Does Social Support Expressed in Post Titles Elicit Comments in Online Substance Use Recovery Forums?

Anietie Andy[1], Sharath Chandra Guntuku[2]
[1]Penn Medicine [2]Computer and Information Science
University of Pennsylvania
{anietie.andy@pennmedicine,sharathg@cis}.upenn.edu

Abstract

Individuals recovering from substance use often seek social support (emotional and informational) on online recovery forums, where they can both write and comment on posts, expressing their struggles and successes. A common challenge in these forums is that certain posts (some of which may be support seeking) receive no comments. In this work, we use data from two Reddit substance recovery forums: */r/Leaves* and */r/OpiatesRecovery*, to determine the relationship between the social supports expressed in the titles of posts and the number of comments they receive. We show that the types of social support expressed in post titles that elicit comments vary from one substance use recovery forum to the other.

1 Introduction

In the United States (US), substance use disorder (SUD) is one of the main causes of premature death (Johnston et al., 2003; Schulte and Hser, 2013). Social media is often used by people with SUD to seek support (MacLean et al., 2015). Online substance use recovery forums such as the Reddit forum, */r/Leaves* - which focuses on discussions around quitting marijuana, provide a "safe space" where members - some of whom may be struggling with substance use, can freely communicate and seek help from other members.

Prior work showed that 10% of support-seeking messages on a forum focused on discussions around cancer received no comments (Wang et al., 2015; Yang et al., 2019a). We find that support-seeking posts (across a 3 months time period) from */r/Leaves* follow a similar trend with 11% of posts receiving no comments, thereby leaving some users seeking support without adequate support.

The following social supports: emotional and informational are crucial in online forums focused on discussions around health and well-being (Wang

et al., 2012; Yang et al., 2017), where similar to prior work (Wang et al., 2012), emotional support sought in posts seek affirmation, encouragement, and compassion, while informational support sought in posts seek information or advice.

Reddit posts are made up of two sections: the *title* - which briefly describes the post, and the *selftext* - which describes the post in more detail. Glenski and Weninger (2017) demonstrated that on Reddit, user interactions such as voting and commenting on posts is guided by the title of the post (post-titles); thereby implying that readers utilize linguistic cues in the post-titles on Reddit to decide whether to respond to posts. Hence, in this work, we focus on analyzing the post-titles in two Reddit substance use recovery forums */r/Leaves* and */r/OpiatesRecovery*. We measure the extent to which these post-titles seek emotional support and informational support and analyze their relationship with the number of comments they receive. We hypothesize that since the interests of users who belong to different forums focused on similar discussions differs Tran and Ostendorf (2016), the type of social support (expressed in post-titles) that elicits comments may differ from one online substance use recovery forum to the other.

Understanding the relationship between the social supports (emotional and informational) sought in post-titles published in substance use recovery forums and the number of comments these posts receive is important partly because users join online health forums when going through a health care event such as cancer (Wen and Rosé, 2012; Yang et al., 2019a) or recovering from substance use (MacLean et al., 2015); therefore, posts not receiving comments in these forums means that some users are not getting the necessary support they need.

2 Related Work

Some prior work focused on non-health related forums while others focused on health related forums.

In non-health related forums, Althoff et al. (2014) studied a Reddit forum to determine the social and linguistic factors associated with posts in the forum that elicit responses. Tran and Ostendorf (2016) explored several Reddit forums and determined that different forums use different language styles and there was a correlation between a forums language style and the responses to comments. Jaech et al. (2015) examined the effect of language use in online forums on the reaction of members of the forums to comments. Hessel et al. (2017) studied if incorporating multimodal features attracted user attention (Horne et al., 2017).

In health related forums, Wang et al. (2012) examined the effect of different social supports users are exposed to in an online cancer forum and its effect on the duration of user membership in the forum. Wang et al. (2015) demonstrated that in an online cancer forum, members tend to respond with emotional support when users self-disclose negative information about themselves. Yang et al. (2019b) studied communication in an online cancer forum and determined that members of the forum tend to disclose more negative information about themselves in public domains provided by the forums compared to the forums private domains. Yang et al. (2017) examined the relationship between the kind of communication received by members of an online cancer forum and their commitment to the forum. Yang et al. (2019a) determined that over time, members of an online cancer forum change roles and that certain roles are predictors of prolonged commitment to the group. Chancellor et al. (2018) examined online weight loss forums and determined how support influences user behavior changes. MacLean et al. (2015) analyzed different phases of opioid addiction in an online forum focused on discussions around opioid use recovery.

