

How social network analysis can help to measure cohesion in collaborative distance-learning

Christophe Reffay, Thierry Chanier

▶ **To cite this version:**

Christophe Reffay, Thierry Chanier. How social network analysis can help to measure cohesion in collaborative distance-learning. Computer Supported Collaborative Learning, Jun 2003, Bergen, Norway. pp.343-352. edutice-00000422

HAL Id: edutice-00000422
https://edutice.hal.science/edutice-00000422

Submitted on 19 Apr 2004

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

C. REFFAY, T. CHANIER

HOW SOCIAL NETWORK ANALYSIS CAN HELP TO MEASURE COHESION IN COLLABORATIVE DISTANCE-LEARNING

Abstract. It has been argued that cohesion plays a central role in collaborative learning. In face-to-face classes, it can be reckoned from several visual or oral cues. In a Learning Management System or CSCL environment, such cues are absent. In this paper, we show that Social Network Analysis concepts, adapted to the collaborative distance-learning context, can help measuring the cohesion of small groups. Working on data extracted from a 10-week distance-learning experiment, we computed cohesion in several ways in order to highlight isolated people, active sub-groups and various roles of the members in the group communication structure. We argue that such processing, embodied in monitoring tools, can display global properties both at individual level and at group level and efficiently assist the tutor in following the collaboration within the group. It seems to be more appropriate than the long and detailed textual analysis of messages and the statistical distribution of participants' contributions.

1. INTRODUCTION

Let's start from the hypothesis that collaborative learning is effective only within communicative groups. As already emphasized in (Homans, 1951), interactions influence cohesion, cohesion reinforces reciprocal appreciation, which encourages interactions. When considering collaborative learning situations, Earl Woodruff (1999) points out "glue factors" as sources of the cohesion.

Cohesion is an important factor that motivates participants to accomplish the requested task. When it does not exist, the collaborative task may be considered by participants as a painful constraint and even an obstacle to learning. This can arise when a lot of participants are absent, leaving too small a group compared to the size of the collaborative task. In such a case, collaborative learning becomes a waste of time and motivation for the remaining participants. For this reason, the tutor's job is to prevent such a situation. In face-to-face situations, numerous oral and visual cues can help the teacher to reckon the activity and the cohesion of the group: present/absent, talkative/silent, active/passive, etc. However, in Computer Supported Collaborative Distance-learning (CSCDL), it is not easy to detect problems of cohesion within a group. Generally focusing on the support of active learners, the tutor may pay less attention to the quiet ones. An experienced tutor would find and care about the latter ones, by using his/her own check lists but if, for example, communication within the group is divided into various cliques, he/she will not be able to detect it. Existing Learning Management Systems (LMS) display limited participation indices such as the quantity of e-mail or discussion forum messages, but they can't give any information on the *communication structure* of the group.

Social Network Analysis (SNA) (Wasserman & Faust, 1994) characterizes the notion of cohesion more accurately. This domain comes from sociometry, group

T.H.E. Editor(s) (ed.), Book title, 1—6.
© yyyy *Kluwer Academic Publishers. Printed in the Netherlands.*

dynamics, graph theory and their intersection with structural and functional anthropology (Scott 2000). SNA, also named *structural analysis*, aims at studying relationships between individuals, instead of individual attributes or properties. A classical starting point is the construction of a sociometric graph defining relationships between individuals. Then, using connexity properties from graph theory, social network analysts defined various types of subsets (cliques, n-cliques, cliques at level c, k-plex, k-core or k-kernel, etc.) where a given level of "cohesion" exists and depends on proximity, frequency, affinity or other properties. Generally speaking, cohesion is an attractive "force" between individuals. For example, it can measure: a) the number of exchanges between 2 individuals, b) the geodesic distance (or proximity) between 2 individuals, the minimum number of cut-points necessary to disconnect 2 individuals, etc.

As clearly explained in (Fjük & Ludvigsen 2001), CSCL research provides limited insight in large-scale distance-learning issues. Experiments are generally based on activities limited in terms of duration, number of learners, reproducibility, etc. Learning activities are often presented and analysed without considering the context in which they were immersed.

