Physica A 610 (2023) 128336

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Cognitive network neighborhoods quantify feelings expressed in suicide notes and Reddit mental health communities

Simmi Marina Joseph [a], Salvatore Citraro [b,c], Virginia Morini [b,c], Giulio Rossetti [c], Massimo Stella [a,*]

[a] CogNosco Lab, Department of Computer Science, University of Exeter, Exeter, UK
[b] Department of Computer Science, University of Pisa, Pisa, Italy
[c] Consiglio Nazionale Delle Ricerche (ISTI-CNR), Pisa, Italy

ARTICLE INFO

Article history:
Received 20 July 2022
Received in revised form 17 September 2022
Available online 23 November 2022

Keywords:
Complex networks
Topic modeling
Emotions
Corpus
Concepts

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 modeling 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 polarization 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 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 network-based comparisons provide quantitative evidence that suicide notes encapsulate different ways of expressing feelings compared to online Reddit boards, the latter acting more like personal diaries and relief valves. Our findings provide an interpretable network-based aid for supporting psychological inquiries of human feelings in digital and clinical settings.

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.

* Corresponding author.
E-mail address: m.stella@exeter.ac.uk (M. Stella).

https://doi.org/10.1016/j.physa.2022.128336
0378-4371/

Complex circumstances can lead to unbearable situations and towards behaviors that are unique in nature, where people could reinforce negative emotions and isolate themselves [3]. The most extreme consequences of such behaviors 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 died by 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]. The online environment changes this bias thanks to users' anonymity, even when posts can be widely read: 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]. This represents 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 (cf. [8,11–13]) and we proceed along this direction.

In this paper, we aim to use the statistical mechanics of complex networks to understand and compare the language used by online users and authors of suicide notes. Our motivation is to use network science as a tool to uncover posts' and letters' underlying emotional patterns, 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 [5,12,14] 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 people who died by suicide with those of individuals discussing mental health conditions. Interpretability stems from the fact that unlike other neural network approaches to natural language processing [15], our networks represent explicitly syntactic, semantic and emotional associations between concepts, providing a directly accessible mapping of associative knowledge [16,17]. 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.

Importantly, our network-based method can model the structure of associative knowledge in dozens of thousands of texts, reconstructing the ways concepts were framed and emotionally perceived over an amount of texts that go way beyond the reach of a single human coder [18], e.g. a single psychiatrist. Our network-based method thus represents a knowledge mining AI that 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 person who died by suicide displayed emotional states and associative knowledge similar to the one reported in massively read online digital platforms. Rather than analyzing the network structure of online discussions, we focus our AI to identify which feelings were expressed by authors of suicide notes and of online posts in mental health communities.

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 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. Background

Text represents an insightful fragment of the psychology of individuals [19–21]. 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 [22–24]. 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 fatal suicides express more self-blame than those from attempted suicides [22]. 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 [25].

More recently, the study of mental health discussions and suicide contents in online pools focuses predominantly on Natural Language Processing (NLP) techniques, involving both supervised and unsupervised machine learning approaches. Among supervised methods, we find studies aiming to extract linguistic features for classification, e.g., predicting whether a textual content contains early signs of suicidal ideation [12,14,26] or more in general, signs of mental health disorders [13,27,28]. For instance, Nikfarjam et al. [29] 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 et al. [28] rely on regression and 90 text features derived from Reddit (e.g., sentiment

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

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 [12,13,27] 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 [30]. 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 [31], that belongs to the family of unsupervised approaches. Topic modeling techniques are able to detect the hidden topics – distributions of word importances across groups – 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 [11,32]. 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 [33]. 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, (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 [16]; Stella, (2021). Leveraging graph theory concepts, cognitive networks allow the analysis of and on complex structures in the human mind apt at storing and processing knowledge [34]. 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 [34–36]. 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 [20]. Another important aspect of cognitive networks is their natural tendency to form modular clusters or communities [16], 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 [37]. 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 [17]. In fact, the associations attributed to a given concept in text represent its so-called semantic frame [38]. Semantic frame theory indicates that meaning and perceptions relative to a concept can be reconstructed by analyzing its semantic frame [38,39] 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 [20,39,40].

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 [cf. 21]. 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 [26] and investigated also in [17]. These English letters were authored by people who died by suicide. On average, a suicide note included 120 words. No additional information (e.g. socio-economic context) was available in the dataset. 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 et al. [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 (https://www.parsehub.com/, Last Accessed: 08/10/2021) and anonymized 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. All notes and posts were fully anonymized and aggregated to protect the individuals' privacy, in adherence with the ethical best practice of the *Declaration of Helsinki* as revised in 1989.

