I’m always in so much pain and no one will understand" - Detecting Patterns in Suicidal Ideation on Reddit

Michal Monselise
Drexel University
Philadelphia, PA, USA
michal.monselise@drexel.edu

Christopher C. Yang
Drexel University
Philadelphia, PA, USA
chris.yang@drexel.edu

ABSTRACT
Social media has become another venue for those struggling with thoughts of suicide. Many turn to social media to express suicidal ideation and look for peer support. In our study we seek to better understand patterns in the behaviors of these users particularly on the social media platform Reddit. This study will explore how Reddit users move or progress between subreddits until they express active suicidal ideation. We also look at these users’ posting pattern in the time leading up to expressing suicidal ideation and the time after. We examined a large dataset of posts from users who created at least one thread on SuicideWatch during January 2019 - August 2019 and collected their posts starting in July 2018 to create a look back period of 6 months. This generated a total of 5,892,310 posts. We defined what it means to progress between subreddits and generated a graph of progressions of all users in our dataset. We found that these users mostly progressed to or from 8 different subreddits and each of these subreddits could point to a particular emotional difficulty that a user was having such as self harm or relationship problems. Furthermore, we examined the volume of posts and the proportion of posts with negative sentiment leading up to the first incident of active suicidal ideation and found that there is an increase in both negative sentiment and volume of posts leading up to the day of the first incident of suicidal ideation on Reddit. However, during the day of first incident of suicidal ideation, there is a precipitous drop in the number of posts which goes back up on the following day. Using this insight, we can better understand these users. This will allow for developing intervention for suicide prevention in social media platforms in the future.

CCS CONCEPTS
• Applied computing → Health informatics.

KEYWORDS
social media, mental health, health informatics

1 INTRODUCTION
According to the WHO, more than 700,000 people die each year by suicide [21]. Suicide is the fourth leading cause of death in young people aged 15–29 years in 2019 according to the WHO and is ranked 12 in the list of leading causes of death in the US in 2019 overall according to the CDC [2]. With the increase in popularity of social media, many individuals turn to social media looking for peer support with mental health condition such as depression and suicide [19]. This presents an opportunity for early detection or intervention on social media platforms when this behavior is observed. The Columbia Suicide Severity Rating Scale (C-SSRS) is a scale developed for the classification of suicidal behaviors [13]. According to the C-SSRS, expressing thoughts of suicide is called suicidal ideation and is considered a less severe behavior than a suicide attempt [1]; however, it is still a concerning behavior that can potentially escalate to a suicide attempt. Furthermore, suicidal ideation can start at the passive stage with abstract thoughts of ending one’s life and move to active suicidal ideation when an individual has more concrete thoughts of suicide and perhaps even a plan to act on these thoughts[4]. This study attempts to identify the signs of suicidal ideation before they occur by analyzing users who have made such posts on social media, and particularly Reddit. Reddit is a social media platform that contains millions of subreddits which are message boards that center around a certain theme or subject. Reddit has subreddits in many areas, and in this study, we will focus on subreddits for mental health support and particularly SuicideWatch. SuicideWatch is a subreddit dedicated to individuals who express thoughts of suicidal ideation and look for peer support. Our study will focus on the users of SuicideWatch and evaluate two hypotheses:

(1) Do users who express active suicidal ideation on SuicideWatch start somewhere else on Reddit and progress to SuicideWatch over time?

(2) Do users who express suicidal ideation on SuicideWatch have unique patterns in their posts that can be detected in a quantitative analysis and allow us to predict their behavior before they express suicidal ideation?

2 LITERATURE REVIEW
In this review, we will discuss research on suicidal ideation on Reddit and on social media in general. We will also discuss predictive models for suicidal ideation.
2.1 Suicidal Ideation on Social Media

There are a number of surveys of literature on suicide prevention on social media. One survey is performed by Robinson et al. [19] and examines 4 types of suicide prevention studies using social media: Studies that described resources on social media aimed at suicide prevention, studies identifying the potential of social media to detect users at risk of suicide, studies that aim to identify or distinguish between users at risk of suicide, and studies that examine how those at risk of suicide receive support online. Another survey was written by Lopez-Castroman and focuses on projects that help prevent suicidal behavior on social media [15]. Coppersmith et al. performed an exploratory analysis to compare people who have made a suicide attempt to those who have not and found a few differences [9]. They have identified that those who have made a suicide attempt tend to have more emotional content in their posts. They also noted that around the time of a suicide attempt, those who attempt to take their life will have a higher volume of social media posts and that volume will go down as we move away from the event. This study only makes some observations based on the temporal nature of the data. Furthermore, Low et al. examine Reddit posts from the subreddit SuicideWatch and examine whether more or less responses to posts of suicidal ideation will result in more suicidal ideation in the future [16].

