Cognitive network science quantifies feelings expressed in suicide letters and Reddit mental health communities

Simmi Marina Joseph
CogNosco Lab, Department of Computer Science
University of Exeter
Exeter, UK
sj504@exeter.ac.uk

Salvatore Citraro
Department of Computer Science, University of Pisa
Largo Bruno Pontecorvo, 3, Pisa
KDD-Lab (ISTI-CNR)
G. Moruzzi, 1, Pisa
salvatore.citraro@phd.unipi.it

Virginia Morini
Department of Computer Science, University of Pisa
Largo Bruno Pontecorvo, 3, Pisa
KDD-Lab (ISTI-CNR)
G. Moruzzi, 1, Pisa
virginia.morini@phd.unipi.it

Giulio Rossetti
KDD-Lab (ISTI-CNR)
G. Moruzzi, 1, Pisa
giulio.rossetti@isti.cnr.it

Massimo Stella
CogNosco Lab, Department of Computer Science
University of Exeter
Exeter, UK
m.stella@exeter.ac.uk

ABSTRACT

Writing messages is key to expressing feelings. This study adopts cognitive network science to reconstruct how individuals report their feelings in clinical narratives like suicide notes or mental health posts. We achieve this by reconstructing syntactic/semantic associations between concepts in texts as co-occurrences enriched with affective data. We transform 142 suicide notes and 77,000 Reddit posts from the r/anxiety, r/depression, r/schizophrenia, and r/do-it-your-own (r/DIY) forums into 5 cognitive networks, each one expressing meanings and emotions as reported by authors. These networks reconstruct the semantic frames surrounding "feel", stem for "to feel" and "feelings", enabling a quantification of prominent associations and emotions focused around feelings. We find strong feelings of sadness across all clinical Reddit boards, added to fear r/depression, and replaced by joy/anticipation in r/DIY. Semantic communities and topic modelling both highlight key narrative topics of "regret", "unhealthy lifestyle" and "low mental well-being". Importantly, negative associations and emotions co-existed with trustful/positive language, focused on "getting better". This emotional polarisation provides quantitative evidence that online clinical boards possess a complex structure, where users mix both positive and negative outlooks. This dichotomy is absent in the r/DIY reference board and in suicide notes, where negative emotional associations about regret and pain persist but are overwhelmed by positive jargon addressing loved ones. Our quantitative comparisons provide strong evidence that suicide notes encapsulate different ways of expressing feelings compared to online Reddit boards, the latter acting more like personal diaries and relief valve. Our findings provide an interpretable, quantitative aid for supporting psychological inquiries of human feelings in digital and clinical settings.

Keywords Complex Networks · Topic Modeling · Emotions · Corpus · Concepts

1 Introduction

Humans' cognition has no equal. Our structured emotions and keen social attitudes lead us to develop sympathy, altruism and cooperation, among the wide range of social ties we can form [1]. However, there exists a downside of such a natural social willingness. It concerns individuals' difficulties to get integrated into the shared values of a culture, an anomie [2] or anomaly that also concerns individuals' inability to understand, express or regulate their own emotions. Complex circumstances can lead to unbearable situations and towards behaviours that are unique in nature, where people could reinforce negative emotions and isolate themselves [3]. The most extreme consequences of such behaviours lead to drastic, often non-reversible, solutions [4]. Among them, the willingness to die is the most alarming and tragic one. According to the World Health Organization, more than 7,00,000 people commit suicide every year, and it is reported as the fourth leading cause of death among 15–29-year-olds[1]. Suicide can be experienced as the ultimate step of a long period of distress caused by clinical pathologies like anxiety or depression or by clinical psychosis like schizophrenia [5]. Victims' reluctance to speak about their feelings to a near person makes suicide hard to predict and understand [6]. Nevertheless, interestingly, people tend to be more inclined to use online social platforms to talk about their own deeply intimate emotions to other people who share similar problems [7, 8], a human tendency analogous to homophily in social networks analysis, i.e. a bias for social ties to be established between similar actors [9]. Due to its forum structure as well as users' anonymity, Reddit comes as one of the most promising social platforms to characterize mental health communities i.e., actual pools where people aim to gain information about their health problems, share their symptoms and difficulties, or give support to others having similar problems [7, 10]. Several recent studies have started using clinical boards on Reddit as data sources for identifying suicide ideation and other disorders linked with emotional distress [8, 11, 12, 13] and we proceed along this direction.

In this paper, we aim to understand both online users' and suicide victims' language to uncover their underlying emotional patterns and quantify their degree of similarity and variability. In the era of Big Data, healthcare systems are increasingly supported by Artificial Intelligence (AI) tools that can allow to analyze suicide note contents and extract patterns and features from texts [12, 14, 5] but often at the cost of losing information about the interpretation of a machine learning classification/regression model [15]. We provide a novel, interpretable approach to understand and compare the thoughts of suicide victims with those of individuals experiencing mental health conditions. In particular, we leverage a simple, interpretable, cognitive network methodology that directly compares individuals' feelings when expressing their narrative in suicide notes and clinical Reddit boards. This method may open new ways for clinicians to deeply understand the relatability of suicide with any mental condition, exploring further possibilities to move towards the direction of predicting suicidal tendencies in the vulnerable category. It can also open additional ways to detect if a suicide victim had any potential mental disorders.

The rest of the paper is organized as follows. In Section 2, we provide a broad background framing this work. Section 3 describes the methods used to exploit our analysis, spanning from novel cognitive network science to classic NLP

---

[1]"Mental Health and Substance Use." `https://www.who.int/teams/mental-health-and-substance-use/suicide-data` (accessed Mar. 22, 2021)

tools, together with the datasets used for the analysis. Section 4 sums up the main results of the paper, while Section 5 discusses the most promising directions of our work as well as its limitations and future lines of research.

