Longitudinal Analysis and Visualization of Participation in Online Courses Powered by Cohesion Network Analysis

Maria-Dorinela Sirbu, University Politehnica of Bucharest, maria.sirbu@cti.pub.ro
Mihai Dascalu, University Politehnica of Bucharest, Cognos Business Consulting S.R.L., mihai.dascalu@cs.pub.ro
Scott A. Crossley, Georgia State University, scrossley@gsu.edu
Danielle S. McNamara, Arizona State University, dsmcnama@asu.edu
Stefan Trausan-Matu, University Politehnica of Bucharest, Cognos Business Consulting S.R.L., stefan.trausan@cs.pub.ro

Abstract: Online learning environments are increasingly used by students and instructors. Cohesion Network Analysis (CNA) can be employed by instructors to analyze discourse structure in terms of cohesive links in order to model student participation and interactions in collaborative learning environments. This paper presents an extended longitudinal analysis together with corresponding visualizations of participation generated for an online math course, powered by CNA. Multiple interactive views centered on the evolution of participation and of interactions between students clustered into three layers are generated using the CNA indices provided by the *ReaderBench* framework. Two types of sociograms are used to show the interactions between learners in the two course weeks that exhibited extreme conditions, namely: a) week 6, when a dramatic decrease of participation was identified, and b) week 16, when the highest number of participants and contributions were logged. In addition to the views centered on participants, we introduce a heatmap depicting the evolution of keyword relevance over time, as well as a Chord diagram for visualizing concept maps based on semantic relatedness. Our analytics dashboard can be used by tutors to monitor students throughout the term and to better ascertain the correlation of course material, schedule, and deadlines with the participation of students, as well as their interactions among themselves and with the tutor.

Introduction

Online courses have become a useful learning platform due to their accessibility in terms of distance and time management. These platforms help students learn through collaboration and interactions with peers. However, tutors are faced with the difficult task of tracking a large audience; thus, automated analyses of students' participation are required. This paper relies on Cohesion Network Analysis (CNA) (Dascalu, McNamara, Trausan-Matu, & Allen, 2018) to automatically assess students' participation. Going beyond Social Network Analysis (SNA) (Scott, 2017) that considers the number of exchanged messages, CNA indices are used to evaluate interactions between participants based the cohesion graph (e.g., out-degree indices reflect a higher participation, in-degree indices indicate a higher collaboration). CNA considers the content of the discourse and participants' interactions by examining semantic cohesion of the links between student forum posts.

This article is a continuation of the studies conducted by Sirbu et al. (2018). The paper introduces a longitudinal analysis and detailed visualizations of participation useful for gaining a better understanding of how students interact during an online math course. The *ReaderBench* framework (Dascalu, Trausan-Matu, McNamara, & Dessus, 2015) was used to generate multiple CNA indices and to extract keywords, which were afterwards used to generate different views to model trends in student participation using the d3.js library, as well as concept maps relying on the semantic relatedness between keywords found in the discussion forums.

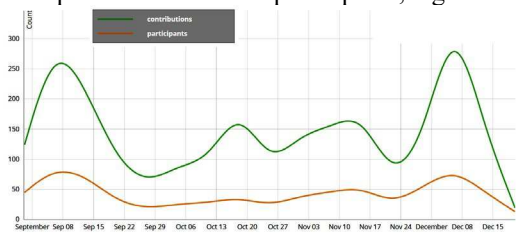
Method

Our corpus was collected from a discrete math course for undergraduate students in a Computer Science Department (Crossley, Barnes, Lynch, & McNamara, 2017). The course combined standard lectures with the support of online tools, which include a standard online question-answer forum. The total number of students on the forum was 250, out of which 169 had at least one contribution. Altogether the students produced more than 2000 posts. The class lasted for 18 weeks (Aug 23, 2013 - Dec 24, 2013).

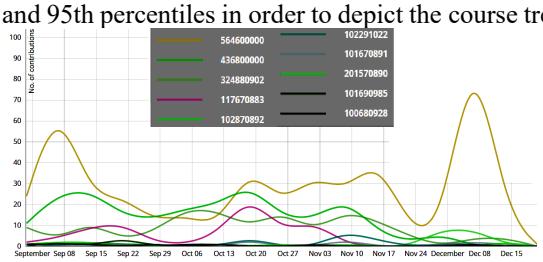
The *ReaderBench* framework was used to analyze the interactions between participants and to evaluate their contribution by generating multiple CNA indices. Using the method proposed by Nistor, Panaite, Dascalu, and Trausan-Matu (2017), we clustered the members based on CNA indegree and outdegree into three clusters in descending order of the average of their indegree scores defined as central, active, and peripheral scores. A hierarchical clustering algorithm was applied using the Ward Criterion for minimizing variance after the merging

of clusters. Based on the indices provided by the *ReaderBench* framework and the clustering method, we created multiples types of views and graphics to illustrate interactions among participants, trends in their participation over time, as well as the most relevant topics discussed.

