

Reviewing Theoretical and Generalizable Text Network Analysis: Forma Mentis Networks in Cognitive Science

Oleksandra Poquet ^{1,2} and Massimo Stella ³

¹ Learning Planet Institute, Universite Paris Cite & INSERM 1284, Paris, France

² C3L, Education Futures, University of South Australia, Adelaide, Australia

³ CogNosco Lab, Department of Computer Science, University of Exeter, Stocker Road, Exeter, UK

Abstract

Recommendations for network studies in learning analytics emphasize that network construction requires careful definitions of nodes, relationships between them, and network boundaries. Thus far, researchers in learning analytics have discussed how to operationalize interpersonal networks in learning settings. Analytical choices used in constructing networks of text have not been examined as much. By reviewing examples of text network analysis in learning analytics we demonstrate that convenience-based decisions for network construction are common, particularly when the ties in the text networks are defined as the co-occurrences of words or ideas. We argue that such an approach is limited in its potential to contribute to theory or generalize across studies. This submission presents an alternative approach to network representations of the text in learning settings, using the concept of Forma Mentis Networks. As reported in previous studies, Forma Mentis Networks are network representations either (1) elicited from individuals through free association tasks that capture valence or (2) constructed by analysts creating shared mental maps derived from text. Forma Mentis Networks is a theory-based and scalable approach complementary to the existing set of tools available for the analysis of teaching and learning.

Keywords

Text analysis; learning analytics; cognitive network science.

1. Introduction

Quantitative text analysis describes a family of the methodologies commonly used in learning analytics, including but not limited to content analysis [1], discourse analysis [2], and natural language processing (NLP) [3]. Computational techniques used within these methodologies leverage various aspects of the text to understand learning and knowledge creation. Some of these techniques capture discourse characteristics, for instance, using measures of coherence [4], others - analyze the meaning of text, for example, by analyzing key concepts identified through supervised or unsupervised machine learning [5].

Automated content analysis is well suited for a quick high-level summary of the frequent concepts across large quantity of texts. However, its utility for nuanced research insights has been challenged in seminal work [6], where Carley pointed out that quantitative content analysis focuses on isolated concepts within the text, for instance, frequency of a keyword. The meaning of the keyword derived from contextual relationships to the other words, concepts, and ideas, is therefore, lost. These makes the texts decontextualized, and the comparisons between them can become biased. In contrast, analyzing texts in ways that preserve inter-word relationships can help account for the lack of context. Such relationships can be defined using semantic, proximal, and linguistic perspectives [7], [8]. Carley suggested using network-based representations of text, so that more nuanced meaning can be quantified and analyzed. The applications of network text analysis, broadly referred to by

Proceedings of the NetSciLA22 workshop, March, 2022, Online

EMAIL: sasha.poquet@cri-paris.org (A. 1); massimo.stella@inbox.com (A. 2)

ORCID: 0000-0001-9782-816X (A. 1); 0000-0003-1810-9699 (A. 2).



© 2020 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

Carley, as *map analysis*, draw on the scholarship in text analysis and network science. Therefore, analyzing text as a network is an inter-methodological endeavor: representing text as a network requires theoretical justifications grounded in text analysis, whereas analyzing such a network requires the knowledge of graph theory.

This short review argues that currently analyzing text networks in learning analytics research can benefit from further theoretical and analytical rigor. This argument is not new. Recommendations for network studies put forward by the participants of NetSciLA21 workshop similarly emphasized that when researchers construct networks in learning settings, they need to carefully define nodes and relationships for the network representation. How to operationalize networks of people who interact in learning settings had been discussed elsewhere [9], [10]. Here, we further argue that scrupulous methodological and theoretical considerations are less common when learning analytics researchers conduct network analysis of text. First, we provide examples of text analysis in learning analytics, highlighting their over-reliance on convenience-based decisions, such as the co-occurrence of concepts. As we explain, such examples are limited in their ability to contribute to theory or generalize across studies. We, then, offer alternatives to network representations of the text in learning settings, using the concept of Forma Mentis Networks (FMN). FMNs are network representations either elicited from individuals through free association tasks that capture valence [11] or constructed by analysts creating shared mental maps through text [12]. In either approaches, network operationalizations are theoretically grounded in mental maps and cognitive knowledge theories [13] and aligned with the theoretical tenets of semantic memory underpinning cognitive structures. We explain how FMNs can contribute to theory, generalizability of findings, external validity, as well as offer a trade-off between scalability and the presence of noise in a network representation. Based on this discussion, we demonstrate that FMN is a theory-based and scalable approach complementary to the existing set of tools available for the analysis of teaching and learning.

2. A Critical Review of the Text Networks in Learning Analytics

Learning analytics research commonly uses networks to represent and analyze texts created by learners. Examples of student-produced text analyzed through networks include student personal reflection essays [14], socially shared annotations [15], and the online messages posted in group discussions [16]. For such analyses, researchers make methodological decisions such as (1) defining a node in a network (e.g., a word, phrase, idea unit); (2) selecting the unit of analysis (e.g., personal essay, personal post, a sentence, a paragraph, a discussion thread); and (3) defining the meaning of an edge, which represents a relationship between the nodes (i.e., the co-occurrence of nodes within the unit of analysis). In this section we explain that some of these decisions are convenience-based, rather than theoretically grounded, and that this compromises the validity of the network representations.