In order to bypass these limitations, we designed an experiment, namely "Simuligne" where 40 learners were involved for 10 weeks in more than 20 activities implemented in lots of different virtual locations (discussion forum, e-mail, chat rooms, drop boxes, publications, etc…) within an existing LMS. We were more interested in the level of activity of a group as a whole than in a fine grain analysis of every participant's message. It is meaningless to measure the cohesion of a group on a micro-activity of very short duration, even if, for this small activity, the cohesion of the group plays an important role in success. Accordingly we had to take into account the whole set of communications over the entire learning session in order to globally reckon cohesion, although its influence may be local.

By using these SNA models, on global communication tracks, we aim at developing a monitoring system (Mbala & al., 2002) according to Jermann & al.'s classification (2001) which distinguishes mirroring-, monitoring- and guiding-systems in CSCL.

In section 2, we present the pedagogical context named Simuligne, from which the data have been extracted. We introduce SNA concepts in section 3 and run cliques and clusters analysis on discussion forum graphs. Both analyses are compared in section 4, before the conclusion.

2. THE SIMULIGNE LEARNING SESSION

Simuligne is a distance French as a foreign language learning session and was born in a trans-disciplinary research project named ICOGAD. In Simuligne, we had 40 learners –English adults in professional training, registered at the Open University–, 10 natives –French teacher trainees from the Université de Franche-Comté–, 4 tutors –teachers of French from the Open University– and one pedagogical coordinator. They can all be classified in one of the three classes of agents of this distance-learning experiment: learners, experts –natives– and teachers –tutors and the

coordinator. All agents were dispatched into four learning groups, namely Aquitania, Lugdunensis, Narbonensis and Gallia. Looking for a collaborative production-oriented project, we decided to adapt the method called "Simulation globale" for the first time to a distance situation. The global simulation method is based on role playing and is often used in intensive face-to-face language learning. Distance was the rule: everybody worked at a distance; no one had ever met before Simuligne, except the natives from Besançon.

Two other important factors of that experiment are sequence and duration. Simuligne spanned over 10 weeks, broken down into 4 parts:

1. (2 weeks): self-introduction to the group and acquisition of technical skills,
2. (3 weeks): designing the place of the simulation, –city, campus map, …
3. (3 weeks): defining the various characters and putting them in various situations to solve some problems –explosion on campus, buses on strike...
4. (2 weeks): discussing and voting for their favoured project.

Three groups out of four achieved the simulation, which is a high ratio in distance-learning. When the Lugdunensis group broke up, its most active learners were transferred to another group. Posters produced by the three remaining groups presented rich and high quality language productions (Chanier, 2001). But what can be assessed afterwards by teachers from interaction and collaboration in the group needs to be more carefully understood if we want to improve learning environments.

The whole Simuligne learning session produced in :

- Discussion forum: 879 015 characters in 2686 messages (which represents 45.11% of the communication flow measured in number of characters);
- E-mail: 834 753 characters in 4062 messages (42.84%);
- Synchronous chat: 234 694 characters in 5680 speech turns (12%).

In Distance-learning (DL), synchronous chat is generally hard to use because of the constraint it imposes. The success of DL is mainly due to the fact that time constraints are relaxed. For this main reason, asynchronous communication tools (like e-mail and discussion forum) are considered as the best communication tools in the adult distance-learning context.

3. FROM MIRRORING TO MONITORING SYSTEM USING SNA MODELS

As mentioned in the introduction, it is not straightforward to extract all the useful information about the messages of asynchronous communication tools such as discussion forum and e-mail, or about speech turns of synchronous communication tools. The reader interested in this data mining issue can find more details in (Reffay & Chanier 2002). We suppose in the current paper that the basic data of messages (communicators, date, time, space, reply_to, size,…) are directly “mirrorable”. We now present how SNA concepts can be used to compute representations that will highlight global information (invisible in raw data): the group communication structure.

3.1. Communication graphs

Participation and communication are different. When evaluating participation of a given agent, you may be interested in the number of messages he/she posted, even if some of them have never been opened by the addressee. When considering communication, you may require that only read messages should be taken into account. Consequently, non-read messages are ignored in our communication graphs.

3.1.1. Definition: e-mail graph

Let $G_e(A,M,P)$ be the directed and valued graph, where A is the set of Agents, M: A×A→N: the relation defining for each couple (a,b) in A×A, the corresponding number of **e-mail** messages **posted by agent a and opened by agent b** during a given period of time P.

