Toward Macro-Insights for Suicide Prevention: Anaylzing Fine-Grain Distress at Scale

Abstract

This paper is an initial exploratory research

1 Introduction

Suicide is among the top ten—top five for indiviuals 10–44 years in age—leading causes of death in the United States (Heron and Tejada-Vera, 2009). Indeed, while mortality rates for most illnesses decreased between 2008 and 2009, the rate of suicide increased by 2.4% (Heron and Tejada-Vera, 2009). The lifetime prevalence for suicidal ideation is between 5.6 and 14.3 percent in the general population, and as high as 19.8–24.0% among youth (Nock et al., 2008).

The first step toward prevention is to identify, ideally in consultation with clinical experts, the risk factors associated with suicide. Due to social stigma and the impact on self -judgement and -awareness due to such risk factors as depression or drug use, individuals with suicidal ideation may not always reach out to professionals or, if they do, provide them with accurate information. They may not even realize their own level of suicide risk before it is too late. Self-reporting, then, is not a reliable means of detecting and assessing suicide risk.

Individuals may be more inclined to seek support from informal resources, such as social media, instead of seeking treatment (Crosby et al., 2011; Bruffaerts et al., 2011; Ryan et al., 2010). Evidence suggests that youth and emerging adults usually prefer to seek help from their friends and families; however, higher levels of suicidal ideation are associated with lower levels of help-seeking from both formal or informal resources (Deane et al., 2001).

These trends in help-seeking behavior suggests that social media might be a rich outlet for learning about support seeking. Internet- and telecommunications-driven activity is revolutionizing the social sciences by providing data—much of it publicly available—on human activity in situ, at volumes and a level of time and space granularity never before approached. Can such data improve clinical preventative measures by providing access to at-risk individuals who would otherwise go undected, and by leading to better science about suicide risk behaviors?

In this paper, we take steps toward the automatic detection of suicide risk among individuals via social media. We use various lexicon-based methods to retreive microblog posts (tweets) from Twitter and compare the performance of human annotators—some of whom are experts, and some of whom are not—to rate the level of distress of each tweet. Distress is an important risk factor in suicide that is relatively easy to observe from microblog text, though admittedly observing suicide risk behavior is a highly subjective and noisy venture. Expert annotation—rather than general-purpose tools for content and sentiment analysis such as LIWC (linguistic inquiry and work count), provides a measure of ground truth for language-based statistical modeling that is free of the sort of self-reporting biases that plague suicidality. We show that a principled approach for retrieving such training data results in better interannotator aggreement between non-experts and between experts and non-experts.

We also discover social and geographic patterns related to mood that microblogging sites such as Twitter reveal. Social support theory suggests that suicide and related mental health problems are strongly affected by one's physical and social environment (Wellman and Wortley, 1990). Twitter provides a very noisy frame into the offline state of its participants. The trick to using it effectively is in aggregating the data in a manner that is appropriate to the task at hand, based on appropriate principles. We show that otherwise weak correlations in the use of affective language between friends become very strong when one uses social networking concepts such as embeddedness to control for friendship strength.

Finally, we use our hand-annotated data to classify, at the level of individual tweets, risk factors for suicidal behavior from a highly-connected, geographically dense collection of over 2.5 million tweets from 6,000 highly active Twitter users. We use time-series analysis of these noisily classified tweets to detecting individuals who express high-distress episodes. Our outperforms very reasonable and baseline approaches.

2 Related Work

Data on suicide is traditionally collected from healthcare organizations, large-scale studies, or self reporting (Crosby et al., 2011; Horowitz and Ballard, 2009). These approaches are limited by sociocultural barriers, such as stigma and shame, among other factors (Crosby et al., 2011). For these reasons, and because of the fundamentally subjective nature of the problem, data on suicide is never particularly reliable.

Approximately one-third of all individuals who reported suicidal ideation in their lifetime made a plan. Nearly three-quarters of those who reported making a suicide plan actually attempted suicide (Kessler et al., 1999). According to Kessler, Borges, and Walters (Kessler et al., 1999), the odds of attempting suicide increased exponentially when individuals endorsed three of more risk factors (e.g., having a mood or substance abused disorder).

Other established risk factors include demographics, previous suicide attempts, mental health concerns (i.e., depression, substance abuse, suicidal ideation, self-harm, and impulsivity), family history of suicide, interpersonal conflicts (i.e., family violence and bullying), mechanism or means for suicidal behavior (e.g., firearms) are commonly cited risk factors for suicidal behavior (Nock et al., 2008; Crosby et al., 2011; Gaynes et al., 2004; Harriss and Hawton, 2005; Shaffer et al., 2004; Shaffer et al., 2004; Brown et al., 2000).