2.1. Co-occurrence Networks

The use of co-occurrence of words in a sentence is a common approach to defining network edges in text networks. This section provides a few examples outlining this approach. The basic steps of such an approach can be found in Wise & Cui [14] where the researchers assess written reflection essays of dental students. The study examines and compares text networks of the ‘top concepts’ used by the students. First, the researchers extracted unigrams most frequently used in the set of student essays. These unigrams were theorized as ‘top concepts that the students reflected on’. A tie between two concepts was created if they were co-located within the same sentence, i.e., based on co-occurrence. Such an approach produces dense networks, where many words co-occur many times, with the times representing the edge weight. To see the underlying network structure clearer, the study further manually filtered the edge weights between co-occurring concepts. The resulting networks, representative of reflections written at different times in a semester, were then qualitatively inspected. Wise, Reza, and Han [17] further extended this approach towards an improved understanding of how students use prominent constructs in their reflections. In their 2020 study, researchers first applied machine learning, to cluster essays based on the similarity of text in them.

Only then, they identified the top concepts. Here, the researchers constructed ego-networks of the prominent top constructs, i.e. each network centered around a top construct within the cluster, connecting with the words that co-occurred with this top construct. These networks at the level of a top construct were analyzed over longer periods of time and were qualitatively interpreted in relation to professional identity formation by the students. On the one hand, this approach was helpful in revealing the trends. On the other hand, the number of qualitative decisions and human interpretation during processing make it somewhat challenging to replicate.

More nuanced automated approaches to text processing that precede the construction of co-occurrence networks are also available. For instance, Joksimovic and colleagues [18] analyzed Twitter exchanges in an open online course, to identify prominent themes discussed by the course participants. Here, instead of using unigrams, the state-of-the-art automated annotation tools were applied to extract keywords from the original student text. Co-occurrence networks were then constructed from these keywords. If two keywords co-occurred within the post produced by a student, they were linked by the tie. Further a graph-clustering algorithm was chosen to separate words into themes, rather than identify them through subjective interpretation. Specifically, Joksimovic et al. applied a graph modularity algorithm to a strongly connected component of the keywords network, to identify prominent clusters. The clusters were then described quantitatively in relation to prominence of keywords within them and thematically interpreted.

A different way of defining nodes and edges in co-occurrence networks has been suggested by van Labeke et al. [19] and Whitelock et al. [20] who also examined text networks to assess the quality of a student essay. In van Labeke et al. [19], the researchers defined sentences as nodes in the network. If a word co-occurred in two sentences, they are linked; each edge between two sentences is weighted to reflect the cosine similarity between the words in a pair of sentences. A graph ranking algorithm is then used to derive prominent essay sentences within such a network. Later work by the group continued exploring the application of network analysis of text in student essays. In Whitelock et al. [20] researchers offered feedback on student essays with network representations of the student text. They presented the networks from student essays, coloring nodes (sentences in the network) in different colors, to show they belonged to different parts of the essay (introduction, conclusion, etc). To validate the effectiveness of these representations, the researchers complemented the networks with grades and human expert evaluation.

A variation of the analysis of co-occurrence networks is offered by Yun and Park [21]. The researchers used the transcripts of science teachers' classroom talk to construct the networks of words co-located in the same sentence. The words prominent in such networks were compared to the words frequently appearing in scientific corpus related to the same subject area. Peculiar to this study is that the researchers used external discourse to evaluate and understand how teachers were explaining content in the science classroom.

As shown through the above examples, analytical decisions around co-occurrence networks vary. The questions remain as to whether some of the automated approaches offer insight into theoretical perspectives, as well as whether the more theory-grounded analytical steps that require human interpretations, can be successfully replicated.

2.2. Epistemic network analysis

Epistemic network analysis (ENA), an approach that has recently gained prominence for the analysis of texts generated in the classroom, essentially is a co-occurrence network too. In ENA, the networks reflect 'the structure of connections among coded elements by quantifying the co-occurrences of codes within a defined segment of data, or stanza' [22]. That implies that in many cases, ENA requires researcher-imposed segmentation of the text (i.e., decision around the unit and level of analysis), as well as qualitative interpretation of text (i.e., content analysis or thematic analysis). Consequently, ENA inherits methodological assumptions required for content analysis that has developed knowledge base and systematic protocols for qualitative decisions around the unit and level of analysis, as well as the assignment of themes. ENA leverages these decisions by applying co-occurrence to link qualitatively derived codes within a network. In addition to these components ENA utilizes innovative technique to create a visual projection of this network, as raw counts of code co-occurrence within the stanzas are transformed. Vectors of code co-occurrence are processed using

single vector decomposition and normalized to control for the more ‘crowded’ stanzas. These transformations allow to reduce the noisy co-occurrence networks to highlight more prominent relations and to visualize the co-occurrences in a more replicable and comparable manner. The ENA-specific transformation enables comparisons between these processed representations at the levels of individuals or groups.

