AQ1
AQ2

Modeling the Semantic Structure of Textually Derived Learning Content and its Impact on Recipients' Response States

David Munoz
Industrial Engineering,
The Pennsylvania State University,
University Park, PA 16802
e-mail: dam395@psu.edu

Conrad S. Tucker
Engineering Design and Industrial Engineering,
The Pennsylvania State University,
University Park, PA 16802
e-mail: ctucker4@psu.edu

In the United States, the greatest decline in the number of students in the STEM education pipeline occurs at the university level, where students, who were initially interested in STEM fields, drop-out or move on to other interests. It has been reported that "of the 23 most commonly cited reasons for switching out of STEM, all but 7 had something to do with the pedagogical experience." Thus, understanding the characteristics of the pedagogical experience that impact students' interest in STEM is of great importance to the academic community. This work tests the hypothesis that there exists a correlation between the semantic structure of lecture content and students' affective states. Knowledge gained from testing this hypothesis will inform educators of the specific semantic structure of lecture content that enhance students' affective states and interest in course content, toward the goal of increasing STEM retention rates and overall positive experiences in STEM majors. A case study involving a series of science and engineering based digital content is used to create a semantic network and demonstrate the implications of the methodology. The results reveal that affective states such as engagement and boredom are consistently strongly correlated to the semantic network metrics outlined in the paper, while the affective state of confusion is weakly correlated with the same semantic network metrics. The results reveal semantic network relationships that are generalizable across the different textually derived information sources explored. These semantic network relationships can be explored by researchers trying to optimize their message structure in order to have its intended effect. [DOI: 10.1115/1.4032398]

Keywords: semantic network, emotional states, engineering design, text mining, Kansei engineering, design for emotion

1 Introduction

Currently, there exists a knowledge gap in terms of how positive or negative affective experiences during classroom instruction, impact students' interest in STEM-related majors and careers. In the United States, the greatest decline in the number of students in the STEM education pipeline occurs at the university level, where students, that were initially interested in STEM fields, drop-out or move on to other interests [1]. Notably, "of the 23 most commonly cited reasons for switching out of STEM, all but 7 had something to do with the pedagogical experience" [2]. A significant challenge facing today's educators is creating lecture content that enhances students' attention and interest during lectures. Lecture content should be structured in a manner that offers enhanced learning experiences for students, while also enhancing students' positive emotional states. In this work, the terms *emotion* and *affective* are used interchangeably to encompass any internal state that influences cognitive and behavioral processes. Several studies have demonstrated the close relationship between emotions expressed by students in a classroom and their course performance [3,4]. Quantifying students' emotional states, such as boredom, frustration, and engagement, in relation to the lecture material being presented, will enable researchers to discover novel, previously unknown correlations that exist between lecture content and students' affective states. Instructors' teaching methods/styles and pedagogical tools could then be modeled and compared in terms of their ability to minimize negative emotional states and maximize positive emotional states during classroom instruction.

This work tests the hypothesis that there exists a correlation between the semantic structure of lecture content and students' affective states. While nonverbal communication such as body language, intonation, and facial expressions are relevant dimensions of expressing emotion, the analysis of the structure of a message attempts to quantify the verbal dimension of communication. This work is limited to the context of information dissemination in an educational learning context. In order to analyze these patterns, the authors employ semantic network measures to characterize the lecture content that is being transmitted. In addition, a self-reported attitudinal survey is employed to quantify emotional state intensities. Based on this information, correlation and regression analyses are conducted to identify interesting patterns and relate emotional states based to semantic network measures.

Quantifying the relationship between the content of a message and the emotional states expressed by recipients of that message will inform both students and educators in STEM of the importance of communication in enhancing learning and decision making; concepts that are of great importance in engineering education and STEM. In the engineering design community, researchers have explored the impact that designs have on eliciting certain human emotions [5,6]. For example, *Kansei engineering* seeks to enhance products and services by translating customers' emotions and feelings about a product's design into tangible design parameters [7,8]. In the context of learning that this work explores, the *product* is analogous to the knowledge gained by the recipient and the *customer* is analogous to the recipient of that knowledge (i.e., in this case, a student).

This paper is organized as follows. In the current section, the authors provide an introduction and motivation into the

Contributed by the Design Education Committee of ASME for publication in the Journal of Mechanical Design. Manuscript received June 18, 2015; final manuscript received November 20, 2015; published online xx xx, xxxx. Assoc. Editor: Andy Dong.

knowledge gap that exists. Section 2 contains a literature review of fields relevant to this research. Section 2 includes a brief review of the emotional states present in the learning process and their role as a key communication channel, the association between emotional states and learning outcomes, and methods to assess individuals' mental states. The methodology is presented in Sec. 3. In Sec. 4, the authors introduce a case study, followed by Sec. 5 that provides a detailed explanation of the main results obtained. Section 6 discusses future research directions and concludes the paper.

2 Literature Review

2.1 Emotional States and Learning Outcomes. There are four main types of evidence about emotions: language, self-reports, behavior, and physiology. However, behavior and physiology evidence concern the consequences of emotional states, instead of its origin. For this reason, some studies of emotional states have been focused mostly on language and self-report evidence [9]. It has been proven that emotions are relevant to the learning process [10]. Some of them such as engagement and interest will positively impact and enhance learning. Furthermore, according to Gal and Ginsburg [11], noncognitive factors such as negative attitude, beliefs, feelings, interest, and motivations could influence individuals' ability to develop understanding. Craig et al. [12] found that there is a significant correlation between emotional states such as boredom, confusion, flow (mental state in which a person is fully immersed and involved in an activity), and learning gains. Table 1 summarizes the second order emotional states and their impact on learning gains.

2.2 Text Data Mining and Semantic Exploration in Engineering Design. Data mining of textual data is an emerging area of research in the design community. For example, Dong proposes a latent semantic approach to studying design team communication in an effort to understand how designers construct knowledge pertaining to a design artifact [19]. Dong et al. propose a latent semantic approach that measures the quality of the design performance using textual descriptions of related design concepts and events [20]. An ontology-based design system is proposed by Li et al. in order to increase the efficiency of information extraction and retrieval in engineering design [21]. Ghani et al. employ an attribute value pair approach to mine product features from unstructured textual data [22]. Liang and Tan employ text mining techniques to analyze product patents in search of product innovations [23]. Kang et al. propose a text mining-driven methodology to search for similarities in End of Life products and components through a process called product resynthesis [24]. In bio-inspired design, Glier et al. employ automated text classification techniques to improve the keyword corpus search results [25]. Fu et al. propose a distance measure, based on latent semantic analysis (LSA) and Bayesian-based models for discovering the structural form of products [26]. Stone and Choi propose machine learning classification models to extract customer preferences from online user generated content [27]. Ren and Papalambros propose a method of eliciting design preferences using crowd implicit feedback [28]. A text mining approach for identifying key product attributes and their importance levels has been proposed by Rai [29]. Tuarob and Tucker propose a latent Dirichlet allocation (LDA) based methodology for mining social network data in an effort to predict emerging product trends [30]. Tucker et al. have proposed text mining models for quantifying students' sentiments in massively open online courses (MOOCs) [31].

While existing text-mining driven techniques have been proposed to solve a wide range of engineering related problems, a fundamental understanding of the correlations that exist between the semantic structure of textual content and individuals' affective states remains an open research area. This work aims to advance the scientific body of knowledge centered on textual data, as it pertains to content formulation, delivery, and reception. While the authors have employed both spatial-based methods such as LSA and probabilistic-based methods such as LDA to solve a wide range of engineering design problems [24,30,32], the proposed semantic network approach to quantifying word associations is better suited for this research because: (i) semantic associations between words can be easily visualized from the semantic network, which will help instructors understand how to optimize their content structure in order to increase desired emotional states exhibited by students and (ii) spatial-based methods such as LSA may violate metric axioms such as (i) symmetry and (ii) triangle inequality [33], which are important characteristics in trying to understand how messages impact receivers' emotional states.

Table 1 Emotional states and their impact on learning gains

Emotional state	Learning gains impact	References
Engagement/interest	Positive	[13] and [14]
Frustration	Negative	[12] and [15]
Boredom	Negative	[16]
Confusion	Positive	[12], [15], and [17]
Delight	Positive	[18]

3 Methodology

3.1 Methodology Overview. This work tests the hypothesis that there exists a correlation between the semantic structure of lecture content and students' affective states. Figure 1 presents a three-phase approach to testing this hypothesis that includes: (i) developing a semantic network of the information (i.e., lecture content) being disseminated, (ii) quantifying emotional states of the receivers (i.e., students), as a response to the information being transmitted, and (iii) identifying interesting patterns between the semantic network and the receivers' emotional states. The main aim of the first phase is to characterize a message (i.e., lecture content) using a set of semantic network metrics. As a reminder, the assumption made in this work is that a message can be automatically transformed into textual data (e.g., speech to text, typing to text, etc.). In this sense, nonverbal communication such as body language, facial expression, gestures, and voice intonation, among others, are not considered in the scope of this methodology. A codification protocol could be included to account for nonverbal communication [34], and therefore, have a more comprehensive framework to scope the multidimensionality nature of communication. The second phase quantifies students' feedback in terms of the impact of the lecture content on their emotional states through a self-reported attitudinal survey. Finally, in the third phase, interesting patterns between the semantic network characteristics of a message and students' emotional states are explored and quantified. In this work, it is assumed that in order for communication to be effective, the encoding and decoding processes are aligned in the same language and using a familiar communication channel.