Suicide is a complex phenomenon with a low base rate; therefore, many researchers tend to focus the relationship between risk factors and suicidal behavior, without relying heavily on theoretical models (Nock et al., 2008). However, Mann, Waternaux, Haas, and Malone (1999), developed the stress diathesis model for suicidal behavior using many of the aforementioned risk factors. Specifically, this framework suggests that objective states such as(e.g., depression and life events) as well as subjective states and traits such as a (e.g., family history of depression and/or suicide as well as substance abuse) were among the risk factors that contributed to suicidal ideation and could eventually lead to either externalizing (e.g., interpersonal violence) or internalizing aggression (e.g., attempting suicide) (Mann et al., 1999).

Since the stress-diathesis model was developed using risk factors for suicidal behavior, it is the ideal framework to analyze publically available linguistic data from social media outlets such as Twitter. While traditional methods are limited by survey questions or clinical reports, data from social media can be used as a natural experiment to examine depression and suicidal ideation without being constrained by such sample biases as individuals who are willing take part in research and/or seek out formal sources of support. Moreover, this natural experiment method may provide information about individuals who are unlikely to engage in formal help-seeking behaviors and eventually it could be used to identify effective methods in natural helping. Hence, this universal approach to screening for suicidal behaviors may have future implications not only for identifying individuals who have a higher prevalence for suicidal behaviors but it could eventually lead to the methods for enhancing protective factors against suicide.

2.1 Affect Detection and Clinical Expertise

2.2 Affect in Social Media

Sentiment analysis has been widely studied in a number of computational settings, including on various social networking sites. Relatively little of this work has focused on suicide or related psychological conditions. (Masuda et al., 2013) study suicide on mixi. (Cheng et al., 2012) consider the ethical and political implications of online data collection for suicide prevention. (Jashinsky et al., 2013) show correlations between frequency in tweets related to suicide and actual suicide in the 50 United States of America. (Sadilek et al., 2014) study depression on Twitter. De Choudhury and collaborators studied depression—in general and post-partem—in Twitter (De Choudhury et al., 2012a; De Choudhury et al., 2012b; De Choudhury et al., 2013; De Choudhury and Counts, 2013) and Facebook (De Choudhury et al., 2014). Homan et al. investigate depression in TrevorSpace (Homan et al., 2014). A number of social theories of suicide have been proposed (Wray et al., 2011). Most of this work was with respect to offline social systems.

A rather substantial body of work already exists on the use of Twitter to study emotion (Bollen et al., 2011b; Dodds et al., 2011; Wang et al., 2012; Pfitzner et al., 2012; Kim et al., 2012; Bollen et al., 2011a; Pfitzner et al., 2012; Bollen et al., 2011c; Mohammad, 2012; Golder and Macy, 2011; De Choudhury et al., 2012a; De Choudhury et al., 2012b; De Choudhury et al., 2013; De Choudhury and Counts, 2013; Hannak et al., 2012; Thelwall et al., 2011; Pak and Paroubek, 2010). For instance, Golder and and Macy study aggregate global trends in "mood," and show, for example, that people wake up in a relatively good mood that decays as the day progresses (Golder and Macy, 2011), Bollen et al. (Bollen et al., 2011c) show that POMS-scored tweets are often tied to current events, such as elections and holidays.

A common theme in social network analysis is that actors who share ties generally share similar properties. A widely-used (Bliss et al., 2012; Coviello et al., 2014; Bollen et al., 2011a) metric for testing the overall similarity between actors in a network for some property X is *assortativity*, defined as the Pearson correlation coefficient (Newman, 2002) of X over all pairs of actors who share a tie. One line of research seeks to discover the mechanism through which such correlations occur (Newman, 2002). At the most fundamental level, this is a matter of whether like individuals seek each other out (called selection, or—confusingly enough—homophily) or whether related individuals influence one another. Teasing apart which of these two processes can be rather challenging and generally requires some level of experimental design (Centola, 2010; Centola, 2011) For instance, Coviello et al. study the spread of mood in Twitter (Coviello et al., 2014). They notice a very small—but statistically significant—spreading of mood over Facebook.