↘	Gt	Gl1	Gl2	Gn1	Gl3	Gl4	Gl5	Gn2	Gl6	Gl9	Gl10
Gt	0	28	19	22	17	13	9	13	14	1	4
Gl1	25	0	2	0	0	0	0	2	0	0	0
Gl2	20	4	0	3	0	0	0	1	0	0	0
Gn1	24	1	3	0	0	5	0	3	0	0	0
Gl3	12	0	0	0	0	0	0	0	2	0	0
Gl4	9	0	0	4	0	3	0	8	0	0	0
Gl5	3	0	0	0	0	0	0	0	0	0	0
Gn2	11	2	2	3	0	10	0	0	0	0	0
Gl6	12	0	0	0	4	0	0	0	0	0	0
Gl9	2	0	0	0	0	0	0	0	0	0	0
Gl10	2	0	0	0	0	0	0	0	0	0	0

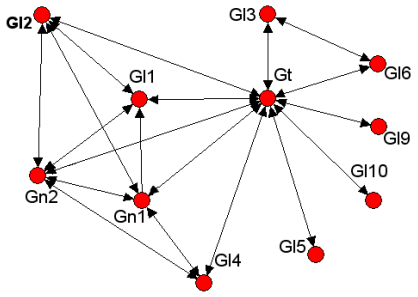


Fig. 1: Matrix and graphical representation of e-mails exchanged within the group named Gallia over the whole training period

Figure 1 shows that each learner (Gl*) exchanges e-mail messages with the tutor (Gt). Some sub-groups (Gt,Gl3,Gl6) and (Gt,Gl1,Gl2,Gn1,Gn2,Gl4) also appear directly on the graphical view. More expressive representations of e-mail graphs, automatically generated by Dot (Graphviz 2000), including the value of lines, comparing the four basic groups are given in (Reffay & Chanier 2002).

3.1.2. Definition: Forum graph

Let $G_f(A,M,P)$ be the directed and valued graph, where A is the set of Agents, M: A×A→N: the relation defining for each couple (a,b) in A×A, the corresponding number of messages **posted by the agent a and opened by agent b** during a given period of time P in the discussion forums.

In fig. 2, the number of communications is so high, that the resulting graph is complete (fully connected). The graphical representation, in this case is not very expressive. A simple text saying that everybody is connected to each other would give the same information. The main purpose of this paper is to show how some

SNA concepts can guide us to provide more expressive graphical representations for forum graphs.

↘	Gt	Gl1	Gl2	Gn1	Gl3	Gl4	Gl5	Gn2	Gl6	Gl9	Gl10
Gt	107	107	99	106	90	107	26	106	107	25	27
Gl1	60	60	52	59	45	60	7	59	60	11	9
Gl2	27	27	27	24	24	27	5	25	27	3	4
Gn1	36	36	33	36	29	36	6	36	36	5	7
Gl3	18	18	18	18	15	18	1	18	18	3	1
Gl4	48	48	42	47	38	48	8	47	48	9	8
Gl5	7	7	6	7	4	7	5	7	7	4	6
Gn2	30	30	27	28	25	30	5	28	30	4	5
Gl6	14	15	14	14	11	15	3	14	15	3	3
Gl9	9	9	7	9	6	9	5	9	9	5	7
Gl10	5	5	3	5	4	5	3	5	5	2	2

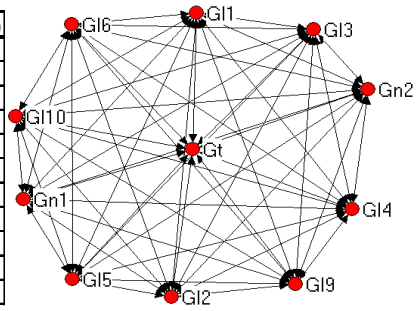


Fig. 2: Matrix and graphical representation of discussion forum messages exchanged within Gallia over the whole training period

3.2. Computing cliques on discussion forum graphs

In SNA literature (Wasserman & Faust, 1994; Scott, 2000) it is clear that cohesion is generally not the goal of a given analysis. The final goal is to find sub-groups of people that are, firstly, closer one to another, and secondly, more connected one to another, or thirdly all connected one to another. The latter case defines the clique. In a graph, a clique is a maximum complete sub-graph. Such a simple definition only represents people who are connected or not connected and ignores the quantity of messages exchanged between agents. Its graphical representation is not informative for discussion forums (cf. fig. 2).

