# Cognitive network science as a framework for detecting structural patterns and emotions in suicide letters

M. Stella<sup>1,\*,+</sup>, T.J. Swanson<sup>2,+</sup>, Y. Li<sup>3</sup>, T.T. Hills<sup>4</sup> and A.S. Teixeira<sup>5</sup>

<sup>1</sup>CogNosco Lab, Department of Computer Science, University of Exeter, UK
<sup>2</sup>Department of Psychology, University of Kansas, USA
<sup>3</sup>Max Planck Institute for Human Development, Germany
<sup>4</sup>Department of Psychology, University of Warwick, UK
<sup>5</sup>Hospital da Luz Learning Health and INESC-ID, Portugal
<sup>+</sup>These authors contributed equally. \*Corresponding Author: m.stella@exeter.ac.uk.

May 27, 2021

#### Abstract

Communicating one's mindset means transmitting complex relationships between concepts and emotions. Using cognitive network science, we reconstruct the mindset around suicide as communicated in 139 genuine suicide notes. Despite their negative context, suicide notes are surprisingly positively valenced and their ending statements are markedly more emotional, i.e. elicit deeper fear/sadness but also stronger joy/trust and anticipation, than their main body. By using emotional states from the Emotional Recall Task, we "open the lid" of suicidal narratives and compare their emotional backbone against emotion recall in mentally healthy individuals. Supported by psychological literature, we introduce emotional complexity as an affective analogue of structural balance theory, measuring how elementary cycles (closed triads) of emotion co-occurrences mix positive, negative and neutral states in narratives and recollections. Both authors of suicide notes and healthy individuals exhibit less complexity and more emotional coherence than expected by chance. However, suicide narratives display higher complexity, i.e. a lower level of coherently valenced triads, than healthy individuals recalling the same states. Entropy measures identified a similar tendency for suicide letters to shift more frequently between contrasting emotional states. Our results demonstrate that suicide notes possess highly contrastive narratives of emotions, more complex than expected by null models and healthy populations.

**Keywords:** data science, psycholinguistics, complex networks, text analysis, emotional profiling, cognitive network science.

#### Introduction

Committing suicide is the outcome of complex, distressing emotional processing. It represents an extreme behavior that often rises from negative perceptions and distorted mindsets about the self, others and life [1, 2]. Whereas the advent of Big Data and natural language processing opened new ways for quickly detecting one's intention to commit suicide through their language [3], more work is required to understand the emotional content and overarching organization of suicidal ideation in the minds of those who commit suicide [4, 5]. This challenging task requires the adoption of transparent techniques, mediating access to mental patterns and ways of thinking as expressed within the final letters and notes written by people who committed suicide.

This article adopts the recent framework of cognitive network science to investigate the emotional structure of suicide notes in view of relevant studies about emotions and human memory [6]. Building on previous approaches using language processing for understanding ways of thinking in texts [7], we aim to detect how emotional words are organized and recalled across sentences in suicide letters. Specifically, our investigation adopts cognitive networks in order to: (i) quantify whether concepts eliciting emotional states tend to concentrate in specific portions of suicide letters, e.g. in the closing remarks, or in association with specific ideas like "family", (ii) introduce the metric of emotional complexity, capturing the tendency for the narrative of suicide notes to transition between positive, negative and neutral emotional states.

Our methodology strongly revolves around cognitive network science, a novel area of research at the fringe of computer science, psychology, linguistics and network science [8, 9, 10, 11, 12, 13, 14]. Cognitive networks are models of cognition where nodes represent concepts/ideas and links indicate conceptual relationships of relevance for language processing in psycholinguistic tasks, e.g. synonyms. In other words, networks are used as models for cognition, capturing the highly structured nature of conceptual associations as stored within the human mind and used with language processing. Research has overwhelmingly shown that conceptual associations and their network structure strongly influence a variety of language-related processes [11, 9], including writing styles [15] and efficient communication processing [12]. Representing text as a cognitive network, where nodes represent words that are connected according to specific semantic/syntactic relationships, has been shown to extract salient features of written corpora in a variety of different settings, including identifying opposing factions in political debates [16], detecting negative stances towards science [10] and social media analysis during extreme events [17, 18].

Our approach merges cognitive networks with recent data [6] about emotional states experienced by healthy and clinical populations of individuals. This synergy between conceptual associations and emotional data leads to automatic assessments of texts in terms of their emotional content and structure without the need for human intervention, analogous to previous approaches [7, 18] and in line with other methods across information-, cognitive-and data science [19, 20, 21, 22, 23].

The task of detecting emotions in text is also known as *emotional profiling* and recently met an unprecedented flourishing thanks to the increasing availability of large-scale datasets about cognitive features of words in the human mind [23, 11, 6, 24, 25].

Emotions are abstract, conceptual categories whose instances are highly variable in perception and yet yoked together by situated, functional similarities [26, 27]. Given this, it is not surprising to find that language goes beyond describing emotions after the fact. It also supports making sensations meaningful as emotions [28]. Despite being highly complex, all emotions vary in terms of valence (often referred to as sentiment in computer science literature) — the most primary and culturally universal component of emotion experience, which recasts concepts along a positive-negative dichotomy [29]. Although context can alter how concepts are perceived by individuals [10, 30], there are generally robust categorisations of concepts as being positive (e.g. freedom) or negative (e.g. inequality) [31, 19, 32]. This stability enables the attribution of positive/negative/neutral valence to concepts as rated by large audiences of individuals, ultimately producing a valence lexicon, i.e. a dataset attributing average normative valence and error bounds to each sampled word. Mohammad built a valence lexicon including 20,000 English words, reporting high reliability scores as expressed by different raters [31]. Beyond valence, other aspects of communication like excitement, alertness, lethargy, passivity, dominance, toxicity and many more, all influence the general emotions attributed to text [33]. This richness makes the emotion-attribution component of emotional profiling more complex than the simpler sentiment analysis.

Emotional profiling requires a complex systems perspective where a text cannot be considered as a bag of disconnected words but rather an interconnected complex system representing the ways of thinking of one or more authors [34, 30, 32, 23]. Understanding a text means activating a network of syntactic and semantic relationships between words [12, 13] that, as a whole, channels emotional stances towards specific concepts or ideas. Neglecting this network of associations can lead to partial or incomplete information. Let us consider an illustrative example. The sentence "This is not fine" relates syntactically the meaning negation "not" with the specification "fine". In a bag-of-words model, individual words possessing meaning by themselves would characterize the sentence, which would be wrongly dominated by the positive concept "fine". This example underlines the importance of considering syntactic relationships in emotional profiling of text, which can be fulfilled by cognitive networks like co-occurrence networks [15], i.e. networks where adjacent concepts in sentences are linked with each other. Patterns of syntactic relationships as captured by co-occurrence networks were found to be informative about writing styles [15] or text genre [16] and captured also key aspects of social discourse about COVID-19 and national lockdowns [30]. Considering co-occurrences can be relevant for emotional profiling in terms of frame semantics [35], a psycholinguistic theory indicating that the perceived meaning of an object is attributed and described by its conceptual associations. Identifying which concepts are linked to a certain word and then assessing the emotions elicited by such associates can both provide insightful information about the connotation attributed to that word in a given context [10, 18].