From the clinical perspective of detecting individuals who exhibit a high risk for committing suicide, determining causality remains a challenge to this multidimensional problem; however, finding patterns in the social interactions of individuals who commit suicide may provide additional insight. For our purposes, then a more relevant theory is perhaps that of social support (Wellman and Wortley, 1990) which seeks to clarify the social forces—protective, preventative, persuasive, or coercive—that affect behavior.

New York City Dataset

Unique users	632,611
Unique geo-active users	6,237
Tweets total	15,944,084
GPS-tagged tweets	4,405,961
GPS-tagged tweets by geo-active users	2,535,706
GPS-tagged tweets by geo-active users that show a symptom of an illness	2,047
"Follows" relationships between geo-active users	102,739
"Friends" relationships between geo-active users	31,874

Table 1: Summary statistics of the data collected from NYC. Geo-active users are people who geo-tag their tweets relatively frequently (more than 100 times per month). Note that the reciprocity rate in the social graph is about 31%, which is consistent with previous findings cite what is twitter.

At the most basic level, one can distinguish between *weak* and *strong* social ties and observe different behavior and effects between them.

Following in the work of Bliss et al. (Coviello et al., 2014) and Bollen et al. (Bollen et al., 2011a) we show that mood is assortative. We additionally consider the predictive power of various measureable notions of tie strength. We study suicide risk factors here but we would expect our methods would apply to other domains.

3 Methods

3.1 Overview

CMH: Mention coding as the method of obtaining ground truth; provide justification for it.

3.2 Data

Twitter is a worldwide popular social networking and microblogging service. Twitter message contains up to 140 words each and such words limit encourages users to update frequently. User's regular posts on Twitter have been used to predict depression[Munmun De Choudhury et al., 2013], influenza-like illnesses[Adam Sadilek et al., 2012] in previous studies. Our research looked into an old Twitter dataset originated from the New York City, which covered a month long period since May 18, 2010, total with about 2.5 million tweets from 6,237 unique users. See Table 1.

3.3 Data Collection

To identify suspected suicide tweets, we created a list of inclusive search terms/phrases according to various risk factors and warning signs linked to suicide. This search methodoly was used first by Jashinsky et al. (2013) [Table 1].

Among these, terms in Sad category were generated from LIWC [http://www.liwc.net/index.php]. All the rest were concluded from depression and other psychological disorders (Lewinsohn, Rohde, & Seely, 1994), prior suicide attempts (Lewinsohn et al., 1994), family violence, family history of drug abuse, firearms in the home, and exposure to the suicidal behavior of others (National Insti- tute of Mental Health, 2012). Other search terms included common antidepressants, as well as phrases that indicated suicide (Hawton, Zahl, & Weaterall, 2003), ideation (American Foundation for Suicide Prevention, 2012a), deliberate self-harm (Zahl & Hawton, 2004), bullying (Klomek, Sourander, & Gould, 2011), feelings of isolation (CDC, 2012), and impulsiveness (American Foundation for Suicide Prevention, 2012b). [Self-directed Violence Surveillance: Uniform Definitions and Recommended Data Elements, report from Megan]

Before searching, we did some kinds of preprocess work: (1) converted all text to lower case (2) stripped out all the punctuations and special characters; (3) built a slang dictionary which contains 5424 text slang based on online resources [http://www.noslang.com/dictionary/], Internet slang, and abbreviations. We replaced all the matching items fount in the dataset with corresponding easy read words. These two steps helped us extract more suicide-related tweets for analysis in the following process.

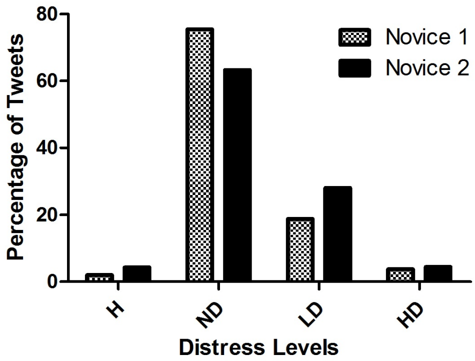
3.4 Preprocess

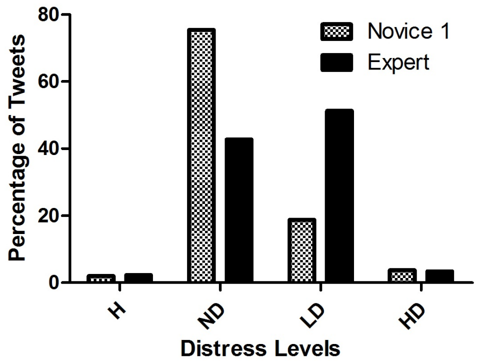
Move this section to other sections:

-stopwords -keep stemming words

Using Scikit-learn package(http://scikit-learn.org/stable/about.html#citing-scikit-learn), it is very simple to remove stop words (like "and", "the", etc.) with parameter "stop_words" tuned as "string {'english'}".

