Réutilisation d'un corpus pour une nouvelle analyse des réseaux sociaux grâce à l'adaptation de l'outil KSV

Christophe Reffay, Christopher Teplovs

Productive re-use of CSCL data and analytic tools to provide a new perspective on group cohesion

Christophe Reffay, STEF, INRP - ENS Cachan - UniverSud, Cachan, France. christophe.reffay@inrp.fr
Christopher Teplovs, University of Toronto, Ontario, Canada. chris.teplovs@utoronto.ca
François-Marie Blondel, STEF, INRP - ENS Cachan - UniverSud, Cachan. francois-marie.blondel@inrp.fr

Abstract: The goals of this paper are twofold: (1) to demonstrate how previously published data can be re-analyzed to gain a new perspective on CSCL dynamics and (2) to propose a new measure of social cohesion that was developed through improvements to existing analytic tools. In this study, we downloaded the Simuligne corpus from the publicly available Mulce repository. We improved the Knowledge Space Visualizer (KSV) to deepen the notion of cohesion by using a dynamic representation of sociograms. The Calico tools have been used and extended to complete this cohesion measure by analyzing lexical markers. These complementary analyses of cohesion, based on clique sizes and communication intensity on the one hand, and lexical markers on the other hand, offer more detailed information on the communication intensity and structures, and the position of their participants. They particularly show strong convergences that were not visible in the previous analysis.

Introduction

Because of their complexity, authentic learning experiences are hard to replicate. This makes comparison and validation of research tools, methods and results in our CSCL field difficult. Research collaboration has been well advocated by Chan et al. (2006) in the context of Technology Enhanced Learning:

> There is urgent need of putting together complementary strengths and contexts and combining our insights as rapidly as possible to make a greater impact and further elevate our research quality at the same time. Research generally has had a small voice in national educational outcomes; we can speak louder if we speak together. (Chan et al., 2006, p. 21)

This issue has been addressed by various projects that have been concerned with data sharing within communities of researchers.

Data sharing

In the research data sharing perspective, the Dataverse Network project (2010) described by King (2007), shows why datasets have to be made available, or at least identified and recorded as persistent, authorized, and verifiable data. When connected to a traditional paper published in a journal or a conference, such a data *publication* would increase the value of the article and of the related journal (or conference proceedings).

For the Intelligent Tutoring System (ITS) field, the PSLC DataShop (2010) presented by Koedinger et al. (2008) provides a data repository including data sets and a set of associated visualization and analysis tools in order to evaluate the action/feedback interaction between learners and (virtual) tutor tools.

In the CSCL community, the DELFOS framework (Osuna, 2000; Osuna et al., 2001) provides an XML based data structure (Martinez et al., 2003) for collaborative actions in order to promote interoperability (between analysis tools), readability (either for human analysts and automated tools) and adaptability to different analyzing perspectives. Some of these authors joined the European research project (JEIRP–IA) on Interaction Analysis and reported in (Martinez et al., 2005) a template describing IA tools and a common format. This project focused on technical interoperability in order to be able to apply any tool on any data.

A major project has been initiated by the Virtual Math Team (2010). Multimodal Chat sessions (from the Virtual Math Forum) have been collected and shared among numerous (28) researchers coming from 11 countries, 18 institutions and 8 different research fields. Every collaborator applied his/her own analysis methods and tools processing these interaction data in order to see what came up. The result is reported in (Stahl, 2009). The same data set has been used also for a Workshop at the CSCL2009 conference.

CSCL environments are capable of recording vast amounts of data that represent both content and process. With the explosive growth in the use of CSCL environments we are faced with potentially more data than can be analyzed. Various projects offer the possibility of re-using data but few studies report on the productive re-use of CSCL data.

The Mulce project (Mulce, 2010) developed a platform to share learning and teaching corpora. This new possibility should deepen our understanding of well contextualized situations and hopefully better validate tools and have a greater impact on the real world of (collaborative online) learning. Even if more than 25 complex objects are already publicly available on this repository, there is still no evidence of productive re-use of these corpora. The purpose of this paper is to show how productive such a re-use can be.

Revisiting Social Network Analysis with new capabilities

Social interactions are an inherent aspect of computer-supported collaborative learning (CSCL). Considering participants as a social network (Wellman, 2001) provides a framework that can help us understand what are often complex patterns of interaction. Several studies have used techniques from social network analysis (Wasserman & Faust, 1997) to examine patterns of interaction among CSCL participants (de Laat, Lally, Lipponen, & Simons, 2007; Liao, Li, Wang, Huang, & Zhang, 2007; Martínez, Dimitriadis, Rubia, Gomez, & de la Fuente, 2003). Nurmela, Lehtinen, & Palonen (1999) suggest that social network analysis can provide useful tools in situations where traditional, statistical methods may not be suitable or where their use may obscure interesting results.

