



Using Topic Modeling to Detect and Describe Self-Injurious and Related Content on a Large-Scale Digital Platform

PETER J. FRANZ, AM, ERIK C. NOOK, AM, PATRICK MAIR, PhD AND MATTHEW K. NOCK, PhD

Objective: Self-injurious thoughts and behaviors (SITBs) are a complex and enduring public health concern. Increasingly, teenagers use digital platforms to communicate about a range of mental health topics. These discussions may provide valuable information that can lead to insights about complex issues like SITBs. However, the field of clinical psychology currently lacks an easy-to-implement toolkit that can quickly gather information about SITBs from online sources. In the present study, we applied topic modeling, a natural language processing technique, to identify SITBs and related themes online, and we validated this approach using human coders.

Method: We separately used topic modeling software and human coders to identify themes present in text from a popular online Internet support forum for teenagers. We then determined the degree to which results from the software's topic model aligned with themes identified by human coders.

Results: We found that topic modeling detected SITBs and related themes in online discussions in a way that accurately distinguishes between relevant and irrelevant human-coded themes.

Conclusions: This approach has the potential to drastically increase our understanding of SITBs and related issues discussed on digital platforms, as well as our ability to identify those at risk for such outcomes.

Self-injurious thoughts and behaviors (SITBs; including both suicidal thoughts/behaviors and nonsuicidal self-injury [NSSI]) are a major health concern worldwide. In the United States, suicide is the second leading cause of death among adolescents and young adults aged 15–34 (CDC, 2017). However, many people considering suicide do not disclose, or outright deny, having suicidal thoughts to others (Busch, Fawcett, & Jacobs, 2003). As a result, it is extremely difficult to

identify those who may be at risk for SITBs. Prior studies suggest that those engaging in SITBs might be more likely to discuss these matters in Internet communities than with mental health professionals, potentially because these communities are less stigmatizing than other sources of support (Burns, Davenport, Durkin, Luscombe, & Hickie, 2010; Joinson, 2004; Vogel, Wade, & Hackler, 2007). Indeed, adolescents who have engaged in SITBs tend to use the Internet

PETER J. FRANZ, ERIK C. NOOK, PATRICK MAIR, AND MATTHEW K. NOCK, Department of Psychology, Harvard University, Cambridge, MA, USA.

Address correspondence to Peter J. Franz, Harvard University, 33 Kirkland Street, Cambridge, MA 02138; E-mail: pfranz@g.harvard.edu.

more and have more close online relationships than adolescents who have not engaged in SITBs (Mitchell & Ybarra, 2007), and increased web-based peer communication is associated with increased risk for SITBs (Tseng & Yang, 2015).

Although the Internet and other digital media may provide useful methods for identifying and studying SITBs, the field currently lacks an efficient and accurate tool for rapidly identifying which individuals are discussing SITBs online. Such a tool would be advantageous for at least two reasons. First, it may be possible to use such a tool to identify individuals at high risk who could benefit from intervention and are unlikely to seek help through traditional conduits. Second, identifying online discussions of SITBs offers a first step in conducting research that would grant greater insight into the role of online peer-to-peer communication in SITB risk.

One promising tool for identifying online discussions of SITBs is topic modeling—a machine-learning method that identifies latent themes present in text. Topic modeling has been used to detect coherent themes frequently discussed by users of an Internet support forum (Carron-Arthur, Reynolds, Bennett, Bennett, & Griffiths, 2016). This approach revealed discussions of a range of mental health issues, such as depression, substance use, and anxiety. To date, little published research has examined SITB content in online support communities using topic modeling. Furthermore, no published research to our knowledge has carefully validated machine-learned topics relevant to SITBs using human coders.

Here, we report on the use of topic modeling to analyze text data from TeenHelp.org, an online forum in which adolescents and adults frequently discuss life stressors, including SITBs in order to address three research questions. First, can an easily implementable topic modeling approach detect SITB and related topics in online text? Second, how well do machine-learned topics correspond to topics identified by human coders? Third, does online SITB discussion vary with age and gender? As such, this project aims both to provide a tool for the rapid

and accurate detection of discussions concerning SITBs and to reveal how online SITB discussion relates to broad demographic factors.

METHOD

Data

Text data were acquired from TeenHelp.org, a popular Internet support forum, using the web scraping software BeautifulSoup (Richardson, 2016). This forum serves as an online support community for young people who are facing challenges in life. Users of the forum can freely post content and respond to others' posts, often creating a thread of reciprocal posts. The TeenHelp.org web site is divided into subforums spanning a wide range of topics. For this project, we scraped only the first post of each thread (i.e., we did not collect responses) from three of the website subforums, titled "Self-harm," "Depression and Suicide," and "Friends and Family." We chose the "Self-harm" and "Depression and Suicide" subforums because of our interest in capturing online discussions of SITBs. We collected discussions in the "Friends and Family" subforum in order to provide some variation in the themes discussed (i.e., non-SITB themes), while ensuring that these posts would be similar in their expression of distressing emotions/experiences as those posted within the other subforums. The Harvard University Institutional Review Board approved analysis of these data, and web site moderators provided permission for use.

Users

Users of TeenHelp.org are given the choice to self-identify their gender and age. For purposes of this study, we only analyzed the posted content of users with a "male" or "female" gender identification and whose age was 25 or under. The average age of the sample was $M = 19.27$ ($SD = 2.47$) and 2,052 (87.1%) were female.

Each user's post, referred to as a "document" in natural language processing research, is comprised of a set of words, referred to as "terms." We analyzed 2,359 posts. These posts contained an average of $M = 43.21$ terms ($SD = 42.99$). We did not collect user names and, thus, cannot confirm that each post was generated by a separate user.