At the same time, we kept original words without any word stemming, i.e.: "computers", "computing", and "compute", neither of these words will be mapped to the same stem "comput". The reason here is: a tweet containing like "sooooooooooo sad!" is supposed to convey stronger emotions than simple "so sad".





3.5 Ground Truth

The dataset is divided into two equal parts, each set is then annotated. The first dataset is annotated by experts in Computing Science and Linguistics while the other set by a clinical expert in Psychology. Each dataset was annotated by male and female pairs. The reasoning behind this method is to analyze the difference in perception of distress between male and female, and, also between experts VS non-experts in psychology.

For annotation process each tweet was provided with a context, i.e. three tweets before and after the tweet to be annotated, along with the timestamp of these tweets and thematic category to which the tweet belonged to. Each tweet was annotated for distress level and thematic category. The distress level was divided into four categories: High Distress(HD), Low Distress(LD), No Distress(ND) and Happy(H), whereas, for the thematic category the annotaters labeled just yes or no based on whether the thematic category applied to the tweet or not.

Add example.

Annotator	H	ND	LD	HD
Annotator 1	2.0	75.4	18.8	3.8
Annotator 2	4.3	63.3	28.0	4.4
Annotator 3	2.3	42.8	51.3	3.4
Annotator 4				

Table 2: percentage of distress labels

3.6 Challenges with Annotation

There are unique challenges in annotating data from Twitter. Aside from having to become familiar with different types of slang and abbreviations that could have multiple meanings, this format provides limited background context to inform the annotation process. Outside of the theoretical informed annotation process, there were emerging themes of aggression, privilege and oppression, and daily struggles, among other. As a result of the aggressive context, personal bias may have impacted annotation decisions. For instance, numerous tweets contained sarcasm and dark humor which may result in annotators underestimating or overlooking actual distress. In addition, by pulling data from Twitter, critical information such as pictures and the context behind information that has been retweeted. Specifically, a few individuals retweeted in a humorous manner about what to say to someone who considering suicide; however, without any knowing the circumstances of the original message it was difficult to classify this tweet. It could represent dark humor or it could be a form of bullying.

3.7 Topic Modeling

Toping modeling is often used to analyze text data by finding topics within a corpus of documents. Each topic consists of words that occur together frequently. These models are capable of connecting words with similar meanings and distinguish words with multiple meanings. We utilize Latent Dirichlet Algorithm (Blei et al., 2003) to create these topics, in this method the documents (in our case tweets) are represented as random mixtures over latent topic where each topic is characterized by a distribution over words (REPHRASE).

We perform topic modeling on our dataset to compare the topics within high distress and everyday tweets. Before performing the topic modeling, the stop words and words that occur only once in the dataset are removed. The LDA algorithm is then applied to create 5 topics using 100 iterations, Table 3 shows the results.

Topic No	**Words**	**random**
Topic 1	miss u, leave alone, sleep forever, win lose, gon lose, left alone, #iconfess hate, lost best, best friend, think insomnia	let-know, don't-want, bout-2, even-tho, right-now, jus-got, gotta-go, don't-wanna, wit-da
Topic 2	hate job, feel sad, don't wanna, feel helpless, bed lonely, feel better, miss you, sad =/, tired everything, miss love	feel-like, look-like, let's-go, last-night, looks-like, don't-get, fuck-wit, show-love, smile-face, gonna-start
Topic 3	miss you!, wanna cry, committing suicide, tired living, miss 2, one person, broke bitches, worst feeling, leave world, bout go	time-get, go-sleep, know-(cont), can't-wait, even-though, hip-hop, big-baby, lil-wayne, listen-2, don't-think
Topic 4	commit suicide, get hurt, miss baby, feel empty, :(miss, lost phone, don't let, drug overdose, can't wait, work	don't-know, make-sure, dont-see, wats-good, hell-yea, r-u?, need-new, yall-niggas, can't-get, don't-care
Topic 5	feel like, tummy hurts, lost friend, ima miss, deserve die, right now..., hurts :(, get fat, every day, like crying	good-morning, happy-birthday, bout-go, what's-good, jus-w, chris-brown, right-now!, 2-da, don't-feel, don't-understand