Wang and Li (2006) provide a brief history of social network analysis and its application to CSCL. They trace the development of SNA from the early 1930s through its growth in popularity in the 1970s and its emergence as a methodology in the American Evaluation Association's conference program in 1998. Social network analysis is often concerned with the detection of individuals' social position or role (Pattison, 1994). To this end, a variety of measures have been proposed. These measures include indegree, outdegree, betweenness, density and cohesion. Cho, Stefanone , and Gay (2002) used two centrality measures to test various hypotheses about the peer behaviors in response to an actor's centrality in a CSCL community.

Aviv, Erlich, Ravid & Geva (2003) combined the use of social network analysis to determine cohesion (through clique analysis) with manual content analysis. They found that although maintaining high degrees of cohesion required substantial effort on the part of participants, such groups were characterized by high phases of critical thinking. Martinez et al. (2003) similarly combined qualitative analysis of content with social network analysis.

We were interested in re-examining a data set that had been previously used for a social network analysis. Reffay & Chanier (2003) analyzed the data set described in the next section in terms of cohesion. The goals of this paper are twofold: (1) to demonstrate how previously published data can be re-analyzed to gain a new perspective on CSCL dynamics and (2) to propose a new measure of social cohesion that was developed through improvements to existing analytic tools. After providing a description of the data we describe how an existing tool was modified to facilitate the development of a more sophisticated measure of cohesion. This analysis (based on cliques) is compared with a completely different method using lexical markers.

The Simuligne data re-used

Simuligne is a distance French as a foreign language learning situation, which has been used as experimentation in 2001, in a trans-disciplinary research project. In this Simuligne learning situation, the global simulation method was generally used for intensive face-to-face language learning courses. It has been adapted to this extensive online learning situation in parallel in 4 basic groups. Distance was the rule: everybody worked at a distance; none of the learners had ever met before Simuligne.

We count 40 learners (English adults in professional training, registered at the Open University, UK), 10 natives (French teacher trainees from the Université de Franche-Comté, Besançon, FR), 4 tutors (teachers of French from the Open University) and one (French) pedagogical coordinator. All agents were dispatched into four basic learning groups, namely: *Aquitania*, *Lugdunensis*, *Narbonensis* and *Gallia*. Each of these groups counted 10 learners, a couple of natives and one dedicated tutor. During the whole period, the special space named "*Formateurs*" was dedicated to a meta-level of monitoring for the 4 tutors, 10 natives and the coordinator.

Three groups out of four achieved the simulation, which is a high ratio in distance- learning. On May, the 31st, the *Lugdunensis* group broke up and two of its most active learners were transferred to *Aquitania* group. The whole Simuligne learning session produced 1 531 086 characters in communication tools: forum (57%), e-mail (27%), chat (15%) through out the 4 basic groups (*Aquitania*, *Gallia*, *Lugdunensis*, *Narbonensis*) plus 2 other discussions: *Formateurs* and *Monde*.

The Simuligne data was recorded as a corpus and shared on the Mulce Platform since 2009. In this study, we focus on the forum exchanges in the 4 basic groups only and pay attention on the period before the *Lugdunensis* group broke up, i.e. from April the 3rd to May the 31st.

The Knowledge Space Visualizer

The Knowledge Space Visualizer (KSV) is a software tool that was originally developed to facilitate the exploration of social and semantic networks in data collected from online discourse environments (Fujita & Teplovs, 2009; Teplovs, 2010; Teplovs & Scardamalia, 2007). The source code is freely available (http://code.google.com/p/ksv). Sha and van Aalst (2009) and van Aalst, Teplovs and Sha (2010) have used the KSV to investigate the relationships between structural networks of documents (i.e. relationships based on replying, referencing, and annotating) and the semantic links between those documents. In the current study the KSV was modified to allow the representation of social links between authors from the Simuligne data set.

For the purposes of conducting a social network analysis of the Simuligne data, a relationship between authors was established if an author had read a posting made by another author in the online forum. The strength of the tie between two authors was based on the number of postings read. The relationships were symmetrical. For example, in the case of two authors, Author A and Author B, they were the sum of the number of posts made by Author A that were read by Author B and the number of posts made by Author B that were read by Author A.

