

# Evaluating CSCL Log Files by Social Network Analysis

Kari Nurmela, Erno Lehtinen and Tuire Palonen

University of Turku, Faculty of Education

## Abstract

In this paper our aim is to present a methodology which can be used in analyzing the interaction processes in a groupware environment. We demonstrate how the social network analysis approach can be used as a method to evaluate the social level structures and processes of a group studying in a CSCL environment. This approach tries to highlight in particular the participatory aspects of collaborative learning processes, but it can also serve as a starting point for more detailed analysis of knowledge building and acquisition processes. The relations between learners and the structure between documents written are the examples studied here.

There are some features that make log files especially important in CSCL systems. First, log files can be used automatically, precisely and effectively for data collection. Second, analyzing this information enables evaluative perspective to the collaborative action as a whole. Third, this feedback can be made available immediately for the learning community. As a contribution, we can assume that the network analysis of log files helps us to understand the working processes in CSCL.

## Introduction

Numerous studies on CSCL not only demonstrate promising improvement in individual level learning but also in the quality of social interaction processes among the community. However, practical experiences also raise questions about the shortages and problems students have when participating in CSCL learning environments. In this paper our aim is to present a methodology which can be used in analyzing the interaction processes which take place in a groupware environment.

Sfard (1998) has made a division into two main metaphors of learning: the acquisition metaphor and the participation metaphor. The questions concerning the learning outcomes belong to the more traditional acquisition paradigm that interprets learning in terms of the acquisition of something in an individual mind and knowledge in terms of property and possession. The ideas of collaborative learning at least partly belong to the emerging participation metaphor. This approach deals with learning as becoming a participant and with knowledge as an aspect of practice, discourse and activity. We agree with Sfard that both of the two metaphors of learning (acquisition and participation) are needed.

Miyake (1986) and Hutchins (1995, 1991), among others, have argued that social interaction (and interaction with the tools of technological culture) provides new cognitive resources for human cognitive accomplishment. According to Miyake's

analysis, understanding emerges through a series of attempts to explain and understand the processes and mechanisms being investigated. In a shared problem-solving process, agents who have partial but different information about the problem in question appear to improve their understanding through social interaction (see also Oatley, 1991; Brown & Palincsar, 1989, Hakkarainen & Lipponen 1998; Wittenbaum & Stasser, 1996). In order to explain one's view to one's peers, an individual student has to cognitively commit himself or herself to some ideas, explicate his or her beliefs, as well as organize and reorganize his or her knowledge (Hatano & Inagaki, 1992).

Computer supported collaborative learning, in which thought processes are externalized in the form of public discourse, provides an agent with access to other participants' processes of thought, thus supporting the development of the agent's metacognitive skills. The cognitive value of externalization in social interaction is based on a process of making internal processes of thought visible (Collins, Brown, & Holum, 1991; Lehtinen & Rui, 1997; Lehtinen & Repo, 1996; Scardamalia & Bereiter, 1989).

A CSCL environment can provide structures and activities that foster monitoring of one's own and the other students' comprehension and reflect advancement of learning and problem-solving processes (Brown & Campione, 1996). One way to do this is to use automatic data collection. The computer records the detailed actions that occur within the program used. The recording is often stored in log files that can be analyzed as part of evaluation process. This analyzing can be made immediately available in a CSCL environment thus helping to reflect the learning process. In addition to analyzing acquisition, participatory analyzing should be used to reflect the learning processes too.

The approach of social network analysis is useful for studying relationships. Instead of studying the behavior of analytically isolated individuals social analysts describe patterns of relationships between actors, analyze the structure of these patterns, and seek to uncover their effect on individual behavior. A useful tool for that is the representation of social structure as a network: a set of nodes and the set of ties connecting these nodes. (Wasserman & Faust, 1995; Scott, 1991.)

The analytic methods of social network analysts have focused on obtaining information about linkages and pointing out the fundamental structural patterns. Because a communication network is a social network - a patterned set of connections linking actors to each other - the network approach is especially useful for CSCL studies.