Replication and scalability of deriving codes (network nodes) within the stanzas (units of analysis) remain in tension with theoretical grounding. For instance, a theory can guide which codes are selected and how, but replication of such analysis requires detailed report of the content analysis approach in the study, whereas automating it requires further human annotation and training of supervised machine learning models. Researchers have been working to create automated approaches to deriving key codes [23]. These techniques (described below) offer a significant advancement of the method, and yet, they are less theoretically grounded than human interpretation of the texts. For instance, Ferreira and colleagues applied natural language processing, Latent Dirichlet Allocation (LDA) in particular, to identify topics within posts made by students. By way of background, LDA estimates the probability of a topic to be associated with a particular unit of analysis, here a post in a discussion forum. Ferreira and colleagues substituted the matrix of binary co-occurrences between codes in stanzas, which is commonly used as input for ENA, with a different matrix, capturing the relationship between the posts and LDA-derived topics, with edge weights representing probability of topic to appear in that post. The authors explain that the ENA pipeline processes this matrix with probability weights, similarly to the co-occurrence matrix, using single value decomposition with the possibility to transform the weights via direct product, square root, or natural log methods.

Computational innovations help ENA evolve to eventually overcome the tensions between theoretical insights, replicability, and scalability it offers. However, so far, the ENA pipeline is methodologically eclectic. Its input can be both derived using ‘interpretivist’ data collection requiring content analysis and thematic analysis, as well as using data that is not interpreted but created by machine learning approaches. In either case, the relationships between the codes, regardless of when derived from human interpretation or via NLP, are operationalized through co-occurrence, suggesting that ENA does share certain methodological and theoretical assumptions with the methods described in the previous section.

2.3. Socio-semantic network analysis and network-text analysis

Learning sciences offer theoretically grounded frameworks to the analysis of learning, where relational thinking (about students and text they produced) is inherent to the theory explaining the learning processes. A prominent example is the work by Oshima and colleagues around modelling knowledge building processes [24]. Oshima and colleagues [16] describe their application of network analysis to student discourse produced in a digital environment. Knowledge building approach aims to help students to collectively develop ideas as well as produce artifacts that help them refine their ideas. Oshima and colleagues use socio-semantic network analysis (SSNA) to understand how students develop ideas. They represented student ideas as ‘clusters of words in the network of words’ (p.1312), therefore, linking the words if they have co-occurred in the discourse exchange units, which are group-level discussion units within Knowledge Building Discourse Explorer digital environment. By showing how centrality of different words changes due to their position in co-occurrence networks, researchers glean insights on how some ideas persist and do not. Although the approach links students and text, the underlying principle of linking co-occurent meaning units is applied here as well.

A dynamic extraction of keywords from text prior to network construction is offered through the Network-Text Analysis (NTA) approach by Taskin, Hecking, and Hoppe [25]. The approach presents yet another graph-based application for filtering the noisy co-occurrences, which requires decisions around window size and thresholds that can be challenged around its theoretical considerations. They proposed a technique for extracting networks of concepts appearing in texts as linked by certain measures of proximity. The authors emphasize that the reduction of the number of relationships between words/keywords to exclude those that are not meaningful are among key challenges for co-occurrence networks. To this end, the authors apply an explicit semantic analysis approach that infers more meaningful entities. Once entities or keywords are identified, the edges between them are

created based on the moving window approach: the words are not separated by more than $k-2$ words, for instance 20 words in an example presented by the authors. Once this step is completed; the authors suggest further filtering of the concept networks based on a threshold that is fine-tuned to the dataset.

2.4. Other network approaches to text analysis and learning analytics

Most of the approaches discussed above use co-occurrence of words to define relationship between the nodes in text networks. However, text analysis has been combined with network analysis in learning analytics methodologies in other ways as well. Some approaches combine text analysis with network analysis to delineate groups of people that shared discourse of particular kind, then analysing relationships between these individuals. We provide three prominent examples that reflect the techniques. Hecking and colleagues [26] apply NLP to semantically identify conversations exclusively focused on the subject matter. Then, the relationships between learners who exchanges only this subject matter content are constructed and analysed via network analysis. Dascalu and colleagues [27] apply the so-called coherence network analysis – an NLP-based approaches that creates links between people based on the similarity between their text. Hecking and colleagues [15] construct bipartite networks of learners and words from learner-contributed video comments. Clustering and analysis are then conducted combining network structure linking learners and text. The latter example is not limited to this particular study (for other examples, see [28]). These applications are out of scope in this review as they do not focus on the knowledge structures, but on the interpersonal structures underpinning knowledge exchanges.

2.5. Critical reflection and under-explored research areas

Across all the above examples we can observe that the more scalable approaches are often less theoretically grounded. In many of these studies analytical decisions are based on convenience and require contextual decisions that impede generalisability. As a result, interpreting text networks, i.e. addressing the question of ‘what these networks represent’, is non-trivial, and its external validity is limited. That is not to say that the problems like these characterise all the examples. Theory-grounded approaches to integrating text into networks include bipartite networks of learners and constructs they contributed to where the presence of text and ties are theoretically justified, as well as graphs derived from concept maps constructed by learners themselves [29]. In the first example, a network representation is an operationalization of an emergent discourse-mediated community, in the second example - it is student knowledge representation. However, theory-based interpretations of text networks built on co-occurrence ties, which is a more commonly used approach, are difficult to infer.

