#02

Casillas, L. (2011). Estimating time between creation and achievement of knowledge objects in learning groups through social network analysis. *eLC Research Paper Series, 3, 16-25.*

ESTIMATING TIME BETWEEN CREATION AND ACHIEVEMENT OF KNOWLEDGE OBJECTS IN LEARNING GROUPS THROUGH SOCIAL NETWORK ANALYSIS

Luis Casillas

Computer Science Department
University of Guadalajara
Mexico
Revolución 1500,
módulo "O", planta baja.
Guadalajara, Mexico.
luis.casillas@red.cucei.udg.mx

Estimating time between creation and achievement of knowledge objects in learning groups through social network analysis

ABSTRACT

Networked collaboration performed over specific platforms designed for such purposes can provide knowledge about roles, intentions and effects regarding participants, their interaction among themselves and the interaction with the available knowledge objects. This study aims to propose a mechanism for discovering temporal behaviour underlying the raw data collected in log files from e-learning activity in specific platforms. The proposal is based on measuring and, subsequently, estimating time spans through social networks analysis (SNA). The main focus of this work is to match different temporal behaviours, shown during collaborative learning, with formal profiles identified inside a complex network of interactions. The final goal is to define a concrete mechanism to measure the response of participants, from the perspective that knowledge objects have been created by the partners in the same learning group.

KEYWORDS

Collaborative learning, Temporal analysis, Time series, Social networks.

1. INTRODUCTION

The analysis of collaborative learning in virtual environments has become a significant task when supervising and monitoring the performance of learning groups. Regular interactions (face-to-face) for collaboration include a diversity of communication elements beyond the spoken language. This complex "message exchange" would imply, hopefully, a synchronization of minds, which is supported by a "shared" network of beliefs and concepts.

The experience of being present *in situ* when collaborating, provides a complete feeling of situation-awareness (Endsley, 1995). Collaboration in virtual environments does not provide this rich experience of having a (full) situation-awareness. Limited channels of communication restrict the full transfer of the messages generated by each member of the learning group. Collaboration analysis is then oriented to discover the intentions and effects of every primitive collaborative action (Carroll et al., 2006). It is evident that mutual understanding among peers is the first step to solve any problem that may appear when different people are involved in the search for a solution. Besides, it is a fact that the Web has become a highly social environment that is sustained by social structures that generate content in networking fashion (Anklam, 2009). Every member of the group might have a different approach to the problem and its possible solution. Although in networking there is no formal and ordered process to follow, the initial stages must be clearly oriented to model a common structure of concepts regarding the problem and its solution.

The main task is to discover the knowledge that is hidden among the large volume of data stored in virtual environments (supporting the networking). This should be performed during collaboration analysis. Hence, the specialists performing the analysis must collect and process data through mining techniques and do the required inference for filling in those indicators defined for measuring the networked collaboration. As explained in Casillas and Daradoumis (2008), the current work is oriented at gathering temporal information from the very same interactions already analysed in previous studies.

2. RELATED WORK

The temporal analysis of social networking activity in specific environments implies an effort with regard to the complexity of gathering knowledge from the raw data stored by these platforms. In particular, the work performed by Soller and Lesgold (2000) 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. Hence, it has been possible to compare current behaviours by matching them with expected behaviours.

Another interesting approach for performing collaborative analysis is presented by Soller and Lesgold (1999), which presents a Model of Effective Collaborative Learning as a framework that provides meaning to collaborative learning acts by classifying the primitive actions into skills, sub-skills and attributes. This model depends on the use of certain "sentence openers" that express the intentions of the collaborators; however, this restricts the flexibility of collaboration due to the mechanization of communication. Nevertheless, it is clearly understandable that the complexity of free-style human communication is an important challenge for current computational approaches. This trend of collaboration analysis has been maintained in subsequent research work by Soller (2001, 2004), focusing on the search for specific messaging structures.

Using models for analyzing collaborative learning is becoming a key aspect. A model gives coherence to the study of collaboration phenomena by classifying primitive acts in predefined slots, and gathering semantics in the process. The study developed by (Daradoumis et al., 2006) deals with the collaboration analysis through a layered model, which breaks down the analysis goals into different levels. The lower level is related to raw events generated by a CSCL tool, whereas the higher level is related to abstract networked collaboration analysis. The problem with this approach is the increasing complexity when trying to infer abstract behaviours from the very raw data collected in log files from CSCL tools. Nevertheless, collaborative analysis under such circumstances is more flexible, though flexibility is always harder to implement.