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 [41] or semantic framing in suicide notes [17]. 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 [19,39]. 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, cf. 42], 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 visualization, built with Gephi, is reported in Fig. 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 how close a node is to all other nodes in a network in terms of its average network distance [cf. 16]. 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 [43]. 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 [20].

Taking inspiration from past approaches in cognitive network science [40], we performed an emotional enrichment of co-occurrence networks. We endowed stems with valence and emotional attributes as coming from the Valence–Arousal–Dominance [44] and from the Emotion Lexicon [45] 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 visualization purposes. Words were also attributed a list of emotions they elicited according to the cognitive data gathered by Mohammad and Turney [45].

3.3. Semantic frame theory and Latent Dirichlet analysis

Co-occurrence networks are representations of the ways authors associate concepts in their own narratives [39,41]. Considering a network neighborhood 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 [38].

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 characterization 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 [46] 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 notes and blog posts we focused over network neighborhoods/semantic frames.

3.4. Emotional profiling

The emotions attributed to individual words [45] 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 neighborhood, we measured the fraction $r_i = m/N$ of m words inspiring emotion i in a neighborhood with N entries. The collection of all fractions r_i constitutes the so-called emotional profile of a given semantic frame [cf. 40]. Notice that word negations were considered in these counts: As in [20], antonyms of words linked with negations (like

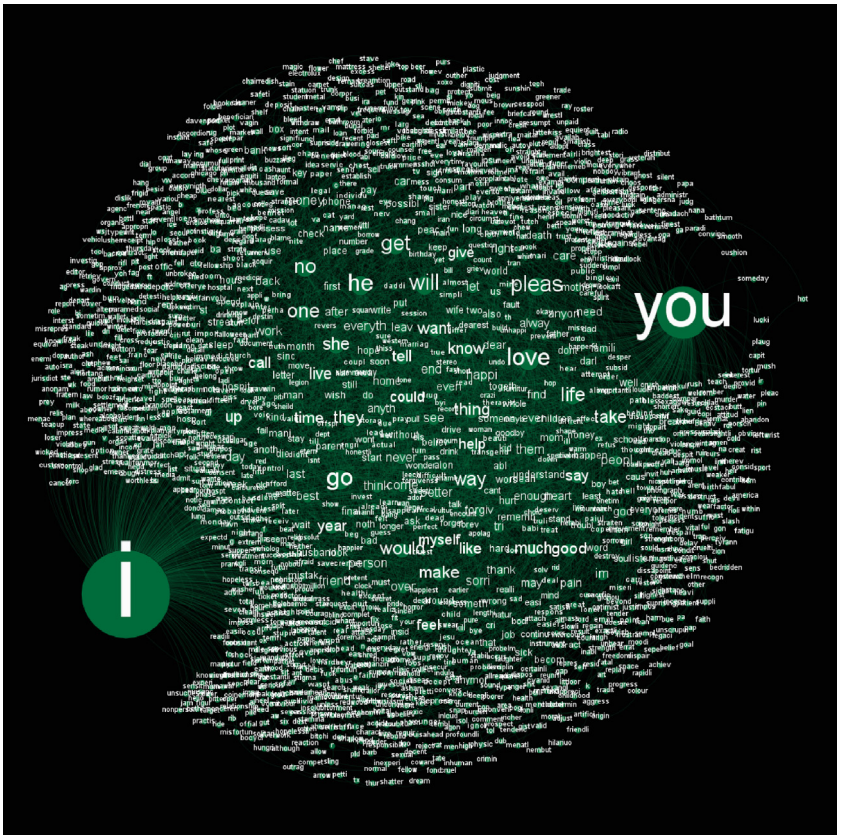

Fig. 1. Co-occurrence network of concepts as expressed in 139 genuine suicide notes.

"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:

$$z_i = \frac{r_i - R_i}{\sigma_i},$$

where R_i and σ_i are the average and the standard deviation of the distribution of number of words sampled at random from the lexicon and eliciting emotion *i*. Z-scores were plotted as petals, in a visualization reminiscent of Plutchik's wheel of emotions [cf. 47] and that we call emotional flower [20]. 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 Methods) 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 died by suicide expressed prominently their love to their dear ones. In the clinical subreddits, key concepts are relative to distress and feelings (e.g. "anxiety"

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. Frequency scores are normalized over the total for each subreddit/corpus to range between 0 and 1.

