Towards Automatic Content Analysis of Social Presence in Transcripts of Online Discussions

Máverick Ferreira
Centro de Informática
Universidade Federal de Pernambuco
Recife, PE, Brazil
madf@cin.ufpe.br

Vitor Rolim
Centro de Informática
Universidade Federal de Pernambuco
Recife, PE, Brazil
vbr@cin.ufpe.br

Rafael Ferreira Mello
Centro de Informática
UFRPE
Recife, Pernambuco, Brasil
rafael.mello@ufrpe.br

Rafael Dueire Lins
Centro de Informática
UFRPE
Recife, Pernambuco, Brasil
rdl@cin.ufpe.br

Guanliang Chen
Faculty of Information Technology
Monash University
Melbourne, VIC, Australia
Guanliang.Chen@monash.edu

Dragan Gašević
Faculty of Information Technology
Monash University
Melbourne, VIC, Australia
dragan.gasevic@monash.edu

ABSTRACT

This paper presents an approach to automatic labeling of the content of messages in online discussion according to the categories of social presence. To achieve this goal, the proposed approach is based on a combination of traditional text mining features and word counts extracted with the use of established linguistic frameworks (i.e., LIWC and Coh-metrix). The best performing classifier obtained 0.95 and 0.88 for accuracy and Cohen's kappa, respectively. This paper also provides some theoretical insights into the nature of social presence by looking at the classification features that were most relevant for distinguishing between the different categories. Finally, this study adopted epistemic network analysis to investigate the structural construct validity of the automatic classification approach. Namely, the analysis showed that the epistemic networks produced based on messages manually and automatically coded produced nearly identical results. This finding thus produced evidence of the structural validity of the automatic approach.

CCS CONCEPTS

• **Information systems → Clustering and classification**; • **Applied computing → E-learning**; **Distance learning**;

KEYWORDS

Community of Inquiry Model, Content Analytics, Online Discussion, Text Classification, Epistemic Network Analysis

ACM Reference format:
Máverick Ferreira, Vitor Rolim, Rafael Ferreira Mello, Rafael Dueire Lins, Guanliang Chen, and Dragan Gašević. 2020. Towards Automatic Content Analysis of Social Presence in Transcripts of Online Discussions. In *Proceedings of Proceedings of the 10th International Conference on Learning Analytics and Knowledge, Frankfurt, Germany, March 23–27, 2020 (LAK '20),* 10 pages.
https://doi.org/10.1145/3375462.3375495

1 INTRODUCTION

Online learning enables students and instructors to participate in the teaching and learning process without being in the same physical space [34]. Due to this characteristic, it has promoted access to education for people located in regions that are difficult to reach or distant from educational institutions. One of the critical challenges facing instructors of online courses is creating a supportive and productive environment for student communication and collaboration through technology [11]. According to the literature, asynchronous online discussion is a resource with high potential for promoting collaboration in online education [46] supporting students' social interactions and social-constructivist pedagogies [1], which encourage the engagement of learners [4].

Within this context, a social constructivist model called Community of Inquiry (CoI) [12] is a frameworks developed to support instructors in online learning environments. The study of CoI is heavily depended on the analysis of messages exchanged in online discussions. The most commonly used approach to this analysis is based on *Quantitative Content Analysis* (QCA) [9, 22] of the transcripts of asynchronous online discussions. Krippendorff [27] states that content analysis is *"a research technique for making replicable and valid inferences from texts (or other meaningful matter) to the contexts of their use"*[p.18]. QCA methods can use predefined coding schemes to analyze text artifacts (i.e., messages in online discussions) with respect to the defined research goals and objectives. The CoI model defines the QCA coding schemes for each of the three presences that can be applied to analyze online discussion messages. As widely done in the social sciences, research of CoI primarily uses QCA for retrospection and research after online discussions are over, without much impact on the actual student learning and outcomes in real-time [43].

However, it is possible to adopt automated methods for text analysis commonly used within learning analytics [17] to perform

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

LAK '20, March 23–27, 2020, Frankfurt, Germany
© 2020 Association for Computing Machinery.
ACM ISBN 978-1-4503-7712-6/20/03…$15.00
https://doi.org/10.1145/3375462.3375495

automatic and real-time analysis of online discussion messages according to the CoI presences [23].

Existing literature reports several approaches to automating the process of content analysis of online discussion according to the coding scheme for cognitive presence as proposed in the CoI model [10, 24, 31], even for languages different than English, like Portuguese [35]. However, methods for automatic content analysis of online discussions for indicators of social presence are not commonly found in the literature.

This paper describes a method that combines several text analytics techniques for automatic content analysis of social presence from online discussion transcripts. The study combines traditional text mining features based on content words with tools that extract different psychological processes indicators, and measures of text coherence and complexity [32, 44]. We developed three classifiers, one for each category of social presence, that use different feature sets and achieved up to 0.95 and 0.88 for accuracy and Cohen's Kappa, respectively. Besides, we proposed a network analytic approach to evaluate the structural validity of our proposal in practice. The results and their implications are also discussed in the paper.

2 BACKGROUND WORK

2.1 The Community of Inquiry Model

Several models of and approaches to understanding students' interactions in online environments have been proposed. Among them, the Community of Inquiry (CoI) model is one of the most researched structures when the objective is to describe the essential facets of social interactions and knowledge construction in online and blended education [14]. CoI proposes three dimensions that explain the processes of social knowledge construction of learners and instructors with the goal of describing promoting effective educational experience. Garrison et al. [12] distinguishes between the three key dimensions of CoI, known as presences, as follows: (i) **Social presence** measures the ability to humanize the relationships among participants in a discussion. It focuses on social interactions and tries to model the social climate within a group of learners (i.e., cohesion, affectivity, and open communication) [41]; (ii) **Cognitive presence** is highly related to the development of learning outcomes. It aims to capture the progress of interactions in students' cognitive processes that support the development of critical thinking, knowledge construction, and problem-solving [13]; (iii) **Teaching presence** concerns teaching role before (i.e., course design) and during (i.e., facilitation and direct instruction) the course [2].

Given the focus of the current study, social presence is further unpacked. Garrison et al. [12] define social presence as *"the ability of participants in a community of inquiry to project themselves socially and emotionally, as 'real' people (i.e., their full personality), through the medium of communication being used"* (p. 94). Besides, in contrast to face-to-face interaction, in online discussions, it is essential to textually express such abilities in order to establish a socio-emotional communication [15]. Social presence, as defined in the CoI model, includes three categories: (i) **Affective**: This category analyses the translation of real emotions into text. It encompasses emotion, feelings, and mood expressions; (ii) **Interactive**: This category focuses on the interactivity of the messages exchanged among participants. The main goal of this category is to enhance open communication among students; (iii) **Group Cohesion**: This category investigates the sense of union and group commitment among students.

Each of the three categories of social presence has several indicators related, as described in Table 1. These indicators are a roadmap to interpret the interactions though the social presence concept.

Table 1: Indicators social presence [41].

