

# Mining Learning Groups' Activities in Forum-type Tools

**Pablo Reyes**

LIUM-CNRS FRE 2730  
Université du Maine, France  
Pablo.Reyes@lium.univ-lemans.fr

**Pierre Tchounikine**

LIUM-CNRS FRE 2730  
Université du Maine, France  
Pierre.Tchounikine@lium.univ-lemans.fr

**Abstract.** Mining data produced by students involved in communication through forum-like tools can help revealing aspects of their communication. In this paper we propose an approach to the construction of models to highlight structural properties of learning groups based on a relational perspective (analysis of chains of references) and the use of Social Network Analysis techniques. These models can be useful both for the tutor and the participants. We begin by introducing the overall approach, then we describe how the models are constructed and finally we present preliminary results from the integration of these ideas in a forum-type tool.

**Keywords:** Link analysis, peer-to-peer support, social network analysis.

## INTRODUCTION

Network technologies have enabled web-learning activities such as learning groups and e-communities that can take place in e-learning platforms. Traditionally, collective activities of these groups take place in what we call a *forum-type tool* (FTT). The FTT describes a mainly text-based and asynchronous electronic conferencing system that makes use of a hierarchical data structure of enchainned messages, called *threads*.

In this paper we are interested on supporting collective activities that take place in learning groups based on a relational model of messages exchanged among participants in a FTT.

We model the message exchanges as a graph where the vertices are the participants of a group and the links are the exchanges among them. This is a basic model broadly used for modeling relationships among users from a social networks analysis point of view. We exploit the group's link structure to contribute to a comprehensive understanding of the group activities.

In this article we propose the use of two models to gather information about group activity from a relational perspective. Each of these models corresponds to a different granularity of the analysis. The first model denotes properties of the group as a whole; the second denotes properties of the individuals in relation with the group to which they belong. Our algorithms construct indicators that allow characterizing the collaboration process, which can be useful for both tutors and students.

The article is organized as follows. First, we describe the idea of mining group activities and their use in a CAL context. Then, we propose and describe two models to gather different characteristics of a group. Finally, we present preliminary results from an empirical study that illustrates the use of our models.

## MINING GROUP ACTIVITIES

### Mining group activities in a learning context

Mining group activities is an active line of research. Current research is mainly focused on the construction of indicators of collaborative group's activities aiming at a theoretical or ethnographic analysis of the group (e.g., (Reffay & Chanier, 2002), (Martínez et al., 2002) or (Butts, 2001)) and other social networks works). Nevertheless, these analyses are rarely used to support online collective learning activities.

The use of mining strategies can be an important element in educational contexts. Mining group activities in a learning context provides quantifiable profiles of the groups, which allows to (1) evaluate the collaborative activity that the participants carry out, (2) analyze the link structure of the group, (3) compare the collaborative performance among different groups and (4) predict behaviors, discover link patterns (Getoor, 2003) and collaboration trends. This knowledge can be used and applied directly to support the collaborative activities. In this sense, link models can be an element that helps (1) the tutor on his tasks of collaboration management and that scaffolds the collaborative learning among the participants without needing an extensive review of each group's interactions. (2) the participants of the group (the students), who use link analysis to discover the

structural features or activities of their group (what has been termed structural awareness (Gutwin, Greenberg, & Roseman, 1996)).

Link analysis provides a new role for tutors in collaborative environments. Instead of their traditional role in the "transfer of information", the role of tutors is shifted to that of establishing the appropriate conditions to allow the students to get connected to the group (the set of relations) through participation (e.g., as part of a community of practice) in the service of an intention (Barab et al., 1999). The structural models that we propose make salient certain profiles of the groups and their participants which can help to be aware of the group activities and can help orienting the pedagogical strategies. Structural models can help participants of a group to create macro-micro links and facilitate peer-to-peer learning. The macro-micro link is a sociological concept that establishes the theoretical foundations for the influence of interaction structure of a community (macro level) with the local interactions among the participants (micro level) (e.g., (Bourdieu, 1988)). The concept of macro-micro link allows us to focus on the interdependency of the structural regularities of the group with the activities of the participants. Indeed, several learning theories emphasize the influence of social interactions on the individual learning. In the perspective of the communities of practice, for example, learning changes "(...) from the individual as learner to learning as participation in the social world" (Lave & Wenger, 1991).