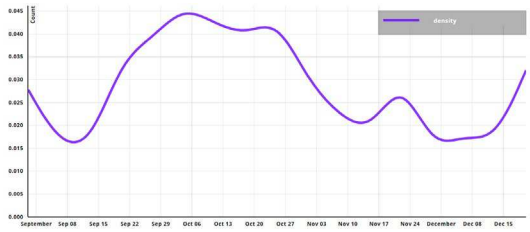
A longitudinal analysis based on CNA was performed to visualize different interactions patterns throughout the course. The visualizations from Figure 1 presents the evolution of participation using Line Chart and Trend Chart. Figure 1.a presents the number of participants (orange line) and contributions (green line) throughout the course, where the X axis represents the period, and Y axis the corresponding number of participants or contributions. A large number of participants/contributions can be observed at the beginning of the course (first 2 weeks), followed by a drastic decrease in the 6th week, and an accelerated increase in the final weeks, with a maximum number of participants and contributions in the 16th week. Figure 1.b presents the density of the community network calculated as $m/(n*(n - 1))$, where m is the number of edges between participants, and n is the number of participants. We also generated a line chart to observe changes in participation over time for specific individuals. Figure 1.c shows the evolution of participation for 10 randomly chosen students. Patterns similar to Figure 1.a are observed, including a decrease at the beginning of the course followed by an increased involvement towards the end (see participant with ID 564600000 marked with yellow). Figure 1.d presents the median weekly participation values for all participants, together with 5th and 95th percentiles in order to depict the course trends.



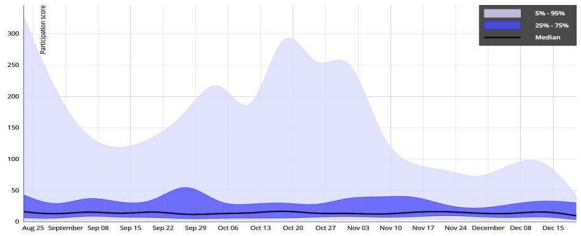
a. Overall evolution of the counts for participants and contributions.



c. Specific participants' evolution.



b. Density evolution.



d. Trends in terms of median CNA participation

Figure 1. Longitudinal Analysis of participation.

In addition to the previous participant-centered views, we examined which topics were most frequent by using the keywords extraction mechanism from *ReaderBench*. We used a Heatmap visualization presented in Figure 2 to represent the 10 most relevant keywords from each week identified in participants' discussions. Words such as "*grade*", "*exam*", and "*assign*" are used intensively in the last 4 weeks of the course, "*homework*" is present during the weeks in which homework was due, while "*number*", "*post*" and "*proof*" were used throughout the entire course.

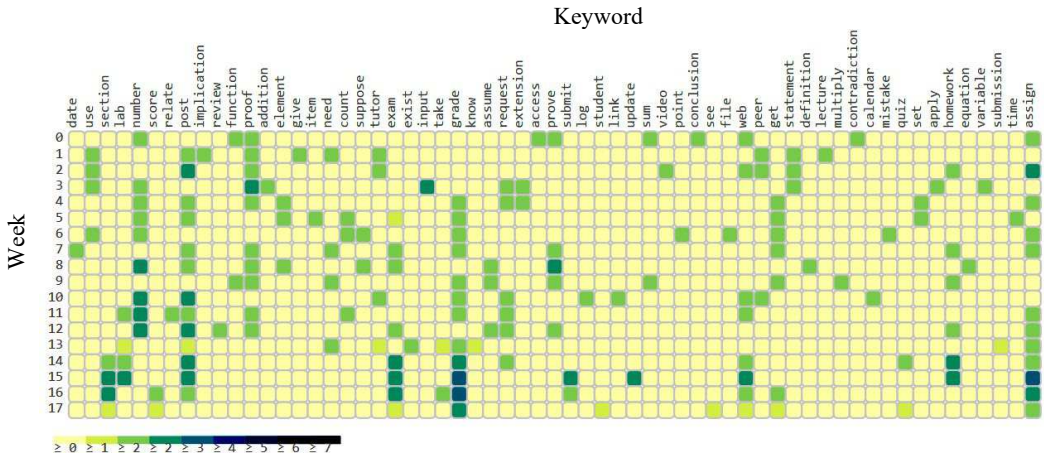
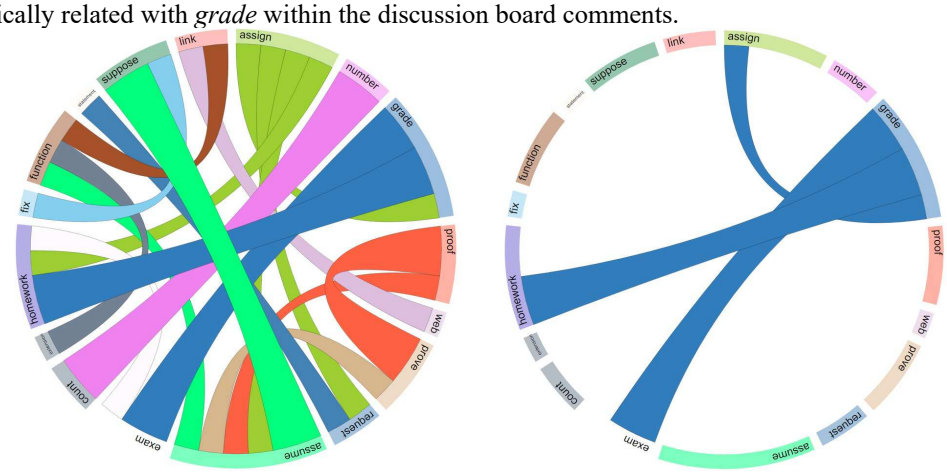