Another approach for performing an analysis of collaborative learning is presented in the 2006 study by Daradoumis and Casillas that looked at the problem from a neural network perspective. Log files are consumed through a mechanism that classifies information using a *sui generis* neural net. This approach can offer useful information inbetween the structure of the neural network and not only in the output layer (as usually happened with conventional neural networks). Every element in this neural net has information regarding an indicator of collaboration activity, while, at the same time, the indicator feeds some of the neurons in the next layer; this process prepares the indicators of the following layer.

This paper proposes performing a social network analysis (SNA) in two stages. The first deals with collecting and processing raw data from a CSCL platform, and applies a quantitative analysis method to fill up a series of indicators. This stage is already satisfied in Casillas and Daradoumis (2008). The second stage – the focus of this paper – sets out to discover the response profiles (from a temporal approach).

3. GATHERING KNOWLEDGE FROM SOCIAL NETWORKING ACTS

The essence of the task is to develop a precise and useful approach for the analysis of the collaborative interactions of participants in a networked environment, involving different analysis dimensions. According to the experience acquired through the review of several cases, we identified different strategies to approach the SNA. These strategies range from the simple collection of data and results in spreadsheets, queries to data bases, data mining and decision trees, to the use of agents with specifications that satisfy different criteria for measuring the networking activity. Naturally, simple approaches are easy to design and implement, but their results are usually weak and they lack some way of automatic interpretation. Developing a more advanced solution implies extra effort, but the benefits are manifold, since we can obtain a variety of in-depth, quality results that provide a more powerful tool for decision making and interaction monitoring.

One may initially consider that the complexity of performing SNA on online human interaction via a computational solution to be apparently low due to restrictions of the communication channels available, which reduce the amount of information elements produced. Indeed, the restrictions in transferring information could erroneously imply that a rather simple solution would suffice; however, it is precisely the absence of certain information elements that is the cause of the complexity, due to the need to fill a number of 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 a number of collaborative issues; therefore, special effort must be made to discover some of the participants' interests and intentions lying beneath their collaboration acts.

Taking these considerations into account, we first need to carry out a preliminary process so that we can understand the way networking interaction evolves and then build a method for gathering higher-level semantics indicators from primitive collaborative actions. To this end, we developed temporal schemata capable of working as an abstract framework for

Table I. **Temporal schemata for measuring collaborative actions**

AXIOM	FORMULA
Let X be any raw action; there is at least one Y which is an interpreted activity resulting from X. Raw actions enable interpreted activity, which are complex behaviours beneath specific sets of raw actions.	∀X (rawAction(X) → ∃Y intdActivity(Y))
Every raw action generated in the past implied a subsequent interpreted activity.	■ (rawAction → ◆ intdActivity)
All raw actions in the past, implied current interpreted activities.	■ rawAction → intdActivity
All raw actions in the past will imply interpreted activities at some future moment.	■ rawAction → ◇ intdActivity
Some raw actions in the past implied current interpreted activities.	◆ rawAction → intdActivity
Some raw actions in the past will imply interpreted activities at some future moment.	◆ rawAction → ◇ intdActivity
Raw actions perfd in the previous moment implied current interpreted activities.	● rawAction → intdActivity
Current raw actions will imply interpreted activities in the next moment.	rawAction → ○ intdActivity
Current raw actions will imply interpreted activities at some future moment.	rawAction → ◇ intdActivity
Current raw actions will imply interpreted activities in the future.	rawAction → □ intdActivity
Future raw actions will imply interpreted activities at some future moment.	□ (rawAction → ◇ intdActivity)
Every raw action performed by any collaborator (Agent) implies an interpreted activity performed by the same collaborator in the same moment.	∀X ((rawAction(X)∧perfd(X,Ag0,t0)) → ∃Y (intdActivity(Y)∧perfd(Y,Ag0,t0)))
A sequence of raw actions performed by a collaborator (Agent) implies an interpreted activity performed by the same collaborator at some later moment.	∀X ((seqRawActions(X)∧perfd(X,Ag0,t0...*tn*)) → ∃Y (intdActivity(Y)∧perfd(Y,Ag0,*tn*+1)))
A sequence of raw actions performed by different collaborators (Agents) implies an interpreted activity performed by the same collaborator at some later moment.	∀X ((seqRawActions(X)∧perfd(X,groupColls (Ag0...Ag*n*),t0...*tn*)) → ∃Y (intdActivity(Y) ∧perfd(Y, groupColls(Ag0...Ag*n*),*tn*+1)))
A sequence of raw actions performed by different collaborators (Agents) implies a sequence of interpreted activities performed by the same group of collaborators at some later moment.	∀X ((seqRawActions(X)∧perfd(X,groupColls (Ag0...Ag*n*),t0...*tn*)) → ∃Y (seqIntdActivity(Y) ∧perfd(Y, groupColls(Ag0...Ag*n*),*tn*+1)))