Assessing syntactic dependencies is important not only for meaning modification but it can also map: (i) how words are located within the scope of a given text, i.e. whether they are peripheral or key concepts of a text and (ii) how words of contrasting emotions are associated with each other along the flow of a given narrative/piece of text. The first point relates to the networked nature of knowledge in the human mind, where conceptual associations lead to some concepts being central hubs connecting several less connected and more specific concepts/ideas, a structure with deep implications for memory search [36]. The second point relates to emotional complexity in the flow of a narrative. Emotion complexity, sometimes referred to as emotion dialecticism [37, 38] or core affect variability

[39], captures individual differences in the tendency to experience either pleasant or unpleasant emotions within a short period of time. Evidence for emotional complexity has been accumulating in various research paradigms, such as recalling prior experience, judging hypothetical scenarios, and summarizing experience across a period of time [40, 41]. One exception is that momentary emotion experience (such as *how do you feel right now*) fails to yield evidence for mixture between positive and negative emotions, even in cultures characterised as prone to dialectic experience (e.g. Chinese culture [38]).

The main innovation of the current manuscript lies in addressing emotional profiling and the above two points of semantic prominence and emotional complexity in specific parts of suicide notes and in relation to psychologically validated emotional datasets [6]. Differently from a previous study, focusing on the overall lexical structure of notes [7], our quantitative results test: (i) whether suicide letters concentrate emotional states within specific portions of text and (ii) how positive/negative emotional states are alternated in suicide notes in comparison to a control group of emotionally healthy individuals, i.e. showing no display of anxiety, depression or stressed mental well-being. By introducing a quantitative measure of emotional complexity based on structural balance from network science, we unveil conflicting emotional transitions in suicide narratives that could not be observed with previous approaches [7] and that are considerably stronger than in recollections from mentally healthy individuals.

# Manuscript Outline

This manuscript has the following structure. The Methods section outlines the construction of the cognitive network representing the selected corpus, as well as the cognitive datasets and statistical methods adopted for emotional profiling. The Results section articulates a quantitative answer to the research question described in the foregoing section. The Discussion identifies key outcomes and limitations of this study, providing also links to future research using emotional profiling in clinical cognitive science.

#### Methods

# Corpus of genuine suicide notes

This study adopted the corpus of Genuine Suicide Notes developed and maintained by Schoene and Dethlefs [3], which includes 143 different suicide notes written by individuals who successfully committed suicide. Notes were written in English, with the names of people changed to random alternatives for privacy purposes. No personally identifiable information was present in the considered textual dataset. Suicide notes consisting of less than two sentences were not considered in the analysis, leading to 139 notes being used in the current analysis. On average, a suicide letter included only 120 words. No contextual information (e.g. socio-economic demographics) was available from the dataset. All the text from different notes was processed together in order to obtain a single cognitive network representing conceptual associations, ideas and emotions as occurring in the minds of people who committed suicide.

#### Language processing and network construction

This study used two different network construction methods for representing conceptual associations as embedded in suicide notes, namely: (i) word-word co-occurrence networks and (ii) emotional co-occurrence networks.

#### Word-word co-occurrence networks

Syntactic relationships between adjacent words can be captured by checking which words appear together or cooccur one after another in text [15, 34]. Analogously to what was done in a previous approach [7], we built a network of word-word co-occurrences, where two words were connected if one occurred immediately after the other in at least one sentence of the considered textual corpus. Notice that these networks contain mostly short-range syntactic relationships and include a variety of different words, e.g. entities ("family"), verbs ("give") and even pronouns ("she") and stopwords (words without meaning when considered in isolation like "of"). In our analysis we kept negations in order to account for them in emotional profiling (see also next section).

For Study 1, we built 2 networks of word-word co-occurrences: one considering all text present in the whole corpus (with 2075 nodes and 8676 links) and another considering only the last sentence of every suicide note in the corpus (with 204 nodes and 342 links). The co-occurrence network extracted from the whole corpus featured 426 negative words, 1054 positive words and 595 neutral words. The network of suicide notes featured a Kendall Tau correlation between the sentiment polarities of words connected by links equal to  $\kappa_t = 0.053$  (p < 0.01), indicating

a weak tendency for authors to assemble together concepts having the same sentiment polarity (e.g. negative words with negative words). The tendency for links in a network to link nodes of the same type is also known as *homophily* in social networks and the understanding of its cognitive impact over the mental lexicon structure remains an open challenge [10].

#### Emotional co-occurrence networks

In Study 2 we focused on extracting the emotional content of suicide notes, focusing our attention on the cooccurrence of emotional words within the same suicide letter.

A dictionary D of emotional words was extracted from a recent experiment using the Emotional Recall Task led by Li and colleagues [6]. In the experiment, 200 individuals recalled their emotional states in the last month and produced exactly 10 words describing it within a fluency task. The resulting dataset included 475 different stems of emotional words (e.g. happy, sad, etc.) in English. Word stems were considered in order to have only a single representation of lemmas in the network despite word declination (e.g. sad and sadness being both present in the data and both representing the stem "sad").

We then processed the suicide notes. For each letter in the corpus, we: (i) stemmed all words in its sentences, (ii) removed all words outside of the emotional dictionary D and (iii) preserved the ordering of emotional words in the letter. Punctuation was discarded. This procedure transformed the 139 suicide notes in the corpus into 139 ordered lists of emotional words. Only 119 notes/lists contained at least 5 emotional words. At the overall level of the suicide note, these lists of emotional words contain information about the order and type of emotional ideas that were mentioned by authors while recalling and writing down their final thoughts. In this way, we considered these lists as the outcome of a language production process, certainly more structured than in word recall and fluency task experiments but also tightly connected with the recall of concepts and emotions surrounding the idea of suicide. Along this interpretation, we considered the ordered lists of emotional concepts as word sequences coming from suicidal ideation, and represented conceptual relationships among them by applying Goni and colleagues' procedure of network construction from word lists [42]. The procedure considers co-occurrences of words in each list within a given window and then builds network connections between concepts that co-appeared more often than random expectation. We used a window considering l=2 words before and (l=2 words) after each entry. Links were built by counting all co-occurrences happening at least twice in the dataset and further statistical filtering was applied by considering a Binomial distribution accounting for word frequency in the dataset. For more details, we refer to the original paper [42].