Figure 2. Keywords used by participants.

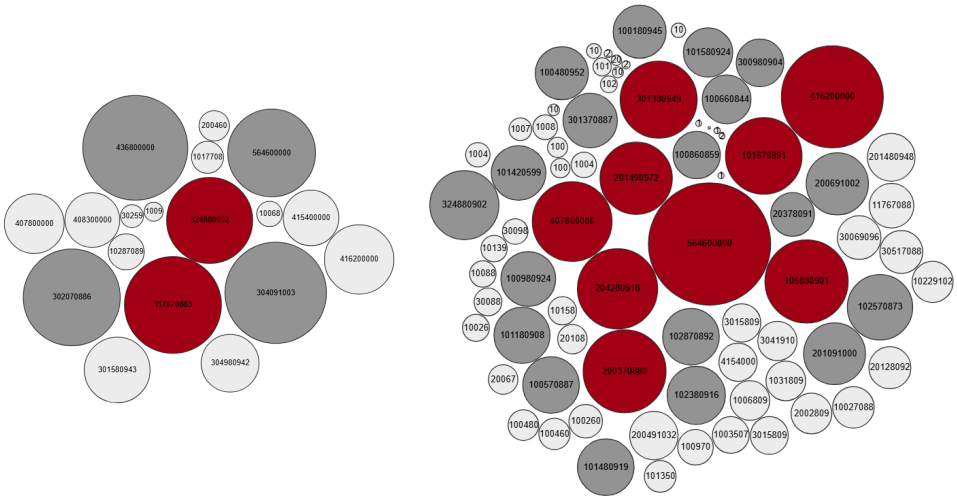
The semantic similarity between the most relevant 20 keywords is represented using a Chord diagram (see Figure 3) in which the external circle is populated with concepts and edges reflect the similarity between two concepts above an imposed threshold (i.e., edge width is proportional to semantic relatedness between concepts). Three concepts were not included in the diagram (i.e., "*lab*", "*section*", and "*post*") because these concepts were isolated in relation to the other extracted keywords. Figure 3.a presents the similarity between the remaining 17 concepts. Mouseover on the concept *grade* results in Figure 3.b, which isolates the visualization to concepts that are semantically related with *grade* within the discussion board comments.



(a) Keywords similarities. (b) Semantically related concepts to "grade".

Figure 3. Concept maps relying on the semantic similarity between concepts.

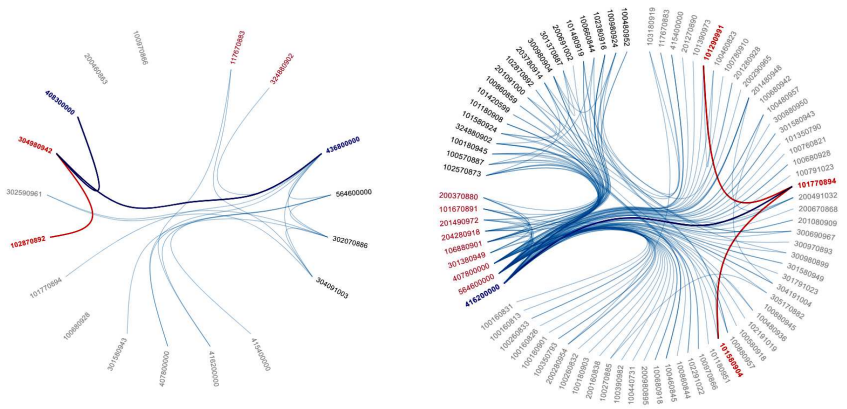
In addition, different sociograms are used to observe the interactions between participants, the participants' distribution over the three clusters, and the evolution of participation over time. Weekly snapshots are generated to examine how participation in the community evolves from one week to the next. Two types of interactive sociograms are generated to display interaction patterns between participants. First, the Clustered Force Layout presented in Figure 4 shows the position of each participant in the hierarchical structure using circles proportional to their CNA participation scores. The sociograms from Figure 4 reflect the participation of course members in week 6 (when a drastic decrease occurred) and in week 16 (when the highest number of participants and contributions was observed). Dark red nodes represent the central participants, active participants are colored in dark grey, whereas light grey denotes peripheral members.