2 Related Works

Text represents an insightful fragment of the psychology of individuals [16, 17, 18]. Within the context of clinical text analysis, several works trying to analyze and understand the content of suicide notes focused on finding dominant emotional words (e.g., "love") or the reason behind the gesture [19, 20, 21]. In most of the analyzed letters, individuals express the willingness to escape from the pain and the anger towards other people and, interestingly, letters from completed suicides express more self-blame than those from attempted suicides [19]. Early studies also tried to characterize the differences between the text features hidden in suicide notes compared to other types of text. As an example, an early study on a corpus of 66 suicide notes where half were genuine and half were simulated aimed to identify hidden features that can classify genuine and fake notes [22]. More recently, the study of mental health discussions and suicide contents in online pools focuses predominantly on Natural Language Processing (NLP) techniques, often aiming to extract linguistic features for classification, e.g., predicting whether a textual content contains early signs of suicidal ideation [23, 12, 14] or more in general, signs of mental health disorders [24, 13, 25]. For instance, Nikfarjam and colleagues [26] defined a framework to label 900 genuine suicide notes with their underlying emotions by firstly extracting syntactic and semantic features, thus using them to define the input rules for a machine learning classifier. Similarly, Low and colleagues [25] rely on regression and 90 text features derived from Reddit (e.g., sentiment analysis, semantic categories, personal pronouns) to verify whether the COVID-19 pandemic was impacting mental balance. It is relevant to notice that most of those works [24, 12, 13] leverage as primary source Reddit mental health communities since they occur as a helpful source for monitoring people's perceptions about their conditions through the emotional language they use to convey feelings.

However, a recent discussion has underlined the risk of dehumanizing individuals while studying mental health prediction on social media [27]. Indeed, using binary classification, as in the previous examples, and thus reducing mental health status to two corresponding machine learning classes (e.g., "positive"/"negative" or "suffering"/"not suffering") produces severe biases that flatten the complexity of mental illness.

Another branch of NLP that has been fruitfully used both in suicide notes as well as in online forums analysis to capture topical patterns in the mental health discourse is Topic Modeling [28]. Topic modeling techniques are able to detect the hidden topics - lists of words - that characterize a set of textual documents, and this can help researchers in finding the events that led to suicide ideation or the most debated issues that grieve on individuals affected by mental health disorders [29, 11]. For instance, Xue and her team leverage a popular topic modeling technique, i.e., Latent Dirichlet Allocation (LDA), to detect individuals' key themes and emotional reactions during the early stage of COVID-19 [30]. The main limitation of this approach is that the definition of "topic" is loose and with no clear cognitive counterpart, since it is merely based on statistical co-occurrences that: (i) crucially depend on the partitioning of text across documents, (ii) can change according to an external parameter, i.e. an arbitrary number of desired topics,

3

(iii) lead to multiple topics sharing significant fractions of words. Despite the presence of heuristics that can partially resolve these issues, LDA should be considered as a technique to use in synergy with other cognitively-grounded methods like word count inquiry or network analysis.

A promising tool for identifying emotional and semantic features in texts are Cognitive Networks [31, 32]. Leveraging graph theory concepts, cognitive networks allow the analysis of and on complex structures in the human mind apt at storing and processing knowledge. Cognitive networks of conceptual associations aim to provide results that are interpretable from a cognitive point of view, e.g., being used in lexical retrieval tasks, early language learning studies, and, in general, whenever a cognitive process can be applied on a complex network structure [33, 31, 34]. Such methods are transparent and help to characterize, among others, the most central words in a network as well as the positive or negative emotions surrounding word neighborhoods [17]. Another important aspect of cognitive networks is their natural tendency to form modular clusters or communities [33], i.e., sets of well-connected nodes more similar to each other rather than the rest of the graph, which can be exploited, among others, to study users' sentiment on social media [35]. Applying cognitive networks on corpora of suicide notes can be useful for further understanding and getting meaningful insights into such rich text data, e.g., analyzing how some concepts tend to cluster and influence the overall structure of the network [36]. In fact, the associations attributed to a given concept in text represent its so-called semantic frame [37]. Semantic frame theory indicates that meaning and perceptions relative to a concept can be reconstructed by analysing its semantic frame [38, 37] and cognitive networks enable this reconstruction in a quantitative way, allowing for researchers to reconstruct the structure of semantic/syntactic associations in texts and to focus attention over specific ideas reported in such networked structure [38, 17, 39].

In the following work we rely on a combination of methods from cognitive network science and NLP to study how people express their feelings through mental health online boards and genuine suicide notes.

3 Methods

This section outlines the key resources and methodologies adopted in this study.

3.1 Textual data: Genuine suicide notes and clinical Reddit posts

This work used letters and posts as personal narratives providing a glimpse into the psychology of individuals [18]. Specifically, we used suicide notes and Reddit blogposts from schizophrenia, depression, anxiety and "Do It Your own" (DIY).

The 142 suicide notes used here were curated by Schoene and Dethlefs [23] and investigated also in [36]. These English letters were authored by people who successfully took their lives. On average, a suicide note included 120 words. No additional information (e.g. socio-economic context) was available in the dataset. All notes are anonymised to protect the individuals' privacy, in adherence with the Declaration of Helsinki.

Building our study on feelings expressed in suicide notes, we searched for other datasets sharing analogous clinical features. We selected a collection of 77,000 blog posts mined from Reddit by Kim and colleagues [13]. These texts indicated online posts on subreddits like r/schizophrenia, r/depression, and r/anxiety and were used by [13] to implement deep learning binary classifiers detecting the presence of clinically negative emotional states. The r/DIY subreddit was included as a reference poll, including texts that were produced online but coming from individuals not engaged in mental health discussions. These posts were scraped using ParseHub[2] and anonymised by removing user details and names. In this instance, we focused only on these forums, though our approach can be extended to include multiple subreddit or online forums.