Traditional evaluative measurements and analysis have certain limitations (for example the statistical assumption that measurement error is independent across subjects) when used to evaluate interventions that involve social action or co-operational learning. Typically the studies concerning individuals overlook measures of social structural factors. Social network techniques permit analysis at both a group and an individual level, and integrate the data on interpersonal relations. (Haythornthwaite, Wellman & Mantel, 1995.) So, network models may provide useful tools for the design and evaluation of CSCL environments.

It is obvious that computer supported collaborative learning environments can facilitate individual learning (Sfard's acquisition metaphor) as well as the development of social systems in which each individual can serve the joint activities and benefit from the shared resources of the group (Sfard's participation metaphor). In this study we demonstrate how the social level structures and learning processes in a CSCL environment can be evaluated with social network analysis. Here we focus especially on the participatory aspects of collaborative learning processes.

## **Method**

### **Participants**

The participants of the experiment were eighteen university students, one tutor (a doctorate student) and two supervisors (a professor of education and a professor of psychology). To combine written communication with face-to-face communication the students worked in pairs:

- teacher training students (two pairs: MiMe, HaSa)
- educational science student & teacher training student (one pair: AnLa)
- educational science students (two pairs: MiRi, MaSa)
- educational science student & psychology student (two pairs: MiEl, TeMi)
- psychology students (three pairs: ToNi, RiMi, AnSu)

### **Materials**

The research focused on an educational psychology course (in particular the diagnosis and training of reading strategies of learning disabled pupils) that took place in the University of Turku, Finland, in spring 1999. The course started with a short introductory lecture (4 hours) that was followed by 18 case based assignments. There was a strict timetable that grouped the assignments in four periods that lasted about a week each. The students then worked with the assignments until the conclusion lecture (4 hours). The collaborative work took place in World Wide Web (WWW) based groupware called WorkMates 4.

Like many other groupware programs, WorkMates 4 (WM) shares documents in the World Wide Web (WWW). The program is being developed in the Educational Technology Unit at the University of Turku. In addition to making new documents with different access types, users can attach files and mark keywords. Comments and questions can be added to the desired part of text. A feature unique to WM is the possibility to add reference links to other documents in WM and mark them as either "for" or "against" the proposal or statement.

All eighteen assignments were given in WorkMates and the students presented their analyses of the cases in the form of socially shared documents in the system. The tutor encouraged the addition of comments and references in documents. Student pairs were also guided to make one document per assignment. The assignments were focused on authentic case studies about different learning strategies and processes. Most of them

were based on video clips describing pupils' reading processes. The clips were digitally compressed MPEG-1 files that created a large space requirement (about 500 megabytes), and therefore the video clips were distributed to the students on CD-ROM. This way the network connection was used only for document sharing which made it possible for the students to work from their homes by using modems.

### **Data collection and preparation**

In addition to the information in the documents, Workmates collects data on the users' actions in a log file. This data collection is one of the only terms that are prerequisites for Workmates usage. The data was used for observing document building and communication structure. Tailored perl scripts were made to analyze the log file and the documents.

The log file for this particular Workmates had more than six thousand actions recorded. Besides the action type, the log files also contained user identification, possible target identification (the target document for example) and the time in seconds. From 26 different action types logged only some seemed important to this research. We picked "Finished making a new document", "Finished editing a document", "Reading a document" and "Added a comment, question, link or keyword to a document" for more detailed examination. These were considered the best actions to describe knowledge building. Study of time usage will be a future complementary analysis.

The main part of the work was placed in over 200 documents, so they were also handled by the scripts. The documents were searched for reference links between them. This resulted in a weighted and directed graph where the documents were viewed as nodes and the reference links as vertices. In a similar way the documents were searched for comments and questions. The resultant graph was made of nodes symbolizing the work pairs and the vertices indicated the network communication. These graphs were then analyzed by social network analysis and multidimensional scaling (MDS).