2.2 Suicidal Ideation Predictive Models

There are a number of studies that use predictive modeling for predicting suicidal ideation using social media data. In a critical review of predictive models for mental health on social media, Chancellor and De Choudhury find a variety of machine learning techniques used for predicting suicidal ideation [7]. Their review identified 22 studies dedicated to the prediction of suicidal ideation on social media. Since their review, new research using more advanced deep learning has been released. One of these is the study by Sawhney et al. which uses a transformer based model to detect suicidal ideation on social media. The researchers take into consideration the temporal dependency between a user’s posts to make their predictions [20]. This study uses the CLPsych 2019 shared task dataset [22]. The study by Mishra et al. examined a user’s profile and posts and created embeddings of different user features to predict suicidal ideation [17]. Coppersmith et al. used a dataset extracted from OurDataHelps.org that combines different social media platforms like Facebook, Reddit, and Twitter [8]. They applied natural language processing models to this data to identify posts containing suicidal ideation. The key differences between this study and other studies cited is that these studies aim to predict the labels of posts using the content of the post. However, this study is looking at the user’s behavior in the time leading up to the suicidal ideation post and extracting features from the temporal patterns. Additionally, this study will use techniques that will require less computing resources than deep learning and NLP models. Finally, De Choudhury et al. studied suicidal ideation with a similar notion of progression from other mental health related subreddits to SuicideWatch [10]. This study examined linguistic features of posts such as post structure and unigram and bigram frequencies, as well as variables like interpersonal awareness. In contrast, our study seeks to quantitatively identify which are the subreddits that users transition from, and we also sought to identify quantitative patterns of behavior rather than a predictive model that is partially based on linguistic patterns.

3 DATA AND METHODS

3.1 Data

This study utilizes data from the Pushshift Reddit dataset [5]. This dataset allows faster retrieval of older Reddit posts, however, this dataset is static and does not update frequently. This dataset also does not have all the fields that can be found in the Reddit API. We chose this dataset since we are not interested in using the dynamic aspects of Reddit data such as upvotes and downvotes of posts and we also don’t need all fields of the Reddit API. Therefore, this dataset is adequate for our analysis. We identified all users who created a thread in SuicideWatch at least once between January 2019 and August 2019 to create a list of users who have expressed suicidal ideation. We then extracted all of these users’ post history during that time period as well as 6 months back for a total of all posts between July 2018 and August 2019. This ensured that we had a 6 month look back period for these users. We chose to identify all users with at least one thread as users who presented suicidal ideation because we assumed that users who make a standalone thread do so to express suicidal ideation as this is in accordance with the guidelines of the subreddit. The guidelines state that users are directed to seek support for themselves primarily and are directed to other resources for support for someone else of for bereavement of someone close lost to suicide [3]. Additionally, the subreddit is moderated and typically irrelevant posts are removed. While it is possible that a user may have expressed suicidal ideation somewhere else on the internet, we chose to focus on those who did so on SuicideWatch because it removes the complications that may arise from trying to deduce this using human annotators or natural language processing. We processed the data by first removing users with only one post overall from the dataset as well as removing all empty posts. This resulted in 5,892,310 posts and 26,493 distinct users. All duplicates were removed; however, many users on Reddit regularly make more than one account to increase anonymity. This cannot be accounted for in the data. Our data contains three types of posts:

1. **Threads** are standalone posts where a user prompts other users to post in response to their statement. In the context of SuicideWatch, this is typically how a user expresses suicidal ideation.

2. **Comments** are posts that respond directly to a thread. They are the first level of nesting of posts.

3. **Responses** are posts that relate to comments or to other responses. They are nested one level below comments and can be further nested and respond to other responses.