Human Coding

The first author and two additional, trained human coders classified each post based on the primary issue discussed. The

primary issue was defined as the central topic of concern in each post. A list of 11 primary issues (NSSI, suicidal ideation, suicide attempt, passive suicidal ideation, suicide planning, depression, social concerns, abuse, medication, substance use/abuse, and other) was generated by reviewing approximately 500 random posts from the corpus and identifying the most common themes. Descriptions for each of these primary issues are presented in Table 1. Human coders assigned each post a single primary issue from this list of 11 primary issues. The primary issue, "other," was assigned to those posts that could not be classified under any of the alternative 10 primary

TABLE 1
List of Human-Coded Primary Issues

Primary issue	Definition	Example
Nonsuicidal self-injury (NSSI)	Discussion of injuring one's own body tissue (e.g., cutting, burning, hitting self) with no expressed or inferred intent to die	I promised my best friend I would stay clean this month and finish the year with no more cuts
Suicidal ideation	Discussion of taking one's own life	I can't stop thinking about killing myself. . . will this ever end?
Suicide attempt	Discussion of purposefully engaging in some harmful act with the intent to die as a result	Last month I was sent to hospital after a suicide attempt
Passive suicidal ideation	Discussion of the desire to die, but no description of a self-inflicted act	I just hope that as I go to sleep tonight I won't wake up, but I always do wake up
Suicide planning	Discussion of a plan for suicide, possibly, but not necessarily including method	The thoughts of suicide are becoming unbearable, I have come close many times to going through with my plan
Depression	Discussion of major depressive episodes or depressed mood	I am getting more depressed and I have no one to talk to about this. I feel so alone
Social concerns	Discussion of problems with family members, romantic partners, or friends	I've been together with my boyfriend since August 2012. We had our hard times but now I'm thinking we rushed it too fast
Abuse	Discussion of physical or emotional abuse	My mom makes my life miserable. She is always putting me down. She tells me I am a bitch, a drama queen, she tells me I am irresponsible and hard to trust
Medication	Discussion of medication usage, side effects, or experiences (especially related to psychotropic medications)	I just started Wellbutrin, it made me very hyper, awake and in a good mood today
Substance	Discussion of drug/alcohol use for nonprescribed purposes	I get high a lot, like every day, but I want to quit. It's been really tough
Other	Anything other than the above-listed six topics	–

issues. For purposes of the analyses that follow, suicidal ideation, passive suicidal ideation, suicide planning, and suicide attempt were collapsed into a single primary issue called “suicide.” Human coders achieved a moderate level of interrater reliability (Fleiss’ $\kappa = .66$).

Text Preprocessing

In line with canonical text mining practices, text from the corpus was transformed through several preprocessing steps to enable more efficient manipulation and analysis. This series of steps entailed splitting contractions into their constituent words, removing punctuation, removing numbers, converting text to lower case, removing excess spaces (in circumstances where a single space would be sufficient), and removing highly common terms (called “stop words”) that would not provide considerable benefit for differentiating between the topics of interest (e.g., “for,” “a,” “the,” “it,” “that,” and “to”). Details for these procedures have been described elsewhere (e.g., Lin & He, 2009; Zhao et al., 2011) and are important steps to produce interpretable topic modeling results. Following preprocessing, we created a document-term matrix (DTM) representing the frequency of terms across each of the documents in the corpus. We removed terms that occurred in less than 10% documents, resulting in a DTM containing the 144 most frequently used terms from the corpus. The threshold of 10% was set in order to ensure the term “suicide” remained. Four documents contained no terms in this final DTM and were thus removed. Topic modeling was then performed on this DTM, comprised of 2,355 documents by 144 terms.

Topic Modeling

We used the *topicmodels* package in R (Grün & Hornik, 2011) to identify and quantify the topics, or themes, discussed by users of TeenHelp.org. Topic modeling is a natural language processing approach that identifies latent thematic structures present in text-

based data sets that would be impossible, or at least very costly for human coders to parse. Topic modeling views each document in a corpus as a mixture of different themes, or topics. This set of topics can then provide insight about the themes discussed within the corpus.

To illustrate topic modeling as an approach to understanding themes in text, let us explain how it could be applied to the articles in an issue of the journal *Suicide and Life-Threatening Behaviors*. Human readers of this journal can detect prominent themes—or topics—that are discussed in this issue, such as *Prediction* and *Treatment*. However, topic modeling can be used to both (1) identify the topics discussed in articles based on clusters of commonly co-occurring terms and (2) estimate how much each article discusses each topic. For example, a human could intuitively categorize an article that uses the terms “risk,” “longitudinal,” “predicts,” and “future” as discussing the topic *Prediction*. Similarly, articles using the terms “outcome,” “intervention,” “CBT,” and “improved” are likely to be categorized as discussing the topic of *Treatment*. Topic modeling merely trains a computer algorithm to make this classification without human input. This occurs in two steps: (1) discovering the latent structure of topics that are defined by terms that tend to occur in close proximity to each other (the exact number of topics is set by the researcher) and (2) quantifying the extent to which terms that define each topic are contained within each article. As such, a paper that examined demographic risk factors that predict suicide attempts would likely contain a high proportion of the topic *Prediction* and a lower proportion of the topic *Treatment*, but the opposite might be true for a paper that examined cross-sectional correlates of suicidal ideation in a group of patients undergoing treatment for major depressive disorder.

In this project, we applied topic modeling to comments posted on TeenHelp.org to (1) identify latent topics that were discussed in these 2,355 posts and (2) score each post based on its discussion of each topic. In particular, we were interested in testing how well the topic modeling approach’s assessment of

suicidal content correlated with human coders' perception of suicidal content in each post.

We employed the Latent Dirichlet Allocation (LDA) topic modeling algorithm. This approach utilizes a probabilistic generative statistical model to infer specific topics based on terms that tend to cluster together, as well as the extent to which topics are present in specific documents (Blei, Ng, & Jordan, 2003). LDA and similar topic modeling algorithms have been used for the purpose of quickly and accurately detecting an underlying thematic structure in text documents across a wide range of contexts. These approaches have been utilized to compare topics discussed on Twitter to those of printed news media (Zhao et al., 2011), to group medications based on topics found in the text of "warnings" printed on manufacturers' drug labels (Bisgin, Liu, Fang, Xu, & Tong, 2011), and to identify scientific articles in a large archive that are relevant to researchers' specific interests (Wang & Blei, 2011). Of greater relevance to our study, topic modeling has more recently been used to identify mental health-related content in various text-based sources. For example, studies have used topic modeling to predict current psychological health of adults from essays written in childhood (Lynn et al., 2017), identify depression-related content in social media posts (Resnik et al., 2015), and identify specific topics discussed by users of the emergency text support service, Crisis Text Line (Dinakar, Lieberman, Chaney, & Blei, 2014).