modelling the chronological phenomena that lie behind the networking activity, based on the temporal-logics tools proposed by Fisher (2006). Table I shows the axioms that involve the formulae used to express and measure the collaborative actions that occur and evolve over time.

Our study and perspective about networking has two dimensions. The first one is related to the analysis of tacit collaborative actions that result from the simple circuit of putting knowledge objects in shared spaces, which are then being accessed by other peers who have the required permission to participate in that shared space. Figure 1 shows a sketch of this networking approach. The second dimension is more oriented to detecting the temporal profiles shown by users throughout the semester. This paper includes some results from that part of the study, and they are considered to reach an integrated solution to the problem of analyzing collaborative interaction and awareness. To this end, we show some important results from of our study.

Our approach was applied in real collaborative learning situations that took place in the Basic Support for Cooperative Work (BSCW) platform (http://bscw.uoc.es/). The BSCW tool is able to record many types of primitive actions (interactions, transactions, accesses, etc.) performed by users in log files. The BSCW was used as a CSCL support. These data provide an important source of information in order to study both the individual and group performance in learning. In fact, there are thousands of data entries recorded from the interactions among users; more than 25 megabytes per semester. The central part of the quantitative-analysis stage consisted of processing the whole data stored in log files to build a social network based on the automatic detection of interactions according to the collaboration model shown in Figure 1.

Figure 1. When a Creation-Access circuit is completed, two networked collaboration models are manifested in a CSCL environment.

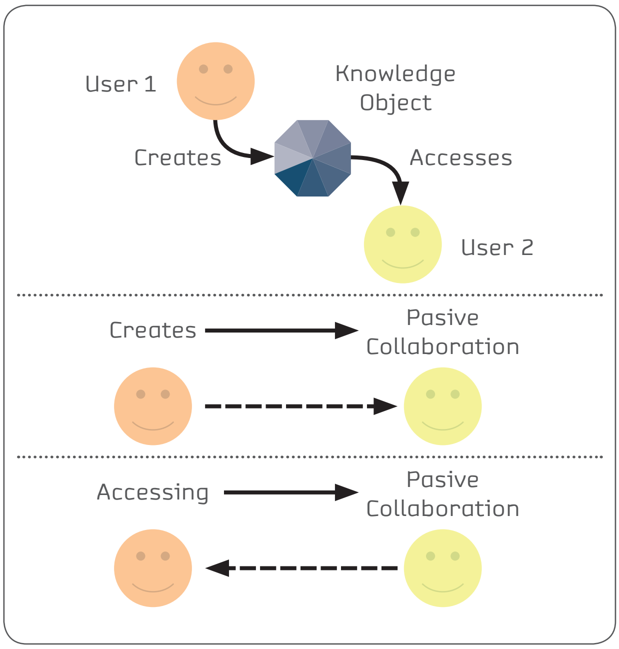
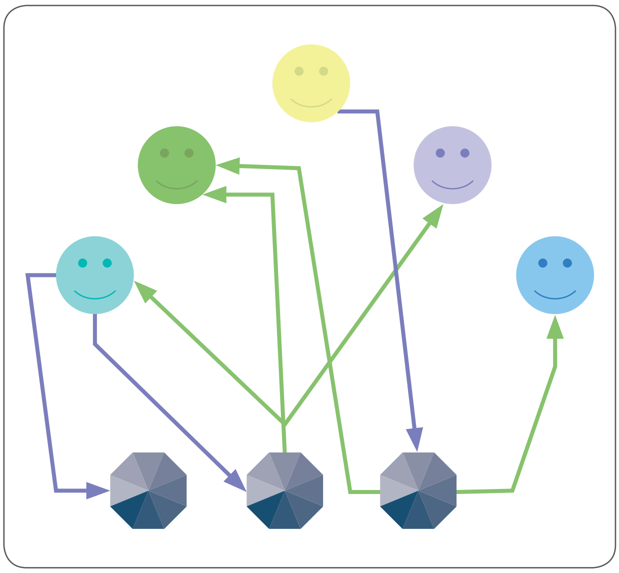


