CODING PROCEDURES TO ANALYSE INTERACTION PATTERNS IN EDUCATIONAL WEB FORUMS

Stefania Manca, Manuela Delfino, and Elvis Mazzoni

aIstituto per le Tecnologie Didattiche, Consiglio Nazionale delle Ricerche, Italy; bSEFoRA Lab, Facoltà di Psicologia, Università di Bologna, Italy

Abstract

Analysis of interaction patterns is one of the most important indicators of quality of learning in educational web forums. Social Network Analysis is gradually assuming importance in the study of interaction patterns as it focuses on the analysis of the interrelationships between individuals thus providing a holistic perspective on group performance. However, most of the studies that use SNA in Computer-Supported Collaborative Learning scenarios derive their data from server log files, on the assumption that this data source reflects the way people really interacted online. This study, the purpose of which is to better understand the communication flows that really occur among users, challenges these assumptions through an experimental study that makes a comparison between the structural method normally used to detect the posting addressees and an approach enriched with semantic coding. Results show that this new coding schema, if compared with traditional structural coding, detected a greater number of addressees thus allowing a greater number of postings to be included in an SNA adjacency matrix.

Keywords: Computer-Supported Collaborative Learning; Asynchronous Learning environment; Interaction patterns; Social Network Analysis.
INTRODUCTION

The analysis of the communication flows that occur in educational web forums may significantly help researchers and tutors to understand the nature and quality of learning processes (e.g., Wegerif 2006; Stahl 2006). Written discourse is central in web forums and helps to determine the educational and socio-psychological dimensions of learning in these settings (Feenberg 1989; Lapadat 2002). Thus the tracking and analysis of interaction patterns may serve many purposes: to evaluate the quality of the learning experience through evaluation of the instructional design measures adopted; to monitor students’ performance so as to provide them with appropriate support; to assess individual learning processes and carry out formative and summative evaluation of students’ performances (Pozzi et al. 2007).

Different methods have been adopted so far to assess participation and to understand different qualities of interaction: Content Analysis, that focuses on classifications of text features into categories and indicators within the text of the single message (De Wever et al. 2006); Exchange Structure Analysis, that analyses moves and turns in dialogue by discourse function and role relationships (Kneser et al. 2001); Speech Act Analysis, focalized on messages according to their speech act composition (Nastri et al. 2006); Argumentation Analysis, based on evidence and warrant for claims and on logical connections (Weinberger & Fischer 2006). Beyond these approaches, other methods take into account the structural or syntactical form of pattern interactions. Among the latter, Social Network Analysis (SNA) focuses on the interrelationships between individuals rather than assuming an individualistic perspective. Through the concepts of nodes and links that constitute the network of participants, SNA displays relationships as graphs, with nodes representing individuals and edges representing interaction types (Scott 1991; Wasserman & Faust 1997).

In the context of Computer-Supported Collaborative Learning (CSCL) environments, SNA has been revealing itself to be a powerful methodological approach through which to measure
the levels and patterns of interaction in online learning environments, especially when it is used in combination with other methods, such as content analysis and quantitative techniques (Daradoumis et al. 2004; Zhu 2006; Martinez et al. 2006; De Laat et al. 2007) or multidimensional scaling and correspondence analysis (Hurme et al. 2006). Approaches aimed at visualizing participation have proved to be successful in fostering and improving participation during CSCL learning (Janssen et al. 2007).

The wide set of indices provided by the SNA technique has been tested and used in a variety of studies. For instance, Cho et al. (2002) investigated the role that key communicators occupy in a social network and how they influence others’ behaviours in the form of social navigation. The authors used centrality measures and Bonacich’s power index to show how learners occupy central or peripheral positions in the e-mail and discussion board networks analysed. In a case study by Reffay & Chanier (2003), SNA provided the possibility, through cohesion analysis of a learning group, to display global properties both at individual and at group level; according to the authors, SNA may efficiently assist the tutor in tracking the collaboration within the group. In another study by Aviv et al. (2005), it emerged that in a structured forum the knowledge construction process developed cohesive cliques and reached a very high phase of critical thinking, whereas in the non-structured discussion areas the knowledge construction process reached a low phase of cognitive activity and most of the students took on the passive role of teacher-followers.