Our work is different from prior work in that we analyze posts published in online substance use recovery forums with the aim to determine if and how emotional and informational support sought in published posts in substance use recovery forums elicit comments from members of the forums. This study received exempt status from the University of Pennsylvania Institutional Review Board.

3 Dataset

Our dataset consists of posts and meta-data from two active Reddit substance use recovery forums, */r/Leaves* and */r/OpiatesRecovery*, which have 147,000 members and 27,000 members, respectively, as of September 2020. */r/Leaves* is self-described as *"a support and recovery community for practical discussions about how to quit pot, weed, cannabis, edibles, BHO, shatter, or whatever THC-related product you're using, and support in staying stopped"*. */r/OpiatesRecovery* is self described as *"We are a group of people dedicated to helping each other kick the habit"*. We chose */r/Leaves* and */r/OpiatesRecovery* because these forums are the Reddit substance use recovery forums focused on marijuana and opioid use, respectively, with the most number of users.

Typically in these forums, a member writes a post (*title* and *selftext*) and other members respond to the post by either voting the post *up* or *down* or writing a comment. In this paper, we focus only on the comment responses to posts. Using Google's BigQuery[1], which is a data warehouse containing data from Reddit, we collected and processed all posts published in */r/Leaves* and */r/OpiatesRecovery* between December 2015 and August 2019 and collected the following data related to each post: the post-title, the user who published the post, the time the post was created, the comments the post received, and the number of comments the post received. Table 1 highlights the summary of our dataset.

Attribute	*/r/Leaves*	*/r/OpiatesRecovery*
Number of unique users	18,100	4,374
Number of posts	35,961	9,900
Number of comments	227,850	129,801

Table 1: Summary of the */r/Leaves* and */r/OpiatesRecovery* datasets

4 Social Support

Several studies have shown the importance of the expression of emotional and informational social supports in forums focused on discussions around health (Wang et al., 2012; Yang et al., 2017, 2019a). Using a similar method from previous work by Wang et al. (2012), we built two models to determine how much emotional and informational support, respectively, is sought in post-titles in our dataset. We had 3 annotators - who were health

---

[1]https://cloud.google.com/bigquery/

care professionals with graduate degrees and familiar with substance use recovery, to rate a sample of 1,000 post-titles from our dataset, on (i) how much informational support each post-title sought and (ii) how much emotional support each post-title sought; where informational support post-titles provide/seek advice or information and emotional support post-titles seek encouragement, understanding, or affirmation (Wang et al., 2012). Similar to prior work, (Wang et al., 2012), the annotators rated these post-titles using a 7-point Likert scale (1 meant "social support was not expressed in a post-title" and 7 meant "social support was expressed a lot in a post-title"). To measure the reliability of the annotators, we used intra-class correlation (ICC) (Bartko, 1966), which measures annotator reliability when each post-title is rated by different groups of annotators; the ICC for informational support sought and emotional support sought were 0.95 and 0.93, respectively. For each post-title, the annotator ratings were averaged, hence each post-title had a score that ranged between 1 and 7 which indicated how much informational and emotional support was sought.

4.1 Features

We extracted several language features from the annotated post-titles.

Using Linguistic Inquiry and Word Count (LIWC) (Pennebaker et al., 2015) - a dictionary comprising different psycho-linguistic categories, we selected the following LIWC categories relevant to informational and emotional support (Wang et al., 2012): "positive emotion", "negative emotion", "she/he", "you", "we", "i",impersonal pronoun", auxiliary" "verb", "past", "present", "future","religion", "death" "they", "cognitive mechanism", "biological processes", "time".

Similar to Wang et al. (2012), from each post-title, we extract the number of sentences, the number of words in each sentence, the number of sentences that contain negation words/phrases such as "not", and the number of sentences phrased as questions.

We also extracted the number of specific parts-of-speech and the number of strong subjectivity words such as "affirmation" and weak subjectivity words such as "abandoned" (Wilson et al., 2005; Wang et al., 2012).

Post-titles seeking advice or involving requests were identified and counted (Wang et al., 2012) by

(i) selecting sentences that began with the word "you" and followed by a Modal verb such as "may" and (ii) selecting sentences that began with the word "please" and followed by a verb.

We collected names of medicines from the Food and Drug Administration website website [2]; also a comprehensive list of nicknames for drugs was compiled (Wang et al., 2012); we counted the number of drug names in each post-title.

We trained a model of 20 Latent Dirichlet Allocation (LDA) (Blei et al., 2003) topics from 45,000 randomly selected post-titles from our dataset. Two physicians, who are familiar with substance use recovery, manually assigned a label to each of the topics, as shown in Table 2.