Category	Indicator	Label
Affective	1. Expression of emotions	Emotions
	2. Use of humor	Humor
	3. Self-disclosure	Self_disclosure
Interactive	4. Continuing a thread	Cont_Thread
	5. Quoting from others' messages	Quoting_Mess
	6. Referring explicitly to others' message	Referring_Mess
	7. Asking questions	Asking_Q
	8. Complimenting, expressing appreciation	Complimenting
	9. Expressing agreement	Agreement
Cohesive	10. Vocatives	Vocatives
	11. Addresses of refers to the group using inclusive pronouns	Group
	12. Phatics, salutations	Salutations

2.2 Analysis of the CoI

The published literature presented two methods for the analysis of the three presences within the CoI perspective through the use of questionnaires and the adoption of content analysis.

Several questionnaires have been proposed and validated in the context of CoI to examine the perception of students about their experience online interactions. The most broadly adopted is the instrument proposed by Arbaugh et al. [3], in which a 34-item survey measures the perception of the students regarding the three presences using a five-point Likert scale (1 = strongly disagree to 5 = strongly agree). This form is adopted by several studies to analyze individual presences [38] and relationship among them [26].

The second approach to the analysis of the CoI presences is content analysis of online discussion transcripts. Rourke et al. [41] and Garrison et al. [13] defined coding schemes to analyze social and cognitive presences. These schemes have widely been adopted for manual content analysis of CoI. For instance, Gašević et al. [16] adopted the cognitive presence scheme to evaluate the improvement of asynchronous online discussions after an instructional intervention. Following a similar idea, Kovanovic et al. [22] used the manual codding to evaluate the association between social presence and social network position.

Initial proposals to automate content analysis according to the coding schemes of the CoI model primarily relied upon features traditionally used in text mining such as word and phrase counts. For instance, Mcklin [31] an artificial neural network based on word frequency features to classify online discussion messages according to their cognitive presence. The classifier reached 0.31 Cohen's κ.

Recent studies examined the use of other features and classifiers. Kovanović et al. [24] examined the use of a combination of bag-of-words (n-gram) and Part-of-Speech (POS) N-gram features for classifying cognitive presence using the Support Vector Machines (SVMs) classifier, achieving 0.41 Cohen's κ. Kovanović et al. [25] and Neto et al. [35] adopted features based on Coh-metrix [32], LIWC [44], latent semantic analysis (LSA) similarity, named entities, and discussion context [45], to identify phases of cognitive presence for messages written in English (0.63 Cohen's κ) and Portuguese (0.72 Cohen's κ). Besides, the authors applied a random forest classifier [6], which also allowed for the analysis of the influence of the different features on the final classification results.

Although there are studies to extract the phases of cognitive presence automatically, to our knowledge, there is no publication that looked at the automatic content analysis of social presence.

3 RESEARCH QUESTIONS

As discussed in Section 2.1, social presence has a key role in the CoI, influencing the development of cognitive presence in online learning environments. It enhances personal relationships and promotes the sense of community among students. Although several studies demonstrated its importance [21], there is no automatic method to code online discussion messages according to the categories of social presence(affective, interactive and cohesive). Hence, our first research question is:

> **Research Question 1 (RQ1):**
> *To what extent can accurately text mining methods automatically code online discussion messages according to the categories of social presence?*

In addition to addressing the above research question by training a supervised machine learning algorithm (i.e., classifier) for social presence, we were also interested in providing additional insights into the features that were more relevant to each of the three categories of social presence. To do so, we explored a method similar to the one applied by Kovanović et al. [25] and Neto et al. [35]. As such, our second research question is:

> **Research Question 2 (RQ2):**
> *Which features do best predict each category of social presence?*

Finally, we were interested in whether the automatically coded messages preserve the same structural properties when associations between social presence and discussion topics were analyzed. That is, we were interested in examining the extent to which the analysis of associations between automatically coded messages produced results similar to the analysis of manually coded messages according to the categories of social presence. Therefore, our third research question is:

> **Research Question 3 (RQ3):**
> *Do automatically coded messages preserve similar structural properties in the analysis of associations between the categories of social presence and discussion topics to the results of the analysis performed with manual manually coded messages according to the categories of social presence?*

4 METHOD

4.1 Data and course design

The dataset used in this study was taken from a fully online master's degree course in software engineering offered by a public university in Canada. The dataset consists of a total of 1.747 posts from the interaction between 81 students during six offers of the course (winter 2008, fall 2008, summer 2009, fall 2009, winter 2010, winter 2011) [16]. The goal of the online discussion was to debate around videos about research papers related to one of the course topics. The participation in the discussion accounted for 15% of the final grade [16].

Table 2: Distribution of social presence categories.

	Messages					
Category	Control		Treatment		*Total*	
Affective	266	14.27%	264	13.84%	530	30.34%
Interactive	825	44.28%	878	46.04%	1,703	97.48%
Cohesive	772	41.45%	765	40.11%	1,535	87.98%
Total	1,863	100%	1,907	100%	3,770	100.00%

During the first two offerings of the course, the participation of the students was primarily driven by the extrinsic motivational factors (i.e., course grade), with limited scaffolding support. The students from the first two offerings are referred to as the *control group*. After the first two course offerings, a scaffolding of discussion participation through role assignments and clear instructions was implemented (*treatment group*). Table 2 shows the number of messages accounting for control and treatment groups. It is important to remark that the same message could have more them one category of social presence.

Two expert coders categorized the dataset, considering the 12 indicators of social presence (see Table 1) [22]. That is, for each post in the dataset, each indicator received the value "one" (has the indicator) or value "zero" (does not have the indicator). The percentage of agreement between the evaluators was 84%, and a third evaluator resolved the cases with disagreements. Following Kovanovic et al. [22], three indicators (Continuing a thread, Complimenting, and Vocatives) were removed because they had a high number of messages.

Finally, as the objective of this study was to construct binary classifiers for each category of social presence, the categories were reorganized to have binary coding (negative 0 or positive 1). For a message to be classified in positive (1), it must have at least one indicator annotated with the value "one" in the respective category. For instance, if a message had for the affective category, the indicators Emotions = 0, Humor = 0, and Self_disclosure = 1, it was coded as positive (1). Finally, we obtained the dataset as shown in Table 3.

Table 3: Final distribution of social presence phases

Category	Negative (0)	Positive (1)
Affective	1217	530
Interactive	717	1030
Cohesive	421	1326

4.2 Training and test data preparation

The classification of texts has been the target of several educational works over the last years. In the systematic review of the literature presented in [11], 343 studies that applied text mining techniques in educational problems were selected, from which 109 (31.77%) studies focus on text classification. These studies applied machine learning algorithms that used a previously labeled training set to generate a model capable of predicting the correct labels of examples whose labels were unknown (future cases). Therefore, the data set adopted in this study was divided into training and test sets; the first (training) one formed by the five initial offerings of the course (winter 2008, fall 2008, summer 2009, fall 2009, winter 2010) and the second (test) for last offer (winter 2011) according to the recommendations by Farrow et al. [10]. As shown in Table 4, the training group had 1,510 (86%) posts and the test group with 237 (14%) posts. The negative and positive classes presented approximate distributions in the Interactive and Cohesive categories, with a greater difference in class distribution only for the Affective category.

