Cognitive Networks and Text Analysis Identify Anxiety as a Key Dimension of Distress in Genuine Suicide Notes

Supporting Information

Massimo Stella, Trevor James Swanson, Andreia Sofia Teixeira, Brianne N. Richson, Ying Li, Thomas T. Hills, Kelsie T. Forbush, and David Watson

S1. Data processing, partitioning and cleaning

This project considered a collection of 200 responses to the Emotional Recall Task. The task measured recall from semantic memory when solicited by an emotional question, i.e., "Which emotions did you feel in the last month?". The authors of the original study showed that these recalls correlate with other psychometric measures of emotional regulation and mental well-being; hence, the data represent a valuable source of insights to explore cognitive and emotional patterns.
Responses were in the form of an ordered list of 10 emotional words. For instance,

R_1 = *("happy", "sad", "tired", "funny", "moody", "content", "angry", "concerned", "scared", "upset")*

Each response represents the recollection of an individual, whose mental well-being has also been measured through different psychometric scales. In this analysis, we used the scale of DASS-21, a psychometric scale measuring levels of emotional distress across three subscales for depression, anxiety and stress, respectively. According to the data gathered by the authors of the ERT study, DASS-21 scores ranged for a minimum of 0 (no distress) to a maximum of 22 (maximum level of distress). These numerical scores were attributed to individuals and thus to their recall lists. In this way, we obtained a set of 200 ordered word lists with annotated numerical values expressing the intensity of anxiety/depression/stress of the individual who produced such words.

We also considered 139 suicide notes, released by a different team of authors and investigated in a previous study. These notes were collected over decades, written in plain English and fully anonymized. We considered each note as a collection of words organized across sentences.

As reported in the flowchart in Figure 1 (main text), we used these two data sources to identify whether the structure of emotional associations expressed in suicide notes resemble those produced by individuals with high or low levels of emotional distress, across the dimensions of anxiety, depression and stress.

Operationally, we applied stemming to ERT recall lists, removing suffixes and word declinations while preserving the root of a word and thus its core meaning, e.g., "sadness" became "sad". As reported in the main text, we used Mathematica 11.3 to perform stemming. This led to a set of 475 unique emotional stems/responses being mentioned in the ERT recalls. We called this collection of unique, stemmed emotional responses the emotional dictionary D. We then used D to investigate the emotional content of suicide notes. For each suicide note, we removed punctuation and tokenized all text, obtaining a sequence of words to which stemming was applied. Finally, we selected only stemmed words, written in suicide notes, that were mentioned in the emotional dictionary D. We preserved the order in which these words appeared in the text.

These also steps transformed suicide notes into lists of ordered emotional words without explicit grammatical connectors (e.g., prepositions were not in *D* and were thus filtered out). This process led to the identification of 139 emotional word sequences from suicide notes. Because of repetitions across letters, suicide notes mentioned only 179 of all the 475 words present in the emotional dictionary of ERT data.

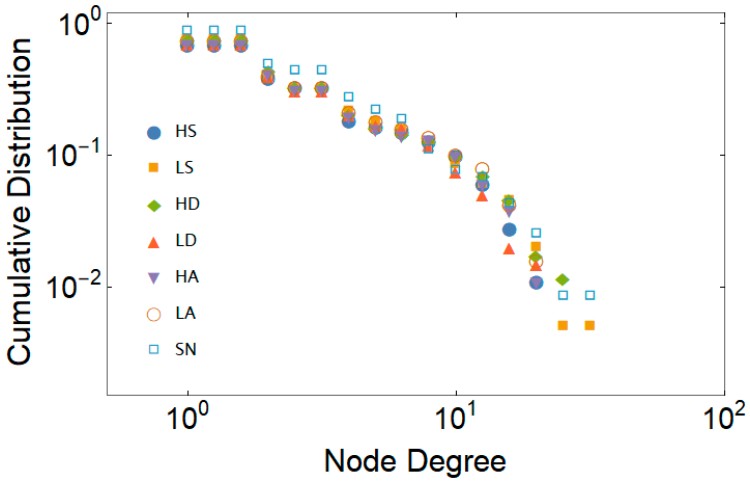
As reported in Figure 1 (main text), after obtaining the regularized word sequences, we proceeded with network construction. Considering the distributions of observed depression (anxiety, stress) levels in the overall ERT lists, we partitioned them into two mutually exclusive subsets, indicating high and low levels of emotional distress, respectively. To obtain a balanced partitioning, where high and low subsamples possess an equal number of recall lists, we used the median score for each subscale as a threshold for distinguishing between high and low levels of distress. Repeating this process across depression (D), anxiety (A) and stress (S), we obtained six partitions and six subsequences of emotional stemmed words from which co-occurrence networks could be built. Also applying the same network construction method to suicide notes led to the creation of seven emotional networks in total, as reported in the main text.

