wedge \text{positiveSentiment}(U2, T) \wedge \text{attendsAA}(U1) \rightarrow \neg \text{recovers}(U1) \\ & \text{retweets}(U2, U1, T) \wedge \text{friendTweetTopic}(U2, T, sober) \wedge \text{positiveSentiment}(U2, T) \wedge \text{attendsAA}(U1) \rightarrow \text{recovers}(U1) \\ & \text{replies}(U2, U1, T) \wedge \text{friendTweetTopic}(U2, T, sober) \wedge \text{positiveSentiment}(U2, T) \wedge \text{attendsAA}(U1) \rightarrow \text{recovers}(U1) \end{aligned}$$

---

### PSL-LIWCSimilarity

**Set E. Collective Rules with LIWCSimilarity:**

$$\begin{aligned} & \text{LIWCSimilarity}(U1, U2) \wedge \text{recovers}(U2) \rightarrow \text{recovers}(U1) \\ & \text{LIWCSimilarity}(U1, U2) \wedge \neg \text{recovers}(U2) \rightarrow \neg \text{recovers}(U1) \\ & \text{LIWCSimilarity}(U1, U2) \wedge \text{recovers}(U1) \rightarrow \text{recovers}(U2) \\ & \text{LIWCSimilarity}(U1, U2) \wedge \neg \text{recovers}(U1) \rightarrow \neg \text{recovers}(U2) \end{aligned}$$

---

**Table 8: Table showing the additional rules in different variations of PSL-Relational 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 4. 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 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 8 gives the different variations and the changes to the different rule sets in the *PSL-Recovery* model in Table 5. 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

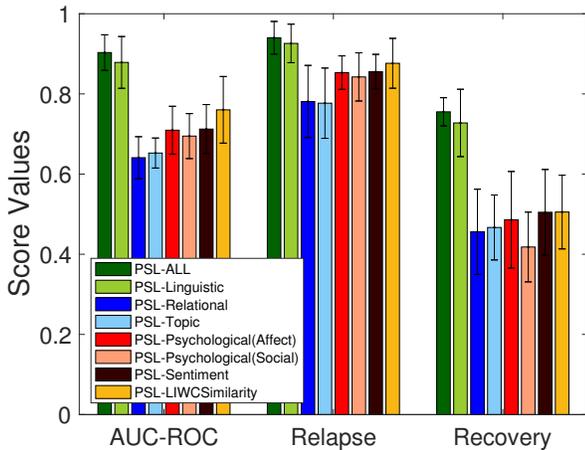


Figure 4: Graph showing area under precision-recall curve and ROC values for recovery and relapse prediction for the different variations of PSL-Recovery model

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 5 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 8.

*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 8 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.

#### 5.4 Linguistic Analysis of Relapsed User Tweets

Analyzing tweets of relapsed users, we uncover tweets on reasons that users mention for drinking and relapse. We capture the most occurring reasons in our data in Table 9. Anxiety and depression are popular reasons with 468 total tweets in relapsed users using anxiety words or the hashtags #anxiety, #depression, and #sadness. The second most popular reason that we observe in our dataset is stress, especially work or school-related stress. We see 318 tweets in this category. Sleep related disorders also have a high incidence in the data, as sleep issues are related to stress, anxiety, and many other psychological ailments [26].

Reason	Example Tweets
Anxiety	I want a delete button in my <b>life</b> . To delete some people, some <b>memories</b> and some feelings. Can’t sleep. <b>Anxiety</b> and <b>sadness</b> . I really need to see a doctor. <b>#anxiety #depression #sleeplessnights</b> I’ve been <b>depressed</b> and skipping AA, so Sponsor is worried I’m going to relapse. I’m on the <b>depression</b> swing of my <b>bipolar</b> cycle. I haven’t thought about drinking, so thats good news.
Stress	I’m so <b>stressed</b> out with <b>work</b> right now. okay ive had a long enough break from <b>work</b> :( <b>Work</b> is so dead. So tired and so bored. I thought life was holding some boring <b>job</b> , no money, being bored and doing nothing else.