This procedure represented the co-occurrence of emotional jargon across suicide notes as a complex network including 179 nodes/emotional words and 303 links. We investigate the structure of this network and its emotional complexity in Study 2. For comparison, we build a network based on the ERT task through the same methodology outlined above. However, from the ERT data we consider only word lists produced by individuals with no signals of anxiety, stress or depression symptoms. We operationalised this by considering only responses coming from individuals who achieved a total score of 0 on the Depression Anxiety Stress Scales (DASS) (cf. [6, 43]) in the ERT task. Notice that the DASS uses a four-point Likert scale, ranging from 0 (never) to 3 (most of the time), indicating how frequently individuals went through specific experiences (items) in a recent time window.

#### Cognitive datasets and emotional profiling

Words were given both a sentiment label and an emotional attribute. A sentiment label indicates whether a word is in the lower quartile (lower valence/negative perception), in the middle quartiles (neutral valence/perception) or in the upper quartile (higher valence/positive perception) of the distribution of valence scores as obtained from [31]. Sentiment labels were used in order to reconstruct the overall sentiment expressed in the considered textual corpus. Sentiment patterns were detected by counting words with positive/negative/neutral sentiment labels.

Emotional profiles were built by considering which emotions were elicited by a given word according to the NRC Emotion Lexicon [24]. Individuals enrolled in a mega-study rated which emotions were inspired by individual words. For example, "pandemic" elicited fear. The emotional profile of a text was extracted by considering how many of its words elicited a certain emotion. We call emotional richness  $r_i$  the fraction  $r_i = m/N$  of m words inspiring emotion i in a text composed of N words. By definition  $r_i$  ranges between 0 (no word in the text elicits emotion i, emotional absence) and 1 (all words elicit emotion i, complete emotional richness). Notice that a single word can elicit multiple emotions. The collection of observed emotional profiles  $r_{ii}$  constitutes the so-called *emotional profile* of a text.

The above approach might look like a bag-of-words approach, where a text is considered as the sum of its words [24, 19]. However, this is not true, as in our case the network structure determines how words are combined together

in text and makes it possible to take into account some mechanisms of meaning modification like negations. For both emotional profiling and sentiment identification in the co-occurrence network representing a given text, the antonyms of words linked with meaning negations (e.g. "not") were added to the count (antonyms were defined via WordNet 3.0 [25]).

In order to compare the observed emotional profiles to a baseline, random word sampling was considered. A collection of 1000 random emotional profiles was built in order to obtain z-scores estimating how likely it was to observe a given emotional profile through random sampling (significance level of 0.05). The collection of 8 z-scores obtained in this way, each one relative to a specific emotion, was then plotted in a sector bar chart visualisation resembling a flower layout (see Figure 2). The rejection area z < 1.96 was plotted as a semi-transparent circle; each emotion was represented as a petal-shaped bar and concentric circles indicated units of z-scores after 2, providing a visual impression for those emotions more prominently featured in a given text, analogously to the well-known wheel of emotions by Plutchik [33]. We call this visualisation an emotional flower and it was introduced in [18].

# Cognitive networks and semantic frames

The combination of word co-occurrences, sentiment and emotional profiles constitute three dimensions outlining the content of a given text. These dimensions are built in an automatic way and provide information on the semantic content (network structure), the overall sentiment (pleasantness) and emotional features of the jargon used in text. The combination of knowledge structure and emotional perceptions embedded in texts provides a rich characterisation for the content of a given text in terms of quantitative measures (see also [18]). As an example, the network neighbourhood of a given concept c identifies words directly co-occurring with c in text, thus approximating the so-called semantic frame of c itself, i.e. a collection of conceptual units characterising the meaning attributed to c in language. Further characterising a given semantic frame [35] by measuring the sentiment and emotions of its words can reconstruct the emotional perception attributed to c in the text itself.

This combination enables a powerful, automatic identification of the semantic frames and emotion perceptions affecting specific stances in a text or in its different sub-components. This automatic structure is a quantitative and approximated reflection of the cognitive component used by the author for writing down that text, i.e. their mental lexicon, the cognitive system appropriate to language acquisition, storage and use [44, 36, 9, 6, 18, 14]. In this way, exploring the network of conceptual/emotional patterns extracted from a text is a way of opening a window into people's minds.

#### Networks can quantify conceptual prominence

Conceptual prominence is measured here through a network approach relying on closeness centrality [11]. Closeness indicates the inverse average distance between one node and all of its neighbours. Nodes connected to their neighbours by few (many) links, on average, will possess higher (lower) closeness centrality. In networks of syntactic relationships, Stella [18] showed that closeness centrality was powerful in identifying words expressing the topic of short texts. In networks of free associations, recent work by [45] and replicated by [46] showed that network distance, which closeness is based on, was more powerful than latent semantic analysis in predicting semantic similarity rates in empirical experiments. Lastly, closeness centrality was highly predictive of words prominently featured in early learning environments and learned by young children during early word acquisition [8].

#### Capturing the emotional complexity of sequences of words

Keeping in mind that in an emotional network all nodes represent jargon relative to one (or more) emotional sphere(s), then links can be interpreted as transitions between emotional states of different valence. Emotional networks like the one introduced here can be applied to the quantification of *emotional complexity* in texts through two synergistic approaches. With the aim of understanding how these emotional aspects of language were expressed in language by those who committed suicide and authored suicide notes, we introduce two concepts:

- Emotional entropy, i.e. the Shannon entropy of positive, negative and neutral emotional states as encountered in suicide notes and as detected by the emotional recall task [32]. This metric is network independent.
- Emotional complexity, i.e. a translation of structural balance theory [47] into the domain of cognitive network science, where links are signed not according to social ties, but rather according to the valence of words connecting them. Thus, signed links indicate connections or flows of emotional jargon in the overall narrative of suicide notes. This metric is network dependent.

Remember that the emotional network observed here encapsulates information about the flow of emotions in the narratives provided by individuals who committed suicide. Emotionally complex stories rarely describe emotional states of the same valence exclusively, as they may transition frequently between positive and negative emotional states such as sadness and hope. Stories with lower emotional complexity, however, will tend to switch less frequently back and forth between positive and negative emotional spheres. This difference can be captured by emotional entropy, i.e.:

$$h_n = -\sum_i p_i \log(p_i),$$

where  $p_i$  is the occurrence probability of a word with valence i in a given list of emotional states as extracted from a single suicide note, n, according to the ERT lexicon. Each suicide note n corresponds to an entropy  $h_n$ . Analogously, each word list m in the ERT dataset gives rise to an entropy  $h_m$ . Statistical testing between the samples of entropies from the suicide notes  $h_{nn}$  and from the recall task  $h_{mm}$  enable further comparison between letters and recall lists.

In addition to comparing suicide letters and recall lists in terms of the entropy with which positive/negative/neutral emotional states appear in them, we also performed comparisons between each corpus and a randomised version of it. We designed two null models:

- A Shuffled model, preserving the length of each list, the dictionary of words present in these lists and their valence, but reshuffling the order in which words appear in the list;
- A Uniform model, preserving the length of each list and the dictionary of words present in these lists, but reshuffling the order in which words appear in the list and replacing their empirical positive, negative or neutral valence with a fictional value selected uniformly at random. This model disrupts correlations between valence and other quantities (e.g. frequency of positive words in a list).

