Managing CSCL Activity through Networking Models

L. Casillas[1], T. Daradoumis[2,3] and S. Caballé[2]
[1] University of Guadalajara, Guadalajara, Mexico
[2] Open University of Catalonia, Barcelona, Spain
[3] University of the Aegean, Mytilene, Greece

Abstract—**This study aims at managing activity carried out in Computer-Supported Collaborative Learning (CSCL) environments. We apply an approach that gathers and manages the knowledge underlying huge data structures, resulting from collaborative interaction among participants and stored as activity logs. Our method comprises a variety of important issues and aspects, such as: deep understanding of collaboration among participants in workgroups, definition of an ontology for providing meaning to isolated data manifestations, discovering of knowledge structures built in huge amounts of data stored in log files, and development of high-semantic indicators to describe diverse primitive collaborative acts, and binding these indicators to formal descriptions defined in the collaboration ontology; besides our method includes gathering collaboration indicators from web forums using natural language processing (NLP) techniques.**

Index Terms—**Cooperative/collaborative learning, evaluation methodologies, interactive learning environments, social networks analysis.**

I. INTRODUCTION

One may initially consider that the complexity of performing a collaboration analysis on human interaction through a computational solution could be apparently low due to the restrictions of the communication channels available, which reduces the amount of information elements produced. Indeed, the restrictions in transferring information could falsely imply that a rather simple solution could suffice; however, it is just the absence of certain information elements which is the cause of complexity, due to the need to fill some important gaps. These gaps refer to certain messaging elements that complement the regular communication pieces that constitute the whole human communication system, such as non-verbal messaging. Some of these non-verbal communication elements could help understand several collaborative issues; therefore special efforts must be made to discover some of the participants' interests and intentions lying beneath their collaboration acts, both physical (tacit) and verbal.

The analysis of collaborative activity in virtual environments has become an important task when supervising and monitoring the performance of learning groups. Regular face-to-face collaborative interaction includes a diversity of communication elements beyond the spoken language. This complex message exchange would imply, hopefully, a synchronization of minds, which supports a shared network of concepts [13]. Collaboration in virtual environments does not provide this rich experience of having a complete situation-awareness. Limited communication channels restrict the full transfer of the messages generated by each member of the workgroup [10].

The main purpose for our approach is to build a comprehensive and robust collaboration analysis model for analyzing, assessing, monitoring and personalizing collaborative learning in web-based learning and working environments in a practical and effective manner. The construction of a flexible framework to support the analysis and evaluation of collaborative interactions is a complex process. For instance [14], in his qualitative interaction analysis model, proposes a three-level approach: dialogue, knowledge and action analysis levels. For dealing with this complexity, our approach distinguishes different levels and dimensions of analysis. First, it takes into account that collaboration analysis is oriented to discover the intentions and effects of every primitive collaborative action [16]. Then, both group members and the evaluator of the collaborative learning process should be able to discern that mutual understanding exists among peers, since this is the first step to solve any problem that may appear when different people are involved in the search of a solution. Every member of the group might have a different approach of the problem and its possible solution. Though collaboration is not a formal and ordered process, the initial stages must be clearly oriented to model a common structure of concepts regarding the problem's understanding and its solution [15, 10].

Our work takes an existing analysis model as a point of departure [3] and extends it by designing and implementing a new model which is based on a refined analysis of the knowledge structures involved and the incorporation of a new unification level that binds the different indicators involved in two collaboration analysis dimensions: analysis of tacit collaborative actions (like create, read, modify objects) and analysis of verbal actions generated in conversations in web forums. Doing so, we do not only achieve that the diverse results that emerge from these dimensions are combined and correlated in a meaningful way, but we also obtain a holistic framework for collaboration analysis through innovative knowledge representations, based on a networked approach.

II. RELATED WORK

The analysis of the collaborative activity in CSCL environments implies a significant effort as concerns the complexity of gathering knowledge from the raw data stored by these platforms. From the turn of the century, several research works have tried to provide a solution toward this

direction. In particular, the work performed by [9] has been oriented to carry out an analysis from data coming from a networked collaborative environment to study the interaction which is organized in a temporal structure. Doing so, it was possible to compare current behaviors by matching them with expected behaviors.

According to [5], we also need to capture the interdependencies of partners in discourse. In this sense, that work claims that group interactions capture the dynamic interplay in meaning-making over time in discourse between participants, what they understand, the material and symbolic resources they use, the types of contributions that they make, and how they are taken up or not in a given discourse. As a consequence, we need to identify aspects of collaborative activity that can be *operationalized*, measured, and analyzed from well-grounded indicators and the appropriate correlations among them.

The study of collaborative learning is a multi-method, multi-disciplinary affair, a complex interplay of the individual, the group, and the context in which learning occurs [10]. Recent research in collaborative learning is assembled in the volume [6] and describes the current efforts in controlled experiments, ethnographic portraits, surveys, and qualitative or quantitative analysis of talk and interaction. It is also observed that hybrid or mixed methods are increasingly used to integrate studies of interactional processes and learning outcomes in collaborative learning, ranging from content analysis of discourse, to social network analysis to multilevel modeling [1, 7]. However, a complete and robust integrated collaboration analysis approach needs to take into account not only the different dimensions in which collaboration occurs, the various levels (stages) in which the analysis is defined but also to interrelate them in a coherent manner so that to obtain a consistent collaboration analysis model capable of providing an effective understanding and measuring of the four variables identified in [7]: individual learning, group processes, interaction sequence and the context (dimensions) in which learning occurs.