Suicide notes				Anxiety				Depression				Schizophrenia				DIY			
Frequency		Closeness		Frequency		Closeness		Frequency		Closeness		Frequency		Closeness		Frequency		Closeness	
I	*0.105*	*I*	*0.570*	*I*	*0.103*	*I*	*0.891*	*I*	*0.101*	*I*	*0.928*	*I*	*0.104*	*I*	*0.898*	*I*	*0.109*	*I*	*0.557*
you	*0.045*	*you*	*0.506*	*anxiety*	*0.021*	*anxiety*	*0.535*	*you*	*0.024*	*you*	*0.537*	*he*	*0.024*	*he*	*0.527*	*wall*	*0.018*	*wall*	*0.425*
love	*0.009*	*love*	*0.440*	*go*	*0.019*	*go*	*0.525*	*she*	*0.015*	*she*	*0.529*	*you*	*0.018*	*you*	*0.523*	*paint*	*0.017*	*want*	*0.424*
go	*0.009*	*go*	*0.435*	*feel*	*0.017*	*feel*	*0.518*	*go*	*0.015*	*go*	*0.527*	*they*	*0.017*	*they*	*0.523*	*need*	*0.017*	*paint*	*0.411*
he	*0.008*	*he*	*0.434*	*you*	*0.016*	*you*	*0.515*	*feel*	*0.015*	*feel*	*0.525*	*go*	*0.017*	*go*	*0.521*	*look*	*0.017*	*need*	*0.410*
will	*0.008*	*will*	*0.431*	*time*	*0.014*	*time*	*0.509*	*time*	*0.013*	*time*	*0.520*	*she*	*0.017*	*she*	*0.517*	*use*	*0.016*	*look*	*0.408*
life	*0.008*	*life*	*0.428*	*make*	*0.012*	*make*	*0.506*	*life*	*0.011*	*life*	*0.519*	*feel*	*0.015*	*feel*	*0.517*	*go*	*0.016*	*use*	*0.408*
please	*0.008*	*please*	*0.427*	*she*	*0.011*	*they*	*0.505*	*depress*	*0.011*	*depress*	*0.517*	*think*	*0.015*	*think*	*0.512*	*new*	*0.009*	*go*	*0.407*
take	*0.007*	*take*	*0.427*	*they*	*0.011*	*she*	*0.505*	*they*	*0.011*	*they*	*0.516*	*people*	*0.014*	*people*	*0.506*	*way*	*0.009*	*new*	*0.406*
way	*0.007*	*way*	*0.426*	*think*	*0.011*	*think*	*0.502*	*friend*	*0.010*	*friend*	*0.515*	*make*	*0.013*	*make*	*0.594*	*no*	*0.009*	*way*	*0.405*
know	*0.006*	*no*	*0.426*	*work*	*0.010*	*work*	*0.502*	*myself*	*0.010*	*myself*	*0.515*	*time*	*0.012*	*time*	*0.504*	*make*	*0.009*	*no*	*0.405*
want	*0.006*	*one*	*0.423*	*really*	*0.010*	*really*	*0.501*	*make*	*0.010*	*make*	*0.514*	*up*	*0.010*	*thing*	*0.502*	*side*	*0.005*	*make*	*0.405*
make	*0.006*	*thing*	*0.421*	*take*	*0.009*	*take*	*0.501*	*people*	*0.010*	*people*	*0.514*	*know*	*0.010*	*know*	*0.502*	*water*	*0.005*	*side*	*0.404*
help	*0.006*	*know*	*0.421*	*he*	*0.009*	*thing*	*0.501*	*even*	*0.010*	*even*	*0.513*	*take*	*0.010*	*take*	*0.501*	*want*	*0.005*	*aer*	*0.400*
tell	*0.005*	*want*	*0.419*	*people*	*0.008*	*he*	*0.500*	*want*	*0.009*	*want*	*0.513*	*really*	*0.010*	*one*	*0.501*	*work*	*0.005*	*water*	*0.496*

being more frequent than "you" in r/Anxiety, "feel" being prominent/frequent across all clinical boards). This indicates a difference between suicide notes, addressing expressions of love, and subreddits, reporting personal feelings and experiences. Bi-gram analysis confirms this difference between love-focused suicide notes and self-focused online posts (see Supplementary Table S2). 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 analyzed 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

Figs. 2 and 3 report the semantic and emotional content associated with "feel" in suicide notes 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, Fig. 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 [48]. 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 [49]. 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 polarization 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 (Fig. 2, middle), where the emotional polarization 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 [cf. 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 [50].