A higher entropy, i.e. higher than random expectation, will be indicative of more uniformly random and less coherent transition rates between positive, negative and neutral words in a given corpus (suicide notes or ERT data).

#### Focus on emotional complexity as a network measure

We draw inspiration from psychology to operationalise emotional complexity as a network metric that quantifies how positive and negative emotional states tend to be associated or contrasted with each other in text. Building on important past attempts like Van Rensbergen and colleagues' emotional assortativity [48], Stella and colleagues emotional homophily [10] and Teixeira and colleagues emotional balance [7], which all quantified the tendency for free associations to link words sharing the same positive/negative/neutral valence in language, emotional complexity considers how triangles of conceptual associations tend to mix emotional words with opposing emotional polarities, e.g. negative and positive emotional states.

A network of linked emotional states will be considered as structurally incoherent, i.e. with a higher emotional complexity, whenever the text underlying that network contains frequent switches between positive and negative emotional jargon along the flow of narration.

Figure 1 compares emotional complexity and structural balance theory applied to networks featuring positive, negative and neutral nodes (e.g. enemies, friends or neutral agents or positive, negative and neutral words in language).

Analogous to the degree of balance in structural balance theory—i.e. the fraction of balanced triangles in a given network configuration—we here introduce the degree of emotional complexity c for a network as the fraction of emotionally incoherent triads:

$$c = \frac{n_i}{n_c + n_i},\tag{1}$$

which is the complement to 1 of the fraction of coherent triads. Notice that "complexity" for us should then be interpreted as a measure of the lack of emotional coherence. For this reason, in the visualizations for this work we will focus on coherence, 1-c, while for the more appropriate psychological interpretation of the measure we will consider complexity, i.e. c. This difference is motivated by the fact that the term "emotional coherence" already appears in other psychometric approaches. Notice that in our definition we consider neutral words as not giving rise to any contrast with either negative or positive words.

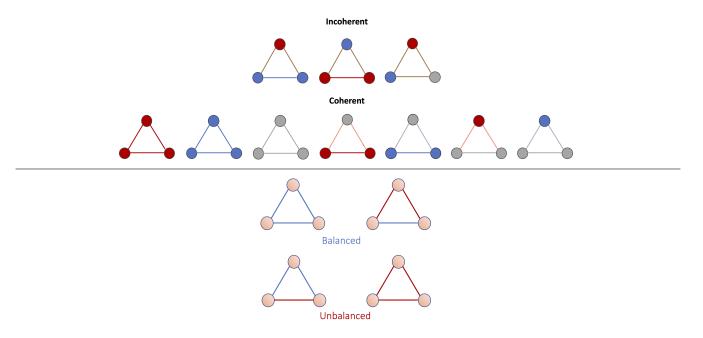


Figure 1: Comparison of emotional complexity in word networks (top) versus structural balance in social networks (bottom). Colours indicate positive (blue), negative (red) or neutral (gray) words (top) or social ties (bottom).

#### Null models for emotional complexity

Analogous to our approach with emotional entropy, in order to identify whether the abundance of coherent/incoherent triangles in the emotional network is statistically significant we identify two different null models which preserve the empirical network structure (i.e. the number of triangles observed in the real network) but reshuffle the valence of individual words in triangles. Additionally, these null models are designed to:

- preserve the frequency of positive/negative/neutral words in the empirical emotional network (Shuffled model);
- make positive/negative/neutral words occur uniformly at random (Uniform model).

Across 500 iterations of each null model, we measured z-scores for all empirical triangle counts and fixed a significance level of 0.05 in order to identify those triangles over- or underrepresented in the network (z < -1.96 or z > 1.96). For each iteration r we also measured the degree of coherence  $c_r$ , thus enabling a statistical comparison between the empirical value observed in the emotional network c and a distribution of coherence values for each null model.

# Results

Results are split in two studies. Study 1 focuses on understanding how the final sentences of suicide notes are structured and which emotions populate them in comparison to the other parts of such texts. Study 2 focuses on highlighting the emotional words of suicide notes in view of suicidal ideation and by using data coming from the emotional recall task for the first time.

In the following we refer to all words included in the last sentence of each suicide note as the "last words", outlining the closing remarks of each text.

#### Study 1: Prominence and emotional content of last words in suicide notes

This section aims to quantify how prominent or peripheral the last words were within the overall landscape of conceptual associations expressed in the suicide notes.

The considered corpus included a total of 204 last words, excluding stop-words and names of people or places. 75 of these words were positive while 39 of them were negative. Figure 2 presents a network visualisation of the largest connected component of the last words as co-occurring with each other in the suicide notes. For easier

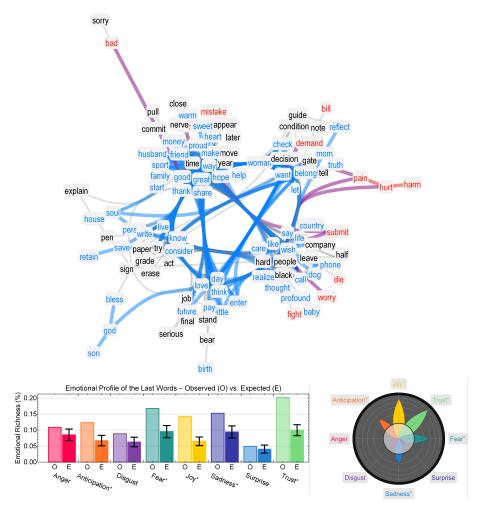


Figure 2: Network visualisation of co-occurrences between last words as expressed in the suicide notes. Cyan highlights positively perceived words whereas red highlights negative concepts. Links between positive (negative) concepts are in cyan (red). Co-occurrences between positive and negative (neutral) concepts are in purple (grey). Neutral words are in black. Left: Emotional profiles of last words observed in suicide notes against the expected profile of 1000 random samples of 204 non-last words. Right: Emotional flower reporting positive z-scores between the observed emotional profiles of last words and random expectation of non-last words.

visualisation, words were clustered according to the Louvain algorithm [49], based only on text co-occurrences. The network visualisation highlights five clusters of words used by authors in the closing remarks of their suicide notes. One cluster features spiritual concepts like "soul", "bless" and "god" while another associates several positive concepts to "time" and "family". Although dominated by concepts generally perceived as positive, the last words also included negative concepts like "harm", "pain" and "hurt". These patterns indicate a tendency for the ending remarks to express concepts of emotional positivity but also emotional pain. But, are these concepts peripheral or prominent in the whole text portrayed by authors of suicide notes?

