The hidden depths of suicidal discourse: Network analysis and natural language processing unmask uncensored expression

Damien Lekkas[1,2] (iD) **and Nicholas C Jacobson**[1,2,3,4]

Abstract

Background: The socially unattractive and stigmatizing nature of suicidal thought and behavior (STB) makes it especially susceptible to censorship across most modern digital communication platforms. The ubiquitous integration of technology with day-to-day life has presented an invaluable opportunity to leverage unprecedented amounts of data to study STB, yet the complex etiologies and consequences of censorship for research within mainstream online communities render an incomplete picture of STB manifestation. Analyses targeting online written content of suicidal users in environments where fear of reproach is mitigated may provide novel insight into modern trends and signals of STB expression.

Methods: Complete written content of $N = 192$ users, including $n = 48$ identified as potential suicide completers/highest-risk users (HRUs), on the pro-choice suicide forum, Sanctioned Suicide, was modeled using a combination of lexicon-based topic modeling (EMPATH) and exploratory network analysis techniques to characterize and highlight prominent aspects of censorship-free suicidal discourse.

Results: Modeling of over 2 million tokens across 37,136 forum posts found higher frequency of positive emotion and optimism among HRUs, emphasis on methods seeking and sharing behaviors, prominence of previously undocumented jargon, and semantics related to loneliness and life adversity.

Conclusion: This natural language processing (NLP)- and network-driven exposé of online STB subculture uncovered trends that deserve further attention within suicidology as they may be able to bolster detection, intervention, and prevention of suicidal outcomes and exposures.

Introduction

Suicide claims the lives of over 700,000 individuals annually with approximately one death by suicide every 40 s.[1,2] Moreover, non-fatal suicidal behaviors occur with a frequency of more than 20 times higher,[2] and many more are likely unreported due to cultural taboos and governmental proscriptions.[3] Undoubtedly, suicidal thought and behavior (STB) is a global public health problem with unique epidemiological and scientific challenges. While traditionally studied on small, non-generalizable cohorts and within clinical settings by necessity, the rapid evolution

[1]Center for Technology and Behavioral Health, Geisel School of Medicine, Dartmouth College, Lebanon, NH, USA
[2]Quantitative Biomedical Sciences Program, Dartmouth College, Hanover, NH, USA
[3]Department of Biomedical Data Science, Geisel School of Medicine, Dartmouth College, Lebanon, NH, USA
[4]Department of Psychiatry, Geisel School of Medicine, Dartmouth College, Hanover, NH, USA

Corresponding author:
Damien Lekkas, Center for Technology and Behavioral Health, Geisel School of Medicine, Dartmouth College, 46 Centerra Parkway, Suite 300, Office #313S, Lebanon, NH 03766, USA.
Email: Damien.Lekkas.GR@dartmouth.edu

of the internet and the seamless integration of mobile and computing technology with day-to-day life have expanded the methodological and sociodemographic scope of suicidal inquiry, effectively creating an unprecedented opportunity to research and model STB on larger and farther-reaching scales.

The research literature on STB has exploded in the past several years with an impressive array of descriptive analyses and modeling efforts aiming to better detect,[4] predict,[5] describe,[6] and explain[7] suicidal phenomenology through the leverage of online social media and the "big data" afforded by web-based communication platforms. Given the nature of the medium, natural language processing (NLP) has emerged as a prominent analytical tool, effectively paired with a variety of other cutting-edge quantitative approaches such as network analysis[8] and machine learning.9 From the vast toolkit of NLP, STB research has prominently featured sentiment analysis/opinion mining (e.g. quantification of emotion or affective valence),[10,11] topic modeling (i.e. categorization of content based on word clustering),[12] and various word-embedding techniques (e.g. vectorized numeric representations of text),[13] often wrapped within more complex machine learning or deep learning architectures.[14] Collectively, these efforts have been promising and have indicated a strong ability to detect and predict suicidal outcomes from non-uniform texts.[15] Indeed, informal written records of what is said, how it is said, and to whom it is said, conveniently and exhaustively captured in forum threads and social media account histories, may contain meaningful patterns that signal STB risk and may shed light on the nature and importance of key theoretical constructs.

The unique scientific value and numerous research opportunities provided by these digital footprints of modern sociality are undeniable; however, prevailing public perceptions of STB may limit the ability of data to represent a complete picture of the suicidal mind. In other words, minority opinion and the accompanying fear of social reproach, which are part of a larger, multifaceted force of censorship on digital platforms,[16–18] may lead to an incomplete understanding of how STB manifests. Studies must address or acknowledge the fact that behavioral mediators linked to social marginalization, as well as competing corporate interests[19] and algorithmic adjustments to information exchange,[20] ultimately influence what is said, what is seen, and what is preserved for posterity. This does not invalidate the data or the studies that use it, but it would be preferable to study STB in an environment that reduces concerns about censorship and encourages naturalistic tendencies.

While any uncensored community of scientifically useful size is exceptionally rare on the surface Web, Sanctioned Suicide (https://sanctioned-suicide.org) presents a staggeringly large (over 30,000 users), publicly facing phenomenon of unmitigated STB discourse and social exchange. Sanctioned Suicide is self-described as a "pro-choice" suicide forum and "a safe space to discuss the topic of suicide without the censorship of other places, as well as a community that can understand you and let you be yourself without judging you or forcing you to do anything."[21] It is intended as a place to "vent, talk to like-minded individuals, share experiences, or to empathize and offer kind words to others who might need them."[21] From its early roots as a fringe Reddit subforum to its current incarnation as a fully independent entity (as of March 2018), Sanctioned Suicide has gradually fashioned a supportive, philosophical, and morbid subculture around STB with its own unique profile of language, content foci, and knowledge generation. Anthropologically speaking, it is replete with valuable ethnographic potential. As NLP has proven to be invaluable as a tool with which to probe the digital signatures of STB, the ability to leverage these techniques within the less restricted setting of Sanctioned Suicide may shed new light on how STB is expressed on what comprises the *de facto* mode of modern communication.