However, current approaches aimed at acquiring data for SNA fail to fully capture the nature of interaction that actually occurs in web forums. In the following, after a presentation of the literature in the field of SNA applied to CSCL environments and of the limits highlighted by coding methods, a coding procedure to analyse educational web forums aiming to disclose the complexity of interaction patterns will be presented.
RATIONALE OF THE STUDY

The nature of interaction in asynchronous learning environments

The public nature of web forums may significantly affect the way we treat the concepts of interaction and communication. According to a classic definition, interactivity in Computer Mediated Communication (CMC) is interpreted as a three-step process, where 1) someone sends a message, 2) someone else replies to this message, 3) the original sender replies in his/her turn (Henri 1992). In web forums interaction is a much more complex phenomenon since messages are rarely addressed to, received by, and responded to by a single person, as they are usually directed to the group of participants as a whole and may be responded to by many people (McDonald & Gibson 1998).

There is no full agreement on the definition of interaction. Martinez et al. (2003) provide a generic view of interaction without restricting it to a particular source of data or analytical perspective: “an action that affects or can affect the collaborative process. The main requirement for an action to be considered a possible interaction is that the action itself or its effect can be perceived by at least one member of the group distinct from the one that performed the action” (p. 380). Their proposal distinguishes between direct interactions with a source and one or more receivers; indirect interactions, mediated by a shared object; and finally, participation-oriented interactions, in situations where no receptor has been identified.

Beuchot & Bullen (2005), for their part, in a study aiming to analyze how students in online courses interact and how groups developed in computer-mediated environments, distinguish between two kinds of communication in CMC environments: 1) interpersonal communication, that is interaction aimed at the creation of relationships among participants; 2) impersonal communication, that is task-oriented communication in which information is offered or requested. In its turn, interaction is categorized according to three dimensions: active, when it does not refer to other messages; reactive, when it refers implicitly or explicitly to a message
posted immediately before the reactive message; and interactive, when messages relate to each other in a sequence and later messages recount the relatedness of earlier messages. Whereas reactivity can be assimilated to one-way feedback, interactivity requires a thread of messages or a chain of interrelated messages. According to this study, while through time initial levels of active content tend to decrease and levels of reactive content tend to increase, overall interactive content in a graduate students’ online forum increases from initial levels over the duration of the study.

Most of the research conducted in the field of SNA and CSCL analysed the social dimension of groups and communities, including interaction patterns that characterize the first phases of a group formation and development. Interaction has been identified as a critical and indispensable component to include when creating an online learning community. According to Jung et al. (2002), interpersonal issues become fundamental especially in the first stages of a group development as the creation of social connectedness reduces feelings of isolation, thus enhancing building of trust processes and the sense of warmth and belonging, both necessary to establish the ideal climate for collaboration and learning.

McDonald (1998) found that interpersonal issues remained prominent through the asynchronous computer conferencing course. Results of her study showed that participants, initially directing at the group as a whole, increasingly related to other group members as individuals, not just as a group entity, in an interactive pattern of communication that suggests collaboration and construction of knowledge.

Finally, Waltonen-Moore et al. (2006) explored how interactive patterns occurred over time in the threaded discussion responses of graduate students participating in an online course. They found that at initial stages Branching (i.e., the direct reply of one person to another’s message) is non existent or very poor and that the postings tend to be largely independent. As the group development progresses, sparse and limited interaction through the form of
branching emerges until the two last stages (Involvement and Inquiry), when branching is frequent and interactive and the most voluminous to date.

**SNA and participation networks in CSCL environments**

Most of the studies that use indices and methods of SNA take server log files or web tool databases as input data and perform specialised SNA by means of software tools, such as Ucinet and NetMiner. With these tools, that treat the relationships obtained in a matrix format, it is possible to apply several techniques to analyze social networks, e.g., find interaction patterns, evaluate the role of an individual in an organization, etc. (van der Aalst et al. 2005).

