Complex Network Theory Approach to the Assessment on Collective Knowledge Advancement in CSCL

Jun Oshima, Ritsuko Oshima, Shizuoka Univ., 3-5-1 Johoku Hmamatsu-shi JAPAN 432-8011
joshima@inf.shizuoka.ac.jp, roshima@inf.shizuoka.ac.jp
Knowledge Forum® Japan Research Group, kfjp@inf.shizuoka.ac.jp

Abstract: This study discusses the possibility of the integrated analytic approach to discourse in CSCL by the combination of macro network analysis and micro analysis of argument on students' written discourse. Although studies have established fine-grained analytic approaches to discourse or argumentation in CSCL environments, we still have difficulty with evaluating collective knowledge advancement. The *Complex Network Theory* would be a promising approach to challenging this difficulty. We can visualize a variety of network structures with identifying ideas as *nodes*, and co-presence of words as *links*. Several indices numerically inform us how a target network is structured. In this paper, we report our attempt to describe how the network of ideas represented in discourse is structured in CSCL environments and its relation to the network structure analysis.

Background and Research Purposes

Although the development of argument analysis in the learning sciences provides us with fine-grained information on cognitive activity by individual learners (e.g., Sampson, & Clark, 2006), we still do not have tools to evaluate learners' collective knowledge advancement. The assessment of collective knowledge advancement is crucial with two reasons. One comes from the perspective of *summative assessment* that the combination of discourse analysis at the individual learner's level and collective or structural analysis of ideas gives us richer interpretation of individual learner's cognitive performance. The analysis of how each individual learner contributes to collective knowledge advancement is an important measure of the knowledge advancement through collaborative learning. Another reason comes from the perspective of *informative assessment*. If we figure out how each learner is contributing to the collective knowledge advancement in the community such as a classroom, we can suggest each learner what ideas they should know or contribute to in their next stage of learning.

We propose an analytic approach to achieving the dynamic assessment on students' collective knowledge advancement, the network structure analysis based on the *complex network theory* (Barabási, Albert, & Jeong, 1999; Watts, 1999; Watts, & Strogatz, 1998). The *complex network theory* is an analytic approach to describing a variety of network structures developed based on statistical physics, and several important features of the network structures around us were discovered (i.e., the Small World, the Scale-Free, etc.). Some of recent studies were focused on the issue of whether the same principles could be applied to the development of the computer-network communication (e.g., Strogatz, 2001). For instance, one study by a Japanese research group (Tagawa, Yasutake, Yamakawa, & Inoue, 2006) attempted to describe remarkable features of computer-mediated communication by university students with WebCT. Although the network structure analysis would give us important resources to evaluate students' collective knowledge advancement, we have not yet known how we can utilize the indices from the network structure analysis in evaluating students' knowledge advancement in collective situation. In this study, we attempt to find answer to this research question by comparing results from both fine-grained and network structure analysis on the same students' discourse in a CSCL context.

Methodology

Target Group of Students

Forty-one fifth grade students (21 females, and 20 males) at a Japanese elementary school were engaged in their collaborative learning on genetically modified foods by using Knowledge Forum®. They discussed whether they should develop GM crops or not and why based on their understanding of GM foods (Oshima, et al., 2005). We analyzed their written discourse in the final phase of learning through the two approaches.

Individual Analysis of Written Discourse

Written discourse in each report was analyzed from the perspective of the argument structure and the epistemic operation (Oshima, Oshima, & Knowledge Forum® Japan Research Group, 2006). In the argument structure analysis, we referred to the simplified framework of Toulmin's (1958) argument structure such as *Data*, *Reasoning*, *Claim, and Rebuttal*. Each written discourse was evaluated with whether each argument component was present. In further analysis of epistemic operations in written discourse, the cognitive levels of the two components (i.e., *Data*, and *Reasoning*) were evaluated by referring to the rubric developed in other studies (e.g., Sandoval, & Millwood, 2005). Two trained undergraduate students independently involved in the evaluation procedure. The inter-rater agreement was over .80. The disagreement was resolved through discussion with the first author.

Network Structure Analysis

Each discourse was decomposed into paragraphs as minimum units of ideas, and each paragraph was further decomposed into morpheme words through the software application of Japanese language morphological analysis. The same procedure was applied to discourse in teaching documents on the genetically modified foods that we created under a domain expert's supervision. We depicted nouns that appeared in both corpuses. Based on the list of noun words, we conducted the complex network structure analysis on the both corpuses by the software called *Pajek*. *Pajek* is a network structure analysis application that provides us with basic descriptions of each node (word in this case) and statistical indices, *clustering coefficient* and *betweeness centrality*. The *clustering coefficient* is a measure in describing network structures. We omit its mathematical explanation here. In short, the coefficient informs us how each node (word in this case) contributes to the development of clustering structures in the network. We can describe how the target network is composed of idea clusters and how powerful the central ideas are in structuring the network. The *betweeness centrality* is another measure that manifests how central each node is in the network structure or clusters. With these two measures, we can describe network structure of ideas learners reported in their written discourse from the perspectives of how their ideas are linked to one another as groups, and which word plays important roles to create clusters of ideas. In this study, we used the network structure of the document we created as the benchmark of idea network. Later, we compared network structures of students' discourse with the benchmark structure to examine how scientifically appropriate the network structure of ideas in students' discourse is.