The current work offers a unique opportunity to explore and analyze suicide-focused discourse within an online community whose mission is to provide an accepting space for individuals to discuss their own STB anonymously and freely without fear of stigma or censorship. Bolstered by the promise and success of NLP methodologies in the study of STB, this research was guided by the following aims:

1. Classify and summarize the topical content of STB discourse on Sanctioned Suicide using a representative cohort of $N = 192$ users.
2. Statistically compare the relative frequencies of topics between users identified as highest risk and controls matched in terms of posting behavior and forum tenure.
3. Employ exploratory network analysis tools to highlight key topics and topic associations among all posts authored by the cohort.
4. Model topic-related token usage—the words contained within posts that belong to one or more topics—to gain more detailed insight into the prominent semantics of this community.

Methods

Data source characteristics

The current investigation focused on written post data from users on the pro-choice suicide forum, Sanctioned Suicide. Sanctioned Suicide provides a stigma-free space for individuals across the globe who struggle with suicide-related thoughts, behaviors, and associated negative experiences such as those stemming from mental and physical comorbidities, trauma, grief, and social alienation. Although the forum

is publicly accessible for viewing, it has strict rules and moderation efforts that prevent disclosure of personally identifying information, and moreover, requires an approved account to post and interact with members of the community. When registering, individuals must acknowledge that they are at least 18 years of age and must explain in detail why they want to join. This study and all associated protocols were deemed to present no greater than minimal risk to subjects and thus "exempt" from further review by the Committee for the Protection of Human Subjects at Dartmouth College (STUDY00032141).

Data collection

All activity within the "Suicide Discussion" subforum, from platform inception on 17 March 2018 to 5 February 2021, was collected and organized into tabular format using a custom Python (v3.8) script that primarily leveraged the *BeautifulSoup* package[22] to parse html code. The resulting dataset consisted of more than 600,000 time-stamped posts across nearly 40,000 threads, ultimately reflecting expressions of STB for 11,583 users. Information on (a) thread title, (b) thread author, (c) post author, (d) post date, (e) post text content, and (f) direct mentions and references to other user comments within the post text were collected. Each username was automatically assigned a randomly generated, 32-character hashed ID. These de-identifying IDs were automatically replaced with all instances of users' online handles in the data prior to subsequent preprocessing.

Cohort selection and data preprocessing

Identifying highest-risk individuals. One goal of this research was to compare the written content of the highest-risk users (HRUs) with all other users on the forum. Accordingly, highest risk was synonymized with active volition,[23] herein translating to a written account of fatal suicidal behavior while active on the forum. A structured approach was devised to select a subset of representative users and identify HRUs based on the findings of the New York Times investigation into Sanctioned Suicide[24] as well as the authors' thorough review of the forum content. To best identify HRUs, data was first filtered by searching for thread titles with the following keywords/phrases: "bus is here," "catch the bus," "fare well," "farewell," "final day," "good bye," "goodbye," "leaving," "my time," "my turn," "so long," and "took SN." For reference, "catch the bus" is a notable euphemism adopted by the community to symbolize suicide, while "SN" is short for sodium nitrate, a popular chemical used in methods for completing suicide. These terms were used to identify what are known as "goodbye threads" on Sanctioned Suicide, and thus have the highest chance of signaling that an individual intends to act on their suicidal thoughts.

From the content of these flagged threads, each thread was manually read in its entirety, and the author of each thread was determined to be a HRU if the following conditions held: (a) there was no record of post or engagement activity by the user on the forum after the date of the last post within the goodbye thread, (b) no other users mentioned seeing the user as "online" in their profile information after the date of the last post in their goodbye thread, and (c) the thread contained "confirmation" of fatal suicidal behavior as stated by other users who either allegedly knew the user personally in real life or who interacted directly with the user in real time during their suicide completion. These conditions echo the investigative findings of the New York Times that linked these behaviors with real-life suicide incidents.[24] This strategy yielded 48 users as engaging in active suicidal behaviors, and due to the nature of these threads, the community dialog surrounding these behaviors, and this study's filtering criteria, likely completing suicide. Given the anonymity of this platform and of online social interactions as a whole, there is of course no complete certainty of veracity; however, this manually- and contextually-selected group of users represents a subset of those for whom this outcome most likely came to pass and therefore undoubtedly represents among the HRUs on the platform.

Control individuals. Expectedly, user activity across Sanctioned Suicide was highly heterogeneous due to factors such as the amount of time spent as a member on the forum, the number of posts, the frequency of posting, and the size (verbosity) of posts. To obtain a suitably sized cohort of users for analysis that is both representative of activity across Sanctioned Suicide and consists of suitable behavioral controls for HRUs (see 2.3.1), each HRU was matched to three control users. A suitable match was primarily determined based on selecting users with equivalent forum tenure duration (within 2 days; determined by the difference in days from a user's most recent and first post on Sanctioned Suicide) and then selecting users with the closest total number of posts on record. With nine exceptions stemming from three HRUs of unusually high activity, this second step yielded controls with a total post volume within one order of magnitude of their respective HRU. In total, 144 users were selected as representative controls. All user-specific activity information, showing which controls were associated with each HRU, is provided for reference in Supplemental File 1 under "Cohort Activity Data."

Summary of cohort activity. The 192 users of Sanctioned Suicide (48 HRUs and 144 matched controls) selected for analysis represented content across 37,136 posts. Post activity across the cohort was summarized in terms of median and range (a) total number of posts, (b) total number of tokens (words), (c) range, in days, of post activity, and (d) percentage of total tokens that mapped to the

NLP lexicon of choice (see 3.1). The Wilcoxon rank-sum test was also employed to compare HRUs with controls, thereby testing for non-significance (parity) across these characteristics.

Data filtering and concatenation. For each user in the cohort, all their respective posts were concatenated into a single "document," resulting in 192 documents. All punctuation (except apostrophes) as well as embedded mentions and quotes of other users were removed from these documents to generate 192 lists of unigram tokens. Each document contained every word originally written by the respective user and therefore represented a comprehensive semantic catalog of their suicide-related discourse on the forum.