It is more interesting to consider cliques defined by a certain level of communication between all members. In (Wasserman & Faust, 1994, p.278), a "clique at level c" is a sub-graph in which the ties between all pairs of agents have values of c or greater. However, level-c cliques are easier to characterise in undirected graphs than in directed ones.

The trick when starting from a directed graph is to derive the non-symmetrical relation in a symmetrical one. In our graph, the value corresponding to line (a,b) is generally different from the value of the opposite line (b,a). The derivation can be operated by min-, max-, mean- or sum-function on both these values. The min-function was a good candidate but would eliminate intensive unidirectional flow of messages sent and read. Then, we chose the "sum" operator to derive our undirected graph. Values of (a,a) ties are reduced to zero. For each pair (a,b) (where $a \neq b$), the tie (a,b) of the resulting non-directed graph will represent the number of messages communicated (sent and read) between a and b in either direction. So, the resulting matrix (cf. fig. 3b) is symmetric. In order to achieve the computation of level-c cliques in our graph, we just have to fix threshold c. In our experiment, given the activities in Simuligne, we consider that an agent is in a cohesive sub-group if he/she has exchanged at least 10 messages with each of the other members of that

sub-group. With c=10, on the forum graph of the Gallia group, UCINETv6 will give the following list of level-10 cliques:

(Gt Gl1 Gn1 Gn2 Gl4 Gl2 Gl6 **Gl3**)

(Gt Gl1 Gn1 Gn2 Gl4 Gl2 Gl6 **Gl5**)

(Gt Gl1 Gn1 Gn2 Gl4 Gl2 Gl6 **Gl9**)

(Gt Gl1 Gn1 Gn2 Gl4 **Gl10**)

Fig. 3a: List of cliques of level 10 in the forum graph of Gallia during the whole Simuligne period

	Gt	Gl1	Gl2	Gn1	Gl3	Gl4	Gl5	Gn2	Gl6	Gl9	Gl10
Gt	0	167	126	142	108	155	33	136	121	34	32
Gl1	167	0	79	95	63	108	14	89	75	20	14
Gl2	126	79	0	57	42	69	11	52	41	10	7
Gn1	142	95	57	0	47	83	13	64	50	14	12
Gl3	108	63	42	47	0	56	5	43	29	9	5
Gl4	155	108	69	83	56	0	15	77	63	18	13
Gl5	33	14	11	13	5	15	0	12	10	9	9
Gn2	136	89	52	64	43	77	12	0	44	13	10
Gl6	121	75	41	50	29	63	10	44	0	12	8
Gl9	34	20	10	14	9	18	9	13	12	0	9
Gl10	32	14	7	12	5	13	9	10	8	9	0

Fig. 3b: Symmetric matrix of forum graph of Gallia during the whole Simuligne period

Figure 3b gives the whole data on forum exchanges, but it is hard to detect who is in the central core of the discussion group from it. The list of level-10 cliques given by UCINETv6 and presented in figure 3a is much more informative. We can see that Gt, Gl1, Gn1, Gn2 and Gl4 belong to all the cliques. They can then be considered as the most central agents. Gl2 and Gl6 belong to 3 of the 4 cliques, they are nearly as central as the previous ones. But Gl3, Gl5, Gl9 or Gl10, belonging to only one clique, could be considered a little bit more peripheral. Let us now compare Gallia with the other groups:

- Aquitania : 5 cliques: (Al1) (Al3) (Al4) (At,An2,Al2,Al5,An3,Al6,Al8,Al10) (At,Al2,Al5,An3,**Ll4**,Al6,**Ll9**,Al10);
- Narbonensis : 6 cliques: (Nl7) (Nl9) (Nn1,Nl10) (Nn1,Nl3,Nl4,Nt,Nl8,Nl1) (Nn1,Nl3,Nl4,Nt,Nl8,Nn2) (Nn1,Nl3,Nl4,Nt,Nl8,Nl5);
- Lugdunensis : 9 cliques: (Ll2) (Lt,Ll5) (Lt,Ll6) (Lt,Ll7) (Lt,Ll8) (Lt,Ll10) (Lt,Ll1) (Lt,Ln2,Ln3) (Lt,Ln1,Ln2,**Ll4**,**Ll9**).