(a) Week 6. (b) Week 16.

Figure 4. Clustered Force Layout.

Second, the interactions between participants are modeled using Hierarchical Edge Bundling sociograms. Figure 5 presents the dependencies between participants in Weeks 6 (left) and 16 (right) in a radial manner. Similar to the previous type of sociograms, the participants are grouped into their corresponding cluster, whereas the dependencies are grouped into spline bundles. The incoming and outgoing links are displayed on mouseover as follows: incoming links are colored in dark blue, whereas outgoing links are marked in red.



(a) Week 6. (b) Week 16.

Figure 5. Hierarchical Edge Building.

Discussions

This paper introduces a longitudinal analysis and new visualizations based on CNA for an online math course generated with the *ReaderBench* framework. The evolution of participation is depicted using Clustered Force Layouts and Hierarchical Edge Building models, while the most frequently used keywords and their semantic similarity are visualized using heatmaps and Chord diagrams.

The provided visualizations facilitate the understanding of course dynamics over time and can be used to derive specific characteristics, namely that: a) peripheral members tend to dominate the interactions – most students had singular inquiries and responses, without being engaged in multiple discussion threads; b) the degree of participation grows from peripheral members to active and central ones; and c) the course is not dominated by a single participant. Our tools are designed to ensure an in-depth longitudinal analysis of course participation based on discourse and text cohesion, while focusing on interaction patterns, clustering of students based on their participation, and monitoring course trends.

As follow-up directions, we plan to integrate the learning analytics tools within the ReadME platform (Botarleanu et al., 2018) in order to facilitate interactions, monitor students, and to automatically identify changes in the community structure (e.g., participants who exceed the peripheral layer and gain a more central role).

References

Botarleanu, R.-M., Dascalu, M., Sirbu, M.-D., Crossley, S. A., & Trausan-Matu, S. (2018). ReadME – Generating Personalized Feedback for Essay Writing using the ReaderBench Framework. In *3rd Int. Conf. on Smart Learning Ecosystems and Regional Development* (pp. 133–145). Aalborg, Denmark: Springer.

Crossley, Scott, Barnes, Tiffany, Lynch, Collin, & McNamara, Danielle S. (2017). Linking Language to Math Success in an On-Line Course.

Dascalu, M., McNamara, D. S., Trausan-Matu, S., & Allen, L.K. (2018). Cohesion Network Analysis of CSCL Participation. *Behavior Research Methods, 50*(2), 604–619. doi: 10.3758/s13428-017-0888-4

Dascalu, M., Trausan-Matu, S., McNamara, D.S., & Dessus, P. (2015). ReaderBench – Automated Evaluation of Collaboration based on Cohesion and Dialogism. *International Journal of Computer-Supported Collaborative Learning, 10*(4), 395–423.

Nistor, N., Panaite, M., Dascalu, M., & Trausan-Matu, S. (2017). Identifying Socio-cognitive Structures in Online Knowledge Communities (OKCs) using Cohesion Network Analysis. In *9th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing*. Timisoara, Romania: IEEE.

Scott, John. (2017). *Social network analysis*. London, UK: Sage.

Sirbu, M.-D., Dascalu, M., Crossley, S. A., McNamara, D. S., Barnes, T., Lynch, C. F., & Trausan-Matu, S. (2018). Exploring Online Course Sociograms using Cohesion Network Analysis. In *19th Int. Conf. on Artificial Intelligence in Education (AIED 2018), Part II* (pp. 337–342). London, UK: Springer.

Acknowledgments

This research was supported by the README project "Interactive and Innovative application for evaluating the readability of texts in Romanian Language and for improving users' writing styles", contract no. 114/15.09.2017, MySMIS 2014 code 119286, the FP7 2008-212578 LTfLL project, the National Science Foundation (DRL-1418378) and the Office of Naval Research (N00014-17-1-2300). We would like to thank Tiffany Barnes and Collin Lynch for their help in conducting the previous experiments.