Cohesion-Centered Analysis of Sociograms for Online Communities and Courses Using ReaderBench

Cohesion-Centered Analysis of Sociograms for Online Communities and Courses Using *ReaderBench*

Mihai Dascalu[1,2(✉)], Maria-Dorinela Sirbu[1], Gabriel Gutu-Robu[1], Stefan Ruseti[1], Scott A. Crossley[3], and Stefan Trausan-Matu[1,2]

[1] University Politehnica of Bucharest,
Splaiul Independenței 313, 60042 Bucharest, Romania
{mihai.dascalu, gabriel.gutu, stefan.ruseti, stefan.trausan}@cs.pub.ro, maria.sirbu@cti.pub.ro
[2] Academy of Romanian Scientists,
Splaiul Independenței 54, 050094 Bucharest, Romania
[3] Department of Applied Linguistics/ESL,
Georgia State University, Atlanta 30303, USA
scrossley@gsu.edu

Abstract. Computer Supported Collaborative Learning (CSCL) environments facilitated by technology have become a viable learning alternative from which valuable data can be extracted and used for advanced analyses centered on evaluating participants' involvement and their interactions. Such automated assessments are implemented within the *ReaderBench* framework, a Natural Language Processing platform that contains multiple advanced text analysis functionalities. The *ReaderBench* framework is based on Cohesion Network Analysis from which different sociograms, relying on semantic similarity, are generated in order to reflect interactions between participants. In this paper, we briefly describe the enforced mechanisms used to compare two Math communities, namely an online knowledge building community and an online course.

Keywords: Cohesion Network Analysis · Sociograms · Text cohesion
Natural Language Processing · *ReaderBench* framework

1 Introduction

Teachers and tutors have a limited amount of time to manually assess and grade student output. Moreover, monitoring and scoring student activities using indicators reflective of their performance in terms of participation or collaboration with peers is a cumbersome process. Hence, there is necessity for automated analyses, which led to the development of the Cohesion Network Analysis (CNA) approach and its integration within the *ReaderBench* framework available online at http://readerbench.com. *ReaderBench* [1, 2] is a fully functional open-source framework centered on discourse analysis that consists of various Natural Language Processing (NLP) techniques designed to support students and teachers in their educational activities. This paper

V. Pammer-Schindler et al. (Eds.): EC-TEL 2018, LNCS 11082, pp. 622–626, 2018.
https://doi.org/10.1007/978-3-319-98572-5_59

presents a brief overview of Computer Supported Collaborative Learning (CSCL) experiments centered on online communities and performed within *ReaderBench*.

2 Performed Experiments

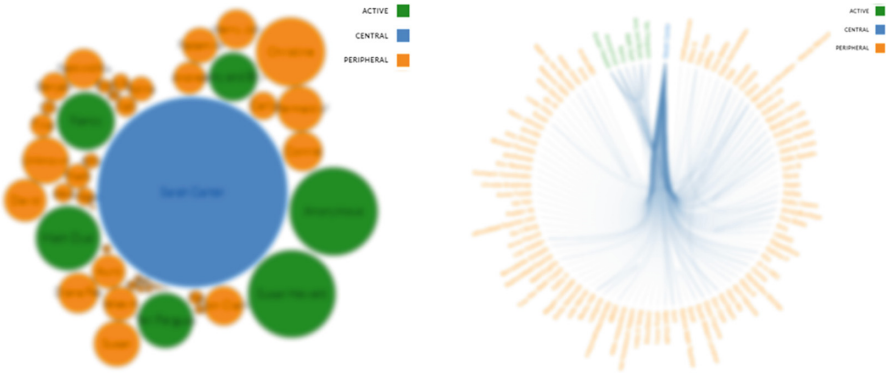
Two experiments in different CSCL environments were conducted. These experiments focused on Online Knowledge Building Communities (OKBC) and online courses. CNA transcends Social Network Analysis (SNA) by taking into account discourse quality reflected in semantic cohesion. CNA models interactions between participants and provides a scoring mechanism within collaborative conversations by combining NLP techniques with SNA. In *ReaderBench*, CNA is used to compute cohesion indices that are based on the discourse structure and which reflect participation and collaboration throughout the conversation [2]. Moreover, CNA is tightly coupled with dialogism and polyphony which define the theoretical framing of CSCL [2]. Moreover, CNA closely resembles SNA by relying on equivalent indices to quantify participation within the generated sociograms [1, 2]. Afterwards, hierarchical clustering is used to extract the community's socio-cognitive structure based on two CNA indices derived from the sociogram: in-degree (reflective of collaboration in terms of inbound messages) and out-degree (highlighting active participation in the community).

Two types of views are used to model the interaction between participants, namely a *Force-Clustered Graph* and a *Hierarchical Edge Bundling* visualization [3]. The views are generated using the d3.js library (https://d3js.org). The *Force-Clustered Graph* view shows the interactions between participants based on a graph in which the nodes represent participants who are clustered by considering the inter-exchanged messages between them. The size of nodes represents the average score of in-degree and out-degree values from the overall sociogram and it is directly proportional with the participant's score. The clustered participants were colored as follows in descending order of average in-degree CNA scores: central members are colored with blue, active members with green, and peripheral members with orange. The *Hierarchical Edge Bundling* view presents the interactions between participants in a branching manner. The participants are organized into their corresponding cluster: active, central or peripheral. The same colors like in the previous view were used.

The first experiment evaluated the involvement of participants in online blog communities, their interactions and evolution throughout the discussion threads [3]. The analysis was performed on a corpus of 85 conversations from 78 members, cumulating 250 contributions extracted from the online Math community http://mathequalslove.blogspot.com. Figure 1 has been blurred in order to anonymize the names of the participants and it depicts the sociogram of the OKBC in which the blog owner is the main person within the community.

