Detecting Changes in Suicide Content Manifested in Social Media Following Celebrity Suicides

Mrinal Kumar[†], Mark Dredze[‡], Glen Coppersmith[‡], Munmun De Choudhury[†]
[†]College of Computing, Georgia Institute of Technology
[‡]Human Language Technology Center of Excellence, Johns Hopkins University
{mkumar73, munmund}@gatech.edu, mdredze@cs.jhu.edu, coppersmith@gmail.com

ABSTRACT

The Werther effect describes the increased rate of completed or attempted suicides following the depiction of an individual's suicide in the media, typically a celebrity. We present findings on the prevalence of this effect in an online platform: r/SuicideWatch on Reddit. We examine both the posting activity and post content after the death of ten high-profile suicides. Posting activity increases following reports of celebrity suicides, and post content exhibits considerable changes that indicate increased suicidal ideation. Specifically, we observe that post-celebrity suicide content is more likely to be inward focused, manifest decreased social concerns, and laden with greater anxiety, anger, and negative emotion. Topic model analysis further reveals content in this period to switch to a more derogatory tone that bears evidence of self-harm and suicidal tendencies. We discuss the implications of our findings in enabling better community support to psychologically vulnerable populations, and the potential of building suicide prevention interventions following high-profile suicides.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

Keywords

social media; suicide; Werther effect; mental health; Reddit

1. INTRODUCTION

The "Werther effect" describes the increased rate of completed or attempted suicide following media reported incidents or depiction of celebrity suicides [19]. Naturally, the Werther effect is a highly regarded phenomenon in media effect research — prior literature has examined ways of exploiting findings on the effect to inform journalistic practices, as well as to guide suicide prevention programs [32].

However, key to the success of these interventions is the ability to measure the prevalence of the phenomenon as well as quantifying its manifestation in psychologically vulnerable populations. Unfortunately, scientific evidence in support of the Werther effect has been either qualitative, specific to particular populations, self-report based, or based on government or agency reported suicide rates [5, 16]. Due to the sensitive nature of suicide, gathering data on suicidal ideation or attempts to corroborate existence of the Werther effect, especially the epidemiological extent of the phenomenon, has been challenging [32].

Online social media platforms, such as Twitter, Facebook, Reddit, and Tumblr, are popular outlets for people seeking information and social support, including issues around a variety of psychological and health challenges [37, 38, 13]. An attractive feature of these platforms is that they allow for anonymous or pseudonymous participation. Thus, they provide individuals with a candid platform of expression, especially around conditions that are considered socially stigmatized, such as suicide and mental illness [6].

In this paper, we leverage data derived from a widely adopted suicide support forum called "SuicideWatch"[1]. This community is hosted on the popular social media Reddit. Through analysis of historical posts and associated metadata, we examine attributes of suicidal ideation and suicide interest, specifically relating to the Werther effect. Our underlying assumption is that such a forum can provide us information about individuals who are likely prone to suicidal thoughts or tendencies. Moreover, this comes at a scale previously unavailable, spanning thousands of individuals, thus enabling rigorous statistical analysis. The ability to analyze the content of the messages shared on this forum further allows us derive language and behavior related attributes associated with the Werther effect. The main contributions of this paper are:

- We show notable increases in the posting frequency on the forum following reports of celebrity suicides. This change is persistent after accounting for baseline expected variability in posting activity, and in contrast to other mental health forums on Reddit.
- We include linguistic measures of behavior, obtained from content shared on the forum. We find that posts following celebrity suicides express greater negativity, raised cognitive bias, increased self-attentional focus, and lowered social integration in the aftermath of celebrity suicides.
- We utilize n-gram and topic model analyses to show that expressions of self-derogatory behavior, depictions of suicidal tendencies, and detachments from the social realm are more frequently discussed after celebrity suicides.

Broadly, we aim to aid suicide prevention, one of the leading causes of death in the United States. While suicide is the 10th leading cause of death[2], it rises to the third position for people aged

HT '15, September 1–4, 2015, Guzelyurt, TRNC, Cyprus.
 ISBN 978-1-4503-3395-5/15/09 ...$15.00.
DOI: http://dx.doi.org/10.1145/2700171.2791026.

[1]http://www.reddit.com/r/suicidewatch
[2]Per data from 2013: http://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm

15-24,[3] a demographic known to extensively use social media.[4] An ability to measure existence and prevalence of the Werther effect from large-scale online data may help moderation efforts on these platforms, especially in periods that succeed high-profile celebrity suicide. Online interventions may also be built to extend careful advice and help to individuals identifying with a particular suicide event and therefore at an increased risk of future suicide.

Privacy Note: All social media data used in this paper are publicly available. At no time did we contact or interact with a user. Given the sensitive nature of this research, we took care to anonymize the data in analysis and presentation, so as to minimize any inadvertent disclosure of personal information, or information that may reveal cues about an individual's online identity. Approval was obtained from the relevant institutional review boards.

2. BACKGROUND

2.1 Werther Effect and Copycat Suicides

A number of studies have found evidence, both anecdotal and statistical, supporting the existence of the Werther effect in news and entertainment media [19, 5, 31, 54]. These findings indicate that the effect varies in its persistence, peaking by about three days and leveling out by about two weeks [17]. The effect is also associated with the amount and prominence of media coverage, and it has consistently been observed to be prevalent in diverse cultures [43, 63, 16]. Hence literature has also critically examined the role of media as a risk factor of suicide [56, 54].

A complementary line of work also found that repetitive or high-profile suicide events led to imitative behaviors [61] — especially with media consumers with similar characteristics to the celebrity. Explicit depiction of particular methods has been known to lead to increases in completed suicides employing that method — this phenomenon is also known as "copycat suicides" [33].

Previous work on the Werther effect has measured actual suicides, which do not lend direct insights into the nature of suicidal ideation. Moreover, little has been investigated in terms of the role of online suicide discussion and support platforms, and their relationship to the manifestation of the Werther Effect. Relatedly, much of the psychology research into suicidality and suicidal ideation relies on surveys and self-report. These methods can be intrusive, expensive, vulnerable to participants' memory bias, vulnerable to experimental demand effects, and often lack temporal granularity. These are further complicated by impediments due to the stigma and sensitivity surrounding suicide. In contrast, social media, especially public online communities geared towards supporting suicidal individuals, can provide us with a rich, real-time source of information about the phenomena, without many of the drawbacks of more traditional approaches. Here, we measure suicide related interest and intent by looking at conversations on a prominent suicide forum on Reddit, called SuicideWatch. This focus on online communities allows us to look at content around suicide interest and intent, allowing complementary insight into the problem.

2.2 Psychology of Suicide

Several studies have examined linguistic and behavioral attributes associated with suicide. Due to the inherent challenge of obtaining information shared by suicidal individuals, most prior work focuses on poetry as the source of content, analyzing suicidal tendencies of poets, or publicly available corpus of suicide notes [52]. [57] leveraged the popular and widely validated psycholinguistic lexicon LIWC to understand the various linguistic constructs manifested in poems written by suicidal poets. An important finding of this work was that the suicidal poets showed higher usage of first-person singular nouns in their writing when compared to a control group of poets, interpreted as lowered social integration or stronger inward focus. In general they found that suicidal thoughts can be associated with detachment from the social realm and higher preoccupation with the self. Along similar lines, others [26, 53, 42, 15, 28, 60] have independently found that suicidal poets showed ambivalence towards death in their writing, exhibited cognitive distortion and emotional inhibition, and over time their linguistic style shifted to a personal, expressive form and stronger inward focus. Authors in [44] found cognitive impairment to be a characteristic predictor of suicide ideation based on a standardized battery of questionnaires. Petrie and Brook found that repeated suicide attempts were associated with lowered coherence and self-esteem, and greater hopelessness [42].