Table 9: Example tweets that mention reasons for consuming alcohol

## 6 CONCLUSION

In this paper, we presented a structured prediction approach that incorporates linguistic and psycholinguistic features from users’ and friends’ tweets and structural interactions with friends and effectively encodes dependencies among them to model and understand recovery and relapse from/into AUD. Our extensive feature analysis on the different groups of features helps in understanding their respective individual ability in predicting recovery and helps in extending our models to similar prediction problems and other data where certain groups of features are unavailable. There are several exciting directions to go from here. Modeling fine-grained linguistic signals will be helpful in understanding the reasons behind AA users getting succumbed to AUD and effectively averting relapses before they occur. These signals will also help in identifying users who are potentially most receptive to interventions, thus helping channel limited resources to help users who will benefit the most.

**Acknowledgements.** 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] American Psychological Association. 2012. Recovery principles. <http://www.apa.org/monitor/2012/01/recovery-principles.aspx>. (2012).
- [2] 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.
- [3] Sairam Balani and Munmun De Choudhury. 2015. Detecting and Characterizing Mental Health Related Self-Disclosure in Social Media. In *Proceedings of the Annual ACM Conference on Human Factors in Computing Systems*.
- [4] Steven Bird. 2006. NLTK: the natural language toolkit. In *Proceedings of the COLING/ACL on Interactive presentation sessions*.
- [5] 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)*.
- [6] 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.
- [7] 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)*.
- [8] 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)*.
- [9] M. Dredze. 2012. How Social Media Will Change Public Health. *Journal of IEEE Intelligent Systems* 27 (2012), 81–84.
- [10] Marica Ferri, Laura Amato, and Marina Davoli. 2006. Alcoholics Anonymous and other 12-step programmes for alcohol dependence. *The Cochrane Library* (2006).
- [11] Christian Greiner, Anne Chatton, and Yasser Khazaal. 2017. Online self-help forums on cannabis: A content assessment. *Journal of Patient Education and Counseling* (2017).
- [12] 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)*.
- [13] 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>. (2010).
- [14] Nabil Hossain, Tianran Hu, Roghaye 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)*.
- [15] 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*.
- [16] George M Hunt and Nathan H Azrin. 1973. A community-reinforcement approach to alcoholism. *Journal of Behaviour Research and Therapy* 11 (1973), 91–104.
- [17] 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.
- [18] 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)*.
- [19] 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.
- [20] 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 International ACM Conference on Computers and Accessibility (SIGACCESS)*.
- [21] 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: WWW and Population Health Intelligence*.
- [22] 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*.
- [23] 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 *In proceedings of the ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW)*.
- [24] 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.
- [25] 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*.
- [26] 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).
- [27] Yelena Mejova, Ingmar Weber, and Michael W Macy. 2015. *Twitter: a digital socioscope*. Cambridge University Press.
- [28] 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.
- [29] 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.
- [30] World Health Organization and World Health Organization. Management of Substance Abuse Unit. 2014. *Global status report on alcohol and health, 2014*. World Health Organization.
- [31] 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)*.
- [32] Michael J. Paul and Mark Dredze. 2014. Discovering Health Topics in Social Media Using Topic Models. *Journal of PLOS ONE* 9 (2014), e103408.
- [33] 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.
- [34] Julie Repper and Rachel Perkins. 2003. *Social inclusion and recovery: A model for mental health practice*. Elsevier Health Sciences.
- [35] 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*.
- [36] Bruce Simons-Morton. 2007. Social influences on adolescent substance use. *American Journal of Health Behavior* 31 (2007), 672–684.
- [37] 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*.
- [38] 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.
- [39] 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 ACM conference on Computer Supported Cooperative Work (CSCW)*.
- [40] Marsha White and Steve M Dorman. 2001. Receiving social support online: implications for health education. *Journal of Health Education Research* 16 (2001), 693–707.
- [41] 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.
- [42] 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.
- [43] 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).