To address this question we measured the closeness centrality of each final word in the complete network of word co-occurrences as extracted from all the considered notes. These 204 centrality scores were then compared against the closeness centrality of all 1871 other words in the network. A statistical test identified the last words as being on average more central, i.e. with a higher closeness centrality, than all other words in the suicide notes (Kruskal-Wallis,  $N_1 = 204$ ,  $N_2 = 1871$ , 77.98, p < 0.001). This difference indicates that the words used by the authors for closing their notes are most prominent in the whole discourse revolving around suicide.

No statistically significant difference was found between last words and non-last words in terms of sentiment polarity, i.e. final and non-last words were similarly distributed among positive/negative/neutral categories ( $\chi^2 = 5.54$ , p = 0.622 > 0.05).

The finer categorisation of words across emotions highlighted statistically significant differences. The emotional

profile of all 204 last words was compared against 1000 random samples of 204 non-last words drawn from the largest connected component of the co-occurrence network built on all sentences in suicide notes [7] excluding stopwords. In this way, we compared the emotional content of conceptual associations present in the last sentences of suicide letters against those links present in other parts of suicide notes. By sampling words from a network of co-occurrences rather than directly from the text, we take into account connectivity between concepts in the absence of stopwords. The resulting z-scores, comparing the emotional profiles of final and non-last words, are reported in Figure 2. As reported in Figure 2 (bottom), last words elicited anger, surprise and disgust to the same extent as other words in suicide notes. In comparison to other concepts recurring in suicide notes, the last words elicited stronger patterns of sadness and fear, but also joy, trust and anticipation.

# Study 2: Emotional content and emotional complexity of emotional words in suicide notes

This study focuses on using two measures of emotional coherence (see Methods) for the investigation of emotional switches between positive, negative and neutral emotional states in suicide notes.

#### **Emotional entropy**

A comparison of the emotional entropy of suicide notes against the reference nulls models (see Methods) identified the key role of word frequency in modulating the switches observed in suicide notes between positive, negative and neutral words. Both null models disrupted emotional associations by randomly reshuffling the same emotional states present in suicide notes. However, one null model preserved the empirical word frequency of emotional states while another implemented emotional states all with the same occurrence rate.

Considering 500 random iterations, the median entropy observed in suicide notes was lower than random expectation only for the null model discarding concept frequency (Kruskal-Wallis, K > 11.082, p < 0.0004) but not for the null model preserving word frequency (Kruskal-Wallis, K < 1.261, p > 0.262). This indicates that the empirical frequency of emotional states and the empirical length of suicide notes (in terms of emotional words) are sufficient to produce the same rate of switches observed in the empirical suicide notes. Let us consider this more deeply. We observe a certain sample or lexicon of words, i.e. a collection of words, each one either neutral or negative or positive. How much of the entropy observed in word lists could be explained by the affective composition of the lexicon? For instance, a lexicon fully composed of positive words would give rise to word lists fully made of positive words and thus with 0 entropy. In case of a lexicon combining all sentiment polarities, we would expect a higher entropy. Our testing tells us that such expected entropy, obtained by sampling words in lists uniformly at random from the lexicon, is still different from the entropy we observe in the empirical word lists. Hence, the composition of the lexicon in terms of positive/negative/neutral words is not sufficient to generate the emotional transitions we observe in the empirical data. Instead, sampling words at random but according to their frequency provides entropic rates of switches between emotional states that are compatible with the entropies observed in the empirical data.

Notice that this result does not provide any additional insight on what type of valence is predominant in the switches observed above. This is why we turn to a network analysis of emotional complexity tracking how positive and negative emotional states are associated in the flow of narrative in notes.

#### Emotional complexity

The network of emotional transitions extracted from suicide notes featured mostly incoherent +, +, - (occurring 80/278 times) and coherent +, +, + triangles (occurring 77/278 times). In order to assess whether the observed occurrences are peculiar, comparisons with suitable null models are required (see Methods). We compare emotional coherence in suicide notes and in the ERT data from mentally healthy individuals by considering two distinct networks (see Methods). For each network we also build random null models by either reshuffling word valence while preserving valence frequency and network structure (Shuffled null models), or by giving words fictional valence labels (positive/negative/neutral) uniformly at random (Uniform null models).

Notice that the set of emotional words extracted from suicide notes was a subset of the emotional states identified by the ERT data. In suicide notes, negative emotional states occurred 31% of the time, positive states were mentioned 52% of the time while the remaining 17% was composed of neutral emotional states. In the ERT data provided by mentally healthy individuals, negative emotional states occurred only 22% of the time, positive states were mentioned 48% of the time, and neutral words occurred 30% of the time. These differences might give rise to different patterns of emotional complexity that should be further investigated.

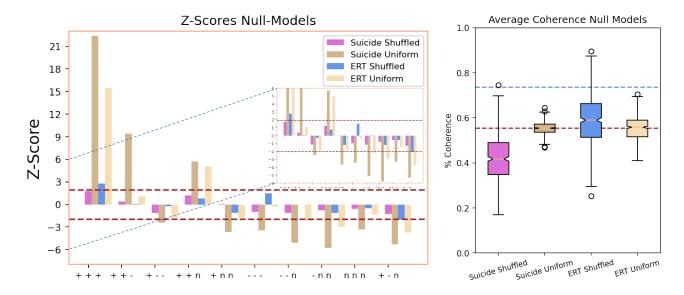


Figure 3: Left: Z-scores indicating the occurrence of signed triangles in the empirical emotional network against 2 random null models. The inset shows more clearly which z-scores are below the rejection threshold. Right: Boxplot comparing the distribution of observed emotional coherence in the null models against the value observed in the emotional network of suicide notes (dashed line).

Figure 3 (left) shows that the occurrence of coherent +,+,+ triangles in the emotional network of suicide notes is compatible with random expectation when reshuffling word valences while preserving valence frequency (Suicide Shuffled model, z < 1.96). In comparison to reshuffling word valences with uniform valence frequencies, suicide notes feature more +,+,+ triangles than expected (Suicide Uniform model, z > 1.96). Analogous patterns are observed for the +,+,- and +,+,n triangles. These differences indicate that the frequency with which emotional states are mentioned in suicide notes is sufficient for determining a high count of +,+,+,+ and +,+,n triangles, suggesting there is no evident influence of network structure in suicide notes on the formation of closed triads of positive emotional states. This is not the case in the ERT network, however, where a higher occurrence of +,+,+ triangles is observed in comparison with both the Shuffled (z = 2.9 > 1.96) and Uniform (z = 15.2 > 1.96) null models. This difference indicates that the observed count of fully positive triangles in the emotional recall from mentally healthy individuals cannot be fully accounted for by considering only valence frequency, but instead requires both valence frequency and network structure to obtain a complete picture.

For other triangles we do not observe differences between the empirical networks and the Shuffled null models. However, notice that these analyses assume that each individual triangle count is a random variable independent of others. In the Shuffled null model, since the empirical frequency of positive/negative/neutral concepts is preserved, triangle counts cannot evidently be treated as independent and this calls for additional analysis in terms of coherence assessment.