To this end, our work focuses on determining classical key issues and problems and sets the basis for providing a complete solution (collaboration analysis model) using an integrated (mixed) approach.

III. COLLABORATION ANALYSIS FROM LARGE DATA SETS

The core task of our work is to develop a sound and efficient model and approach for the analysis of the collaborative interactions of participants in a CSCL environment, involving different analysis dimensions, levels and indicators. CSCL research has both analytic and design components. As [11] state, "analysis of meaning making is inductive and indifferent to reform goals. It seeks only to discover what people are doing in moment-to-moment interaction, without prescription or assessment. Design, on the other hand, is inherently prescriptive—any effort toward reform begins from the presumption that there are better and worse ways of doing things. To design for improved meaning making, however, requires some means of rigorously studying praxis. In this way, the relationship between analysis and design is a symbiotic one—design must be informed by analysis, but analysis also depends on design in its orientation to the analytic object".

To achieve a well-grounded and consistent collaboration analysis model which takes into consideration differ-

ent knowledge structures at various levels, a quantitative analysis is performed through an automated mechanism which is fed with raw data to produce a social network. Through its structural details, this network models the semantics of the collaborative activity, by collecting knowledge within its nodes and its edges, as well as by clustering participants around knowledge objects. In fact, since students may belong to more than one teams and thus have access to a variety of shared knowledge objects, the social network developed for carrying out the quantitative analysis level is not limited to isolated learning teams. Instead, we faced the challenge to build and manage a huge social network that functions as a *Knowledge Server* and provider.

Problem complexity is broken down to gather the main system elements. Data are originally composed by means of primitive events from the CSCL tool. Through mining techniques, these data are translated into primitive interaction acts. Finally interaction acts are processed to produce the collaboration network which models the whole collaborative activity *a posteriori*. Reaching at this point, it is possible to calculate different ontology indicators.

Using modern programming paradigms (such as object oriented programming, agents programming and distributed programming), we developed a system which builds and manages the social network that supports our *Knowledge Server*. A further important feature and contribution of our approach is the incorporation of two agents to support a search module. Due to the multiple redundancies which are present in the primitive interaction-acts file, which is a natural consequence of the different accesses to objects along the time, it is necessary to handle the multiple references to the same user, group or knowledge-object (thus avoiding, throughout this approach, multiple insertions of such elements to the social network). To do so, we employ two agents that take charge of managing the specific names recovery from alternative files; these agents can ran in different networked computers, decreasing the demand of computer resources. This system is outlined in Figure 1. The collaboration network has to be constructed only the first time the data are loaded from the log files. These logs include a collection of diverse activity indicators collected along the collaboration act among participants, workgroups and knowledge objects sharing common spaces. The analyzed activity comes from real performance of participants in different subjects for semesters Spring-2003 and Fall-2004. Participants have produced, unconsciously, data in log files; by using the BSCW environment for their tasks. Indicators are fulfilled with different lines and stripes of data from these logs. Such indicators accumulate primitive events performed by environment users. Log files are built automatically by CSCL for diverse supervision goals, most of these goals oriented to the correct operation. Nevertheless, we use these logs to supervise collaboration. This is possible because of the type of collaborative analysis we are performing; the analysis is carried out a posteriori, that is, after a collaborative work and learning phase is finished.

Regarding the need for providing high level meaning to the results extracted from this model, a specific ontology had to be defined. Our ontology constitutes a natural extension of an ongoing effort to provide a rich representation scheme that supports collaboration analysis, which started from the proposal of [8] and continued through the work suggested by [2]. As shown in Figure 2, a hierar-
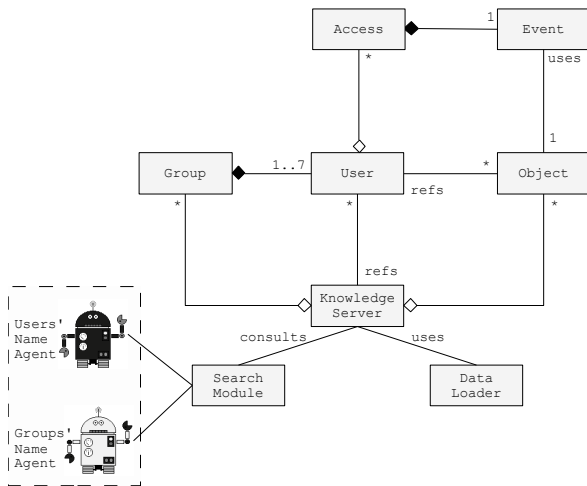
Figure 1. Diagram of classes that model the basic structure of the social network
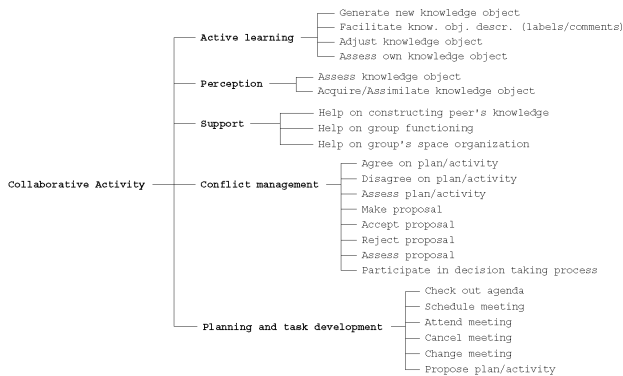