Indeed, server programs record significant events in log files, which is the most widespread method of collecting data. Tracking of raw data and subsequent analysis of log files are often used to gain on-the-fly information about the participative dimension. Examples of raw data are: the number and duration of sessions per participant; the number of messages sent (sorted according to participant, area/conference, date, etc.); the size of the messages sent by each participant; the number of messages read (i.e., opened) by each participant; the number of threads activated by students and tutors; the number of documents produced; the number of chats in which participants took part, etc. (Pozzi et al. 2007).

Among studies that emphasize the role of log files, Nurmela et al. (1999) look at them as the basis for data collection through which to elaborate perspectives on how the collaboration has worked. They state that log files can be used effectively to evaluate relevant behaviours in a CSCL environment as they are a quick way to select and organize large amounts of information. In another study, Martinez et al. (2003) developed a tool called SAMSA (System for Adjacency Matrix and Sociogram-based Analysis), that processes interaction from log files. This tool takes as input the interaction data represented in XML syntax and builds a sociomatrix representing the social network and computes the indices chosen to describe the
participatory patterns of the network. It also shows the sociogram based on multidimensional scaling statistics and allows for the visualization of the actors’ attributes. A similar tool called SLM (Synergeia Log Miner) was developed by Calvani et al. (2005) to extrapolate SNA adjacency matrices from the log file of Synergeia Platform.

What characterizes these methodological approaches is the assumption that the kind of data source recorded in log files reflects the way people have really interacted. However, the meaning of the information contained in these files depends on the type of forum used. For instance, posting a message in a sequential forum is very different from posting a message in a threaded-discussion based forum where all messages are arranged in branches of trees alternating messages with their responses (Hewitt 2001). In a sequential web forum users read all the messages posted and then send their message or their answer to the forum. With this interface visualization the reader answers the solicitations coming from the previous messages, so it is very difficult to know if answers are individually or collectively directed. Another difficulty arising with this type of visualization regards the notion of conversation. Since discussion takes the form of an interaction one-to-many, the sender speaks to the group of participants in a sort of monologue that requires being read, but not necessarily being replied to. In threaded-discussion platforms, such as Knowledge Forum, Synergeia or Moodle\(^1\), messages are hierarchically organized into threads or main headings, and subsequent responses are displayed in subheadings. Users can read only one message at a time and their answer is a reply to that precise message. This structural organization induces an interaction that is more similar to face-to-face conversation, in which the addressee replies to a specific speaker and each reply is addressed to a specific message. Despite this organization, the exchange structure is very complex and we cannot always rely on message threading to determine which messages relate to others (Guldberg & Pilkington, 2006).

---

\(^1\) Moodle platform allows participants to choose four different types of message visualization: two sequential visualizations and two hierarchized visualizations.
The outlined comparison between sequential and hierarchized visualization interface of messages is a critical aspect for SNA. On one hand, it is not clear how to consider the messages sent to the whole group of participants in a web forum with sequential interface visualization: are these messages sent to all the participants, are they sent only to those who read them or only to the participants that reply to them? These three different solutions generate three different SNA adjacency matrices and interpretations of the interactive structure of a group. On the other hand, different visualizations imply different tracking of the user’s action. In a sequential web forum there is no chance to identify the message(s) actually read within the entire message list. In contrast, in a hierarchized visualization interface the tracking system can differentiate between individual messages read or not by a specific user. This is a crucial aspect in SNA, especially in the cases in which the chosen solution for data matrix is to consider each message as if it was addressed only to its readers.

Another critical aspect of a web forum for applying SNA is the definition of interaction and the kind of interactions allowed and facilitated by the interface visualisation. This is often related to the user expertise in relation to the visualization interface of messages. Observing different web forums, we note that a lot of people, especially beginners, face difficulties in “properly” using the communication functions provided by the environment. For instance, in the hierarchized visualization of messages, it is common to find messages posted in a wrong thread, or reply messages posted as if they were thread starters (Wiley 2004). This may be due to the absence of turn-taking procedures in place that encourages the proliferation of threads that sometimes compete for the attention of group members (Brooks & Jeong 2006). All these problems affect the way interaction is recorded and stored in the system tracking and, consequently, the relations visualized in an SNA sociogram and the relative structural indices. On these premises and considerations, the objective of this study is to provide answers to the problems usually faced when coding interaction patterns in web forums. The proposed
solution is a coding schema whose aim is to identify the addressees of messages through a qualitative analysis of the postings exchanged. According to the literature on content analysis (Bauer 2000; Job-Sluder & Barab 2004; Herring 2004), we will call this procedure “semantic coding”, since it requires a process of interpretation of what the author of messages meant to say. The results of its application are compared with those that typically emerge from SNA matrices, that are automatically built on the basis of tracking and log file data by machines or, in absence of an automatic tool, by a human coder who simulates the machine process by referring to the thread structure. We will call this latter procedure “structural coding”.