Figure 2. Knowledge objects are the elements which bind users as a social network.



The way of collecting and organising the information regarding the transactions made inside the BSCW provides important elements for building a social network. This network represents the users and the knowledge objects stored in the shared learning spaces by binding users (students) to those knowledge objects that they access. Students are organised in learning groups, which is an abstract social network. Our aim is to discover and analyse the interactions among these students using different SNA techniques. Figure 2 shows a plain perception of such a social network.

Since students may belong to more than one team and, thus, have access to a variety of shared knowledge objects, the social network developed for carrying out the quantitative analysis stage is not limited to isolated learning teams. Instead, we faced the challenge of building and managing a huge social network that functions as a knowledge

Figure 3. A schema for processing collaborative learning interactions from log files.

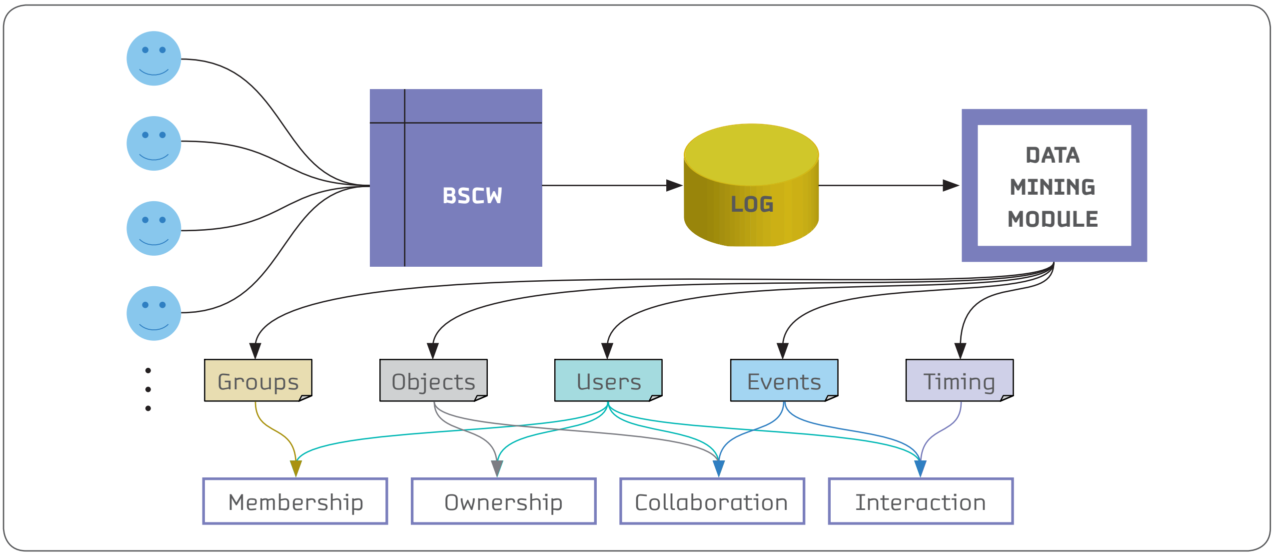
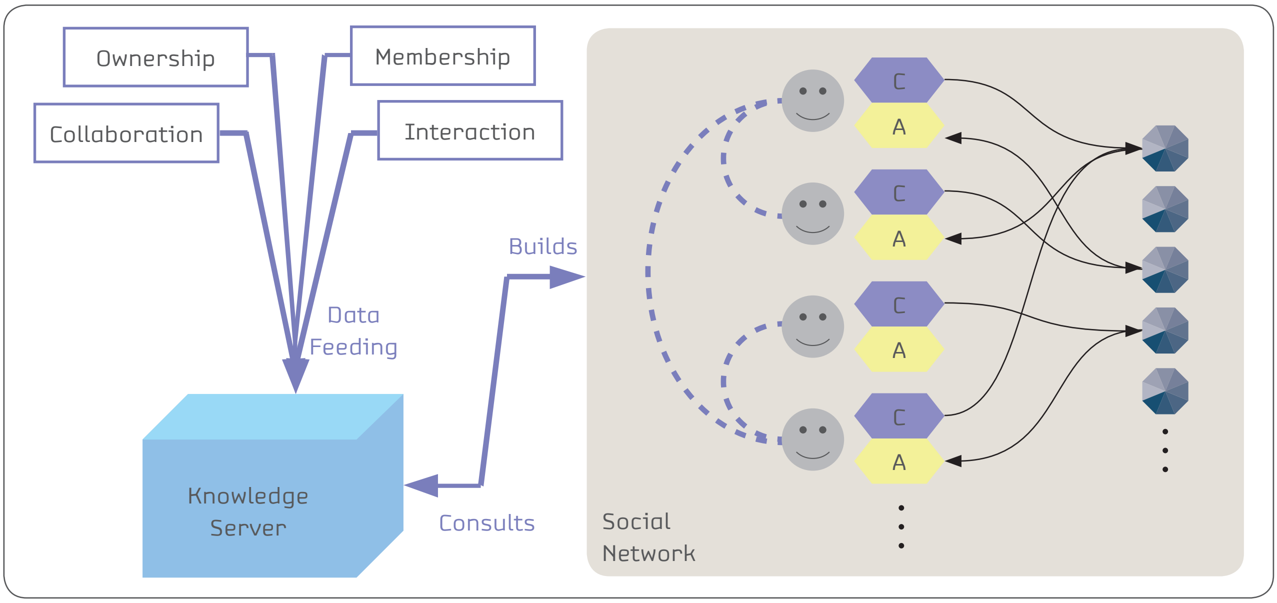


Figure 4. The Knowledge Server module is fed by the coupled model built from activity logs. In the social network, "C" stands for "creates" and "A" for "access".



server and provider. In fact, the network itself represents a semantic model of the collaboration that takes place among the whole class of participating users. Figure 3 presents a simplified schema of the process used for gathering significant knowledge from the raw data stored in log files, which is a sub-process of the quantitative approach employed. As a matter of fact, the collaborative learning experience held in the BSCW produced a huge amount of raw data. These data were processed through mining techniques, in order to discover relationships, and through a model that handles the involved elements via multiple queries and coupling operations over the database. This allows us to relate users based on their accesses to knowledge objects.

Once the log files are processed, the information regarding collaboration is used for feeding the Knowledge Server, a module which first builds the social network that represents the collaborative activity among users and then exploits the social network data to model a variety of interaction phenomena *a posteriori*. It is possible to figure out the functionality of this module by reviewing the schema shown in Figure 4.

In particular, the Knowledge Server becomes the data provider for the agents that are further employed to fulfil the activity indicators, which explain different aspects and issues of collaboration according to specifications defined in an ontology that is specifically designed to support our approach of collaboration analysis.

4. TEMPORAL ANALYSIS OVER COLLABORATION

Our approach constitutes a natural extension of an ongoing effort to provide a rich representation scheme that supports collaboration analysis from CSCL, which started from the proposal of Soller and Lesgold (1999) and went through to the work suggested by Daradoumis et al. (2006). The studies from Casillas and Daradoumis (2008, 2009) provided the guidelines for discovering semantics from the collaborative learning acts, although some analysis over temporal performance is pending.

This analysis is not directly connected to any specific CSCL tool; therefore an additional mechanism needed to be inserted between the log-files database and the knowledge server. This inner mechanism was implemented using extraction-transformation-load (ETL) techniques, which unify, contract or expand the stored data in order to achieve the required format that matches them with the ontology.

At first, the understanding regarding time influence on the process was that the reaction period would be in between the global average time for reactions observed in the whole activity. This reaction refers to any act over a knowledge object after its creation. Nevertheless, this approach was not founded on any formal probe. The works from Casillas and Daradoumis (2008, 2009) were aimed at detecting another kind of behaviour, and studying the reaction time was not their main goal. This paper, however, does have such a goal. The initial hypothesis of this paper is that reaction time would vary throughout the semester according to the increase of time pressure over students. This specific phenomenon is clearly observed in students working in traditional academic environments. Thus, the question is:

Will participants in CSCL (organised as learning groups) vary their response according to the proximity of deadlines?

In order to deal with this matter, we have considered a time series analysis. The data from reaction periods have been collected from the knowledge server. This information was collected for the whole group and for specific learning groups. Due to non uniform presentation in timing data, two normalization steps were needed. The first was oriented at present dates and hours from events in a single data set. It was used an internal representation based on floating point numbers. This form is the same used by Microsoft Excel to handle timing information in numerical forms, i.e. the date "03/02/2003 08:16:47 p.m." is translated as "37682.8449884259". This representation has the capacity to maintain the distance among reaction events. The second normalization step consisted of calculating the average for the normalized timing, as well as the standard deviation these data were used to calculate:

$$z=(x-\mu)/\sigma$$

Where ***z*** is the normalized value to be calculated, ***x*** is the original form of the value, ***μ*** is the average of the whole values for ***x*** and σ is the standard deviation from those values of ***x***.

Once the data from reaction time were normalized, we started the regression analysis for the time series. The equations used for the calculation were:

$$m = \frac{n\sum(xy) - \sum x \sum y}{n\sum(x^2) - \left(\sum x\right)^2}$$

$$b = \frac{\sum y - m\sum x}{n}$$

$$r = \frac{n\sum(xy) - \sum x \sum y}{\sqrt{[n\sum(x^2) - (\sum x)^2][n\sum(y^2) - (\sum y)^2]}}$$

Where ***n*** is the quantity of data available in the time series, ***x*** refers to the time events occurred, ***y*** refers to the time measurements (reaction period) corresponding to the events in ***x***, ***y=f(x)***, and finally the constants ***m*** (slope) and ***b*** (y-intercept) of the equation ***y=mx+b***. In additionl, ***r*** is the correlation coefficient between ***x*** and ***y***. The indicators ***m***, ***b*** and ***r*** are calculated for the global performance collected from the activity in the BSCW in the Spring 2003 semester (a semester with high activity levels in different dimensions). The results from this action are rather interesting. Table II shows the main data and results for this calculation.

The slope is clearly flat and the correlation is almost void. At this point, there is no apparent influence of the date or proximity to deadline, over the period to react after a knowledge object has been created in a shared space. Nevertheless, the data is insufficient to draw immediate conclusions. The different spaces are bound to different subjects and every subject has its own rules, according to the academic programme.

Hence, the same calculations were carried out on the data collected from specific learning groups. Unfortunately the results are very similar to the results in the analysis conducted on the global data. One of these analyses is shown in table III.

Though there are significant differences regarding the volume of data analysed and the intermediate results, the relationship discovered by the linear regression shows no significant changes from the performance shown by participants during the whole experience.

On the one hand, there is no apparent influence from the passage of time and/or closeness to deadlines in the performance of participants' reactions. The flatness of the slope in both cases, global and local to workgroup, indicates that response will not vary with time in the social network. On the other hand, although there is no linear model, ***y=mx+b***, to predict the performance, there is indeed the confidence to use the average reaction span, gathered from global behaviour, to measure specific reactions in learning groups and even specific

Table II. Main data and results for the calculation to build a linear regression in the time series for the global events performed in the BSCW tool considering the collaboration through objects in shared spaces. Specifically, this time series is bound to the period stated from the creation of the object, until the moment in which it is accessed.

n	∑x	∑y	∑(xy)	∑x²	∑y²	(∑x)²	(∑y)²
87670	130662.5617	44539.19186	71995.9945	23345917.3	110285.984	17072705021	1983739612

m: 0.000242545	**b:** 0.507670815	**r:** 0.003941674

Table III. Main data and results for the calculation to build a linear regression in the time series for the events performed by one workgroup in the BSCW tool. This was calculated considering the collaboration through objects in the shared space. Specifically, this time series is bound to the period stated from the creation of the object, until the moment in which it is accessed.

n	∑x	∑y	∑(xy)	∑x²	∑y²	(∑x)²	(∑y)²
1036	-0.000143194	-1.25762E-08	82.588223	1035	1849.70442	2.05044E-08	1.58162E-16

m: 0.079795385	**b:** 1.1017E-08	**r:** 0.05968934

participants. The measuring can be performed irrespective of which part of the semester is being observed.