Statistical analysis

EMPATH topic modeling

To classify and summarize the topic domains of STB discourse on the forum, each document was analyzed using the EMPATH tool and lexicon in Python.[25] The topics in the lexicon and their associated seed words were created by leveraging the ConceptNet framework.[26] The resulting seed words were then modeled using a deep learning, neural vector space model (VSM) architecture[27] to derive sets of tokens that represent each topic. These tokens were then validated through expert-level crowdsourcing to arrive at a final list of tokens for each topic that define the published EMPATH lexicon. At the time of analysis, the lexicon provided mapped tokens across 194 topic/emotion categories.

EMPATH was selected over other popular methods such as Latent Dirichlet Allocation (LDA) to capitalize on the powerful hybridization of deep learning and expert curation, as well as to avoid a purely unsupervised approach whose results would be entirely driven by the context and idiosyncrasies of the current dataset. Through EMPATH, this study utilized topic frequency data across 188 topics. The frequency of a topic in a document (for a user) was based on the document's total count of tokens that mapped to that topic in the lexicon and was normalized by the total number of tokens in the document. Due to consistently low or no representation across the cohort, six topics ("anonymity," "exotic," "farming," "medieval," "superhero," and "terrorism") from the available 194 were filtered out of consideration in downstream modeling. This equated to the removal of the bottom 10% of topics ranked by median normalized frequency across users (rank ≥ 174.6).

Description and statistical comparison of topic frequencies

To describe and summarize the topic landscape of content on the forum prior to modeling, relative topic frequencies were examined both qualitatively and quantitatively. At the cohort level, a word cloud was first created to easily visualize the overall relative magnitude of topic representation. The most frequently mapped tokens for each topic were also calculated and reported. Statistically significant differences in topic frequency between HRUs and controls (see 2.3) were then ascertained through implementation of the Wilcoxon rank-sum test to highlight topics that may signal for active suicidality. Because this work was exploratory and because the primary aims regarded network connectivity, correction of multiple comparisons was not performed. A $P < 0.05$ was considered statistically significant.

Network analysis of topics, topic clusters, and words

Exploratory graph analysis. Many EMPATH topics have some degree of lexical overlap with one another. Since topic frequencies were derived through the presence of defining tokens in text, and tokens need not be mutually exclusive to any single topic, the frequencies of some conceptually synonymous topics are similar to one another due to the large degree of overlap in the tokens that define them. The main goal of this step was therefore to first reduce the topic space in a statistically informed manner using a method that also allows for an intuitive visualization of the results. To this end, Exploratory Graph Analysis (EGA) was performed on the topic frequency data from 3.1 using the *EGAnet* package (v1.0.1) in R (v4.0.2).[28,29] Briefly, this package was used to first construct a Triangulated Maximally Filtered Graph (TMFG)[30] representing the zero-order correlations between variables and then perform community detection using the Louvain algorithm[31] to specify the number and topic constituency of dimensions (communities) present in the graph. The TMFG and the associated community membership for each topic were visualized for qualitative and comparative introspection with the subsequent modeling step.

Partial correlation network of topic dimensions. Using the results obtained in the section 'Exploratory graph analysis', the normalized frequencies belonging to all topics within their respective dimension were summed for each user to obtain the overall normalized frequency of each dimension in the text. A Gaussian Graphical Model (GGM) was then estimated using the *bootnet* package (v1.5) in R[32] to observe associations among topic dimensions across the cohort. This approach utilizes regularization techniques alongside the Extended Bayesian Information Criterion (EBIC) to model the partial correlational associations among continuous variables. For the current model, the EBIC tuning parameter, "tuning," was set to "0.3" with all other parameters set to their default values. Non-parametric bootstrap resampling of the estimated network was performed $k = 1000$ times to generate 95% confidence intervals around the network edge estimations. Associations between dimensions (nodes) were considered

significant if the confidence interval did not cross zero. Node strength, betweenness, and expected influence were also reported. The GGM was visualized with attention given to significant edges.

For more information on the theoretical and practical aspects of implementing a GGM in *bootnet*, interested readers are encouraged to consult an excellent tutorial paper authored by the creators of the package.[32]

Partial correlation network of topic tokens. To further investigate the relational context of STB discourse at a more detailed level of analysis, we modeled individual token frequencies. To achieve this in a directed, phenomenologically interesting, and statistically feasible manner, we focused on a subset of all tokens. To determine this subset, we first calculated raw counts of tokens across all posts in the dataset. Then, we mapped these tokens to their associated topics, which were then mapped to their designated topic dimension (see 'Exploratory graph analysis'). From this, we selected the top 2% of tokens in each dimension to lexically represent their respective cluster of topics. We obtained a list of 136 tokens and modeled their normalized frequencies for each user as a GGM. Given the exploratory nature of the analysis and the high variable-to-sample ratio (136:192), we set the "tuning" parameter to "0" to maximize sensitivity at the expense of specificity. All other parameters were set to their default values. We calculated node strength, betweenness, and expected influence and reported them alongside the estimated token network. The complete lists of token counts and dimension affiliations are available in Supplemental File 1 under "Token Dimension Data."

Results

Description and statistical comparison of cohort forum activity

Table 1 summarizes the key characteristics of forum activity across the cohort. Users, on average, posted 194 times, wrote 10,779 tokens, and had posts spanning 97 days of activity. Overall, the data represent a wide range of post activity, verbal representation, and forum tenure. Due to successful efforts in matching HRUs with appropriate controls, cohort-wide heterogeneity was not recapitulated between groups. This was indicated by non-significant results for all Wilcoxon rank-sum tests between HRU and control groups. Importantly, this implies that any differences uncovered between the groups in downstream analyses would likely not be due to unequal representation in the data (i.e. number of posts, number of tokens). The data successfully capture many archetypes on Sanctioned Suicide, from the intermittent "lurker" to the consistently and highly engaged, from the laconic to the verbose.