Beyond suicide, literature in psychology has focused on attributes of mental illness and other behavioral health concerns [1]. Pennebaker and colleagues used LIWC in a number of different populations and scenarios to identify language related markers of anxiety, stress, neurotic tendencies and psychiatric disorders [41, 47].

Taken together, this suggests the potential of leveraging behavioral and linguistic cues for understanding vulnerability in individuals and populations. While this paper does not focus on inferring suicidality of individuals based on the content they share online, we do examine how such behavioral and linguistic cues may be mined from data gathered from social platforms, with particular focus on the Werther Effect.

2.3 Health, Well-being and Social Media

Social media research has indicated that psychological states, health, well-being, and social support status of individuals may be gleaned via analysis of language and conversational patterns. These include utilizing social media to understand conditions and symptoms related to diseases [37, 38], influenza propagation [49] and prediction [39], cyberbullying and teenage distress [14], substance abuse [29, 30], postpartum depression [10], mental health [12, 36, 22, 8, 9, 50], and insomnia [24].

Contrastively, research on suicide in social media is limited. Authors in [62] focused on South Korean blogs to predict nationwide suicide rate data. An interesting aspect of this work is that economic, meteorological, and celebrity suicide count variables were used as controls in the prediction model. Similarly, there is some evidence of announcements of suicidal thoughts on Facebook [48], and public health consequences of suicide related content on social media was examined in [27].

This paper builds on this emergent body of research. We utilize data shared on a social media platform, Reddit, to probe the psyche and social milieus of individuals contemplating suicide or seeking support to fight such tendencies.

3. DATA

3.1 Social Media Data

As mentioned earlier, we examine data from Reddit. We provide a description of the features of this social media platform, which are important to understand the context of our research problem. Reddit allows users to submit content in the form of links or text posts, organized by areas of interest or sub-communities called "subreddits" (e.g., politics, programming, science). Users can voice their

[3] http://www.cdc.gov/injury/wisqars/pdf/10lcid_all_deaths_by_age_group_2010-a.pdf

[4] http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet/

It just feels like death is the only option but also so conflicted and wanting to live. I cannot think anymore.
I desperately need some help from this subreddit! I feel I am right on the edge.
It will be one more week, and I'll be done.
There's just no one I can talk to if my life is worth living. I fear they'd think it is a suicide threat and manipulation. Maybe I should just do it.
Just want to let you know, I'll be ending it this Weekend

Table 1: Example titles of postings in SW dataset, modified to protect the privacy of the authors.

opinion on the post via a voting mechanism which allows more popular submissions to be featured more prominently according to their score: the difference between the "upvotes" and "downvotes" cast on each post (also known as "score"). Users can also engage with each other via a comment thread attached to each post. In 2014, Reddit had 71 billion page views, over $8,000$ active communities, 55 million posts, and 535 million comments[5].

In this paper, we focus specifically on the subreddit called "SuicideWatch", a forum for users contemplating suicide and who seek help, advice, and support. It is a strong support community with (currently) about 35K subscribers. The community is highly moderated, with many of its moderators and active subscribers adopting prominent roles in providing support to individuals showing vulnerability. In this subreddit, votes on posts are used as a proxy for support and to increase or decrease a post's prominence, rather than as a statement of agreement/disagreement.

We used Reddit's official API[6] to collect posts, comments, and associated metadata from r/SuicideWatch (hereafter SW). Our analysis in this paper is based on all posts made to SW between October 16, 2013 and December 19, 2014 – 66,059 posts from 19,159 unique users. Example (paraphrased) titles of posts from r/SuicideWatch are given in Table 1. We collected the title of the post, the body or textual content, ID, timestamp, author ID, the number of upvotes and downvotes the post obtained, including the difference between upvotes and downvotes on the post (i.e. score.)

3.2 Wikipedia Data

Next we compiled a list of reported celebrity suicides which fell within the time range of our Reddit data. Defining who is a "celebrity" is nontrivial, so we refer to the Wikipedia page listing celebrity suicides,[7] as a way to measure who has sufficient celebrity status for inclusion. We obtained 10 reported celebrity suicides in the same period as our Reddit data; their names and reported suicides are shown in Table 2.

We measure the prominence of a celebrity's death by measuring the change in Wikipedia page views for the celebrity's Wikipedia page. Wikipedia provides daily page view statistics for each page.[8] We compare the number of page-views in the two weeks prior to their death with the two weeks following their death in terms of z-score (Figure 1). Here z-scores are computed by converting the page views to standard normal variable with 0-mean and standard deviation of 1. For 9/10 of the cases, we see a notable spike in number of views, showing that the suicides of these individuals were well-known enough to be viewable on such a macro scale and for examining the presence of Werther Effect in social media.

We note two aspects related to the above analysis, and which will be used through the rest of this paper. First, since we are focusing on different types of data sources—Wikipedia and Reddit, we use

[5]http://www.redditblog.com/2014/12/reddit-in-2014.html
[6]http://www.reddit.com/dev/api
[7]http://en.wikipedia.org/wiki/List_of_suicides
[8]http://en.wikipedia.org/wiki/Wikipedia:Pageview_statistics

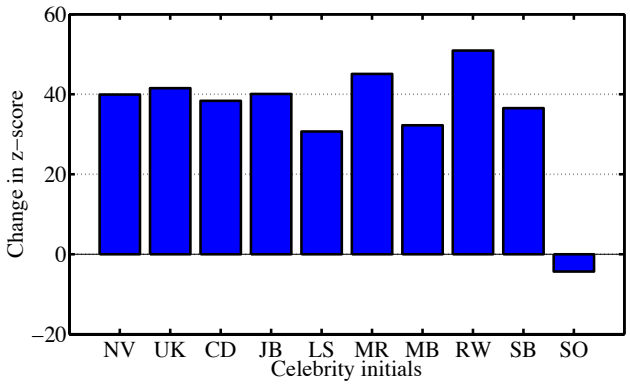

Figure 1: z-score change for Wikipedia celebrity page views succeeding a celebrity suicide.

Celebrity	Profession	Reported suicide
Ned Vizzini	Writer	12/19/2013
Uday Kiran	Actor	1/5/2014
Charlotte Dawson	Television personality	2/22/2014
Jurgen Brummer	Gymnast	2/25/2014
L'Wren Scott	Stylist/Model	3/17/2014
Michael Ruppert	Writer/Activist	4/13/2014
Malik Bendjelloul	Filmmaker/Journalist	5/13/2014
Robin Williams	Actor/Comedian	8/11/2014
Simone Battle	Actress/Singer	9/5/2014
Sean O'Haire	Wrestler	9/8/2014

Table 2: List of Wikipedia derived celebrity suicides.

z-score conversion as a normalization technique for the Wikipedia page views and Reddit's SW posting activity volume. Further, the above observation in Wikipedia data, and the analyses that ensue, focus on observing changes over a two week window preceding and succeeding a celebrity suicide; this choice is motivated by our initial analyses and from the literature on Werther Effect [17].