The Knowledge Space Visualizer differs from other commonly used social network analysis software such as UCINET (Borgatti, Everett, & Freeman, 2002) in that it facilitates the manipulation of various parameters such as dates, authors, and layout. Of particular importance is the ability to manipulate various thresholds. For social network analysis, the threshold that determines whether two actors are linked by a tie can be manipulated. In previous work with the data set used for this study, the threshold was 10 interactions. That is, actors had to have interacted (by reading each others' postings) at least 10 times for them to be considered linked. The KSV allows the linking threshold to be easily adjusted to any value from zero to the maximum number of interactions between actors. Manipulating the linking threshold allows the user to investigate the nature of the social network at varying levels of communication intensity. The KSV also allows the social network to be examined for any given date. For the purposes of this study, the social network was considered to be cumulative from the beginning of the course.

The KSV, when used in this exploratory fashion, allows the user to freely manipulate the representation of the network. By doing so, the user can find interesting patterns within the network structures and then conduct additional manipulations of the representations.

Calico tools to analyse computer mediated discussions

The Calico platform (Calico, 2010) was developed for sharing and analyzing discussion forum objects (Giguet & al. 2009). Calico is a shared workspace that proposes several ways to display the contents of a forum (ShowForum), to compute quantitative and qualitative indicators about authors, interactions and topics (Authagora, Volagora, Colagora) and to offer new ways to display global or local information about a forum (Anagora, Bobinette). For the purpose of our analysis which focuses here on specific lexical markers, the Colagora and Bobinette were used to give both general and local measures and views on the utterances of these markers.

Colagora allows the user to build up lexical topics, defined with lists of words (or regular expressions) taken from the messages and then to explore the forums through these highlighting filters. Colagora displays word occurrences and highlights every matching word of the forum with the colour linked to the topic.

Bobinette was developed by Huynh Kim Bang and Bruillard (2005) to facilitate reading large forums with usual interfaces. Bobinette displays a forum on a grid where the threads are drawn on horizontal lines and days of posts are indicated in columns. Bobinette computes statistics about word topics for each post, each thread, each day and the whole forum. The content of a selected post can be displayed with the topic words highlighted.

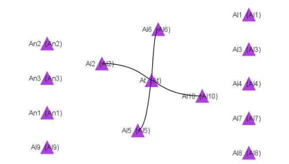
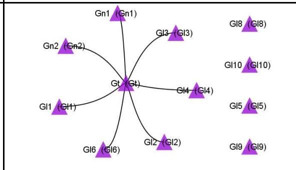
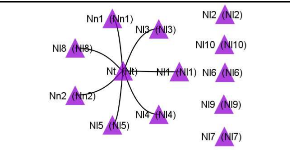
Results

Our exploration across intensity led us to the core / periphery model of (Borgatti & Everett, 1999) and more specifically the Freeman star (Freeman, 1979), i.e.: for a given threshold: Who is at the center of the star? Who is a branch of it? Who is not connected?

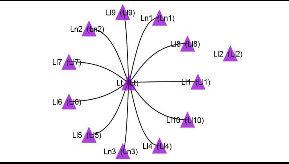
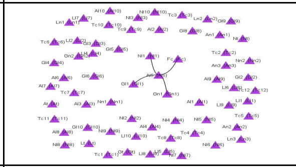
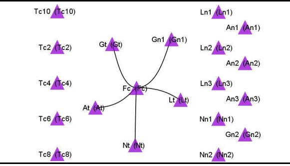
Social Network Analysis using KSV

KSV gave us the opportunity to observe these phenomena across all the possible intensity values, observing cliques formation along this dimension. This star shape was chosen because it occurs in most situations and shows the most central actor in the discussion. This characteristic threshold is exactly 1 linking threshold unit more than the threshold that yields the first 3-clique(s). The maximum level of a 3-clique shows the maximum intensity of interaction for the triad, and we know that this is much more interesting for group cohesion and collaboration than the maximum intensity between any pair of actors. The "star" shapes and thresholds for each group are shown in Tables 1 & 2. The period of the study is bounded by the starting date and the date of official closure of the *Lugdunensis* group.