Table 4: Distribution of posts in training and testing groups

	Group	negative (0)	positive (1)	Total
Affective	Train	1038 (69%)	472 (31%)	1510
	Test	179 (76%)	58 (24%)	237
Interactive	Train	620 (41%)	890 (59%)	1510
	Test	97 (41%)	140 (59%)	237
Cohesive	Train	362 (24%)	1148 (76%)	1510
	Test	59 (25%)	178 (75%)	237

4.3 Feature extraction

This work combines traditional text mining features, like word frequencies, with the linguistic tools LIWC and Coh-Metrix to extract indications of social presence from textual contributions. These tools are widely validated in the literature as suitable extractors of cohesive, psychological, and social aspects of texts [37]. The remainder of the subsection provides an overview of these features as well as justifications for their use to automate the identification of social presence in asynchronous online discussions.

4.3.1 LIWC features. Linguistic Inquiry Word Count (LIWC) is a linguistic text analysis resource that extracts 93 features divided into the categories as follows: summary of language variables, linguistic dimensions, grammar, and psychological processes. The last category has words that express Affective Processes, Positive Emotion, Negative Emotion, Anger, Sadness, Social Processes, and others. By relating the definition of social presence (students' ability to demonstrate that they are real people [37]), and the social indicators proposed in the CoI [12] model, we hypothesized the use of this language resource might contribute to the construction of classifiers capable of correctly discriminating messages with or without evidence of social presence. LIWC was already used in a previous study focusing on automating the identification of cognitive presence, reaching high levels of accuracy [25, 35]. Thus, in this study, the LIWC 2015 version was used to extract features from the messages in our dataset.

4.3.2 Coh-Metrix features. According to [19], Coh-Metrix uses lexicons, POS Tagger, LSA, among other natural language processing (NLP) techniques to analyze cohesion and textual coherence. Several studies have reported good results when using Coh-Metrix to generate text cohesion indicators [18, 30]. Therefore, we hypothesized that the existence/absence of social presence indicators could be related to the index of textual cohesion and complexity proposed in Coh-Metrix. For instance, in the message "You got it right!" there is a hint of social interaction through the pronominal "You" cohesion. Hence, when considering the use of cohesive words as a way of demonstrating group projection, Coh-Metrix was also adopted as a component of the analysis in the process of automatic identification of social presence.

4.3.3 Word frequency. Finally, we adopt a bag-of-words vector, a traditional text mining technique, as the last set of features to extract social presence. More specifically, we adopt the method which transforms the textual documents (in this case, online discussion message) into an array consisting of the terms count. A problem found in this type of technique is the high dimensionality of the generated vector of features because as the text itself is used as discriminant the size of the vocabulary of the documents will correspond to the size of the matrix. Therefore, three techniques were used to reduce the dimensionality of the term count matrix. The first was a spelling correction of the texts. The second was the removal of stopwords, which consisted of removing words of little significance in a text such as articles, conjunctions, and prepositions [29]. Finally, the last technique was stemming, which seeks to reduce words to their respective radicals [36]. For example, the words "engineer" and "engineering" become "engine".

4.4 Data preprocessing

The main focus of machine learning is the creation of inductors based on past data (training set) and with the ability to generalize learned patterns to future examples [6]. One of the challenges in machine learning is dealing with datasets that have unbalanced class distributions [20]. According to He and Garcia [20], in these cases, the generated inductors usually prioritize the majority class. As presented in section 4.2, the negative and positive classes in all categories (Affective, Interactive, and Cohesive) were unbalanced. The Cohesive category of social presence was particularly highly unbalanced, where the negative class presented approximately 25% of the data and the positive class approximately 75%, suggesting that the classifier could prioritize the positive class. According to Chawla et al. [7], there are basically two approaches to solving the data imbalance problem: (i) cost-sensitive classification, i.e., penalizing the majority and minority prediction errors in different ways in order to force the algorithm prioritizing classes with fewer examples; and (ii) resampling of the data, with the options of undersampling the majority class in order to balance the number of examples which has the negative point of data loss or oversampling of the minority class. Thus, we decided to use the oversampling technique in the data of the training sets of each category. For this, we adopted the SMOTE algorithm, which is used in several works to create artificial data of the minority class (oversampling) [20].

4.5 Model Selection and Evaluation

To address research question 1, we trained three machine learning classifiers – one for each category of social presence. Recent studies show that combining machine learning classifiers tends to yield better results compared to those obtained by individual classifiers [6]. Ensembles can be performed by combining several distinct algorithms or by using only one algorithm with different training sets. Random Forest, one of the most widely used ensembles in the literature, combines decision trees using a technique called bagging which randomly samples characteristics. In other words, each tree is trained with different views (feature sets) of the same problem. Finally, all decisions are combined using the majority vote decision rule [6]. Random Forest is also often used to estimate the importance of an individual feature where it considers metric Mean decrease gini impurity index (MDG) [6], categorizing this technique as a white-box algorithm. Thus, Random Forest was the algorithm chosen for this study.

In educational research, to measure the performance of a supervised machine learning algorithm, the accuracy and Cohen's kappa[8] metrics are used [22, 35]. Hence, these metrics were also used in this work.

As highlighted in [6], the main parameters of the Random Forest algorithm are the number of input variables randomly chosen from each division (max_features) and the number of trees in the forest (n_estimators). To optimize the final performance, we performed a tunning in the parameters (max_features and n_estimators) of the Random Forest classifier through executions using the cross-validation technique for each training set (Affective, Interactive, and Cohesive). Each validation fold consisted of one of the course offerings in the respective training set (winter 2008, fall 2008, summer 2009, fall 2009, winter 2010). The first parameter to be adjusted was *max_features*. The literature considers that the Random Forest performance stabilizes after a certain number of trees [25]. Therefore, we set the number of trees at 1500 to ensure the convergence of the algorithm and, consequently, choosing the best *max_features* parameter. To obtain a deterministic behavior during parameter setting, one seed (five) was established. In each cross-validation execution, a total of 140 values were verified for the *max_features* parameter, which was set at random, respecting the maximum number of possible characteristics (6418) and without repetition. At the end of the executions, the average performances and the standard deviations obtained for each parameter were reported. In Figure 1, it is shown that the average accuracy obtained in each of the three categories (Affective, Interactive, and Cohesive) began to stabilize when the number of characteristics was about 2000.

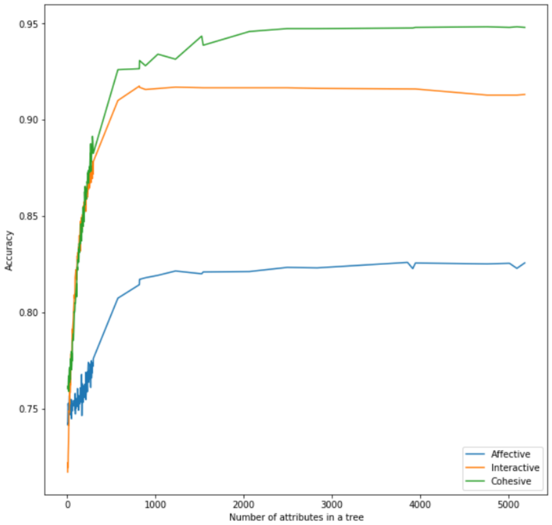


Figure 1: Random forest parameter tuning results

After setting the *max_features parameter*, the next step was to estimate the appropriate number of trees *n_estimators* for the problem. For this, the (Out-of-Bag) OOB error was used and it was calculated in a set of observations that were not used to build the current tree. As shown in Figure 2, it was not possible to notice drastic differences when changing the n_estimators parameter, so we set the number of trees to 800.