It is important to mention that in LDA, the model's output is created by a random process that iteratively generates and refines its own solution in order to find a statistical solution that best fits the data (Blei et al., 2003). The random nature of this process is important in order to generate unbiased results. However, because of the randomness inherent in topic modeling, exact replication of a topic model's output, even using the same data, would be unlikely without the exact code used to generate that model.

Additionally, the results presented below were acquired through an entirely

unsupervised modeling approach. In other words, we did not utilize a "training set" of data, which would enable the researcher to have direct input for the topic modeling process. As we will show, unsupervised modeling of these text data returned an easily interpretable topic solution.

Statistical Analyses

We used a series of logistic regressions to examine the relationship between LDA topics and the four human-coded issues most relevant to our research questions (Suicide, NSSI, Social Concern, and Depression). For each regression, we dichotomized the human-coded issue as the outcome of interest (absent vs. present; 0, 1). Thus, in each logistic regression, we model the likelihood of a given human-coded primary issue, relative to all other primary issues, as a function of the corresponding LDA topic. For example, in the model examining the relationship between the Friends LDA topic and the human-coded primary issue of NSSI, all posts that were human-coded as NSSI were assigned a value of "1," whereas all other posts were assigned a "0." Odds ratios representing the association between each document's LDA topic probability estimates for specific topics and odds of a post being human-coded as a given issue were log-transformed in order to reduce the overall range of values. We used bivariate correlations to examine the relationships between LDA topics and self-reported age, and *t* tests to examine the relationships between LDA topics and gender.

RESULTS

Can Topic Modeling Detect SITB and Related Themes in Online Content?

We examined 10-, 15-, and 20-topic LDA models and determined that the 15-topic model provided enough specificity to detect coherent topics that represented constructs of interest (e.g., suicide, NSSI, depression, and social issues) and also minimized the

number of uninterpretable topics. The full 15-topic model and the top 20 terms with the highest probabilities of occurring in all 15 topics are included in Table 2. The model identified six topics that were relevant to the hypotheses for our study, including *Suicide*, *Friends*, *Family*, *Depression*, *NSSI*, and *Social*.

Each of these six topics was defined by a coherent set of terms that represent the theme of the topic. For example, the *Suicide* topic includes the terms, “want,” “life,” “hate,” “die,” “live,” and “kill.” The remaining nine topics included four that were intuitively interpretable as *Past*, *Time*, *Negative Affect*, and *Effort* based on the terms shown in Table 2. However, the final five topics were

less-intuitively interpretable, which often occurs as a result of the latent nature of topic modeling (Blei et al., 2003; Chang, Gerrish, Wang, & Blei, 2009). Prior research has commonly labeled topics based on the term(s) with the highest probability of occurring in each topic (e.g., Alsumait, Barbar, & Carlotta, 2009; Grimmer & Stewart, 2013). Because each of the final five less-intuitively interpretable topics included a single term with a considerably higher probability than all other terms, we labeled

these five topics with that single term (*Just**, *Even**, *Really**, *Like**, and *Will**) in accordance with prior work. An asterisk has been added to each of these labels to remind the reader

TABLE 2
Top 20 Terms Accounting for Each LDA Topic

Topic label	Top 20 terms
1. Past	Last years now since months ago night tried past days stil times week thought two back time gettting suicide well
2. Suicide	Want life hate anymore care happy die live everything kill nothing end take love worse suicide right thinking depression hurt
3. Just*	Just now right much everything point away getting anything everyone someone last tried know like hard bad anymore thinking get
4. Even*	Even never anything something though always good say reason ever wrong felt things done without actually like thought else still
5. Time	Time day much every first long thing still back used make love little get one way made see end away
6. Really*	Really know just think thing bad sure felt actually talking thinking long much makes lot getting talk time happy times
7. Friends	Friends friend one school best year talk also well made around good said started make two since talking never told
8. Family	Told got mom said dad started talk went home wanted family parents tell never well today talking let back little
9. Negative affect	Not want know just cannot scared have give away hurt get see thinking work last going stop hate alone makes
10. Like*	Like going things get lot come bad something always just getting wrong back way away time makes everything one think
11. Will*	Will know get tell see think going parents better sure back make ever still actually long without come used try
12. Depression	Feel like feeling depression depressed better alone makes make thinking way worse today end enough lot come happy also getting
13. Effort	Can keep trying try work hard anymore get enough take anything one nothing see point know find worse today give
14. NSSI	Help cut need stop cuting self want please done without bad started something days today enough give come week hurt
15. Social	People someone anyone way person say one everyone else find family get around hurt think good let help always also

that these topics were labeled based on a single term. The 10 terms with the highest probabilities of occurring in each of the 15 topics are shown with their specific probability estimates in Figure 1.

How Well Do Machine-Learned Topics Correspond to Topics Identified by Human Coders?

Results of logistic regressions testing relations between the six LDA topics most relevant to our hypotheses (*Suicide, Friends, Family, Depression, NSSI, and Social*) and the four human-coded primary issues hypothesized to correspond with those topics are presented in Table 3. As expected, each of these six LDA topics was significantly associated with increased odds of being human-coded as the primary issue corresponding with that LDA topic. Further, none of these six LDA topics were significantly associated with increased odds of the post being coded as any other primary issue.

We also examined similar, but exploratory logistic regressions among each of the remaining nine LDA topics and the four human-coded primary issues. Significant

relationships were found between the LDA topic *Just** and the primary issue *Suicide* (Log Odds = 11.74, 95% CI = [7.2, 16.24]), *Even** and *Social* (Log Odds = 6.56, 95% CI = [2.89, 10.23]), *Negative Affect* and *Suicide* (Log Odds = 6.85, 95% CI = [3.24, 10.38]), *Negative Affect* and *NSSI* (Log Odds = 5.15, 95% CI = [2.05, 8.22]), *Like** and *Social* (Log Odds = 7.51, 95% CI = [3.66, 11.39]), and *Effort* and *Suicide* (Log Odds = 8.67, 95% CI = [4.51, 12.78]).