This brief review highlights that, so far, the potential of text networks in learning analytics to contribute to theory or generalize beyond specific examples, is limited, despite their pragmatic utility in deriving insight for specific cases. Defining ties between words or phrases through co-occurrence is a practical and convenient decision that requires little theoretical knowledge of what is being represented. However, such a decision can lead to oversimplifying the patterns captured through the network. Co-occurrence, particularly when it comes to representing networks of text, is a crude tie definition. Its power to derive an insight comes at a cost: the noise in the network that treats multiple types of linguistic relationships similarly [30]. When further coarse graining takes place, for instance, if the concepts used in a network as nodes are derived computationally using machine learning algorithms, the nuances in meaning that individuals assign to perceived language, are further washed out.

3. Forma Mentis Networks in Cognitive Network Science as a Suitable Theoretical Framework

Cognitive science has evolved to include network approaches with rigorous and theory-grounded network definitions, offering alternative inspiration to researchers in learning analytics [31]. Cognitive networks are conceptualized as the mental reflection of language and associative

knowledge in the human mind [32]. Accessing cognitive networks means reading people's minds, accessing people's perceptions as associations of ideas about environments, opinions, and emotions. Overwhelming empirical research [13], [33] has supported the importance of cognitive networks. Prior work showed that these structures of human perceptions and construction organization can influence different cognitive processes, such as early word learning [34], cognitive impairments [35], writing styles [36], individual creativity levels [37] and estimates of curiosity [38]. In education settings, recent studies showed that maps of conceptual associations can be informative of students' performance [29], [33], [39].

Representing human associative knowledge as networks is advantageous over black-box machine learning tools [40]. Firstly, it facilitates generalizability - metrics of network representations, operationalized in ways that reflect human associations or semantic memory, allows to use metrics and null models from network science (see also [41]). For instance, network distance out matches semantic latent analysis in predicting similarity rates [42], whereas network growth models provide evidence for the preferential acquisition hypothesis in word learning, i.e. a tendency for children to acquire first the most semantically prominent concepts in the language they are exposed to [43], [44]. Secondly, when the interpretation of these cognitive structures is consistent, the analysis can power cognitive and psycholinguistic theories through data. For instance, checking which concepts were associated with a target idea provides contextual information about how that idea was semantically framed by a given text. This reconstruction of contextual information from associations is formally described by semantic frame theory [45] and operationalized by cognitive networks, where semantic frames become network neighborhoods or communities of tightly related concepts (see also [46]).

Here, to offer an alternative to approaches used in learning analytics, we describe a specific framework to operationalizing cognitive networks. This framework of *forma mentis networks* (FMN, from *forma mentis*, Latin for "mindset") can capture, reconstruct, and explore perceptions in individuals or groups. FMNs combine artificial intelligence, cognitive psychology and complex systems to explore both explicit/conscious and implicit/subconscious knowledge and emotional perception of individuals or groups of individuals toward a given topic (see [12], [47], [48]). FMNs combine conceptual links and emotional/affective perceptions to offer a scalable approach for accessing the human mind.

3.1. Behavioral forma mentis networks as memory recall patterns

Behavioral forma mentis networks (BFMNs) represent knowledge as a web of concepts interconnected by memory links and rated in terms of sentiment/valence, i.e., "positive", "negative" or "neutral". From a methodological point of view, this means that a link between the concepts is based on free associations [49] for establishing conceptual links: participants are given cue words, for example "bird", to which they might respond with the word "dove". These two concepts, "bird" and "dove" are then associated and linked with each other. Positive or negative valence between the concepts is elicited from the individuals, embedded in the network as an edge attribute. Such tie definitions have strong theoretical roots. Free associations represent conceptual knowledge about the external world as embedded in the so-called semantic memory [50] and are consequently powerful proxies for predicting language learning [34], creativity levels [37] and even personality traits [51]. FMNs rely on such powerful psycholinguistic tools but include an emotional aspect as well, quantified via word valence, e.g., how positively, negatively, or neutrally a given concept is perceived [13].

An example is reported in Figure 1 (reproduced from [47]), which features the FMNs around "mathematics" as reconstructed by 159 high-schoolers and 59 STEM researchers. Importantly, participants provided free associations and valence norms, tagging nodes/concepts as positive, negative or neutral. Concepts tagged as negative and linked mostly with other negative concepts were found to elicit higher levels of anxiety in an external dataset of affective psycholinguistic norms [52]. Hence, words mostly surrounded by negative associates in FMNs were also found to correspond to anxiety-eliciting concepts and were therefore used to identify signals of STEM anxiety [47], test anxiety and implicit negative biases such as stereotype threats [48] present in the students' perception and absent in researchers'.