In this work, the emotional state intensities are in part, a function of different semantic network measures of the message itself. Thus, the intensity of emotional state i can be expressed as

$$E_i = f(O, U, V) \tag{1}$$

where O is a set of network metrics characterizing the entire semantic network, U is a set of cluster-related semantic network metrics, V is a set of vertex-related semantic network metrics.

The detailed definition of each of these metrics is presented in the subsequent sections. AQ3

3.2 Creating a Semantic Network of Information

3.2.1 Defining the Set of Words in a Semantic Network. Semantic networks are a representation of the semantic relationship between concepts of language at different levels that include word, phrases, sentences, paragraphs, and other language units [35]. Typically, semantic networks are used to represent knowledge graphically based on the patterns of interconnected nodes (words) and arcs (relationship between words). In order to generate the semantic network, the set of words to be used in the textual analysis needs to be defined.

The first step in phase 1 is to characterize the content of a lecture in terms of the words that it is comprised of. This represents the main input needed to create the adjacency matrix. The adjacency matrix is a matrix representation of a graph that is used to create the semantic network graph of a given lecture. Given the transcripts or textual representation of a lecture, the set W is a set of N sequentially ordered words represented by

$$W: \{w_1, w_2, w_3, \ldots, w_N\}$$

Additionally, C is defined as a set containing M common words that could be omitted from the textual analysis as

$$C: \{c_1, c_2, c_3, \ldots, c_M\}$$

For example, the set of words C could be the 250, 500, or 1000 most used words in a given language. This set is used as a way of classifying those words that are commonly used in a given language, as these words (e.g., the, and, etc.) will not add much value to the understanding of the message or topic. Common connectors (words) such as prepositions, conjunctions, pronouns, and common verbs can be omitted from the message, depending on the application of the textual analysis [36]. Finally, a set T is generated that contains the "topic" words. T is defined as the set of L words that are *meaningful* for defining the topic under consideration. Therefore, T is a subset of W that contains the elements of W, except those elements also included in C. Then:

$$T: \{t_1, t_2, t_3, \ldots, t_L\}$$

where $T \subseteq W$.

The number of elements in these sets, also referred to as the *size* or *order*, is given by $|W|$, $|C|$, and $|T|$, respectively. For example, let us assume that we have the following sentence from a lecture: "A fundamental attribute of the engineering design process is information exchange." Then, using the guide for set generation provided, the resultant sets are

W:{a, fundamental, attribute, of, the, engineering, design, process, is, information, exchange}

Let us also assume that the words *a, of, the,* and *is* are some elements of the set C.

C:{a, of, the, is,…}

Thus, the set of topic words is defined as

T:{fundamental, attribute, engineering, design, process, information, exchange}

By removing the set C from the textual data, textual noise is reduced. Therefore, a cleaner set of words, T, is obtained for its use in generating the adjacency matrix.

3.2.2 Generating the Adjacency Matrix. The sets described above are used to generate a co-occurrence matrix among words, also called adjacency matrix in network analysis [37]. This matrix contains the frequency in which two words appear sequentially in a given transcript or textual data. Their sequential appearance will give a quantifiable indication of the relationship between two words. In this study, two words are said to be close or related if each of them appear in the set T within a given window size of Z elements, where Z should be selected such that $Z < |T|$ for nontrivial cases. It must be noted that the larger the value of Z, the more non-null values in the adjacency matrix. Therefore, the semantic network becomes denser. Consequently, as Z approaches T, the number of null values in the adjacency matrix approaches zero. The concept of density is explained in Sec. 3.2.3.1.

In order to generate the adjacency matrix, a new set of words (T^*) must be defined. This new set is an unordered subset of T ($T^* \subseteq T$) including only unique elements of T, i.e., $t_1^* \neq t_2^* \neq t_3^* \neq \ldots \neq t_k^*$. Therefore, T^* can be written as

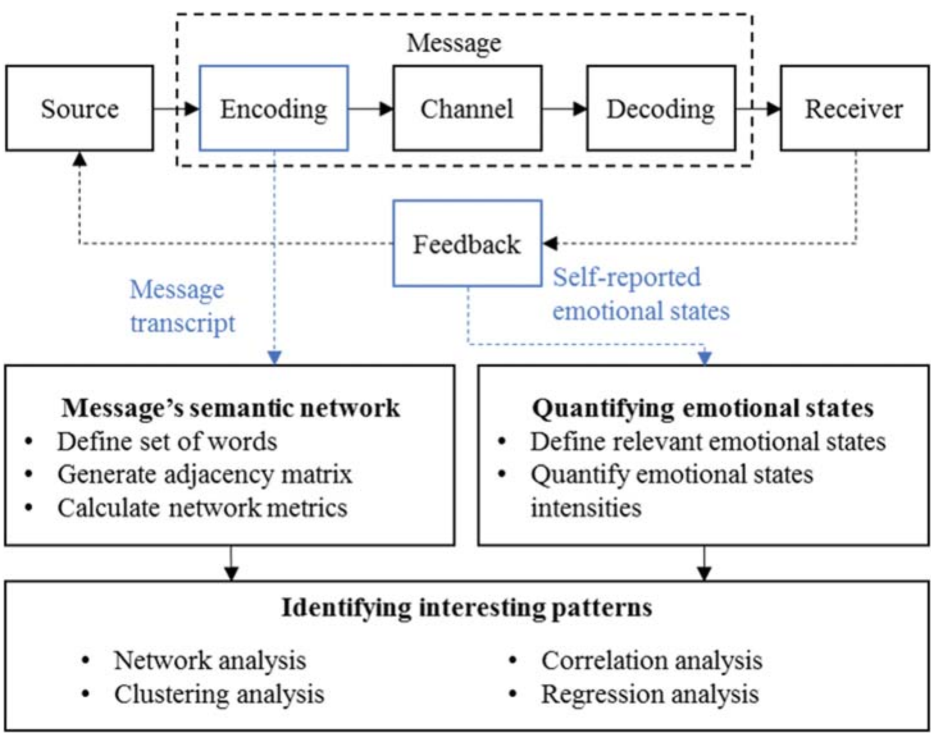

Fig. 1 Methodology for quantifying the correlation between the semantic structure of lecture content and students' affective states

$$T^* : \{t_1^*, t_2^*, t_3^*, \ldots, t_k^*\}$$

The adjacency matrix A is represented as

$$A = \begin{array}{c} \\ t_{1*} \\ t_{2*} \\ t_{3*} \\ \vdots \\ t_{k*} \end{array} \begin{array}{c} \begin{array}{cccccc} t_{1*} & t_{2*} & t_{3*} & \ldots & t_{k*} \end{array} \\ \begin{bmatrix} - & x_{12} & x_{13} & \ldots & x_{1k} \\ x_{21} & - & x_{23} & \ldots & x_{2k} \\ x_{31} & x_{32} & - & \ldots & x_{3k} \\ \vdots & \vdots & \vdots & - & \vdots \\ x_{k1} & x_{k2} & x_{k3} & \ldots & - \end{bmatrix} \end{array}$$

where x_{ij} represents the frequency or number of times in which both words i and j appear in a windows size Z. For undirected graphs, a triangular symmetric matrix is obtained, i.e., $x_{ij} = x_{ji}$. Therefore, the number of cells that must be calculated to complete the adjacency matrix is: $|T^*|*(|T^*| - 1)/2$. Similar approaches to generating an adjacency matrix for semantic networks, based on windows for assessing word co-occurrence of words, have been previously validated in the literature [36].

The adjacency matrix constructed provides a matrix representation of the lecture's semantic network. Therefore, the set of words (nodes) and their relationships (edges) are the input for creating the semantic network graph.

3.2.3 Network Analysis and Metrics. In this section, the main network measures that characterize the message are defined. Let us define a semantic graph $G: (T^*, E)$, where T^* is a set of nodes (i.e., unrepeated topic words) and E is a set of edges representing the relationship between two consecutive nodes. In this case, E contains unordered pairs of words extracted from the adjacency matrix A, specifically from non-null cells.

In order to characterize the semantic network of a message, various network metrics are defined that comprise of the feature set of the semantic network itself, consistent with the literature [38–40]:

(1) overall network-related metrics,
(2) cluster-related metrics
(3) vertex-related metrics.

Network metrics such as *density* and *geodesic distance* can be calculated for the overall network or its clusters. On the other hand, the most used vertex-related metrics that can be calculated are the *degree centrality*, *betweenness centrality*, and *eigenvector centrality* [40].

3.2.3.1 Density. The *density* of a network represents the proportion of existing edges out of the potential edges within the network. This metric can be calculated for the entire network or parts of it, also called subnetworks or clusters. The maximum number of edges in an undirected semantic network is given by $|T^*|*(|T^*| - 1)/2$. The density of the network can be defined as

$$\text{Density} = \frac{2|E|}{|T^*|(|T^*| - 1)} \tag{2}$$

Networks with density equal to one are called complete networks. In practice, complete semantic networks are not common, i.e., there is little semantically meaningful knowledge in a graph if every word is connected to every other word. This metric becomes relevant to understanding how connected the words of the message are in the network or clusters.