Do Topics of Online Discussions Differ by Age or Gender?

T tests revealed that the posts of male users contained significantly greater use of the LDA topics *Suicide* ($t[385.47] = -2.43, p = .016, 95\% \text{ CI} = [-0.008, -0.001], \text{Cohen's } d = -.15$), *Friends* ($t[361.25] = -2.02, p = .044, 95\% \text{ CI} = [-0.008, -0.0001], d = -.12$), and *Social* ($t[376.05] = -2.67, p = .008, 95\% \text{ CI} = [-0.008, -0.001], d = -.16$). Posts of female users contained significantly greater use of the LDA topics *Family* ($t[443.98] = 4.96, p < .001, 95\% \text{ CI} = [0.005, 0.011], d = .31$), *Negative Affect* ($t[430.17] = 3.31, p = .001, 95\%$

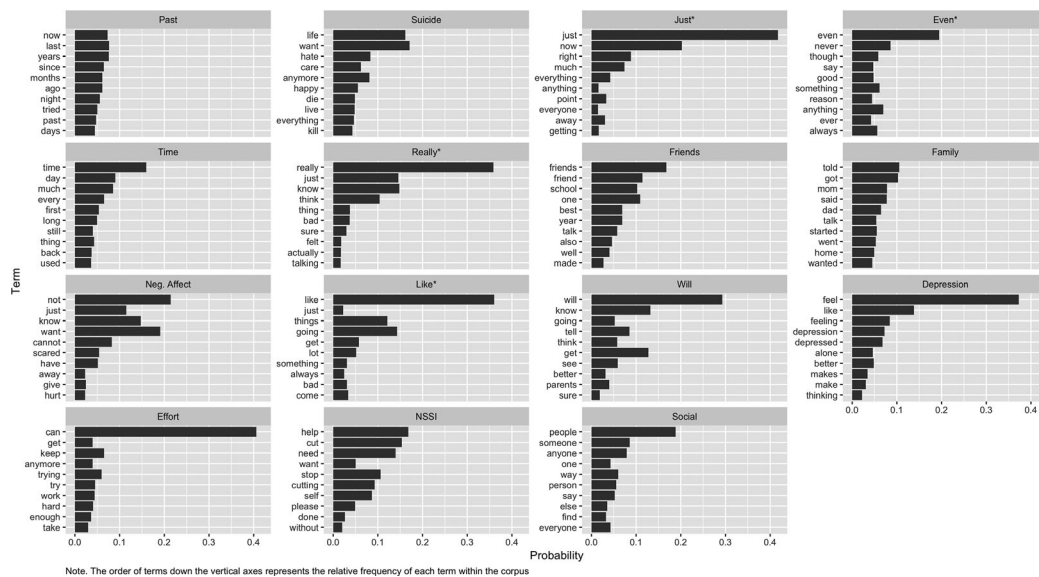


Figure 1. Terms with the highest probabilities in each of the LDA Topics.

TABLE 3
Relationships Between LDA Topics and Human-Coded Primary Issues

LDA topic	Human-coded primary issue							
	Suicide		NSSI		Depression		Social concern	
	Log odds	95% CI	Log odds	95% CI	Log odds	95% CI	Log odds	95% CI
Suicide	29.61	[25.74, 33.57]	-	-	4.50	[-0.43, 9.22]	-	-
NSSI	-	-	41.78	[37.66, 46.05]	-	-	-	-
Depression	3.16	[-1.08, 7.29]	-	-	31.50	[26.64, 36.45]	-	-
Friends	-	-	-	-	-	[-, -0.82]	24.81	[21.28, 28.45]
Family	-	-	-	-	-	[-, -1.90]	13.85	[10.90, 16.87]
Social	0.81	[-3.91, 5.13]	-	[-, -3.22]	-	[-, -6.14]	4.17	[0.67, 7.67]

Cells with “-” are the natural log of odds ratios of 0.00. Bolded values indicate $p < .05$.

CI = [0.002, 0.008], $d = .20$), and *NSSI* (t [408.6] = 4.13, $p < .001$, 95% CI = [0.003, 0.010], $d = .16$). There were no significant gender differences for any other LDA topic.

Age was significantly positively associated with the LDA topics *Time* ($r = .05$, $p = .013$, 95% CI = [0.01, 0.09]), *Negative Affect* ($r = .07$, $p < .001$, 95% CI = [0.03, 0.11]), *Like** ($r = .08$, $p < .001$, 95% CI = [0.04, 0.12]), and *Effort* ($r = .05$, $p = .01$, 95% CI = [0.01, 0.09]). Age was significantly negatively associated with the LDA topics *Friends* ($r = -.12$, $p < .001$, 95% CI = [-0.16, -0.07]), *Family* ($r = -.07$, $p < .001$, 95% CI = [-0.11, -0.03]), and *NSSI* ($r = -.09$, $p < .001$, 95% CI = [-.12, -.04]). There were no significant associations between age and any of the other LDA topics.

DISCUSSION

In this study, we used a topic modeling approach to identify latent topics discussed in a large corpus of online forum posts. There were three main findings in this study. First, we found that a 15-topic model detected topics representing discussions of SITBs and related issues such as families, friends, and negative affect. Second, each of these LDA topics distinguished between the relevant human-coded primary issue and all other issues. Third, we observed both gender- and age-related differences in several of these topics. Namely, males used more of the *Suicide*, *Friends*, and *Social* LDA topic than females, whereas females used more of the *Family*, *Negative Affect*, and *NSSI* topics. Age was associated with reduced use of the *Friends*, *Family*, and *NSSI* topics, and greater use of the *Time*, *Negative Affect*, *Like**, and *Effort* topics. Together, these results both validate the use of topic modeling to quantify SITB discussions in online forums and provide initial evidence of age and gender differences in these discussions.