We consider the opening post of the thread to be the first expression of suicidal ideation. Some examples of opening posts of threads in SuicideWatch are detailed below:

“I feel like i want to kill myself in public idk whether its a sickly need for attention or revenge on someone. I feel compelled to blow my brains out in my high school cafeteria in front of the others”

“Even with people around my im lonely, im miserable. My only wish for 2019 is to finally have the courage. Nothing means anything, everything is worthless.”
3.2 Defining Progression

In our first hypothesis, we examine the progression of users through different subreddits. Progression through stages as a method for studying change in behavior is a well documented phenomenon in psychology including the transtheoretical model of behavior change (TTM) [18]. In their study, Prochaska et al. present a framework to identify behavior stages using social media [14]. They examined behavior stages for conditions like drinking alcohol and used their data to identify behavior stages. In our study, we look at progression of suicidal behavior from passive suicidal ideation to active suicidal ideation. We ask whether users who expressed active suicidal ideation on SuicideWatch started in SuicideWatch or whether they progressed in severity by starting at another subreddit such as depression and then moved on to SuicideWatch. To perform this analysis, we will assume that given two subreddits A and B, progressing from subreddit A to subreddit B means that a user initially posted on subreddit A and then posted in subreddit B within a window of posts of size n. In other words, we define our user as having progressed from A to B. For example, for a window of 5, if a user posted in the depression subreddit, and then made 3 posts on the memes subreddit and then posted on the SuicideWatch subreddit, we consider our user to have progressed from the depression subreddit to the memes subreddit, from memes to SuicideWatch, and also from depression to SuicideWatch. We created this window to account for situations where a user may have been wrapping up a response to another user from before or posting casually on a common subreddit like memes or AskReddit. We create a window so that we don’t lose that relationship between two subreddits that would be otherwise lost if we only look at sequential progressions. Once a user progresses from A to B, they cannot progress back from B to A. So we consider this to be a directed relationship. Choosing our n requires caution. An n too small might lose important progression relationships while an n that is too large will capture relationships that are not actually directly connected because too much time has passed between the first and last post in the window. For our analysis, we select a window of size n = 20. In our dataset, a lag of 20 is on average approximately one day between posts. Having created a dataset of progressions, we construct a graph with these directed relationships between subreddits. We define our graph \( G = (V, E) \) with weight function \( w : E \rightarrow \mathbb{R} \) where:

1. \( V \) is the set of vertices – in this case the list of all subreddits containing more than 50 posts in the sample.
2. \( E \) is the set of edges connecting two subreddits.
3. \( w \) is the count of users who progressed from one subreddit to another.

3.3 Sentiment Analysis

Our text was preprocessed by removing non letter characters, tokenizing, stemming, and then assigning a sentiment using the VADER (Valence Aware Dictionary and sEntiment Reasoner) Python library [12]. The VADER library applies a rule based model to text for sentiment analysis. VADER is based on existing and well-established sentiment lexicons such as LIWC (Linguistic Inquiry and Word Count). Additionally, VADER is supplemented using lexical features that are common in social media for expressing sentiment. VADER produces a score between -1 and 1 for each piece of text. We classify our text as containing negative sentiment if the score is below -0.05.

4 RESULTS

4.1 Progressing Between Subreddits

Using the relationships between the subreddits in our dataset of progressions, we plot the graph \( G \) in figure 1.

We note that even though we scraped the entire post history of all users who created at least one thread in SuicideWatch between January and August 2019, there are only 8 subreddits that users progressed to or from SuicideWatch: AskReddit, depression, advice, relationship advice, teenagers, memes, self harm, and off my chest. We also note that the weights are about equal for both directions for all but one subreddits connected to SuicideWatch - off my chest. In other words, approximately an equal number of users progressed from one of the 8 subreddits to SuicideWatch as users who progressed from SuicideWatch to a different subreddit. The only exception is off my chest. There are no users who progressed from off my chest to SuicideWatch, only those who progressed from SuicideWatch to off my chest. Therefore, we see no distinct progression for all users, instead, different users start in different subreddits.

4.2 Changes in User Behavior Over Time

To predict whether a user will express suicidal ideation in the future, it is crucial to identify posting behaviors that are unique to users who we identified as making at least one thread in SuicideWatch. We examine changes in posting behavior over time.

4.2.1 Changes in Post Volume and Sentiment Over Time. To evaluate whether there is a change in users’ posting behavior as they approach the first time they express suicidal ideation, we look for that first occurrence and call that day zero. We then aggregate our data to the day level using this adjusted relative date. We computed the mean and the median of number of posts by date relative to the first time a user expressed suicidal ideation on SuicideWatch and plotted them in figures 2 and 3.

We observe an increase in posts leading up to the day of the first suicidal ideation and then a precipitous drop on that day followed by an increase back to the previous peak and then a gradual drop back to baseline. While there are heavy tailed users who inflate the mean number of daily posts, when we examine the median number of daily posts, we see a similar trend. We verify that these observations are significant by performing a t-test and find that there is a significant difference between average daily posts 20 days before and 20 days after the first expression of suicidal ideation.
compared to days 20-50 before and after the first expression of suicidal ideation (95.36 ± 6.76 for 20 days before and after, 67.34 ± 1.99 for days 20-50 before and after, p < 0.001).