### **Social network analysis**

Social network analysis is focused on uncovering the patterning of people's interaction. We chose these special tools as the methods in standard statistics were incapable of handling relationships. Standard methods use individual attributes of actors, not relationships among actors. Instead of only describing the structure with general terms, we have gathered information of the ties in the network and looked at the tie structure. Theoretically we have two distinct forms of social interaction: cohesion and structural equivalence. (Scott, 1991; Wasserman & Faust, 1995.)

Here we concentrate on the concept of cohesion, which refers to the extent of direct interaction among individuals in the learning environment. This means an attempt to quantify the prominence of an individual actor embedded in a network. We have used the analysis for three purposes: 1) to seek the central actors of the CSCL environment, 2) to study the connections between actors and to scale them by MDS and, 3) to study the

structure of documents. All analyses are calculated using the Ucinet program (Borgatti, Everett and Freeman 1996a).

Centrality is used to find the most visible, remarkable or impressive actors in the field. There are several analyses that emphasize centrality in somewhat different perspectives. We have used three of them: Freeman's degree and betweenness, and Stephenson's and Zele's information measure. (Borgatti, Everett & Freeman 1996b, 82-89.) Freeman's degree measures the network activity of individual actors. It is possible to use asymmetric data (like received versus sent comments). Freeman's betweenness tells about actor's possibility to control information. Stephenson's and Zele's information measure shows in this case how far an actor is situated from the central actors of a graph. The path distance between two actors is measured based on the number of lines between actors.

The basic idea behind MDS is that of using the concepts of space and distance to map relational data. Contrary to the previous path distance, this distance is closer to the physical one. It includes an attempt to convert chart measures into metric measures (Scott 1991, 151-156.). Thus a complex social structure can be presented only by some, in this case with three dimensions. The quality of a MDS map can be measured by a value of stress, where the greater value means a worse model. The value is, however, dependent on data: the number of actors and the scale of measures.

The component analysis was used when the document structure of the CSCL environment was studied.

## **Results**

### **Describing learning activities**

The four main document action types consisted of more than half of the logged actions in total (Table 1.). From these types, document reading was clearly the largest (85 %). This was influenced by the type of working process: the students were accomplishing the recommendation to read each others' documents. There are, however, quite big differences in the amount of document reading actions between the student pairs.

The numbers of documents created are pretty similar with all the work pairs. This is connected to the amount of assignment documents. On the other hand, the amount of editing is quite dissimilar: some students apparently worked more with document building than others did. Likewise, in document reading, the amount here describes only the number of actions. With time usage analysis we could have found out more about the different learning strategies; for example several short periods of editing versus few more intensive editing. The indicator for groupware communication was also quite different between each student pair. There is no statistically significant correlation between the measured action types and credit due to the small number of student pairs. The two student pairs AnLa and HaSa do not follow the general trend that the activity in CSCL environment results better credits.

Table 1. The selected action types from the WM log file and credits

	1. Student pair	Finished making a new document	Finished editing a document	Reading a document	Added comment, question or link to a document	Credit (max 3)
MiMe		24	21	415	34	2.5
AnLa		20	4	480	39	2
MiRi		19	13	243	14	Not yet passed
TeMi		21	27	426	10	2.25
MaSa		21	12	333	19	2.25
MiEl		22	17	458	49	2.75
ToNi		22	11	254	13	1.75
RiMi		20	0	210	9	Not yet passed
HaSa		23	6	184	13	2.25
AnSu		0	0	72	0	Suspended course
Tutor		2	0	545	129	
Supervisor 1 (psychology)		0	0	16	0	
Supervisor 2 (education)		23	15	442	11	
Total		<b>217</b>	<b>126</b>	<b>4078</b>	<b>340</b>	

From all the students that were enrolled in this course, one pair (AnSu) did not take part in assignments because of timetable problems. They were dropped from the student pair list, likewise, supervisor 1, because their work did not have affect on the results.