**METHOD**

*Context and participants*

The research context is a twelve-week post-graduate course on Educational Technology addressed to student teachers of the Postgraduate School for Secondary School-teaching (SSIS) of the University of Genoa, in the 2004/2005 academic year. The course adopted the blended approach combining five face-to-face lessons and twelve weeks of online activities delivered at a distance via a computer conferencing system (Centrinity FirstClass). Online activities relied on text-based interaction, mostly asynchronous. The course plan and structure were characterized by experiential learning and the activities required collaboration among participants, peer-reviewing activities and cooperative production of artefacts. Online instructional activities were arranged in a number of phases and the environment was organized in different areas of interaction, comprising some familiarization and socialization facilities, and a meta-cognitive area to encourage reflection. The interactions were based on conversations that took place in the online conferences aimed at making students debate specific topics in order to collaboratively achieve the given tasks. For all activities the whole cohort of students was arranged in subgroups under the guidance of experienced tutors.
The group of participants was composed of 95 student teachers and 7 tutors. Among the students 77 were female and 18 were male. Their mean age was 31.3 years (SD=5.3), from 24 to 45 years old. They belonged to different streams according to their graduate degrees: 29 (30.5%) to Humanities and Arts, 39 (41.1%) to Maths and Sciences, 11 (11.6%) to Human Science and 16 (16.8%) to Foreign Languages. A survey carried out before course starting revealed that 92.0% of them had a personal computer at their disposal, 72.3% had an Internet home connection and 77.3% accessed web resources regularly. However, for most of them (89.8%) this was the first exposure to CMC in formal learning.

Data

The pattern of postings chosen for the analysis concerned the postings exchanged in the familiarization phase (week 1-3) and in the final activity (week 12). In both activities students were split into seven subgroups. The familiarization activity was based on a metaphor: the course was described as a sea-voyage in which each participant-sailor was supposed to choose a metaphorical boat on which to run the experience. The group of participants who chose the same boat-conference had to collaborate on the realization of an assigned task (i.e., collaborate for three weeks to decide, through negotiation, a name for their boat, a motto and a symbol). The rationale for this activity resided in the importance given to help participants to overcome their sense of estrangement and to make them feel at ease with the communication platform and with the pursuing of collaborative tasks through asynchronous written communication (Delfino et al. 2007). During the Final activity, in order to properly conclude their navigation, students were asked to choose a boat again.

The reasons for choosing these data derive from two considerations: 1) most of the students were CMC beginner users in educational web forums and the first weeks of course would be devoted to familiarization with the thread-based conferencing system in use; 2) after several
weeks of use, learners would be expected to be more confident with the communication protocol and to use the technical features of the system properly. In this way initial inexperience would be balanced with final greater expertise in the use of the system.

The total number of postings examined was 1302: 961 (73.8%) of them were produced during the familiarization phase and 341 (26.2%) during the final activity.

**Approach and procedure**

Systems that automatically build matrices for SNA analysis by use of tracking and log file data normally follow a procedure according to which any posting, independently of its role in a hierarchized tree (i.e., first posting of the thread, isolated posting or reply), is coded as a posting with or without reply (see Table 1).

In this way, a number of interaction patterns, mainly those that take into account the communication flows between senders and addressees of communication, that are not detectable and not eligible for the adjacency matrix\(^2\), are neglected.

[Insert Table 1 here]

In our study a content analysis approach was used in combination with exogenous data (i.e., properties of postings, such as name of sender, name of the area, date and time of dispatch, etc.) to detect senders and repliers/responders to the postings, so that interaction patterns usually neglected by traditional structural coding might also be included in an adjacency matrix for SNA purposes.