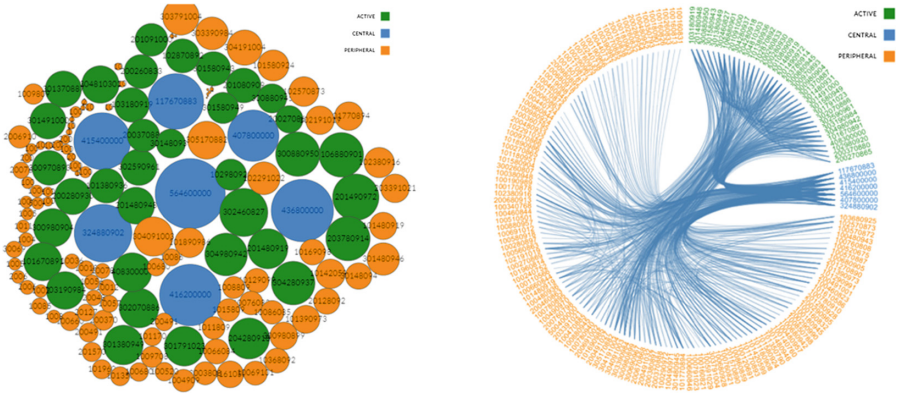
The second experiment was used to predict students' completion rates in the context of an online Math course [4]. Based on the generated cohesion graph and a longitudinal analysis, CNA indices were computed for each of the 157 students (from 250 students, only 157 made posts on the forum). The method showed that students who are active on forums are more likely to finish the course, while the number of days spent on the forum and the consistency of posts are predictors for math success.

As the number of participants is considerably larger and the force directed view becomes too cluttered, we opted to include a force-clustered view besides the hierarchical edge bundling view (see Fig. 2). In this case, we can observe a higher presence of central and active members in contrast to the online community in which the discourse is centered around the blog owner who sustains the community (see Fig. 1), while the interactions from other participants are quite limited.



(a) Force-Clustered Graph view (b) Hierarchical Edge Bundling view

Fig. 1. Sociograms corresponding to the OKBC.



(a) Force-Clustered Graph view (b) Hierarchical Edge Bundling view

Fig. 2. Sociograms corresponding to the online course.

In addition to the global views, weekly timeframes are also generated in the performed longitudinal analysis in order to represent interactions and connections between participants in an interactive and intuitive manner. While participation in the online community exhibits little fluctuations between adjacent weeks, we can observe specific traits within the online course. Figure 3 shows participants' activity in the first, 9th

(mid-semester), and the last week of the online course. Some interaction patterns can be observed using the newly introduced Force-Clustered Graph views, as follows: (a) the peripheral members play a more important role in the community; (b) a decrease in involvement can be observed towards the end of the course; (c) the discussions are not dominated by a single member. These findings are aligned with the observations from our previous study [5].

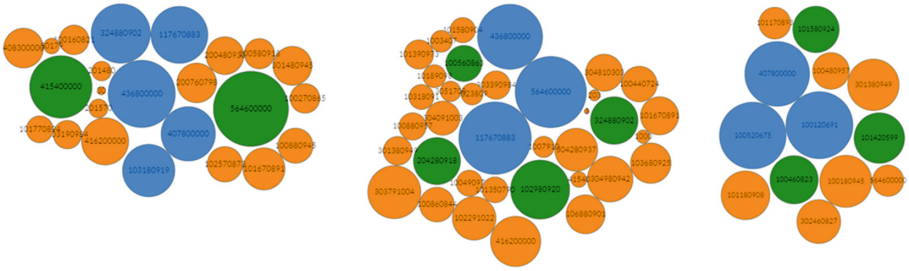


Fig. 3. Sociograms within Math Course - Snapshot views of weeks 1, 9 and 18.

3 Discussion and Conclusion

These experiments demonstrate the capability of the *ReaderBench* framework to analyze different online CSCL environments and to perform in-depth, cohesion-centered analyses of interaction patterns. The generated sociograms provide valuable insights with regards to different interactions patterns and can be used in follow-up experiments to provide personalized feedback to learners in order to actively engage them.

Acknowledgments. This research was partially supported by the 644187 EC H2020 RAGE project, the FP7 2008-212578 LTfLL project and the National Science Foundation (DRL-1418378). We would like to thank Nicolae Nistor, Tiffany Barnes, Collin Lynch and Catalina Durbala for their help in conducting the previous experiments.

References

1. Dascalu, M., Trausan-Matu, S., McNamara, D.S., Dessus, P.: ReaderBench – automated evaluation of collaboration based on cohesion and dialogism. Int. J. Comput. Support. Collab. Learn. **10**(4), 395–423 (2015)
2. Dascalu, M., McNamara, D.S., Trausan-Matu, S., Allen, L.K.: Cohesion network analysis of CSCL participation. Behav. Res. Methods **50**(2), 604–619 (2018)
3. Sirbu, M.-D., Panaite, M., Secui, A., Dascalu, M., Nistor, N., Trausan-Matu, S.: ReaderBench: building comprehensive sociograms of online communities. In: 9th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC 2017), Timisoara, Romania. IEEE (2017)

4. Crossley, S.A., Dascalu, M., Baker, M., McNamara, D.S., Trausan-Matu, S.: Predicting success in massive open online courses (MOOC) using cohesion network analysis. In: 12th International Conference on Computer-Supported Collaborative Learning (CSCL 2017), Philadelphia, PA, pp. 103–110. ISLS (2017)
5. Sirbu, M.-D., et al.: Exploring online course sociograms using cohesion network analysis. In: Penstein Rosé, C., et al. (eds.) AIED 2018, Part II. LNCS (LNAI), vol. 10948, pp. 337–342. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-93846-2_63