4. METHODS

Our goal is to measure the change in both the quantity and quality of posts to SW following a celebrity suicide. First, we will measure the volume of posts on SW preceding and succeeding a celebrity suicide to obtain a measure of increased interest in the topic of suicide. Next, we will use a series of content analysis techniques to examine the nature of these posts: how the topic of posts changed in the wake of the suicide.

4.1 Measuring Post Volume

4.1.1 Developing a Baseline

We begin by constructing a baseline as to the expected variability in posts by measuring pairs of subsequent two week periods. Deviations from these expected trends following a celebrity suicide

depression	mentalhealth	Anger
traumatoolbox	psychoticreddit	EatingDisorders
getting_over_it	survivorsofabuse	alcoholism
rapecounseling	bipolarreddit	addiction
hardshipmates	StopSelfHarm	socialanxiety
MakeMeFeelBetter		ptsd
depersonalization, derealization		feelgood
Borderline Personality Disorder		panicparty

Table 3: Control group mental health subreddits.

would provide evidence for the Werther effect.

Our goal is to identify, per celebrity, a set of k consecutive 2-week time period pairs in the entire timeframe of our data. We refer to the first item of the pair (i.e., the first 2-week window) as the "preceding" window, and the immediately following 2-week window as "succeeding". Collectively, these baseline pairs yield an empirical distribution of the expected variation when there is not a celebrity suicide. Specifically, each of the k two-week window pairs (1) have no reported celebrity suicide, and (2) take periods that start on the same as the day of week. This accounts for day-of-week related variations on SW. For the purposes of this paper, we choose k as 20.

4.1.2 Developing a Control

Next, we develop a control to establish that changes in volume of posts in SW succeeding a celebrity suicide compared to that preceding it is attributed to the topic of suicide in particular, and such changes are not part of a broader shift in interest in mental health topics. For this purpose, we identified a set of "control group" subreddits, which are on topics related to mental health, but are unlikely to be specifically about suicide or suicidal ideation. These mental health subreddits (henceforth referred to as MH subreddits) were compiled based on our prior work in [11]; refer to the paper for details on how these subreddits are identified and crawled. Table 3 lists the control subreddits, crawled in the same timeframe as SW. We obtained 32,509 posts from 23,807 unique users. Like SW, all of these subreddits are public.

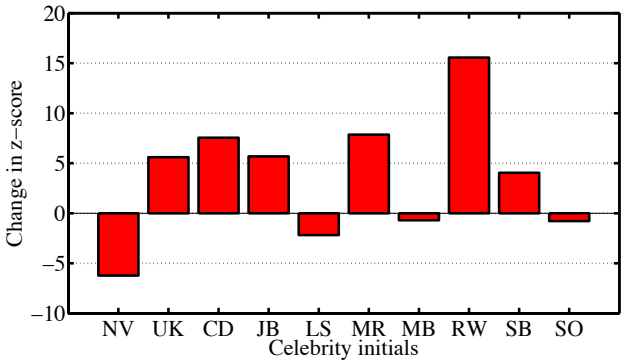

Figure 3: z-score changes in SW post volume succeeding and preceding a celebrity suicide.

4.2 Content Analysis

4.2.1 Linguistic Measures

We propose four categories of linguistic and non-linguistic attributes to examine preceding/succeeding celebrity suicides. These are: (1) **affective attributes**, (2) **cognitive attributes**, (3) **linguistic style attributes**, and (4) **social attributes**. Measures belonging to all of these attribute categories are largely based on the psycholinguistic lexicon LIWC [40], and were motivated from prior literature that examine associations between the behavioral expression of individuals and their responses to traumatic context and crises, including vulnerability due to mental illness [7, 11]. Note that LIWC has been extensively validated to perform well on Internet language [7, 18].

(1) We consider two measures of affect derived from LIWC: positive affect (PA), and negative affect (NA), and four other measures of emotional expression: *anger*, *anxiety*, *sadness*, and *swear*.

(2) We use LIWC to define the cognitive measures as well: (a) cognition, comprising *cognitive mech*, *discrepancies*, *inhibition*, *negation*, *death*, *causation*, *certainty*, and *tentativeness*; and (b) perception, comprising set of words in LIWC around *see*, *hear*, *feel*, *percept*, *insight*, and *relative*.

(3) Next, we consider four measures of linguistic style: (a) **Lexical Density**: consisting of words that are *verbs*, *auxiliary verbs*, *nouns*, *adjectives* (identified using NLTK's [2] POS tagger), and *adverbs*. (b) **Temporal References**: consisting of *past*, *present*, and *future* tenses. (c) **Social/Personal Concerns**: words belonging to *family*, *friends*, *social*, *work*, *health*, *humans*, *religion*, *bio*, *body*, *money*, *achievement*, *home*, and *sexual*. (d) **Interpersonal Awareness and Focus**: words that are *1st person singular*, *1st person plural*, *2nd person*, and *3rd person* pronouns.

(4) For social attributes we utilized a variety of content sharing, social interaction, and social support indicators. These are: *post length*, *number of comments*, *vote difference* (difference between upvotes and downvotes, divided by total upvotes and downvotes), *comment arrival rate* (average time difference between any two subsequent comments in a post's comment thread), *time to first comment* (time elapsed between the first comment and the timestamp of the corresponding post), and *median comment length*[9].

We compute each of the above linguistic measures of behavior at the post level – the value of a measure is given by the ratio of the number of words in a post that match words belonging to the measure, to the total number of words in the post. For each measure we take the average across all celebrities, to ensure each suicide event is equally weighted, i.e., to avoid skew due to a single suicide. For statistical comparison, we used the Welch t-test; a negative t-statistic value means the measure increased after suicide.

4.2.2 n-gram Analysis

We also present an analysis of the usage of various n-grams in posts shared succeeding and preceding celebrity suicides. Specifically, we focus on uni-, bi-, and tri-grams — we refer to them as n-grams throughout the paper. For comparison of the post and pre-celebrity suicide periods, we compute log likelihood ratios of each n-gram given as the logarithm of the ratio of the probability of occurrence of the n-gram in the post-suicide period, to the probability of the same n-gram over the pre-suicide period. Thus, when a n-gram is equally frequent, its log likelihood ratio will be zero; it would be greater than 1 if it is more frequent post celebrity suicides, whereas less than 1 if the reverse.

4.2.3 Topic Model Analysis

Our final content analysis method used a topic model, which have been commonly employed to analyze health data [45, 37, 38]. We obtain topics by running Latent Dirichlet Allocation (LDA) [3]

[9]Although comments in Reddit can be nested, we do not consider the nesting structure, rather focus on all comments associated with a post ordered by their timestamps.