Our first finding, that the LDA topic model identified topics representing themes of suicide and NSSI, highlights that an LDA machine-learning approach can clearly

identify online discussions of SITBs. There is a general increase in the use of web forums, social media, and other Internet-based peer-to-peer communication platforms among teens (Lenhart, Purcell, Smith, & Zickuhr, 2010), and so, the ability to detect these types of themes in the online content of individual users provides an opportunity to identify specific individuals who may be at risk for self-injurious behavior. This is especially important given that the prevalence and accessibility of suicide-related content appears to be increasing dramatically (Biddle et al., 2016). Furthermore, this method could help researchers elucidate additional psychological and social factors that may be associated with SITB risk, as the exact causes of self-injurious behaviors have not been clearly understood (Cha et al., 2018; Nock, 2009). Individuals who engage in SITBs are less likely to seek help than those facing other problems (Ciarrochi, Deane, Wilson, & Rickwood, 2002; Wilson, Deane, Ciarrochi, & Rickwood, 2005). As such, further development of the procedures described here may help provide a tool for rapidly identifying high-risk individuals who may be unwilling to seek support through more traditional means.

It is noteworthy that our topic model solution included separate topics for *Depression* and *Suicide*. Although depression and suicide are strongly related and share a number of risk factors, such as hopelessness (Rosellini & Bagge, 2014; Weersing et al., 2016) and are often comorbid conditions (Nock, Greif, Hwang, McLaughlin, & Sampson, 2013), depression is not a perfect predictor of suicidal thoughts or behaviors (Franklin et al., 2017; Ribeiro, Huang, Fox, & Franklin, 2018). Accordingly, our field has strived to develop tools that can identify suicide risk above and beyond risk for commonly co-occurring syndromes, especially among youth (Cha et al., 2018). That our unsupervised topic model solution included separate *Depression* and *Suicide* topics suggests that this type of machine-learning approach could hold promise for further refining our ability to detect communication about suicide separate from communication about depression.

Many of the topics include terms that may provide an intuition about how these topics tend to be discussed by users of the forum. For example, the *NSSI* topic includes the term “help.” Previous research indicates that individuals who engage in NSSI tend to seek help online (Whitlock, Powers, & Eckenrode, 2006), and seeking positive social feedback, such as help, support, or attention, is a common function of actual engagement in NSSI (Nock & Prinstein, 2004). Thus, one potential explanation for this result is that those who engage in NSSI tend to both engage in this behavior and communicate about this behavior for social reasons, such as seeking help. This possibility merits further research to better understand whether and why differences in help-seeking online communication exist between those discussing suicide, depression, and NSSI.

Similarly, the *Negative Affect* topic included the term “scared,” and this topic was associated with the human-coded primary issues Suicide and NSSI. This finding suggests that fear-related affective states may be relevant to SITBs or discussing SITBs online. The well-known interpersonal theory of suicide posits that one’s natural fear of death stands as a barrier to suicide attempts (Van Orden et al., 2011). The fact that a topic that includes the term “scared” is especially common in posts primarily dealing with SITBs may reflect the natural fear of death evoked by these issues. It is also possible that a topic which includes the term “scared,” as well as many other terms related to negative affective states (e.g., “hurt,” “hate,” alone”), is related to SITBs because individuals who engage in SITBs tend to struggle with difficult negative affective states (Nock, Wedig, Holmberg, & Hooley, 2008) and seek to communicate about these states in order to recruit support (Prinstein, Guerry, Browne, & Rancourt, 2009).

Our second finding, that LDA machine-learning topics are associated with relevant issues determined by human coders, validates that an LDA algorithm can accurately model issues that people intuitively observe. Although previous studies have utilized this algorithm (Carron-Arthur et al.,

2016), none to our knowledge have validated SITB-related LDA topics directly against similar issues identified by human coders. This finding indicates that this type of topic modeling approach can not only save considerable effort in identifying points of online discussion, but that it can also do so with a high degree of agreement with humans.

In addition to the hypothesized agreement between the LDA topics and the corresponding human-coded primary issues, we also found that the *Effort* and *Just** LDA topics were associated with the primary issue Suicide, the *Past* and *Really** topics were associated with the primary issue NSSI, and the *Even** and *Like** topics were associated with the primary issue Social. We hesitate to overinterpret these results, given their exploratory nature and our lack of clear hypotheses about them. Nonetheless, we provide some initial conjecture below.

The *Effort** topic seems to reference the theme of struggling toward a goal. One prominent theory suggests that suicide serves as an escape from intolerable distress about disappointment or unmet goals (Baumeister, 1990). The relationship between the topic *Effort* and the primary issue Suicide may reflect the challenging experience of failing to meet goals, which has been shown to be related to suicidal thinking (Schneider et al., 2011; Tang, Wu, & Miao, 2013). The *Effort* and *Just** topics include several similar terms, including “know,” “hard,” “get,” and “trying”/“try” (*Effort*) and “tried” (*Just**). The overlapping term structure of these topics may account for the fact that *Just** was also associated with the primary issue Suicide. Regarding the relationship between the topic *Past* and the primary issue NSSI, it is possible that individuals on this forum often reference specific episodes of NSSI that have occurred. Anecdotal evidence from our data suggests this may be the case. The relationships between the topic *Really** and the primary issue NSSI, as well as between the topics *Even** and *Like** and the primary issue Social Concerns, may not have clear theoretical explanations. These topics were less-intuitively interpretable than many of the other

topics, and so, we caution against drawing conclusions based on these exploratory analyses. They nonetheless represent a part of the lexicon of individuals discussing NSSI and Social Concerns on this Internet forum. Additional studies can help clarify whether these LDA topics replicate across other similar web forums, and whether the LDA topics show similar relationships to human-coded issues. Additionally, by administering targeted questionnaire measures to users of online forums, future studies could provide insight around the exact nature of LDA topics like *Effort*.