Table 3: Topic Analysis on bigrams of High Distress and Random Tweets

3.8 Analysis

3.9 Features

To prepare the tweets for the classifier, we extract features using the unigram, bigram and trigram model. For example, a simple tweet "I am so happy" is represented as the following feature vector: {I, am, so, happy, I am, am so, so happy, I am so, am so happy}. The tf-idf values were calculated for each attribute: tf-idf stands for "term frequency - inverse document frequency", which is a numerical statistic to reflect how important a tokenization is to a document in a collection or corpus. The tf-idf value increases proportionally to the number of times a tokenization appears in the dataset, but is offset by the frequency of the tokenization in the corpus, which helps to control for the fact that some words are generally more common than others. With the help of Scikit-learn package(http://scikit-learn.org/stable/about.html#citing-scikit-learn), there values can be easily acquired.

3.10 feature selection

chi2 feature selection

3.11 Prediction Model

We use Support Vector Machines (SVM) for our prediction model. SVM are proven to perform better, while working with text data (Joachims, 1998).

Describe the model [WIP]

3.12 Network Features

One of the fundamental properties of social networks appears to be *tie strength*, or how "close"

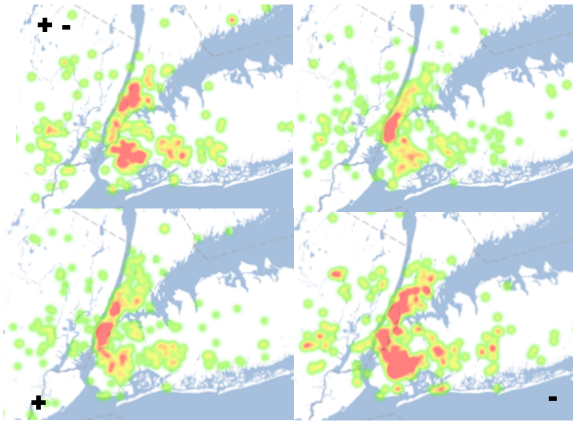


Figure 1: Heatmaps showing the most common tweet location for those individuals who (clockwise from upper left) are among the top 25% in using negative and positive emotional language, in the bottom 25% of both categories, in the top 25% of negative language and bottom 25% of positive language, and in the top 25% of positive language and bottom 25% of negative language.

socially two people are. A large body of literature suggests that people are more likely to share personal information with stronger ties, and that weak ties play an important role in providing new information. Measuring tie strength is problematic, as there is no gold standard here. Social networking services seem to exacerbate the disparity between strong and weak ties, as many have "friends" or "followers" whom they may not even know personally, and also create their own problems and opportunities for estimating tie strength. In large-scale network analysis, researchers have sometimes characterized tie strength by the *embeddedness* of an edge, which is the number of friends in common that two actors sharing a tie have. Highly embedded links are part of a strong social fabric, and represent strong ties. Another method of estimating tie strength is to measure the amount of activity between users. In this study, we investigate both the role that embeddness and activity play in corrolating suicide-related language.

4 Results

4.1 Geography and Emotional Language

4.2 Strength of Ties

Table **??** shows how the use of sad language (measure by the LIWC sad feature) correlates among users sharing different tie strength and between personal and broadcast messages.

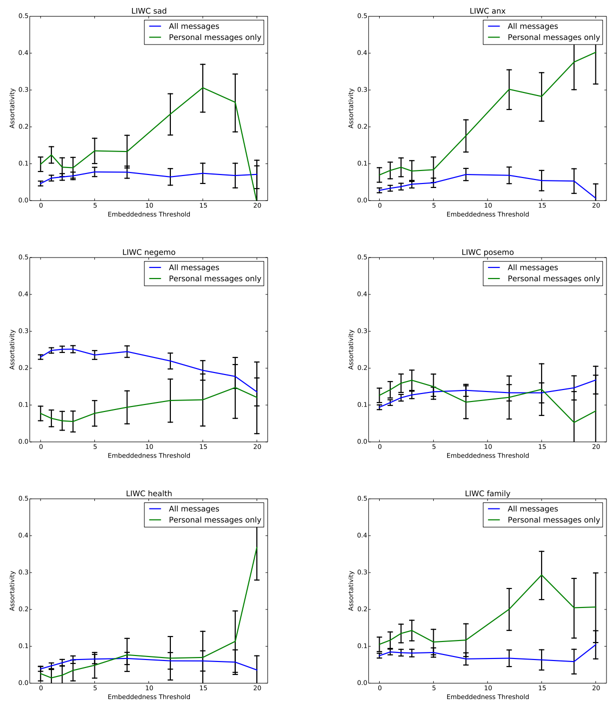


Figure 2: Correlations between twitter friends of various LIWC scores increase as the strength of ties increase, while others behave very differently. Interestingly, two of the LIWC scores most closely associated with distress, sadness and anxiety, are among those most strongly affected by ignoring those friendships having fewer than a certain number mutal friends (i.e., embeddedness).