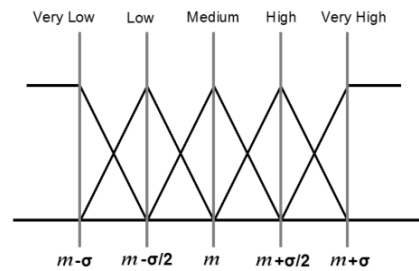Figure 2. Proposed ontology designed to support the collaboration analysis

Figure 3. Fuzzy model for automatically assign meaning to quantitative results from the Knowledge Server

TABLE I.
DIFFERENT DETAILS OF GLOBAL ACTIVITY DISCOVERED USING THE KNOWLEDGE SERVER

Global Activity Distribution	
Total # of Events	77157
Avg Events/User	221.716
Global % of Events/Category	
Event Category	*%*
BranchEvent	0.005
ChangeDateEvent	0.01
ChangeDescEvent	0.818
ChangeEvent	0.086
CheckinEvent	0.021
CheckoutEvent	0.022
ChvinfoEvent	0.009
CopyEvent	0
CreateEvent	18.079
CutEvent	0.999
DeleteEvent	1.581
DropEvent	0.708
EditEvent	0.614
ForgetEvent	0.001
LinkEvent	0.118
RateEvent	0.417
ReadEvent	74.001
RenameEvent	0.903
ReplaceEvent	1.196
ReviseEvent	0.248
UndeleteEvent	0.03
VersionEvent	0.133

Total of Objects Created	13949
% of Creative Activity	18.079
Avg Objects Created/User	40.083
Global % of Object/Type	
Object Type	*%*
Appointment	0.652
Document	46.756
Folder	3.843
Note	47.394
Noteboard	0
Notes	1.341
WSCalendar	0.014

Total of Objects Accessed	63208
% of Perceptive Activity	81.921
Avg Objects Accessed/User	181.632
Global % of Object/Type	
Object Type	*%*
Appointment	0.196
Document	38.963
Folder	1.321
Note	59.005
Noteboard	0
Notes	0.511
WSCalendar	0.003

chical model is built from the general collaborative activity entity which branches to five principal activity indicators: active learning, perception, support, planning and task development, and conflict management. Each branch is subsequently divided into specific aspects of collaboration.

Primitive events, as those shown in Table I, are bound to leaves in the ontology tree of figure 2. For instance, the primitive CreateEvent provides body to leaf "Generate new knowledge object". As well as primitive DeleteEvent influences the leaf "Help on group's space organization". There is a complete designation binding primitive events to leaves and branches of the ontology, but they are not fully included here due to space restrictions. Nevertheless, the already provided examples suggest the politics followed up to define such designations.

In order to provide an adaptive response to different sets of data studied, we have further extended and enriched the quantitative analysis, by developing a fuzzy-logic machinery oriented to classify the quantitative results produced by the *Knowledge Server*, thus enabling an automatic interpretation for such numerical results. This fuzzy-logic model has been designed to perform the following tasks: for every activity item the average (m) is calculated, which is settled as the centre of the fuzzy-sets structure, as well as the standard deviation (σ), which is used as the measurement to establish the fuzzy sets around the middle fuzzy set, as shown in figure 3.

Once the network is loaded, the *Knowledge Server* is able to respond to queries. Table I shows the profile of the global activity discovered throughout the awareness information collected by the *Knowledge Server*, using the data collected from the BSCW platform during the realization of a real collaborative learning practice that was carried out in our university during a whole semester in two undergraduate courses that were offered in parallel.

By reviewing the specific indicators that make up the global activity, shown in Table I, we noted the existence of significant trends which cannot be disregarded. In particular, the global activity follows a standard distribution; the most common action performed by users is ReadEvent with a frequency of 74%, followed by CreateEvent with 18%.

In addition, object creation spans to almost a 50% for both Create Document and Create Note actions. Finally, access to objects represents almost a 60-40% relation of Note-Document access, being Note objects the most accessed ones (60%). The reader must consider that a Note is a contribution to the Web forums provided by the tool, which means that group members were interested in accessing forums to see their peers' contributions. The rest of indicators have a low significance. As a result, these data offer a clear picture both of the way group activity
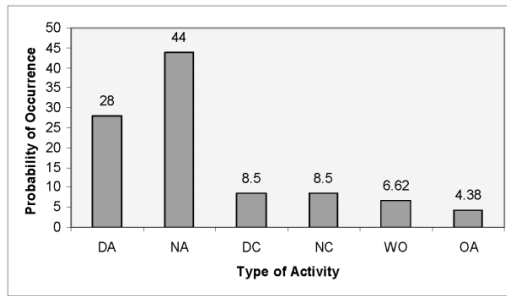
Figure 4. Distribution of the users' activity that shows the participation-behaviour during collaboration. Where DA: Doc Access, NA: Note Access, DC: Doc Creation, NC: Note Creation, WO: Workspace Organization, OA: Other Activities.

tends to be distributed and the way the CSCL tool is used. Figure 4 recomposes the results of Table I and shows the tendencies of users' activity distribution in a more graphical and concise way.

Going a step further, we proceeded to compare the activity information of an individual user shown in Table II with the general pattern of collaborative activity identified in Table I. In this example, the supervisor of the collaborative activity can discern that the user has created a number of documents (68.3%) which is far above the average of the global document creation (46.7%). Thus, the distribution of document and note creation of this user does not follow the balanced 50-50% global tendency of the general pattern; instead, the user gives more importance to document rather than note creation. The evaluator will then have to figure out the quality of user's documents as well the user's true intentions underlying his/her notes through a qualitative assessment carried out by means of direct reading over documents. In this sense, the conversation analysis approach, provided by the second dimension of our model, supports the evaluator's decision making process more efficiently. Moreover, the user's document and note access activity does not follow the general 40-60% pattern; in fact, it is more balanced, being closer to 50-50%.