Moreover, the social constructivist point of view highlights also that knowledge is socially constructed by interactions among individuals (McCarthy, 1992). Indeed, the theory of learning of Piaget (Piaget, 1926) states as a fundamental assumption that the interaction among peers while performing tasks facilitates the learning of concepts. By making salient the structure of interactions in a group we allow the participants to be aware of an important element of learning.

### Techniques for mining group activities in a relational perspective

The mathematical tools we use for mining group activities come from the Social Network Analysis (SNA) models. Many methods have been proposed in this field to obtain knowledge about the group from its relational ties. The SNA uses as data the connections among units, which relates them in a system.

We use a graph theory to mining group activities by analyzing the *sociograms* associated to a given group. In FTT a sociogram is a graph where the participants are represented as vertices and the messages that they exchange are represented as the links of the graph. Sociograms can be handled as sociomatrices which are the matricial representation of the graph (more information on the construction of sociograms and sociomatrix can be found in (Wasserman & Faust, 1997)).

## PROPOSED MODELS FOR SOCIAL INTERACTIONS

We model two characteristics of social interactions: the status of participants and group cohesion. These models gather information about the group activity at different granularity levels. The status belongs to a family of models that reveal the role of a given participant in the group. The cohesion belongs to a family of models that reveal structural properties of groups: it provides information about the group and not about the participants of the group.

### Status

#### *Status definition*

In a community, the concept of status represents the "prestige" of a specific participant. The status of a participant is related to his participation in a community as well as the status of the participants which s/he communicates with (Wasserman & Faust, 1997). This concept is not a simple account of the number of user interventions, because it also considers the prestige of his entire neighborhood.

Starting from the participant's status we can find the each participant position in relation to the whole community, and the social structure of this community. Moreover, this indicator can be related to a concept of learning in the communities of practice (Lave & Wenger, 1991), where learning is conceived in terms of participation. In the context of the communities of practice, *learning can be interpreted as an evolution of the status* of a participant from a peripheral participation (low status) towards a central participation (high status) within its community. Through the status indicator, we can measure these evolutions. This model gives participants and tutors an element for comparison among their position in the group and a quantitative measure of their evolution.

#### *Status model*

There are several models to obtain the status of participants in a group: *Betweenness-centrality*, *Closeness-centrality*, *Degree-centrality* and *Eigenvector-centrality* (Wasserman & Faust, 1997). Here, we will concentrate on the *Eigenvector-centrality* model because it is the only status model that establishes the value of a participant status taking into account the other participant's status. Consequently, "an actor's status is increased more by nominations from those who themselves have received many nominations" (Bonacich & Lloyd, 2001).

In spite of the precision of this method to obtain the status values of participants in a group, it makes sense only for the symmetrical sociomatrices ( $A \rightarrow B \equiv B \rightarrow A$ , for example, somebody's brother is a symmetrical

relation). In our case, the matrices of interactions in the FTT are asymmetrical ( $A \rightarrow B \neq B \rightarrow A$ , answering a message is an asymmetric relation). The alpha-centrality model introduced by Bonacich (Bonacich & Lloyd, 2001) presents a generalization of the eigenvectors' model for asymmetrical matrices. Bonacich makes the assumption that the status of a participant depends on two parameters: the external initial status of a participant and the status that is formed starting from the interactions among the participants. By adding the external idea of status to the traditional concept of eigenvectors we obtain equation (1). The complete description of this method is found in (Bonacich & Lloyd, 2001).

$$c = \alpha A^T c + e \quad (1)$$

In this equation  $A$  is a sociomatrix.  $A^T$  is the transposed matrix of  $A$ .  $c$  is a vector of participant's status for the asymmetrical relations. Each component of this eigenvector  $c$  corresponds to the status of each participant. That is, if  $c = (c(v_1), c(v_2), \dots, c(v_n))$ , the status of participant  $i$  is  $c(v_i)$ .  $\alpha$  is a parameter which reflects the relative importance of the external status versus the internal status to determine the final status. We can also interpret  $\alpha$  as the status degree of transference from one person to another.