3.2.3.2 Geodesic distance. The *geodesic distance* is defined as the shortest path or route between two nodes. In nonweighted edge networks such as the case presented in this work, the geodesic distance between two nodes is the minimum number of edges connecting them. This metric indicates how reachable a particular node is for the other nodes. Typically, this metric is used to evaluate the cohesion of a network. In order to characterize networks or clusters, the *maximum* and *average geodesic distances* are used. In semantic networks, the geodesic distance indicates the level of reachability between the words of the network or cluster. This becomes especially useful in evaluating how well or "close" the clusters (subtopics) or ideas of a message are developed. More specifically, large geodesic distances will indicate that the words or subtopics are far apart or not closely related.

3.2.3.3 Degree centrality. While the density and geodesic distance metrics are related to the whole network or cluster, the *degree centrality* is a vertex-related network metric. In an undirected network, it measures the number of direct connections of a particular node to other nodes in the network. Consequently, the *degree centrality* can be used as an indicator of the importance of a node. In directed graphs, this metric is separated into two; *indegree* and *outdegree* centrality which represent the number of edges toward or from a node, respectively. In this work, semantic networks are treated as undirected, and hence, these last two metrics are not considered. The degree centrality of a node v is usually written as $C_d(v) = \deg(v)$. In semantic networks, this metric is used to identify the main topic words of a message. For example, words that have five direct connections to other words are said to have a degree centrality of five. Consequently, those words with comparatively larger degree centrality can be interpreted as the main topic words, as they are central for the topic or textual data.

3.2.3.4 Betweenness centrality. The betweenness centrality quantifies the number of times that a node serves as a bridge along the shortest path between other pairs of words within the network. The betweenness centrality of a node v is expressed as

$$C_b(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{3}$$

where σ_{st} is the total number of shortest paths between word s and word t, and $\sigma_{st}(v)$ is the number of those shortest paths that pass through word v. In semantic networks, this metric is relevant in identifying words that bridge subtopics. Consequently, the words with a comparatively high betweenness centrality are the connecting words among the other words of a message.

3.2.3.5 Eigenvector centrality. Another relevant centrality measure of a node is the eigenvector centrality [41]. This metric is typically used to quantify the influence of a given node in a network. Thus, the words with comparatively high eigenvector centrality are said to be accessible by other well connected nodes and have a larger influence on the message's network. Those nodes with a high eigenvector centrality are well connected to other nodes, which are also well connected. The eigenvector is mathematically defined as the principal eigenvector of the adjacency matrix A. Hence, the defining equation of an eigenvector is $Av = \lambda v$, where λ is the eigenvalue and v is the eigenvector.

As outlined in Fig. 1, the above metrics can be considered the candidate features of the semantic network. By exploring correlations with emotional states, the relevant semantic network features will be discovered.

3.3 Quantifying Students' Affective States. This work postulates that students' affective state intensities are, in part, a function of the message's semantic network structure. In order to capture these emotional states, techniques such as observation or self-reported attitudinal surveys can be used. According to Kort et al., expert communicators (i.e., instructors) are proficient at recognizing and addressing the emotional states of the receivers of information (i.e., students) [15]. Based on observation, communicators can take actions to positively impact the learning experience. However, some barriers including the experience of the instructor, size of the audience, and cultural barriers can impact the detectability of students' emotional states through visual observation alone. In order to overcome the described limitations, self-emotional attitudinal surveys can be used to quantify receiver's emotional states, and hence provide feedback to disseminators of learning content so that they can update their course material in such a way that students' positive emotional states are improved. The main objective of such self-reported instruments is to capture data directly from the recipient of that information and minimize observer's bias. However, studies have criticized its use due to risk of reporter's bias [16]. Fernandez-Ballesteros presents a series of tips to avoid inaccurate information from self-reported questionnaires [42]. Anonymity, which has been suggested in the literature as a method of minimizing bias in self-reports, has been employed by the authors of this work.

The set of emotional states included in the survey depends on the nature of the information exchange that is being considered. For example, in the classroom setting, evidence has shown that second-order emotions such as engagement, interest, delight, boredom, frustration, and confusion are more relevant to the learning experience. In order to quantify the emotional state intensities, a Likert scale is recommended. In the study presented in Sec. 4, a survey, including the six emotional states mentioned, is filled out by the receiver right after the message is transmitted.

3.4 Quantifying Interesting Patterns of a Semantic Network. In order to explore the relationship between the semantic structure of lecture content and students' corresponding affective states, correlation and regression analyses are investigated. The network parameters will be derived from the three main network metric groups; overall graph metrics, clustering metrics, and vertex-related metrics. The first group includes metrics related to the whole network such as number of vertices (words), number of edges, geodesic distance, density, and modularity. The second group includes metrics that are related to the cluster. These clusters can be obtained by employing traditional data mining clustering algorithms [43]. The clustering-related metrics included in this group are similar to those in the overall graph-related metrics group but applied to clusters (subnetworks). For the vertex-related, the most used and representative metrics are included; degree centrality, betweenness centrality, closeness centrality, and eigenvector centrality [40]. From all the metrics stated, relevant parameters are extracted based on descriptive statistics. Those include minimum, mean, maximum, and standard deviation. A summary of the specific parameters proposed is presented in Table 2.

3.4.1 Correlation Analysis. The validation step of this methodology evaluates whether there exists a relationship between two or more parameters through correlation analysis. Typically, this analysis includes the use of the correlation coefficient (r), also known as the Pearson product-moment correlation coefficient. This coefficient measures the linear relationship between two variables. The values of r range from -1 to $+1$. A correlation coefficient of -1 represents a perfectly negative relationship between the two variables. On the other hand, a correlation coefficient of $+1$ represents a perfectly positive relationship between the two variables. A correlation value of 0 indicates that there is no relationship between the variables. Intermediate values can be interpreted using the Salkind scale [44]. The correlation coefficient for a sample can be calculated as follows:

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \qquad (4)$$

where n : samplesize, X_i : value of ith observation from sample X, $i : 1$ to n,
$\bar{X}$: average value of all observations from sample X,
Y_i : value of ith observation from sample Y, $i : 1$ to n, and
$\bar{Y}$: average value of all observations from sample Y.

Paired samples X and Y are represented by semantic network metrics and emotional state intensities, respectively. Therefore, the correlation analysis is focused on calculating whether the messages' semantic network metrics are linearly correlated to students' emotional states. This analysis serves as the initial baseline to identify whether interesting patterns can be found to explain the variability of emotional state intensities based on the message's semantic network metrics. It should be noted that even though this analysis explores linear relationships, other nonlinear analyses can also be conducted. However, there is no previous work stating that a nonlinear analysis is better in this case. Future work will explore more complicated relationships.

3.4.2 Regression Analysis. Regression analysis involves the identification of the relationship between a dependent variable and a set of independent variables. In this case, the message's (i.e., lecture content) semantic network metrics are the independent variables and the receiver's (i.e., student) emotional states are the dependent variables. First, the significant semantic network metrics are identified; thus, the set of relevant parameters is reduced. Second, regression models could be used to estimate what would be the value of the different emotional states, based on the message's semantic network structure. Finally, the signs of the significant parameters can be interpreted for each one of the semantic network metrics. Therefore, insights can be gained about the positive or negative impact of those metrics on the receiver's emotional states. This can guide the design of messages based on the evidence gained from the regression analysis.

Linear regression is proposed in this work, as a first step approximation of the relationships between the semantic structure of messages and individuals' emotional states. More complex, nonlinear relationships could exists, which would be inferred, based on the performance of the linear regression models. In the linear regression model, the model assumes that the dependent variable is a linear combination of the independent variables. For example, intensity of the emotional state i can be expressed as a function of P semantic network parameters

$$E_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_P x_P + \varepsilon_i \qquad (5)$$

where β_0 is the intercept or constant, β_i is the slope or contribution of the semantic network parameter x_i, and ε_i is the error term.

4 Application

4.1 Participants and Lecture Selection. Participants from diverse fields volunteered to attend a lecture composed of five short video-lectures (lessons). The experiments conducted included the participation of 22 students from different fields such as engineering, psychology, finance, accounting, biology, and education. Out of the 22 participants, 18 of them were pursuing bachelor degrees and 4 of them pursuing graduate degrees. Their ages range from 18 to 35, with a mean age of 22. The experiment included 3 females and 19 males. Participants were randomly partitioned into separate groups. Each of the groups was asked to attend a video-lecture of about 30 min. In order to reduce external sources of bias, the experiment took place in a regular classroom of standard size where noise and disturbing factors were minimized. In Fig. 2, an illustration of the participants' learning environment is shown.