<u>Table 1: Star shapes and thresholds for 3 basic well functioning groups for the first 8 weeks of the Simuligne learning session.</u>

Star (threshold)	*Aquitania*	*Gallia*	*Narbonensis*
Star shape			
Intensity (threshold)	129	36	49
Nb of branches	4	7	7
Who is the center?	Tutor	Tutor	Tutor
Who is around?	4 Learners	5 Learners, 2 Natives	5 Learners, 2 Natives

<u>Table 2: Star shapes and thresholds for the braking up group Lugdunensis and non basic groups, during the first 8 weeks of the Simuligne learning session.</u>

Star (threshold)	*Lugdunensis*	*Monde*	*Formateurs*
Star shape			
Intensity (threshold)	12	77	121
Nb of branches	11	4	5
Who is the center?	Tutor	Al5 (Learner)	Coordinator
Who is around?	8 Learners, 3 Natives	2 L, 1 N, 1 coordinator	4 Tutors, 1 Native

In tables 1 & 2, we can observe that:

Each group show a single well formed star (with more than 2 branches);

Among the 4 basic groups, *Gallia* and *Narbonensis* are very similar and *Aquitania* and *Lugdunensis* very different. These comments are equally true for the shapes, the center and the thresholds;

The center of each of the 4 basic group's star is the tutor. The center for the "*Formateurs*" is the coordinator. However, the center of the "*Monde*" is a learner;

Aquitania's star is the only one where no native appear in the star branches;

The threshold value is a good indicator of the intensity of the exchanges between members for each group. Extreme's values are from *Lugdunensis* and *Aquitania*.

On Figure 1, we show 4 curves (one for each basic group). The vertical axis represents the intensity and the horizontal one the maximum size of cliques. The first point of *Aquitania's* curve is (k=3, intensity=128). This means that the highest intensity reached by any 3-clique in *Aquitania* is 128. Remember that a k-clique is a graph with k fully connected nodes. We can see that the value of the star's threshold (of tables 1 & 2) is the top of the curves for each group. This particular diagram does not indicate precisely which actors are involved in these cliques, because its purpose is to compare the groups.

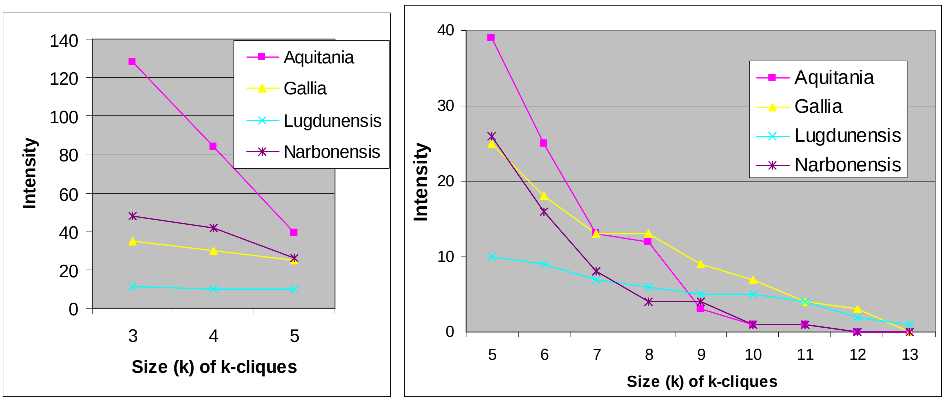


Figure 1a & 1b: Maximum intensity of exchange for each size of cliques on the 4 basic groups.

In order to get a more precise view of these curves for k-cliques with k≥5, the scale of intensity has been changed from the left part (1a) to the right part (1b). We can observe that the intensity of the *Aquitania's* internal kernel (up to 4-cliques) is twice greater than the best (*Narbonensis*) of the other 3 kernels. Up to 7-cliques, *Aquitania* shows the greater intensity. But for bigger cliques (k>7), *Gallia's* intensities dominate the graph. This clearly means that *Gallia's* exchanges where reasonably high in intensity but engaged much more members than *Aquitania*. Another interesting value is given for 9-cliques where *Aquitania's* and *Narbonensis* curves (that are very high on fig. 1a) become lower than the *Lugdunensis* one as if the very high number of exchanges between the core members were discouraging the peripheral members. The more reasonable amount of communication between core members of *Gallia* seems having kept more members in. The shape of *Aquitania's* and *Narbonensis* curves are very similar (especially on fig. 1b): they start very high and decrease rapidly. *Lugdunensis* curve has exactly the opposite characteristics and *Gallia's* curve is between both of these shapes.

Finally, this analysis of cliques and density for the 4 basic groups shows that (1) *Lugdunensis* cliques (even little ones) have very low density, (2) Aquitania and *Narbonensis* have very similar cliques characteristics across density: a restricted core very active and very few communication exchanged in medium and large cliques. The *Gallia's* core (little cliques) shows a lower intensity but this is the group where medium and large cliques have the more intensive exchanges.

The next part proposes a study of cohesion based on the same data, by using lexical markers (provided by Calico). The discussion will show convergences and discrepancies between SNA and lexical analysis of cohesion.