Among the outstanding characteristics of this indicator we emphasize that the status is not related to participation (e.g., in the star graph (see figure 1), the most central participant has the greatest value of status and s/he has never participated), the status is associated with the impact of the interventions of a specific participant on the activity of the group (participant's visibility).

## Cohesion

### *Cohesion definition*

Cohesion is a concept related to the diffusion of information in a group (Wasserman & Faust, 1997). In a cohesive group the information is extremely likely to be distributed for the entire group. This fact improves the communication, the coordination and the influence within the group. Cohesion gives a measure of how strong the social relations are in order to maintain the group together (Moody & White, 2000).

From this indicator, users can perceive the ability of the group to hold their members. A group with a high value of cohesion is a group that holds social relations among almost all participants. Consequently, the group could face the departure of some of its participants without destroying it.

### *Cohesion model*

There are several models to obtain the degree of cohesion of a group (Wasserman & Faust, 1997). Bock and Husain propose to iteratively build sub-groups so that the proportion between the number of links in the sub-group and links between the sub-groups does not decrease with the addition of new members. Reffay and Chanier (Reffay & Chanier, 2002) obtain the group cohesion by measuring the degree of reciprocal relations that take place in a forum among participants. James Moody and White (Moody & White, 2000) introduce another concept of cohesion, which is defined as the minimal number of participants who, if removed from the group, would disconnect it. This approach led to obtain hierarchically nested groups, where highly cohesive groups are built over less cohesive ones. We seek to make salient this notion, which corresponds to the definition of k-connectivity (a graph is k-connected if there are at least k independent paths connecting every pair of participants in the graph) in the graph theory. This indicator expresses the property of certain groups to hold their members. Yet, this model of cohesion is very sensitive to participants slightly connected in the group. For example, a group with a complete network configuration (see figure 1) with 6 participants have k-connectivity value equal to 5. Nevertheless, if we add another participant to this group with only one link, the k-connectivity value decreases to 1, i.e., a very low cohesion degree for a group that is still highly connected. So, the real group cohesion is hidden. Thus, we modify the original cohesion model (the minimal number of participants who, when removed from the group, disconnect it) in favor of a concept of cohesion as the minimal number of participants who when removed from the group, disconnect it completely. This model provides a more robust measure of cohesion, even for groups with weakly connected participants.

To calculate cohesion, we apply the original algorithm in an iterative way to the groups that remain connected. The summa of the values of the k-connectivity of each iteration will give the final measurement of cohesion. In order to compare the cohesion values for groups of different sizes and structures of participants, these values are normalized. Two normalization methods are necessary: first, in relation to the number of iterations ( $j$ ) executed in the algorithm (i.e., the number of iterations to obtain a group completely disconnected). Second, a normalization regarding the number of participants ( $n$ ). Equation (2) shows the normalized cohesion value:

$$\bar{C}(G_n) = \frac{C(G_n)}{(n-1) * j} \quad (2)$$

## Hypothetical social networks

Figure 1 illustrates four hypothetical social networks with six participants each, with strength of each link equals to 1. The associated status values for each participant ( $c = (c(v_a), c(v_b), \dots, c(v_f))$ ), where  $c(v_i)$  is the status of participant  $i$ . For simplicity, participants in all these social networks, have the same initial status, that is,  $e$  is a vector of '1'.

In the star network we can observe the central position of participant "6" in it. This fact is reflected by his/her high status value (3.5). The same result is obtained in the hierarchical network. Nevertheless, the low value of cohesion of the star graph structure allows us to suppose that it is fragile, given that all interactions pass through participant "6". In the circular network, all participants have the same status values because each of them has the same link number and structure. The highest cohesion value is obtained for a complete graph. This fact represents a group highly robust, with multiple channels of communication among participants. We note that status indicator is sensitive to strength of links, but the cohesion indicator does not, because of the nature of the algorithm to obtain the cohesion.