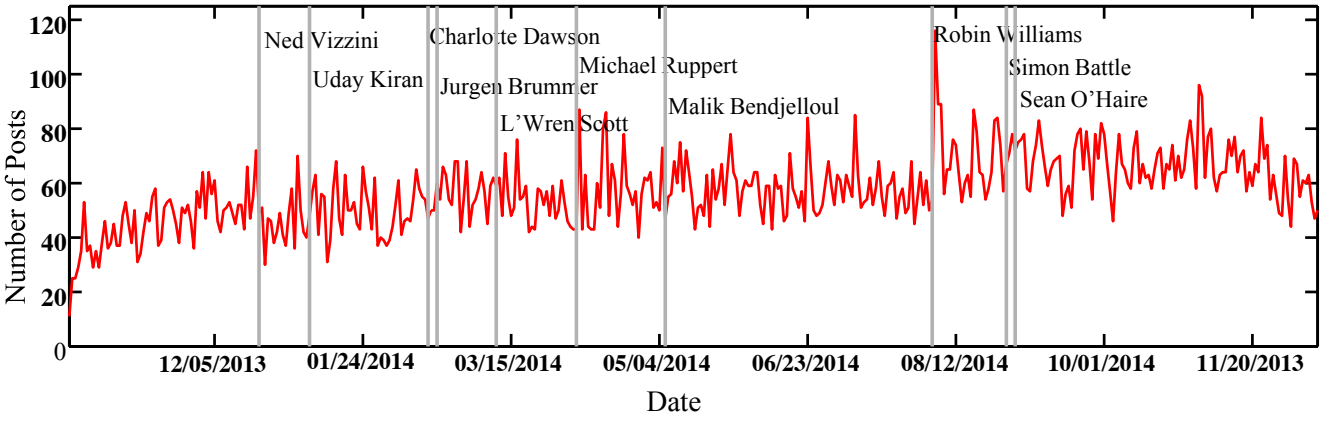

Figure 2: Number of posts on SW with markers of celebrity suicides.

	Actual	Baseline				Control	
Celebrity	**Change in z-score**	**Change in z-score**	**t-stat.**	**Std. Err.**	**p**	**Change in z-score**	**Std. dev.**
Ned Vizzini	-6.2247	4.8708	-31.362	2.5018	$< p^{-10}$	0.7822	±2.50
Uday Kiran	5.6022	3.1108	3.7312	4.7215	$< .001$	2.5407	±4.72
Charlotte Dawson	7.5475	1.9654	8.8301	4.4706	$< p^{-10}$	-0.8699	±1.47
Jurgen Brummer	5.6801	0.8948	7.5050	4.5085	$< p^{-10}$	-1.2648	±2.50
L'Wren Scott	-2.1786	2.2798	-7.0382	4.4792	$< p^{-8}$	0.4181	±2.47
Michael Ruppert	7.8587	1.7304	8.4508	5.1277	$< p^{-10}$	-0.1192	±1.12
Malik Bendjelloul	-0.7002	1.7444	-3.5361	4.8887	$< .001$	-0.9398	±4.88
Robin Williams	15.561	0.6006	22.842	4.6314	$< p^{-10}$	1.6456	±3.63
Simone Battle	4.0460	1.3554	4.2296	4.4981	$< .001$	-0.4795	±2.49
Sean O'Haire	-0.7781	0.9866	-2.6803	4.6554	$< .01$	2.6393	±1.65
Mean	3.6414	1.9539				0.4353	

Table 4: *Left:* Actual change in z-scores of post volume over a two-week period following each celebrity suicide, compared to the same period before. *Center:* Paired t-tests comparing z-score changes after each celebrity suicide and that over 20 similar windows with no celebrity death (baseline). *Right:* changes compared to the 21 mental health subreddits (control).

over the combined set of posts shared in a two week period preceding and succeeding the celebrity suicide events. We use Gensim's [46] implementation of online LDA from [21]. We use the default hyper-parameter settings and 50 topics, which we found to work well in initial experiments. To measure topic increases in post-suicide posts we first compute the posterior probability of each topic separately for the pre-suicide and post-suicide posts. We then compute the rate of increase for each topic as the difference between the posterior probability post-suicide and pre-suicide, divided by the probability of topic pre-suicide.

5. RESULTS

5.1 Measuring Post Volume

Figure 2 shows trends in posting activity on SW (raw frequencies of posts), overlaid with the times of reported celebrity suicides. To quantify changes in the aftermath of the celebrity suicides, we use z-scores of the number of posts in the pre celebrity suicide and post celebrity suicide two week periods[10]; change per celebrity is given by the difference between post-suicide z-scored #posts and the pre-suicide z-scored #posts (Figure 3). Posting activity increased after most celebrity suicides; combining all suicides, we find a strong increase of SW posts (Table 4).

[10]Use of z-scores for the daily volume of posts helps us account for an observed gradual increase in the number of posts over time.

Compared to the observed changes in the empirical background distribution (§4.1.1) we find that the actual change combined across all 10 celebrities is 3.64 (mean difference between post-celebrity suicide z-score of posts and pre-suicide z-score across all celebrities), whereas, although positive, the same change is nearly half the mean baseline change (1.95) (Table 4). Further, paired t-tests indicate that the change associated with each celebrity suicide event is significantly different from the corresponding baselines ($p < .001$). Thus the observed increase in SW posts is unlikely due to a random fluctuation in posting volume.

Finally, comparing the z-score posting change of SW (3.64) to the MH subreddits control group (0.43) we find that these increases are specific to SW (Table 4). These three findings support the manifestation of the Werther Effect in the forum.

5.2 Content Analysis

5.2.1 Linguistic Measures

Table 5 summarizes the linguistic measures of behavior derived from SW content, comparing the two week period following celebrity suicides to a period of same length before.

(1) Affective Attributes: *Post-suicide content is more negative, angry, sad, and anxious*. Prior literature indicates increased negative affect to be associated with increased emotional vulnerability such as mental instability, helplessness, loneliness, and restlessness [59].

(2) Cognitive Attributes: *Post-suicide content shows greater cog-*

Category	Pre-	Post-	t-stat.	p
Affective				
negative emotion	0.0317	0.0587	-5.821	**
anger	0.0121	0.0656	-7.431	**
anxiety	0.0082	0.0348	-8.628	**
sad	0.0122	0.0219	-3.276	*
Cognitive				
negation	0.0311	0.0573	-4.234	*
certainty	0.0409	0.0170	3.094	*
inhibition	0.0533	0.0042	7.388	**
death	0.0173	0.0925	-7.942	**
feel	0.0013	0.0092	-5.845	**
Linguistic Style				
Lexical Density				
nouns	0.2032	0.1145	6.934	**
adjectives	0.1263	0.0668	5.437	**
adverbs	0.0362	0.0121	1.696	*
Temporal Ref.				
past tense	0.0937	0.1733	-6.374	**
future tense	0.2478	0.1035	9.852	**
Social/Personal Concerns				
social	0.1841	0.0686	5.268	**
work	0.0782	0.0440	2.364	*
humans	0.0191	0.0176	1.257	*
Interpersonal Awareness				
1st pp singular	0.1184	0.3615	-11.548	***
2nd pp	0.0445	0.0122	8.270	**
3rd pp	0.0729	0.0083	8.482	**
Social				
post length	107.37	324.02	-14.521	***
#comments	6.3914	10.757	-7.431	**
vote difference	5.1800	1.7526	3.864	*
comm. arrival rate	12.423	6.699	8.548	***
med. comment len.	37.236	86.935	-10.448	***

Table 5: Welch t-test between content preceding and succeeding celebrity suicides; negative values mean the measure increased after suicide. We report results for α=.05(*), .01(**), .001(***). p values are adjusted after adopting the Bonferroni correction (number of measures m=49).

nitive biases. Posts are less certain, show increased negation, and use more perception centric words, such as words in the category 'feel'. The psychology literature indicates such cognitive biases to be associated with lower emotional stability and increased self-consciousness [44, 25, 12]. Additionally, *post-suicide content has more death related conversation and shows lowered inhibition*. These are known to be associated with greater health concerns as well as suicidal thoughts [58, 7]. Lowered inhibition also indicates an increased tendency for self-disclosure [42, 57] – suicidal ideation or suicide interest, lowered self-esteem and display of self-derogatory thoughts are extreme forms of self-disclosure and inhibition.