Hence, we shift our attention from individual triangle counts to the measurements of emotional complexity c (see Methods). Figure 3 (right) reports the value of c for the emotional suicide network (brown dashed line,  $c_S = 0.55$ ) and for the healthy ERT network (blue dashed line,  $c_S = 0.71$ ). Healthy individuals devoid of stress, anxiety and depression (based on their DASS subscale scores) produced conceptual associations between mental states with a coherence almost 30% higher than the one observed in emotional associations from suicide notes. Combining this result with the finding that healthy individuals recalled neutral emotional states almost twice as frequently as suicide notes, as well as considering how neutral triangles contribute to our definition of emotional complexity, the observed difference indicates that healthy individuals tend to feature more elementary cycles - i.e. triangles - of neutral emotional states in comparison to the authors of suicide notes.

Our null models can test two potential sources of emotional complexity: (i) valence-network correlations between words, e.g. trends for words of similar valence to originate cycles together, (ii) valence frequency, e.g. positive words being more frequent than negative ones and creating more coherent triangles.

Disrupting any correlation between the valence of a concept and its network location reduces the overall emotional coherence in both the considered networks. Even when preserving the empirical frequencies of valence, the Shuffled models produce networks with evidently lower median emotional complexity than their respective empirical values.

Hence, valence frequencies by themselves are not enough to reproduce the emotional coherence observed in both the suicide and ERT emotional networks. These patterns of emotional coherence are therefore to be attributed to an interplay between affect and network structure which ultimately makes emotional shifts in either suicide notes or recall tasks less "switchy" and more confined within elementary cycles of words sharing the same valence.

Disrupting the empirical valence frequency in the observed network leads to the same average value of emotional complexity. Notice how preserving valence frequency also preserves the observed gap in emotional complexity between suicide notes and healthy individuals. This indicates that whereas the coherence of each network stems from correlations between network structure and valence, the observed gap in coherence between suicide notes and the ERT data has to be attributed mainly to a different frequency of positive/negative/neutral emotional states.

To summarise, the above results indicate that the narratives in suicide notes are more structurally coherent than expected by a null model where valenced words are sampled according to the same frequency observed in the empirical dataset but independently from network structure. Furthermore, the narratives in suicide notes are less emotionally coherent than observed in emotional associations provided by healthy individuals, a difference that can be explained by a higher occurrence of neutral emotional states in healthy individuals in comparison to suicide notes.

# **Discussion and Conclusions**

This study reconstructed the collective mindset expressed by authors of suicide notes [3, 23] using cognitive network science. This mindset was represented as a network of word co-occurrences, capturing syntactic associations from text, enriched with emotional data. The network structure enabled a quantitative comparison between different groups of words. The last words used by authors at the end of their letters were shown to be highly prominent concepts in the overall discourse of suicide notes, as quantified by closeness centrality. Authors might therefore use these last words in order to reinforce their narrative, repeat and summarise their points or express general-level concepts highly relevant to what they described in their notes. In line with recent advancements in natural language processing using cognitive data for improving model interpretability [50], the cognitive network structure provides a transparent snapshot of the semantic content and relationships between such words in text. Semantic analysis reveals that these prominent last words are mostly clustered around the topics of family, religion and painful emotional processing, in agreement with previous human coding studies of suicidal ideation [51]. In comparison to other less prominent concepts expressed in suicide notes, last words elicited more intense feelings of sadness and fear, but also elicited joy, trust and anticipation.

Trust and joy are emotions in line with the above semantic analysis exposing clusters of family-related jargon in the closing remarks of suicide notes. The stronger intensity reported here indicates that the closing remarks contain concepts strongly expressing positive emotions, providing a trustful ending to notes. However, in the considered collection of suicide notes, this positive remark co-existed with anticipation, fear and surprise, a combination that expresses uncertainty about the future [33]. The co-existence of fear and trust in last words can also be related to the above cluster of words which express painful emotional processing. Once combined, these patterns indicate that suicide notes also express jargon of trustful relief from fearful and painful concepts, providing quantitative support to psychological theories indicating suicidal ideation as a pain relief mechanism [51, 3].

This preliminary study outlines the quantitative potential for cognitive network science to give structure to language and its perception. This framework relies on the assumption that every author possesses a cognitive system, a mental lexicon [9, 11, 10], where linguistic knowledge and emotional perceptions are highly structured in an interdependent fashion. The approach outlined here might potentially be integrated with natural language processing at the stage of network creation in order to achieve more realistic knowledge representations expressing syntactic dependencies [18] or sociolinguistic features of texts [23]. Under the umbrella of complex systems, recent work has successfully used co-occurrence networks for the investigation of authorial writing style [15, 34] and social media discourse [17]. Co-occurrences were merged with emotional features in recent investigations of social media discourse around the gender gap in science [18] and the COVID-19 pandemic [30, 32], opening the way to future mindset reconstructions of different authors through additional research.

This work also introduced the network metric of emotional complexity in terms of counting triads connecting positive, negative and neutral emotional words. Emotional complexity considers different kinds of signed triangles in comparison to structural balance. In structural balance theory, triangles are identified as unbalanced whenever there is an odd number of negative links, i.e. an odd number of tense relationships. This rule captures tension in social connections and is a phenomenon supported by relevant social science literature (e.g. [52, 53, 54]). One possibility is to extend structural balance also to valenced words, as was done in previous work [7], in order to quantify tension in the outline of conceptual associations featuring negative words. However, when considering

words made of emotional states only, as in the present case, it is possible to shift the attention from an overall level of tension between concepts to the specific tendency for elemental cycles in the network to feature transitions between emotional states of opposing valence. This shift is the key difference between the current approach with emotional complexity and the one by Teixeira and colleagues in terms of emotional balance [7]. Whereas in emotional balance triads are categorised according to their links, in emotional complexity triads are classified according to the valence of their nodes. Configurations that are emotionally balanced can end up being treated as incoherent in terms of emotional complexity. Considering coherence/balance in terms of nodes rather than in terms of links is evidently different from structural balance and more akin to emotional complexity in cognitive psychology [37, 29, 38]. In other words, whereas structural balance focuses on the positive/negative nature of social ties—operationalising the balance of concepts like "the enemy of my enemy is my friend"—emotional complexity focuses on the valence of individual emotional states. As opposed to social ties, when considering emotional complexity and its network counterpart a triangle connecting two negative and one positive emotional concept must be considered as incoherent, as it captures an emotional transition between negative and positive emotional states.