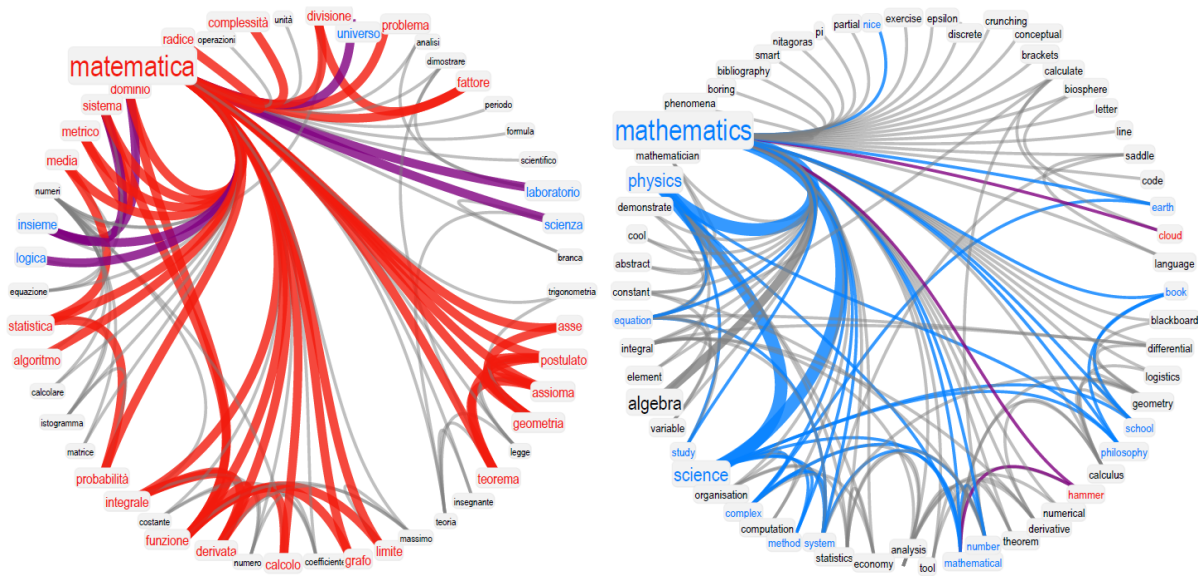


Figure 1: Forma mentis networks for “mathematics” of 159 Italian high school students (left) and 59 international researchers (right). Negative (positive, neutral) concepts are in red (blue, black). Links indicate that a word reminded of another one in the free association task. Connections between negative (positive, different valence) concepts are highlighted in red (blue, purple). Data from Stella et al. (2019).

3.2. Textual forma mentis networks as socially shared texts

Textual forma mentis networks (TFMN) adopt natural language processing to build links between concepts. Language can describe experiences (“I moved to another town”, “He passed away”, etc.) and inner emotional states (“I feel relaxed”, “I feel like I am slowly healing”). Such language can be communicated through text, e.g., in social media posts. In essence, the premise for TFMN is similar to that used in text networks in learning analytics. The difference, however, lies in the way network ties are operationalized in TFMN. TFMNs are sensitive to syntactic and semantic associations of words in language, linking two words that follow each other within the unit of analysis and combining this with the automatically detected sentiment within the concepts to denote valence.

From a methodological perspective, TFMNs can identify syntactic relationships in text with the help of NLP and AI. Stella [53] implemented TFMNs based on sentence parsing from the Stanford NLP universal parser [54], as implemented in Mathematica 11.3 through the TextStructure routine. Powered by a recurrent neural network architecture, the dependency parser is trained to identify grammatical dependencies between words in sentences. The parser scans words in linear time with sentence length and, at every step, it maintains: (i) a stack of words being processed so far in the sentence, and (ii) a buffer of words yet to be processed. By transitioning its inner state, the parser is trained to identify words as either syntactically dependent/independent according to their features and the features of both the stack and the buffer. Sequentially, the parser empties the buffer and captures the dependency structure of words in the stack. The parser is taught to apply the correct transitioning through training data, that is an annotated corpus for the English language [55]. Once the syntactic dependencies are identified, TFMNs can be constructed by connecting non-stop word nodes that share a path of syntactic dependencies lower than a given threshold (that is manually selected by the researcher to regulate the density of syntactic links in TFMNs). This refined, condensed structure, is semantically enriched by further linking synonyms, using a dictionary such as WordNet [56]. Finally, words are endowed with sentiment/valence and emotional norms, e.g. “love” is a positive word or inspires “joy” according to psycholinguistic mega-studies (see [46]). Notice that TFMNs are flexible enough to feature different machine learning outcomes or linguistic models under the hood, a feature that is promising in field-specific linguistic models like FLAIR (for fashion-

related language [57]) or spaCy (for news-related language, see <https://spacy.io/>, accessed 18/05/2022).