3.2 Text cleaning, network construction and centrality, affective enrichment

With NLTK in Python, texts were tokenized and reduced to sequences of sentences, each one containing a sequence of words. Word ordering was preserved but punctuation, symbols, hyperlinks and numbers were discarded. Individual words were stemmed, thus reducing the occurrence of multiple word forms relative to the same concept/idea (e.g. "sadness" and "sad" both became the stem "sad"). For stemming we used Porter's algorithm as implemented in NLTK. With NLTK, stopwords excluding meaning negations - like "no" and "not" - were removed from all word lists. We then drew co-occurrence links between adjacent words, as performed in previous approaches using co-occurrence networks for authorship identification [40] or semantic framing in suicide notes [36]. Co-occurrence links are also known as bi-grams, as they indicate the occurrence of two lexical items together. From a network perspective, counting the frequency of bi-grams was useful for filtering infrequent co-occurrences. However, once thresholded, we treated co-occurrence links as undirected and unweighted.

Co-occurrence networks capture the syntactic structure linking concepts in sentences and attributing meaning to narratives [38, 16]. However, when working with corpora differing in size and length (of words), it can be difficult to build networks possessing comparable/analogous connectivity among the same set of nodes. To overcome this limitation, we used suicide notes - our smallest dataset - as a baseline. It contained 6465 bi-grams/links. We then performed the same co-occurrence/bi-gram counts for other corpora and considered only the top-ranked 6465 bi-grams to build co-occurrence networks representing the narratives of such online clinical populations. This hard constraint enabled the construction of networks possessing equivalent link connectivity and it filtered strong conceptual associations in Reddit polls. While we are aware of the existence of a variety of methods for filtering word-word relationships (like minimum spanning trees [41]), it is difficult to gain insights about the effects that these methods have over selecting conceptual links, so that we resorted to a simpler frequency-based pruning, instead. This led to the creation of 5 different co-occurrence networks, each one representing the structure of conceptual associations as expressed by authors in their notes/posts. A network visualisation, built with Gephi, is reported in Figure 1 as an example. Notice how "I" and "you" are central in the semantic/syntactic structure extracted from suicide notes. Then, to further characterize the co-occurrence networks, we leverage closeness centrality. Closeness is a well-known centrality measure that identifies

---

[2]https://www.parsehub.com/, Last Accessed: 08/10/2021

how close a node is to all other nodes in a network in terms of its average network distance [33]. In cognitive network science, network distance was shown to strongly correlate with semantic relatedness, so that concepts further apart on a semantic network were found to be less semantically related [42]. Hence, nodes with a higher closeness centrality are expected to be more semantically related to other connected concepts as mentioned in text, which was further confirmed by more recent studies [17]. Taking inspiration from past approaches in cognitive network science [39], we performed

Figure 1: Co-occurrence network of concepts as expressed in 139 genuine suicide notes.

an emotional enrichment of co-occurrence networks. We endowed stems with valence and emotional attributes as coming from the Valence-Arousal-Dominance [43] and from the Emotion Lexicon [44] datasets, respectively. Valence is a psychological metric estimating how pleasant/unpleasant concepts are perceived. Words were attributed "positive" (upper quartile), "negative" (lower quartile) or neutral (otherwise) labels, purely for visualisation purposes. Words were also attributed a list of emotions they elicited according to the cognitive data gathered by [44].

3.3 Semantic frame theory and Latent Dirichlet analysis

Co-occurrence networks are representations of the ways authors associate concepts in their own narratives [38, 40]. Considering a network neighbourhood for a word W identifies all concepts associated to it in text, i.e. the semantic frame of associates providing meaning and content to W itself [37].

Another way to reconstruct a semantic frame is to identify a topic, i.e. a class of highly co-occurring concepts in a set of documents. To implement this second characterisation of semantic frames we adopted Latent Dirichlet analysis, as implemented in the Gensim and pyLDAvis packages in Python. To fix the number K of topics to search for we adopted information coming from network structure. We used the Louvain method [45] to count how many communities C populated the semantic frame of "feel" in a given network. We then used this number, |C|, as input for detecting K = |C| topics via LDA, but only in documents mentioning at least once either "feel" or its different word forms (e.g. "feeling"). Since in LDA the same concept can be repeated across different topics, we selected as potential semantic frame the topic where "feel" occurred with the highest frequency. Notice that this LDA analysis offers only a limited understanding of which words were associated with "feel", so that we used it only to validate results coming from the network analysis. In our further investigations of the emotions populating the semantic frame of "feel" across suicide letters and blog posts we focused over network neighbourhoods/semantic frames.

3.4 Emotional profiling

The emotions attributed to individual words [44] were used to explore how individuals felt in their online narratives. Emotional profiling was applied to the semantic frame of "feel" across all reconstructed networks. For each semantic frame/network neighbourhood, we measured the fraction $r_i = m/N$ of m words inspiring emotion i in a neighbourhood with N entries. The collection of all fractions r_i constitutes the so-called emotional profile of a given semantic frame []. Notice that word negations were considered in these counts: As in [17], antonyms of words linked with negations (like "not") were added to the count. Then, the emotional profile was matched against 1000 random counterparts, all built by sampling uniformly at random from the Emotion Lexicon the same number of words eliciting at least one emotion observed in the empirical profile. The empirical and random scores were used to compute z-scores indicating the strength of the observed emotion in a given semantic frame. Z-scores were plotted as petals, in a visualisation reminiscent of Plutchik's wheel of emotions [46] and that we call emotional flower [17]. Petals falling outside of a rejection region $z < 1.96$ (semi-transparent circles) are indicative of strong emotional intensities populating a given frame.

4   Results

We present our results starting from simple, unstructured frequency analyses in texts and semantic prominence rankings in cognitive networks. Finding evidence that authors of texts expressed their feelings in different ways across corpora, we move to more specific investigations of semantic frames for "feel". We present key findings about different emotional profiles of "feel" between suicide notes and Reddit forums. We conclude by validating our results via a combination of community analysis and LDA.