The emotional complexity (coherence) of suicide narratives was found to be higher (lower) than that of healthy individuals recalling recent emotions over the past month. Emotion complexity captures an individual's tendency to experience a mixture of positive and negative emotions over time [55], and therefore unsurprisingly relates to the time frame within which people recall their emotional experience. That is, people are more likely to report a greater mixture of positive and negative emotions when describing their experience over a long period of time than their experience at the moment [55]. The higher degree of complexity in suicide notes may suggest that these notes refer to emotions over a long period in life instead of emotions at the moment of writing. Indeed, as we read though the suicide notes, many notes refer both to relatively recent sadness (e.g. "I just can't carry on the despair") and an overview of life in general (e.g in the same note: "I have looked at our 25 years mainly as happy years and had hope to spend old age together"). Moreover, high complexity in suicide notes may also suggest that people who chose suicide were not simply dominated by negative feelings; instead, they still have access to their positive memories.

These results suggest new topics for further research, particularly to improve an understanding of the relationships between emotional complexity, suicidal ideation (SI), and suicidal behavior (SB). If emotional complexity is predictive of SI or SB, it would be important to study the degree to which it reflects an individual trait or is relatively stable over time. Some clinical psychological studies with a similar aim have shown that affective variability (i.e., the extent to which a person experiences fluctuations in mood) and variability in SI may reflect stable traits [1] that are associated over time [2], and that affective variability may be an effective predictor of SI [5]. These studies used experience sampling methods to see how such constructs relate over time on a day-to-day scale, however such relationships have only been demonstrated in certain high-risk populations. Extending this work to investigate how emotional complexity and emotional entropy relate with risk of SI or SB would thus be a potentially valuable topic to study in larger populations.

The approach outlined here has several limitations. The most prominent is the necessity of using sentiment and emotional labels extracted from large populations for reconstructing an author's mindset. This methodology neglects subjective interpretations, e.g. high school students might perceive "physics" as a negative concept, whereas large populations might perceive it as neutral [10]. A way to overcome this limit would be the adoption of language processing techniques identifying the context around a given concept and inferring its emotional perception [56]. This is possible even with cognitive networks by considering valence and emotions in the semantic frame of associates around a concept. In this way, focus should be given not to the specific sentiment or emotional labels attributed to a given word but rather to its structural position among its neighbors in the text. Another limitation of our study is relative to emotional complexity focusing only on elementary cycles, i.e. closed triads, and disregarding other network patterns that might arise from emotional disassortativity. In order to address this point we considered not only emotional complexity but also emotional entropy, the latter being a non-network metric considering shifts between emotional states in the whole datasets. Future research could integrate emotional complexity and entropy with recent measures like conformity [57], mixing network structure and node features.

# Competing interests

The authors declare that they have no competing interests.

# Acknowledgements

The authors acknowledge Cynthia S. Q. Siew (National University of Singapore) and Szymon Talaga (University of Warsaw) for constructive feedback about the manuscript scope and structure. A.S.T. acknowledges national funds through FCT, Fundação para a Ciência e a Tecnologia, under project UIDB/50021/2020.

# References

- [1] Oquendo, M.A., Galfalvy, H.C., Choo, T.-H., Kandlur, R., Burke, A.K., Sublette, M.E., Miller, J.M., Mann, J.J., Stanley, B.H.: Highly variable suicidal ideation: a phenotypic marker for stress induced suicide risk. Molecular psychiatry, 1–8 (2020)
- [2] Rizk, M.M., Choo, T.-H., Galfalvy, H., Biggs, E., Brodsky, B.S., Oquendo, M.A., Mann, J.J., Stanley, B.: Variability in suicidal ideation is associated with affective instability in suicide attempters with borderline personality disorder. Psychiatry 82(2), 173–178 (2019)
- [3] Schoene, A.M., Dethlefs, N.: 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, pp. 128–133 (2016)
- [4] Hallensleben, N., Glaesmer, H., Forkmann, T., Rath, D., Strauss, M., Kersting, A., Spangenberg, L.: Predicting suicidal ideation by interpersonal variables, hopelessness and depression in real-time. an ecological momentary assessment study in psychiatric inpatients with depression. European Psychiatry **56**(1), 43–50 (2019)
- [5] Palmier-Claus, J., Taylor, P.J., Gooding, P., Dunn, G., Lewis, S.: Affective variability predicts suicidal ideation in individuals at ultra-high risk of developing psychosis: An experience sampling study. British Journal of Clinical Psychology 51(1), 72–83 (2012)
- [6] Li, Y., Masitah, A., Hills, T.T.: The emotional recall task: Juxtaposing recall and recognition-based affect scales. Journal of Experimental Psychology: Learning, Memory, and Cognition (2020)
- [7] Teixeira, A.S., Talaga, S., Swanson, T.J., Stella, M.: Revealing semantic and emotional structure of suicide notes with cognitive network science. arXiv preprint arXiv:2007.12053 (2020)
- [8] Stella, M., Beckage, N.M., Brede, M.: Multiplex lexical networks reveal patterns in early word acquisition in children. Scientific reports 7, 46730 (2017)
- [9] De Deyne, S., Kenett, Y.N., Anaki, D., Faust, M., Navarro, D.J.: Large-scale network representations of semantics in the mental lexicon. Big data in cognitive science: From methods to insights, 174–202 (2016)
- [10] Stella, M., De Nigris, S., Aloric, A., Siew, C.S.: Forma mentis networks quantify crucial differences in stem perception between students and experts. PloS one **14**(10), 0222870 (2019)
- [11] Siew, C.S., Wulff, D.U., Beckage, N.M., Kenett, Y.N.: Cognitive network science: A review of research on cognition through the lens of network representations, processes, and dynamics. Complexity **2019** (2019)
- [12] Lynn, C.W., Papadopoulos, L., Kahn, A.E., Bassett, D.S.: Human information processing in complex networks. Nature Physics, 1–9 (2020)
- [13] Lydon-Staley, D.M., Zhou, D., Blevins, A.S., Zurn, P., Bassett, D.S.: Hunters, busybodies and the knowledge network building associated with deprivation curiosity. Nature human behaviour 5(3), 327–336 (2021)
- [14] Mehler, A., Gleim, R., Gaitsch, R., Hemati, W., Uslu, T.: From topic networks to distributed cognitive maps: Zipfian topic universes in the area of volunteered geographic information. Complexity **2020**
- [15] Amancio, D.R.: A complex network approach to stylometry. PloS one 10(8), 0136076 (2015)
- [16] Corrêa Jr, E.A., Marinho, V.Q., Amancio, D.R.: Semantic flow in language networks. arXiv preprint arXiv:1905.07595 (2019)
- [17] Stella, M., Ferrara, E., De Domenico, M.: Bots increase exposure to negative and inflammatory content in online social systems. Proceedings of the National Academy of Sciences 115(49), 12435–12440 (2018)