\(^2\) An adjacency matrix is a square matrix in which each row represents the sender of messages and each column the receiver. The cross between row and column represents the number of messages sent from sender to receiver.
The aim was to identify the communication patterns between participants in a web forum of subscribed and known participants, (1) differentiating when a posting was addressed to the whole group or to a specific participant (or subgroup of participants), and (2) detecting the presence of postings written in order to reply to a previous message, either by the use of the Reply function or by the creation of a new message.

As there was no pre-defined schema to apply for posting codification, an iterative and inductive procedure was followed. Each step in the validation process was carried out by two researchers, who were also designers and tutors of the course, as well as authors of this paper. The steps taken in order to ensure inter-rater reliability helped in the definition of a coding schema to analyze the postings’ addressees. Inter-rater reliability was validated using Krippendorff’s alpha (Krippendorff 2003).

After a preliminary coding conducted together, in the first phase the two coders blindly examined autonomously a subset of 341 postings (in particular those of the final activity, corresponding to 26.2% of the total) on the basis of the following criteria: 1) name of sender; 2a) name of explicit addressee(s) declared within the message or 2b) name of addressee(s) inferable from it; 3) name of replier(s) to the posting.

During this phase, a number of “irregular” behaviours were detected and treated in a systematic way. For instance, it emerged that many participants replied to previous postings not using the Reply function provided by the platform, but creating a new message. For this reason it was decided to differentiate the role of repliers (i.e., those who used the Reply function provided by the platform) from the one of responders (i.e., those who replied to the posting creating a new message). To do so, date and time of dispatch were useful features to track isolated messages and virtually insert them in a semantic thread of discussion. Hence, in order to capture the interaction component activated by the author who initiates the thread of a dialogue, the traditional structural coding (henceforth, TSC) procedure was strengthened,
considering also the directionality of communication. We define the part of the analysis that can be automated as “potentiated structural analysis”, as it includes all the communication flows of an interactive chain, both from the sender and the replier side. The resulting approach is a semantic coding combined with the potentiated structural analysis (henceforth, S/SC, i.e., Structural/Semantic Coding), as shown in Table 2. In this phase, the measure of agreement between the coders turned out to be $\alpha=0.67$.

In a second phase, the coders independently analyzed all the postings, following the coding procedure previously negotiated. The inter-rater reliability coefficient of agreement reached was $\alpha=0.91$.

[Insert Table 2 here]

To better understand the coding procedure, a few excerpts will illustrate the analysis conducted by the coders. The following examples show postings sent to the whole group (Excerpt 1) or addressed to specific participants (Excerpts 2-6).

All the excerpts belong to postings sent in the Familiarization conferences. They are introduced by a short comment aiming to illustrate the different interpretations that would be attributed to interaction patterns with a TSC procedure or with a S/SC procedure.

**Excerpts 1 (a2):** Luca (author S) sends a posting addressed to the whole group of participants. Nicola (replier R) replies to the posting. According to our schema there would be a double communicative flow ($S \rightarrow R; R \rightarrow S$), while a TSC would recognize only the communication flow $R \rightarrow S$.

*From: Luca - Date: 18 March 2005 – 2.53 p.m. - Subject: presentation*
Text: I am not capable of writing anything intelligent in this area!!!!!!!!!

From: Nicola - Date: 20 March 2005 – 10.58 a.m. - Subject: Re: presentation
Text: go on with you... we are in the same boat actually what I’m writing doesn’t make much sense!

Excerpt 2 (a2): Claudio (author S) sends a posting to express intolerance towards spices. Camilla (responder R), responds to his message creating a new posting. A TSC would ignore the posting because written without using the proper reply function; according to our schema this communication flow must be treated as if the respond was a reply (R → S; S → R).

From: Claudio - Date: 18 March 2005 – 10.23 a.m. - Subject: Fish, shellfish and ... jellyfish!
Text: Hi everybody! You could make a start on lunch, but... no herbs, please, I don’t like them! Claudio

From: Camilla - Date: 18 March 2005 – 11.23 a.m. - Subject: message to Claudio
Text: hello Claudio, I’m worried by your intolerance of herbs: how can you conceive that fish is cooked without them? I think you should get accustomed to herbs! Bye for now!
Excerpt 3: Gabriella (author S) sends a posting. Floriana uses the reply function, but the reply posting is empty. A TSC would consider the posting as part of the threaded discussion (R → S), recognizing as a Reply a posting sent by mistake through the reply function. According to our schema, the posting was ignored.