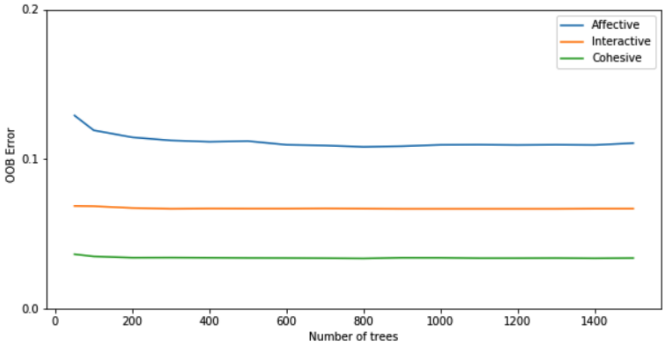


Figure 2: Best random forest configuration performance

Besides, the development of the classifier, we also provide additional insights on which features are more relevant for each category. One popular measure to calculate the feature importance from a Randon Forest is Mean Decrease Gini (MDG) which is based on the reduction in Gini impurity measure [6]. In this paper, we adopted MDG address the second research question, the evaluation of the relevance of different features to the outcome of our classifiers.

4.6 Epistemic Network Analysis

We applied Epistemic Network Analysis (ENA) [42] to address research question 3 and provide insights into the validity of the results produced by the proposed classifier. ENA is a graph-based technique used for analysis of associations between different concepts (called *codes*) used for coding textual datasets. Within ENA, a network of relationships among different *codes* is created for each *unit of analysis* (e.g., student). Two codes are considered related if they co-occur in the same chunk of text, called *stanza* (or *conversation*).

In this work, we reproduced the experiment proposed by Rolim et al. [39] using the automatic classifier to generate the social presence codes. Rolim et al. [39] adopted LDA to extract topics from the dataset (15 topics in total). Then, the authors used the students as *units of analysis*, social presence categories and course topics as *codes*, and individual students' discussion messages as *stanza*. Again, the goal of this analysis is to compare the graphs plotted based on the humans' annotation, and those automatically generated by the proposed classifier.

ENA mainly provides three graphical outcomes: (1) Projection graph; (2) Epistemic Network; and (3) Subtraction network. A relevant characteristic of ENA is that all these graphical outcomes are represented in the same bi-dimensional space, called *analytic space*, composed by X and Y axes. So it is possible to analyze different aspects at the same time. Each graph produced by ENA has elements to be analyzed, as follows: (i) **Projection graph** presents the units of analysis (i.e., the students in the current study) distributed in the analytic space. In this graph, each unit of analysis is represented as a pair (x,y) that are related to their position in the axes X and Y. (ii) **Network graph** is undirected graph and contains three important elements: the *size* of the nodes represents the frequency of occurrence of the nodes; the *nearness* of the node represents the similarity among them; and the *strength* of the code relationship represents the frequency of their co-occurrence; and (iii) **Subtraction network** captures the differences between two ENA networks and only shows edges that are different between the two networks. Moreover, it presents the same aspects of the regular ENA networks.

The graphs produced by ENA provide a promising approach to assessing the validity of the results of the classifier, mainly for allowing us to compare the results with our previous work [39], which provides insights into how the students' social presence was related to the course topics; moreover, all relationships developed by the students can be quantified. Several other studies have already applied ENA in the context of CoI. For example, Rolim et al. [40] presents an approach that uses ENA to understand the relationships between cognitive and social presences and uncover how students progress over time in their social inquiry. In another paper, Rolim et al. [39] analyze, using ENA, the relationships between course topics and indicators of social presence categories.

Aiming to quantify the comparison of the proposed classifier output with the manual codded data using ENA, we used the Pearson correlation coefficient (PCC), which measures the linear relationship between two variables [5]. In our study, we measured the correlation between the projections, the pair (x,y) of each student, comparing the data generated with human-annotated and automatically generated data. Thus, we wanted to measure the extent to which the epistemic networks with automatically and manually assigned codes were similar.

In order to measure the PCC between the subtraction network of the automatically generated and manually coded data we used the following three variables: (i) two variables related to the nodes positions, divided into the position in relation to axes X and Y; and (ii) the strength of the links between the nodes. In this case, we report three values of PCC.

5 RESULTS

5.1 Model training and evaluation – RQ1

Initially, we evaluated the influence of parameter tuning in the final classification. Table 5 shows the average results reported of the performance of the Random Forest classifiers with the default parameters and tuned parameters using the training set and the cross-validation approach. The results, in terms of accuracy, increased by 12.5%, 22.6% and 21.79% for the affective, interactive and cohesive categories, respectively. Regarding Cohen's kappa, it achieved even higher improvements of 145%, 66% and 114.6% for the same three categories. These results demonstrate the importance of fine-tuning the algorithm parameters.

Table 5: Random forest parameter tuning results

Category	Optimization	Accuracy	Kappa
Affective	Default parameters	0.72 (0.05)	0.20 (0.10)
	Tuned parameters	0.81 (0.04)	0.49 (0.08)
Interactive	Default parameters	0.75 (0.03)	0.50 (0.05)
	Tuned parameters	0.92 (0.02)	0.83 (0.05)
Cohesive	Default parameters	0.78 (0.06)	0.41 (0.11)
	Tuned parameters	0.95 (0.03)	0.88 (0.06)

After parameter optimization, the Random Forest classifier was ran ten times for each social presence category. In each execution, the training set examples (1.510 posts – initial five runs of the courses in our data) were used to generate a binary classifier, and its generalization capacity was verified in the respective test sets (237 posts – the last run of the courses in our data). Thus, the following results were obtained for the test sets of each category: Affective – accuracy = 0.80 (0.01) and Cohen's kappa 0.34 (0.02); Interactive – accuracy = 0.92 (0.0) and Cohen's kappa 0.85 (0.0); and Cohesive – accuracy = 0.97 (0.0) and Cohen's kappa 0.93 (0.0). Despite the high dimensionality of the feature vectors and fine parameter adjustments (max_features and n_estimators), the results achieved in the test sets approximate the average accuracy obtained in the validation sets (cross-validation). Therefore, it shows that there were no overfitting in the training set. Instead, the models demonstrated good generalizability for unknown examples.

Table 6 shows the confusion matrix generated for each category. Corroborating with the average values of accuracy and Cohen's kappa presented, it is possible to notice that the highest occurrences of false positives occurred in the Affective class achieving 37 examples. On the other hand, there were only 9% occurrences of false positives for the Interactive category and 2% for Cohesive.