Overall, TFMNs follow a nuanced approach to reconstructing the associations between ideas encoded in texts if compared to the networks built on the co-occurrence of ideas. TFMNs capture semantic, syntactic, and emotional structures underpinning the associations of concepts, without the need to interview individuals. Crucially, whereas in BFMNs there can be different affective perceptions for the same word across groups, TFMNs rely on external data for producing the valence labels attributed to words (e.g. the EmoLex dataset [58]). This means that in two different BFMNs, the same word “mathematics” might be perceived as a negative concept in one network and as a positive concept in another one (see also Figure 1). This difference is due to the fact that in the behavioral data behind those networks, individuals rated “mathematics” with lower valence scores in one case compared to the other. In all TFMNs based on the EmoLex dataset, “mathematics” would always be represented as a neutral concept. This is due to a limitation in the way TFMNs are constructed: TFMNs can only access textual data, and so syntactic relationships, but they do not consider meta-data about valence like behavioral forma mentis networks do. As discussed also in [53] and in [59], also TFMNs can portray the same concept along difference valence connotations but only through contextual information: In a given text, “mathematics” could be syntactically linked with mostly negative concepts and thus acquire a negative affective connotation within its own semantic frame/network neighborhood. Consequently, TFMNs put even more emphasis over the importance of going beyond node-level quantifications of valence and reconstruct affective perceptions of concepts by checking how they are interconnected with each other. In other words, whereas BFMNs already present variability at node level (through valence scores), TFMNs should be analyzed always in terms of their network structure and how it relates with the identified valence/emotional labels [59].

4. Discussion

Our reflection presents concerns about the rigor, theoretical grounding, generalizability, and scalability of approaches to text networks in learning analytics. First, prominent approaches used for text networks in learning analytics can improve their use of theory. Text networks based on co-occurrence are not theory grounded. Epistemic networks are also less theory grounded, particularly in the instances when automated detection of codes is in place. Second, the generalizability of text networks is limited, due to varying and convenience-based decisions around the unit of analysis and operationalizations of the networks. Analytical choices in co-occurrence networks enable scalability but result in the presence of the noise and lack of sensitivity to the context within the text, i.e., it is not evident to the experimenter whether a co-occurrence expresses a syntactic, semantic, or phonological association between words.

We presented an approach to text networks rooted in cognitive network science, Forma Mentis Networks (FMNs) to demonstrate how the issues currently present in many instances of text networks can be overcome. FMNs are theory-based and can be applied to data sources that are either elicited from individuals or collected from written text, where network ties can be interpreted as associations or as sequences. FMNs are based on previous research in cognitive networks and allow representing knowledge structures. Yet, they also enable scalability - as automated approaches are used to derive networks [54]. FMNs contain clusters of nodes with similar semantic characteristics that can be compared with the models of mental lexicon. The tight interrelationship between theory, a network representation, and analytical techniques that enable generalizability (comparison to external models of language) offer an example of a methodology that is both scalable and informative to theory and practice. We hope that this critical and reflective argument opens for a discussion about other possibilities of analyzing text networks in the learning and teaching settings.

Acknowledgements

The authors acknowledge the feedback of the workshop participants at NetSciLA22 that was integrated into this submission, as well as thank anonymous reviewers for their feedback.

References

- [1] F. Henri, "Computer conferencing and content analysis," in *Collaborative learning through computer conferencing*, Springer, 1992, pp. 117–136. Accessed: Oct. 12, 2016. [Online]. Available: http://link.springer.com/chapter/10.1007/978-3-642-77684-7_8
- [2] A. De Liddo, S. Buckingham Shum, and I. Quinto, "Discourse-centric learning analytics," 2011.
- [3] D. S. McNamara, L. Allen, S. Crossley, M. Dascalu, and C. A. Perret, "Natural language processing and learning analytics," *Handbook of learning analytics*, vol. 93, 2017.
- [4] L. K. Allen, E. L. Snow, and D. S. McNamara, "Are You Reading My Mind?: Modeling Students' Reading Comprehension Skills with Natural Language Processing Techniques," in *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*, Poughkeepsie, New York, 2015, pp. 246–254. doi: 10.1145/2723576.2723617.
- [5] V. Kovanović *et al.*, "Towards automated content analysis of discussion transcripts: A cognitive presence case," in *Proceedings of the sixth international conference on learning analytics & knowledge*, 2016, pp. 15–24.
- [6] K. Carley, "Extracting culture through textual analysis," *Poetics*, vol. 22, no. 4, pp. 291–312, 1994.
- [7] K. Carley, "Coding choices for textual analysis: A comparison of content analysis and map analysis," *Sociological methodology*, pp. 75–126, 1993.
- [8] K. Carley and M. Palmquist, "Extracting, representing, and analyzing mental models," *Social forces*, vol. 70, no. 3, pp. 601–636, 1992.
- [9] O. Poquet, L. Tupikina, and M. Santolini, "Are Forum Networks Social Networks? A Methodological Perspective," Frankfurt, Germany, 2020. doi: <https://doi.org/10.1145/3375462.3375531sharma>.
- [10] A. F. Wise, Y. Cui, and W. Q. Jin, "Honing in on social learning networks in MOOC forums: examining critical network definition decisions," *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, pp. 383–392, 2017.
- [11] M. Stella, S. De Nigris, A. Aloric, and C. S. Siew, "Forma mentis networks quantify crucial differences in STEM perception between students and experts," *PloS one*, vol. 14, no. 10, p. e0222870, 2019.
- [12] M. Stella, "Forma mentis networks reconstruct how Italian high schoolers and international STEM experts perceive teachers, students, scientists, and school," *Education Sciences*, vol. 10, no. 1, p. 17, 2020.
- [13] J. Aitchison, *Words in the mind: An introduction to the mental lexicon*. John Wiley & Sons, 2012.
- [14] A. F. Wise and Y. Cui, "Top concept networks of professional education reflections," in *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, 2019, pp. 260–264.
- [15] T. Hecking, V. Dimitrova, A. Mitrovic, and U. Ulrich Hoppe, "Using network-text analysis to characterise learner engagement in active video watching," in *ICCE 2017 Main Conference Proceedings*, 2017, pp. 326–335.
- [16] J. Oshima, R. Oshima, and S. Saruwatari, "Analysis of students' ideas and conceptual artifacts in knowledge-building discourse," *British Journal of Educational Technology*, vol. 51, no. 4, pp. 1308–1321, 2020.
- [17] A. Wise, S. Reza, and R. Han, "Becoming a dentist: Tracing professional identity development through mixed-methods data mining of student reflections," in *14th International Conference of the Learning Sciences: The Interdisciplinarity of the Learning Sciences, ICLS 2020*, 2020, pp. 294–301.
- [18] S. Joksimović, V. Kovanović, J. Jovanović, A. Zouaq, D. Gašević, and M. Hatala, "What do cMOOC participants talk about in social media? A topic analysis of discourse in a cMOOC," in *Proceedings of the fifth international conference on learning analytics and knowledge*, 2015, pp. 156–165.
- [19] N. Van Labeke, D. Whitelock, D. Field, S. Pulman, and J. T. Richardson, "OpenEssayist: extractive summarisation and formative assessment of free-text essays," 2013.