5 Discussion

As previously mentioned, many of the risk factors for suicidal behavior may be linked to other expressions of distress such as aggression and interpersonal violence (Mann et al., 1999). The goal of this study is to classify whether or not tweets were related to distress in order to determine the feasibility of classifying suicidal behaviors. However, due to the overlap between internal and external expressions of anger, it is difficult to classify suicidal behavior without more contextual information. Consistent with the stress diathesis model for suicidal behavior, aggression was an emerging theme that arose from the data. A number of individuals tweeted about feeling empty, hopeless, angry, frustrated, and alone. While these are risk factors for internalizing aggression (i.e., suicidal behavior); these states are also associated with externalizing aggression. In addition to overt expressions of anger and violence, many of the humorous tweets had an aggressive undertone. Individuals often exert dominance by using pejorative and derogatory language, and the content of these tweets was often suggestive of past or current distress.

5.1 Limitations

As ground truth, we rely on tweets hand-annotated by experts and non-experts. However, the mental state of another individual, observed from a line or two of text often written in an informal register is necessarily hard to discern and, even under less noisy conditions, extremely subjective; even the observers' personal understandings of such concepts as "distress" may differ drastically. This makes inter-annotator agreement quite a challenge, to say nothing of observation in some objective fashion of the true mental state.

Higher levels of suicidal ideation have an inverse relationship with all types of help-seeking and a positive correlation with the decision to not seek support (Deane et al., 2001). Thus we would expect suicidal individuals to generally be less active on social media than those who are not. (One ray of sunshine is that a number of studies have shown a positive correlation between online social network use and negative mood. Perhaps this means in part that individuals who are depressed are slower to disengage on- rather than off-line.) Part of the problem in assessing the effectiveness of self-reporting is the relative rareness by which suicide occurs, and by the inherent subjectivity of the act, which makes any data on suicide fuzzy.

5.2 Tools Used

6 Conclusion and Future Work

Acknowledgments

References

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022, March.

Catherine A Bliss, Isabel M Kloumann, Kameron Decker Harris, Christopher M Danforth, and Peter Sheridan Dodds. 2012. Twitter reciprocal reply networks exhibit assortativity with respect to happiness. *Journal of Computational Science*, 3(5):388–397.

Johan Bollen, Bruno Gonçalves, Guangchen Ruan, and Huina Mao. 2011a. Happiness is assortative in online social networks. *Artificial life*, 17(3):237–251.

Johan Bollen, Huina Mao, and Xiaojun Zeng. 2011b. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8.

Johan Bollen, Alberto Pepe, and Huina Mao. 2011c. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, pages 450–453.

Gregory K Brown, Aaron T Beck, Robert A Steer, and Jessica R Grisham. 2000. Risk factors for suicide in psychiatric outpatients: a 20-year prospective study. *Journal of consulting and clinical psychology*, 68(3):371.

Damon Centola. 2010. The spread of behavior in an online social network experiment. *science*, 329(5996):1194–1197.

Damon Centola. 2011. An experimental study of homophily in the adoption of health behavior. *Science*, 334(6060):1269–1272.

Qijin Cheng, Shu-Sen Chang, and Paul SF Yip. 2012. Opportunities and challenges of online data collection for suicide prevention. *The Lancet*, 379(9830):e53–e54.

Lorenzo Coviello, Yunkyu Sohn, Adam D. I. Kramer, Cameron Marlow, Massimo Franceschetti, Nicholas A. Christakis, and James H. Fowler. 2014. Detecting emotional contagion in massive social networks. *PloS one*, 9(3):e90315.

Alex E Crosby, LaVonne Ortega, and Cindi Melanson. 2011. *Self-directed violence surveillance: Uniform definitions and recommended data elements*. Centers for Disease Control and Prevention, National Center for Injury Prevention and Control, Division of Violence Prevention.

Munmun De Choudhury and Scott Counts. 2013. Understanding affect in the workplace via social media. In *16th ACM Conference on Computer supported cooperative work and Social Media (CSCW 2013)*, pages 303–316. ACM.