Table 6: Confusion matrix for the best performing models

	Affective		Interactive		Cohesive	
	neg*	pos*	neg	pos*	neg	pos*
neg*	170	9	91	6	56	3
pos*	37	21	12	128	3	175

* pos = positive and neg = negative

5.2 Feature importance analysis – RQ2

Although the same feature vectors were used to discriminate the classes (positive and negative) of the three categories, each classifier considered different variables as the most important. The Random Forest uses the Mean Decrease Gini impurity index (MDG) measure to define the degree of relevance of a feature. Tables 7, 8 and 9 present the top-15 features for the classifier of each category (Affective, Interactive and Cohesive).

The most important set of variables for the Affective category shows six from the word frequency, eight LIWC, and one Coh-Metrix features. Besides, the two most important variables are the words "hope" and "happi" (without applying the stemming technique, "happy"), reaching 11.68 and 10.33 and MDG, respectively. It is also noteworthy that two characteristic cm.WRDPRP1s concerning the number of first-person pronouns (e.g., I, me, mine) is in the top-15 set of features.

The most predictive features of the Interactive category were divided into word frequency (three), LIWC (eight), and Coh-Metrix (four). Table 8 shows that the most important was the number of Question marks which achieved MDG of 44.2. The presence of the word "agre" (stemmed from the word agreement) and the liwc.assent which measures the agreement was also noticeable.

Finally, Table 9 presents the main features of the Cohesive class being 10 from word frequency, five from LIWC, and none from Coh-Metrix. It may be highlighted that the presence of words commonly used to greet ("hi" - MDG of 54.18 and "hello" MDG of 3.64) and of socially biased variables like liwc.affiliations (e.g., ally, friend, and social) and Liwc.social (e.g., mate, talk, and they). Therefore, the most predictive feature listings for each category demonstrate the importance of the three language resources used in this study.

Table 7: Fifteen most important variables for the Affective category according to MDG

Variable	Description	MDG	Negative	Positive
hope	Word frequency	11.68	0.02 (0.14)	0.24 (0.46)
happi	Word frequency	10.33	0.0 (0.05)	0.22 (0.51)
liwc.i	1st pers singular	3.88	2.3 (2.04)	3.39 (2.19)
cm.WRDPRP1s	Incidence score of pronouns, first person, single form	3.31	22.86 (19.94)	33.53 (21.68)
liwc.Apostro	Number of Apostrophes	2.07	0.7 (1.25)	1.08 (1.4)
liwc.Exclam	Number of Exclamation marks	1.18	0.21 (0.84)	0.66 (3.58)
liwc.negemo	Number of negative emotion	1.18	0.63 (1.04)	0.83 (1.03)
work	Word frequency	1.14	0.25 (0.64)	0.54 (0.98)
bb	Word frequency	1.14	0.35 (0.48)	0.55 (0.5)
experi	Word frequency	1.07	0.09 (0.37)	0.26 (0.78)
develop	Word frequency	0.83	0.53 (1.1)	0.79 (1.43)
liwc.power	words with power idea	0.79	1.47 (1.59)	1.55 (1.33)
liwc.hear	hear	0.72	0.29 (0.66)	0.41 (0.79)
liwc.negate	Number of negations	0.7	0.98 (1.16)	1.35 (1.17)
liwc.we	1st pers plural	0.62	0.21 (0.65)	0.37 (0.81)

Table 8: Fifteen most important variables for the Interactive category according to MDG

Variable	Description	MDG	Negative	Positive
liwc.QMark	Number of question marks	44.2	0.07 (0.4)	1.5 (1.38)
agre	Word frequency	9.21	0.01 (0.1)	0.25 (0.5)
present	Word frequency	5.98	0.54 (0.9)	1.29 (1.23)
liwc.assent	Number of assent	1.77	0.26 (1.43)	0.3 (0.71)
liwc.auxverb	Auxiliary verbs	0.81	7.58 (4.2)	8.82 (2.77)
liwc.you	2nd person	0.8	1.57 (2.91)	2.25 (1.88)
liwc.Period	Number of periods	0.79	7.64 (5.2)	5.51 (2.54)
liwc.AllPunc	Number all punctuation	0.78	17.22 (14.72)	13.3 (10.48)
cm.WRDPOLc	Number of senses (core meanings) of a word	0.73	3.72 (0.83)	3.98 (0.57)
cm.WRDPRP2	Incidence score of pronouns, second person	0.69	15.56 (28.84)	22.26 (18.65)
liwc.Dic	Dictionary words	0.65	75.83 (14.12)	80.5 (7.4)
cm.DESWLltd	Mean number of letters in the words within the text	0.64	3.34 (1.75)	2.98 (0.81)
cm.SYNSTRUTt	Proportion of intersection tree nodes between all sentences	0.51	0.07 (0.06)	0.06 (0.03)
did	Word frequency	0.49	0.0 (0.04)	0.15 (0.42)
liwc.function.	Total function words	0.47	44.29 (12.67)	48.66 (6.74)

Table 9: Fifteen most important variables for the Cohesive category according to MDG

Variable	Description	MDG	Negative	Positive
hi	Word frequency	54.18	0.0 (0.07)	0.86 (0.35)
liwc.affiliation	Number of affiliations	10.48	0.98 (2.46)	1.97 (2.17)
regard	Word frequency	5.02	0.05 (0.23)	0.29 (0.52)
hello	Word frequency	3.64	0.0 (0.0)	0.05 (0.21)
cheer	Word frequency	1.12	0.0 (0.05)	0.07 (0.26)
bb	Word frequency	1.01	0.52 (0.5)	0.38 (0.49)
liwc.social	Social processes	0.96	5.49 (5.34)	7.12 (3.94)
thank	Word frequency	0.76	0.6 (0.64)	0.77 (0.68)
grant	Word frequency	0.55	0.1 (0.3)	0.03 (0.18)
liwc.Clout	Number of clout-related words	0.5	48.4 (20.16)	58.07 (19.09)
present	Word frequency	0.5	0.76 (1.08)	1.05 (1.19)
hey	Word frequency	0.32	0.0 (0.0)	0.01 (0.1)
liwc.Apostro	Number of Apostrophes	0.31	1.23 (1.77)	0.68 (1.1)
liwc.we	1st pers plural	0.3	0.19 (0.68)	0.28 (0.72)
sy	Word frequency	0.3	0.04 (0.19)	0.01 (0.11)

5.3 Epistemic Network Analysis – RQ3

We reproduced the work by Rolim et al. [39] in order to evaluate the impact on the validity of the automatic classification in producing codes that are used in ENA. Figure 3 presents the projection graph for both manually and automatically coded datasets. In this graph, each node represents a student, and the squares are the mean values for the two groups; control and treatment group are represented as red and blue nodes, respectively. Although the first two SVD dimensions (i.e., x- and y-axes) small differences in explained variance, it is possible to visually identify a high level of similarity between the two graphics (Fig. 3a and 3b); for instance, the positions of the group centroids (red and blue squares), and the distribution of the units (red and blue points) along the axis. The Mann-Whitney test showed that along the X-axis the students from the control and treatment groups were statistically significantly different at the alpha=0.05 level for both manually (U=1497.00, p=0.00, r=0.76) and automatically coded (U=164.00, p=0.00, r=0.81) data with very similar effect sizes as reflected by the r values.