In general, in an effort to connect and interpret these results according to the proposed ontology (in Figure 2), we can say that the user's creative activity (24.4%), which forms part of his/her active learning behavior, is above the average global one (18%), whereas the user's perceptive attitude seems more balanced, though he/she shows a somewhat lower reading activity (69.6%) as opposed to the 74% of the general pattern. Making this comparison, the evaluator is able to analyze specific details of every user's activity as well as to measure the user's performance in terms of the general expected behavior of the whole class.

As regards the analysis of learning group performance, the *Knowledge Server* is able to extract precise details for the activity of a specific group and build a detailed report for the group's social network. Table III shows the report generated for a given group. The information available for a learning group goes beyond the one available for a single user. In particular, the Knowledge Server provides detailed information about the way activity is distributed, including information about the members' creative and perceptive behavior. Besides, there are plenty of details regarding the interaction volume produced by every pair of members when they access common objects; for instance, the interaction "User-2695683 => User-2697485:

183" means that "User-2695683" accessed the objects created by "User-2697485" 183 times. In addition, we also count the number of accesses that every member produces outside their group. As in the case of individual performance analysis, group performance analysis can be also compared to the performance shown by the total number of collaborative groups.

TABLE II.
DETAILED INFORMATION ACTIVITY FOR A SPECIFIC USER, EXTRACTED FROM THE KNOWLEDGE SERVER

Activity review for participant: *User-2761880*	* Total of Events (168) Frequency of Events Performed:
* Created Objects (41) **Frequency of Created Objects Types:** Appointment (1): 2.44% Document (28): 68.29% Folder (0): 0% Note (10): 24.39% Noteboard (0): 0% Notes (2): 4.88% WSCalendar (0): 0%	BranchEvent (0): 0% ChangeDateEvent (0): 0% ChangeDescEvent (0): 0% ChangeEvent (0): 0% CheckinEvent (0): 0% CheckoutEvent (0): 0% ChvinfoEvent (0): 0% CopyEvent (0): 0% CreateEvent (41): 24.4% CutEvent (0): 0% DeleteEvent (9): 5.36% DropEvent (0): 0%
* Total # of Accesses (127) **Average Time spent in object access (2.52 hours)** **Frequency of Object Types Accessed:** Appointment (1): 0.79% Document (65): 51.18% Folder (0): 0% Note (59): 46.46% Noteboard (0): 0% Notes (2): 1.57% WSCalendar (0): 0%	EditEvent (0): 0% ForgetEvent (0): 0% LinkEvent (1): 0.6% RateEvent (0): 0% ReadEvent (117): 69.64% RenameEvent (0): 0% ReplaceEvent (0): 0% ReviseEvent (0): 0% UndeleteEvent (0): 0% VersionEvent (0): 0%

TABLE III.
REPORT FOR A SPECIFIC GROUP PRODUCED BY THE KNOWLEDGE SERVER

G: Group-2757651	*** Activity Distribution ***
Members: User-2695683 User-2697485 User-2697789 User-2749058	*** Creative Activity per Member *** User-2695683(197): 36.08% User-2697485(192): 35.16% User-2697789(98): 17.95% User-2749058(59): 10.81%
* Objects Created by Group (546) **Average # of Objects per Member (136.5)** **Frequency of Object Types Created:** Appointment (8): 1.47% Document (252): 46.15% Folder (120): 21.98% Note (680): 124.54% Noteboard (0): 0% Notes (32): 5.86% WSCalendar (0): 0%	*** Perceptive Activity per Member *** User-2695683(686): 31.98% User-2697485(658): 30.68% User-2697789(367): 17.11% User-2749058(434): 20.23% *** INTERACTION ACTIVITY *** *** Internal Interaction *** *mbr => mbr : # intercts. through objs* * . * @ User-2695683 => User-2695683: 341 User-2695683 => User-2697485: 183 User-2695683 => User-2697789: 66 User-2695683 => User-2749058: 40
* Total # of Events of Group (5382) **Average # of Events per Member (1345.5)** **Frequency of Events Performed:** BranchEvent (0): 0% ChangeDateEvent (0): 0% ChangeDescEvent (206): 3.83% ChangeEvent (18): 0.33% CheckinEvent (0): 0% CheckoutEvent (0): 0% ChvinfoEvent (0): 0% CopyEvent (0): 0% CreateEvent (1092): 20.29% CutEvent (76): 1.41% DeleteEvent (122): 2.27% DropEvent (62): 1.15% EditEvent (70): 1.3% ForgetEvent (0): 0% LinkEvent (8): 0.15% RateEvent (2): 0.04% ReadEvent (3350): 62.24% RenameEvent (60): 1.11% ReplaceEvent (312): 5.8% ReviseEvent (0): 0% UndeleteEvent (4): 0.07% VersionEvent (0): 0%	* . * User-2697485 => User-2695683: 142 @ User-2697485 => User-2697485: 430 User-2697485 => User-2697789: 43 User-2697485 => User-2749058: 29 * . * User-2697789 => User-2695683: 95 User-2697789 => User-2697485: 91 @ User-2697789 => User-2697789: 150 User-2697789 => User-2749058: 24 * . * User-2749058 => User-2695683: 130 User-2749058 => User-2697485: 133 User-2749058 => User-2697789: 60 @ User-2749058 => User-2749058: 85 * . * *** External Interaction *** User-2695683: 56 accesses outside User-2697485: 14 accesses outside User-2697789: 7 accesses outside User-2749058: 26 accesses outside
* Total # of Accesses of Group (2145) **Average # of Accesses per Member (536.25)** **Average Time spent in object access (2.5 hours)** **Frequency of Object Types Acceded:** Appointment (8): 0.37% Document (1664): 77.58% Folder (262): 12.21% Note (2280): 10.62% Noteboard (0): 0% Notes (76): 3.54% WSCalendar (0): 0%	