Description and statistical comparison of EMPATH topic frequencies

The relative frequency of 188 EMPATH topics across all posts made by the cohort is represented as a word cloud in Figure 1. As the largest in the cloud, "negative emotion," "positive emotion," "pain," "suffering," "death," and "violence" qualitatively stand out as the most frequent topics of conversation. Additionally, Table 2 summarizes the top 20 topics in terms of their median rank (ranging from the most frequent median rank of "1" to the least frequent median rank of "24.5"). Each of these top-ranked topics is accompanied by a descriptive list of the top five most frequent words used by the cohort that represent the topic, as defined by the EMPATH lexicon. In separate consideration of HRUs and controls, five of these 20 topics were found to be statistically significantly different ($P < 0.05$) in their frequency of representation between groups. "Positive emotion" and "optimism" were found to be more frequently represented among posts made by HRUs relative to controls, while "violence," "speaking," and "hate" were found to be less frequently represented among posts made by HRUs relative to controls. All other topics, while frequently represented, did

Table 1. Descriptive statistical summary of cohort activity.

| Characteristic median [range] | All users | Controls | HRUs | |
	$N = 192$	$n = 144$	$n = 48$	Wilcoxon *P*-value
Total number of posts	111.5 [7, 2318]	105.5 [7, 976]	135.5 [7, 2318]	0.254
Total number of tokens	5642 [186, 135510]	5456.5 [186, 53932]	7391 [207, 135510]	0.238
Range of post activity (days)	57 [1, 442]	57 [1, 442]	57 [1, 442]	0.996
Tokens mapped to EMPATH (%)	16.23 [9.63, 42.89]	16.32 [12.49, 26.61]	16.02 [9.63, 42.89]	0.134

Note: Median and range are reported for each characteristic. Group comparisons were carried out using the non-parametric, two-sided Wilcoxon rank-sum test. HRU and control groups were not significantly different ($P > 0.05$) in terms of forum engagement or in the percentage of tokens mapped to EMPATH.
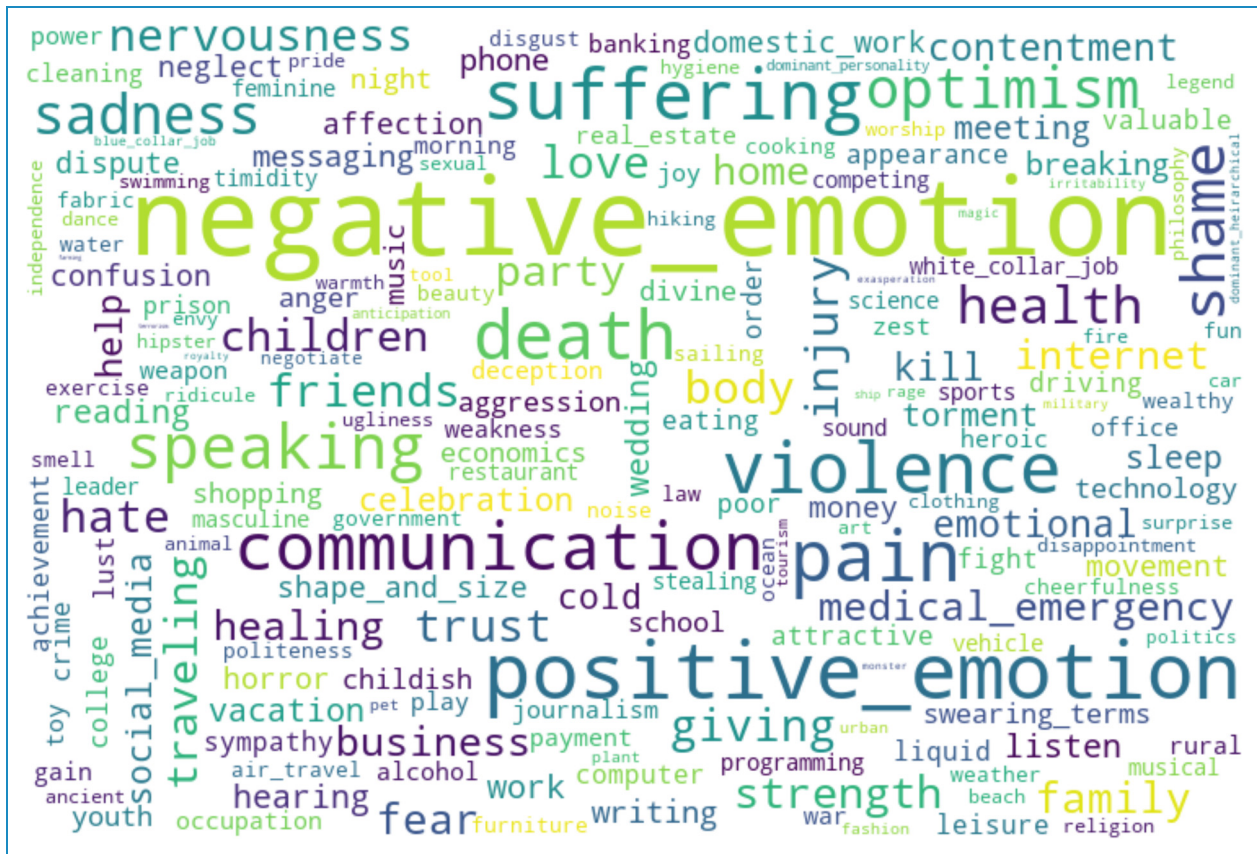
Figure 1. Word cloud of EMPATH topics.
Note: The word cloud displays the relative frequency of 188 EMPATH topics across all 192 users' posts. Larger fonts indicate higher frequencies.

not differ significantly between HRUs and controls. All 188 topic frequencies, significance between groups, and most representative tokens are presented in Supplemental File 1 under "EMPATH Topic Data."

EMPATH topic and topic cluster network analysis

The exploratory TMFG, built on 192 frequencies of 188 topics, is presented in Figure 2(A). Through the implementation of the Louvain community detection algorithm, nine distinct clusters or dimensions of topics were identified. A GGM was constructed to capture the associations among the frequencies of these dimensions, as shown in Figure 2(B). Bootstrapping yielded five statistically significant associations between pairs of topic dimensions. Notably, these associations suggested that the corpus of a user's posts tended to be characterized by (a) the juxtaposition of both negative and positive sentiment (D2-D4), (b) a primarily positive dialog regarding themes of social interactivity and the arts (D2-D6), and (c) a tendency toward the non-concurrence of content regarding digital life and several negative sentiments such as pain, suffering, and shame (D4-D8). Topic membership for each of these

dimensions is enumerated and color-coded for reference in Figure 2(C). Additional results concerning edge stability analysis and topic centrality are available in Supplemental File 1 under "TopicNet EdgeStab Centrality."