#### Communication structure: comments between participators

The different measures for centrality (degree, betweenness and information) are presented in Table 2. As we can see the received and sent comments are very asymmetric. The students have more received comments than sent comments. This is mainly because the tutor was very actively commenting their documents. We can also see that students have participated in discussion with a different activity: some pairs (ToNi, TeMi, MaSa, RiMi, MiRi) have written only a few comments while some have commented on tens of documents (AnLa, MiEl). The number of sent messages tells how active the student pair has been in commenting on the other's documents and the number of received messages tells how many comments their own documents have collected all together. The asymmetric values have been used only with degree measure. Two latter analysis are calculated with symmetric values so that the received and given comments are summed up. None of these measures show statistically significant correlation to the course achievement of the student pairs.

Table 2. Centrality of the student pairs (comments, questions), calculated with students, the tutor and the supervisor

Student pair	Degree		Betweenness	Information	Credit (max 3)
	Received messages	Sent messages			
MiMe	30	16	0.29	7.69	2.5
AnLa	25	33	0.41	7.92	2
MiRi	18	8	0.41	6.95	Not passed
TeMi	31	7	0.41	7.47	2.25
MaSa	30	6	0.13	7.08	2.25
MiEl	43	33	0.41	8.16	2.75
ToNi	27	7	0.29	7.31	1.75
RiMi	26	3	0.13	7.08	Not passed
HaSa	10	13	0.13	6.75	2.25
Tutor	1	114	9.41	8.49	
Supervisor 2	0	1	0.00	1.07	

The betweenness measures give the probability of how often a specific actor occurs on the shortest information flow between some actors. Using this perspective the tutor gets a special position while the student pairs are not so different from each other. The information measure of Stephenson and Zele serves another view: there is one outsider (supervisor 2) and all other participants are near each other. This measure does not give any particular position for the tutor either. The information measure also gives a very similar result as can be seen in Figure 1.

The MDS calculation resulted in coordinates in three dimensions with a good stress value (0.04). These coordinates symbolizing the student pairs are then drawn in a VRML (Virtual Reality Modeling Language) model that can easily be studied (Figure 1.). To further visualize the CSCL communication we added the strongest links (cut point 3) between the student pairs.



Figure 1. The participatory approach: Tutor, supervisor and student pairs, 3D MDS model from two perspectives

red: class teacher education; yellow: educational science & class teacher education; blue: educational science; green: educational science & psychology; pink: psychology, black: the tutor & supervisor 2

The star-like shape of the chart shows that the structure of the communication is very centralized. In the chart the tutor has all the students around her and the supervisor does not seem to participate in CSCL communication.

As the tutor has such a central position we calculated the MDS coordinates (stress value 0.02) without the tutor and the supervisor. In this way we can take a closer consideration on students' interrelationships (see Figure 2). In the result VRML model we can see one pair (MiEl) located in the center.

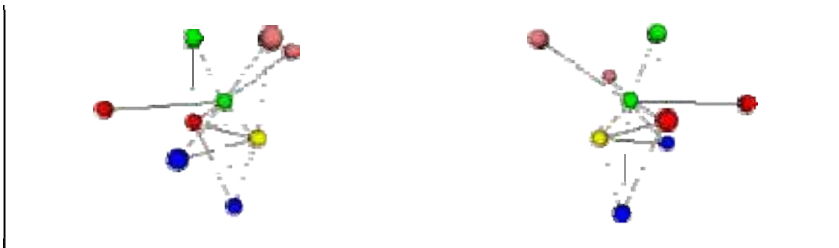


Figure 2. The participatory approach: Student pairs, 3D MDS model from two perspectives

Surprisingly we cannot find many communication links between the student pairs in the same domain or discipline. It was assumed that discussions would have been easier within the same disciplines that form more coherent knowledge cultures.

#### **Knowledge building: links between documents**

Of the more than 200 documents, only 34 were linked to at least one other document. These 34 documents form altogether 11 components (subgroups of linked documents) that consisted of 6 document pairs, 3 components of 3 documents, one component of 5 documents and one component of 8 documents (Figure 3). As nearly all components were



small (only 2 or three documents) and in the component of 5 documents the links were made between the same student pair's documents we considered the most important component as the one with 8 documents.