In the r/DIY board (Fig. 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 polarized, featuring equally strong levels of sadness and trust simultaneously (see Fig. 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 [cf. 51]. 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" and

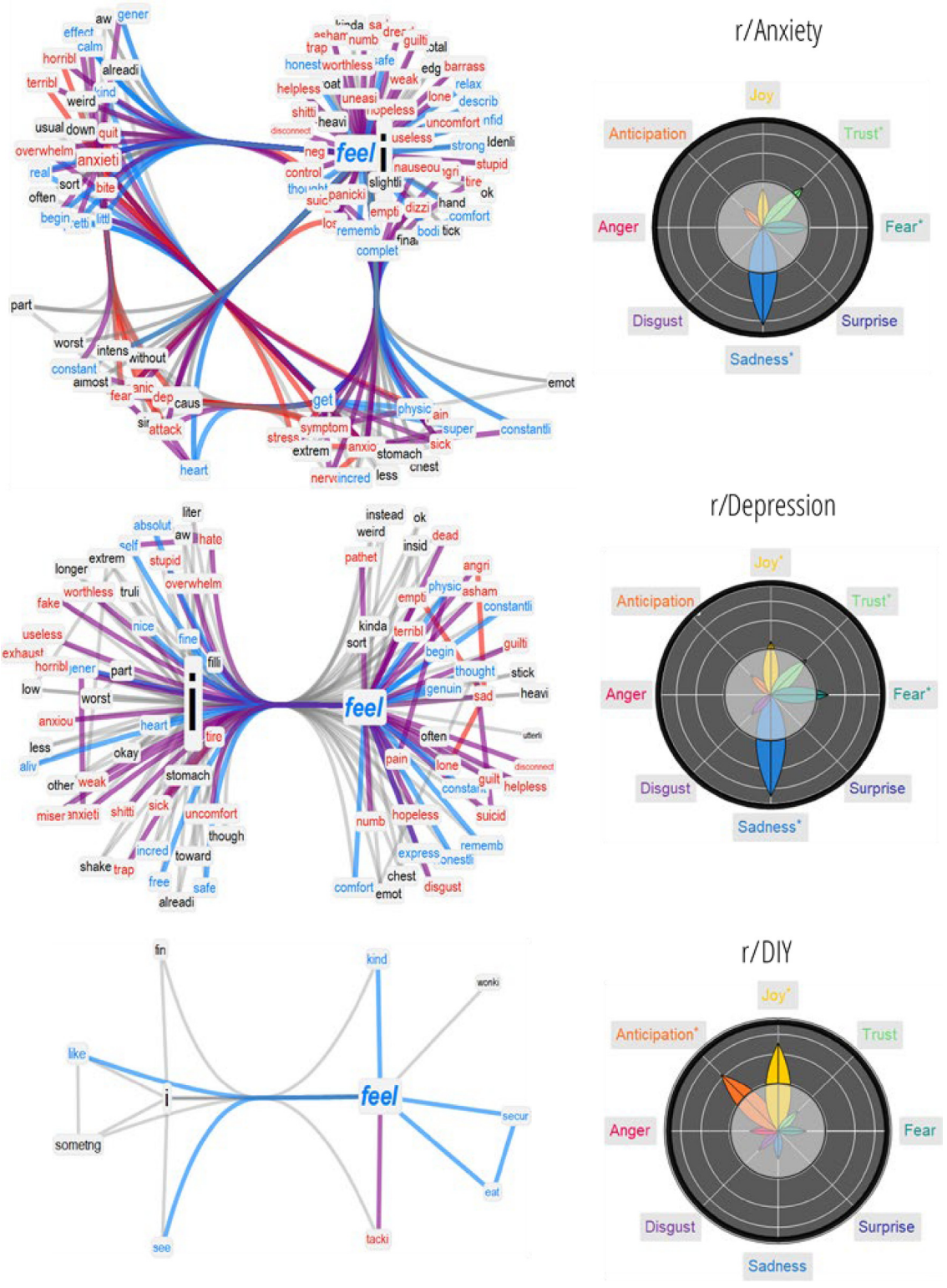

Fig. 2. Cognitive networks (left) and emotional flowers (right) for the semantic frames of "feel" across r/Anxiety (top), r/Depression (middle) and r/DIY (bottom). Links in cognitive networks indicate concept co-occurrence. Colors indicate valence, i.e. positive (cyan), negative (red) or neutral (black) 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.