Network analysis of EMPATH topic tokens

The GGM estimated from the most prominent topic-related tokens is presented in Figure 3. Several key observations arose from this exploration. First, several tokens belonging to seemingly innocuous topics such as "drink"/"drinking," "taste," "bag," "exit," "dose," and "help" were among the most central in the network. This underlines the prominence of a major content theme across Sanctioned Suicide, namely the appeal for, and sharing of, information related to suicide-related methods.[33] This is secondarily supported by the close association of "information" and "thread" in the network. Relatedly, the central tokens "bus," "exit," and "bag," along with other less central tokens in association with each other, such as "peaceful" + "journey," shine a light on some of the prominent, unconventional, and previously unreported jargon/language that is often used among members to discuss and describe

Table 2. Group differences in EMPATH topic frequencies.

| Topic	Median rank	Top 5 most frequent tokens	Wilcoxon P-value		+/−
negative_emotion	1	death; pain; die; hard; bad	0.736		
positive_emotion	4	hope; love; better; wish; feeling	0.016	*	+
pain	6.5	feel/feeling; pain; bad; suffering; kill	0.517		
suffering	7	death; pain; bad; feeling; suffering	0.355		
death	7	life; death; die; wish; suffering	0.607		
violence	7	feel; death; bad; mean; suffering	0.030	*	−
communication	9	understand; read; talk; tell; idea	0.484		
speaking	11	say/saying; understand; talk; tell; ask	0.010	*	−
shame	12	feel/feeling; pain; bad; sad; hurt	0.599		
optimism	13.25	hope; love; better; feeling; happy	0.035	*	+
sadness	14	pain; feeling; suffering; depression; sad	0.977		
health	17	mental; depression; health; anxiety; hospital	0.205		
friends	21	love; always; best; friends/friend; nice	0.867		
nervousness	23	feel/feeling; anxiety; fear; scared; afraid	0.467		
giving	23	need; thank; give; agree; job	0.360		
love	23	feel; love/loved; experience; sense; wanting	0.616		
body	23.25	feel; body; head; face; neck	0.754		
hate	23.5	bad; feeling; suffering; kill; worse	0.019	*	−
trust	24	care; matter; friends; wrong; true	0.918		
injury	24.5	pain; cause; painful; blood; hurt	0.231		

Note: The table presents the 20 most frequent topics in ascending order of median rank frequency. Rank 1 represents the most frequent, while rank 24.5 represents the least frequent. Significant differences ($P < 0.05$) in topic frequency between HRU and control users as determined through the Wilcoxon rank-sum test are denoted with an asterisk (*). When significant, a "+" denotes higher frequency among HRUs, while "−" denotes lower frequency among HRUs relative to controls.

suicide-related concepts in euphemistic terms. Other close token associations also hint at other content themes such as loneliness ("alone" + "feel" + "pain" + "tired"), perceived adversity ("life" + "suffering" + "hell" + "hate"), and comorbidity ("mental" + "health" + "depression" + "anxiety" + "worse"). Lastly, "hope" is among the most central tokens; however, it connects to tokens that may suggest a darker etiology—hope that stems not from a renewed investment in life, but from reaching the decision to end it, to escape the trials and tribulations of the current life, and to achieve peace in the hereafter ("journey" + "decision").

Discussion

Using topic modeling alongside exploratory network analysis techniques, the written expressions of $N = 192$ users on the online pro-choice forum, Sanctioned Suicide, were explored to describe and uncover key patterns that may be
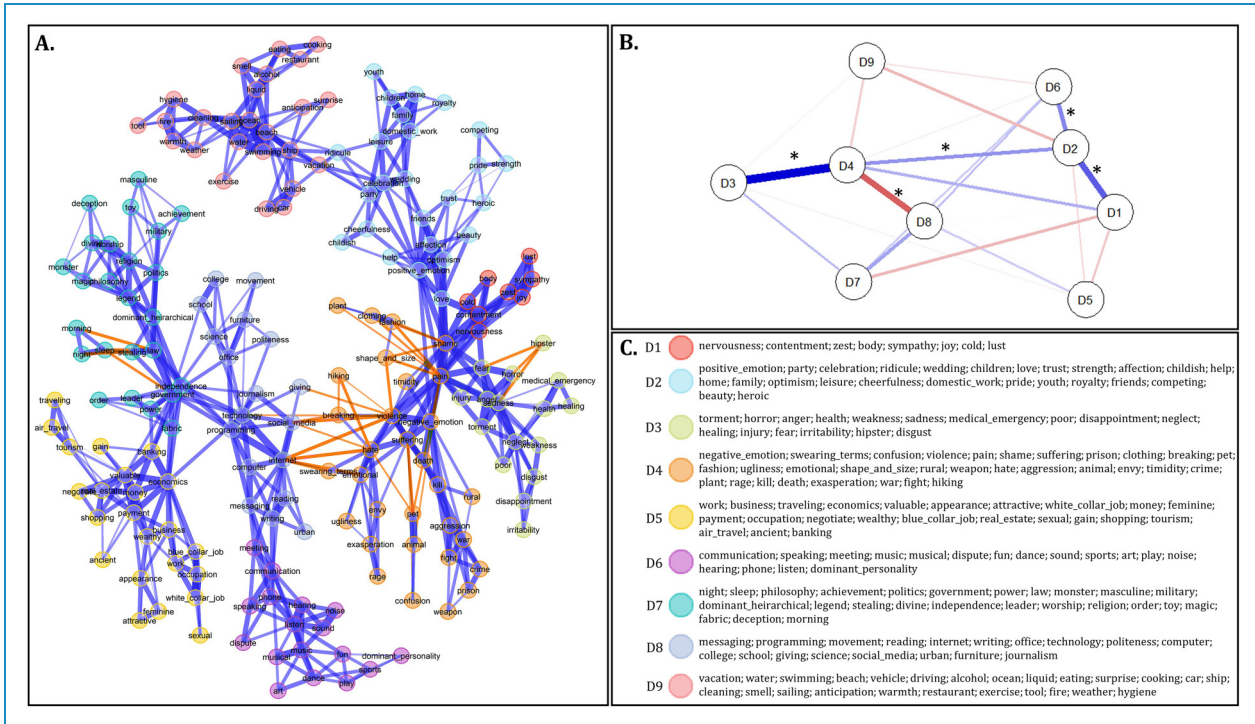
Figure 2. Topic dimension reduction and associated network models.
Note: (A) The TMFG of 188 EMPATH topic frequencies across $N = 192$ users resulted in the identification of nine distinct clusters of topics using Louvain community detection. Each node in the graph represents a topic and is colored according to its cluster affiliation. The edges represent correlative strength, with blue edges denoting positive correlations and orange edges denoting negative correlations. The width of each edge is proportional to its correlative strength. (B) The GGM of the nine topic clusters. Each node represents a cluster of topics identified by the TMFG. The blue edges denote positive partial correlations, while the red edges denote negative partial correlations. An asterisk (*) denotes an edge that was found to be statistically significant through the bootstrapping procedure. (C) The EMPATH topic membership by cluster. The clusters are color-coded and labeled D1–D9, which correspond to the node colors in Panel A and the node labels in Panel B, respectively.