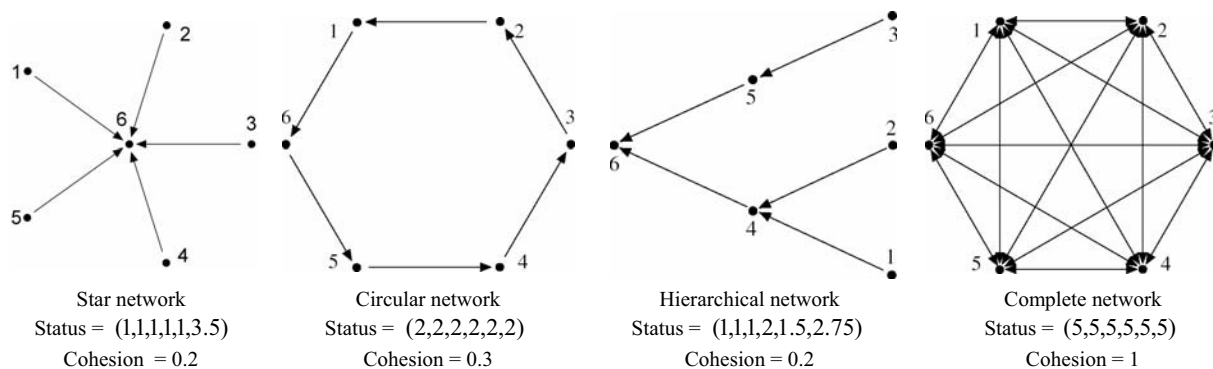


Figure 1. Hypothetical social networks

## INTEGRATION OF COHESION AND STATUS IN A FTT

The results of the proposed structural models are integrated in a FTT called "Mailgroup". In this environment, the participants can maintain a discussion by exchanging messages. Mailgroup has been designed according to the objective of supporting learning conversations taking place in forums (Reyes & Tchounikine, 2003). The support provided by such a FTT tool is enhanced by allowing the participants and the tutor to access at any time, through a menu item, to the values of status and cohesion. Mailgroup shows a single bar representing the group cohesion value (group-level indicator), and individual bars representing the each participant status value (participant-level indicator).

## EMPIRICAL STUDY AND RESULTS

An empirical study was designed in order to collect feedback on the actual characteristics of the group models from the user's perspective. In this study, 15 participants were recruited. The participants were teachers who, during one and a half months, carried out a distance collaborative activity as part of training course on ICT. During the study, they used Mailgroup as medium of communication and discussion (Reyes & Tchounikine, 2003). The goal of the activity was to carry out a collaborative analysis of the integration and utilization of ICTs in education.

In a first stage of experimentation, indicators are showed only to the monitor of this activity in order to test out the validity of SNA models used in these indicators. Yet, the tutor was able to use these indicators to gather information about the groups' activities. Table 1 shows values of cohesion and status obtained in some real conversations (each conversation equals to different threads) that took place in the carried out experience in Mailgroup. The Cohesion is a single value (in percentage) that represents this property of the whole group. The status is a vector where each component represents the status of a single participant. Consequently the vector has many components as participants in a specific conversation.

For example, from the analyses of indicators of conversation number 3, the tutor saw as an outstanding fact the low value of the cohesion indicator. Analyzing the status of participants we can deduce that there is an unbalanced participation since two users carry out almost the whole conversation. Their participation and central position (high status value) indicate that they lead the conversation. A potential absence of these participants can imply the ending of this conversation or a radical change of interaction structures. This way, both indicators indicate to a tutor (or to users) that it is necessary to change their current social structure: based on the indicators

provided by Mailgroup, the tutor might introduce different strategies in order to orient the group towards a more reliable structure, with a more important and balanced implication of the participants in the common task.

Conversations	Cohesion	Status
1	9%	(12, 1, 13, 9, 1, 11, 14, 1)
2	56%	(17, 12, 6, 3, 8)
3	13%	(5, 11, 1, 4, 4, 17, 8)

**Table 1.** Values of cohesion and status.

## CONCLUSIONS

The use of models for mining group activities is an active line of research. In this paper we have presented how we have adapted them for their use for pedagogical purposes: The tutor management and orientation of participant exchanges that take place in a learning group through the tracking of its structural properties.

The pedagogical use of these models is inspired by learning theories and models that emphasize the importance of peer-to-peer interactions and the social structure that they generate. These models can facilitate and even automate the work of tutors in tracking the group activities, helping in focus the attention of the tutor in groups with low levels of cohesion or unbalanced structures of participation.

In this article we have presented two methods for mining group activities based on models for status and cohesion inside a group. The new cohesion model that we have introduced takes into account the general structure of a group, thus overcoming the problem of sensitivity to groups with weakly connected participants. We consider these models as complementary given that they focus on different levels of granularity in the analysis: the group-level in the case of cohesion and the participant-level in the case of status, allowing the analysis of groups in a complementary way.

We showed the results of a test that aimed to corroborate the proposed link models. We obtained that these models describe certain structural properties of a group. Moreover, for the tutor this information can be an element that improves the effectiveness of its pedagogical.

Finally, the models presented in this work are implemented as a part of a peer-to-peer support system: "Structural awareness". The objective of structural awareness is to make salient the structural properties of a group to its participants in order to promote collaborative interactions and allowing tutors the management of learning interactions and tracking collaborative processes.

## REFERENCES

- Barab, S. A., Cherkes-Julkowski, M., Swenson, R., Garrett, S., Shaw, R. E., & Young, M. (1999). Principles of Self Organization: Learning as Participation in Autocatakinetic Systems. *Journal of the Learning Sciences*, 8(3 & 4).
- Bonacich, P., & Lloyd, P. (2001). Eigenvector-like measures of centrality for asymmetric relations. *Social Networks*, 23(3).
- Bourdieu, P. (1988). *La distinción*. Madrid: Taurus.
- Butts, C. T. (2001). The complexity of social networks: theoretical and empirical findings. *Social Networks*, 23(1), 31–71.
- Getoor, L. (2003). Link mining: A new data mining challenge. *ACM SIGKDD Newsletter*, 5(1), 84–89.
- Gutwin, C., Greenberg, S., & Roseman, M. (1996). *Workspace Awareness in Real-Time Distributed Groupware: Framework, Widgets and Evaluation*. HCI 1996.
- Lave, J., & Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge, UK: University Press.
- Martínez, A., Dimitriadis, Y., Rubia, B., Gómez, E., Garachón, I., & Marcos, J. A. (2002). *Studying social aspects of computer-supported collaboration with a mixed evaluation approach*. CSCL, Boulder, Colorado.
- McCarthy, S. J., & McMahon, S. (1992). From convention to invention: Three approaches to peer interaction during writing. In R. Hertz-Lazarowitz & N. Miller (Eds.), *Interaction in cooperative learning groups. The theoretical anatomy of group learning*. Cambridge, UK: Cambridge University Press.
- Moody, J., & White, D. R. (2000). *Social Cohesion and Embeddedness: A hierarchical conception of social groups*. Unpublished manuscript, Ohio State University.
- Piaget, J. (1926). *The language and thought of the child*. New York: Harcourt Brace.
- Reffay, C., & Chanier, T. (2002). *Social Network Analysis Used for Modelling Collaboration in Distance Learning Groups*. ITS 2002, Biarritz.
- Reyes, P., & Tchounikine, P. (2003). *Supporting Emergence Of Threaded Learning Conversations Through Augmenting Interactional And Sequential Coherence*. CSCL.
- Wasserman, S., & Faust, K. (1997). *Social network analysis methods and applications* (Vol. 8). Cambridge: Cambridge University Press.