Use of Calico to analyze cohesion through pronoun markers

Studying the interpersonal relationships in textual communication, Yates (1996) has found that participants use first and second pronouns more often in on-line discussions than in usual written communication. This study was made by counting personal pronouns. Following the same technique, we defined four markers that may reflect group cohesion: "first-person singular" (FPS) markers (I, my, me... / je, j', me, m', mon, ma, mes…), "second-person plural" (SPP) markers (you, your… / vous, votre, vos…) and "first-person plural" (FPP) markers (we, our, ours... / nous, notre, nos…). It should be noticed that in French the second-person plural ("vous") is different from the second-person singular ("tu").

The Colagora tool computes the occurrences of each marker and highlights related words in messages and the Bobinette tool visualizes theses occurrences among the different threads and periods, giving sub totals of occurrences by periods and threads.

We first consider the use of pronouns in the messages of all groups (Table 3) and their potential relations with cohesion. Figures from the *Lugdunensis* group should be considered with precaution because of the low number of messages.

The "first person singular" (FPS) markers are very frequent as already observed by Yates (1996) and the percentages are very similar for all groups.

Table 3: Messages and lexical markers in the four basic groups.

	messages	% of messages with FPS (I)	% of messages with SPP (you)	% of messages with FPP (we)
Aquitania	348	77%	30%	24%
Gallia	159	81%	51%	20%
Narbonensis	175	77%	29%	25%
Lugdunensis	73	70%	41%	18%

When examining interactivity in discussion groups, Rafaeli and Sudweeks (1997) found that 24.5% of the messages they qualified as "interactive" contain first-person plural pronouns, which is significantly greater than the percentage calculated for the entire corpora of messages (only 9.3%).

Except for the *Lugdunensis* group, the three groups have percentages very similar to those found by Rafaeli and Sudweeks. We can assume that the participants of these groups are in a rather similar situation because they are invited to interact with each others in the same group. The lower percentage of *Lugdunensis* can be interpreted as a possible indicator of lower group cohesion (note that the number of messages of this group is also the lowest).

The use of "second-person plural" markers (SPP) is different among groups and higher in the *Gallia* group.

Using Bobinette we can explore which actors and what messages contain the most significant number of markers in different groups. On Figure 2 is a visualization of 8 threads of the *Aquitania* forum. The FPS markers are in blue, SPP in orange and FPP in red. One line of the grid stands for a thread, one row for a day.

Messages appear as circles followed by the occurrences for each marker with its corresponding color. The message posted on May, the 20th, displayed by clicking on the circle, shows the highlighted words: 3 FPS (*j'*, *je*), 2 FPP (*nous*) and no SPP (*vous*). The horizontal margins of the table cumulate the occurrences for the whole thread, whereas the vertical ones make sub-totals for each day.