(3) Linguistic Style Attributes: *Post-suicide content has lower lexical density*. Greater mental health challenges and suicidal tendency is known to show this characteristic – such content is mostly about the self, hence people attribute less to things, happenings or people around them [60, 11]. The literature also associates lower lexical density to high drive states, which are typical in suicidal individuals [35]. *Post-suicide content is less concerning about the future and more fixated on the past* – likely due to the manifestation of suicidal tendencies [55]. Lowered future orientation is a known attribute of negative attitude towards one's own life and actions. Additionally, *post-suicide content shows little social and personal concerns*. It is known that suicidal thoughts are accompanied by thoughts about the self and self-occupation, hence they are less likely to talk about words relating to social, work, and humans [7].

(4) Social Attributes: *Post-suicide content shows more use of first person singular pronouns and fewer second and third person pronouns*. This suggests that posters are less socially concerned or bothered. Suicidal ideation is associated with *greater self-attentional focus* [57, 60, 12]. In fact, together with the fact that they also exhibit fewer social and personal concerns, it is likely that the post-suicide cohort in SW share more personal stories and in general, high self-preoccupation [4]. *Post-suicide content is longer* – literature on self-disclosure on stigmatized topics, such as suicide interest, shows that greater self-disclosure is associated with longer and more verbose content [20, 23, 11]. Further, the SW community seems to provide more support through a greater number of and longer comments, likely because of their high degrees of expressed vulnerability. However, they get fewer upvotes – suicide interest or ideation related content are unlikely to garner positive approval. People also tend to comment faster – we presume due to the increased sensitive content, the community volunteers to provide help and advice quickly.

5.2.2 n-gram Comparison

Going beyond the linguistic measures, we now investigate whether and how usage of n-grams change following celebrity suicides. Per section 4.2.2, we obtained frequencies and log likelihood ratios of all uni-, bi-, and tri-grams from the SW posts in the two weeks following the suicides and also that in the same period before.

An analysis of the n-grams that exhibit the greatest changes pre- and post- celebrity suicide demonstrate a similar trend (Section 4.2.2). We also measured the adjusted mutual information between the frequencies of occurrence of the n-grams (frequencies > 50) after celebrity suicides and that before: yielding a score of 0.21. This indicates that there is little correlation in terms of frequencies of these n-grams before and after celebrity suicide reports. Further, a Welch t-test informs that this difference to be statistically significant – the t-statistic is found to be 3.7 ($p < .001, df = 6892$).

Table 6 gives a list of 75 n-grams and their associated log likelihoods — we present them in three categories, 25 n-grams each with highest and lowest log likelihoods, and most frequent 25 n-grams with log likelihood zero. The first column indicates those n-grams more frequent in the post celebrity suicide period versus before, the second indicates the reverse, and the third column is a set of n-grams which equally co-occur in both categories.

Our findings align with those from the previous subsection — qualitative inspection of the n-grams in the first column suggest SW content shifts to a more vulnerable tone following the celebrity suicides (first column) compared to before (second column). We organize the inspected themes into various broad categories in line with literature on analysis of content in suicide notes [35, 51, 57].

In the aftermath of the celebrity suicides, we find evidence of expression of anxiety and depression, sense of guilt and regret ("i hate it", "piece of shit", "hate myself so", "give a shit", "without me. i"), hopelessness ("i gave up", "i ended"), sorrow ("alone i", "leave me"), and explicit desire to end their life ("'to hang myself", "wanting to kill", "of suicide", "tired of living"). During this period we also observe heightened conflicting thoughts ("but right now", "i'm probably"), and a sense of urgency and help seeking ("really just want", "to let me", "help me").

On the other hand, n-grams that are more frequent in the period before the celebrity suicides tend to be less sensitive or vulnerable. In many cases use of n-grams like "be happy. i", "a good person", "be happy", "hope" indicate a tendency for individuals to strive towards maintaining a positive spirit. We also observe greater intent to seek help from the SW community — "if anyone has", "feel free to", "want people to", "tell my parents". Mentions of close friends

n-gram (post>pre)	LLR	n-gram (pre>post)	LLR	n-gram (pre=post)	LLR
few weeks ago	0.717	i feel ,	-0.780	thinking of	0
leave me	0.712	be happy. i	-0.675	they would n't	0
days i	0.693	hope	-0.571	a therapist ,	0
of suicide ,	0.693	my life with	-0.511	phone	0
without me. i	0.673	a cry for	-0.375	badly	0
anxiety and depression	0.634	even worse	-0.375	that point	0
hate myself so	0.631	a good person	-0.346	any kind of	0
tired of living	0.623	i feel bad	-0.288	we 're all	0
really just want	0.623	tired of the	-0.288	i 've just	0
i ended up	0.620	if anyone has	-0.323	i 've struggled	0
alone , i	0.619	thought it would	-0.465	i 'll have	0
i 'm extremely	0.619	friends , no	-0.464	cope with	0
i 'm probably	0.609	tell my parents	-0.452	call me	0
to let me	0.609	family , i	-0.448	i believe i	0
i gave up	0.606	she says she	-0.442	at times	0
to hang myself	0.598	my father is	-0.421	miss me ,	0
i 've lived	0.598	people , but	-0.419	i feel really	0
give a shit	0.575	a friend of	-0.386	reason to live.	0
wanting to kill	0.574	my dad and	-0.533	dont even know	0
help me ,	0.573	no one would	-0.357	feel bad for	0
i ended	0.569	be happy.	-0.357	unhappy	0
but right now	0.566	to her. i	-0.357	put up	0
piece of shit.	0.565	want people to	-0.343	helped me	0
year and a	0.563	feel free to	-0.343	love me and	0
i hate it	0.556	so hard for	-0.323	going through with	0

Table 6: Log likelihood ratios (LLR) of n-grams more frequent in posts after celebrity suicides (left); those of n-grams more frequent in posts before celebrity suicides (center); and those equally frequent in both sets (right).

and family are also widely common ("friends, no", "family, i", she says she", "my father is", "people, but", "my dad and", "to her. i"). To contrast with the n-gram usage following the suicides, it thus seems that individuals are typically keen to derive help and support from the SW community, however this reduces significantly in the short time period following the celebrity suicides.

