In Proceedings of the Intelligent Tutoring Systems Conference (ITS 2010)

A Network Analysis of Student Groups in Threaded Discussions

Jeon-Hyung Kang, Jihie Kim and Erin Shaw

University of Southern California Information Sciences Institute
4676 Admiralty Way, Marina del Rey, CA, U.S.A
{jeonhyuk, jihie, shaw}@isi.edu

Abstract. As online discussion boards become a popular medium for collaborative problem solving, we would like to understand patterns of group interactions that lead to collaborative learning and better performance. In this paper, we present an approach for assessing collaboration in online discussion, by profiling student-group participation. We use a modularity function to compute optimal discussion group partitions and then examine usage patterns with respect to high-versus low-participating students, and high- versus low-performing students as measured by grades. We apply the profiling technique to a discussion board of an undergraduate computer science course with three semesters of discussion data, comprising 142 users and 1620 messages. Several patterns are identified, and in particular, we show that high achievers tend to act as 'bridges', engaging in more diverse discussions with a wider group of peers.

Keywords: Student online discussions, group detection in discussions

1 Introduction

Online discussion boards play an important role in distance education and web enhanced courses. Studies have shown online discussion to be a promising strategy for promoting collaborative problem solving and discovery-oriented activities, however, student participation can vary highly; some students post only one or two messages during the whole semester, while others participate more often and interact with many other students. Some students communicate with only a limited number of peers, while others interact with a wider group of students and participate in more varied discussion topics. It is difficult to understand the many different types of group interactions that occur in online discussions and even more difficult to understand how they affect collaborative learning.

We would like to identify and understand patterns of group interactions that lead to collaborative learning and better performance. The patterns might be used to develop pedagogical strategies for promoting more desirable interactions and increasing student learning. Most of the existing computational work on qualitative discussion analysis has focused on analyzing dialogue patterns in individual discussion threads (Ravi & Kim 2007; Feng, Kim & Shaw, 2006, McLaren et al., 2008) or analyzing the

impact of tutors in student discussions (Light et al., 2000; Shaw 2005). Although these results provide good hints about student behavior within online discussions, they have not yet yielded insightful information about collaborative learning, such as how groups form during online interactions over multiple discussion threads, and how group interactions affect student learning.

To help profile discussion participation, we define three new terms: *high-participating students*, *active group participants (AGPs)* and *bridge students*. High-participating students are defined as those who participate in many discussion threads and AGP are students who participated in many discussion threads in the same group. We also define bridge students as those who participated in multiple discussion threads across several different groups.

2 Modeling Discussion Group

A discussion group is modeled based on information about its students and the discussion threads in which they participate. The relationship can be represented as a directed graph, where nodes are either users or threads, and edges connect users who participate in threads. The graph is then partitioned to detect optimal communities of discussants. This is done by first representing the graph as a discussion (student-thread) matrix, and then finding a partitioning that maximizes the strength of the connections within a group, called the modularity (Girvan and Newman, 2002; Ghosh and Lerman, 2008). The modularity measures how good the given partition is by comparing the difference between the number of edges that lie within groups in the given partition and the same quantity when the edges are placed randomly and the vertices have the same degree. Once an ideal community is found, we study patterns between and within groups. To set the threshold, we analyzed the degree of participation by individual students in our largest dataset, in which 50 discussants participated in 100 discussion threads. We applied the power-law to set the threshold. In the dataset for this work, the top 20% students participate in at least five discussion threads and 80% of the messages were written by top 20% of the students. We labeled the top 20% students high-participating, and the rest low-participating. We applied the same threshold for AGPs. AGPs participated in at least five different discussion threads in any given group. Since our model produces 4 groups in spring, 3 groups in summer, and 7 groups in fall 2008 semester, we used the smallest number to define bridge students. Bridge students participated in more than three threads that belong to more than three different groups.

3 Preliminary Results

Each class discussion forum we studied corresponded to a unique class project. ‘Project 2’ discussion communities were modeled because the forum had the highest number of participants. We then compared the corresponding ’Project 2’ grades of *high- and low-participating students*, *AGPs*, and *bridge students*. The results are shown in Table 1. To evaluate the differences in grades between different pairs of

groups we applied the *t*-test. There was no significant difference in the grades of high- and low-participating students; however, AGPs received lower average grades than non-AGP, high- and low-participating students. This is consistent with our earlier findings that high-participating students may be help seekers and not necessarily high performers. A full understanding will require an in-depth investigation of thread features, taking into account the technical quality of messages, dialogue patterns, technical terms used, and discussion topics.

Interestingly, we found that high-performing students, however, tend to act as bridge students across several different community groups, engaging in more diverse discussions with a wider group of peers. Specifically, they participated in multiple discussion threads across several different groups. A *t*-test for the comparison between bridge student and non-bridge student grades shows that the difference is significant at a 99.9% level with $t(56) = 4.39$, $p < 0.001$.

Table 1. Student grades for Project 2 in three different semesters.

Groups	Spring 2008	Summer 2008	Fall 2008	Year 2008
Bridge Students	**37.3**	**-**	**39.3**	**38.4**
Non-Bridge Students	33.4	34.2	34.6	34.3
High Participating Students	28.3	37.0	35.8	34.8
Low Participating Students	36.1	22.2	23.8	34.8
Active Group Participants	24.0	37.6	36.3	34.4
Non-AGP	36.1	33.4	34.9	34.9
All Students	34.3	34.5	35.1	34.8

Acknowledgments. This work was supported by National Science Foundation CCLI Phase II (#0618859) and CISE IIS (#0917328) grants.

References

1. Feng, D., Kim, J.,Shaw, E., and Hovy E.: Towards Modeling Threaded Discussion using Induced Ontology Knowledge. In: National Conference on Artificial Intelligence (2006)
2. McLaren, B., Scheuer, O., De Laat, M., Hever, R., De Groot, R., and Rose, C.: Using Machine Learning Techniques to Analyze and Support Mediation of Student E-Discussions. In: AIED (2007)
3. Girvan, M., and Newman, M. E. J.: Community structure in social and biological networks. In: Natl. Acad. Sci. USA 99, 7821-7826 (2002)
4. Ghosh, R. and Lerman, K.: The structure of heterogeneous networks. In: 1st IEEE Social Computing Conference (2009)
5. Ravi, S. and Kim, J.: Profiling Student Interactions in Threaded Discussions with Speech Act Classifiers (2007)