- [18] Stella, M.: Text-mining forma mentis networks reconstruct public perception of the stem gender gap in social media. Peer J Computer Science 6, 295 (2020)
- [19] Mohammad, S.M.: Sentiment analysis: Detecting valence, emotions, and other affectual states from text. In: Emotion Measurement, pp. 201–237. Elsevier, ??? (2016)
- [20] Acerbi, A., Lampos, V., Garnett, P., Bentley, R.A.: The expression of emotions in 20th century books. PloS one 8(3), 59030 (2013)
- [21] Nguyen, D., Liakata, M., DeDeo, S., Eisenstein, J., Mimno, D., Tromble, R., Winters, J.: How we do things with words: Analyzing text as social and cultural data. arXiv preprint arXiv:1907.01468 (2019)
- [22] Montefinese, M., Ambrosini, E., Visalli, A., Vinson, D.: Catching the intangible: A role for emotion? The Behavioral and brain sciences 43, 138 (2020)
- [23] Boyd, R.L.: Psychological text analysis in the digital humanities. In: Data Analytics in Digital Humanities, pp. 161–189. Springer, ??? (2017)
- [24] Mohammad, S.M., Turney, P.D.: Crowdsourcing a word–emotion association lexicon. Computational Intelligence **29**(3), 436–465 (2013)
- [25] Fellbaum, C.: Wordnet. The encyclopedia of applied linguistics (2012)
- [26] Hoemann, K., Devlin, M., Barrett, L.F.: Comment: Emotions are abstract, conceptual categories that are learned by a predicting brain. emotion review. Emotion Review 12(4)
- [27] Hoemann, K., Xu, F., Barrett, L.F.: Emotion words, emotion concepts, and emotional development in children: A constructionist hypothesis. Developmental Psychology **55**(9)
- [28] Lindquist, K.A., MacCormack, J.K., Shablack, H.: The role of language in emotion: Predictions from psychological constructionism. Frontiers in Psychology 6, 444 (2015)
- [29] Barrett, L.F.: Solving the emotion paradox: Categorization and the experience of emotion. Personality and Social Psychology Review 10(1), 20–46 (2006)
- [30] Stella, M., Restocchi, V., De Deyne, S.: #lockdown: Network-enhanced emotional profiling in the time of covid-19. Big Data and Cognitive Computing 4(2), 14 (2020)
- [31] Mohammad, S.: 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), pp. 174–184 (2018)
- [32] Li, Y., Hills, T., Hertwig, R.: A brief history of risk. Cognition 203, 104344 (2020)
- [33] Plutchik, R.: Emotion. A psychoevolutionary synthesis (1980)
- [34] de Arruda, H.F., Marinho, V.Q., Costa, L.d.F., Amancio, D.R.: Paragraph-based representation of texts: A complex networks approach. Information Processing & Management 56(3), 479–494 (2019)
- [35] Fillmore, C.J., et al.: Frame semantics. Cognitive linguistics: Basic readings 34, 373–400 (2006)
- [36] Hills, T.T., Todd, P.M., Lazer, D., Redish, A.D., Couzin, I.D., Group, C.S.R., et al.: Exploration versus exploitation in space, mind, and society. Trends in cognitive sciences 19(1), 46–54 (2015)
- [37] Bagozzi, R.P., Wong, N., Yi, Y.: The role of culture and gender in the relationship between positive and negative affect. Cognition Emotion 13(6), 641–672 (1999)
- [38] Scollon, C.N., Diener, E., Oishi, S., Biswas-Diener, R.: An experience sampling and cross cultural in- vestigation of the relation between pleasant and unpleasant affect. Cognition Emotion 19(1), 27–52 (2005)
- [39] Kuppens, P., Van Mechelen, I., Nezlek, J.B., Dossche, D., Timmermans, T.: Individual differences in core affect variability and their relationship to personality and psychological adjustment. Emotion 7(2), 262 (2007)
- [40] Robinson, M.D., Clore, G.L.: Episodic and semantic knowledge in emotional self-report: Evidence for two judgment processes. Journal of Personality and Social Psychology 83(1), 198–215 (2002)

- [41] Barrett, L.F.: Computing global structural balance in large-scale signed social networks. Personality and Social Psychology Bulletin **23**(10), 1100–1110 (1997)
- [42] Goñi, J., Arrondo, G., Sepulcre, J., Martincorena, I., De Mendizábal, N.V., Corominas-Murtra, B., Bejarano, B., Ardanza-Trevijano, S., Peraita, H., Wall, D.P., et al.: The semantic organization of the animal category: evidence from semantic verbal fluency and network theory. Cognitive processing 12(2), 183–196 (2011)
- [43] Lovibond, P.F., Lovibond, S.H.: The structure of negative emotional states: Comparison of the depression anxiety stress scales (dass) with the beck depression and anxiety inventories. Behaviour research and therapy 33(3), 335–343 (1995)
- [44] Vitevitch, M.: Can network science connect mind, brain, and behavior. Network science in cognitive psychology, 184 (2019)
- [45] Kenett, Y.N., Levi, E., Anaki, D., Faust, M.: The semantic distance task: Quantifying semantic distance with semantic network path length. Journal of Experimental Psychology: Learning, Memory, and Cognition 43(9), 1470 (2017)
- [46] Kumar, A.A., Balota, D.A., Steyvers, M.: Distant connectivity and multiple-step priming in large-scale semantic networks. Journal of Experimental Psychology: Learning, Memory, and Cognition 46(12), 2261 (2020)
- [47] Moradimanesh, Z., Khosrowabadi, R., Gordji, M.E., Jafari, G.: Altered structural balance of resting-state networks in autism. Scientific reports 11(1), 1–16 (2021)
- [48] Van Rensbergen, B., Storms, G., De Deyne, S.: Examining assortativity in the mental lexicon: Evidence from word associations. Psychonomic bulletin & review **22**(6), 1717–1724 (2015)
- [49] Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. Journal of statistical mechanics: theory and experiment **2008**(10), 10008 (2008)
- [50] Škrlj, B., Martinc, M., Kralj, J., Lavrač, N., Pollak, S.: tax2vec: Constructing interpretable features from taxonomies for short text classification. Computer Speech & Language 65, 101104 (2021)
- [51] Sanger, S., Veach, P.M.: The interpersonal nature of suicide: A qualitative investigation of suicide notes. Archives of Suicide Research 12(4), 352–365 (2008)
- [52] He, X., Du, H., Cai, M., Feldman, M.W.: The evolution of cooperation in signed networks under the impact of structural balance. PloS one 13(10), 0205084 (2018)
- [53] Aref, S., Wilson, M.C.: Balance and frustration in signed networks. Journal of Complex Networks 7(2), 163–189 (2019)
- [54] Doreian, P., Mrvar, A.: Structural balance and signed international relations. Journal of Social Structure 16, 1 (2015)
- [55] Diener, E., Emmons, R.A.: The independence of positive and negative affect. Journal of personality and social psychology 47(5), 1105–1117 (1985)
- [56] Van Rensbergen, B., De Deyne, S., Storms, G.: Estimating affective word covariates using word association data. Behavior Research Methods 48(4), 1644–1652 (2016)
- [57] Rossetti, G., Citraro, S., Milli, L.: Conformity: A path-aware homophily measure for node-attributed networks. IEEE Intelligent Systems **36**(1), 25–34 (2021)