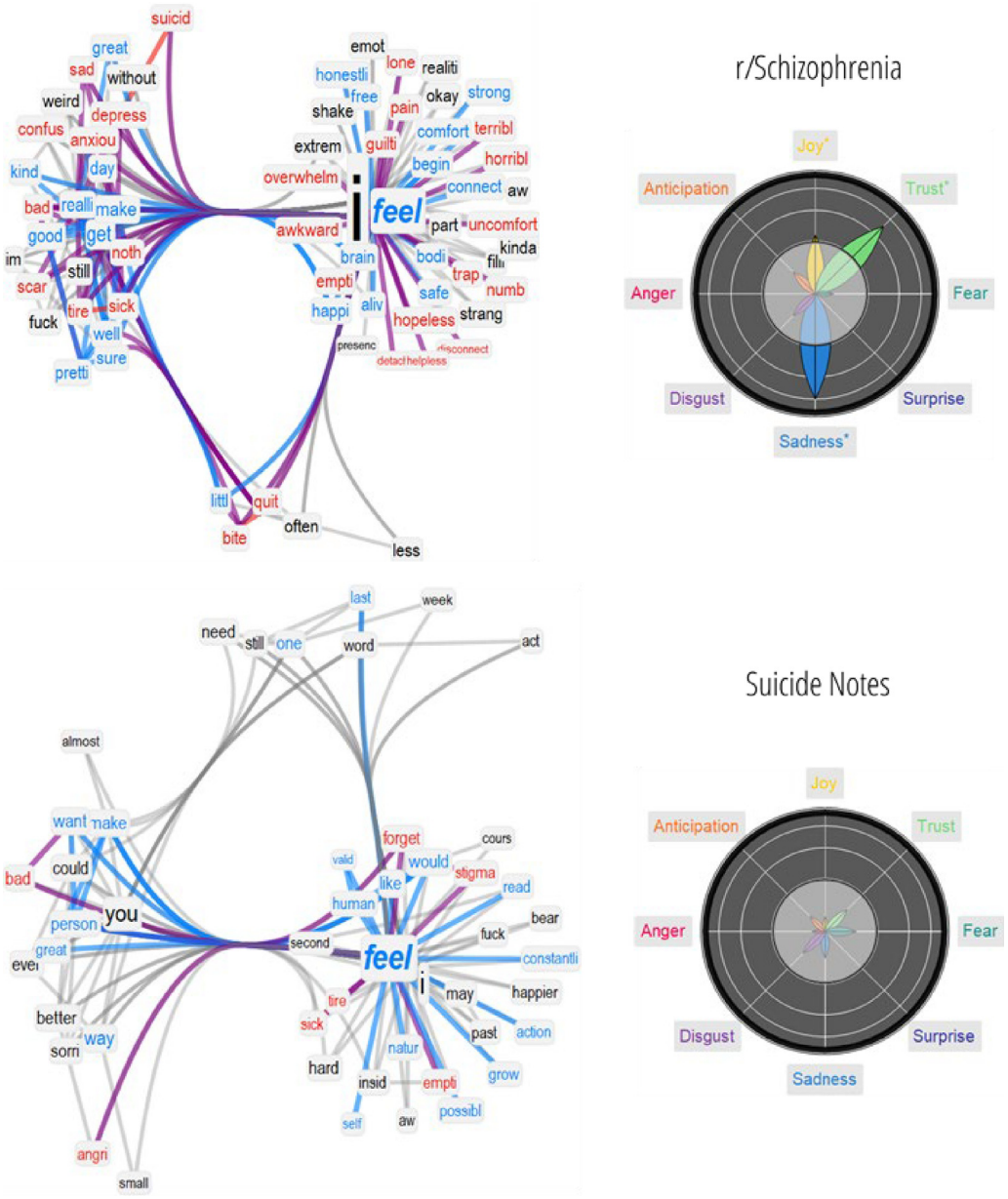

Fig. 3. Cognitive networks (left) and emotional flowers (right) for the semantic frames of "feel" across r/Schizophrenia (top) and suicide notes (bottom). Links in cognitive networks indicate concept co-occurrence. Colors indicate valence, i.e. positive (cyan), negative (red) or neutral (black) valence. 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.

"disconnected", which relate to the well-documented sense of isolation that hallucinations and paranoia can create in patients affected by schizophrenia [52].

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

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. The parentheses report individual word scores reflecting how important words are for that topic.