From: Gabriella - Date: 16 March 2005 – 7.50 p.m. - Subject: Unbelievable!
Text: I can’t believe it: has anybody chosen the motorboat? It is the quickest boat to reach your destination, it is fast, it doesn’t make you waste your time and... it lets you get tanned without suffocating... and yet nobody has chosen it! Humph!

From: Floriana - Date: 16 March 2005 – 9.10 p.m. - Subject: Re: Unbelievable!
Text: [empty]

Excerpt 4 (b1): Stefania (author S) sends a posting addressed also to Tati (addressee A is explicitly mentioned). Nobody, Tati included, will reply/respond to this posting. A TSC would ignore the posting because with no reply; according to our schema the communication flow that involves sender and addressee (S → A) must be considered.

From: Stefania - Date: 16 March 2005 – 9.44 p.m. - Subject: Hi!
Text: hi, I’ve just arrived on the fishing-boat. I suffer from seasickness and I hate the smell of fish. Tati, where are you?

3 The reply function was misused by the participant who sent the posting by mistake and wasn’t able to delete it (the delete message function was inhibited).
**Excerpts 5 (b2):** Roberto (author S), to suggest a name for the boat, sends a posting addressed to Giulio (addressee A), who does not reply. Paola (replier R), instead, replies to Roberto. According to our schema there would be a complex communicative flow \((S \rightarrow A; S \rightarrow R; R \rightarrow S)\), while a TSC would recognize only the communication flow between replier and sender \((R \rightarrow S)\).

*From: Roberto - Date: 29 March 2005 – 2.31 p.m. - Subject: vote updating*

*Text:* My vote for the name of the boat is still missing. Giulio, could you please add it to the list “Caronte”?

*From: Paola - Date: 30 March 2005 – 6.54 p.m. - Subject: Re: vote updating*

*Text:* CARONTE CARONTE!!!!!! Let’s hope it wins Caronte!!!

**Excerpts 6 (b3):** Valeria (author S) sends a posting, as specified in the Subject field, addressed to Giovanna (A), who replies to the message. In this case the addressee A, continuing the thread, is a replier as well. According to our schema there would be a double communicative flow \((S \rightarrow A; A \rightarrow S)\), while a TSC would recognize only the communication flow between replier and sender \((R \rightarrow S)\).

*From: Valeria - Date: 3 April 2005 – 00.12 p.m. - Subject: message to Giovanna*

*Text:* Hi Gio! Let’s meet at 7.45 at the station… btw, I’ll phone you! Vale

*From: Giovanna - Date: 3 April 2005 – 5.51 p.m. - Subject: Re: message to Giovanna*

*Text:* Thanks a lot for the message and for the SMS as well! See you tomorrow!
RESULTS

The coding procedure based on S/SC revealed its potential to detect flows of communication that would be neglected through a TSC. In this section we will especially focus on two key results of the study, in terms of the advantages achieved by our coding procedure: the greater number of postings included in the adjacency matrix, and the greater number of identified addressees, both implicit and explicit.

**Number of postings included in the adjacency matrix**

From the described coding analysis, it occurred that many of the postings that did not receive any reply/response could be virtually introduced in the group of postings that took part of the communication flow as it was possible to detect a semantic relation. It was also found that some postings within a thread should not be included in the matrix because their authors were not replying to previous postings: they actually used the Reply function to start a new thread of conversation.

As shown in Table 3, only a total amount of 530 postings (40.7%) were included in the adjacency matrix according to the TSC approach, whereas 956 (73.4%) were the postings with addressees in the S/SC approach proposed. This means that S/SC allowed us to include in the adjacency matrix 426 postings (32.7% of the total amount of postings) that otherwise would not have been considered.