To build networks out of stem sequences, we follow the protocol described by Goñi et al. (2011). The protocol considers an initial link creation stage and a subsequent link filtering phase. In the link creation stage, all word sequences are processed sequentially, one after another. For each sequence, the algorithm starts from the first word and then builds links between any two words up to *L* positions forward in the sequence, including the starting word. Then, the algorithm passes onto the second word on the list and builds links between any two words up to *L* positions either before or after the second word. In this way, *L* identifies a window of possible co-occurrences between contiguous words, i.e., words separated by, at most, *L-1* words. For instance, in the above formula for R1, *"sad","tired","funny","moody", and "content"* would all be included in the window centered in "funny" and with $L = 2$. In this window, centered around *"funny"*, its co-occurrences would then be "*sad*" - "*funny*", "*tired*" - "*funny*", "*moody*" - "*funny*" and "*content*" - "*funny*". Consistent with the analysis of animal fluency data by Goñi et al. (2011), we select $L = 2$ as our window parameter. Notice that co-occurrences are iteratively built and appended to a repository considering all observed co-occurrences across lists. This generative process continues until the end of the sequence is followed. At this point, counting for each and every co-occurrence takes place, the generative phase ends and the filtering phase starts. Like we mentioned in the main text, idiosyncratic co-occurrences are discarded, i.e., co-occurrences appearing only $T = 1$ times. A subsequent discarding of co-occurrences takes place through binomial filtering. The aim of this filtering is to discard co-occurrences that can be explained by mere frequency effects, e.g., a co-occurrence is observed just because its endpoints are highly frequent in the data.

A table with a few network measures identifies the overall structure of these networks (see SI Table 1). Interestingly, SN is considerably more clustered than other target networks and this might indicate a stronger tendency for suicide letters to create triangles including the same sets of concepts. Triadic closure in SN networks was already investigated in a previous study (Teixeira et al. 2021).

Adopting the information from these networks, we then explore our main research question (RQ 1):

Using network science, which among HS/HA/HD or LS/LA/LD is closer to the emotional associations observed in suicide notes and encoded in SN? (RQ1)



SI Figure S1: *Cumulative distributions $P(X \geq k)$ for node degree k in each of the 7 emotional networks described in the main text.*

As evident from the cumulative degree distribution in SI Figure 1, all emotional networks share very similar topological structures, i.e., heavy-tailed degree distribution with short average network distance and relatively high clustering coefficient; these are all indications of a small-world structure. Using these features will not help us with the exploration of our RQ1.

In the main text, we mainly implemented this approach by following two methods: (1) by using the emotional auras (see main text) and (2) by using the "words not said" analysis (see SI Figure 1, SI and main text). While the first technique has a clear interpretation, i.e., using Jaccard similarity to identify how consistently nodes with given valence/arousal labels tend to link with other nodes of similar valence/arousal (see main text), the second technique requires more steps. Let us focus on the technical implementation and performance of this second technique.