Note that each member of Gallia belongs to at least one clique and there are at least 6 members in each of the 4 cliques. This must be considered as excellent cohesion. The Lugdunensis group is in a very different situation. It counts many small cliques centralised around the tutor "Lt". It means that there are too few exchanges between learners (except for Ll4 and Ll9). Lugdunensis never reached a sufficient level of cohesion to achieve the Simuligne collaborative learning programme. This group stopped and its most active learners Ll4 and Ll9 were moved to the Aquitania group. Note that Ll4 and Ll9 are also emphasized in the representation of level-10 cliques for Aquitania.

The resulting lists of (level-10 cliques) obtained for each group can be used to reckon the global group cohesion (few large cliques contain a lot of agents) and, more precisely to detect which agent is central (belongs to most of the large cliques)

or which is isolated (belongs only to few and small cliques). In other words, the computation of level-c cliques is a valid tool to answer the following question :

> For a given intensity of communication, what is the structure of the group?

When we make c vary, the corresponding communication structure is modified. Such analysis demonstrates how sensitive the value of the selected threshold c is. This value should be meaningful to participants of the learning session and even be made explicit in the pedagogical contract when learners register. In a collaborative learning session, the tutor/teacher necessarily has a precise idea of the value of c.
An agent may want to get a response to the dual question :

> For a given structure of the group, what is the intensity of communication?

This is what clusters highlight as we will see in the following section.

3.3. Computing clusters on discussion forum graphs

In SNA, rationales for using one type of cluster in a given situation are unclear and presented in confusing ways. However the way to compute clusters is very simple. Hierarchical clusters representation is the result of the following algorithm:

> *Given a symmetric n-by-n [matrix] representing similarities among a set of n items, the algorithm finds a series of nested partitions of the items. The different partitions are ordered according to decreasing levels of similarity. The algorithm begins with the identity partition (in which all items are in different clusters). It then joins the pair of items most similar, which are then considered a single entity. The algorithm continues in this manner until all items have been joined into a single cluster (the complete partition).* (Johnson, 1967) cited in (Borgatti & al. 2002)

At each step of this algorithm, the goal is to select the most similar pair of clusters in order to join them in a single cluster. The words proximity and similarity have similar meanings. But we will use “proximity” when referring to clusters and “similarity” when referring to members. The proximity of two clusters can be defined in different ways based on the similarities between the members of one cluster compared with the members of another cluster. Among them, UCINETv6 proposes:

1. **Single link**: Also known as the "minimum" or "connectedness" method. Proximity between two clusters is defined as the largest similarity between members.
2. **Complete link**: Also known as the "maximum" or "diameter" method. Proximity between two clusters is defined as the smallest similarity between members.
3. **Average**: Proximity between clusters defined as average similarity between members.

Let us choose the complete link definition. Then, at step i, we join the pair of clusters whose proximity value is maximum. Let k_i be this proximity: k_i will define the diameter of the new joined cluster. We are then sure that all the members of the new cluster have exchanged at least k_i communications with each of the other

members of this cluster. Note that a cluster whose diameter is k_i is not exactly a level-k_i clique of but would necessarily be entirely contained in such a clique.

The result of this algorithm is the list of clusters (of diameter k_i) joined at each step i. Figure 4 presents the result of this analysis on the non-directed graphs derived from forum graphs of the four basic groups of Simuligne during the whole period.

On each diagram, you can read for a given line: the diameter of the cluster, and contiguous 'X' represent the cluster containing the agents of the corresponding columns. For example, we can see in the Gallia group that Gl1 and Gt exchanged 167 messages, and each of them exchanged at least 108 messages with Gl4, and each of the last three agents exchanged at least 83 messages with the native Gn1, etc…

These diagrams provide interesting information. First, if you globally compare only the maxima (representing the maximum quantity of messages exchanged between the most communicative agents of the group) reached by each group, you immediately find that Lugdunensis reaches a very low level of 40 messages and Aquitania a very high level of 314! You can also see immediately who the most communicative agents for each group are. It is not surprising to find here that, for all groups, the tutor (At, Lt, Gt, Nt) is always in the most communicative pair of agents.