Rank	Suicide notes	r/Anxiety	r/Depression	r/Schizophrenia
1	*love (0.11)*	*go (0.07)*	*know (0.08)*	*want (0.10)*
2	*go (0.08)*	*know (0.06)*	*want (0.07)*	*know (0.08)*
3	*know (0.06)*	*time (0.06)*	*go (0.05)*	*time (0.06)*
4	*want (0.05)*	*think (0.06)*	*depressed (0.05)*	*year (0.06)*
5	*please (0.05)*	*want (0.06)*	*think (0.05)*	*day (0.04)*
6	*life (0.05)*	*really (0.06)*	*time (0.04)*	*really (0.03)*
7	*take (0.03)*	*thing (0.05)*	*really (0.03)*	*take (0.02)*
8	*live (0.03)*	*year (0.03)*	*life (0.03)*	*try (0.02)*
9	*time (0.03)*	*say (0.03)*	*day (0.02)*	*work (0.02)*
10	*leave (0.03)*	*anxiety (0.02)*	*year (0.02)*	*life (0.02)*
11	*way (0.02)*	*life (0.02)*	*try (0.01)*	*even (0.02)*
12	*money (0.02)*	*try (0.01)*	*work (0.01)*	*tell (0.01)*
13	*god (0.02)*	*need (0.01)*	*help (0.01)*	*help (0.01)*
14	*last (0.01)*	*friend (0.01)*	*people (0.01)*	*friend (0.01)*
15	*sorry (0.01)*	*tell (0.01)*	*never (0.01)*	*need (0.01)*

suicide notes. 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 polarization and relationships of feelings with clinical conditions, we examined our textual corpora through a Latent Dirichlet Analysis (see Methods). 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. We performed a simple stability analysis of the results by testing how topics change structurally when K (the number of topics) is perturbed by $+1$ or -1 unit around the number $|C|$ of communities found by network analysis (see Methods). When $K = |C| + 1$ or $K = |C| - 1$, the first 8 positions of all rankings in Table 2 remained stable while words in lower rankings underwent swaps by at most two positions. This simple check indicates that the importance of words identified when $K = |C|$ is mostly stable under local perturbations of model parameters (i.e. the number of topics).

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 notes 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 [19]. 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 [47]: its semantic frame is populated by pairs of unexpected, often contrasting emotions, as in the polarized "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" [47], an attitude that may fit 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 [50]. 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 notes 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" [17], people who are intentioned to die by 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 ([27,28]; Stella, 2021). Indeed, preliminary studies already exploited Reddit boards for monitoring shifts from mental health disorders to suicidal ideation [5,8,14]. 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 et al. [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"). All such mentioned works could rely on the ultimate purpose of predicting early stages of suicide ideation. The features extracted from a complex network could improve the characterization of this mental state (see also [53]). Accordingly, these features could be used to improve classification tasks in the same way as word embeddings improve sentiment analysis [21]. However, the risk of individual dehumanization [30] is high, and makes a prediction task difficult not only in terms of methodology but in terms of ethics as well.

Nevertheless, there are other 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.

CRediT authorship contribution statement

Simmi Marina Joseph: Data Curation, Methodology, Software, Formal analysis, Visualisation. **Salvatore Citraro:** Conceptualisation, Methodology, Software, Writing – original draft, Writing – review & editing. **Virginia Morini:** Investigation, Validation, Writing – original draft, Writing – review & editing. **Giulio Rossetti:** Conceptualisation, Writing – original draft, Writing – review & editing. **Massimo Stella:** Supervision, Project administration, Conceptualisation, Methodology, Software, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work is supported by the European Union's Horizon 2020 research and innovation programme under grant agreements No. 871042 'SoBigData++: European Integrated Infrastructure for Social Mining and Big Data Analytics'

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.physa.2022.128336.