Finally, our system enhances the social network formed by a group with information that includes connection weights for the creative and perceptive collaborative categories. This is done by relating the Internal Interaction values with the Activity Distribution ones, as shown in Table III. For instance, the "User-2695683" created 197 knowledge objects which were accessed 367 (142+95+130) times by the other group partners (averaging 122.3 times); this means that 62.08% of his creation activity was accessed by partners.

As regards the utility of the added fuzzy-logic machinery level, that involves fuzzy sets (shown in Figure 3), it allows us to set the centre of the model at the average value of a specific activity item. For instance, let's consider the average value 18.079 of the activity item "CreateEvent" from the global activity shown in Table I. The standard deviation, which measures the dispersion for a set of values, represents the rate of vibration for the performance shown by participants in a specific activity item. We used the notion of standard deviation for a discrete random variable or data set. Figure 5 shows the fuzzy model adapted to the activity item "*CreateEvent*". In particular, for the current activity item (*CreateEvent*) the global average (μ) of objects created by a certain participant is 40.083 and the standard deviation (σ) is 34.404, thus $\sigma/2$ is 17.202. Triangle shaped fuzzy-sets then populate the inner region of the model, whereas trapezoid sets are settled in the extremes of the model. The limits of the fuzzy-sets are calculated using, again, the standard deviation; the goal is to use σ as the width for the base of sets.

As such, once the fuzzy sets are established, it is possible to measure specific performance indicators for the activity item considered. The triangle shaped fuzzy sets will respond according to the following behavior: limits for the triangle are 'a' and 'c', while 'b' is at the middle and refers to the top, hence the middle set has the parameters: a = 22.881, b = 40.083 and c = 57.285. For the trapezes, we have two parameters, one at top (a) and the other at bottom (b); for the left trapeze, we have a = 5.679 and b = 22.881. Triangles receive a value x and will respond attending the formula: if *(x<=a or x>=c)* the response is zero, if (x=b) the response is one, finally if *(x<b and x>a)* the response is *(x-a) / (b-a)* and if *(x>b and x<c)* the response is *(c-x) / (c-b)*. The same idea works for the trapeze, using only the side of interest. Therefore, the model is enabled to give response to specific values.

For example, let's consider User-2749058 (Table III) who has created 59 objects (Figure 5). This level of creation produces a membership of 0.9003022 for the "High" set and 0.0996977 for the "Very High" set. Our approach has included the use of fuzzy logic in order to provide results with higher semantics. In this example, we can imply that User-2749058 had a "high" production of objects. Moreover, we are using a specific approach for fuzzy logics, which allows us to classify the numerical amounts produced by the knowledge server. This classification process provides a qualitative perspective to the performance shown by participants or groups. By automating this process, we were able to give sense to all numerical results produced by the quantitative level of analysis.

IV. CONVERSATION ANALYSIS IN WEB FORUMS

It is very common that a computer supported cooperative environment (both CSCW and CSCL) includes discussion forums and chat-rooms. Consequently, a complete collaboration analysis model cannot disregard this important aspect of collaboration. Our work faces the analysis of these collaborative learning interactions by means of discourse analysis, involving NLP at certain level, thus providing a qualitative analysis dimension to our model by examining the quality of participants' contributions in conversations carried out in forums and chat rooms.

Our conversation analysis model is constructed in different layers [4]. When a query for an agent is issued, the middle layer is consulted (this layer stores the network of agents' location). Then this layer provides the reference to a thread that handles the socket bound to the specific agent involved in the current query. Figure 6 sketches the functionality layers of the conversation analysis model.

To test the conversation analyzer, we present the analysis of an example forum taken from the BSCW platform, where Spanish language was used. The messages produced were transported to a plain TXT formatted file, identifying the entry points for every message. Then a pre-processing program was used to reorder the file dividing it in separated sentences. Next, the sentences file is used to feed the word separator, which consults each word in every agent on-line. The agents produce their "natural responses" and the word separator produces a new file containing the Certainty Levels provided by every agent. Certainty Levels are produced by specific machinery which uses pattern recognition: every word is analyzed, looking for similitude to words in dictionaries. There are several dictionaries, one for every grammatical category, managed by a specific agent. These agents have retrieval trees inside, which were fed with information from the study performed by [17].
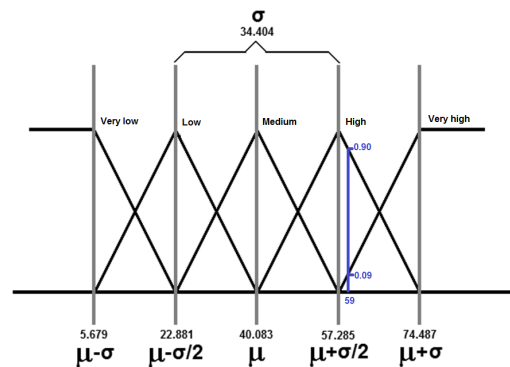