4.1 Prominent concepts across suicide notes and clinical Reddit posts

Table 1 reports the top-20 most semantically prominent concepts as measured by closeness centrality and frequency (see Section 3) across all the corpora. Results highlight that the first-person pronoun is the most central concept across all the considered networks, either according to frequency (of mentions) or closeness (i.e. semantic prominence). In suicide notes, "I love you" is the set of most prominent/frequent words: People who committed suicide expressed prominently their love to their dear ones. In the clinical subreddits, key concepts are relative to distress and feelings (e.g. "anxiety" being more frequent than "you" in r/Anxiety, "feel" being prominent/frequent across all clinical boards). This indicates a difference between suicide letters, addressing expressions of love, and subreddits, reporting personal feelings and experiences. Interestingly, this difference supports previous studies underlining strong self-attentional focus in narratives produced by clinically anxious individuals [11, 36]. Bi-gram analysis confirms this difference

Suicide Notes		Anxiety		Depression		Schizophrenia		DIY	
Freq.	Clos.	Freq.	Clos.	Freq.	Clos.	Freq.	Clos.	Freq.	Clos.
I	I	I	I	I	I	I	I	I	I
you	you	anxiety	anxiety	you	you	he	he	wall	wall
love	love	go	go	she	she	you	you	paint	want
go	go	*feel*	*feel*	go	go	they	they	need	paint
he	he	you	you	*feel*	*feel*	go	go	look	need
will	will	time	time	time	time	she	she	use	look
life	life	make	make	life	life	*feel*	*feel*	go	use
please	please	she	they	depress	depress	think	think	new	go
take	take	they	she	they	they	people	people	way	new
way	way	think	think	friend	friend	make	make	no	way
know	no	work	work	myself	myself	time	time	make	no
want	one	really	really	make	make	up	thing	side	make
make	thing	take	take	people	people	know	know	water	side
help	know	he	thing	even	even	take	take	want	aer
tell	want	people	he	want	want	really	one	work	water

Table 1: Top 15 concepts ranked in the order of decreasing frequency in corpora and closeness centrality in networks. The concept "feel" is highlighted in red.

between love-focused suicide letters and self-focused online posts (see Supplementary Table S4). In addition, suicide notes are found to frequently mention bi-grams expressing hope and regret (e.g. "I - hope" or "I - sorry", which misses the stopword "am" because of network construction), which are missing in subreddit texts. The latter narratives focus more on expressing individual needs and mental distress (e.g. "I - depress", "I - feel" and "feel - like").

Overall, the above indicates that the texts analysed here crucially reported how individuals expressed their own feelings, perceptions and states, thus motivating further analysis.

4.2 Cognitive networks quantify different feelings across suicide notes and clinical Reddit posts

Figures 2 and 3 report the semantic and emotional content associated with "feel" in suicide letters and Reddit posts. All figures report the semantic network community of "feel" (left), i.e. focusing on concepts most tightly associated with "feel", and the emotional flower relative to all associations framing "feel" (right). These results are based on our cognitive network of co-occurrences/emotional labels, capturing how text authors associated ideas and concepts in their narratives.

In r/Anxiety, Figure 2 (top), authors use "feel" to express an overwhelmingly negative jargon, including several negative emotional states like "loneliness", "guilt" and "helplessness". Semantic content relative to "panic", "hopelessness", "stress", concepts like being "worthless" and "overwhelmed" and mentions of body parts like "stomach" and "hand", all strongly indicate mentions of panic attacks, which are frequently experienced by individuals suffering from anxiety disorders [47]. These concepts elicit strong levels of sadness, as reported in the emotional flower. However, this emotion co-exists with trust, i.e. a feeling of confidence and security in an idea [48]. Trust spawns from positive jargon linked with "feel", like "calm", "begin", "kind" and "relax", which indicate positive language mixing with negative ideas in subreddit posts. The co-existence of positive and negative concepts leads to the emotional polarisation between sadness and trust observed in the semantic frame of "feel" for users of r/Anxiety. Similar semantic content is observed for online users in the r/Depression board (Figure 2, middle), where the emotional polarisation is stronger, contrasting stronger-than-expected levels of sadness and fear against joy and trust. Fear spawns from jargon indicating uselessness and emptiness, e.g "fake", "pathetic", "dead", and including mentions to other disorders like "anxiety" and "suicide". These communicative patterns indicate the presence of a strong overlap in feelings between depression and other clinical conditions, as supported by clinical comorbidity studies [4]. Despite this commonality, users in r/Depression express their feelings with stronger fear and joy than users in r/Anxiety, underlining an important difference between these distress signals that is only recently being remarked by psychologists through network psychometrics [49].

In the r/DIY board (Figure 2, bottom), "feel" is more peripheral than in other boards: it has a lower degree and a more sparse semantic frame. "Feel" in r/DIY is framed with joy and anticipation, e.g. cheerful planning. This is expected from a board where users report non-clinical stories about home repairing and it indicates that our methods are not biased to always detect negative emotional states.