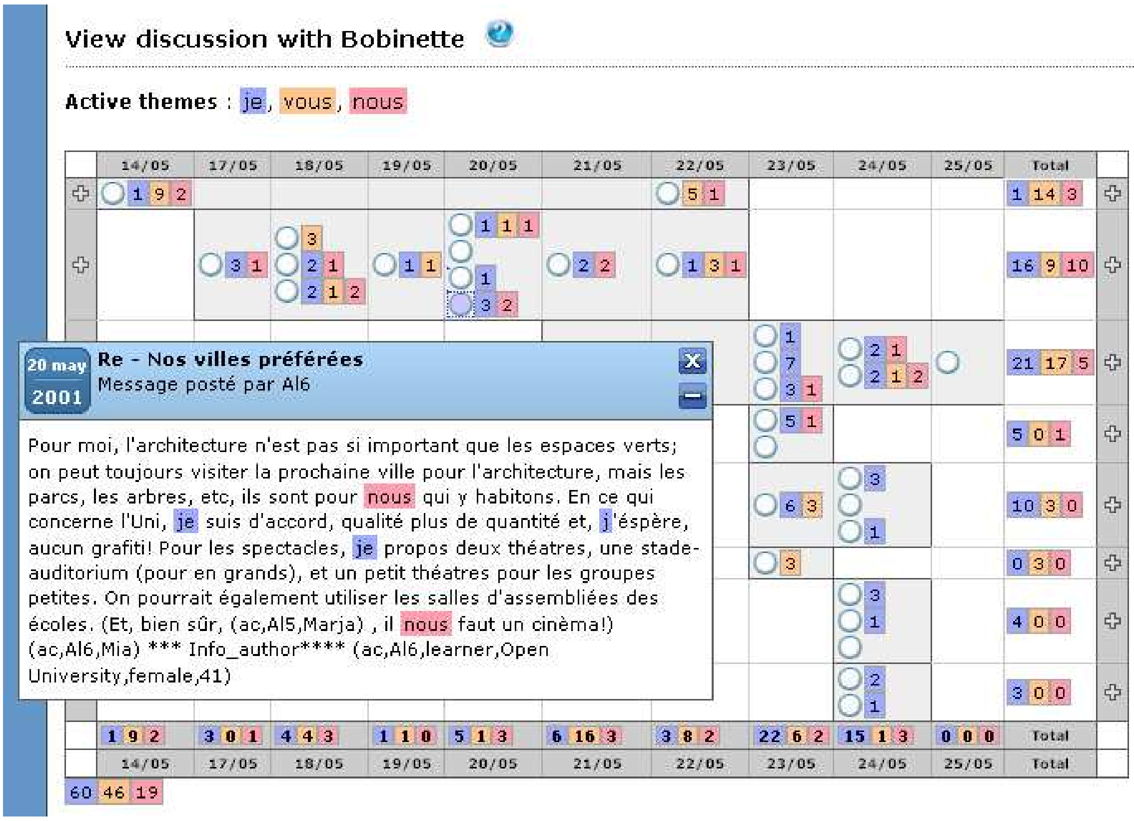


Figure 2: Visualization of 8 threads of the *Aquitania* forum with Bobinette.

Looking more closely at the number of messages for each actor, we found that the 3 well functioning groups counted more messages (from 159 to 384) and at least 50% of messages posted by learners. Conversely, *Lugdunensis* only got 73 messages and only 36% of them were posted by learners. The part of messages posted by natives is more important in Gallia (16%) than in *Aquitania* (7%) and *Narbonensis* (5%).

Using restriction on authors, Bobinette shows that *Aquitania's* and *Narbonensis* tutors wrote quite the same high number (80) of FPP ("we") (only 23 "we" written by *Gallia's* tutor). Furthermore, this abundance of "we" for both of these groups is concentrated in the tutors messages. In *Aquitania*, 58% of the "we" are written by the tutor. This proportion rises to 72% for *Narbonensis*, whereas it is only 44% for *Gallia*.

From these results emerges a strong similarity between *Aquitania* and *Narbonensis*, that was not visible in the previous analysis made by Reffay & Chanier (2003).

Discussion

As a first result, and if we consider the tools adaptation or improvements, we can say that this sharing experience was useful to increase robustness and enrich the initial model for Bobinette (Calico) and KSV. In Bobinette, sub-totals have been added in the margins of the output table (Fig. 2) and restriction by authors is now fully operational. KSV has been more significantly modified to be able to deal with very different graphs (Sociograms of forum members instead of semantic distance between notes of the Knowledge Forum). The major interest of KSV is still the ability to tune the threshold, but the adaptation to new relationships (unbounded integers instead of normalized cosines between -1 and 1).

The comparison of group cohesion has been made on the same experiment with 2 very different techniques: SNA and lexical markers. These analyses showed some convergences. They both conclude that intensity of exchanges (for SNA) or number of messages (lexical) are very (too?) low for the *Lugdunensis* group to be considered. Lexical analysis shows that the use of "we" is similar in *Aquitania* and *Narbonensis* groups, and both groups also show similar cliques structures across density.

An important improvement has been made on cohesion analysis on the Simuligne experiment (see Tables 1 & 2 and Figure 1). While the result in (Reffay & Chanier, 2003) selected a given threshold and draw the cliques only for that value, the KSV allowed us in this study to characterize cliques for all density values.

Conclusion

We have demonstrated that existing data, if described in sufficient detail, can be used with other tools. Moreover, we have shown that it is possible to adapt existing tools – which represents considerable savings in terms of time and money compared with developing a tool from scratch – for use with data that they were not initially designed to analyze. Admittedly the process of data re-use and tool modification was somewhat easier than can typically be expected because the data provider worked with the tool developer. Nevertheless, we believe that such developments will become easier and perhaps even commonplace with the current trend to openness: "open source" in the case of tools and "open data" in the case of CSCL data. There are problems with openness. For example, institutions may assert ownership of patents or copyrights associated with code, and data may be unavailable for publication as a result of stipulations imposed by ethical review procedures. We see these as challenges to be addressed rather than reasons to abandon this approach.

References

Aviv, R., Erlich, Z., Ravid, G., & Geva, A. (2003). Network analysis of knowledge construction in asynchronous learning networks. *Journal of Asynchronous Learning Networks, 7*(3).