References

[1] R.I. Dunbar, Coevolution of neocortical size, group size and language in humans, Behav. Brain Sci. 16 (4) (1993) 681–694.
[2] E. Durkheim, Le Suicide : Étude de sociologie, 1897.
[3] M.M. Tugade, B.L. Fredrickson, L. Feldman Barrett, Psychological resilience and positive emotional granularity: Examining the benefits of positive emotions on coping and health, J. Pers. 72 (6) (2004) 1161–1190.
[4] B. Batinic, G. Opacic, T. Ignjatov, D.S. Baldwin, Comorbidity and suicidality in patients diagnosed with panic disorder/agoraphobia and major depression, Psychiatria Danubina 29 (2) (2017) 186–194.
[5] M. Gaur, V. Aribandi, A. Alambo, U. Kursuncu, K. Thirunarayan, J. Beich, A. Sheth, Characterization of time-variant and time-invariant assessment of suicidality on reddit using C-SSRS, PLoS One 16 (5) (2021) e0250448.
[6] U. Pavalanathan, M. De Choudhury, Identity management and mental health discourse in social media, in: Proceedings of the 24th international conference on world wide web, 2015, pp. 315–321.
[7] M. De Choudhury, S. De, Mental health discourse on reddit: Self-disclosure, social support, and anonymity, in: Eighth International AAAI Conference on Weblogs and Social Media, 2014.
[8] M. De Choudhury, E. Kiciman, M. Dredze, G. Coppersmith, M. 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, 2016, pp. 2098–2110.
[9] M. McPherson, L. Smith-Lovin, J.M. Cook, Birds of a feather: Homophily in social networks, Annu. Rev. Sociol. 27 (1) (2001) 415–444.
[10] G. Gkotsis, A. Oellrich, T. Hubbard, R. Dobson, M. Liakata, S. Velupillai, R. Dutta, The language of mental health problems in social media, in: Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology, 2016, pp. 63–73.
[11] M. Yoo, S. Lee, T. Ha, Semantic network analysis for understanding user experiences of bipolar and depressive disorders on Reddit, Inf. Process. Manage. 56 (4) (2019) 1565–1575.
[12] S. Ji, C.P. Yu, S.F. Fung, S. Pan, G. Long, Supervised learning for suicidal ideation detection in online user content, Complexity 2018 (2018).
[13] J. Kim, J. Lee, E. Park, J. Han, A deep learning model for detecting mental illness from user content on social media, Sci. Rep. 10 (1) (2020) 1–6.
[14] M.M. Tadesse, H. Lin, B. Xu, L. Yang, Detection of suicide ideation in social media forums using deep learning, Algorithms 13 (1) (2020) 7.
[15] C. Rudin, Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, Nat. Mach. Intell. 1 (5) (2019) 206–215.
[16] C.S. Siew, D.U. Wulff, N.M. Beckage, Y.N. Kenett, Cognitive network science: A review of research on cognition through the lens of network representations, Proc. Dyn. Complexity (2019) (2019).
[17] A.S. Teixeira, S. Talaga, T.J. Swanson, M. Stella, Revealing semantic and emotional structure of suicide notes with cognitive network science, Sci. Rep. 11 (1) (2021) 1–15.
[18] M. Stella, Cognitive network science for understanding online social cognitions: A brief review, Topics Cogn. Sci. 14 (1) (2022) 143–162.
[19] M. Murray, Narrative Psychology and Narrative Analysis, American Psychological Association, 2003.
[20] M. Stella, Text-mining forma mentis networks reconstruct public perception of the STEM gender gap in social media, PeerJ. Comput. Sci. 6 (2020) e295.
[21] J. Jackson, J. Watts, J.M. List, R. Drabble, K. Lindquist, From text to thought: How analyzing language can advance psychological science, Perspect. Psychol. Sci. (2021).
[22] A. Brevard, D. Lester, B. Yang, A comparison of suicide notes written by suicide completers and suicide attempters, Crisis: J. Crisis Interv. Suicide Prevent. (1990).
[23] T. Foster, Suicide note themes and suicide prevention, Int. J. Psychiatry Med. 33 (4) (2003) 323–331.
[24] L.D. Handelman, D. Lester, The content of suicide notes from attempters and completers, Crisis 28 (2) (2007) 102–104.
[25] E.S. Shneidman, N.L. Farberow, Some comparisons between genuine and simulated suicide notes in terms of Mowrer's concepts of discomfort and relief, J. General Psychol. 56 (2) (1957) 251–256.
[26] A.M. Schoene, N. 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, 2016, pp. 128–133.
[27] J.H. Shen, F. Rudzicz, Detecting anxiety through reddit, in: Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology—from Linguistic Signal to Clinical Reality, 2017, pp. 58–65.
[28] D.M. Low, L. Rumker, T. Talkar, J. Torous, G. Cecchi, S.S. Ghosh, Natural language processing reveals vulnerable mental health support groups and heightened health anxiety on reddit during covid-19: Observational study, J. Med. Internet Res. 22 (10) (2020) e22635.
[29] A. Nikfarjam, E. Emadzadeh, G. Gonzalez, A hybrid system for emotion extraction from suicide notes, Biomed. Inform. Insights 5 (2012) BII–S8981.
[30] S. Chancellor, E.P. Baumer, M. De Choudhury, Who is the human in human-centered machine learning: The case of predicting mental health from social media, Proc. ACM Hum.-Comput. Interact. 3 (CSCW) (2019) 1–32.
[31] R. Rehurek, P. Sojka, Software framework for topic modelling with large corpora, in: Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks, 2010.
[32] B.S. Fraga, A.P.C. da Silva, F. 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, IEEE, 2018, pp. 568–573.
[33] J. Xue, J. Chen, C. Chen, C. Zheng, S. Li, T. Zhu, Public discourse and sentiment during the COVID 19 pandemic: Using latent Dirichlet allocation for topic modeling on Twitter, PLoS One 15 (9) (2020) e0239441.