Moreover, we calculated the PCC between the same students with the data manually coded and the automatically generated. As it student is projected as a pair (x,y) we analyzed the PCC variables separately for each dimension (X and Y axes). The final values of PCC reached 0.93 and 0.84 for axes X and Y, respectively, which demonstrate a high correlation between the distribution of the students.

Figure 4 shows the subtraction network between the control (red) and treatment(blue) groups for both manually and automatically coded data. Visually analyzing, we can highlight the arrangement of the codes, where the course topics are mostly in the bottom-left corner of both networks, whereas the social presence categories are plotted in the upper-central part o both networks. Also, the connections between codes have a similar strength among the codes. Figure 4b also reveals a slight change in the sizes of “Interactive”, “Cohesive” and “const.meth” codes. Another difference is the position of the three social presence categories with the most significant one for the “Affective” code, which before was positioned closer to the course topics, and in Figure 4b it is closer of the other two categories; closeness in the project graphs means higher similarity.

We also evaluated the PCC between links strength and the node positions in the subtraction network. In this case, we did not evaluate the individual students from the projection graph (as reported previously in this section), but the position of the codes in Figure 4. The results for the positions of the nodes achieved 0.96 and 0.89 for the axes X and Y, respectively. Regarding the PCC of the strength between codes the results reached 0.93. The high PCC values indicate the convergence of the two networks.

6 DISCUSSION

In addressing research question 1, the evaluation of the automatic classification of social presence revealed that the combination of traditional text mining features and word counts extracted from LIWC and Coh-Metrix were effective in classifying online discussion messages message in all categories (affective, interactive and cohesive). Cohen’s κ of 0.49, 0.83 and 0.88, for affective, interactive and cohesive, respectively, represent a medium to substantial interrater agreement [28], and in two out of three categories it is above 0.70, which is the CoI research commonly used as the threshold limit required before manual coding results are considered valid. The optimization of the max_features (i.e., the number of attributes used in each tree of the forest) and n_estimators (i.e., the number of trees used in each iteration) parameters improved the final result in all cases (Table 5).

Although we did not find any other related work which performed a similar analysis of social presence to compare to, it is important to mention that the approach presented here reached accuracy results better than the classifiers of cognitive presence developed for English [24, 25, 45].

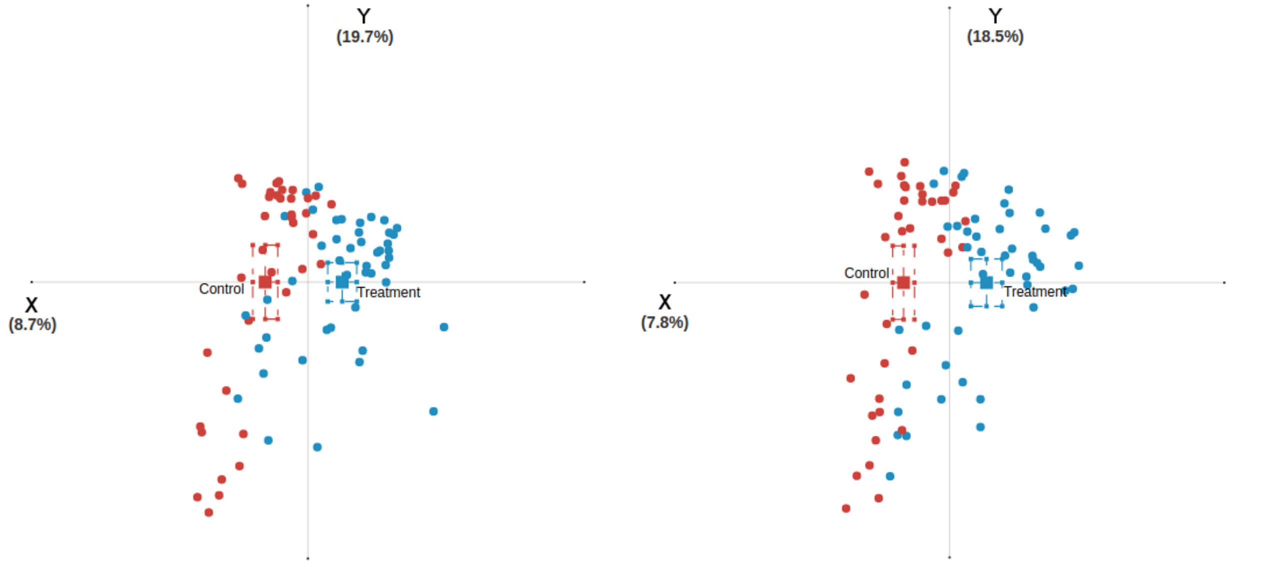
In addressing research question 2, this study conducted a detailed analysis of the features used. By analyzing the features provided in tables 7, 8 and 9, we can draw two conclusions: (i) for every category, there were feature related to word frequency and the tools LIWC and Coh-Metrix, showing the importance of both aspects; (ii) although the features related to word frequency could lead to overfitting depending on the domain, in the case of the current study, the words with high information gain were general ones like hope, happy, hear, agree, hi, hello, among others. Thus, it decreases the chances of overfitting because these words can happen in messages of different domains.

The analysis of the feature importance also highlighted a possible correlation between the main features identified in this study and the indicators considered the most predictive of social presence [41]. For instance, among those selected by the Affective category classifier, the features: *hope*, *happi* (happy), liwc.exclam (number of exclamation points) and liwc.negemo (number of negative emotions) are related to the expression of emotions and the use of humor. While the feature cm.WRDPRP1s (pronoun incidence score, first-person singular) may be associated with the self-disclosure indicator since the student demonstrates self-disclosure when presenting details of life outside the home, classroom, or express vulnerability [41].

For the Interactive category, the features liwc.you (2nd person word count) and cm.WRDPRP2 (second person pronoun incidence score) were related to the indicators of the Interactive category (see Table 1) by citing and referencing openly other messages or people in the discussion. Moreover, in nonverbal interaction, when asking a question it is common to use the question punctuation mark. Thus, the Interactive category indicator named Asking Questions is represented in the list of most essential features by the features liwc.QMark (number of question marks) and liwc.interrog (number of interrogative sentences). Another demonstration of interaction (based on the CoI model) is expressions of agreement students' messages; in this respect, the central features were the word *agre* (agree) and the number of nods per post (liwc.assent).

Finally, the presence of the feature liwc.we (number of first-person plural words) in table 9 corroborate the relevance of using inclusive pronouns (us, ours) as a way of demonstrating group cohesion. Another indicator of the cohesive category, the demonstration of salutations, can be recognized by the characteristics *hi*, *hello*, liwc.affiliation (number of affiliations) and liwc.social (number of social processes).