Finally, we found several gender and age differences in some LDA topics. First, the *Suicide* topic was higher for males on average. Female adolescents are shown to be at greater risk for suicide (Nock et al., 2013). Therefore, we were surprised that males rated higher in the *Suicide* topic. A range of factors could potentially explain this difference. For example, perhaps those males who are motivated to utilize TeenHelp.org do so because they experience higher-than-average distress with these issues. Future studies should seek to determine whether this gender difference replicates across online samples.

Further, we found that the *NSSI* topic was higher for females on average. This finding matches the gender difference typically observed in rates of NSSI in other samples (Bresin & Schoenleber, 2015; Sornberger, Heath, Toste, & McLouth, 2012), and so, this result is in line with our expectations. In addition, the *Family* topic was also higher for females, but the other two social topics, *Friends* and *Social* topics, were each higher for males. The finding that females used more of the *Family* topic accords with previous research showing that family discord is more strongly related to adolescent maladjustment among females than males (Davies & Windle, 1997). However, that males were higher in the *Friends* and *Social* topics was surprising. Numerous studies have suggested that female adolescents are more susceptible to the deleterious effects of social stress (e.g., Meiser & Esser, 2019), and so, we would have expected females to be higher in all three social LDA

topics, *Friends*, *Social*, and *Family*. However, it is unclear whether increased use of a particular topic in an online forum actually indicates that these individuals experience greater distress with that topic. A possible alternative explanation is that increased use of the topic simply implies that users are more likely to communicate about that topic in an online forum. One way that future research can clarify this is by soliciting information about real-world distress regarding specific issues and from users of an Internet forum and comparing these to the results of topic modeling.

Regarding age, the results that the *Friends*, *Family*, and *NSSI* topics were more prevalent among younger users accord with the notion that adolescence is a period of heightened attention to social concerns (for a review, see Somerville, 2013) and adolescents are more likely than adults to engage in *NSSI* (Taliaferro & Muehlenkamp, 2015). Social concerns, such as bullying (Whitney & Smith, 1993), are shown to decline with age. Similarly, conflict with parents seems to be highest in early adolescence, remain high through the middle adolescent years, and then decrease in late adolescence (Laursen, Coy, & Collins, 1998; Paikoff & Brooks-Gunn, 1991; Smetana, 1989). Further, *NSSI* is shown to increase during the early teenage years (Barrocas, Hankin, Young, & Abela, 2012), but decreases during the later teenage years and early twenties (Taliaferro & Muehlenkamp, 2015). Each of these trends aligns with this study's age-related findings. However, effect sizes were very small for each of these relationships and may not be detectable in smaller samples. Further, it remains unclear whether the individuals posting content on TeenHelp.org are similar to more traditionally studied community samples of youth, and how the discussion of these issues online relates to real-world experiences, as mentioned above. Again, additional studies should seek to assess the relationship between real-life behavior and concurrent discussion of related issues online.

We also found that the topics *Negative Affect*, *Time*, *Like**, and *Effort* were each positively associated with age. Given the small effect sizes of these relationships and a lack

of atheoretical explanation for why these effects might exist, we hesitate to speculate about why use of these topics might increase with age. For now, each of these age-related results (including those dealing with the topics *Friends*, *Family*, and *NSSI*) should be considered preliminary, though we do hope to test the replicability of these findings in future studies.

Future research should address three additional limitations of the current study. First, because we did not collect user names associated with each post, it is possible that multiple posts in the corpus were generated by the same user(s). This may have artificially increased similarity among posts, leading to inflated consistency within topics and strengthened correlations between human and LDA topics. Second, although the current study establishes that topic modeling can detect discussions of SITBs, it is not certain that having a higher level of the *Suicide* or *NSSI* LDA topic in a user's post is associated with a greater severity or frequency of real-life SITBs, or even indicates that that user has engaged in SITBs. In addition, we did not explore the use of supervised machine-learning approaches here, which may improve detect/predicting even further. Developing the current tool is the first step to future research that can address this gap by assessing SITB severity in users recruited from Internet support forums and future research should address some of these remaining questions. Third, this study's gender- and age-related findings relied on users' self-identification of age and gender on the forum. Although we have no reason to suspect that individuals misrepresented either age or gender, it is of course possible that the way an individual represents his/herself online differs from their true identity. Future research should assess age and gender, as well as other relevant demographic characteristics through other means than public self-identification on the web site. Fourth, the data we used for this project suffer from some degree of selection bias in that they were collected from a web forum where users have the ability to post content on the Internet and choose to do so.

Future research should test whether these findings generalize beyond teenagers and young adults who use Internet support forums and other digital media to voice concerns about mental health.

Finally, we hope in future studies to use this method to better understand the ways that discussions of SITBs and related content online change as a function of time and other clinically relevant demographic characteristics. Because we did not collect time stamps for posts in this data set, these analyses could not be examined. However, this could be done in future research to better understand how topics tend to shift within person, providing additional insight in the longitudinal processes that may give rise to increased SITB risk.