[34] N. Castro, C.S. Siew, Contributions of modern network science to the cognitive sciences: revisiting research spirals of representation and process, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 476 (2238) (2020) 20190825.
[35] M. Stella, Y.N. Kenett, Viability in multiplex lexical networks and machine learning characterizes human creativity, Big Data Cogn. Comput. 3 (3) (2019) 45.
[36] O. Valba, A. Gorsky, S. Nechaev, M. Tamm, Analysis of English free association network reveals mechanisms of efficient solution of remote association tests, PLoS One 16 (4) (2021) e0248986.
[37] D. Wang, J. Li, K. Xu, Y. Wu, Sentiment community detection: exploring sentiments and relationships in social networks, Electron. Commerc. Res. 17 (1) (2017) 103–132.
[38] C.J. Fillmore, C. Baker, A frames approach to semantic analysis, in: The Oxford Handbook of Linguistic Analysis, 2010.
[39] K. Carley, Coding choices for textual analysis: A comparison of content analysis and map analysis, Sociol. Methodol. 7 (1993) 5–126.
[40] M. Stella, V. Restocchi, S. De Deyne, # Lockdown: Network-enhanced emotional profiling in the time of Covid-19, Big Data Cogn. Comput. 4 (2) (2020) 14.
[41] C. Akimushkin, D.R. Amancio, O.N. Oliveira Jr., Text authorship identified using the dynamics of word co-occurrence networks, PLoS One 12 (1) (2017) e0170527.
[42] Y.N. Kenett, D.Y. Kenett, E. Ben-Jacob, M. Faust, Global and local features of semantic networks: Evidence from the hebrew mental lexicon, PLoS One 6 (8) (2011) e23912.
[43] Y.N. Kenett, E. Levi, D. Anaki, M. Faust, The semantic distance task: Quantifying semantic distance with semantic network path length, J. Exp. Psychol: Learn. Mem. Cogn. 43 (9) (2017) 1470.
[44] S. 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), 2018, pp. 174–184.
[45] S.M. Mohammad, P.D. Turney, Crowdsourcing a word–emotion association lexicon, Comput. Intell. 29 (3) (2013) 436–465.
[46] V.D. Blondel, J.L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks, J. Stat. Mech. Theory Exp. 2008 (10) (2008) P10008.
[47] A. Semeraro, S. Vilella, G. Ruffo, Pyplutchik: Visualising and comparing emotion-annotated corpora, Plos One 16 (9) (2021) e0256503.
[48] L.B. Allen, K.S. White, D.H. Barlow, M.K. Shear, J.M. Gorman, S.W. Woods, Cognitive-behavior therapy (CBT) for panic disorder: Relationship of anxiety and depression comorbidity with treatment outcome, J. Psychopathol. Behav. Assess. 32 (2) (2010) 185–192.
[49] J.R. Dunn, M.E. Schweitzer, Feeling and believing: the influence of emotion on trust, J. Personal. Soc. Psychol. 88 (5) (2005) 736.
[50] C. O'Driscoll, J.E. Buckman, E.I. Fried, R. Saunders, Z.D. Cohen, G. Ambler, S. Pilling, The importance of transdiagnostic symptom level assessment to understanding prognosis for depressed adults: analysis of data from six randomised control trials, BMC Med. 19 (1) (2021) 1–14.
[51] R.A. McCutcheon, T.R. Marques, O.D. Howes, Schizophrenia—an overview, JAMA Psychiatry 77 (2) (2020) 201–210.
[52] M.F. Green, W.P. Horan, J. Lee, A. McCleery, L.F. Reddy, J.K. Wynn, Social disconnection in schizophrenia and the general community, Schizophrenia Bull. 44 (2) (2018) 242–249.
[53] A. Fatima, Y. Li, T.T. Hills, M. Stella, Dasentimental: Detecting depression, anxiety, and stress in texts via emotional recall, cognitive networks, and machine learning, Big Data Cogn. Comput. 5 (4) (2021) 77.