AQUITANIA

```
                                A
        A A L L A A A A A   A A l
        l l l l l l n n l A l l 1
Level   3 4 4 9 1 8 2 3 2 t 5 6 0
------  - - - - - - - - - - - - -
  314   . . . . . . . . . XXX . .
  191   . . . . . . . . . XXXXX .
  126   . . . . . . . . . XXXXXXX
   64   . . . . . . . . XXXXXXXXX
   33   . . . . . . . XXXXXXXXXXX
   23   . . XXX . . . XXXXXXXXXXX
   13   . . XXX . . XXXXXXXXXXXXX
   12   . . XXX . XXXXXXXXXXXXXXX
    3   . . XXX XXXXXXXXXXXXXXXXX
    2   XXX XXX XXXXXXXXXXXXXXXXX
    0   XXXXXXXXXXXXXXXXXXXXXXXXX
```

NARBONENSIS

```
                                N
        N N N N N N   N N N N l
        l l n n l l N l l l l 1
Level   7 4 2 1 1 3 t 5 8 9 0
-----   - - - - - - - - - - -
  163   . . . . . XXX . . . .
   71   . . . . XXXXX . . . .
   54   . . . . XXXXXXX . . .
   30   . . . . XXXXXXXXX . .
   16   . . . XXXXXXXXXX . .
    8   . . XXXXXXXXXXXX . .
    5   . XXXXXXXXXXXXXX . .
    3   XXXXXXXXXXXXXXXX . .
    1   XXXXXXXXXXXXXXXXXX .
    0   XXXXXXXXXXXXXXXXXXX
```

LUGDUNENSIS

```
                                L
        L L L L L L L L   L L L l
        l l l l l n n n L l l l 1
Level   2 5 7 1 6 3 1 2 t 4 9 8 0
-----   - - - - - - - - - - - - -
   40   . . . . . . . XXX . . . .
   11   . . . . . . XXXXX . . . .
   10   . . . . . . XXXXXXXX . .
    9   . . . . . XXXXXXXXXX . .
    7   . . . . XXXXXXXXXXXX . .
    6   . . . . XXXXXXXXXXXX XXX
    5   . . . XXXXXXXXXXXXXXXXX
    4   . . XXXXXXXXXXXXXXXXXX
    2   XXX XXXXXXXXXXXXXXXXXX
    1   XXXXXXXXXXXXXXXXXXXXXX
```

GALLIA

```
                                G
        G G G G   G G G G G G l
        l l n l G l n l l l l 1
Level   3 2 1 1 t 4 2 6 5 9 0
-----   - - - - - - - - - - -
  167   . . . XXX . . . . . .
  108   . . . XXXXX . . . . .
   83   . . XXXXXXX . . . . .
   64   . . XXXXXXXXX . . . .
   52   . XXXXXXXXXX . . . .
   42   XXXXXXXXXXXX . . . .
   29   XXXXXXXXXXXXXX . . .
    9   XXXXXXXXXXXXX XXXXX
    5   XXXXXXXXXXXXXXXXXXX
```

Fig. 4: Hierarchical clusters on forum graphs of Simuligne during the whole period.

Then, inside each group, you can identify, for each agent, level k at which he/she enters the cluster. The decrease of k for each group is also a good indicator. For example, the maximum k value of Narbonensis and Gallia are around 165, but the decrease of Narbonensis is much faster than Gallia's: for 7 members in the cluster, the value of k for Gallia is 42 when for Narbonensis it falls to 8 messages.

In order to build a comparison with level-10 cliques, we materialized the threshold of 10 chosen in the previous section by a horizontal dotted line. However, we want to emphasize that hierarchical clusters analysis does not need any parameter other than the symmetrical matrix of the graph of the discussion forum.

4 CHOOSING BETWEEN CLIQUE AND CLUSTER?

In the previous sections, we presented two different analyses: level-10 cliques and hierarchical clusters. The validity of the results produced by each will be compared in this section.

Let us first recall that these concepts and tools are to be used by tutors or teachers. We assume they can easily define the value of threshold c. But, the main drawback of level-c cliques is the fact that, once threshold c has been fixed, there are only two categories of relationships: those that are eliminated because their value is less than c and the others that are kept because their value is greater than or equal to c. This is the reason why the choice of threshold c is sensitive.

The main difference between clusters and cliques is that cliques are maximal sub-sets at a given level k. In hierarchical cluster construction, 2 items joined in a cluster at an earlier stage cannot belong to different clusters at a later stage. This is also the reason why they are named hierarchical clusters, i.e. they are nested subsets. The result of hierarchical clustering gives a lot of quantitative indices when level-c cliques do not give any. The velocity of the decrease of index k in each group in cluster analysis is also a meaningful information.