REFERENCES

- ALSUMAIT, L., BARBAR, D., & CARLOTTA, G. (2009). Topic significance ranking of LDA generative models. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 67–82).
- BARROCAS, A. L., HANKIN, B. L., YOUNG, J. F., & ABELA, J. R. Z. (2012). Rates of nonsuicidal self-injury in youth: Age, sex, and behavioral methods in a community sample. *Pediatrics*, *130*(1), 39–45. <https://doi.org/10.1542/peds.2011-2094>
- BAUMEISTER, R. F. (1990). Suicide as escape from self. *Psychological Review*, *97*(1), 90–113.
- BIDDLE, L., DERGES, J., MARS, B., HERON, J., DONOVAN, J. L., POTOKAR, J., ET AL. (2016). Suicide and the internet: Changes in the accessibility of suicide-related information between 2007 and 2014. *Journal of Affective Disorders*, *190*, 370–375. <https://doi.org/10.1016/j.jad.2015.10.028>
- BISGIN, H., LIU, Z., FANG, H., XU, X., & TONG, W. (2011). Mining FDA drug labels using an unsupervised learning technique – Topic modeling. *BMC Bioinformatics*, *12*(Suppl 10), S11. <https://doi.org/10.1186/1471-2105-12-S10-S11>
- BLEI, D. M., NG, A. Y., & JORDAN, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*, 993–1022. <https://doi.org/10.1162/jmlr.2003.3.4-5.993>
- BRESIN, K., & SCHOENLEBER, M. (2015). Gender differences in the prevalence of nonsuicidal self-injury: A meta-analysis. *Clinical Psychology Review*, *38*, 55–64. <https://doi.org/10.1016/j.cpr.2015.02.009>
- BURNS, J. M., DAVENPORT, T. A., DURKIN, L. A., LUSCOMBE, G. M., & HICKIE, I. B. (2010). The internet as a setting for mental health service utilization by young people. *Medical Journal of Australia*, *192*(11), 22.
- BUSCH, K. A., FAWCETT, J., & JACOBS, D. G. (2003). Clinical correlates of inpatient suicide. *Journal of Clinical Psychiatry*, *64*(1), 14–19. <https://doi.org/10.4088/JCP.v64n0105>
- CARRON-ARTHUR, B., REYNOLDS, J., BENNETT, K., BENNETT, A., & GRIFFITHS, K. M. (2016). What's all the talk about? Topic modelling in a mental health Internet support group. *BMC Psychiatry*, *16*(1), 1–12. <https://doi.org/10.1186/s12888-016-1073-5>
- Centers for Disease Control and Prevention: National Center for Injury Prevention and Control. (2017). Retrieved from www.cdc.gov/ncipc/wisqars
- CHA, C. B., FRANZ, P. J. M., GUZMÁN, E., GLENN, C. R., KLEIMAN, E. M., & NOCK, M. K. (2018). Annual Research Review: Suicide among youth – Epidemiology, (potential) etiology, and treatment. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *59*(4), 460–482. <https://doi.org/10.1111/jcpp.12831>
- CHANG, J., GERRISH, S., WANG, C., & BLEI, D. M. (2009). Reading tea leaves: How humans interpret topic models. *Advances in Neural Information Processing Systems*, 288–296.
- CIARROCHI, J., DEANE, F. P., WILSON, C. J., & RICKWOOD, D. (2002). Adolescents who need help the most are the least likely to seek it: The relationship between low emotional competence

- and low intention to seek help. *British Journal of Guidance and Counselling*, 30(2), 173–188. <https://doi.org/10.1080/03069880220128047>
- DAVIES, P. T., & WINDLE, M. (1997). Gender-specific pathways between maternal depressive symptoms, family discord, and adolescent adjustment. *Developmental Psychology*, 33(4), 657–668. <https://doi.org/10.1037/0012-1649.33.4.657>
- DINAKAR, K., LIEBERMAN, H., CHANEY, A. J. B., & BLEI, D. M. (2014). *Real-time topic models for crisis counseling*, 1–4.
- FRANKLIN, J. C., RIBEIRO, J. D., FOX, K. R., BENTLEY, K. H., KLEIMAN, E. M., HUANG, X., ET AL. (2017). Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research. *Psychological Bulletin*, 143(2), 187–232. <https://doi.org/10.1037/bul0000084>
- GRIMMER, J., & STEWART, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21, 267–297. <https://doi.org/10.1093/pa/nmps028>
- GRÜN, B., & HORNIK, K. (2011). topicmodels: An R package for fitting topic models. *Journal of Statistical Software*, 40(13), 1–30. <https://doi.org/10.18637/jss.v040.i13>
- JOINSON, A. (2004). Self-esteem, interpersonal risk, and preference for e-mail to face-to-face communication. *Cyber Psychology and Behavior*, 7(4), 479–486.
- LAURSEN, B., COY, K. C., & COLLINS, W. A. (1998). Reconsidering Changes in Parent-Child Conflict across Adolescence: A Meta-Analysis. *Child Development*, 69, 817–832. <https://doi.org/10.1111/j.1467-8624.1998.00817.x>
- LENHART, A., PURCELL, K., SMITH, A., & ZICKUHR, K. (2010). Social media & mobile internet use among teens and young adults. *Millennials Pew Internet and American Life Project*, 01, 1–16. <https://doi.org/10.3898/NEWF.78.02.2013>
- LIN, C., & HE, Y. (2009). Joint sentiment/topic model for sentiment analysis. *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, 375–384.
- LYNN, V. E., GOODMAN, A., NIEDERHOFFER, K., LOVEYS, K., RESNIK, P., & SCHWARTZ, H. A. (2017). *CLPsych 2018 SHARED task: Predicting current and future psychological health from childhood essays*.
- MEISER, S., & ESSER, G. (2019). Interpersonal stress generation—A girl problem? The role of depressive symptoms, dysfunctional attitudes, and gender in early adolescent stress generation. *The Journal of Early Adolescence*, 39, 41–66. <https://doi.org/10.1177/0272431617725197>
- MITCHELL, K. J., & YBARRA, M. L. (2007). Online behavior of youth who engage in self-harm provides clues for preventive intervention. *Preventive Medicine*, 45(5), 392–396. <https://doi.org/10.1016/j.ypmed.2007.05.008>
- NOCK, M. K. (2009). Why do people hurt themselves? New insights into the nature and function of self-injury. *Current Directions in Psychological Science*, 18(2), 78–83.
- NOCK, M. K., GREIF, J., HWANG, I., McLAUGHLIN, K., & SAMPSON, N. (2013). Prevalence, correlates, and treatment of lifetime suicidal behavior among adolescents: Results from the National Comorbidity Survey Replication Adolescent Supplement. *JAMA Psychiatry*, 70(3), 300–310. <https://doi.org/10.1001/2013.jamapsychiatry.55.Prevalence>
- NOCK, M. K., & PRINSTEIN, M. J. (2004). A functional approach to the assessment of self-mutilative behavior. *Journal of Consulting and Clinical Psychology*, 72(5), 885–890. <https://doi.org/10.1037/0022-006X.72.5.885>
- NOCK, M. K., WEDIG, M. M., HOLMBERG, E. B., & HOOLEY, J. M. (2008). The emotion reactivity scale: Development, evaluation, and relation to self-injurious thoughts and behaviors. *Behavior Therapy*, 39, 107–116.
- PAIKOFF, R. L., & BROOKS-GUNN, J. (1991). Do parent-child relationships change during puberty? *Psychological bulletin*, 110(1), 47. <https://doi.org/10.1037/0033-2909.110.1.47>
- PRINSTEIN, M. J., GUERRY, J. D., BROWNE, C. B., & RANCOURT, D. (2009). Interpersonal models of nonsuicidal self-injury. In M. K. Nock (Ed.), *Non-suicidal self-injury: Current science and practice* (pp. 139–156). Washington, DC: American Psychological Association.
- RESNIK, P., ARMSTRONG, W., CLAUDINO, L., NGUYEN, T., NGUYEN, V., & BOYD-GRABER, J. (2015). Beyond LDA: Exploring supervised topic modeling for depression-related language in twitter. *Proceedings of the 52nd Workshop Computational Linguistics and Clinical Psychology*, 1 (2014), 99–107. <https://doi.org/10.3115/v1/w15-1212>
- RIBEIRO, J. D., HUANG, X., FOX, K. R., & FRANKLIN, J. C. (2018). Depression and hopelessness as risk factors for suicide ideation, attempts and death: Meta-analysis of longitudinal studies. *The British Journal of Psychiatry*, 212(5), 279–286. <https://doi.org/10.1192/bjp.2018.27>
- RICHARDSON, L. (2016). *Beautiful soup documentation 4.0.0*.
- ROSELLINI, A. J., & BAGGE, C. L. (2014). Temperament, hopelessness, and attempted suicide: Direct and indirect effects. *Suicide and Life-Threatening Behavior*, 44(4), 353–361.
- SCHNEIDER, B., GREBNER, K., SCHNABEL, A., HAMPEL, H., GEORGI, K., & SEIDLER, A. (2011). Impact of employment status and work-related factors on risk of completed suicide. A case-control psychological autopsy study. *Psychiatry Research*, 190(2–3), 265–270. <https://doi.org/10.1016/j.psychres.2011.07.037>