Ten video-lectures were selected from YouTube assuring variety of topics and complexity of content. Two lists of five videos each were generated. The 22 participants were randomly partitioned into two groups of 11. Each one of these groups watched a list of five videos. Therefore, a total of 110 data points pertaining to source (i.e., lecture content) and receiver (i.e., student) are collected. The 110 data points generated is sufficient for our study, given an anticipated effect size (f^2) of 1 (assuming $r^2 = 0.5$), statistical power of 0.9, the 37 network parameters tested for inclusion in the regression, and a p-value of 0.05. Under these parameters, the minimum number of data points suggested is 65 [45]. The authors acknowledge that correlations may exist based on the 110 data points generated, given that only 22 participants were used. However, it is possible for an individual to express mutually exclusive emotions, given their emotional regulation strategies of reappraisal and suppression [46]. Future work will explore a larger participant sample size and its impact on subsequent regression models. For the subsequent analysis, the lists of video-lectures were named List A and List B. The source of videos was retrieved from the "Big Think" channel. The selected videos have a duration between 5 to 7 min. The main reason for selecting videos from "Big think" sources is to ensure that speech tools are controlled, as this channel provides a relatively standardized speech-delivery format. Videos from a range of topics were selected to evoke different stimuli from the participants, and therefore different emotional states as a response. While other mechanisms for delivering the lecture could have been used, e-learning technologies, including videos, have been found to be increasingly used for knowledge dissemination [47]. In addition, the textual representation of each video (captions) was extracted directly from the YouTube platform. The list of videos is presented in Table 3.

Participants were asked to fill-out the background form at the beginning of the experimental session. Additionally, during each experimental session, at the end of each short video, a self-reported emotional state survey was given to the participants. A Likert scale (1: Strongly disagree to 5: Strongly agree) was used to report the intensity of the emotional states faced during the lecture by completing the following statement: *I felt [engaged, bored, etc.] during the video-lecture.*

In summary, from the communication diagram presented in Fig. 1, it can be said that the *source* is the lecturer and the *receivers* are the individual participants from different fields. The message is *encoded* verbally, transmitted through a video-recorded *channel*, the participants *decode* the message based mostly on the information presented, previous understanding, and academic knowledge of the topic. Finally, the *feedback* is captured through a self-reported questionnaire based on participants' emotional states as a response to the lecture.

Table 2 Overall clustering and vertex-related parameters

Overall graph metrics	
$Graph_{num\_vertex}$	Number of vertices (words) in the whole network
$Graph_{num\_edges}$	Number of edges in the whole network
$Graph_{max\_geodesic}$	Maximum geodesic distance of the whole network
$Graph_{avg\_geodesic}$	Average geodesic distance of the network
$Graph_{density}$	Density of the whole network
$Graph_{modularity}$	Modularity of the whole network
Vertex metrics	
$Degree_{mean}$	Mean degree centrality of the network
$Degree_{stdev}$	Standard deviation of the degree centrality of the network
$Degree_{max}$	Maximum degree centrality value of the network
$Betweenness_{mean}$	Mean betweenness centrality of the network
$Betweenness_{stdev}$	Standard deviation of the betweenness centrality of the network
$Betweenness_{max}$	Maximum betweenness centrality value of the network
$Eigenvector_{mean}$	Mean eigenvector centrality of the network
$Eigenvector_{stdev}$	Standard deviation of the eigenvector centrality of the network
$Eigenvector_{max}$	Maximum eigenvector centrality value of the network
$Clustering_{mean}$	Mean clustering coefficient of the network
$Clustering_{stdev}$	Standard deviation of the clustering coefficient of the network
Cluster metrics	
$Vertices_{mean}$	Mean number of vertices of the clusters
$Vertices_{stdev}$	Standard deviation of the number of vertices of the clusters
$Vertices_{max}$	Maximum number of vertices in a clusters
$Vertices_{min}$	Minimum number of vertices in a clusters
$Edges_{mean}$	Mean number of edges of the clusters
$Edges_{stdev}$	Standard deviation of the number of edges of the clusters
$Edges_{max}$	Maximum number of edges in a clusters
$Edges_{min}$	Minimum number of edges in a clusters
$Max\_geodesic_{mean}$	Mean of the maximum geodesic distance of the clusters
$Max\_geodesic_{stdev}$	Standard dev. of the maximum geodesic distance of the clusters
$Max\_geodesic_{max}$	Maximum of the maximum geodesic distance of the clusters
$Max\_geodesic_{min}$	Minimum of the maximum geodesic distance of the clusters
$Avg\_geodesic_{mean}$	Mean of the mean geodesic distance of the clusters
$Avg\_geodesic_{stdev}$	Standard dev. of the mean geodesic distance of the clusters
$Avg\_geodesic_{max}$	Maximum of the mean geodesic distance of the clusters
$Avg\_geodesic_{min}$	Minimum of the mean geodesic distance of the clusters
$Density_{mean}$	Mean density of the clusters
$Density_{stdev}$	Standard deviation of the densities of the clusters
$Density_{max}$	Maximum density of a cluster
$Density_{min}$	Minimum density of a cluster

5 Results

5.1 Semantic Networks of Content Knowledge. In order to analyze the lecture's semantic networks, a window size of ten was used. This is consistent with what has been recommended in the literature and used in the literature across a wide range of domains [36,48,49]. Danowski used a radius-based windows size and tested radius sizes of up to twenty words on either side of a word. In terms of computing resources, a radius of 3 (7 words is the equivalent to our method) was recommended if the objective is to only identify word clusters [36]. The Clauset–Newman–Moore algorithm was employed in this work [37]. This clustering algorithm is helpful for inferring large community structure and extract meaningful communities from the network based on the optimization of its modularity. A list of common connectors was used to reduce the size of the textual data and at the same time, avoid including irrelevant nodes in the network analysis. Similar approaches have been used in the literature [20,36]. This list of excluded words (prepositions, conjunctions, pronouns, numbers, apostrophes, and common verbs taken from WORD*ij*, a more recent version of WORDLINK [36]) represented an average of 62% of the textual data used (minimum of 57% and maximum of 68%). It is important to note that these lists of excluded words can be employed, depending on the context and purpose of the semantic network evaluation.

As stated previously, one of the main advantages of generating semantic networks is to visualize how different ideas are shaped. For example, in Fig. 3, the semantic network for one of the lectures is shown, including a filter visualization for only those edges whose value is greater than or equal to two (i.e., the pair of words appear together at least twice within the given windows size). From this semantic network, it can be seen that the main topic of the speech was related to teachers and sciences (relatively large nodes and number of connections compared to the other words). Additionally, some subtopics can be visualized from the clusters that are represented by different colors in Fig. 3. For example, in the cluster colored light-blue in Fig. 3, most of the words are AQ4 related to emotions and levels of comfort. It can be said that if the intention of the instructor was to include this subtopic, he/she was able to properly structure it. There may be instances where words may seem grammatically similar (e.g., kid and kids in Fig. 3) but semantically different (e.g., an instructor in a lecture using the word "kid" to mean "joking," while another lecture may use the term "kids" to represent children). Manually clustering these two words together may introduce errors in the semantic structure of

Fig. 2 Participants' layout in the classroom

the network. Instead, semantic relationships are discovered quantitatively, based on the proposed methodology.

The semantic network graph can be complemented by its metrics. The size of the network (number of words) is 181, and the number of edges is 1853. This network has a *density* of 0.0513, indicating that about 5% of the maximum potential edges exist in the whole network. The maximum *geodesic distance* is 6, meaning that at most, five other words separate each pair of words in the network. For example, the words *communication* (dark green left side node) and *problem* (orange right side node) have a geodesic distance of 5, as they are separated by 5 nodes in their shortest path (i.e., *communication* → *important* → *science* → *teachers* → *better* → *course* → *problem).* The average *geodesic distance* is 2.676, which is fairly low, given the size of the network. This metric can be thought of as a measure of reachability or connectivity of the topics of the network. The results reveal that the semantic network is composed of 6 clusters that contain between 11 to 54 words each, and between 44 to 432 edges. The maximum *geodesic distance* for these groups of clusters ranges between 4 and 5, while its average ranges from 1.719 to 2.445. Finally, these clusters have densities that range from 0.146 to 0.400 each.

Vertex-related metrics are also interesting to analyze and complement the semantic network graph. For example, the words *science*, *teachers*, *students*, *questions*, and *training* have a degree centrality of 57, 54, 39, 33, and 32, respectively. This indicates that this set of words represents the central topic of the lecture, as these values are relatively large, compared to the other words in the network. In addition to the previous set of words, others such as *universe* and *kids* have a large betweenness centrality, indicating that this set of words serves to bridge different ideas between the topic and subtopics.

Through the semantic network graph, the encoder is able to visualize whether the message was structured as intended in terms of the main topic, subtopics, and how they are related. Therefore, the visual information extracted from the semantic network graph can be used to calibrate the structured design of the message as intended by the encoder. Moreover, visual text analytics are useful for knowledge building, analytical reasoning, and explorative analysis [50].

In addition to the information extracted through the visualization of the message, the subsequent sections in this study assess the correlation between semantic network metrics and emotional states intensities. From this analysis, some insights can be obtained about AQ5 the network metrics impacting the participants' emotional responses. For instance, if the parameter $graph_{max\_geodesic}$ is negatively impacting the set of positive emotional states that impact students' learning, the communicator might try to incorporate a new word or tie to the semantic network in such a way that the maximum distance between words (maximum geodesic distance) is reduced. More insights about these practical implications are given in Secs. 5.2 and 5.3.

Table 3 List of videos for the classroom experiment

List	Author	Title
A	M. Kaku	Will mankind destroy itself?
	L. Smolin	Physics envy and economic theory
	E. Kandel	Creativity, your brain, and the aha! moment
	L. Krauss	Should science teachers be paid more than humanities teachers
	M. Kaku	The dark side of technology
B	M. Kaku	Escape to a parallel universe
	S. Žižek	Don't act. Just think
	R. Mckee	Bad writers have nothing to say
	K. Dutton	Are you a psychopath? Take the test
	M. Kaku	What if Einstein is wrong?