4.2 Social Support Prediction Model

We built two Random Forest models that each output a numerical value that indicates how much emotional support and informational support is expressed in a post-title based on the annotations. We experimented with SVM, logistic regression, and Random Forest; Random Forest performed better. We randomly partitioned the annotated posts into a training set (80%), a validation set (10%), and a test set (10%). We used the validation set to evaluate the performance of the models; when the performance on the validation set was satisfactory, the models were then evaluated on the test set. Similar to Wang et al. (2012), Pearson's correlation was used to measure the correlation of the models with the annotated data.

These models correlated with the average annotator ratings with Pearson's correlation r = 0.46 and r = 0.51 for seeking emotional support and informational support, respectively. We then applied these models to the posts in our dataset. Table 3 shows the 5 most important features, as ranked by the Random forest model.

5 Does social support expressed in post-titles elicit comments?

A challenge in substance use recovery forums is that several support-seeking posts do not receive any comments, thereby leaving some posters potentially without adequate support. In this section, we aim to determine if emotional and informational support expressed in post-titles elicit comments in */r/Leaves* and */r/OpiatesRecovery*. For our analysis, we select posts by users with 5 or more published

---

[2]https://www.fda.gov/Drugs/

Table 2: Summary of LDA topic themes and top 5 highly correlated words associated with each topic

LDA topic themes	Highly correlated words
Time sober	months, weeks, sober, clean, year
Wanting reasons to stop	stop, can't, high, anymore, friends
Feelings: anxious and depressed	feeling, depressed, tired, high, anxious
Time sober before relapse	days, sober, clean, hours, free
Withdrawal symptoms and addiction	withdrawal, drug, test,addiction,job,
Cravings and relapse	night, strong, cravings, weekend, relapsed
Advice for quitting	quitting, advice, tips,support, benefits
Ready to quit	time, quit, finally, hard, stop
Making the decision to quit	made, make, life, things, friends
Feelings of quitting	anxiety, quitting, depression, deal,pains
mood and feelings	good, today, bad, high,start
Years of using	years, daily, heavy, user, habit
Reason for quitting	quitting, brain, back, motivation, fog
Impact of addiction	life, addiction, love, hate, relationship
Quitting	quit, ago, months,stop, haven't
Side effects	dreams, sleep, night, nightmares, insomnia
Struggling with relapse	day, end, relapse, struggling, thoughts
Time to start sobriety	day, today, start,journey, sobriety
Thinking of quitting	cold, turkey, thinking, cut, habit
Sharing stories	story, thought, wanted, share, addiction

Emotional Support	Informational Support
Thinking of quitting- (0.76)	Share stories - (0.69)
Word length - (0.033)	Thinking of quitting-(0.046)
Sharing stories - (0.032)	Word length - (0.044)
Noun - (0.021)	Noun (0.035)
Strong subjectivity - (0.016)	Strong subjectivity - (0.028)

Table 3: Top 5 most important features as ranked by the Random forest model

posts in our dataset i.e. 12,960 posts published by 1,285 users from */r/Leaves* and 4,055 posts published by 335 users from */r/OpiatesRecovery*. For each user, we calculate the average number of comments they received for all their posts. We also calculated the average extent of emotional and informational support scores for posts published by each user. We correlate the mean extent of emotional and informational support, respectively with the mean number of comments received per user.

Feature	Pearson r
Emotional Support Sought	0.17
Informational Support Sought	0.22

Table 4: Correlation between social support sought and number of comments: */r/Leaves*. $p < 0.001$. Number of users = 1,285

Feature	Pearson r
Emotional Support Sought	0.15
Informational Support Sought	- 0.13

Table 5: Correlation between social support sought and number of comments: */r/OpiatesRecovery*. $p < 0.001$. Number of users = 335

Results and Discussion: Tables 4 and 5 show the correlation between social support sought in post-titles and number of comments in */r/Leaves* and */r/OpiatesRecovery*, respectively. We observed that in */r/Leaves*, the average informational support sought by users in post-titles correlates more with the average number of comments received by these users compared to the average emotional support sought. Also, we observed that in */r/OpiatesRecovery*, the average emotional support sought had a positive correlation with the average number of comments received compared the average informational support sought which had a negative correlation with the average number of comments received.