Whereas feelings expressed in r/Depression and r/Anxiety feature sadness as being prevalent, r/Schizophrenia is more deeply polarised, featuring equally strong levels of sadness and trust simultaneously (see Figure 3, top). Schizophrenia is a severe mental illness where thinking and perceptions are strongly distorted, with symptoms like hallucinations, paranoia and disarticulated speech in addition to comorbidities with anxiety and depression [50]. This clinical connotation is validated by the narratives expressed by online users, which mention anxiety, depression and hopelessness. Interestingly, these negative concepts are mixed with positive/trusted ones, like "happy", "alive", "comfort", "strong" and "good". Notably, the semantic community of "feel" in r/Schizophrenia includes links with "connect", "detached"
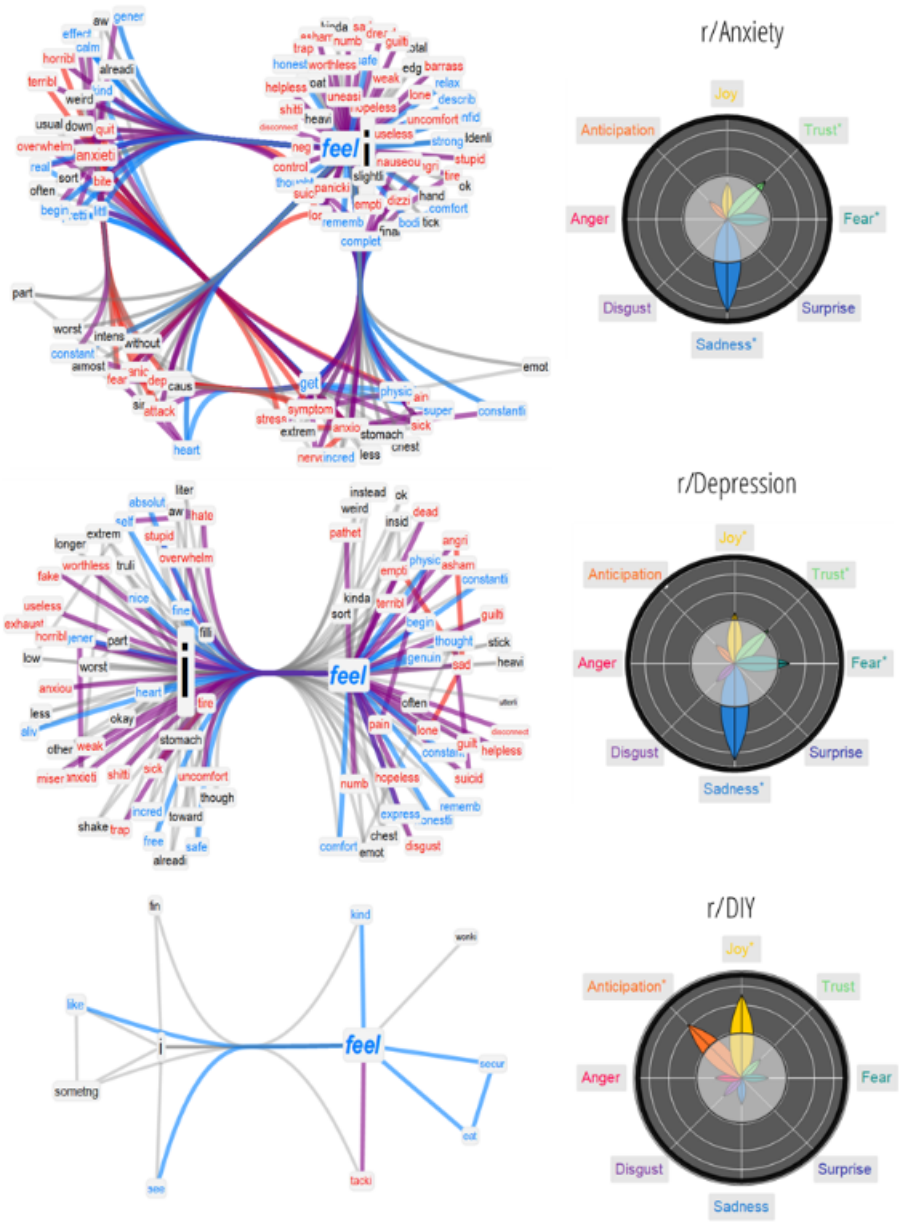
Figure 2: Cognitive networks (left) and emotional flowers (right) for the semantic frames of "feel" across r/Anxiety (top), r/Depression (middle) and r/DIY (bottom). Links in cognitive networks indicate concept co-occurrence. Colours indicate valence, i.e. positive (cyan), negative (red) or neutral (black) valence. Links between positive/negative concepts are in cyan/red. Purple links connect concepts of contrasting valence. Flowers are made of petals representing z-scores of emotional counts (see Methods). Petals falling outside the semi-transparent rejection region (z < 1.96) indicate stronger than expected concentrations of concepts eliciting a given emotion.

and "disconnected", which relate to the well-documented sense of isolation that hallucinations and paranoia can create in patients affected by schizophrenia [51].

Suicide letters frame "feel" in a different way compared to all the above clinical subreddits (see Figure 3, bottom). No strong emotional signal is found in the semantic frame of "feel", which features only a few negative associations

10

Figure 3: Cognitive networks (left) and emotional flowers (right) for the semantic frames of "feel" across r/Schizophrenia (top) and suicide notes (bottom). Links in cognitive networks indicate concept co-occurrence. Colours indicate valence, i.e. positive (cyan), negative (red) or neutral (black) valence. Links between positive/negative concepts are in cyan/red. Purple links connect concepts of contrasting valence. Flowers are made of petals representing z-scores of emotional counts (see Methods). Petals falling outside the semi-transparent rejection region (z < 1.96) indicate stronger than expected concentrations of concepts eliciting a given emotion.

(with "empty", "tired" and "sick") overwhelmed by generally positive jargon (e.g. "want", "make", "person"). This lack of emotional content indicates that suicide notes are substantially different from clinical subreddits. Our results indicate that whereas online forums showcase emotional signals and semantic content relative to personal diaries, where people report how they feel in relationship with their conditions, suicide notes are not informative about how individuals directly feel. In fact, no expected distress pattern relative to suicide ideation, like anxiety or depression [4], is found in narratives from suicide letters. These last notes frame "feel" in general terms, mentioning mostly people and actions, as expected in reports of events.

4.3 Results of Topic Analysis

To further support the above results about emotional polarisation and relationships of feelings with clinical conditions, we examined our textual corpora through a Latent Dirichlet Analysis (see Section 3). Table 2 reports words in the topic where "feel" was the most frequent, ranking them according to their frequency. The r/DIY board was discarded from the analysis as it showcased distinct patterns from other boards.