5.2 Mining Interesting Correlations Between Network Metrics and Emotional States. When considering the correlation analysis for the 110 data points (22 participants reporting emotions for five video-lectures each), a subset of relationships are found between the semantic network metrics and students' emotional states. According to the Salkind scale [44] (±[0.0 to 0.2]: weak or no relationship; ±[0.2 to 0.4]: weak relationship; ±[0.4 to 0.6]: moderate relationship; ±[0.6 to 0.8]: strong relationship; ±[0.8 to 1.0] very strong relationship), engagement, boredom, and interest, are the three emotional states that are more correlated, in general, to the semantic network metrics. However, approximately half of these correlations are not significant at the 0.05 level. Table 4 presents the Pearson coefficient for those network metrics that were at least moderately correlated to any emotional state in our study.

These relatively low correlation values could be argued to be the result of the high variability of responses within the same lecture. The average within lecture standard deviation of the responses was 0.755, ranging from 0.422 to 1.270. In addition, within the same lecture and same emotional state, a wide range of answers might be obtained. For instance, for lecture 2 in list A, the values reported for frustration ranged from 1 to 5. Therefore, one might be interested in understanding the correlations for the average emotional state intensities and the semantic network metrics, similar to how researchers explore correlations between average student ratings and instructor effectiveness [51]. This analysis is presented in the following paragraphs. It must be recalled, however, that one limitation of this approach is that the measures of within lecture variability are lost, as individual observations are averaged by video-lecture. In contrast, a better understanding of

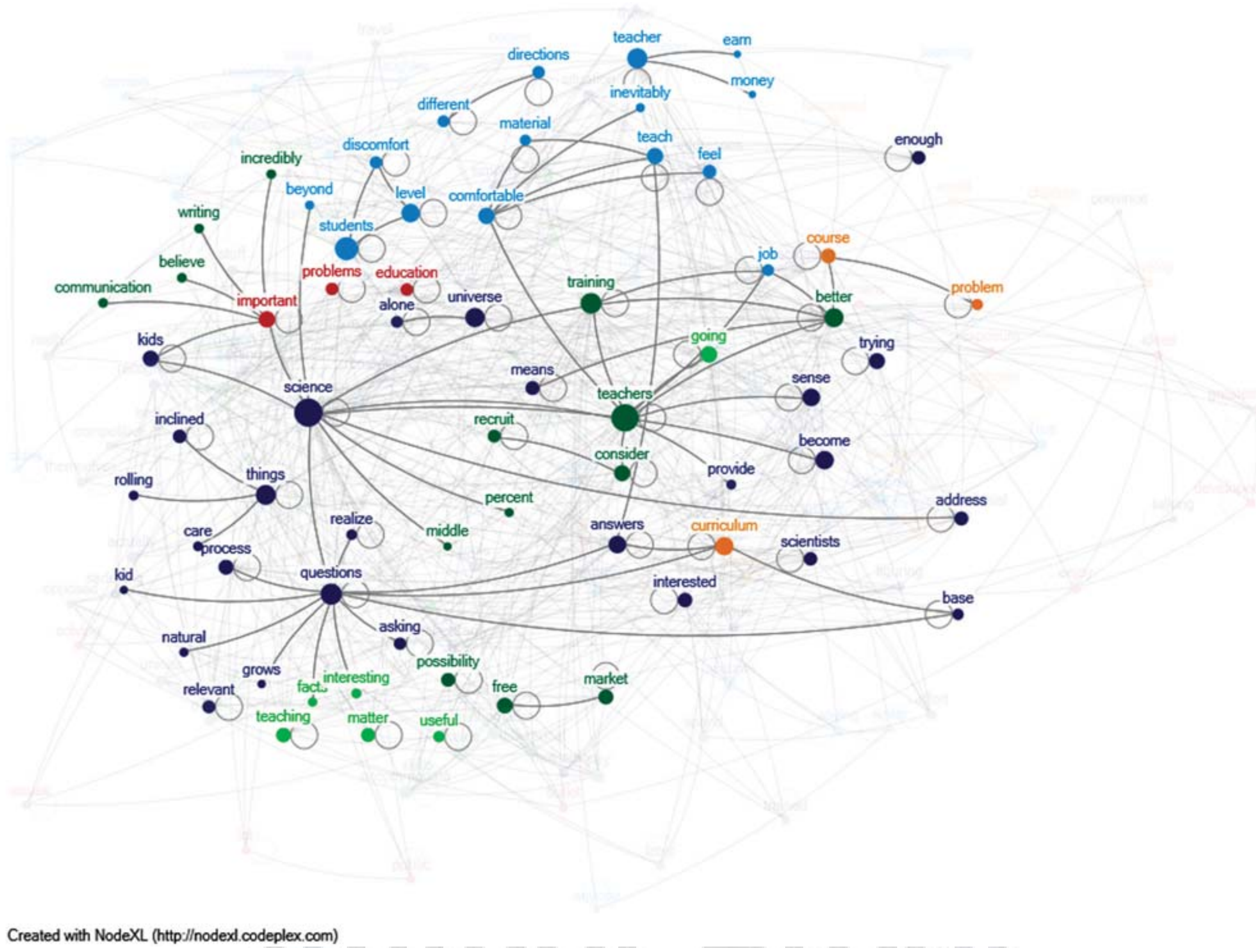

Fig. 3 Semantic network of lecture A4 using windows size ten

how the semantic network metrics are related to students' emotional states as a whole is obtained.

According to the correlation analysis for the average emotional state intensities per lecture ($n = 10$), various interesting relationships between network metrics and emotional states are found (significant at the 0.05 level). A detailed correlation analysis indicated that the overall graph metrics that are more related to the emotional states are $graph_{density}$, $graph_{max\_geodesic}$, and $graph_{avg\_geodesic}$. For example, the $graph_{max\_geodesic}$ is very strongly positively correlated to boredom ($r = 0.813$) and frustration ($r = 0.831$), very strongly negatively correlated to engagement ($r = -0.855$), strongly negatively correlated to interest ($r = -0.766$) and delight ($r = -0.676$), and moderately positively correlated to confusion ($r = 0.411$). One of the interpretations that can be given in this example is that for those lectures whose semantic networks are not well connected, participants' negative emotional states tend to increase while positive emotional states decrease. In this sense, metrics such as density and geodesic distance play a major role in impacting individuals' emotional states and therefore could have a significant impact on their achievement outcomes.

For the clustering metrics, there are various network metrics that are related to the emotional states. Some metrics such as $vertices_{mean}$, $vertices_{min}$, $edges_{mean}$, $edges_{min}$, $max\_geodesic_{stdev}$,

Table 4 Pearson coefficient for moderately correlated semantic network metrics

	ENG	BOR	INT	FRU	DEL	CON
$Degree_{mean}$	0.377[a]	−0.450[a]	0.365[a]	−0.174	0.321[a]	−0.187[b]
$Betweenness_{max}$	−0.399[a]	0.470[a]	−0.411[a]	0.239[b]	−0.415[a]	0.243[b]
$Eigenvector_{stdev}$	−0.326[a]	0.404[a]	−0.315[a]	0.147	−0.265[a]	−0.013
$Eigenvector_{max}$	−0.324[a]	0.428[a]	−0.290[a]	0.086	−0.264[a]	0.058
$Vertices_{mean}$	0.410[a]	−0.446[a]	0.380[a]	−0.139	0.306[a]	−0.180
$Vertices_{min}$	0.458[a]	−0.448[a]	0.393[a]	−0.181	0.237[b]	−0.217[b]
$Edges_{min}$	0.456[a]	−0.450[a]	0.428[a]	−0.264[a]	0.300[a]	−0.184
$Max\_geodesic_{stdev}$	−0.406[a]	0.368[a]	−0.343[a]	0.220[b]	−0.182	0.203[b]
$Avg\_geodesic_{stdev}$	−0.423[a]	0.362[a]	−0.355[a]	0.223[b]	−0.138	0.235[b]
$Avg\_geodesic_{min}$	0.414[a]	−0.372[a]	0.340[a]	−0.185	0.169	−0.188[b]
$Graph_{max\_geodesic}$	−0.535[a]	0.515[a]	−0.534[a]	0.379[a]	−0.371[a]	0.168

[a]Correlation is significant at the 0.01 level (2-tailed).
[b]Correlation is significant at the 0.05 level (2-tailed).