These findings can benefit members of substance use recovery forums; for example, users seeking to elicit comments to their posts published on */r/Leaves* may use linguistic cues associated with higher informational support in their post-titles. Also, the findings from this work can benefit substance use recovery forum moderators; given that there is a negative correlation between the informational support sought in post-titles in */r/OpiatesRecovery*, this potentially means that some users seeking informational support in this forum are not receiving support. Hence moderators of the forum can come up with ways in which these informational support seeking posts receive comments; for example informational support seeking posts not receiving comments may be sent to moderators or users familiar with the support sought.

Limitations and Future Work:

In our analysis, we focused on two subreddits - */r/Leaves* and */r/OpiatesRecovery*, which are the fo-

rums focused on recovery from marijuana use and opiod use with the most number of members on Reddit. In the future, we would explore the relationship between social support expressed in posts and the responses (comments and votes) they receive, in other substance use recovery forums. While the results in this work indicate statistically significant correlations, in the future, we would look at the affect of author tenure and Reddit karma (reputation) - all of which could potentially contribute to the response rates of posts.

6 Conclusion

We built two models to measure the extent of informational and emotional social support expressed in post-titles in two substance use recovery. We used these models to show the social supports that elicit comments in these forums.

References

Tim Althoff, Cristian Danescu-Niculescu-Mizil, and Dan Jurafsky. 2014. How to ask for a favor: A case study on the success of altruistic requests. In *Eighth International AAAI Conference on Weblogs and Social Media*.

John J Bartko. 1966. The intraclass correlation coefficient as a measure of reliability. *Psychological reports*, 19(1):3–11.

David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.

Stevie Chancellor, Andrea Hu, and Munmun De Choudhury. 2018. Norms matter: contrasting social support around behavior change in online weight loss communities. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 666. ACM.

Maria Glenski and Tim Weninger. 2017. Predicting user-interactions on reddit. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, pages 609–612.

Jack Hessel, Lillian Lee, and David Mimno. 2017. Cats and captions vs. creators and the clock: Comparing multimodal content to context in predicting relative popularity. In *Proceedings of the 26th International Conference on World Wide Web*, pages 927–936. International World Wide Web Conferences Steering Committee.

Benjamin D Horne, Sibel Adali, and Sujoy Sikdar. 2017. Identifying the social signals that drive online discussions: A case study of reddit communities. In *2017 26th International Conference on Computer Communication and Networks (ICCCN)*, pages 1–9. IEEE.

Aaron Jaech, Victoria Zayats, Hao Fang, Mari Ostendorf, and Hannaneh Hajishirzi. 2015. Talking to the crowd: What do people react to in online discussions? *arXiv preprint arXiv:1507.02205*.

Lloyd D Johnston, Patrick M O'Malley, and Jerald G Bachman. 2003. Monitoring the future: National results on adolescent drug use: Overview of key findings. *Focus*, 1(2):213–234.

Diana MacLean, Sonal Gupta, Anna Lembke, Christopher Manning, and Jeffrey Heer. 2015. Forum77: An analysis of an online health forum dedicated to addiction recovery. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 1511–1526. ACM.

James W Pennebaker, Ryan L Boyd, Kayla Jordan, and Kate Blackburn. 2015. The development and psychometric properties of liwc2015. Technical report.

Marya T Schulte and Yih-Ing Hser. 2013. Substance use and associated health conditions throughout the lifespan. *Public Health Reviews*, 35(2):3.

Trang Tran and Mari Ostendorf. 2016. Characterizing the language of online communities and its relation to community reception. *arXiv preprint arXiv:1609.04779*.

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 2012 conference on Computer Supported Cooperative Work*, pages 833–842. ACM.

Yi-Chia Wang, Robert E Kraut, and John M Levine. 2015. Eliciting and receiving online support: using computer-aided content analysis to examine the dynamics of online social support. *Journal of medical Internet research*, 17(4):e99.

Miaomiao Wen and Carolyn Penstein Rosé. 2012. Understanding participant behavior trajectories in online health support groups using automatic extraction methods. In *Proceedings of the 17th ACM international conference on Supporting group work*, pages 179–188. ACM.

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*.

Diyi Yang, Robert Kraut, and John M Levine. 2017. Commitment of newcomers and old-timers to online health support communities. In *Proceedings of the 2017 CHI conference on human factors in computing systems*, pages 6363–6375. ACM.

Diyi Yang, Robert E Kraut, Tenbroeck Smith, Elijah Mayfield, and Dan Jurafsky. 2019a. Seekers, providers, welcomers, and storytellers: Modeling social roles in online health communities. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, page 344. ACM.

Diyi Yang, Zheng Yao, Joseph Seering, and Robert Kraut. 2019b. The channel matters: Self-disclosure, reciprocity and social support in online cancer support groups.