Conversely, a cluster does not define a set of individuals as clearly as a clique of level c does. For example, an individual belonging to a cluster may exchange many more messages with an external individual than he does with any of the other members of this cluster. In cliques analysis, these two individuals would necessarily belong to (at least) one common level-c clique if they shared more than c-1 messages.

We can see that these two methods gives complementary information. And the best way to use them is probably to apply clusters analysis first in order to select threshold c among the various k values, and then compute the analysis of level-c cliques. Thanks to the definition of level-c cliques, the cohesion of the group is more accurately reckoned from the overlapping of cliques than from the arrangements of clusters.

5. CONCLUSION

It has been argued that cohesion is crucial in collaborative learning, but many interpretations of cohesion are possible. CSCL research has not yet given any

precise definition of this notion, and existing LMS does not give any representation of the cohesion of a group. We showed that SNA literature gives many precise and computable definitions for cohesion. We chose "level-c cliques" and "hierarchical clusters" analysis to confront them with data collected from a large-scale distance-learning situation. We based our computations on discussion messages and showed that cliques and clusters give complementary information. On the one hand, cliques highlight the communication structure and the position of the agents for a given intensity of communication. On the other hand, clusters emphasize the various intensity levels. By applying cluster analysis first, it is possible to choose the appropriate intensity c and then run an analysis of level-c cliques that gives the communication structure. We will now work at integrating these analyses in a monitoring system to be added to a distance-learning platform.

LIFC : Laboratoire d'Informatique de l'Université de Franche-Comté, France
http://lifc.univ-fcomte.fr

6. ACKNOWLEDGEMENT

Special thanks to the French Ministry of Research (MRT) and its cognitive science program (Programme Cognitique 2000) which supports the ICOGAD project.

7. REFERENCES

Borgatti S.P., Everett M.G., Freeman L.C. (2002): UCINET 6 for Windows. Harvard: Analytic Technologies. http://www.analytictech.com/

Chanier T. (2001): « Créer des communautés d'apprentissage à distance ». *Les dossiers de l'Ingénierie Educative*, n°36, Centre National de Documentation Pédagogique(CNDP), pp 56-59, France.

Degenne A., Forse M. (1994): *Les réseaux sociaux*, Collection U, A. Colin, Paris, France.

Fjük A., Ludvigsen S. (2001): « The Complexity of Distributed Collaborative Learning: Unit of Analysis ». *European Conference on Computer Supported Collaborative Learning*, Maastricht.

GraphViz (2000). Home page. Last visited December 2001 at , http://www.graphviz.org

Jermann P., Soller A., Muehlenbrock M. (2001): « From mirroring to guiding: a review of state of art technology for supporting collaborative learning ». *Proceedings of the European Computer Supported Collaborative Learning Conference*. (EU-CSCL'01), Maastricht.

Johnson S.C. (1967): "Hierarchical clustering schemes". *Psychometrika*, 32, p. 241-253

Mbala A., Reffay C., Chanier T. (2002): « Integration of automatic tools for displaying interaction data in computer environments for distance-learning », *Intelligent Tutoring System Conference*, in Cerri, S.A. , Guardères, G. & Paraguaçu, F.(dirs), Biarritz, France, p. 841-850.

Reffay C., Chanier T. (2002): « Social Network Analysis Used for Modelling Collaboration in Distance-learning Groups », *Intelligent Tutoring System Conference*, in Cerri, S.A., Guardères, G. & Paraguaçu, F.(dirs), Biarritz, France, p. 31-40.

Scott J. (2000): *Social Network Analysis : a handbook*, 2nd ed., SAGE, London.

Wassermann S., Faust K. (1994): *Social Network Analysis : Methods and Applications*, Cambridge University Press, New York.

Woodruff E. (1999): « Concerning the Cohesive Nature of CSCL Communities », *Computer Supported Collaborative Learning Conference*, Palo Alto, CA: Stanford University, p. 677-680.

Wortham D.W. (1999): « Nodal and matrix analyses of communication patterns in small groups ». Proceedings CSCL'1999 Conference, Palo Alto, CA: Stanford University, p. 681-686.