S2. Further notes on the mathematical characterization of the "words not said" analysis

The "words not said" analysis originates from the finding that emotional associations in the SN network are mostly positive (see the Results Section in the main text). When Jaccard similarity is employed on the whole SN network, the positive association patterns between words in the SN network make it most similar to the LA (low anxiety) network. The key factors behind this similarity are already discussed in the main text. This strong dominance of positive patterns might still hide additional emotional trends encoded in the SN network, which could not be detected via a Jaccard similarity approach. In other words, our "words not said" analysis stems from the need to investigate network structures at a deeper level, addressing our secondary research question RQ2:

Are there other negative emotional trends that are being obfuscated by a dominance of positive words and associations in suicide letters? *(RQ2)*

To address RQ2, we devise the "words not said" as a subtractive analysis based on network topology. Let us consider the largest connected component (LCC) of the SN network, i.e., the largest set of nodes for which there is a sequence of links connecting any two of them. As reported in SI Table 1, the LCC(SN), i.e., the largest connected component of the SN network, coincides with the whole network

itself. This means that, for any two words in SN, there is always a sequence of associations that connects them. This is not the case for the other emotional networks, where the LCC is smaller than the original network.

For our "words not said" analysis, we compute pseudo z-scores over the residual sets *R(s)* and cumulate all those $z_e > 1.96$. The higher the sum of all strong emotional z-scores, the stronger the emotions in the residual *R(s)* and thus the further/more dissimilar *SN* is from the target emotional network.

As reported in the main text, with z-scores being visualized as emotional flowers and the rejection region as petals on a radar plot, it is the network of *HA* that displays the weaker emotional residuals and thus the closest similarity to the SN network.

S3. Measuring the correctness and performance of the "words not said" analysis

As a binary classification task (high/low), the "words not said" analysis can either correctly identify that AN is high/low (true positive, TP), correctly identify that AN is not high/low (true negative, TN), or misclassify either AN as high when it is low (false positive, FP) or as low when it is high (false negative, FN). For the sake of simplicity, let us denote the numerosity of these cases by omitting the $|.|$ notation which we have used so far. Then, we can state that accuracy measures the proportion of correct classifications, be them either true positives (AN is high or low, as expected) or true negatives (AN is not low or high, as expected):

$$a = \frac{TP+TN}{TP+TN+FP+FN}. \tag{5}$$

The F1 score is, instead, the harmonic mean between two quantities—precision p and recall c—which are defined as follows:

$$p = \frac{TP}{TP+FP}\,, \quad c = \frac{TP}{TP+FN}\,, \tag{6}$$

Precision represents the fraction of positives that have been correctly identified when false positives are accounted for, while recall indicates the fraction of positive elements correctly identified when false negatives are accounted for. In the following formula, the F1 score becomes

$$F1 = 2(p^{-1} + c^{-1})^{-1} = \frac{2TP}{2TP+FP+FN}\,. \tag{7}$$

In the main text, Table 2 reports the results of the classification tasks across different dimensions of emotional distress once aggregated for each scenario.

REFERENCES

1. Goñi, J., Arrondo, G., Sepulcre, J., Martincorena, I., Vélez de Mendizábal, N., Corominas-Murtra, B., ... & Villoslada, P. (2011). The semantic organization of the animal category: evidence from semantic verbal fluency and network theory. *Cognitive processing*, *12*(2), 183-196.
2. Teixeira, A. S., Talaga, S., Swanson, T. J., & Stella, M. (2021). Revealing semantic and emotional structure of suicide notes with cognitive network science. *Scientific reports*, *11*(1), 1-15.

3. Stella, M. (2020). Text-mining forma mentis networks reconstruct public perception of the STEM gender gap in social media. *PeerJ Computer Science*, *6*, e295.
4. Newman, M. (2018). *Networks*. Oxford University Press, Oxford UK.
5. Mohammad, S., & Turney, P. (2010, June). Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text* (pp. 26-34).
6. Golino, H. F., & Epskamp, S. (2017). Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PloS One*, *12*(6), e0174035.