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

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

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

Copyright © 2016 by ASME
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. Authors introduce a novel study, followed by Sec. 4 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]. Ghan 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 innovation [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 elicit 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]. Tuorob 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-driven mining 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.

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

\[
E_i = f(O, U, V)
\]

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.

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

\[
W = \{ \text{a, fundamental, attribute, of, the, engineering, design, process, is, information, exchange} \}
\]

\[
C = \{ \text{a, of, the, is,} \}
\]

Thus, the set of topic words is defined as

\[
T = \{ \text{fundamental, attribute, engineering, design, process, information, exchange} \}
\]

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 \( T \) including only unique elements of \( T \), i.e., \( t'_1 \neq t'_2 \neq t'_3 \neq \ldots \neq t'_L \). Therefore, \( T' \) can be written as

![Fig. 1 Methodology for quantifying the correlation between the semantic structure of lecture content and students’ affective states](image-url)

**Journal of Mechanical Design**

MONTH 2016, Vol. 00 / 00000-3
The adjacency matrix $A$ is represented as

$$A = \begin{bmatrix}
  t_{11} & t_{12} & t_{13} & \ldots & t_{1n} \\
  t_{21} & t_{22} & t_{23} & \ldots & t_{2n} \\
  t_{31} & t_{32} & t_{33} & \ldots & t_{3n} \\
  \vdots & \vdots & \vdots & \ddots & \vdots \\
  t_{n1} & t_{n2} & t_{n3} & \ldots & t_{nn}
\end{bmatrix}$$

where $x_{ij}$ represents the frequency or number of times in which both words $i$ and $j$ appear in a window 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: $|T^*|^m(|T^*| - 1)/2$. 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., unreported 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 $|T^*|^m(|T^*| - 1)/2$. The density of the network can be defined as

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

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 messages 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 two last 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

$$C_b(v) = \sum_{s \neq v \neq t} \sigma_{st}(v) \cdot \sigma_{st} \quad (3)$$

where $\sigma_{st}$ 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 $\lambda$ 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 dis-
seminators 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 learn-
ing experience. In order to quantify the emotional state intensities,
we use a Likert scale, as 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 Net-
work. 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 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 clus-
ters can be obtained by employing traditional data mining cluster-
ing 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 meth-
ology 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 vari-
able. The values of r range from −1 to +1. A correlation coeffi-
cient 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 rela-
tion between the variables. Intermediate values can be inter-
terpreted using the Salkind scale [44]. The correlation coefficient
for a sample can be calculated as follows:

\[ r = \frac{\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \overline{X})^2 \sum_{i=1}^{n} (Y_i - \overline{Y})^2}} \]  

where: 
- \( n \): sample size,
- \( X_i \): value of ith observation from sample X, \( i = 1 \) to \( n \),
- \( Y_i \): value of ith observation from sample Y, \( i = 1 \) to \( n \), and
- \( \overline{X} \) and \( \overline{Y} \): average value of all observations from sample X and 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 mes-
sages’ semantic network metrics are linearly correlated to stu-
dents’ 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 mes-
sage’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 indepen-
dent parameters and the receiver’s 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 for
approximation of the relationships between the semantic structure
of messages and individuals’ emotional states. More complex,
nonlinear relationships could exist, 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

\[ E_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_P X_P + \epsilon_i \]  

where \( \beta_0 \) is the intercept or constant, \( \beta_i \) is the slope or contribution
of the semantic network parameter \( x_i \), and \( \epsilon_i \) 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 part-
tioned 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 mini-
mized. In Fig. 2, an illustration of the participants’ learning envi-
ronnement is shown.

Ten video-lectures were selected from YouTube assured vary-
ty of topics and complexity of content. Two lists of five videos
each were generated. The 22 participants were randomly parti-
tioned 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) were col-
lected. The 110 data points generated is sufficient for our study,
given an anticipated effect size (\( f^2 \)) of 1 (assuming \( r^2 = 0.5 \)), sta-
tistical 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.