Compared to the above two categories of n-grams, we observe that certain types of content are prevalent in SW posts irrespective of the time of reports of celebrity suicides. These include reflection ("thinking of", "don't even know", they wouldn't", "we're all", "reason to live"), illustration of one's experiences of coping with distress ("i've struggled", "cope with", "going through with", "a therapist ,"), and negative tone ("badly", "unhappy", feel bad for"). We also observe some n-grams suggesting call for help ("call me", "helped me", "phone"). This indicates the manner in which the broader SW community might be catering to some of the prime needs of this population in a consistent fashion irrespective of external agents or events.

5.2.3 Topic Model Analysis

Finally, we provide a topic analysis to determine how topics discussed in SW prior to celebrity suicides differ from those following the suicides. For the purpose, we use the method described in section 4.2.3. From Figure 4, we observe that the mean (absolute) change across all topics is 2.34% ($\pm$9.16%) in the two-week period succeeding the suicides compared to that before. Specifically, we observe that topics #4, 50, 30, 8, 17 (five topics with highest absolute change) show notable changes in use in the post-celebrity suicide period, compared to that preceding the suicides. Now we analyze change in the content of these topics, like with n-grams, based on prior literature on analysis of suicide notes [35, 51, 57]. We present top changing five topics along with a set of their representative words that capture the essence of the topics.

(1) *lost, useless, done, poisoning, alone, fucking, angry, shitty, hate, suffocate, damn*: This topic (#4) describes self-derogatory and self-critical thoughts relating to self destruction (increase: 13%).

(2) *suicidal, sorry, lifeless, kill, death, withdraw, horrible, anxiety, rough, afraid, hotline, flashes, numb, scars, harsh, scared*: This topic (#50) manifests confessions and regrets of individuals and their desire to commit suicide, particularly via distress expression (increase: 7%).

(3) *maybe, though, probably, except, however, if, but, admit, reason, finally, support, wanted, hurt, bad, sure, nothing, pain, best*: This topic (#30) illustrates the struggling and conflicting feelings that are often known to ensue suicidal ideation [51] — cognitive functioning under competing motives, e.g., self-criticism vs. self-protection, aggression vs. affection towards others (increase: 6%).

(4) *talk, help, ask, emergency, advice, happy, sober, smile, trying, comforted, believe, promise, hopeful, love, sure, friend, family, relationship, parents, school, domestic, guy, undergrad*: This topic (#8) mixes requests for help, including constructs related to demand, command, and request. It expresses needs of the individual and requires some behavior on the part of the listener for their satisfaction [20]. We conjecture such requests are a prime reason why individuals join the SW community. We also notice manifestation of positive and compassionate cognitive thoughts, as well as content around societal, practical and familial concerns. Alarmingly, this topic decreases after celebrity suicide (decrease: 6%).

(5) *really, anything, want, always, never, everything, unable, still, every, unwilling*: This topic (#17) reflects information under high drive or emotion and tend to be more extreme, polarized or ambivalent in their assertions [34]. This is indicated by use of terms that permit no exception (increase: 5%).

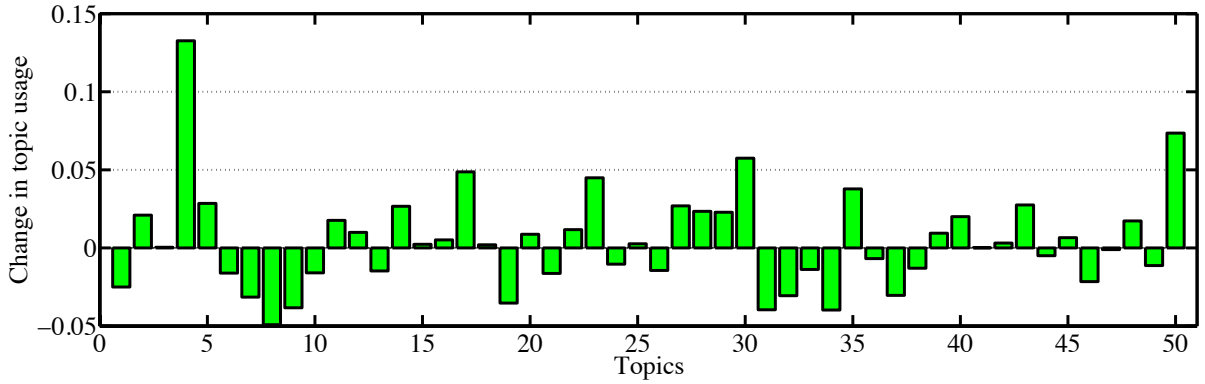

Figure 4: Changes in use of topics over two weeks succeeding a celebrity suicide compared to that over two weeks before. Positive change indicates value increased in post-suicide period.

6. DISCUSSION

Our empirical findings suggest that social media (Reddit in this work) contain subtle yet significant changes in language and activity around suicide content following celebrity suicide events. To the best of our knowledge, this is one of first studies examining the prevalence of the Werther effect centered around suicidal ideation manifested online.

However, we urge caution in deriving causal implications from our findings. While SW is a great resource for suicide related support, it is a platform wherein individuals discuss *why* they desire to end their lives, instead of a channel for communicating or sharing suicide notes. We cannot assess how increased activity on SW compares with increased attempted or completed suicides. In fact, participation in such online support groups may decrease suicide attempts. Nevertheless, our study provides insight into a previously difficult facet of suicide to explore — discussion and ideation.

Next, one presumed explanation of the observed SW post changes might be that the users are, in fact, discussing the celebrity suicides themselves. To examine the validity of this conjecture, we calculated the number of times each celebrity was mentioned — 8/10 celebrities were never mentioned, L'Wren Scott was mentioned once, and Robin Williams was mentioned 52 times. This suggests that the ensuing linguistic changes of post content are not attributable to discussion of the suicide events of the celebrities themselves.

Additionally, we note that one celebrity suicide is perhaps qualitatively different than the rest (as is evidenced in the preceding paragraph) – that of Robin Williams. To ensure that our observed effects were not driven by this singular event, we repeated our empirical investigation excluding his suicide; our findings remain true. For instance, even after disregarding this suicide, the changes in the period succeeding suicides compared to that before was significantly higher than the baselines and the control group (MH subreddits), although it decreased to some extent (32% decrease for the former, and 8% for the latter). The different LIWC categories continued to show significance, while the difference between the posterior probabilities of the pre-celebrity suicide and post-celebrity suicide topics was also distinct.

We also comment on the generalizability of our findings. We acknowledge that the results presented in this paper are limited to those individuals who participate in the SuicideWatch support forum on Reddit. It is possible there is a self-selection bias in this population. For instance, it is likely it is a set of individuals who are seeking help on this sensitive issue. Second, it is also a group who are choosing an *online* platform for seeking help, instead of, or in addition to other (offline) modalities of suicide support. Caution is advised given the known bias of Reddit user base as well—the average redditor is a 20-something male[11], and perhaps more "tech-savvy" than the general set of individuals contemplating suicide. Finally, even though some of the linguistic and content cues that we mined are Reddit-specific (language characteristic of the platform's culture), our methods could be extended to other social media platforms as well—especially to those online platforms which possess similar attributes implicitly or explicitly, and which allow sharing of textual content.