Figure 5.   Fuzzy model for the activity item "*CreateEvent*", which has an average value of **40.083** and a standard deviation of **34.404** from the global perspective. Membership levels are also shown (0.09 in "Very high" and 0.90 in "High") for the **59** objects created by **User-2749058**.
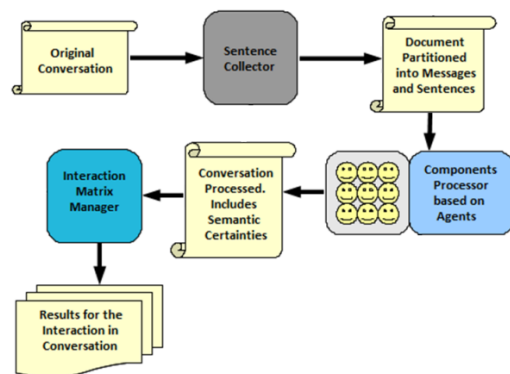

Figure 6.   Different layers for the conversation analysis designed.

Besides, the distance of edition is calculated in order to detect and solve the several *miskeying* errors present in Web forums. Finally, the produced complex file is processed by the conversation analyzer in its last layer, binding up the messages through the use of common elements.

Once the associations are made among messages, we are able to automatically discover the type of interactions that occurred in the forum. Table IV shows the adjacent matrix resulting from binding up messages. The resulting matrix is symmetric. The labels M1, M2, etc. refer to the messages involved in the whole conversation. In order to bind up these messages, the system does not compare plain strings; instead, it involves the grammatical role for the words and the similitude degree among the elements in different messages. Hence, number 7 relating M4 and M6 (marked in table IV) means that there are seven common elements -from the grammatical and synthetic approach-between these messages. The rest of numbers in table are calculated in the same way. The TV line and the TH column simply collect the sum of these elements. For instance, the number 26 for M1 means that this message has accumulated 26 elements which are common with the rest of the messages.

When messages are collected from the forum, every message includes some data such as author, date, time and forum's name. In this case, the authors' names have been replaced by codes such as UZZ, UMM, etc. Consequently, by reviewing the authorship of the messages involved in this conversation, it is possible to verify the authors of the messages as follows: UZZ(M1, M4, M8), UMM(M2, M5), UPR(M3), UPA(M6), UUV(M7), UIN(M9), and UBA(M10).

By knowing the authorship of messages and the way these messages are connected through the accumulation of common concepts (discovered among the messages in Table IV), the system is capable of constructing Table V automatically. Table V contains the number of interactions among participants characterized by the use of common elements in their messages. In the conversation analyzed, seven users participated and are represented by the codes UZZ, UMM, UPR, etc. in table V. For instance, let's consider user UZZ. This user started the conversation with a question asking to know "the goals of every member in the group". The following contributions in the forum will, hopefully, try to answer that question. In fact, the big numbers in UZZ column show a very active interaction between the other group members and UZZ, which means highly connected responses to UZZ's question. Even the most indifferent participant in conversation, UBA in the last row, had 3 interactions with UZZ. As a consequence, UZZ's question proves to be the axis of the whole conversation, a fact that can be validated by a human if he/she reads up the conversation. The final row in this Table contains the total number of interactions that involve the user who is at the top row.

As a result, the association of messages through common elements (shown in Table IV) and the knowledge of authorship of messages enable us to discover and establish a complete association among participants, based on the use of common elements with common roles in messages. We believe that there is a distribution for the ability to interact with counterparts in conversation, and we named Interaction Capacity (IC) to such distribution. Hence, participants in conversation can be bound up confidently as a network.

Furthermore, in order to identify the social roles, additional associations are built through the analysis of conversation. Table VI (a) shows the way each participant has distributed its IC with his/her counterparts; for instance, UZZ had 31% of its IC mingling with UMM and UMM had 46.3 of its IC mingling with UZZ, or UPR had 11% of its IC mingling with UIN, and so forth. Besides, Table VI (b) shows the overall IC percentage based on the total number of micro-interactions in the whole IC; hence, the interaction between UZZ with UMM represents the 9.5% of all the interactions occurred. Table VI (b) also shows the total number of interacting acts in conversation per participant. For instance, the final row indicates that UZZ had participated in the 31% of the conversation acts, whereas UPR had only a 4.5% participation.

TABLE IV.
ASSOCIATIONS DETECTED AMONG MESSAGES IN CONVERSATION THROUGH THE USE OF COMMON REFERENCES TO: COURSES, OBJECTS, SITUATIONS AND USERS' NAMES. THIS AMOUNT REPRESENTS THE NUMBER OF TIMES A COMMON REFERENCE HAS BEEN DETECTED IN THE REVIEWED MESSAGES.