Munmun De Choudhury, Scott Counts, and Michael Gamon. 2012a. Not all moods are created equal! exploring human emotional states in social media. In *Sixth International AAAI Conference on Weblogs and Social Media*.

Munmun De Choudhury, Michael Gamon, and Scott Counts. 2012b. Happy, nervous or surprised? classification of human affective states in social media. In *Sixth International AAAI Conference on Weblogs and Social Media*.

Munmun De Choudhury, Scott Counts, and Eric Horvitz. 2013. Major life changes and behavioral markers in social media: Case of childbirth. In *Proc. CSCW*.

Munmun De Choudhury, Scott Counts, Eric J Horvitz, and Aaron Hoff. 2014. Characterizing and predicting postpartum depression from shared facebook data. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pages 626–638. ACM.

Frank P Deane, Coralie J Wilson, and Joseph Ciarrochi. 2001. Suicidal ideation and help-negation: Not just hopelessness or prior help. *Journal of clinical psychology*, 57:901–914.

Peter Sheridan Dodds, Kameron Decker Harris, Isabel M Kloumann, Catherine A Bliss, and Christopher M Danforth. 2011. Temporal patterns of happiness and information in a global social network: Hedonometrics and twitter. *PloS one*, 6(12):e26752.

Bradley N Gaynes, Suzanne L West, Carol A Ford, Paul Frame, Jonathan Klein, and Kathleen N Lohr. 2004. Screening for suicide risk in adults: a summary of the evidence for the us preventive services task force. *Annals of Internal Medicine*, 140(10):822–835.

S.A. Golder and M.W. Macy. 2011. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science*, 333(6051):1878–1881.

Aniko Hannak, Eric Anderson, Lisa Feldman Barrett, Sune Lehmann, Alan Mislove, and Mirek Riedewald. 2012. Tweetinin the rain: Exploring societal-scale effects of weather on mood. In *Proceedings of the 6th International AAAI Conference on Weblogs and Social Media (ICWSM12)Dublin2012*.

Louise Harriss and Keith Hawton. 2005. Suicidal intent in deliberate self-harm and the risk of suicide: the predictive power of the suicide intent scale. *Journal of Affective Disorders*, 86(2):225–233.

Melonie Heron and Betzaida Tejada-Vera. 2009. Deaths: leading causes for 2005. *National vital statistics reports: from the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System*, 58(8):1–97.

Christopher M Homan, Naiji Lu, Xin Tu, Megan C Lytle, and Vincent Silenzio. 2014. Social structure and depression in trevorspace. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pages 615–625. ACM.

Lisa M Horowitz and Elizabeth D Ballard. 2009. Suicide screening in schools, primary care and emergency departments. *Current opinion in pediatrics*, 21(5):620–627.

Jared Jashinsky, Scott H Burton, Carl L Hanson, Josh West, Christophe Giraud-Carrier, Michael D Barnes, and Trenton Argyle. 2013. Tracking suicide risk factors through twitter in the us. *Crisis*, pages 1–9.

Thorsten Joachims. 1998. Text categorization with support vector machines: Learning with many relevant features. In Claire Ndellec and Cline Rouveirol, editors, *Machine Learning: ECML-98*, volume 1398 of *Lecture Notes in Computer Science*, pages 137–142. Springer Berlin Heidelberg.

Ronald C Kessler, Guilherme Borges, and Ellen E Walters. 1999. Prevalence of and risk factors for lifetime suicide attempts in the national comorbidity survey. *Archives of general psychiatry*, 56(7):617–626.

Suin Kim, J Bak, and Alice Oh. 2012. Do you feel what I feel? social aspects of emotions in Twitter conversations. In *Proceedings of the AAAI International Conference on Weblogs and Social Media*.

Naoki Masuda, Issei Kurahashi, and Hiroko Onari. 2013. Suicide ideation of individuals in online social networks. *PloS one*, 8(4):e62262.

Saif M Mohammad. 2012. # emotional tweets. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation*, pages 246–255. Association for Computational Linguistics.

Mark EJ Newman. 2002. Assortative mixing in networks. *Physical review letters*, 89(20):208701.

Matthew K Nock, Guilherme Borges, Evelyn J Bromet, Christine B Cha, Ronald C Kessler, and Sing Lee. 2008. Suicide and suicidal behavior. *Epidemiologic reviews*, 30(1):133–154.

Alexander Pak and Patrick Paroubek. 2010. Twitter as a corpus for sentiment analysis and opinion mining. In *Proceedings of LREC*, volume 2010.