All clinical boards and suicide notes express feelings as related with the passage of time (e.g. "day", "hour", "time"). Whereas clinical boards consistently mention social environments (e.g. "work", "friend", "people"), suicide notes prominently express feelings of love that are missing in clinical boards. This confirms that suicide letters are more personal and less formal narratives, targeting loved ones rather than general online audiences. The co-occurrence of "feel", "want", "help" and "tell" in clinical subreddits indicate a willingness for authors to seek aid and comfort, which further confirms the semantic/emotional signals of distress and trust found in the previous section. As indicated by narrative psychology, narratives can help people understand and confront their issues [16]. Our network and LDA analyses provide strong evidence that this help is explicitly mentioned in subreddits when individuals report their well being or support each other.

Suicide notes mention requests for help to a lesser extent than clinical subreddits, and rather focus feelings towards being "sorry" for having to "leave" and depart from "love"(d) ones (cf. Table 2). This distinctive trait shifts emotional connotations from expressions of feelings to final salutations, providing a reason for the absence of strong emotional signals from the network analysis. This also indicates that suicide notes happen at a later stage than mere diaries, when individuals have already processed their mental distress and do not want to express it anymore in their final words bidding farewell to loved ones.

5 Discussion

In the present work, we used a combination of methods from cognitive network science and NLP to study how people express their feelings through mental health online boards (i.e., r/anxiety, r/depression, r/schizophrenia, and r/do-it-your-own) and 142 genuine suicide notes. In this final Section, we focus our attention on the emotional profiles derived from the word "feel" in these two different scenarios. Indeed, the rich emotional structure conveyed by this word reveals non-trivial, heterogeneous associations within the mental health communities but a different, more neutral profile in the suicide notes.

Focusing on the Reddit boards firstly, we observe that "feel" is not related to a unique, mono-dimensional emotion, or even to a primary dyad [46]: its semantic frame is populated by pairs of unexpected, often contrasting emotions, as in the polarised "sadness/trust" example from r/schizophrenia or, to a smaller extent, the "sadness/joy" pair from r/depression. It is also interesting to mention that only a "homogeneous" emotional dyad appears in r/DIY, i.e., the "anticipation/joy" dyad, indicating how mental health communities are more complex than other thematic subreddits where people ask questions to solve problems on their own. Such a dyad indeed is a marker of "optimism" [46], an attitude that may fit

Rank	Suicide Notes	r/Anxiety	r/Depression	r/Schizophrenia
1	love	go	know	want
2	go	know	want	know
3	know	time	go	time
4	want	think	depressed	year
5	please	want	think	day
6	life	really	time	really
7	take	thing	really	take
8	live	year	life	try
9	time	say	day	work
10	leave	anxiety	year	life
11	way	life	try	even
12	money	try	work	tell
13	god	need	help	help
14	last	friend	people	friend
15	sorry	tell	never	need

Table 2: Top 15 concepts in the topic including "feel" across text corpora, as obtained from Latent Dirichlet Allocation. Stopwords were removed to focus on semantic content.

well with people communicating their desire to do things by themselves. Conversely, the heterogeneous emotional profiles of "feel" in the clinical boards highlight how interactions on digital platforms are not all the same [6], and a variety of users could be present within them for different reasons and needings. For instance, there can be both people asking for support and people giving that support, where the former ones amplify the "negative" petal of "feel"'s frame, and the latter ones represent the opposite emotional flower. An addendum for this, and also a parallel interpretation, can account for the presence of several users at different stages of a mental condition [49]. Hence, users who need to express their feelings or ask for support can listen and perhaps learn from the previous experiences of other users. It is clear that only further analyses could validate such interpretations through additional data mining. Moving to suicide notes analysis, we find that people tend to express their feelings in a quite different way with respect to Reddit boards. No strong emotional concepts emerge in the semantic frame of "feel" while studying the network of suicide notes, neither positives nor negatives. The fact that people tend to express feelings in an emotionless way could suggest that suicide letters are not meant to report how individuals who are going to end their lives really feel. Indeed, as emerging from previous analyses of the semantic frames of "life" and "love" [36], people who are going to commit suicide are more prone to express their gratitude and love to their families and friends instead of explicitly mention how they feel.

Accordingly, our analysis provides quantitative evidence that if we want to predict suicide ideation through expressions of feelings, we should concentrate on earlier stages of mental distress, as they can be revealed more prominently from online interactions in mental health communities rather than from suicide notes. The "online diary" nature of subreddit boards, for instance, makes them helpful tools for the early-stage prediction of mental distress [24, 25, 36]. Indeed,

13

preliminary studies already exploited Reddit boards for monitoring shifts from mental health disorders to suicidal ideation [8, 14, 5]. A list of distinct markers characterizing these shifts was suggested in this regard, involving heightened self-attentional focus, poor linguistic coherence, and strong manifestation of anxiety, impulsiveness, hopelessness, and loneliness [8]. Some of them can fit with the negative emotional petals found in each board, i.e., hopelessness and loneliness moods, while others might explain the heterogeneity of the semantic frames, e.g., the poor linguistic coherence, whereas such positive emotions come from the same individuals sharing also their emotional distress. However, to the best of our knowledge, the only work that leverages semantic networks to analyze mental health discourse on Reddit boards is the one by Yoo and colleagues [11]. Like us, the authors define the semantic networks of three mental health-related subreddits and thus rely on topic modeling techniques to find discourse patterns. Even if they do not analyze specific semantic frames, the results obtained in r/depression are in line with what we have observed. Indeed, on the one hand, they found a prevalence of "sadness" feeling all over the network but also a significant amount of joyful and positive expressions. On the other, as in our study, topic analysis suggests the particular attention given by mental health users to time-related words (e.g., "weeks", "past", "years") as well as school and friends issues (e.g., "no friends", "social skills", "high school").

Nevertheless, there are some limitations to our current research that also motivates future works. Our analysis aggregates time and individual users, providing a collective quantification of semantic frames that can be opened in future research. Considering a longitudinal approach of time-evolving social interactions (e.g., users remove their accounts, new friendships are made) and accounting for feelings expressed by individuals both represent interesting directions for future research. Although we plan to extend our work by including the "time" features as well as tracking individual emotional evolutions over time by considering individual semantic networks, such a fine-grained perspective could lead to ethical and privacy issues as to preserve users' anonymity, especially in online mental health studies. Regarding the current work, the networks that we build aggregate multiple groups from different perspectives, so we overcome the anonymity issue by making it impossible to identify individual users. Then, future research calls for identifying valid approaches aiming at combining the fine-grained analyses we need for validating our interpretation and the issues related to user anonymity.