$avg\_geodesic_{stdev}$, and $avg\_geodesic_{max}$ present at least a moderate relationship to at least five out of the six emotional states under analysis. For example, $vertices_{mean}$, $vertices_{min}$, $edges_{mean}$, $edges_{min}$ are positively correlated to engagement ($r = 0.567$, $r = 0.657$, $r = 0.533$, and $r = 0.685$, respectively), interest ($r = 0.479$, $r = 0.542$, $r = 0.492$, and $r = 0.579$, respectively), and delight ($r = 0.620$, $r = 0.479$, $r = 0.622$, and $r = 0.554$, respectively), and negatively correlated to boredom ($r = -0.632$, $r = -0.637$, $r = -0.613$, and $r = -0.673$, respectively), frustration ($r = -0.456$, $r = -0.561$, $r = -0.427$, and $r = -0.619$, respectively), and confusion ($r = -0.309$, $r = -0.407$, $r = -0.341$, and $r = -0.384$, respectively). As can be seen from the correlations, emotional states such as engagement and boredom are consistently strongly correlated to the network metrics described. On the other hand, confusion is typically weakly correlated to these four network metrics. From the analysis, it could be argued that the size of the clusters in a network (i.e., mean and minimum number of vertexes and edges per cluster) is positively correlated to positive emotional states and negatively correlated to negative emotional states. On the other hand, other network metrics such as $max\_geodesic_{stdev}$, $avg\_geodesic_{stdev}$, and $avg\_geodesic_{max}$ are positively correlated to negative emotional states and negatively correlated to positive emotional states. In this case, these three metrics are positively correlated to boredom ($r = 0.523$, $r = 0.510$, and $r = 0.509$, respectively), frustration ($r = 0.541$, $r = 0.576$, and $r = 0.438$, respectively), and confusion ($r = 0.404$, $r = 0.458$, and $r = 0.492$, respectively), and negatively correlated to engagement ($r = -0.587$, $r = -0.602$, and $r = -0.517$, respectively), interest ($r = -0.493$, $r = -0.542$, and $r = -0.554$, respectively), and delight ($r = -0.337$, $r = -0.264$, and $r = -0.345$, respectively). From this set of correlation analyses, it can be said that $max\_geodesic_{stdev}$, $avg\_geodesic_{stdev}$, and $avg\_geodesic_{max}$ are at least moderately correlated to boredom and engagement. However, they are only weakly correlated to delight. The practical interpretation of these patterns is similar to the interpretation for the overall graph metrics.

The clustering metrics that were found to be at most weakly correlated to the emotional states include $vertices_{stdev}$, $edges_{stdev}$, $edges_{max}$, $max\_geodesic_{min}$, $max\_geodesic_{max}$, and $avg\_geodesic_{mean}$. Finally, the vertex metrics that were found to be at least moderately correlated to at least five emotional states are $betweenness_{stdev}$ $betweenness_{max}$, and $eigenvector_{stdev}$. Out of this group, $betweenness_{max}$ is strongly negatively correlated to engagement ($r = -0.648$), interest ($r = -0.638$), and delight ($r = -0.745$), strongly positively correlated to boredom ($r = 0.750$), and moderately positively correlated to frustration ($r = 0.454$) and confusion ($r = 0.547$). The betweenness centrality metrics were typically positively correlated to positive emotional states, while eigenvector centrality metrics were typically positively correlated to negative emotional states. The emotional state that is least correlated to these metrics is frustration. These results could be interpreted in terms of how topic words are used to connect the topic to make the information flow in a coherent manner. Betweenness metrics can be seen as words that help to connect the ideas of a topic. Usually, these words are also the main words in a topic and serve as a central idea to explore different subtopics, represented in this case by the clusters of the network. Therefore, lectures that have a weak structure in terms of their betweenness centrality could have a negative impact on participants' emotional states.

Table 5 ANOVA for engagement

Model	Sum of squares	df	Mean square	F	Sig
Regression	48.839	8	6.105	10.289	0.000
Residual	59.925	101	0.593		
Total	108.764	109			

Additionally, there are other vertex-related metrics that presented some interesting relationships. For example, $degree_{mean}$ was strongly negatively correlated to boredom ($r = -0.645$), and moderately positively correlated to the positive emotional states; engagement ($r = 0.531$), interest ($r = 0.522$), and delight ($r = 0.563$). Some other metrics such as $degree_{stdev}$ and $clustering_{mean}$ do not appear to be significantly correlated to the different emotional states.

5.3 Network Metrics to Estimate Emotional State Intensities. Regression analysis was conducted to identify significant independent variables in order to estimate emotional states. The analysis was conducted using *IBM SPSS Statistics 22*. Each emotional state was separately used as the dependent variable. Consequently, six regression models were created. An interesting pattern found through regression was that the best model for all emotional states included the same eight variables when using the *backward method* with criterion of probability of F-to-remove ≥ 0.100 (vertex-related: $degree_{stdev}$ and $degree_{max}$; cluster-related: $vertices_{mean}$, $vertices_{stdev}$, $edges_{min}$, and $graph_{num\_vertex}$; and overall network-related: $graph_{max\_geodesic}$ and $graph_{modularity}$). Interestingly, five out of the eight metrics were not found to be relevant in the correlation analysis (at least moderately correlated to any emotional state). For instance, in Table 5, the ANOVA is presented for the emotional state ENG. This model provides a good fit for the data ($p < 0.05$), had an $r = 0.670$, and was able to explain 45% of the variability of ENG mean (Table 6). It is important to recall that not all variables included are statistically significant, as some *p*-values are greater than the 0.05 level, as shown in Table 6. However, these variables were selected by the models to predict each one of the emotional states (backward method). Other regression methods such as the stepwise can be used to provide more conclusive results, as the generated models only include statistically significant variables. When using the stepwise method, at most three variables are found to be statistically significant per model at the 0.05 level and entry and removal probabilities-of-F of 0.05 and 0.10, respectively (ENG: $r = 0.663$, $graph_{max\_geodesic}$, $vertices_{stdev}$, and $graph_{num\_edges}$; INT: $r = 0.668$, $graph_{max\_geodesic}$, $degree_{mean}$, and $graph_{num\_edges}$; DEL: $r = 0.507$, $graph_{max\_geodesic}$, and $degree_{max}$; BOR: $r = 0.660$, $graph_{max\_geodesic}$, $degree_{mean}$, and $graph_{num\_edges}$; FRU: $r = 0.429$, $graph_{max\_geodesic}$ and $max\_geodesic_{mean}$; and CON: $r = 0.407$, $graph_{num\_edges}$ and $edges_{min}$). The results and discussions to be presented are based on the models generated from the backward method. These models are interesting to be compared, given that the same variables are used. In future works, however, each emotional state could be modeled using a more conclusive model using only the corresponding statistically significant variables.

The sign (+ or −) of contribution of each one of the variables included in the models could explain the relationship between those variables and emotional state intensities. This becomes relevant as a mechanism to explore how different lectures can impact students' affect in a classroom setting. In most cases, the sign of the variables are consistent with what was already explained in the correlation analysis. For instance, there are three variables that negatively impact students' engagement: $degree_{max}$, $graph_{num\_vertex}$, and $graph_{max\_geodesic}$. The last two metrics can be seen as proxies of "complexity" in the structure of the message. In this sense, $graph_{num\_vertex}$ (number of unique topic words) serves as an approximation of the size of the message while $graph_{max\_geodesic}$ (maximum distance between two words) could represent how far apart two words or concepts are within the message. Interestingly, the sign contribution of $graph_{max\_geodesic}$ is statistically significant in five of the six models, even when testing the stepwise regression method. In addition, there are variables that positively impact students' engagement: $degree_{stdev}$, $vertices_{mean}$, $vertices_{stdev}$, $edges_{mean}$, and $graph_{modularity}$. Interesting insights can be obtained from interpreting these variables. The first variable, $degree_{stdev}$, indicates that the words used in the message should be heterogeneous in terms of their use, and hence, a larger standard deviation of degree centrality will positively

Table 6 Regression model for engagement

	Unstandardized coefficients		Standardized		
Model	B	Std. error	B	t	Sig.
(Constant)	2.255	4.964		0.454	0.651
$Degree_{stdev}$	0.245	0.190	0.243	1.294	0.199
$Degree_{max}$	−0.036	0.022	−0.289	−1.659	0.100
$Vertices_{mean}$	8.494	6.828	0.155	1.244	0.216
$Vertices_{stdev}$	3.727	7.193	0.051	0.518	0.606
$Edges_{min}$	0.006	0.008	0.117	0.796	0.428
$Graph_{num\_vertex}$	−0.009	0.003	−0.260	−2.658	0.009
$Graph_{max\_geodesic}$	−0.615	0.134	−0.463	−4.597	0.000
$Graph_{modularity}$	3.684	1.264	0.256	2.914	0.004

Table 7 Pearson correlation coefficients and the sign of contribution of each variable included in the regression models

	ENG	INT	DEL	BOR	FRU	CON
r	0.670	0.669	0.547	0.676	0.454	0.506
$Degree_{stdev}$	+	+	+	−	+	−[a]
$Degree_{max}$	−	−[a]	−[a]	+[a]	−	+[a]
$Vertices_{mean}$	+	+[a]	+[a]	−	−	−[a]
$Vertices_{stdev}$	+	+	+	+	+	−[a]
$Edges_{min}$	+	+	+	−	−	−
$Graph_{num\_vertex}$	−[a]	−[a]	−	+	+[a]	+[a]
$Graph_{max\_geodesic}$	−[a]	−[a]	−[a]	+[a]	+[a]	+
$Graph_{modularity}$	+[a]	+[a]	−	−[a]	+	−[a]

[a]Contribution sign is significant at the 0.05 level.

impact students' engagement. In the lecture context, this indicates that the key words should be used more frequently and have more connections than other words. The three cluster-related variables that positively impact students' engagement also provide some useful insights. For instance, $vertices_{mean}$ and $edges_{min}$ are proxies of the size of the clusters. The positive contribution indicates that larger clusters (subtopics) are better at influencing the emotional state of engagement. However, $vertices_{stdev}$ indicates that these clusters should be of heterogeneous sizes, i.e., the subtopics should not have the same importance (number of words) within the lecture. Finally, the sign of the contribution of $graph_{modularity}$ indicates that a stronger division of a network into its clusters is better at impacting students' engagement state. These discovered insights are inputs for designers of lectures seeking to increase students' positive emotional states. A similar analysis can be made for the remaining emotional state regression models.