Borgatti, S. P., Everett, M. G. (1999). Models of core/periphery structures, *Social Networks, 21*, 375-395.

Borgatti, S. P., Everett, M. G., & Freeman, L. C. (2002). *UCINET 6 For Windows: Software for Social Network Analysis*. Harvard, MA.

Calico (2010). *The Calico platform*. http://www.crashdump.net/calico/

Chan, T., Roschelle, J., Hsi, S., Kinshuk, Sharples, M., Brown, T., Patton, C., Cherniavsky, J., Pea, R., Norris, C., Soloway, E., Balacheff, N., Scardamalia, M., Dillenbourg, P., Looi, C., Milrad, M., Hoppe, U. (2006). One-to-one technology-enhanced learning: An opportunity for global research collaboration. In *Research and Practice in Technology Enhanced Learning, 1*(1), 3-29.

Cho, H., Stefanone, M., & Gay, G. (2002). Social network analysis of information sharing networks in a CSCL community. In Gerry Stahl (Ed.), *Proceedings of the Conference on Computer Support for Collaborative Learning: Foundations for a CSCL Community (CSCL '02)* (pp. 43-50). International Society of the Learning Sciences.

Dataverse (2010). *The Dataverse Network Project*: http://thedata.org/

de Laat, M., Lally, V., Lipponen, L., & Simons, R.-J. (2007). Investigating patterns of interaction in networked learning and computer-supported collaborative learning: A role for Social Network Analysis. *International Journal of Computer-Supported Collaborative Learning, 2*(1), 87-103.

Fujita, N., & Teplovs, C. (2009). Automating the Analysis of Collaborative Discourse: Identifying Idea Clusters. In A. Dimitracopoulou, C. O'Malley, D. Suthers & P. Reimann (Eds.), *Computer Supported Collaborative Learning Practices: CSCL 09 Community Events Proceedings* (pp. 162-164). International Society of the Learning Sciences.

Freeman, L. C. (1979). Centrality in social networks conceptual clarification. *Social Networks, 1*(3), 215-239.

Giguet, E., Lucas, N., Blondel, F.-M., & Bruillard, É. (2009). *Share and explore discussion forum objects on the Calico website*. In A. Dimitracopoulou, C. O'Malley, D. Suthers & P. Reimann (Eds.), *Computer Supported Collaborative Learning Practices: CSCL 09 Community Events Proceedings* (pp. 174-176). International Society of the Learning Sciences.

Huynh Kim Bang, B., & Bruillard, E. (2005). Vers une nouvelle interface de lecture pour des forums de discussion dédiés à des élaborations collectives. In Saleh I. & Clement T. (Eds.), *Créer, jouer, échanger, Actes de H2PTM'05* (pp. 43-56). Paris: Lavoisier.

King, G. (2007). An Introduction to the Dataverse Network as an Infrastructure for Data Sharing. *Sociological Methods & Research*, 36(2). http://gking.harvard.edu/files/dvn.pdf

Koedinger, K.R., Cunningham, K., Skogsholm, A. (2008). An open repository and analysis tools for fine-grained, longitudinal learner data. In *Proceedings of the First International Conference on Educational Data Mining* (pp. 157-166). http://www.educationaldatamining.org/EDM2008/uploads/proc/16_Koedinger_45.pdf

Liao, J., Li, Y., Wang, J., Huang, R., & Zhang, Z. (2007). A systemic model of interaction analysis in CSCL. In V. Uskov (Ed.) *Proceedings of the 10th IASTED International Conference on Computers and Advanced Technology in Education* (pp. 463-469). Anaheim, CA, USA. ACTA Press.

Martinez, A., De la Fuente, P., Dimitriadis, Y.(2003). Towards an xml-based representation of collaborative action. In Wasson, B., Ludvigsen, S., & Hoppe, U. (Eds.), *Designing for change in networked learning environments: Proceedings of the International Conference on Computer Support for Collaborative Learning 2003*. Dordrecht, Netherlands: Kluwer Academic Publishers. http://hal.archives-ouvertes.fr/hal-00190429/fr/

Martínez, A., Dimitriadis, Y., Rubia, B., Gomez, E., & de la Fuente, P. (2003). Combining qualitative evaluation and social network analysis for the study of classroom social interactions. *Computers & Education, 41*(4), 353-368.