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, avoiding 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 [50] and the 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 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 term “kids” 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

Table 2 Overall clustering and vertex-related parameters

<table>
<thead>
<tr>
<th>Overall graph metrics</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graph density</td>
<td>Number of vertices (words) in the whole network</td>
</tr>
<tr>
<td>Graph edges</td>
<td>Number of edges in the whole network</td>
</tr>
<tr>
<td>Max_geodesic distance</td>
<td>Maximum geodesic distance of the whole network</td>
</tr>
<tr>
<td>Max_geodesic mean</td>
<td>Mean geodesic distance of the whole network</td>
</tr>
<tr>
<td>Avg_geodesic min</td>
<td>Minimum number of edges in a clusters</td>
</tr>
<tr>
<td>Avg_geodesic max</td>
<td>Maximum number of edges in a clusters</td>
</tr>
<tr>
<td>Max_geodesic stdev</td>
<td>Standard deviation of the number of edges of the clusters</td>
</tr>
<tr>
<td>Avg_geodesic mean</td>
<td>Mean of the maximum geodesic distance of the clusters</td>
</tr>
<tr>
<td>Avg_geodesic max</td>
<td>Maximum of the maximum geodesic distance of the clusters</td>
</tr>
<tr>
<td>Max_geodesic mean</td>
<td>Minimum of the maximum geodesic distance of the clusters</td>
</tr>
<tr>
<td>Densities edge</td>
<td>Mean density of the edges</td>
</tr>
<tr>
<td>Densities max</td>
<td>Maximum density of a cluster</td>
</tr>
</tbody>
</table>

5.2 Vertex Metrics: Number of vertices (words) in the whole network, Number of edges in the whole network, Average geodesic distance of the whole network, Maximum geodesic distance of the whole network, Minimum of the maximum geodesic distance of the clusters, Maximum of the maximum geodesic distance of the clusters, Standard deviation of the maximum geodesic distance of the clusters, Mean of the mean geodesic distance of the clusters, Minimum of the mean geodesic distance of the clusters, Maximum density of a cluster, Minimum density of a cluster.
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 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.

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

Table 3 List of videos for the classroom experiment

<table>
<thead>
<tr>
<th>List</th>
<th>Author</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>M. Kaku</td>
<td>Will mankind destroy itself?</td>
</tr>
<tr>
<td></td>
<td>L. Smolin</td>
<td>Physics easy and economic theory</td>
</tr>
<tr>
<td></td>
<td>E. Kandel</td>
<td>Creativity, your brain, and the aha! moment</td>
</tr>
<tr>
<td></td>
<td>L. Krauss</td>
<td>Should science teachers be paid more than humanities teachers</td>
</tr>
<tr>
<td></td>
<td>M. Kaku</td>
<td>The dark side of technology</td>
</tr>
<tr>
<td>B</td>
<td>M. Kaku</td>
<td>Escape to a parallel universe</td>
</tr>
<tr>
<td></td>
<td>S. Zizek</td>
<td>Don’t act. Just think</td>
</tr>
<tr>
<td></td>
<td>R. Mckee</td>
<td>Bad writers have nothing to say</td>
</tr>
<tr>
<td></td>
<td>K. Dutton</td>
<td>Are you a psychopath? Take the test</td>
</tr>
<tr>
<td></td>
<td>M. Kaku</td>
<td>What if Einstein is wrong?</td>
</tr>
</tbody>
</table>
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 $vertices_{mean}$, $vertices_{min}$, $edges_{mean}$, $edges_{min}$, $max_{geodesic_{stdev}}$, $avg_{geodesic_{stdev}}$, $avg_{geodesic_{min}}$, and $graph_{max_{geodesic}}$ are moderately correlated to emotional states. Table 4 shows the Pearson coefficient for moderately correlated semantic network metrics.