[Insert Table 3 here]
**Number of posting addressees**

TSC procedures analyzing the threaded interactions inevitably take into account the surface of message exchanged, in terms of replies to postings starting the threads of communication, thus completely ignoring the complexity of interpersonal communication that emerges reading the posting contents.

In TSC, replies to a previous posting are treated as if the repliers were the effective addressees: in this way the communication flow between sender and repliers is equivalent to the total number of replies. Through S/SC, it was possible to recover the real transition from sender to addressee, through a great number of addressees, most of them defined internally within the text of postings.

Comparing the data in relation to the first and second phase of activity, a more adequate representation of the richness of interaction patterns emerges thanks to S/SC. On one hand, application of the Mann-Whitney test reveals that, according to TSC, the thread average length in the first activity (1.97, SD=2.20) is not significantly different from the final activity (2.18, SD=1.75): U=5605.5, p>0.2. On the other, the average number of addressees revealed by the S/SC is significantly different in the first (1.44, SD=1.09) and in the final activity (1.92, SD=2.04): U=76112.0, p<0.002. This means that the proposed method is able to capture better the phenomenon according to which participants, initially directing at the group as a whole, increasingly relate to other group members as individuals, not just as a group entity, thus increasing the number of specific addressees (see McDonald, 1998).

Moreover, to better appreciate the differences between TSC and S/SC, it is useful to observe the diagrammatic representations of the social networks produced by the two coding methods. Figure 1, in particular, shows the communication flows identified in a specific area of interaction, the conference-boat Caravel during the Familiarization phase. Sociogram based on the TSC is indicated on the left, whereas the S/SC one is on the right.
One can observe that with TSC the communication structure of the group seems to be not very dense and presents a lot of subjects in weak positions (i.e., with little connections). On the contrary, S/SC emphasizes a higher aggregation and more connections among the group members.

CONCLUSIONS AND FUTURE DIRECTIONS

Results of the study show that if log files represent unquestionable data, the way data are currently sought and represented through TSC is not sufficient to reflect properly either the interactions between users or their patterns of communication. A qualitative approach of coding, enriched with the potentiated structural coding we envisioned in our study, is able to detect a greater number of postings that can be included in an adjacency matrix for SNA purposes, since it is able to capture the complexity of communication patterns. As emerged in this study, our approach allowed us to include 32.7% of the total amount of postings that otherwise would not have been considered. Indeed, communication flows between participants in a web forum are a complex matter. The nature of interaction itself, where most messages are directed to the group of participants as a whole rather than to specific addressees, tends to evolve over time (Waltonen-Moore et al. 2006). In our study both the complexity and the evolution of communication patterns have been enlightened better by the adopted S/SC, the only method able to reveal how much the number of addressees tend to increase over time. Among approaches that tend to show the complexity of interaction patterns as a means by which to foster and improve participation during CSCL learning, SNA sociograms are helpful during the communication process both for students and tutors. For the former, also previous
research showed that students who have access to participation tools report higher awareness of group process and activities, collaborate differently and perform better than students with no access to these tools (Janssen et al. 2007). On the tutors’ side, tracking and analysis of collaboration, through interaction graphical representation, may serve to monitor students’ performance so as to provide them with appropriate support; to assess individual learning processes and carry out formative and summative evaluation of students’ performances (Pozzi et al. 2007).

Going beyond the prompt outcomes of this study, further development could see implementation of linguistic cues and discourse structure analysis, aiming to find different types of participation according to the possible communicative roles - such as participants more keen on initiating a thread, in reinitiating a new one or in responding, etc. In this way it could be possible to analyse asynchronous turn-takings both on the students’ and on tutors’ side, thus creating the premises to evaluate the quality of exchanged dialogues (Kneser et al. 2001).

However, some limitations of the exploratory nature of the study are to be underlined. Sample sizes and the threaded discussion type of the CMC environment (in our case, Centrinity FirstClass) used for investigation deeply affected the results of the study. The latter should be verified in other conferencing systems that are based on different communication structures and reinforced with reflections on what makes a topic a new topic, as well as on the difficulties in identifying the difference between initiating, reinitiating or responding to a posting.