considered in future studies of STB risk. In a comparison of HRUs and controls, users who engaged in active suicidal behaviors (and likely successfully completed) suicide tended to write more frequently with positive emotion and optimism than their control counterparts. Additionally, HRU content revolved less around hate and violence. These findings may run counter to expectations given how suicidal ideation is often documented alongside constructs like depression, which itself is comprised and often measured in part through negative affect-related items (e.g. CES-D). In addition to evidence hinting at the comforting properties of suicidal cognition,[34] these results support findings that suicidal thoughts can be reinforcing and serve to positively regulate affect.[35] It may be that a noticeable shift toward a more positively-valenced context in suicidal expression signals a heightened risk and an advanced development of suicidal volition. Indeed, this comparison of HRUs with controls allowed for an indirect capture of the transition process from ideation-to-action, which is a prominent modern framing of STB.[23,36,37] From this perspective, the results echo the theoretical importance placed on moderators that promote suicidal volition in the transition from intention

formation to behavioral enaction,[23] and specifically highlight aspects of suicidal imagery as especially prominent in this community. Some detail of this prominence on Sanctioned Suicide is seen among the network of most representative tokens (Figure 3), where concepts such as "peaceful journey" and actions such as catching the "bus" are metaphorically promoted and ingrained into the collective suicidal imagery that exists in the minds of vulnerable users.

Other topical patterns uncovered through network modeling (Figure 2(B)) indicate a dichotomous juxtaposition of positive and negative emotions, which may speak in part to the aforementioned dynamic of affect regulation but also may be a repercussion of Sanctioned Suicide's pro-choice, community-fostering nature. Users share their past, engage in philosophical discourse, and gain a sense of belonging. Many are understood and validated in their negative attitudes, which is naturally accompanied by commiseration and positive dialog. In fact, another association highlighted in Figure 2(B) is the tendency toward mutual exclusion of negative emotions and digital topics (e.g. "messaging," "internet," "computer"). This may signal how virtual spaces

Figure 3. Network model of most frequent tokens across topic dimensions.
Note: The GGM network presents the partial correlational associations between each token, based on their frequencies across user-specific corpora of posts. Solid edges represent positive associations, while dashed edges represent negative associations. These 136 tokens (nodes) represent the most frequently used words by the cohort and belong to one or more of the nine topic dimensions identified through TMFG-based EGA. Circled nodes represent a subset of the most central tokens and whose network statistics are presented in Table 3.

(including Sanctioned Suicide) are not seen as primary sources of negative sentiment and may be respites from the trials and tribulations of the real world. This finding is important since it highlights the selection biases that may exist in studies that leverage social media data, biases that stem from analyses conducted on an overrepresentation of individuals more likely to feel familiar and comfortable with the qualities, expectations, and norms of online social communication.[38,39] In addition, positive emotions tended to co-occur with entertainment topics such as "music," "dance," and "sports." This association may suggest that such topics serve as personally beneficial, protective, or community-fostering. "Reasons for Living" inventories have items relating to aspects of life that would be missed.[40,41] Thus, the emergent tie of positive emotion with these topics may highlight important focal points in intervention akin to what is exploited within behavioral activation-based therapies[42]—the reinforcement of personally enjoyable pastimes and the reminder that there is community and belonging through others who share similar interests.

Upon diving into the specific tokens of topics (Figure 3 and Table 3), a prominent representation of methods-related content and sharing activity was uncovered. This finding has also been observed in a separate analysis of the

Table 3. Network node statistics of the most frequent tokens across topic dimensions.

Token	Expected influence	Degree centrality	Betweenness centrality
face	1.435	1.435	3014
death	1.057	1.057	2366
taste	1.044	1.044	966
drinking	1.002	1.018	1274
bus	0.967	0.967	724
hell	0.938	0.938	2692
hope	0.933	1.047	1846
live	0.932	0.968	2386
sad	0.905	0.905	1808
drink	0.895	0.895	1918
talk	0.853	0.853	1034
help	0.824	0.824	720
dose	0.813	0.813	1388
agree	0.806	0.806	2452
alone	0.796	0.818	1564
thank	0.784	0.821	1252
old	0.753	0.753	2806
life	0.746	0.761	1936
head	0.733	0.752	1698
hour	0.724	0.724	844
love	0.721	0.721	248
bag	0.686	0.686	262

Note: Node statistics corresponding to a subset of the most central tokens in the estimated token-level GGM and labeled (circled) in Figure 3.

forum.[33] The presence of a close token association between "information" and "thread" is a consequence of the ubiquity of suicidal preparation discourse. While several additional examples of the topic-token-method associations are highlighted in Supplemental File 1 under the "EMPATH Topic Data" tab, the frequency and centrality of words in Figure 3 and Table 3 such as "drink"/"drinking," "dose,"

"hour," "taste," and "bag" represent prominent topical exposures within the community. This includes users who have some ideation and are looking to learn more about their options, to users who have already solidified plans to end their life but need clarification on one or more details. Moreover, several methods-related EMPATH topics were found to differ significantly between HRUs and controls with those topics presenting at a higher frequency in the former compared to the latter. While Sanctioned Suicide provides a place for sufferers to find social solace, it also hosts a substantial array of methods-related resources. This latter characteristic importantly indicates that methods searching is a popular online behavior among suicidal individuals, and thus it may help to flag such terminology within search engines to prevent exposure.