Figure 3. The linked documents with their identification number grouped by student pair and assignment

#### Combining two perspectives: knowledge network

The component of eight documents was linked together by four student pairs. Only one of these references was "against" while the rest were supportive "for" references. When using only references as indicators, one document is emphasized in the cluster (document 83). This document, however, is referenced only by the documents made by one student pair.

We combined reference information of the component to information about student pairs that made comments or questions (Figure 4). The account of comments and questions highlights document 55 in the component. This document had only one reference from one document of the component, but four comments from separate student pairs. The student pair that made this document also commented the reference document emphasized earlier. It should also be noted that the tutor did not play an important role here.

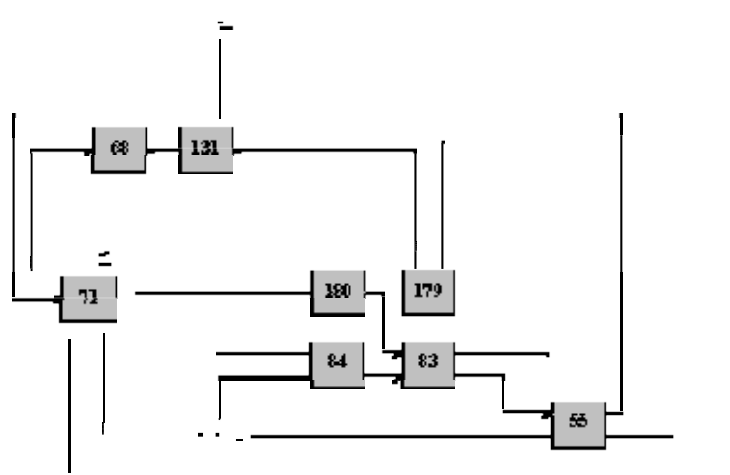


Figure 4. The biggest document component made of references, the comments and commentaries to the documents

There can be many documents with great importance to the learning community that this analysis does not emphasize (trivial examples are assignments and reference material). While the technique used seems to be incapable of making sure of the role of these documents, it can produce more information when revealing the structure of the work done in the learning environment. To get a deeper evaluation of the CSCL environment, the contents of the documents should also be studied. Here we focus only on the content of the eight most strongly interrelated documents.

When examining the contents of the documents, some remarks can be made. The content of the document that had the most comments and links (document 55) does not seem to be very remarkable. It doesn't contain any theory, but it did contain small comments, descriptions and interpretations as well as one new idea that had been overlooked by the others. The document could not be considered mature or done from the theoretical perspective. The link turned to refer to similarities between the documents, not dissimilarities. The comments were more like encouragement ("Good point!", "We found that too!") or elaborations of the idea. In addition, the comments were not real dialogues between the student pairs.

## Conclusions

According to our results, log files can facilitate the CSCL evaluation process in many ways. First, it can be a rather quick way to select and organize even large amounts of information. Second, the summaries of different operations tell about learning activities as a whole. Thus the log files facilitate the evaluation process when the structure of a learning community, a general view of the learning process or the subject of the most central content are to be found.

Social network analysis is appropriate when studying structures and relationships. These methods provide an approach that is not easily achieved by other tools. Some methods could be particularly valuable when evaluating CSCL environments. There are useful techniques to find the central actors or issues. These tools can be used to scale the environment, for example in the basis of the frequencies of interaction between the actors. Social network analysis could also focus on dynamics, where visualizations do play an important role with colors, three-dimensional representation and animation. Graphical figures are particularly useful, since they enable researchers to display, in a compact way, the position of individual actors: how they are related to each other, and what the overall structure looks like.

As a result of this special case we can assume that the network analysis of log files helps us to understand the working processes in CSCL at least from the participation perspective. The knowledge acquisition processes did not seem to get such a visible form that we could argue for it in the same way. However, the acquisition approach should not be fully replaced by the emerging participation approach. Apart from the description of activities and discourse processes, we should also consider the knowledge acquisition in CSCL environments. This means that better tools to follow the elaboration of documents and interrelated groups of documents should be developed.