Martinez, A., Harrer, A., Barros, B. (2005). *Library of Interaction Analysis Tools. Deliverable D.31.2 of the JEIRP IA* (Jointly Executed Integrated Research Project on Interaction Analysis Supporting Teachers & Students' Self-regulation). KaleidoScope. http://www.rhodes.aegean.gr/ltee/kaleidoscope-ia/Publications/D31-02-01-F%20Lirbary%20of%20IA%20tools%20.pdf

Mulce (2010). *Mulce: French national research project 2006-2010 (ANR-06-CORP-006)*, coordinated by T. Chanier. http://mulce.org

Nurmela, K., *Lehtinen*, E., & Palonen, T. (1999). Evaluating CSCL log files by social network analysis. In C. M. Hoadley, J. Roschelle, (Eds.), *Proceedings of the Computer Support for Collaborative Learning 1999 Conference*. (pp. 434-444). Palo Alto, CA: Stanford University.

Osuna, C. (2000). *DELFOS: A Telematic and Educational Framework based on Layer oriented to Learning Situations.* PhD Dissertation, Universidad de Valladolid, Valladolid, Spain

Osuna, C., Dimitriadis, Y., Martínez, A. (2001). Using a Theoretical Framework for the Evaluation of Sequentiability, Reusability and Complexity of Development in CSCL Applications. In P. Dillenbourg, A. Eurelings & K. Hakkarainen (Eds.), *Proceedings of EuroCSCL 2001: European Perspectives on Computer-Supported Collaborative Learning*, Maastricht, NL.

Pattison, P. E. (1994). Social cognition in context: Some applications of social network analysis. In S. Wasserman & J. Galaskiewicz (Eds.), *Advances in social network analysis in the social and behavioral sciences* (pp. 79-109). Newbury Park, CA: Sage.

PSCL Datashop (2010). *The Pittsburgh Science of Learning Center DataShop*. https://pslcdatashop.web.cmu.edu/

Rafaeli, S., & Sudweeks, F. (1997). Networked Interactivity. *Journal of Computer-Mediated Communication*, *2*(4). http://jcmc.indiana.edu/vol2/issue4/rafaeli.sudweeks.html

Reffay, C., & Chanier, T. (2003). How social network analysis can help to measure cohesion in collaborative distance-learning. In B. Wason, S. Ludvigson & U. Hoppe (Eds.), *Designing for change in networked learning. Proceedings of the international conference on Computer Supported Collaborative Learning* 2003 (pp. 343-352). Bergen, Norway: Kluwer Academic Publishers.

Reffay, C., & Betbeder, M.-L. (2009). Sharing corpora and tools to improve interaction analysis. In proceedings of the EC-TEL 2009, In *4th European Conference on Technology Enhanced Learning, Learning in the Synergy of Multiple Disciplines, LNCS 5794*. Springer.

Sha, L., & van Aalst, J. (2009). A 2x3 model of student-directed formative assessment in collaborative knowledge creation. In *Proceedings of the 17th International Conference on Computers in Education* (pp. 205-212*)*. Hong Kong: Asia-Pacific Society for Computers in Education.

Stahl, G. (2009). *Studying Virtual Math Teams*. New York, NY: Springer.

Teplovs, C. (2010). *Visualization of knowledge spaces to enable concurrent, embedded and transformative input to knowledge building processes*. Unpublished Ph.D., University of Toronto, Toronto.

Teplovs, C., & Scardamalia, M. (2007). Visualizations for knowledge building assessment. *Paper presented at the AgileViz workshop, CSCL2007.*

van Aalst, J., Sha, L., & Teplovs, C. (2010). A visualization of group cognition: Semantic network analysis of a CSCL community. In K. Gomez, L. Lyons, and J. Radinsky (Eds.) *Proceedings of the 9th International Conference of the Learning Sciences - Volume 1 (ICLS '10)* (pp. 929-936). International Society of the Learning Sciences.

Virtual Math Team (2010). http://www.ischool.drexel.edu/faculty/gerry/vmt/index.html

Wang, Y., & Li, K. (2006). An application of social network analysis in evaluation of CSCL. In R. Mizoguchi, P. Dillenbourg & Z. Zhu (Eds.), *Learning by effective utilization of technologies: Facilitating intercultural understanding*: IOS Press.

Wasserman, S., & Faust, K. (1997). *Social network analysis: Methods and applications*. Cambridge, UK: Cambridge University Press.

Yates S. J. (1996). Oral and Written Linguistic Aspects of Computer Conferencing: A Corpus Based Study. In S. C. Herring (Eds.), *Computer-Mediated Communication. Linguistic, Social and Cross-Cultural Perspectives* (pp. 29-46). Amsterdam: John Benjamins.