Implications of Findings. There are many practical implications of this work. Community moderators may develop strategies for the aftermath of celebrity suicides, allowing them and other interested/committed volunteers to be more proactive with the help they offer, as well as provide some unstated context relevant to support during the period immediately following a celebrity suicide report. Awareness of this phenomenon on SW will also allow moderators and volunteers to pay specific attention to redditors who show increased signs of suffering exacerbated by the suicide event. Additionally, individuals whose content contain phrases and other linguistic constructs of high suicidality may be connected to community members who have volunteered to provide help and support—social support and higher levels of social capital can help individuals fight such vulnerable tendencies. Thus, adequately deployed interventions following celebrity suicides can actually motivate potentially suicidal individuals to steer away from their final decision, and encourage them to continue living.

Ethics. Finally, we identify the ethical challenges of this line of research. The interventions outlined above need to honor the privacy of the individuals and those who volunteer to provide help and support. Further, beyond the design suggestions outlined above, *how* to actually intervene, deploy, and offer support to individuals of high likelihood of suicide is a research and ethical question of its own. Especially given the (semi)-anonymous ecosystem that Reddit's SW forum provides to this sensitive population, we need to inculcate utmost care in the manner in which help and support are catered to them following the evidence of existence of the Werther Effect. Unthoughtful interventions may actually lead to counter-helpful outcomes—for instance, chilling effects in participation in the community, or suicide ideation related expression moving on to other alternative or peripheral platforms online where such populations might be difficult to discover and therefore extend help to. Broadly, the ethical dimensions of interventions and their deploy-

[11]http://www.pewinternet.org/files/old-media/Files/Reports/2013/PIP_reddit_usage_2013.pdf

ment need to ensure that communities like the SW can continue to be safe and powerful platforms for seeking help, advice, and support around suicidal tendencies, and for online psychotherapy.

7. CONCLUSION

We have presented preliminary findings on the manifestation of the Werther effect in the prominent Reddit suicide support forum r/SuicideWatch. We found significant changes manifested following reports of several celebrity suicides – impacting the frequency of posting activity as well as the nature of content shared. Our findings are among the first to demonstrate the Werther effect on suicidal ideation and have implications for building suicide prevention interventions following high-profile suicides.

There are several interesting directions to future work. Examination of other suicide support forums and social media, observations over longer periods of time and more celebrities and across varied cultural contexts will help us generalize our findings on presence of Werther Effect online. We also intend to investigate to what extent Werther Effect in social media relates to actual changes in completed (government-reported) suicide numbers in regions where the Reddit platform is widely adopted. Causal relationships between the change observed following celebrity suicides and the reporting of these suicide events can be better inferred through a predictive setting, which also constitutes a promising future research direction. Finally, given our observation that more prominent suicidal thoughts are expressed in posts succeeding the celebrity suicides, automated or semi-automated suicide ideation detectors may be developed using machine learning approaches, that can be used to bring timely help and support to these vulnerable communities.

ACKNOWLEDGEMENTS

We thank Shawndra Hill for suggesting the idea of investigating the Werther Effect in social media. De Choudhury was partly supported by a National Institutes of Health grant #1R01GM11269701.

8. REFERENCES

[1] A. G. Billings and R. H. Moos. Coping, stress, and social resources among adults with unipolar depression. *Journal of personality and social psychology*, 46(4):877, 1984.

[2] S. Bird. NLTK: the natural language toolkit. In *COLING/ACL Interactive presentation sessions*, 2006.

[3] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research (JMLR)*, 3:993–1022, 2003.

[4] A. Boals and K. Klein. Word use in emotional narratives about failed romantic relationships and subsequent mental health. *Journal of Language and Social Psychology*, 24(3):252–268, 2005.

[5] A. T. Cheng, K. Hawton, T. H. Chen, A. M. Yen, C.-Y. Chen, L.-C. Chen, and P.-R. Teng. The influence of media coverage of a celebrity suicide on subsequent suicide attempts. *Journal of Clinical Psychiatry*, 68(6):862–866, 2007.

[6] A. Chester and A. O'Hara. Image, identity and pseudonymity in online discussions. *International Journal of Learning*, 13(12), 2007.

[7] C. Chung and J. W. Pennebaker. The psychological functions of function words. *Social Comm.*, pages 343–359, 2007.

[8] G. Coppersmith, M. Dredze, and C. Harman. Quantifying mental health signals in twitter. In *ACL Workshop on Computational Linguistics and Clinical Psychology*, 2014.

[9] G. Coppersmith, C. Harman, and M. Dredze. Measuring post traumatic stress disorder in twitter. In *International Conference on Weblogs and Social Media (ICWSM)*, 2014.

[10] M. De Choudhury, S. Counts, E. Horvitz, and A. Hoff. Characterizing and predicting postpartum depression from facebook data. In *Computer-Supported Cooperative Work and Social Computing (CSCW)*, 2014.

[11] M. De Choudhury and S. De. Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In *International Conference on Weblogs and Social Media (ICWSM)*, 2014.

[12] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz. Predicting depression via social media. In *International Conference on Weblogs and Social Media (ICWSM)*, 2013.

[13] M. De Choudhury, M. R. Morris, and R. W. White. Seeking and sharing health information online: Comparing search engines and social media. In *Human factors in computing systems (CHI)*, 2014.

[14] K. Dinakar, B. Jones, H. Lieberman, R. Picard, C. Rose, and M. T. R. Reichart. You too?! mixed initiative lda story-matching to help teens in distress. In *International Conference on Weblogs and Social Media (ICWSM)*, 2012.

[15] R. M. Fernquist and P. Cutright. Societal integration and age-standardized suicide rates in 21 developed countries, 1955–1989. *Social Science Research*, 27(2):109–127, 1998.

[16] K.-w. Fu and C. Chan. A study of the impact of thirteen celebrity suicides on subsequent suicide rates in south korea from 2005 to 2009. *PloS one*, 8(1):e53870, 2013.

[17] K.-w. Fu and P. S. Yip. Long-term impact of celebrity suicide on suicidal ideation: Results from a population-based study. *Journal of Epidemiology and Community Health*, 61(6):540–546, 2007.

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

[19] M. S. Gould. Suicide and the media. *Annals of the New York Academy of Sciences*, 932(1):200–224, 2001.

[20] S. H. Hemenover. The good, the bad, and the healthy: Impacts of emotional disclosure of trauma on resilient self-concept and psychological distress. *Personality and Social Psychology Bulletin*, 29(10):1236–1244, 2003.

[21] M. Hoffman, F. R. Bach, and D. M. Blei. Online learning for latent dirichlet allocation. In *Neural Information Processing Systems (NIPS)*, 2010.

[22] C. M. Homan, N. Lu, X. Tu, M. C. Lytle, and V. Silenzio. Social structure and depression in trevorspace. In *Computer-Supported Cooperative Work and Social Computing (CSCW)*, 2014.

[23] D. J. Houghton and A. N. Joinson. Linguistic markers of secrets and sensitive self-disclosure in twitter. In *Hawaii International Conference on System Science (HICSS)*, 2012.

[24] S. Jamison-Powell, C. Linehan, L. Daley, A. Garbett, and S. Lawson. I can't get no sleep: discussing# insomnia on twitter. In *Human factors in computing systems (CHI)*, 2012.

[25] N. Lapidot-Lefler and A. Barak. Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition. *Computers in human behavior*, 28(2):434–443, 2012.

[26] M. A. Long. As if day had rearranged into night: suicidal tendencies in the poetry of anne sexton. *Lit Psychol*, 39:26–41, 1993.