4.2.2 Changes in Sentiment Over Time. Using the VADER Python package, we are able to examine the proportion of negative sentiment every day up until the first occurrence of suicidal ideation in our users. This change is depicted in figure 4.

We observe in this graph that there is a slight change of negative sentiment. The proportion of daily negative sentiment goes from about 28% to about 30% as the first occurrence of suicidal ideation happens and then it goes back down to baseline after a few days. This change is not statistically significant (p = 0.875).
5 DISCUSSION

5.1 Progressing Between Subreddits

Our first hypothesis asked whether users progress from one subreddit to another in terms of severity of their symptoms until they eventually express suicidal ideation. While we found that this was not the case for many users, some users do start off on subreddits like depression or self harm and then move on to express suicidal ideation on SuicideWatch. Additionally, we found that users who express suicidal ideation on SuicideWatch primarily post on 8 other subreddits. The subgraph for SuicideWatch is relatively isolated and most of the subreddits connected to SuicideWatch are for some type of support - support for relationship issues, self harm, or depression.

5.2 Changes in User Behavior Over Time

Through our second hypothesis, we tried to identify patterns in the posting behavior of users who have expressed suicidal ideation on Reddit. We find that on average, users increased the number of daily posts leading up to the first expression of suicidal ideation and then had a drop in posts on the actual day of expressing the suicidal ideation. In the following days, the number of posts first went right back up and then slowly went down to baseline. We also observed a similar change in post sentiment. Negative sentiment went up and peaked on the day of the first expression of suicidal ideation and then went back down to previous levels. Coppersmith et. al noticed similar patterns in their study of social media and suicidal ideation [9]. However, they did not observe a drop in posts on the day of the first suicidal ideation, only an increase followed by a decrease. One possible explanation for the drop on the day of the first expression of suicidal ideation was most likely due to a heightened emotional state. Studies of suicidal behavior suggest that for the more severe behavior of suicide attempt, there is significantly more consumption of alcohol [6, 11]. Therefore, these users might also be under the influence and therefore, less likely to post online. However, we notice that if the attempt failed, eventually posting will return to baseline. This could be quantitative evidence that discussing these
emotions and receiving support in the form of comments from other users can help resolve these emotions temporarily.

5.3 Limitations

We sampled the data and manually examined the sample to ensure that all threads on SuicideWatch expressed suicidal ideation. We also noted that the guidelines of the subreddit direct users to use the subreddit as a place for suicidal ideation and that the subreddit is moderated to ensure only threads containing suicidal ideation appear in the subreddit. However, since there are 45,894 threads in our dataset that contain an expression of suicidal ideation by a user, we could not verify all threads to ensure that they contain suicidal ideation.

Additionally, there are 5,646 users whose last post was on the day they expressed suicidal ideation. Of those, 3,067 users only posted during that day, this means that they were throwaway accounts. However, this leaves us with 2,579 distinct users who stopped posting for good after expressing suicidal ideation. There is no way for us to know whether these users self harmed or whether they stopped posting using that account. This is a serious limitation caused by the anonymity that Reddit allows its users.

6 CONCLUSION

Our study sought to examine the quantitative features of the behavior of Reddit users who express suicidal ideation. Through the analysis of their posts, found that these users progressed primarily either to or from one of 8 subreddits and two of those subreddits are common subreddits that most users on Reddit frequent (memes and AskReddit). The remaining 6 subreddits may point to a specific problem that the individual is dealing with like relationship advice or self harm. Identifying where these users progressed from can help us provide more tailored interventions for their particular emotional situation. We can also see a clear trend in their post volume as we approach the time of their first expression of suicidal ideation and as we pass that time. Users have a distinct increase in the volume of posts and a small increase in the proportion of negative posts. Identifying these patterns can allow us to perform research and develop a predictive model for predicting suicidal ideation. Using such a model, we can perform interventions such as display a suicide hotline number on their screen and thus prevent these users from progressing from suicidal ideation to suicide attempt and potentially save lives.

ACKNOWLEDGEMENT

This work was supported in part by the National Science Foundation under Grant NSF-1741306 and Grant IIS-1650531. Any opinions and conclusions or recommendations expressed in this study are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES