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	TH
M1	0	3	1	0	4	6	6	0	4	2	26
M2	3	0	0	7	0	3	3	0	1	2	19
M3	1	0	0	2	1	1	2	1	1	0	9
M4	0	7	2	0	4	7	6	0	4	1	31
M5	4	0	1	4	0	5	2	1	3	2	22
M6	6	3	1	7	5	0	4	2	2	1	31
M7	6	3	2	6	2	4	0	1	3	1	28
M8	0	0	1	0	1	2	1	0	0	0	5
M9	4	1	1	4	3	2	3	0	0	1	19
M10	2	2	0	1	2	1	1	0	1	0	10
TV	26	19	9	31	22	31	28	5	19	10	

TABLE V.
NUMBER OF INTERACTIONS ATTRIBUTED TO EACH CONVERSATION PARTICIPANT

	UZZ	UMM	UPR	UPA	UUV	UIN	UBA
UZZ	0	19	4	15	13	8	3
UMM	19	0	1	8	5	4	4
UPR	4	1	0	1	2	1	0
UPA	15	8	1	0	4	2	1
UUV	13	5	2	4	0	3	1
UIN	8	4	1	2	3	0	1
UBA	3	4	0	1	1	1	0
	62	41	9	31	28	19	10

TABLE VI.
INTERACTION DISTRIBUTION FOR EACH PARTICIPANT: (A) SHOWS THE % DISTRIBUTION PER PARTICIPANT (REGARDING ONLY THE ACTIVITY FOR THAT PERSON), AND (B) THE % DISTRIBUTION PER PARTICIPANT WITH REGARD TO THE WHOLE NUMBER OF INTERACTION ACTS DETECTED.

(a)

	UZZ	UMM	UPR	UPA	UUV	UIN	UBA
UZZ	0	46.3	44	48.4	46.4	42	30
UMM	31	0	11	25.8	17.9	21	40
UPR	6.5	2.44	0	3.23	7.14	5.3	0
UPA	24	19.5	11	0	14.3	11	10
UUV	21	12.2	22	12.9	0	16	10
UIN	13	9.76	11	6.45	10.7	0	10
UBA	4.8	9.76	0	3.23	3.57	5.3	0
	100	100	100	100	100	100	100

(b)

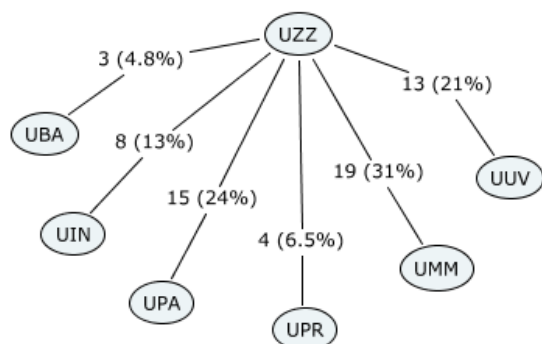|	UZZ	UMM	UPR	UPA	UUV	UIN	UBA
UZZ	0	9.5	2	7.5	6.5	4	1.5
UMM	9.5	0	0.5	4	2.5	2	2
UPR	2	0.5	0	0.5	1	0.5	0
UPA	7.5	4	0.5	0	2	1	0.5
UUV	6.5	2.5	1	2	0	1.5	0.5
UIN	4	2	0.5	1	1.5	0	0.5
UBA	1.5	2	0	0.5	0.5	0.5	0
	31	20.5	4.5	15.5	14	9.5	5	100

Figure 7. Representation of the interaction amounts and their proportions among user UZZ and his/her peers in the conversation analyzed.

Combining the analysis results of Tables IV, V and VI, interesting conclusions can be drawn about the behavior and collaboration of participants in the conversation. As it was expected, the first message in conversation acts as a trigger. In the analyzed case study, the first message was sent by participant UZZ. As shown in Figure 7, most of the interaction among participants has been organized as a star around the participant who is the owner of the first message. The following step consists in measuring the interaction levels in order to discover dialectical pairs. Higher number of common elements among participants implies higher interaction levels among the same participants. Table V shows these interaction amounts between pairs.

Besides, Table VI reorganizes those results and presents them as proportions for every participant (Table VI.a) and for the whole conversation studied (Table VI.b). These results (amounts and proportions) provide valuable awareness regarding the roles undertaken by participants along the conversation. In particular, Table IV provides timing awareness whereas Tables V and VI facilitate awareness regarding participants' interaction and commitment to the conversation's goals. Hence, for the given interaction scenario, the activity supervisor can use the quantitative data collected in the Tables and, at the same time, employ the graph (such as the one shown in Figure 7) to draw qualitative interpretations of participants' behavior regarding roles and interactivity intentions and thus can successfully understand the participants' collaborative behavior that takes place in conversation.

V. COMBINING RESULTS FROM QUANTITATIVE AND CONVERSATION ANALYSES

To obtain a complete picture of all the collaborative interactions that took place in the different shared workspaces, we need to combine the results coming from the quantitative and conversation analyses.

Based on an initial effort to interpret the results that come from both analysis dimensions, the quantitative dimension shows that it is a regular phenomenon that one of the team members acts as the leader –even if s/he has not being designated as such. Besides, another member used to act as his/her dialectic pair by actively responding to the leader's actions. The rest of team members act as supportive ones with more or less commitment. It is interesting to note that a similar individual and group behavior has been also shown by the conversation analysis dimension.

Another interesting fact that we observed concerns the distinct analysis approach we had to employ in the two dimensions. Interaction analysis in shared non-verbal workspaces, where the basic actions are object sharing such as those in Table I, presents specific coding rules for messages and turns which are strongly influenced by the course plan. In contrast, conversation-based interaction follows up the conventional rules defined for chat rooms or web forums. Certainly, more work is needed to draw more solid conclusions.