Identification of pertinent terminology is not straightforward, but this work highlights the critical importance of studying the *digital subculture* that has evolved around STB. Sanctioned Suicide is undoubtedly a major hub, and the data and results found phrases such as "catch the bus" and "exit bag," as well as intentionally obfuscating methods-related abbreviations such as "SN," which are part of common parlance. No studies to date have reported on these prominent verbal phenomena, which means that predictive and preventative models are not attuned to account for their presence and use. More research is needed on these fringe communities as they hold current linguistic and topical idiosyncrasies that may prove useful in uncovering risky online behaviors.

This investigation revealed several interesting patterns in the written expressions of STB online. However, there are important limitations to note. First, the use of EMPATH positioned the employed topic modeling procedure as a lexicon-based approach. While EMPATH is one of the most comprehensive, publicly available lexica available for topic modeling, competing with other highly popular and successful tools such as Linguistic Inquiry and Word Count (LIWC),[43] it suffers from issues of contextual agnosticism, like all lexica-based methods. Indeed, the employed pipeline was unable to account for irony, sarcasm, negations, unconventional meanings of words, or spelling errors. Naturally, any words that were not present in EMPATH were also ignored. This is reflected in the fact that only about 10–43% of a user's written content was used for topic modeling. While other studies may employ deep learning to counteract these limitations, such an approach would sacrifice the transparency and interpretability that lexicon-based methods like EMPATH provide in exchange.

Secondly, this study was exploratory, and the results are intended to be used for hypothesis generation and further consideration in follow-up works. While not necessarily a limitation, it is important to recognize that this research was intended to document and underline content as potential risk-signals and not to provide unequivocal evidence that certain language or topic emphases are reliable

markers of STB. Accordingly, interpretations of the data should be tempered to match the limitations of an exploratory framework and to seed subsequent research efforts.

Thirdly, the results of this work were useful in providing a general description of the conversational emphases on Sanctioned Suicide. They also serve as a cautionary tale to the broader dangers posed to at-risk suicidal individuals on the World Wide Web. However, it is unclear to what degree that findings made within this platform generalize beyond the borders of its unique "safe haven" of free speech, and whether they can be leveraged to describe and detect risk in other online communities. Moreover, the anonymous nature of the forum prevented the ability to profile the sociodemographic attributes of users, thus limiting insight into the cohort's constituency and representativeness. It is reasonable to assert that the platform's global and easily accessible presence as a surface web entity would attract a wide variety of individuals. However, any study that utilizes social media information is potentially impacted by selection and participation biases.[44] The behaviors of forum users and those who provide the majority of available data are likely not representative of a broader population consisting of individuals who do not actively engage with online communities or only do so under more specific conditions.

Conclusions

Sanctioned Suicide provides a rare opportunity to study STB in its most unmitigated form. This work presents a semantic analysis of its content and identifies themes that would likely be more difficult to uncover within more mainstream online communities. Positive language among the most willful and risk-prone, emphasis on methods seeking and sharing behaviors, use of euphemistic jargon, and themes of loneliness and life adversity describe shades of STB that are muted, silenced, or simply go unnoticed within the broader public eye. These facets deserve further attention within the suicidology research community as they may be able to bolster online detection and intervention efforts that rely on expert-curated digital markers of STB. Moreover, an understanding of how best to help suicidal individuals includes an appreciation of how they perceive, parse, discuss, and come to terms with their evolving state of mind. Sanctioned Suicide is a potent written account of these struggles and can serve as a means to this end. The current work represents one means of analytical distillation, one example of how, through creative methodology, the scientific community can work toward a more complete picture of this pernicious and debilitating condition.

ORCID iD: Damien Lekkas [iD] https://orcid.org/0000-0002-6995-9223

References

1. Ritchie H, Roser M and Ortiz-Ospina E. Suicide. *Our World in Data*, https://ourworldindata.org/suicide (2015, accessed 17 August 2022).
2. World Health Organization. Suicide Fact Sheet, https://www.who.int/news-room/fact-sheets/detail/suicide (2022, accessed 21 February 2023).
3. World Health Organization. Quality of Suicide Mortality Data. *WHO*, https://www.who.int/teams/mental-health-and-substance-use/data-research/suicide-data-quality (2022, accessed 28 February 2023).
4. Renjith S, Abraham A, Jyothi SB, et al. An ensemble deep learning technique for detecting suicidal ideation from posts in social media platforms. *J King Saud Univ-Comput Inf Sci* 2022; 34: 9564–9575.
5. Roy A, Nikolitch K, McGinn R, et al. A machine learning approach predicts future risk to suicidal ideation from social media data. *Npj Digit Med* 2020; 3: 1–12.
6. Sierra G, Andrade-Palos P, Bel-Enguix G, et al. Suicide risk factors: a language analysis approach in social media. *J Lang Soc Psychol* 2022; 41: 312–330.
7. Unruh-Dawes EL, Smith LM, Krug Marks CP, et al. Differing relationships between Instagram and Twitter on suicidal thinking: the importance of interpersonal factors. *Soc Media Soc* 2022; 8: 20563051221077028.
8. Sawhney R, Agarwal S, Neerkaje AT, et al. Towards suicide ideation detection through online conversational context. In: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. Madrid Spain: ACM, pp.1716–1727.
9. Castillo-Sánchez G, Marques G, Dorronzoro E, et al. Suicide risk assessment using machine learning and social networks: a scoping review. *J Med Syst* 2020; 44: 205.

10. Burnap P, Colombo G, Amery R, et al. Multi-class machine classification of suicide-related communication on Twitter. *Online Soc Netw Media* 2017; 2: 32–44.