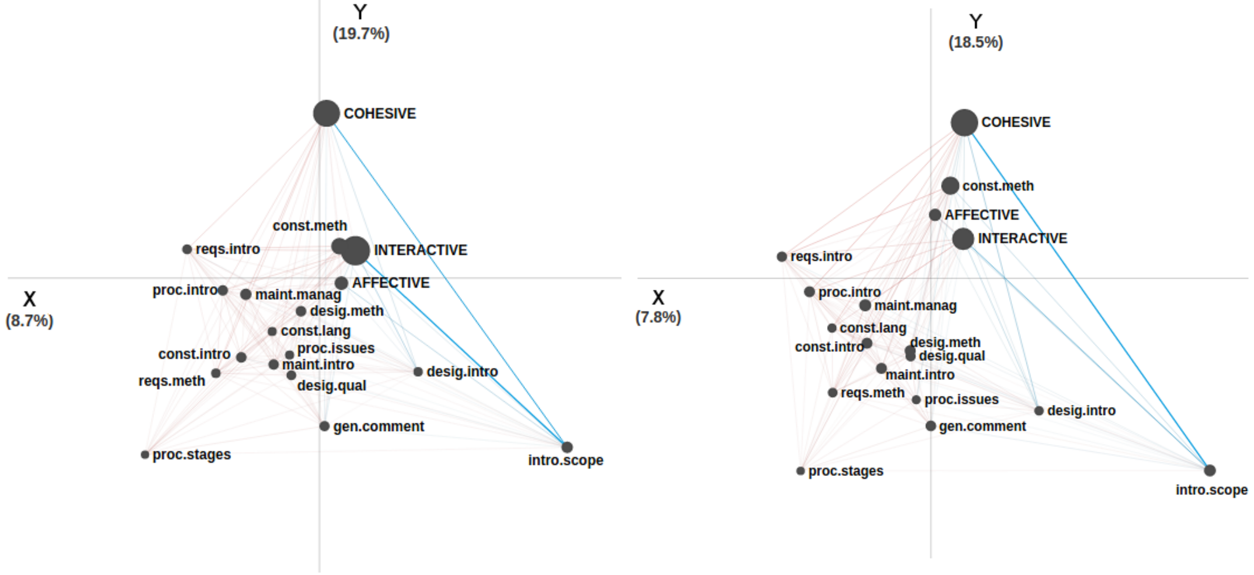
In addressing research question 3, we adopted ENA to investigate the similarities between associations between discussion topics and social presence categories as generated with the manually and automatically coded data. As can be seen in section 5.3, we obtained similar results after performing ENA with manually and automatically assigned codes for both individual projections of students and the subtraction network of the two groups. Also, we demonstrated that these outcomes are correlated using Pearson Correlation Coefficient. Thus, we conclude that analyses performed with automatically assigned codes can reproduce the results of analyses based on manually coded data on a reasonable level of confidence that can preserve structural properties of the associations of social presence with other relevant constructs [33]. This also offers additional reassurance in the validity of the results of analyses that are based on automatically coded messages.



(a) Using manual codded labels.

(b) Using automatic generated labels.

Figure 3: ENA projection of the networks of the students related to social presences and course topics between control (red) and treatment (blue) groups.



(a) Using manual codded labels.

(b) Using automatic generated labels.

Figure 4: Subtraction mean network between control (red) and treatment (blue) groups.

7 CONCLUSIONS

This paper has three contributions. First, the proposal of three binary Random Forests classifiers, using the LIWC, Coh-Metrix and word frequency linguistic resources, to automatically classify online discussion messages into social presence categories (Affective, Interactive and Cohesive). Every category reached Cohen's kappa values of more than 0.49, a medium to substantial inter-rater agreement. Second, the results provide insights into the psycho- and socio-linguistic features that are more relevant for each social presence indicator, linking each of them with the CoI literature. These results additionally clarify the nature of each social presence indicator, which have not been previously reported in the literature. Finally, the use of automatically coded discussion messages in analysis of associations of social presence with other relevant constructs (e.g., discussion topics) produced nearly identical results to the analyses performed with manually assigned codes of social presence.

Despite promising results, some limitations can be identified, such as the small number of message examples used in the current study (1.747 posts). Next, the training and test sets were divided based on different offerings of the same course, making it difficult to generalize the results presented to other contexts. Finally, using word frequency to compose feature vectors can mean a strong bias of the classification models created in the training context.

Future work should seek to optimize the approach proposed in this study to reduce the dimensionality of the feature vectors while maintaining the promising results already obtained, which is important to avoid overfitting. Besides, we also aim to conduct experimentation with the approach and possible evaluation with larger sample sizes composed of data from different domains.

ACKNOWLEDGEMENT

This research was partially supported by the fellowship 310060/2018-4 of the National Council for Scientific and Technological Development – CNPq.

REFERENCES

[1] Terry Anderson and Jon Dron. 2010. Three generations of distance education pedagogy. *The International Review of Research in Open and Distance Learning* 12, 3 (2010), 80–97.

[2] Terry Anderson, Liam Rourke, D. Randy Garrison, and Walter Archer. 2001. Assessing Teaching Presence in a Computer Conferencing Context. *Journal of Asynchronous Learning Networks* 5 (2001), 1–17.

[3] J Ben Arbaugh, Martha Cleveland-Innes, Sebastian R Diaz, D Randy Garrison, Philip Ice, Jennifer C Richardson, and Karen P Swan. 2008. Developing a community of inquiry instrument: Testing a measure of the community of inquiry framework using a multi-institutional sample. *The internet and higher education* 11, 3-4 (2008), 133–136.

[4] Jason J Barr. 2016. Developing a Positive Classroom Climate. *IDEA Center, Inc.* (2016).

[5] Jacob Benesty, Jingdong Chen, Yiteng Huang, and Israel Cohen. 2009. Pearson correlation coefficient. In *Noise reduction in speech processing*. Springer, 1–4.

[6] Leo Breiman. 2001. Random forests. *Machine learning* 45, 1 (2001), 5–32.

[7] Nitesh V Chawla, Nathalie Japkowicz, and Aleksander Kotcz. 2004. Special issue on learning from imbalanced data sets. *ACM Sigkdd Explorations Newsletter* 6, 1 (2004), 1–6.

[8] Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement* 20, 1 (1960), 37–46.

[9] Roisin Donnelly and John Gardner. 2011. Content analysis of computer conferencing transcripts. *Interactive learning environments* 19, 4 (2011), 303–315.

[10] Elaine Farrow, Johanna Moore, and Dragan Gašević. 2019. Analysing discussion forum data: a replication study avoiding data contamination. In *LAK' 2019*. 170–179.

[11] Rafael Ferreira-Mello, Máverick André, Anderson Pinheiro, Evandro Costa, and Cristobal Romero. 2019. Text mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* (2019), e1332.

[12] D. Randy Garrison, Terry Anderson, and Walter Archer. 1999. Critical Inquiry in a Text-Based Environment: Computer Conferencing in Higher Education. *The Internet and Higher Education* 2, 2-3 (1999), 87–105.

[13] D. Randy Garrison, Terry Anderson, and Walter Archer. 2001. Critical Thinking, Cognitive Presence, and Computer Conferencing in Distance Education. *American Journal of Distance Education* 15, 1 (2001), 7–23.