## References

Borgatti, S., Everett, M. & Freeman, L. (1996a). *UCINET IV Version 1.64*. Natick, MA: Analytic Technologies.

Borgatti, S., Everett, M. & Freeman, L. (1996b). *UCINET IV Version 1.64 Reference Manual*. Natick, MA: Analytic Technologies.

Brown, A. L. & Campione, J. C. (1996) Psychological theory and the design of innovative learning environments: On procedures, principles, and systems. In L. Schauble. & R. Glaser (Eds.) *Innovations in learning. New environments for education*. (pp. 289-325). Mahwah, NJ: Lawrence Erlbaum.

Brown, A. L., & Palincsar, A. S. (1989) Guided, cooperative learning and individual knowledge acquisition. In L. Resnick (Ed.), *Knowing, learning, and instruction: Essays in Honor of Robert Glaser*. (pp. 393-451) Mahwah, NJ: Lawrence Erlbaum.

Collins, A., Brown, J.S. & Holum, A. (1991). Cognitive Apprenticeship: Making Thinking Visible. *American Educator*, 6-11, 38-46.

Hakkarainen, K. & Lipponen, L. (1998, April). Epistemology of inquiry and computer-supported collaborative learning. Paper presented April 13-17 1998, at the Annual Meeting of the American Educational Research Association, San Diego.

- Hatano, G. & Inagaki, K. (1992) Desituating cognition through the construction of conceptual knowledge. In P. Light & G. Butterworth (Eds.) *Context and cognition. Ways of knowing and learning*. (pp. 115-133). New York: Harvester.
- Haythornthwaite, C., Wellman, B. & Mantei M. (1995). Work Relationships and Media Use: A Social Network Analysis. In *Group Decision and Negotiation*, 4: 193-211.
- Hutchins, E. (1991) The social organization of distributed cognition. In L. B. Resnick, J. M. Levine & S. D. Teasley (Eds.). *Perspectives on socially shared cognition*. (pp. 283-307). Washington, DC.: American Psychological Association.
- Hutchins, E. (1995) *Cognition in the wild*. Cambridge, MA: The MIT Press.
- Lehtinen, E. & Repo, S. (1996). Activity, social interaction and reflective abstraction: Learning advanced mathematics in a computer environment. In S. Vosniadou, E. De Corte, R. Glaser & H. Mandl (Eds.), *International perspectives on the design of technology supported learning environments* (105-128). Mahwah, NJ: Lawrence Erlbaum.
- Lehtinen, E. & Rui, E. (1996). Computer supported complex learning: An environment for learning experimental method and statistical inference. *Machine Mediated Learning* 5 (3&4), 149-175.
- Miyake, N. (1986). Constructive interaction and the iterative process of understanding. *Cognitive Science*, 10, 151-177.
- Oatley, K. (1991) Distributed cognition. In H. Eysenck, A. Ellis, E. Hunt, & P. Johnson-Laird (eds.) *The Blackwell dictionary of cognitive psychology*. (pp. 102-107). Oxford: Blackwell Reference.
- Scardamalia, M., & Bereiter, C. (1989). Schools as knowledge-building communities. Paper presented at the Workshop on Development and Learning Environments, University of Tel Aviv, Tel Aviv, Israel, October 1989.
- Scott, J. 1991. *Social Network Analysis*. A handbook. London: SAGE Publications.
- Sfard, A. 1998. On two metaphors for learning and the dangers of choosing just one. *Educational Researcher* 27(2), 4-13.
- Wasserman, S. & Faust, K. 1995. *Social network analysis. Methods and applications*. Cambridge university press.
- Wittenbaum, G.M. & Stasser, G. 1996. Management of information in small groups. In J. L. Nye & A.M. Brower (Eds.) *What's social about social cognition? Research on socially shared cognition in small groups*. SAGE. Thousand Oaks. London. New Delhi. Pp. 3-28.