Acknowledgments

---

[3]Last Accessed: Oct, 28, 2021

References

[1] Robin IM Dunbar. Coevolution of neocortical size, group size and language in humans. *Behavioral and brain sciences*, 16(4):681–694, 1993.

[2] Emile Durkheim. *Le suicide: Etude de sociologie*. République des Lettres, 2020.

[3] Michele M Tugade, Barbara L Fredrickson, and Lisa Feldman Barrett. Psychological resilience and positive emotional granularity: Examining the benefits of positive emotions on coping and health. *Journal of personality*, 72(6):1161–1190, 2004.

[4] Borjanka Batinic, Goran Opacic, Tijana Ignjatov, and David S Baldwin. Comorbidity and suicidality in patients diagnosed with panic disorder/agoraphobia and major depression. *Psychiatria Danubina*, 29(2):186–194, 2017.

[5] Manas Gaur, Vamsi Aribandi, Amanuel Alambo, Ugur Kursuncu, Krishnaprasad Thirunarayan, Jonathan Beich, Jyotishman Pathak, and Amit Sheth. Characterization of time-variant and time-invariant assessment of suicidality on reddit using c-ssrs. *PloS one*, 16(5):e0250448, 2021.

[6] Umashanthi Pavalanathan and Munmun De Choudhury. Identity management and mental health discourse in social media. In *Proceedings of the 24th International Conference on World Wide Web*, pages 315–321, 2015.

[7] Munmun De Choudhury and Sushovan De. Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In *Eighth international AAAI conference on weblogs and social media*, 2014.

[8] Munmun De Choudhury, Emre Kiciman, Mark Dredze, Glen Coppersmith, and Mrinal Kumar. Discovering shifts to suicidal ideation from mental health content in social media. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 2098–2110, 2016.

[9] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1):415–444, 2001.

[10] George Gkotsis, Anika Oellrich, Tim Hubbard, Richard Dobson, Maria Liakata, Sumithra Velupillai, and Rina Dutta. The language of mental health problems in social media. In *Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology*, pages 63–73, 2016.

[11] Minjoo Yoo, Sangwon Lee, and Taehyun Ha. Semantic network analysis for understanding user experiences of bipolar and depressive disorders on reddit. *Information Processing & Management*, 56(4):1565–1575, 2019.

[12] Shaoxiong Ji, Celina Ping Yu, Sai-fu Fung, Shirui Pan, and Guodong Long. Supervised learning for suicidal ideation detection in online user content. *Complexity*, 2018, 2018.

[13] Jina Kim, Jieon Lee, Eunil Park, and Jinyoung Han. A deep learning model for detecting mental illness from user content on social media. *Scientific reports*, 10(1):1–6, 2020.

[14] Michael Mesfin Tadesse, Hongfei Lin, Bo Xu, and Liang Yang. Detection of suicide ideation in social media forums using deep learning. *Algorithms*, 13(1):7, 2020.

[15] Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5):206–215, 2019.

[16] Michael Murray. Narrative psychology and narrative analysis. 2003.

[17] Massimo Stella. Text-mining forma mentis networks reconstruct public perception of the stem gender gap in social media. *PeerJ Computer Science*, 6:e295, 2020.

[18] Joshua Jackson, Joseph Watts, Johann-Mattis List, Ryan Drabble, and Kristen Lindquist. From text to thought: How analyzing language can advance psychological science. 2020.

[19] Alison Brevard, David Lester, and Bijou Yang. A comparison of suicide notes written by suicide completers and suicide attempters. *Crisis: The Journal of Crisis Intervention and Suicide Prevention*, 1990.

[20] Tom Foster. Suicide note themes and suicide prevention. *The International Journal of Psychiatry in Medicine*, 33(4):323–331, 2003.

[21] Lori D Handelman and David Lester. The content of suicide notes from attempters and completers. *Crisis*, 28(2):102–104, 2007.

[22] Edwin S Shneidman and Norman L Farberow. Some comparisons between genuine and simulated suicide notes in terms of mowrer's concepts of discomfort and relief. *The Journal of general psychology*, 56(2):251–256, 1957.

[23] Annika Marie Schoene and Nina Dethlefs. Automatic identification of suicide notes from linguistic and sentiment features. In *Proceedings of the 10th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, pages 128–133, 2016.

[24] Judy Hanwen Shen and Frank Rudzicz. Detecting anxiety through reddit. In *Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology—From Linguistic Signal to Clinical Reality*, pages 58–65, 2017.

[25] Daniel M Low, Laurie Rumker, Tanya Talkar, John Torous, Guillermo Cecchi, and Satrajit S Ghosh. Natural language processing reveals vulnerable mental health support groups and heightened health anxiety on reddit during covid-19: Observational study. *Journal of medical Internet research*, 22(10):e22635, 2020.

[26] Azadeh Nikfarjam, Ehsan Emadzadeh, and Graciela Gonzalez. A hybrid system for emotion extraction from suicide notes. *Biomedical informatics insights*, 5:BII–S8981, 2012.

[27] Stevie Chancellor, Eric PS Baumer, and Munmun De Choudhury. Who is the" human" in human-centered machine learning: The case of predicting mental health from social media. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–32, 2019.

[28] Radim Rehurek and Petr Sojka. Software framework for topic modelling with large corpora. In *In Proceedings of the LREC 2010 workshop on new challenges for NLP frameworks*. Citeseer, 2010.

[29] Barbara Silveira Fraga, Ana Paula Couto da Silva, and Fabricio Murai. Online social networks in health care: a study of mental disorders on reddit. In *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, pages 568–573. IEEE, 2018.