[14] D. Randy Garrison, Terry Anderson, and Walter Archer. 2010. The first decade of the community of inquiry framework: A retrospective. *The Internet and Higher Education* 13, 1-2 (2010), 5–9.

[15] D Randy Garrison and J Ben Arbaugh. 2007. Researching the community of inquiry framework: Review, issues, and future directions. *The Internet and Higher Education* 10, 3 (2007), 157–172.

[16] Dragan Gašević, Olusola Adesope, Srećko Joksimović, and Vitomir Kovanović. 2015. Externally-facilitated regulation scaffolding and role assignment to develop cognitive presence in asynchronous online discussions. *The internet and higher education* 24 (2015), 53–65.

[17] Dragan Gašević, Vitomir Kovanović, and Srećko Joksimović. 2017. Piecing the learning analytics puzzle: a consolidated model of a field of research and practice. *Learning: Research and Practice* 3, 1 (2017), 63–78.

[18] Arthur C Graesser, Danielle S McNamara, and Jonna M Kulikowich. 2011. Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational researcher* 40, 5 (2011), 223–234.

[19] Arthur C Graesser, Danielle S McNamara, Max M Louwerse, and Zhiqiang Cai. 2004. Coh-Metrix: Analysis of text on cohesion and language. *Behavior research methods, instruments, & computers* 36, 2 (2004), 193–202.

[20] Haibo He and Edwardo A Garcia. 2009. Learning from imbalanced data. *IEEE Transactions on knowledge and data engineering* 21, 9 (2009), 1263–1284.

[21] Srećko Joksimović, Dragan Gašević, Vitomir Kovanović, Bernhard E Riecke, and Marek Hatala. 2015. Social presence in online discussions as a process predictor of academic performance. *Journal of Computer Assisted Learning* 31, 6 (2015), 638–654.

[22] Vitomir Kovanovic, Srecko Joksimovic, Dragan Gasevic, and Marek Hatala. 2014. What is the source of social capital? The association between social network position and social presence in communities of inquiry. In *Workshop at Educational Data Mining Conference*. EDM.

[23] Vitomir Kovanović, Dragan Gašević, and Marek Hatala. 2014. Learning analytics for communities of inquiry. *Journal of Learning Analytics* 1, 3 (2014), 195–198.

[24] Vitomir Kovanović, Srećko Joksimović, Dragan Gašević, and Marek Hatala. 2014. Automated cognitive presence detection in online discussion transcripts. In *LAK'14*. Indianapolis, IN.

[25] Vitomir Kovanović, Srećko Joksimović, Zak Waters, Dragan Gašević, Kirsty Kitto, Marek Hatala, and George Siemens. 2016. Towards automated content analysis of discussion transcripts: A cognitive presence case. In *LAK'16*. ACM, New York, NY, USA, 15–24.

[26] Kadir Kozan and Jennifer C Richardson. 2014. Interrelationships between and among social, teaching, and cognitive presence. *The Internet and higher education* 21 (2014), 68–73.

[27] Klaus Krippendorff. 2018. *Content analysis: An introduction to its methodology*. Sage publications.

[28] J Richard Landis and Gary G Koch. 1977. The measurement of observer agreement for categorical data. *biometrics* (1977), 159–174.

[29] Rachel Tsz-Wai Lo, Ben He, and Iadh Ounis. 2005. Automatically building a stopword list for an information retrieval system. In *Journal on Digital Information Management*, Vol. 5. 17–24.

[30] Philip M McCarthy, Gwyneth A Lewis, David F Dufty, and Danielle S McNamara. 2006. Analyzing Writing Styles with Coh-Metrix.. In *FLAIRS Conference*. 764–769.

[31] Thomas E. Mcklin. 2004. *Analyzing Cognitive Presence in Online Courses Using an Artificial Neural Network*. Ph.D. Dissertation. Atlanta, GA, USA. Advisor(s) Harmon, Stephen W. AAI3190967.

[32] Danielle S McNamara, Arthur C Graesser, Philip M McCarthy, and Zhiqiang Cai. 2014. *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press.

[33] Samuel Messick. 1995. Standards of validity and the validity of standards in performance asessment. *Educational measurement: Issues and practice* 14, 4 (1995), 5–8.

[34] Natalie B Milman. 2015. Distance education. (2015).

[35] Valter Neto, Vitor Rolim, Rafael Ferreira, Vitomir Kovanović, Dragan Gašević, Rafael Dueire Lins, and Rodrigo Lins. 2018. Automated analysis of cognitive presence in online discussions written in portuguese. In *European Conference on Technology Enhanced Learning*. Springer, 245–261.

[36] Viviane Moreira Orengo and Christian Huyck. 2001. A stemming algorithm for the portuguese language. In *Proceedings. Eighth International Symposium on.* IEEE, 186–193.

[37] James W Pennebaker, Ryan L Boyd, Kayla Jordan, and Kate Blackburn. 2015. *The development and psychometric properties of LIWC2015*. Technical Report.

[38] Oleksandra Poquet, Vitomir Kovanović, Pieter de Vries, Thieme Hennis, Srećko Joksimović, Dragan Gašević, and Shane Dawson. 2018. Social presence in massive open online courses. *International Review of Research in Open and Distributed Learning* 19, 3 (2018).

[39] Vitor Rolim, Rafael Ferreira Leite de Mello, Vitomir Kovanovic, and Dragan Gaševic. 2019. Analysing Social Presence in Online Discussions Through Network and Text Analytics. In *2019 IEEE 19th International Conference on Advanced Learning Technologies (ICALT)*, Vol. 2161. IEEE, 163–167.

[40] Vitor Rolim, Rafael Ferreira, Rafael Dueire Lins, and Dragan Găsević. 2019. A network-based analytic approach to uncovering the relationship between social and cognitive presences in communities of inquiry. *The Internet and Higher Education* 42 (2019), 53–65.

[41] Liam Rourke, Terry Anderson, D. Randy Garrison, and Walter Archer. 1999. Assessing Social Presence In Asynchronous Text-based Computer Conferencing. *The Journal of Distance Education* 14, 2 (1999), 50–71.

[42] David Williamson Shaffer, David Hatfield, Gina Navoa Svarovsky, Padraig Nash, Aran Nulty, Elizabeth Bagley, Ken Frank, André A. Rupp, and Robert Mislevy. 2009. Epistemic Network Analysis: A Prototype for 21st-Century Assessment of Learning. *International Journal of Learning and Media* 1, 2 (2009), 33–53.

[43] Jan-Willem Strijbos. 2011. Assessment of (computer-supported) collaborative learning. *IEEE transactions on learning technologies* 4, 1 (2011), 59–73.

[44] Yla R Tausczik and James W Pennebaker. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology* 29, 1 (2010), 24–54.

[45] Zak Waters, Vitomir Kovanović, Kirsty Kitto, and Dragan Gašević. 2015. Structure matters: Adoption of structured classification approach in the context of cognitive presence classification. In *Information Retrieval Technology*. 227–238.

[46] C Xia, John Fielder, and Lou Siragusa. 2013. Achieving better peer interaction in online discussion forums: A reflective practitioner case study. *Issues in Educational Research* 23, 1 (2013), 97–113.