- [20] D. Whitelock, A. Twiner, J. T. Richardson, D. Field, and S. Pulman, "What does a 'good' essay look like? Rainbow diagrams representing essay quality," in *International Conference on Technology Enhanced Assessment*, 2017, pp. 1–12.
- [21] E. Yun and Y. Park, "Extraction of scientific semantic networks from science textbooks and comparison with science teachers' spoken language by text network analysis," *International Journal of Science Education*, vol. 40, no. 17, pp. 2118–2136, 2018.
- [22] S. S. Fougat, A. Siebert-Evenstone, B. Eagan, S. Tabatabai, and M. Misfeldt, "Epistemic network analysis of students' longer written assignments as formative/summative evaluation," in *Proceedings of the 8th international conference on learning analytics and knowledge*, 2018, pp. 126–130.
- [23] R. Ferreira, V. Kovanović, D. Gašević, and V. Rolim, "Towards combined network and text analytics of student discourse in online discussions," in *International conference on artificial intelligence in education*, 2018, pp. 111–126.
- [24] M. Scardamalia and C. Bereiter, "Computer Support for Knowledge Building Communities," in *CSSL: Theory and Practice of an Emerging Paradigm*, T. Koschmann, Ed. Mahwah, New Jersey: Lawrence Erlbaum Associates Inc. Publishers, 1996.
- [25] Y. Taskin, T. Hecking, and H. U. Hoppe, "ESA-T2N: a novel approach to network-text analysis," in *International conference on complex networks and their applications*, 2019, pp. 129–139.
- [26] T. Hecking, I. A. Chounta, and H. U. Hoppe, "Role modelling in MOOC discussion forums," *Journal of Learning Analytics*, vol. 4, no. 1, pp. 85–116, 2017.
- [27] M. Dascalu, S. Trausan-Matu, P. Dessus, and D. S. McNamara, "Discourse Cohesion: A Signature of Collaboration," in *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*, Poughkeepsie, New York, 2015, pp. 350–354. doi: 10.1145/2723576.2723578.
- [28] U. Hoppe, "Computational Methods for the Analysis of Learning and Knowledge Building Communities," in *Handbook of Learning Analytics*, First., C. Lang, G. Siemens, A. Wise, and D. Gasevic, Eds. Society for Learning Analytics Research (SoLAR), 2017, pp. 23–33.
- [29] I. Koponen and M. Nousiainen, "Koponen, I. T., & Nousiainen, M. (2018). Concept networks of students' knowledge of relationships between physics concepts: finding key concepts and their epistemic support.," *Applied network science*, vol. 3, no. 1, pp. 1–21, 2018.
- [30] A. Ninio, "Syntactic networks, do they contribute valid information on syntactic development in children?. Comment on" Approaching human language with complex networks" by J. Cong and H. Liu," *Physics of life reviews*, vol. 11, no. 4, pp. 632–634, 2014.
- [31] C. Siew, "Investigating cognitive network models of learners' knowledge representations," *Journal of Learning Analytics*, vol. 9, no. 1, pp. 120–129, 2022.
- [32] C. S. Siew, D. U. Wulff, N. M. Beckage, and Y. N. Kenett, "Cognitive network science: A review of research on cognition through the lens of network representations, processes, and dynamics," *Complexity*, vol. 2019, 2019.
- [33] C. S. Siew, "Using network science to analyze concept maps of psychology undergraduates," *Applied Cognitive Psychology*, vol. 33, no. 4, pp. 662–668, 2019.
- [34] T. T. Hills, M. Maouene, J. Maouene, A. Sheya, and L. Smith, "Longitudinal analysis of early semantic networks: Preferential attachment or preferential acquisition?," *Psychological science*, vol. 20, no. 6, pp. 729–739, 2009.
- [35] N. Castro and M. Stella, "The multiplex structure of the mental lexicon influences picture naming in people with aphasia," *Journal of Complex Networks*, vol. 7, no. 6, pp. 913–931, 2019.
- [36] D. R. Amancio, "A complex network approach to stylometry," *PloS one*, vol. 10, no. 8, p. e0136076, 2015.
- [37] Y. N. Kenett, D. Anaki, and M. Faust, "Investigating the structure of semantic networks in low and high creative persons," *Frontiers in human neuroscience*, vol. 8, p. 407, 2014.
- [38] D. M. Lydon-Staley, D. Zhou, A. S. Blevins, P. Zurn, and D. S. Bassett, "Hunters, busybodies and the knowledge network building associated with deprivation curiosity," *Nature human behaviour*, vol. 5, no. 3, pp. 327–336, 2021.
- [39] I. T. Koponen and M. Pehkonen, "Coherent knowledge structures of physics represented as concept networks in teacher education," *Science & Education*, vol. 19, no. 3, pp. 259–282, 2010.