A summary of the Pearson correlation coefficient (r) and the sign of contribution of each variable included in the regression models are shown in Table 7. As can be seen from Table 7, the emotional states that can be better explained by the regression models are INT ($r = 0.669$), BOR ($r = 0.676$), and ENG ($r = 0.670$). In contrast, FRU was the least explained emotional state ($r = 0.454$). The sign of contributions for all the variables, except for $graph_{modularity}$, is consistent among the positive emotional states, i.e., the variable had either a positive or negative contribution consistently for ENG, INT, and DEL. For the set of negative emotional states (BOR, FRU, and CON), four variables are found to be consistent in terms of their contribution ($vertices_{mean}$, $edges_{min}$, $graph_{num\_vertex}$, and $graph_{max\_geodesic}$). The others present some degree of inconsistency. Moreover, if we analyze the signs of the most explained positive (INT) and negative (BOR) emotional states, we can also infer a practical degree of consistency in the contribution of the variables. All the variables, except for $vertices_{stdev}$, have different signs of contribution for these two emotional states. In Table 7, the consistent semantic network metrics within the positive or negative sets of emotional states are colored in gray. All the Pearson correlation coefficients are significant at the 0.05 level.

Some inconsistencies were also found when assessing the contribution of some of the network metrics on the negative and positive emotional states. One might expect that those network metrics that positively contribute to the positive emotional states should have the opposite impact on negative emotional states. These inconsistencies make the practical interpretation of the impact of these network metrics challenging. Nevertheless, the consistent semantic network metrics provide insights for designing superior messages. For instance, $edges_{min}$ was found to be consistent among the positive and negative emotional states. This indicates that a well-designed message should ensure that the minimum number of edges in a cluster should not be low. This design feature seeks to increase the size of the smaller cluster, based on its number of edges. From a course instructor's perspective, a low $edges_{min}$ value might indicate that at least one of the clusters or subtopics was not properly developed in terms of its size (measured in terms of its edges), and hence, not enough value can be extracted by the audience from that cluster. Similarly, the $vertices_{mean}$ also provides some insights about the size of the clusters. In this case, larger clusters (on average) positively impact students' emotional states. This finding is tied to the previous one in the sense that the average size of the subtopics should at least reach a certain level, in this case, measured by the number of words composing a topic or cluster. It must be noted however that more research is needed in order to determine the minimum and maximum thresholds to design each subtopic, given a general message or communication being transmitted.

Interest, *engagement*, and *boredom* were the emotional states that were most explained by the regression models. An interesting finding of the analysis was that all six emotional states analyzed had the same eight significant predictors. Additionally, the contribution of seven out of eight semantic network metrics is consistent for the positive emotional states. In contrast, half were consistent for the negative emotional states. When considering the overall consistency (different impact on the positive and negative emotion states), $vertices_{mean}$, $edges_{min}$, $graph_{num\_vertex}$, and $graph_{max\_geodesic}$ are consistent, and hence, their practical implications are more easily interpreted when designing course content. For instance, $graph_{max\_geodesic}$ has a negative impact on positive emotional states and a positive impact on negative emotional states. This informs the source (i.e., course instructor) that a smaller $graph_{max\_geodesic}$ is desirable. Therefore, some strategies to make this semantic network parameter smaller could be, for instance, incorporating a new link between words in the shortest path (shortest distance) between the two most separated words. From the example presented in Fig. 3, for lecture A4, the $graph_{max\_geodesic}$ is six. One of the paths with distance six is the path between the words *communication* and *problem* (*communication* → *important* → *science* → *teachers* → *better* → *course* → *problem*); hence, the message can be improved by making this path and other paths with distance six shorter, for instance, by directly connecting *communication* with *problem*, or through other words in such a way that the distance between *communication* and *problem* is less than six, which is the current $graph_{max\_geodesic}$ of the semantic network.

6 Conclusion and Future Work

In this work, the authors test the hypothesis that there exists a correlation between the semantic structure of lecture content and students' affective states. According to our results, when considering the set of 110 data points, some network metrics are moderately correlated to a subset of the emotional states analyzed. The overall graph metric $graph_{max\_geodesic}$ was found to be moderately correlated with students' emotional states. Additionally, this variable was statistically significant for all emotional states except for confusion, when selecting a regression model using the stepwise method. Cluster-related metrics, including $vertices_{mean}$, $vertices_{min}$, $edges_{min}$, $max\_geodesic_{stdev}$, $avg\_geodesic_{stdev}$, and

$avg\_geodesic_{min}$, were at least moderately correlated with any of the six emotional states. The vertex-related metrics found to be more relevant in explaining emotional states include $degree_{mean}$, $betweenness_{max}$, $eigenvector_{stdev}$, and $eigenvector_{max}$. It might be noted that, in practice, achieving a strong correlation is difficult as the emotions are based on each student's individual feedback, and hence, are likely to differ from one to another (high within lecture variability in the emotional states intensities). In addition, following the principles of the Arrow's theorem [52]), it may not be possible to satisfy all individuals (students in the case of a classroom setting) by a single product or service design. Therefore, future efforts should be made to design lectures that maximize the demand, or in other words, the number of students that experience positive emotional states during the lecture.

Although the methodology presented can be used to evaluate other types of semantic networks, the set of significant semantic network metrics might change. In this study, we report the results based on the regression model using a *backward method*, and hence, not all the variables used are statistically significant. An interesting discovery was that the same set of semantic network variables was relevant in providing insights about each emotional state. However, for more conclusive interpretations, other regression methods such as the stepwise method could have been used to select only the set of variables that are statistically significant for each emotional state. It should be recalled that the results presented in this study are intended to provide insights about how the semantic network structure can affect the emotional states related to the learning process for a very specific lecture characteristics. Therefore, in other contexts (e.g., MOOCs), different results might be obtained. Although 110 data points were used for the regression analysis, which is over the required sample size according to the Cohen's f^2 score [45], we acknowledge that some of these data points might be correlated as only 22 participants were included in the experiments. In the future, more participants will be included to support more powerful insights from the analyses. We also acknowledge that the semantic network features of a message can explain only one part of the communication processes. In future work, mechanisms to codify informal verbal/textual communication (e.g., jargons), impact of specific topics or strong words on emotional states (e.g., profanity, abuse, etc.), nonverbal communication features, such as body language, intonation, facial gestures, presenters' style, and others, could be used to improve our ability to design messages incorporating the multidimensionality nature of communication. Knowledge gained from exploring the relationships between the semantic structure of lecture content and students' emotional states will inform educators of the specific semantic structure of lecture content that enhance students' affective states and interest in course content, toward the goal of increasing STEM retention rates and overall positive experiences in STEM majors.

Acknowledgment

This research was funded in part by the National Science Foundation: NSF DUE #1449650: Investigating the Impact of Co-Learning Systems in Providing Customized, Real-Time Student Feedback. Any opinions, findings, or conclusions found in this paper are those of the authors and do not necessarily reflect the views of the sponsors.

References

[1] "Increasing the Number of STEM Graduates: Insights From the U.S. STEM Education and Modeling Project," Business-Higher Education Forum, http://www.bhef.com, last accessed 2010. AQ6

[2] Atkinson, R. D., and Mayo, M., 2010, *Refueling the U.S. Innovation Economy: Fresh Approaches to STEM Education*, The Information Technology and Innovation Foundation, Washington, DC.

[3] Schutz, P. A., and Lanehart, S. L., 2002, "Introduction: Emotions in education," Educ. Psychol., **37**(2), pp. 67–68.

[4] Singh, K., Granville, M., and Dika, S., 2002, "Mathematics and Science Achievement: Effects of Motivation, Interest, and Academic Engagement," J. Educ. Res., **95**(6), pp. 323–332.

[5] Crilly, N., Moultrie, J., and Clarkson, P. J., 2004, "Seeing Things: Consumer Response to the Visual Domain in Product Design," Des. Stud., **25**(6), pp. 547–577.

[6] Govers, P., Hekkert, P., and Schoormans, J. P., 2003, "Happy, Cute and Tough: Can Designers Create a Product Personality That Consumers Understand," *Design and Emotion*, ■, ■, pp. 345–349. AQ7

[7] Nagamachi, M., 1995, "Kansei Engineering: A New Ergonomic Consumer-Oriented Technology for Product Development," Int. J. Ind. Ergon., **15**(1), pp. 3–11.

[8] Nagamachi, M., 2002, "Kansei Engineering as a Powerful Consumer-Oriented Technology for Product Development," Appl. Ergon., **33**(3), pp. 289–294.

[9] Pour, P. A., Hussain, M. S., AlZoubi, O., D'Mello, S., and Calvo, R. A., 2010, "The Impact of System Feedback on Learners' Affective and Physiological States," *Intelligent Tutoring Systems*, ■, pp. 264–273.