11. Sarsam SM, Al-Samarraie H, Alzahrani AI, et al. A lexicon-based approach to detecting suicide-related messages on Twitter. *Biomed Signal Process Control* 2021; 65: 102355.

12. Ren F, Kang X and Quan C. Examining accumulated emotional traits in suicide blogs with an emotion topic model. *IEEE J Biomed Health Inform* 2016; 20: 1384–1396.

13. Gaur M, Alambo A, Sain JP, et al. Knowledge-aware assessment of severity of suicide risk for early intervention. In: WWW: International World Wide Web Conference. New York, NY, USA: Association for Computing Machinery, pp.514–525.

14. Tadesse MM, Lin H, Xu B, et al. Detection of suicide ideation in social media forums using deep learning. *Algorithms* 2020; 13: 7.

15. Young J, Bishop S, Humphrey C, et al. A review of natural language processing in the identification of suicidal behavior. *J Affect Disord Rep* 2023; 12: 100507.

16. Gillespie T. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. New Haven, CT, USA: Yale University Press, 2018.

17. Patty M. Social media and censorship: rethinking state action once again. *Mitchell Hamline Law J Public Policy Pract* 2019; 40: 99–136.

18. Sleeper M, Balebako R, Das S, et al. The post that wasn't: exploring self-censorship on Facebook. In: Proceedings of the 2013 Conference on Computer Supported Cooperative Work. San Antonio, Texas, USA: Association for Computing Machinery, pp.793–802.

19. Cobbe J. Algorithmic censorship by social platforms: power and resistance. *Philos Technol* 2021; 34: 739–766.

20. Just N and Latzer M. Governance by algorithms: reality construction by algorithmic selection on the internet. *Media Cult Soc* 2017; 39: 238–258.

21. Sanctioned Suicide: Rules & FAQs. *Sanctioned Suicide*, https://sanctioned-suicide.org/threads/rules-and-faq.4/ (2018, accessed 18 August 2022).

22. Richardson L. Beautiful soup documentation.

23. O'Connor RC and Kirtley OJ. The integrated motivational–volitional model of suicidal behaviour. *Philos Trans R Soc B Biol Sci* 2018; 373: 20170268.

24. Barbaro M. Kids Are Dying. How Are These Sites Still Allowed?, https://www.nytimes.com/2021/12/09/podcasts/the-daily/suicide-investigation.html (accessed 9 December 2021).

25. Fast E, Chen B and Bernstein M. Empath: understanding topic signals in large-scale text. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. San Jose, California: Association for Computing Machinery, pp.4647–4657.

26. Liu H and Singh P. Conceptnet–A practical commonsense reasoning tool-kit. *BT Technol J* 2004; 22: 211–226.

27. Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality. In: *Proc. NIPS 2013*. Lake Tahoe, NV, USA: Curran Associates, Inc.,

pp. 3111–3119. https://papers.nips.cc/paper/2013/hash/9aa42b31882ec039965f3c4923ce901b-Abstract.html (2013, accessed 14 November 2022).

28. Golino HF and Christensen AP. EGAnet: Exploratory Graph Analysis–A framework for estimating the number of dimensions in multivariate data using network psychometrics, github.com/hfgolino/EGA (2022).

29. Golino HF and Epskamp S. Exploratory graph analysis: a new approach for estimating the number of dimensions in psychological research. *Plos One* 2017; 12: e0174035.

30. Massara GP, Di Matteo T and Aste T. Network filtering for big data: triangulated maximally filtered graph. *J Complex Netw* 2017; 5: 161–178.

31. Blondel VD, Guillaume J-L, Lambiotte R, et al. Fast unfolding of communities in large networks. *J Stat Mech Theory Exp* 2008; 2008: P10008.

32. Epskamp S, Borsboom D and Fried EI. Estimating psychological networks and their accuracy: a tutorial paper. *Behav Res Methods* 2018; 50: 195–212.

33. Lekkas D, Oh J, Matsumura R, et al. Profiling the digital mosaic of uncensored suicidal thought and behavior: A theory-driven network analysis of online written expression. Epub ahead of print *PsyArXiV* 2023: 1–50.

34. Crane C, Crane RS, Eames C, et al. The effects of amount of home meditation practice in mindfulness based cognitive therapy on hazard of relapse to depression in the staying well after depression trial. *Behav Res Ther* 2014; 63: 17–24.

35. Kleiman EM, Coppersmith DDL, Millner AJ, et al. Are suicidal thoughts reinforcing? A preliminary real-time monitoring study on the potential affect regulation function of suicidal thinking. *J Affect Disord* 2018; 232: 122–126.

36. Klonsky ED and May AM. The Three-Step Theory (3ST): a new theory of suicide rooted in the "ideation-to-action" framework. *Int J Cogn Ther* 2015; 8: 114–129.

37. Van Orden KA, Witte TK, Cukrowicz KC, et al. The interpersonal theory of suicide. *Psychol Rev* 2010; 117: 575–600.

38. Baeza-Yates R. Bias on the web. *Commun ACM* 2018; 61: 54–61.

39. Hargittai E. Potential biases in big data: omitted voices on social media. *Soc Sci Comput Rev* 2020; 38: 10–24.

40. Linehan MM, Goodstein JL, Nielsen SL, et al. Reasons for staying alive when you are thinking of killing yourself: the reasons for living inventory. *J Consult Clin Psychol* 1983; 51: 276–286.

41. Osman A, Kopper BA, Barrios FX, et al. The brief reasons for living inventory for adolescents (BRFL-A). *J Abnorm Child Psychol* 1996; 24: 433–443.

42. Kanter JW, Manos RC, Bowe WM, et al. What is behavioral activation?: a review of the empirical literature. *Clin Psychol Rev* 2010; 30: 608–620.

43. Pennebaker JW, Francis ME and Booth RJ. *Linguistic inquiry and word count: LIWC 2001*. Mahwah, NJ, USA: Erlbaum Publishers, 2001.

44. Pokhriyal N, Valentino BA and Vosoughi S. Quantifying participation biases on social media. *EPJ Data Sci* 2023; 12: 1–20.