VI. CONCLUSIONS AND FUTURE WORK

The current study presents the results of an approach that is based on network models to analyze and manage collaborative learning interactions that take place in different media (shared non-verbal workspaces and web chats or forums). As such, the analyses was performed on two dimensions and concerned, on the one hand, the study of interaction activity expressed by primitive acts stored in log files in a CSCL environment and, on the other hand, conversation analysis carried out through a NLP approach. Both analysis types are based on networked structures. The quantitative analysis realized on CSCL log files is based on an active element called *knowledge server*, which is supported by a complex network that models human interaction and the knowledge objects they gather, share and act upon. The qualitative analysis establishes an interaction network which models the conversation carried out by work team members. Though both networks are built separately, they use data collected in same time period and involve the same team members. Hence, the discovered connections and the obtained results can be bound together to provide a more coherent interpretation and management of members' collaborative activity. Networking models constitute a promising approach to this end, however further work is in progress to build a holistic collaboration analysis model that will enable multi-level knowledge indicators to explain complex collaboration acts in a well-grounded and effective manner.

The currently achieved results show an important correlation between indicators coming from different types and dimensions of the analysis. A proposal as the one we are providing has included innovative approaches involving complex networks analysis fed with diverse types of data. Discovering correlations among indicators coming from different sources with specific coding is a novel style of collaboration analysis.

REFERENCES

[1] B. Barron, R. Pea & R. Engle. "Advancing understanding of collaborative learning with data derived from video records". The International Handbook of Collaborative Learning. Hmelo-Silver, C.E., et al. (Eds.), New York: Routledge, pp. 203-219. 2013

[2] L. Casillas & T. Daradoumis, "A Quantitative Treatment to Data from Computer-Supported Collaboration: An Ontological Approach". In proceedings of the 2nd International Conference on Complex, Intelligent and Software Intensive Systems (CISIS-2008), Barcelona, Spain, March 4-7, 2008, IEEE Computer Socie-

ty, Los Alamitos, CA, USA, ISBN: 0-7695-3109-1, pp. 226-232. 2008

[3] L. Casillas & T. Daradoumis, "Knowledge Extraction and Representation of Collaborative Activity through Ontology-based and Social Network Analysis Technologies". International Journal of Business and Data Mining (IJBIM). Volume 4, Issue 2. Inderscience Publishers, Switzerland. ISSN: 1743-8195, pp 141-158. 2009

[4] L. Casillas & T. Daradoumis, "Discovering Social Relationships and Intentions in Web Forums Using Natural Language Processing Techniques". Proceedings of the International Conference WebScience'09: Society On-Line, Athens, Greece, March 18–20, 2009. [on-line]: http://journal.Webscience.org/111/ 2009

[5] H. Clark. "Using language". Cambridge: Cambridge University Press. 1996 http://dx.doi.org/10.1017/CBO9780511620539

[6] C. Hmelo-Silver, C. Chinn, C. Chan & A. O'Donnell (Eds.). "The International Handbook of Collaborative Learning". New York: Routledge. 2013

[7] S. Puntambekar. "Mixed methods for analyzing collaborative learning". The International Handbook of Collaborative Learning. Hmelo-Silver, C.E., et al. (Eds.), New York: Routledge, pp. 220-249. 2013

[8] A. Soller & A. Lesgold, "Analyzing Peer Dialogue from an Active Learning Perspective". Proceedings of the AI-ED 99 Workshop: Analyzing Educational Dialogue Interaction: Towards Models that Support Learning, pp 63-71; LeMans, France. 1999

[9] A. Soller & A. Lesgold. "Modeling the Process of Collaborative Learning". International Workshop on New Technologies in Collaborative Learning. Computer-Supported Collaborative Learning Series. Volume 9, The Role of Technology in CSCL, pp 63-86, Springer, USA. 2000

[10] G. Stahl. *"Theories of Collaborative Cognition: Foundations for CSCL and CSCW Together"*. Computer-Supported Collaborative Learning at the Workplace: CSCL@Work. CSCL Series. Vol. 14, 43-63. Springer. 2013

[11] G. Stahl, T. Koschmann & D. Suthers. *"Computer-supported collaborative learning: An historical perspective"*. In R. K. Sawyer (Ed.), Cambridge handbook of the learning sciences (pp. 409-426). Cambridge, UK: Cambridge University Press. 2006

[12] J. W. Strijbos & F. Fischer. "Methodological challenges for collaborative learning research". Learning and Instruction, 17(4), 389-393. 2007 http://dx.doi.org/10.1016/j.learninstruc.2007.03.004

[13] S. Yang & I. Chen. *"A social network-based system for supporting interactive collaboration in knowledge sharing over peer-to-peer network"*. Int. J. Human-Computer Studies, 66, pp. 36–50. 2008 http://dx.doi.org/10.1016/j.ijhcs.2007.08.005

[14] J. Lonchamp. *"A three-level analysis of collaborative learning in dual-interaction spaces"*. International Journal of Computer-Supported Collaborative Learning, September 2009, Volume 4, Issue 3, pp 289-317. 2009

[15] B. Barron, R. Pea & R. Engle. *"Advancing understanding of collaborative learning with data derived from video records"*. The International Handbook of Collaborative Learning. Hmelo-Silver, C.E., et al. (Eds.), New York: Routledge, pp. 203-219. 2013

[16] J. Carroll, M. Rosson, G. Convertino, & C. Ganoe. *"Awareness and teamwork in computer-supported collaborations"*. Interacting with Computers, volume 18, pp 21–46. Elsevier Science Ltd. 2006

[17] Almela, R. "Frecuencias del español, diccionario y estudios lexicos y morfologicos". Ed. Universitas, Spain. 2005

AUTHORS

L. Casillas is with University of Guadalajara, Guadalajara, Mexico.

T. Daradoumis is with Open University of Catalonia, Barcelona, Spain and University of the Aegean, Mytilene, Greece.

S. Caballé is with Open University of Catalonia, Barcelona, Spain.