[10] Marchand, G. C., and Gutierrez, A. P., 2012, "The Role of Emotion in the Learning Process: Comparisons Between Online and Face-to-Face Learning Settings," Internet Higher Educ., **15**(3), pp. 150–160.

[11] Gal, I., and Ginsburg, L., 1994, "The Role of Beliefs and Attitudes in Learning Statistics: Towards an Assessment Framework," J. Stat. Educ., **2**(2), pp. 1–15.

[12] Craig, S., Graesser, A., Sullins, J., and Gholson, B., 2004, "Affect and Learning: An Exploratory Look Into the Role of Affect in Learning With AutoTutor," J. Educ. Media, **29**(3), pp. 241–250.

[13] Csikszentmihalyi, M., 2014, "Toward a Psychology of Optimal Experience," *Flow and the Foundations of Positive Psychology*, Springer, ■, pp. 209–226. AQ8

[14] Akey, T. M., 2006, "School Context, Student Attitudes and Behavior, and Academic Achievement: An Exploratory Analysis," MDRC Report No. ■. AQ9

[15] Kort, B., Reilly, R., and Picard, R. W., 2001, "An Affective Model of Interplay Between Emotions and Learning: Reengineering Educational Pedagogy-Building a Learning Companion," IEEE International Conference on Advanced Learning Technologies, p. 0043.

[16] Baker, R. Sj., D'Mello, S. K., Rodrigo, M. M. T., and Graesser, A. C., 2010, "Better to be Frustrated Than Bored: The Incidence, Persistence, and Impact of Learners' Cognitive–Affective States During Interactions With Three Different Computer-Based Learning Environments," Int. J. Human-Comput. Stud., **68**(4), pp. 223–241.

[17] D'Mello, S., and Graesser, A., 2014, "Confusion and its Dynamics During Device Comprehension With Breakdown Scenarios," Acta Psychol., **151**, pp. 106–116.

[18] Kumar, A., Olshavsky, R. W., and King, M. F., 2001, "Exploring Alternative Antecedents of Customer Delight," J. Consum. Satisfaction Dissatisfaction Complaining Behav., **14**, pp. 14–26.

[19] Dong, A., 2005, "The Latent Semantic Approach to Studying Design Team Communication," Des. Stud., **26**(5), pp. 445–461.

[20] Dong, A., Hill, A. W., and Agogino, A. M., 2004, "A Document Analysis Method for Characterizing Design Team Performance," ASME J. Mech. Des., **126**(3), pp. 378–385.

[21] Li, Z., and Ramani, K., 2007, "Ontology-Based Design Information Extraction and Retrieval," Artif. Intell. Eng. Des. Anal. Manuf., **21**(02), pp. 137–154.

[22] Ghani, R., Probst, K., Liu, Y., Krema, M., and Fano, A., 2006, "Text Mining for Product Attribute Extraction," SIGKDD Explor. Newsl., **8**(1), pp. 41–48.

[23] Liang, Y., and Tan, R., 2007, "A Text-Mining-Based Patent Analysis in Product Innovative Process," *Trends in Computer Aided Innovation*, Springer, ■, pp. 89–96.

[24] Kang, S. W., Sane, C., Vasudevan, N., and Tucker, C. S., 2014, "Product Resynthesis: Knowledge Discovery of the Value of End-of-Life Assemblies and Subassemblies," ASME J. Mech. Des., **136**(1), p. 011004.

[25] Glier, M. W., McAdams, D. A., and Linsey, J. S., 2014, "Exploring Automated Text Classification to Improve Keyword Corpus Search Results for Bioinspired Design," ASME J. Mech. Des., **136**(11), p. 111103.

[26] Fu, K., Chan, J., Cagan, J., Kotovsky, K., Schunn, C., and Wood, K., 2013, "The Meaning of 'Near' and 'Far': The Impact of Structuring Design Databases and the Effect of Distance of Analogy on Design Output," ASME J. Mech. Des., **135**(2), p. 021007.

[27] Stone, T., and Choi, S.-K., 2013, "Extracting Consumer Preference From User-Generated Content Sources Using Classification," ASME Paper No. DETC2013-13228, p. V03AT03A031.

[28] Ren, Y., and Papalambros, P. Y., 2012, "On Design Preference Elicitation With Crowd Implicit Feedback," ASME Paper No. DETC2012-70605, pp. 541–551.

[29] Rai, R., 2012, "Identifying Key Product Attributes and Their Importance Levels From Online Customer Reviews," ASME Paper No. DETC2012-70493, pp. 533–540.

[30] Tuarob, S., and Tucker, C. S., 2015, "Quantifying Product Favorability and Extracting Notable Product Features Using Large Scale Social Media Data," ASME J. Comput. Inf. Sci. Eng., **15**(3), p. 031003.

[31] Tucker, C., Pursel, B., and Divinsky, A., ■, "Mining Student-Generated Textual Data in MOOCS And Quantifying Their Effects on Student Performance and Learning Outcomes," ASEE Comput. Educ. J. (CoEd), **5**(4), pp. 84–95. AQ10

[32] Tuarob, S., and Tucker, C. S., 2015, "Automated Discovery of Lead Users and Latent Product Features by Mining Large Scale Social Media Networks," ASME J. Mech. Des., **137**(7), p. 071402.

[33] Griffiths, T., and Steyvers, M., 2002, "A Probabilistic Approach to Semantic Representation," 24th Annual Conference of the Cognitive Science Society, pp. 381–386.

[34] Behoora, I., and Tucker, C., 2015, "Machine Learning Classification of Design Team Members' Body Language Patterns for Real Time Emotional State Detection," Des. Stud., **39**(1), pp. 100–127.

[35] Hoser, B., Hotho, A., Jäschke, R., Schmitz, C., and Stumme, G., 2006, *Semantic*
1037 *Network Analysis of Ontologies*, Springer, ■.

[36] Danowski, J. A., 1993, "Network Analysis of Message Content," Prog. Com-
1038 mun. Sci., **12**, pp. 198–221.

[37] Clauset, A., Newman, M. E., and Moore, C., 2004, "Finding Community Struc-
1039 ture in Very Large Networks," Phys. Rev. E, **70**(6), p. 066111.

[38] Scott, J., and Carrington, P. J., 2011, *The SAGE Handbook of Social Network*
1040 *Analysis*, SAGE Publications, ■.

[39] Borgatti, S. P., Mehra, A., Brass, D. J., and Labianca, G., 2009, "Network Anal-
1041 ysis in the Social Sciences," Science, **323**(5916), pp. 892–895.

[40] Hansen, D., Shneiderman, B., and Smith, M. A., 2010, *Analyzing Social Media Net-*
1042 *works With NodeXL: Insights From a Connected World*, Morgan Kaufmann, ■.

[41] Estrada, E., and Rodriguez-Velazquez, J. A., 2005, "Subgraph Centrality in
1043 Complex Networks," Phys. Rev. E, **71**(5), p. 056103.

[42] Fernandez-Ballesteros, R., 2004, "Self-Report Questionnaires," *Comprehensive*
1044 *Handbook of Psychological Assessment*, ■, Vol. 3, pp. 194–221.

[43] Han, J., and Kamber, M., 2006, *Data Mining, Southeast Asia Edition: Concepts*
1045 *and Techniques*, Morgan kaufmann, ■.

[44] Salkind, N. J., 2006, *Encyclopedia of Measurement and Statistics*, Sage Publi-
1046 cations, ■.

[45] Cohen, J., Cohen, P., West, S. G., and Aiken, L. S., 2013, *Applied Multiple*
1047 *Regression/Correlation Analysis for the Behavioral Sciences*, Routledge, ■.

[46] Gross, J. J., and John, O. P., 2003, "Individual Differences in Two Emotion
Regulation Processes: Implications for Affect, Relationships, and Well-Being," 1048
J. Pers. Soc. Psychol., **85**(2), pp. 348–362. 1049

[47] Dixson, M. D., 2012, "Creating Effective Student Engagement in Online
Courses: What do Students Find Engaging?," J. Scholarship Teach. Learn., 1050 AQ11
10(2), pp. 1–13. 1051

[48] Zywica, J., and Danowski, J., 2008, "The Faces of Facebookers: Investigating
Social Enhancement and Social Compensation Hypotheses; Predicting Face- 1052
book ™ and Offline Popularity From Sociability and Self-Esteem, and Map- 1053
ping the Meanings of Popularity With Semantic Networks," J. Comput.- 1054
Mediated Commun., **14**(1), pp. 1–34. 1055

[49] Smith, R. A., and Parrott, R. L., 2011, "Mental Representations of HPV in
Appalachia: Gender, Semantic Network Analysis, and Knowledge Gaps," J. 1056
Health Psychol., p. 1359105311428534. 1057

[50] Drieger, P., 2013, "Semantic Network Analysis as a Method for Visual Text AQ12
Analytics," Proc.-Social Behav. Sci., **79**, pp. 4–17. 1058

[51] Marsh, H. W., 1987, "Students' Evaluations of University Teaching: Research
Findings, Methodological Issues, and Directions for Future Research," Int. J. 1059
Educ. Res., **11**(3), pp. 253–388. 1060

[52] Hazelrigg, G. A., 1996, "The Implications of Arrow's Impossibility Theorem
on Approaches to Optimal Engineering Design," ASME J. Mech. Des., **118**(2), 1061
pp. 161–164. 1062