Moreover, content analysis is a very time consuming approach and this aspect must be taken into account when deciding to carry out a qualitative analysis of posting addressees. The benefits must be counterbalanced with its labour-intensive nature and efforts at applying text classification technology to the corpus of data should be made (Rosé et al. in press).
Another limit presented by the qualitative approach is that detecting posting addressees often involves an interpretation of the senders’ communicative intentions and requires searching for latent (cross-)references between participants. For this reasons, in order to establish reliable interpretations not only a plurality of coders is needed, but also a combination of external coders with coders that have a deep knowledge of the context to be analyzed, in particular who know the silent and implicit communication that sometimes occurs between participants (e.g., use of nicknames, indirect personal references, etc.).

However, a number of measures could be adopted to reduce uncertainty of coding interpretation and heavy workload. For instance, more accurate and precise structural coding tools, able to signal for instance the presence of blank postings (see Excerpt 3), and to emphasize the communication flows, showing the directionality of communication (from the sender to the replier and vice versa). Furthermore, message labelling and diachronic graphical representations of the interaction flows may allow one to specify and identify receivers of messages in a unique way, both when the latter are more than one and when they differ from the original sender. This may also have the advantage of enabling participants themselves to see the overall structure of senders/receivers so as to be aware of the participants’ network they are contributing to build. Message constraints and labels can indeed be used to facilitate analysis of the sequential nature of exchanged messages and enable a process-oriented approach to studying student interaction in online discussions (see Brooks & Jeong 2006; Jeong & Joung 2007).

Finally, to reinforce the detection of posting addressees and to gain a better understanding of the communication flows, coding of communication patterns should also take into account data regarding the silent participation expressed, for instance, by data related to posting reading (Hewitt 2003). These may enrich our comprehension of communication flows in web forums.
REFERENCES


<table>
<thead>
<tr>
<th>Situation</th>
<th>Procedure</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Sender S posts a message in the web forum</td>
<td>a) if nobody replies to the posting</td>
<td>→ the posting is ignored</td>
</tr>
<tr>
<td></td>
<td>b) if a participant R replies to the posting</td>
<td>→ the Reply posting is treated as directed to the Sender (R → S)</td>
</tr>
</tbody>
</table>

Table 1: Traditional structural coding procedure (TSC).
<table>
<thead>
<tr>
<th>Situation</th>
<th>Procedure</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1st step</strong></td>
<td>a) if the posting is addressed to the whole group <strong>and</strong> a participant R replies to the posting</td>
<td>→ the posting is ignored</td>
</tr>
<tr>
<td></td>
<td>b) if the posting is addressed to a specific participant or a subgroup (either explicitly mentioned or inferable) <strong>and</strong> b1) nobody replies/responds to the posting</td>
<td>→ the posting is treated as directed to the addressee (S → A)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b2) a participant R replies/responds to the posting</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b3) the addressee A replies/responds to the posting</td>
</tr>
</tbody>
</table>

Table 2: Coding procedure based on semantic analysis combined with potentiated structural analysis (S/SC).
<table>
<thead>
<tr>
<th>Activity</th>
<th>Postings included by TSC</th>
<th>Postings excluded by TSC</th>
<th>Postings with addressee(s) in S/SC</th>
<th>Postings without addressee(s) in S/SC</th>
<th>Postings recovered by S/SC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Familiarization</td>
<td>397 (41.4%)</td>
<td>563 (58.6%)</td>
<td>721 (75.0%)</td>
<td>240 (25.0%)</td>
<td>323 (33.6%)</td>
</tr>
<tr>
<td>Final activity</td>
<td>133 (38.9%)</td>
<td>209 (61.1%)</td>
<td>235 (68.9%)</td>
<td>106 (31.1%)</td>
<td>103 (30.2%)</td>
</tr>
<tr>
<td>Total</td>
<td>530* (40.7%)</td>
<td>772* (59.3%)</td>
<td>956* (73.4%)</td>
<td>346* (26.6%)</td>
<td>426 (32.7%)</td>
</tr>
</tbody>
</table>

* Chi-square = 284.446, df=1, p<0.001

Table 3: Number and percentage of postings included/excluded by TSC and by S/SC, and number of postings recovered by the latter.
Figure 1: TSC vs. S/SC coding: diagrammatic representations of the same social networks coded in the two ways.