- SMETANA, J. G. (1989). Adolescents' and parents' reasoning about actual family conflict. *Child development*, 1052–1067. <https://doi.org/10.2307/1130779>.
- SOMERVILLE, L. H. (2013). Special issue on the teenage brain: Sensitivity to social evaluation. *Current Directions in Psychological Science*, 22(2), 127. <https://doi.org/10.1177/0963721413476512>
- SORNBERGER, M. J., HEATH, N. L., TOSTE, J. R., & McLOUTH, R. (2012). Nonsuicidal self-injury and gender: Patterns of prevalence, methods, and locations among adolescents. *Suicide and Life-Threatening Behavior*, 42(3), 266–278. <https://doi.org/10.1111/j.1943-278X.2012.0088.x>
- TALIAFERRO, L. A., & MUEHLENKAMP, J. J. (2015). Factors associated with current versus lifetime self-injury among high school and college students. *Suicide and Life-Threatening Behavior*, 45(1), 84–97. <https://doi.org/10.1111/sltb.12117>
- TANG, J., WU, S., & MIAO, D. (2013). Experimental test of escape theory: Accessibility to implicit suicidal mind. *Suicide and Life-Threatening Behavior*, 43(4), 347–355. <https://doi.org/10.1111/sltb.12021>
- TSENG, F. Y., & YANG, H. J. (2015). Internet use and web communication networks, sources of social support, and forms of suicidal and nonsuicidal self-injury among adolescents: Different patterns between genders. *Suicide and Life-Threatening Behavior*, 45(2), 178–191. <https://doi.org/10.1111/sltb.12124>
- VAN ORDEN, K. A., WITTE, T. K., CUKROWICZ, K. C., BRAITHWAITE, S., SELBY, E. A., & JOINER, T. E., JR. (2011). The interpersonal theory of suicide. *The Psychology Review*, 117(2), 575–600. <https://doi.org/10.1037/a0018697>
- VOGEL, D. L., WADE, N. G., & HACKLER, A. H. (2007). Perceived public stigma and the willingness to seek counseling: The mediating roles of self-stigma and attitudes toward counseling. *Journal of Counseling Psychology*, 54(1), 40–50. <https://doi.org/10.1037/0022-0167.54.1.40>
- WANG, C., & BLEI, D. M. (2011). *CTM for recommending scientific articles*. *Kdd*. <https://doi.org/10.1145/2020408.2020480>
- WEERSING, V. R., SHAMSEDDEN, W., GARBBER, J., HOLLON, S. N., CLARKE, G. N., BEARDSLEE, W. R., ET AL. (2016). Prevention of depression in at-risk adolescents: Predictors and moderators of acute effects. *Journal of the American Academy of Child & Adolescent Psychiatry*, 55(3), 219–226.
- WHITLOCK, J. L., POWERS, J. L., & ECKENRODE, J. (2006). The virtual cutting edge: The Internet and adolescent self-injury. *Developmental Psychology*, 42(3), 407–417. <https://doi.org/10.1037/0012-1649.42.3.407>
- WHITNEY, I., & SMITH, P. K. (1993). A survey of the nature and extent of bullying in junior/middle and secondary schools. *Educational Research*, 35(1), 3–25. <https://doi.org/10.1080/0013188930350101>
- WILSON, C. J., DEANE, F. P., CIARROCHI, J., & RICKWOOD, D. (2005). Measuring help-seeking intentions: Properties of the general help-seeking questionnaire. *Canadian Journal of Counselling*, 39(1), 15–28. Retrieved from <http://cjc-rcc.ucalgary.ca/cjc/index.php/rcc/article/view/265>
- ZHAO, W. X., JIANG, J., WENG, J., HE, J., LIM, E.-P., YAN, H., ET AL. (2011). Comparing twitter and traditional media using topic models. *European Conference on Information Retrieval* (pp. 338–349). [https://doi.org/10.1016/0952-8180\(89\)90027-5](https://doi.org/10.1016/0952-8180(89)90027-5)

Manuscript Received: August 14, 2018

Revision Accepted: March 4, 2019