<table>
<thead>
<tr>
<th>Metric</th>
<th>ENG</th>
<th>BOR</th>
<th>INT</th>
<th>FRU</th>
<th>DEL</th>
<th>CON</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Degree_{mean}$</td>
<td>0.377*</td>
<td>-0.450*</td>
<td>0.365*</td>
<td>-0.174</td>
<td>0.321*</td>
<td>-0.187b</td>
</tr>
<tr>
<td>$Betweenness_{max}$</td>
<td>-0.399*</td>
<td>0.470*</td>
<td>-0.411*</td>
<td>0.239b</td>
<td>-0.415*</td>
<td>0.243b</td>
</tr>
<tr>
<td>$Eigen_{advers}$</td>
<td>-0.326*</td>
<td>0.404*</td>
<td>-0.315*</td>
<td>0.147</td>
<td>-0.265*</td>
<td>-0.013</td>
</tr>
<tr>
<td>$Eigen_{avex}$</td>
<td>-0.324*</td>
<td>0.428*</td>
<td>-0.290*</td>
<td>0.086</td>
<td>-0.264*</td>
<td>0.058</td>
</tr>
<tr>
<td>$Vertices_{mean}$</td>
<td>0.410*</td>
<td>-0.446*</td>
<td>0.380*</td>
<td>-0.139</td>
<td>0.306*</td>
<td>-0.180</td>
</tr>
<tr>
<td>$Vertices_{min}$</td>
<td>0.458*</td>
<td>-0.448*</td>
<td>0.393*</td>
<td>-0.181</td>
<td>0.237b</td>
<td>-0.217</td>
</tr>
<tr>
<td>$Edges_{mean}$</td>
<td>0.456*</td>
<td>-0.450*</td>
<td>0.428*</td>
<td>-0.264*</td>
<td>0.300*</td>
<td>-0.184</td>
</tr>
<tr>
<td>$Max_{geodesic_{stdev}}$</td>
<td>-0.406*</td>
<td>0.368*</td>
<td>-0.343*</td>
<td>0.220*</td>
<td>-0.182</td>
<td>0.203b</td>
</tr>
<tr>
<td>$Avg_{geodesic_{index}}$</td>
<td>-0.423*</td>
<td>0.362*</td>
<td>-0.355*</td>
<td>0.223*</td>
<td>-0.138</td>
<td>0.235b</td>
</tr>
<tr>
<td>$Avg_{geodesic_{min}}$</td>
<td>0.414*</td>
<td>-0.372*</td>
<td>0.340*</td>
<td>-0.185</td>
<td>0.169</td>
<td>-0.188b</td>
</tr>
<tr>
<td>$Graph_{max_{geodesic}}$</td>
<td>-0.535*</td>
<td>0.515*</td>
<td>-0.534*</td>
<td>0.379*</td>
<td>-0.371*</td>
<td>0.168</td>
</tr>
</tbody>
</table>

*Correlation is significant at the 0.01 level (2-tailed).

bCorrelation is significant at the 0.05 level (2-tailed).
720 edgesstdev, edges max, max_geodesic min, max_geodesic max, states represented in this case by the clusters of the network. Therefore, in a topic and serve as a central idea to explore different subtopics, could be interpreted in terms of how topic words are used to con-

733 that is least correlated to these metrics is frustration. These results negatively correlated to negative emotional states. The emotional state states, while eigenvector centrality metrics were typically posi-

741 states. 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_geodesicstdev, avg_geodesicstdev, and avg_geodesicmax are positively correlated to negative emotional states and negatively correlated to positive emotional states. In this case, these 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 = 0.537, r = 0.264, and r = 0.345, respectively). From this set of correlation analyses, it can be said that max_geodesicstdev, avg_geodesicstdev, and avg_geodesicmax 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 verticesstdev, edgesstdev, edgesmax, max_geodesicmin, max_geodesicmax, and avg_geodesicmean. Finally, the vertex metrics that were found to be at least moderately correlated to at least five emotional states are betweenness, betweenness, and eigenvector. Out of this group, betweenness 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 posi-

5.3 Network Metrics to Estimate Emotional State Inten-

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, verticesstdev, and graphnum_edges; INT: r = 0.668, graph_max_geodesic, degree(mean, and graphnum_edges; DEL: r = 0.507, graph_max_geodesic, and degree_max; BOR: r = 0.660, graph_max_geodesic, degree(mean, and graphnum_edges; FRU: r = 0.429, graph_max_geodesic and max_geodesic(mean, and CON: r = 0.407, graphnum_edges and graphnum_edges.). 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 engagement: verticesstdev, verticesedges, edgesmax, and graphmodularity. Interestingly, insights can be obtained from interpreting these variables. The first variable, degreestdev, 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>
<thead>
<tr>
<th>Table 5 ANOVA for engagement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td>Regression</td>
</tr>
<tr>
<td>Residual</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>
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\textsubscript{mean} and edges\textsubscript{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\textsubscript{max} 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\textsubscript{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\textsubscript{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\textsubscript{mean}, edges\textsubscript{max}, graph\textsubscript{max}\_geodesic, and graph\textsubscript{max}\_geodesic\_stddev). 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\textsubscript{max}, 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\textsubscript{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\textsubscript{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\textsubscript{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 emotional states), vertices\textsubscript{mean} edges\textsubscript{max} graph\textsubscript{max}\_geodesic, and graph\textsubscript{max}\_geodesic\_stddev are consistent, and hence, their practical implications are more easily interpreted when designing course content. For instance, graph\textsubscript{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\textsubscript{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\textsubscript{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\textsubscript{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\textsubscript{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\textsubscript{mean}, vertices\textsubscript{min}, edges\textsubscript{min}, max\_geodesic\_stddev, avg\_geodesic\_stddev, and...
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