René Pfitzner, Antonios Garas, and Frank Schweitzer. 2012. Emotional divergence influences information spreading in twitter. *AAAI ICWSM*, 2012:2–5.

Adam Sadilek, Christopher Homan, Walter S. Lasecki, Vincent Silenzio, and Henry Kautz. 2014. Modeling fine-grained dynamics of mood at scale. In *WSDM 2014 Workshop on Diffusion Networks and Cascade Analytics*.

David Shaffer, Michelle Scott, Holly Wilcox, Carey Maslow, Roger Hicks, Christopher P Lucas, Robin Garfinkel, and Steven Greenwald. 2004. The columbia suicidescreen: Validity and reliability of a screen for youth suicide and depression. *Journal of the American Academy of Child & Adolescent Psychiatry*, 43(1):71–79.

Mike Thelwall, Kevan Buckley, and Georgios Paltoglou. 2011. Sentiment in twitter events. *Journal of the American Society for Information Science and Technology*, 62(2):406–418.

Wenbo Wang, Lu Chen, Krishnaprasad Thirunarayan, and Amit P Sheth. 2012. Harnessing twitter 'big data' for automatic emotion identification. In *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)*, pages 587–592. IEEE.

Barry Wellman and Scot Wortley. 1990. Different strokes from different folks: Community ties and social support. *American journal of Sociology*, pages 558–588.

Matt Wray, Cynthia Colen, and Bernice Pescosolido. 2011. The sociology of suicide. *Annual Review of Sociology*, 37:505–528.

Suicide Risk Factor Category	Search Terms and Phrases
Depressive Feelings	me abused depressed, tired of living,so depressed,leave this world, wanna die,me hurt depressed, feel hopeless depressed, feel alone depressed, i feel helpless, i feel worthless, i feel sad, i feel empty, i feel anxious, hate my job, feeling guilty, deserve to die, desire to end own life, feeling ignored, tired of everything, feeling blue, have blues
Depression Symptoms	sleeping pill, sleeping a lot, i feel irritable, i feel restless, have insomnia, sleep forever,sleep disorder
Drug Abuse	depressed alcohol, sertraline, zoloft, prozac, pills depressed, clonazepam, drug overdose, imipramine
Prior Suicide Attempts	suicide once more, me abused suicide, pain suicide, tried suicide
Suicide Around Individual	mom suicide tried, sister suicide tried, brother suicide tried, friend suicide, suicide attempted, suicide attempt
Suicide Ideation	commit suicide,committing suicide,feeling suicidal, suicide thought about, thoughts suicide, think suicide, thought killing myself, used thought suicide, once thought suicide, past thought suicide, multiple thought suicide, want to suicide, shoot myself, a gun to head, hang myself, intention to die
Self Harm	stop cutting myself, hurt myself, cut myself
Bullying	i am being bullied, i have been cyber bullied, was bullied, feel bullied, stop bullying me, keeps bullying me, always getting bullied
Gun Ownership	gun suicide, shooting range went, gun range my
Psychological Disorders	diagnosed schizophrenia, diagnosed anorexia, diagnosed bulimia, i diagnosed ocd, i diagnosed bipolar, i diagnosed ptsd, diagnosed borderline personality disorder, diagnosed panic disorder, diagnosed social anxiety disorder, diagnosed post traumatic stress disorder, sleep apnea
Family Violence Discord	dad fight again, parents fight again, lost my friend, argument with wife, argument with husband, shouted at each other
Impulsivity	i impulsive, i am impulsive
Sad	abandon, ache, aching, agoniz, agony, alone, broke,cried, cries, crushed, cry, crying, damag, defeat, depress, depriv, despair, devastat, disadvantage, disappoint, discourag, dishearten, disillusion, dissatisf, doom, dull, empt, fail, fatigu, flunk, gloom, grave, grief, griev, grim, heartbreak, heartbroke, helpless, homesick, hopeless, hurt, inadequa, inferior, isolat, lame, lone, longing, lose, loser, loses, losing, loss, lostlow, melanchol, miser, miss, missed, misses, missing, mourn, neglectoverwhelm, pathetic, pessimis, piti, pity , regret, reject, remorse, resign, ruin, sad, sadde, sadly, sadness, sob, sobbed, sobbing, sobs, solemn, sorrow, suffer, suffered, sufferer, suffering, tears, traged, tragic , unhapp,unimportant, unsuccessful, useless, weep, wept, whine, whining, woe, worthless, yearn

Table 4: Search phrases for categories