[30] Jia Xue, Junxiang Chen, Chen Chen, Chengda Zheng, Sijia Li, and Tingshao Zhu. Public discourse and sentiment during the covid 19 pandemic: Using latent dirichlet allocation for topic modeling on twitter. *PloS one*, 15(9):e0239441, 2020.

[31] Nichol Castro and Cynthia SQ Siew. Contributions of modern network science to the cognitive sciences: revisiting research spirals of representation and process. *Proceedings of the Royal Society A*, 476(2238):20190825, 2020.

[32] Massimo Stella. Cognitive network science for understanding online social cognitions: A brief review. *Topics in Cognitive Science*, 2021.

[33] Cynthia SQ Siew, Dirk U Wulff, Nicole M Beckage, and Yoed N Kenett. Cognitive network science: A review of research on cognition through the lens of network representations, processes, and dynamics. *Complexity*, 2019, 2019.

[34] Olga Valba, Alexander Gorsky, Sergei Nechaev, and Mikhail Tamm. Analysis of english free association network reveals mechanisms of efficient solution of remote association tests. *PloS one*, 16(4):e0248986, 2021.

[35] Dong Wang, Jiexun Li, Kaiquan Xu, and Yizhen Wu. Sentiment community detection: exploring sentiments and relationships in social networks. *Electronic Commerce Research*, 17(1):103–132, 2017.

[36] Andreia Sofia Teixeira, Szymon Talaga, Trevor James Swanson, and Massimo Stella. Revealing semantic and emotional structure of suicide notes with cognitive network science. *Scientific reports*, 11(1):1–15, 2021.

[37] Charles J Fillmore and Collin Baker. A frames approach to semantic analysis. In *The Oxford handbook of linguistic analysis*. 2010.

[38] Kathleen Carley. Coding choices for textual analysis: A comparison of content analysis and map analysis. *Sociological methodology*, pages 75–126, 1993.

[39] Massimo Stella, Valerio Restocchi, and Simon De Deyne. # lockdown: Network-enhanced emotional profiling in the time of covid-19. *Big Data and Cognitive Computing*, 4(2):14, 2020.

[40] Camilo Akimushkin, Diego Raphael Amancio, and Osvaldo Novais Oliveira Jr. Text authorship identified using the dynamics of word co-occurrence networks. *PloS one*, 12(1):e0170527, 2017.

[41] Yoed N Kenett, Dror Y Kenett, Eshel Ben-Jacob, and Miriam Faust. Global and local features of semantic networks: Evidence from the hebrew mental lexicon. *PloS one*, 6(8):e23912, 2011.

[42] Yoed N Kenett, Effi Levi, David Anaki, and Miriam Faust. The semantic distance task: Quantifying semantic distance with semantic network path length. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(9):1470, 2017.

[43] Saif Mohammad. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 english words. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 174–184, 2018.

[44] Saif M Mohammad and Peter D Turney. Crowdsourcing a word–emotion association lexicon. *Computational intelligence*, 29(3):436–465, 2013.

[45] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.

[46] Alfonso Semeraro, Salvatore Vilella, and Giancarlo Ruffo. Pyplutchik: visualising and comparing emotion-annotated corpora. *arXiv preprint arXiv:2105.04295*, 2021.

[47] Laura B Allen, Kamila S White, David H Barlow, M Katherine Shear, Jack M Gorman, and Scott W Woods. Cognitive-behavior therapy (cbt) for panic disorder: Relationship of anxiety and depression comorbidity with treatment outcome. *Journal of Psychopathology and Behavioral Assessment*, 32(2):185–192, 2010.

[48] Jennifer R Dunn and Maurice E Schweitzer. Feeling and believing: the influence of emotion on trust. *Journal of personality and social psychology*, 88(5):736, 2005.

[49] Ciarán O'Driscoll, Joshua EJ Buckman, Eiko I Fried, Rob Saunders, Zachary D Cohen, Gareth Ambler, Robert J DeRubeis, Simon Gilbody, Steven D Hollon, Tony Kendrick, et al. The importance of transdiagnostic symptom level assessment to understanding prognosis for depressed adults: analysis of data from six randomised control trials. *BMC medicine*, 19(1):1–14, 2021.

[50] Robert A McCutcheon, Tiago Reis Marques, and Oliver D Howes. Schizophrenia—an overview. *JAMA psychiatry*, 77(2):201–210, 2020.

[51] Michael F Green, William P Horan, Junghee Lee, Amanda McCleery, L Felice Reddy, and Jonathan K Wynn. Social disconnection in schizophrenia and the general community. *Schizophrenia bulletin*, 44(2):242–249, 2018.

Supplementary Information

	Mean Local Clustering	Assortativity	No. of topics/ communities (Louvain)
Suicide Notes	0.198	-0.075	18
Anxiety	0.426	-0.209	21
Depression	0.432	-0.213	14
Schizophrenia	0.401	-0.209	22
DIY	0.12	-0.108	18

Table S3: Network statistics as defined in Siew et al. (2019) for the different co-occurrence networks analysed in the main text.

Suicide Notes	Anxiety	Depression	Schizophrenia	DIY
(love, you)	(I, know)	(I, know)	(I, know)	(I, need)
(I, love)	(feel, I)	(feel, I)	(I, think)	(I, want)
(I, know)	(I, think)	(I, want)	(feel, I)	(I, use)
(I, you)	(get, I)	(I, think)	(I, want)	(I, try)
(I, sorry)	(I, like)	(feel, like)	(get, I)	(I, look)
(hope, I)	(feel, like)	(I, like)	(go, I)	(go, I)
(I, want)	(go, I)	(get, I)	(I, like)	(find, I)
(go, I)	(I, want)	(go, I)	(feel, like)	(I, would)
(know, you)	(anxiety, I)	(I, time)	(I, say)	(get, I)
(I, think)	(I, time)	(I, really)	(I, time)	(I, know)

Table S4: Most frequent bi-grams after stopword removal and stemming in texts.