[27] D. D. Luxton, J. D. June, and J. M. Fairall. Social media and suicide: A public health perspective. *American Journal of*

Public Health, 102(S2):S195–S200, 2012.

[28] T. R. Lynch, J. Cheavens, J. Q. Morse, and M. Rosenthal. A model predicting suicidal ideation and hopelessness in depressed older adults: The impact of emotion inhibition and affect intensity. *Aging & Mental Health*, 8(6):486–497, 2004.

[29] M. A. Moreno, D. A. Christakis, K. G. Egan, L. N. Brockman, and T. Becker. Associations between displayed alcohol references on facebook and problem drinking among college students. *Archives of Pediatrics & Adolescent Medicine*, 166(2):157–163, 2011.

[30] E. L. Murnane and S. Counts. Unraveling abstinence and relapse: smoking cessation reflected in social media. In *Human factors in computing systems (CHI)*, 2014.

[31] T. Niederkrotenthaler, K.-w. Fu, P. S. Yip, D. Y. Fong, S. Stack, Q. Cheng, and J. Pirkis. Changes in suicide rates following media reports on celebrity suicide: a meta-analysis. *Journal of epidemiology and community health*, 66(11):1037–1042, 2012.

[32] T. Niederkrotenthaler, A. Herberth, and G. Sonneck. The "werther-effect": legend or reality? *Neuropsychiatrie: Klinik, Diagnostik, Therapie und Rehabilitation: Organ der Gesellschaft Osterreichischer Nervenarzte und Psychiater*, 21(4):284–290, 2006.

[33] T. Niederkrotenthaler, B. Till, N. D. Kapusta, M. Voracek, K. Dervic, and G. Sonneck. Copycat effects after media reports on suicide: A population-based ecologic study. *Social science & medicine*, 69(7):1085–1090, 2009.

[34] I. O'Donnell, R. Farmer, and J. Catalan. Suicide notes. *The British Journal of Psychiatry*, 163(1):45–48, 1993.

[35] C. E. Osgood and E. G. Walker. Motivation and language behavior: A content analysis of suicide notes. *The Journal of Abnormal and Social Psychology*, 59(1):58, 1959.

[36] M. Park, D. W. McDonald, and M. Cha. Perception differences between the depressed and non-depressed users in twitter. In *International Conference on Weblogs and Social Media (ICWSM)*, 2013.

[37] M. J. Paul and M. Dredze. You are what you tweet: Analyzing twitter for public health. In *International Conference on Weblogs and Social Media (ICWSM)*, 2011.

[38] M. J. Paul and M. Dredze. Discovering health topics in social media using topic models. *PLoS ONE*, 9(8), 2014.

[39] M. J. Paul, M. Dredze, and D. Broniatowski. Twitter improves influenza forecasting. *PLOS Currents Outbreaks*, 2014.

[40] J. W. Pennebaker, M. E. Francis, and R. J. Booth. Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates*, 71:2001, 2001.

[41] J. W. Pennebaker, T. J. Mayne, and M. E. Francis. Linguistic predictors of adaptive bereavement. *Journal of personality and social psychology*, 72(4):863, 1997.

[42] K. Petrie and R. Brook. Sense of coherence, self-esteem, depression and hopelessness as correlates of reattempting suicide. *British Journal of Clinical Psychology*, 31(3):293–300, 1992.

[43] J. E. Pirkis, P. M. Burgess, C. Francis, R. W. Blood, and D. J. Jolley. The relationship between media reporting of suicide and actual suicide in australia. *Social science & medicine*, 62(11):2874–2886, 2006.

[44] D. W. Prezant and R. A. Neimeyer. Cognitive predictors of depression and suicide ideation. *Suicide and Life-Threatening Behavior*, 18(3):259–264, 1988.

[45] K. W. Prier, M. S. Smith, C. Giraud-Carrier, and C. L. Hanson. Identifying health-related topics on twitter. In *Social computing, behavioral-cultural modeling and prediction*, pages 18–25. Springer, 2011.

[46] R. Řehůřek and P. Sojka. Software Framework for Topic Modelling with Large Corpora. In *LREC Workshop on New Challenges for NLP Frameworks*, 2010.

[47] S. Rude, E.-M. Gortner, and J. Pennebaker. Language use of depressed and depression-vulnerable college students. *Cognition & Emotion*, 18(8):1121–1133, 2004.

[48] T. D. Ruder, G. M. Hatch, G. Ampanozi, M. J. Thali, and N. Fischer. Suicide announcement on facebook. *Crisis: The Journal of Crisis Intervention and Suicide Prevention*, 32(5):280–282, 2011.

[49] A. Sadilek, H. A. Kautz, and V. Silenzio. Modeling spread of disease from social interactions. In *International Conference on Weblogs and Social Media (ICWSM)*, 2012.

[50] H. A. Schwartz, J. Eichstaedt, M. L. Kern, G. Park, M. Sap, D. Stillwell, M. Kosinski, and L. Ungar. Towards assessing changes in degree of depression through facebook. In *Association for Computational Linguistics (ACL)*, 2014.

[51] E. S. Shneidman. Suicide notes reconsidered. *Psychiatry*, 36(4):379–394, 1973.

[52] E. S. Shneidman and N. L. Farberow. Clues to suicide. *Public Health Reports*, 71(2):109, 1956.

[53] M. A. Silverman and N. P. Will. Sylvia plath and the failure of emotional self-repair through poetry. *The Psychoanalytic Quarterly*, 1986.

[54] M. Sisask and A. Värnik. Media roles in suicide prevention: a systematic review. *International journal of environmental research and public health*, 9(1):123–138, 2012.

[55] J. M. Smyth. Written emotional expression: effect sizes, outcome types, and moderating variables. *Journal of consulting and clinical psychology*, 66(1):174, 1998.

[56] S. Stack. Media coverage as a risk factor in suicide. *Journal of epidemiology and community health*, 57(4):238–240, 2003.

[57] S. W. Stirman and J. W. Pennebaker. Word use in the poetry of suicidal and nonsuicidal poets. *Psychosomatic Medicine*, 63(4):517–522, 2001.

[58] J. Suler. The online disinhibition effect. *Cyberpsychology & behavior*, 7(3):321–326, 2004.

[59] Y. R. Tausczik and J. W. Pennebaker. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1):24–54, 2010.

[60] K. M. Thomas and M. Duke. Depressed writing: Cognitive distortions in the works of depressed and nondepressed poets and writers. *Psychology of Aesthetics, Creativity, and the Arts*, 1(4):204, 2007.

[61] I. M. Wasserman. Imitation and suicide: A reexamination of the werther effect. *American sociological review*, pages 427–436, 1984.

[62] H.-H. Won, W. Myung, G.-Y. Song, W.-H. Lee, J.-W. Kim, B. J. Carroll, and D. K. Kim. Predicting national suicide numbers with social media data. *PloS one*, 8(4):e61809, 2013.

[63] P. S. Yip, K.-W. Fu, K. C. Yang, B. Y. Ip, C. L. Chan, E. Y. Chen, D. T. Lee, F. Y. Law, and K. Hawton. The effects of a celebrity suicide on suicide rates in hong kong. *Journal of affective disorders*, 93(1):245–252, 2006.