- [40] C. Rudin, “Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead,” *Nature Machine Intelligence*, vol. 1, no. 5, pp. 206–215, 2019.
- [41] A. A. Kumar, “Semantic memory: A review of methods, models, and current challenges,” *Psychonomic Bulletin & Review*, vol. 28, no. 1, pp. 40–80, 2021.
- [42] Y. N. Kenett, E. Levi, D. Anaki, and M. Faust, “The semantic distance task: Quantifying semantic distance with semantic network path length,” *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 43, no. 9, p. 1470, 2017.
- [43] T. T. Hills, M. Maouene, J. Maouene, A. Sheya, and L. Smith, “Longitudinal analysis of early semantic networks: Preferential attachment or preferential acquisition?,” *Psychological science*, vol. 20, no. 6, pp. 729–739, 2009.
- [44] E. A. Karuza, “The value of statistical learning to cognitive network science,” *Topics in Cognitive Science*, vol. 14, no. 1, pp. 78–92, 2022.
- [45] C. J. Fillmore and C. F. Baker, “Frame semantics for text understanding,” in *Proceedings of WordNet and Other Lexical Resources Workshop, NAACL*, 2001, vol. 6.
- [46] A. Semeraro, S. Vilella, G. Ruffo, and M. Stella, “Writing about COVID-19 vaccines: Emotional profiling unravels how mainstream and alternative press framed AstraZeneca, Pfizer and vaccination campaigns,” *arXiv preprint arXiv:2201.07538*, 2022.
- [47] M. Stella, S. De Nigris, A. Aloric, and C. S. Siew, “Forma mentis networks quantify crucial differences in STEM perception between students and experts,” *PloS one*, vol. 14, no. 10, p. e0222870, 2019.
- [48] M. Stella and A. Zaytseva, “Forma mentis networks map how nursing and engineering students enhance their mindsets about innovation and health during professional growth,” *PeerJ Computer Science*, vol. 6, p. e255, 2020.
- [49] S. De Deyne, D. J. Navarro, and G. Storms, “Better explanations of lexical and semantic cognition using networks derived from continued rather than single-word associations,” *Behavior research methods*, vol. 45, no. 2, pp. 480–498, 2013.
- [50] M. Steyvers and J. B. Tenenbaum, “The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth,” *Cognitive science*, vol. 29, no. 1, pp. 41–78, 2005.
- [51] A. P. Christensen, Y. N. Kenett, K. N. Cotter, R. E. Beaty, and P. J. Silvia, “Remotely close associations: Openness to experience and semantic memory structure,” *European Journal of Personality*, vol. 32, no. 4, pp. 480–492, 2018.
- [52] M. Montefinese, E. Ambrosini, B. Fairfield, and N. Mammarella, “The adaptation of the affective norms for English words (ANEW) for Italian,” *Behavior research methods*, vol. 46, no. 3, pp. 887–903, 2014.
- [53] M. Stella, “Text-mining forma mentis networks reconstruct public perception of the STEM gender gap in social media,” *PeerJ Computer Science*, vol. 6, p. e295, 2020.
- [54] D. Chen and C. D. Manning, “A fast and accurate dependency parser using neural networks,” in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 740–750.
- [55] N. Silveira *et al.*, “A gold standard dependency corpus for English,” in *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC’14)*, 2014, pp. 2897–2904.
- [56] G. A. Miller, “WordNet: a lexical database for English,” *Communications of the ACM*, vol. 38, no. 11, pp. 39–41, 1995.
- [57] A. Akbik, T. Bergmann, D. Blythe, K. Rasul, S. Schweter, and R. Vollgraf, “FLAIR: An easy-to-use framework for state-of-the-art NLP,” in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)*, 2019, pp. 54–59.
- [58] S. M. Mohammad and P. D. Turney, “Crowdsourcing a word–emotion association lexicon,” *Computational intelligence*, vol. 29, no. 3, pp. 436–465, 2013.
- [59] M. Stella, M. S. Vitevitch, and F. Botta, “Cognitive Networks Extract Insights on COVID-19 Vaccines from English and Italian Popular Tweets: Anticipation, Logistics, Conspiracy and Loss of Trust,” *Big Data and Cognitive Computing*, vol. 6, no. 2, p. 52, 2022.