5. CONCLUSIONS

This paper presents an approach for analysing online collaborative learning through social networks by integrating different strategies for coping with a variety of issues of the problem using mainly a quantitative technique.

This paper focuses on the quantitative method of time analysis which was based on SNA and consists of building a network using the information stored in log-files. The social network is the kernel of a Knowledge Server which is capable of answering different queries involving performance, interaction and collaboration for distributed problem solving in e-learning situations.

Temporal logics were defined to measure collaborative learning actions, whereas the ontology represents and classifies the primitive actions performed by participants. By imposing order on primitive events and applying semantics to specific sequences of primitive actions, collaboration analysis acquires a formal basis.

Finally, there is no apparent influence from the time span and/or closeness to *deadlines* in the performance of participants' reactions. The flatness of the slope in both cases, global and local to learning groups, indicates that response will not vary with time. Although there is no linear model, ***y=mx+b***, to predict the performance, there is indeed the confidence to use the average reaction span, gathered from global behaviour, to measure specific reactions in learning groups and even specific participants. The measuring can be performed irrespective of which part of the semester is being observed.

This is an interesting point. The first studies conducted on these data used an average of the reaction time, but a time series analysis was needed to determine whether there was a trend behind data. Now, it has been verified that reaction span is not bound to time passage or closeness to deadlines. Users of the CSCL will tend to follow regular behaviours during collaborative experience in social networking, which is more bound to personal customs regarding access to the Internet. Hence, the average for reaction time had been used correctly, although it was restricted to specific circumstances; now this has been proven to be the case.

References

Anklam, P. (2009). The Connection <--> Collaboration Continuum. The{app}gap. Retrieve May 18th, 2011. From http://www.theappgap.com/the-connection-collaboration-continuum.html

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

Casillas, L. & Daradoumis, T. (2008). 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 Society*, Los Alamitos, CA, USA, ISBN: 0-7695-3109-1, pp. 226-232.

Casillas, L. & Daradoumis, T. (2009). 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.

Daradoumis, T., Martínez A., & Xhafa, F. (2006). A Layered Framework for Evaluating Online Collaborative Learning Interactions. *International Journal of Human-Computer Studies*. Special Issue on "Theoretical and Empirical Advances in Groupware Research", Academic Press: Elsevier Ltd., ISSN: 1071-5819, Volume 64, Issue 7, July 2006, pp. 622-635.

Daradoumis, T. & Casillas, L. (2006). A Neural Approach for Modeling the Inference of Awareness in Computer-Supported Collaboration. Lecture Notes in Computer Science (LNCS 4227). *Innovative Approaches for Learning and Knowledge Sharing*. W. Nejdl and K. Tochtermann (Eds.). Springer-Verlag Berlin Heidelberg, ISBN: 3-540-45777-1, pp. 464 – 469. (2006).

Endsley, M. (1995). Toward a theory of situation awareness in dynamic systems. *Human factors*, Volume 37 Issue 1. ISSN 0018-7208. Texas, USA. (1995).

Fisher, M. (2006). Implementing Temporal Logics: Tools for Execution and Proof. Proceedings of the Sixth International Workshop on Computational Logic in Multi-Agent System (CLIMA); Lecture Notes in Artificial Intelligence 3900, Springer-Verlag; London, U.K.

Soller, A. (2001). Supporting Social Interaction in an Intelligent Collaborative Learning System. *International Journal of Artificial Intelligence in Education*, hosted by the School of Informatics at The University of Edinburgh; Edinburgh, U.K.

Soller, A. (2004). *Computational Modeling and Analysis of Knowledge Sharing in Collaborative Distance Learning*. User Modeling and User-Adapted Interaction, number 14; Kluwer Academic Publishers. Netherlands.

Soller, A., & Lesgold, A. (1999). Analyzing Peer Dialogue from an Active Learning Perspective. Proceedings of the AI-ED 99 Workshop: Analysing Educational Dialogue Interaction: Towards Models that Support Learning, LeMans, France.

Soller, A., & Lesgold, A. (2000). Modeling the Process of Collaborative Learning. *International Workshop on New Technologies in Collaborative Learning*, Awaji-Yumebutai, Japan.