Stepwise Network Structure Analysis

For examining the relationship between measures from the discourse analysis and the network structure analysis, we conducted what we call *stepwise network structure analysis*. The stepwise network structure analysis is the procedure that compares the network structure of nodes from total reports with that excluding a target single report for examining its contribution to the total network structure. Our assumption was that a network structure would be significantly changed by excluding cognitively important discourse. We detected several reports that produced crucial changes in the two coefficients of ten most important words in the corpus. Then, we attempted to characterize discourses evaluated as important in idea network structures by using their argument structures of written discourse.

Results and Discussion

First, we conducted the network structure analysis on the two corpuses of discourse by students and researchers. The number of nodes was 101. The mean *clustering coefficients* across all nodes were 0.769713 for students' and 0.760392 for researchers'. The mean *betweeness centralities* were 0.009211 and 0.007735, respectively. If we looked at ten most influential words in structures, only three were appeared in the both lists. In sum, the network of ideas by students' discourse and researchers' discourse were structurally similar, but cognitively different. We further conducted network structure analyses by separating the discourse by positive and negative opinions. The differences in the structures were remarkable (Figure 1). The mean *clustering coefficients* were 0.811 in positive opinion and 0.811 in negative opinion. The *betweeness centralities* were 0.018 in positive opinion and 0.011 in negative opinion. Results manifest that network structures of students' discourse at each side were more different from the researchers' than students' total structure. Based on results, we think that students' idea network structure came to be closer to the researchers' as their learning went on, but students focused their attention to limited range of learning materials.

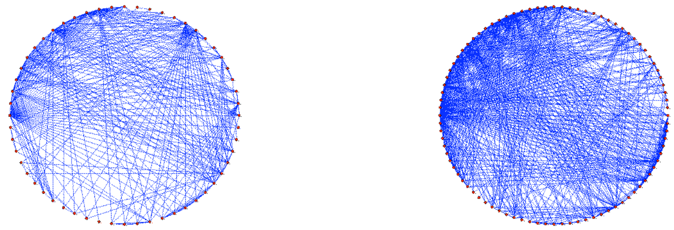

Figure 1. The Network Structures Based on Students' Discourse with Positive (left) or Negative (right) Opinions.

Second, we compared students' discourses detected as strongly influential to their idea network structure with those evaluated as robust in the argument structure analysis. We found no systematic relation between the two results of analyses, but concluded that the two analyses provides us with more fine-grained characteristics by covering different perspectives of students' knowledge advancement in CSCL environments. Discourses highly evaluated by the both analyses are considered to play central roles in their collective knowledge advancement and have robust argument structure. Therefore, instructors can use those as benchmarks of how students' knowledge advancement is evolving. Discourses highly evaluated by the argument analysis but not by the network structure analysis were further divided into two categories. One category included discourse showing ideas appeared in other notes. Idea networks in these discourses can be easily replaced by other discourse and they do not influence the network structure at all. Instructors should suggest authors to merge their reports in one integrated note and further think of its relation to other notes. The other type of discourse was discussing very local ideas so that they created small clusters in the network and no links to other ideas. The instructor can suggest students to think of how their ideas should be related to others' ideas. Particularly, the benchmark notes that are highly rated by both analyses would be good for them to see. The last category was discourses highly evaluated by the network analyses but not by the argument analysis. It may be difficult for other learners to comprehend ideas appeared in this type of discourse because the argument structure is not clear enough. If instructor thinks that ideas in the discourse are valuable to further develop in collaboration with other learners, s/he has to help authors create more robust structures of arguments.

References

Barabási, A-L., Albert, R., & Jeong, H. (1999). Meanfield theory for scale-free networks. *Physica A*, *272*, 173-187.

Oshima, J., Oshima, R., & Knowledge Forum® Japan Research Group. (2006). Scientific argument in written discourse with task requirements triggering different epistemic agency. *Paper presented at the annual meeting of the American Educational Research Association*, San Francisco, CA.

Oshima, J., Oshima, R., Murayama, I., Horino, R., Inagaki, S., Yamamoto, T., Takeshita, Y., Fujimoto, M., Takenaka, M., Yamaguchi, E., Nakayama, H., & Sakamoto, M. (2005). Progressive Refinement of a CSCL-Based Lesson Plan for Improving Student Learning as Knowledge Building in the Period for the Integrated Study. *Proceedings of CSCL2005*.

Pajek Program for Large Network Analysis. http://vlado.fmf.uni-lj.si/pub/networks/pajek/.

Sampson, V. D., & Clark, D. B. (2006). Assessment of argument in science education: A critical review of the literature. *Proceedings of ICLS 2006*, 655-661. Mahwah, NJ: Lawrence Erlbaum.

Sandoval, W. A., & Millwood, K. A. (2005). The quality of students' use of evidence in written scientific explanations. *Cognition and Instruction*, *23*(1), 23-55.

Strogatz, S. H. (2001). Exploring Complex Networks. Nature, 410, 268-276.

Tagawa, T., Yasutake, K., Yamakawa, O., & Inoue, H. (2006). An analysis of a communication network structure in an e-learning course. Proceedings of the Fourth User Conference on WebCT, Japan. (in Japanese with English abstract)

Toulmin, S. (1958). *The uses of argument.* Cambridge, England: Cambridge University Press.

Wattz, D. (1999). Networks, dynamics, and the small world phenomenon. *American Journal of Sociology*, *105(2)*, 493-527.

Wattz, D., & Strogatz, S. (1998). Collective dynamics of small-world networks. *Nature*, *393*, 440-442.