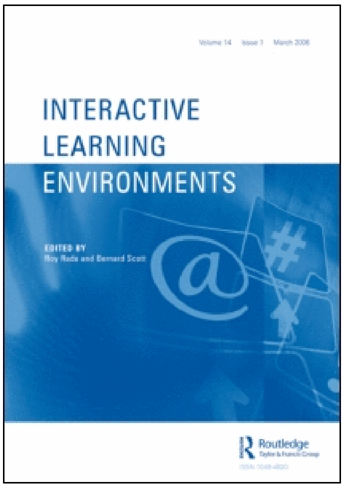


**Interactive Learning Environments**
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t716100701

# Exploring affiliation network models as a collaborative filtering mechanism in e-learning

Daniel Rodríguez [a]; Miguel Ángel Sicilia [a]; Salvador Sánchez-Alonso [a]; Leonardo Lezcano [ab]; Elena García-Barriocanal [a]
[a] Department of Computer Science, University of Alcalá, Alcalá de Henares, Madrid, Spain [b] Department of Computer Science, Open University of Madrid (UDIMA), Collado Villalba, Madrid, Spain

First Published on: 17 August 2009

**To cite this Article** Rodríguez, Daniel, Sicilia, Miguel Ángel, Sánchez-Alonso, Salvador, Lezcano, Leonardo and García-Barriocanal, Elena(2009)'Exploring affiliation network models as a collaborative filtering mechanism in e-learning',Interactive Learning Environments,99999:1,

**To link to this Article: DOI:** 10.1080/10494820903148610

**URL:** http://dx.doi.org/10.1080/10494820903148610

*Interactive Learning Environments*
2009, 1–15, i*First* article



# Exploring affiliation network models as a collaborative filtering mechanism in e-learning

Daniel Rodríguez[a]*, Miguel Ángel Sicilia[a], Salvador Sánchez-Alonso[a], Leonardo Lezcano[a,b] and Elena García-Barriocanal[a]

[a]*Department of Computer Science, University of Alcalá, Alcalá de Henares, Madrid, Spain;* [b]*Department of Computer Science, Open University of Madrid (UDIMA), Collado Villalba, Madrid, Spain*

(*Received 21 December 2008; final version received 22 June 2009*)

The online interaction of learners and tutors in activities with concrete objectives provides a valuable source of data that can be analyzed for different purposes. One of these purposes is the use of the information extracted from that interaction to aid tutors and learners in decision making about either the configuration of further learning activities or the filtering of learning resources. This article explores the use of an affiliation network model for such kind of purposes. Concretely, the use of techniques such as *blockmodeling* – a technique used to derive meaningful patterns of relationships in the network – and the analysis of *m-slices* – a technique helpful to study cohesion in relationships – are explored as tools to decide on the configuration of topics and/or learner groups. In particular, the results of the case study show that such techniques can be used to (i) filter participants for rearranging groups; (ii) rearrange topics of interest; and (iii) dynamically change the structure of a course. The techniques presented can be considered a case of collaborative filtering based on social network structure.

**Keywords:** social network analysis; e-learning; affiliation networks; *blockmodeling*; collaborative filtering

## 1. Introduction

As learning technologies spread out, an increasing number of people are getting involved in online learning activities. Learning designs nowadays emphasize the organization of learning experiences around the concept of *activities* which, in many cases, are shared by more than one individual (Allert, 2004). The interaction of learners takes place through different kinds of *services* such as newsgroups, forums and chats. In consequence, people interact with such services for learning about a particular *objective* (e.g. learning about a topic, acquiring or exercising a given competency), which might be of different granularities. The relationship between these elements, Activities – Objectives – People (AOP), can be used for the empirical analysis of social interaction and for the examination of common interests through technology-enhanced learning.

*Corresponding author. Email: daniel.rodriguezg@uah.es

ISSN 1049-4820 print/ISSN 1744-5191 online

DOI: 10.1080/10494820903148610
http://www.informaworld.com

There are several methods for analyzing AOP interaction. Some authors use qualitative analysis of the discourse (Paz Dennen, 2008; Nussbaum, 2008), while others measure the number of interventions or the users' satisfaction regarding the "social atmosphere" (Kreinjs, Kirschner, Jochems, & van Buuren, 2007). However, these techniques require intensive effort from the tutors to categorize and examine each of the interventions, and are in some cases to some extent subject to the subjectivity in the interpretation of the tutors. As the number of communication events (e.g. messages in online forums) is often large, it seems reasonable to look for indicators of actual social interaction that could be computed with the help of mathematical tools that can be implemented in computer software to gather some measures and indicators automatically, thus helping tutors in decision making.

Social network analysis (SNA) techniques provide a quantitative way to analyze AOP interaction, as demonstrated by Cho, Gay, Davidson, and Ingraffea (2007). SNA can be used for different purposes in e-learning settings, including among others:

(1) *Hypothesis testing* or exploratory studies aimed at finding correlations between social structure elements and different aspects of learning processes. For example, applying statistical test to the learning outcomes obtained using different strategies in group formation.
(2) The *summative assessment* of learners, i.e. the final evaluation of what students have really learned when compared with the expected learning outcomes for different configurations of social structure.
(3) Help tutors to *understand the informal structure of the class* answering questions such as who trusts whom; who offers what kind of expertise and advice; who links sub-groups; who is most central in class activities.
(4) *Reconfiguring the learning contents and/or activities*, for instance, including new activities, changing the future course structure or taking other kind of actions based on the analyzed data.
(5) *Reconfiguring the learning environment* by means of different strategies based on social structure for group formation, or rearranging groups once the course is being delivered.

Here we are mainly concerned with (4) and specially with (5). Concretely, the purpose of *reconfiguring learning environments* includes an attempt to personalize the learning process or to find data that could be used to guide the learner to more interesting or useful activities from the learner's point of view, using social structure as an element that can be adapted. In collaborative learning in general, where learners share knowledge goals, and in Computer Supported Collaborative Learning (CSCL) (Dillenbourg, 1999) in particular, such intelligent group formation functionality is of paramount importance as pointed out by Wessner and Pfister (2001) among other authors. In other words, instead of allowing arbitrary communication and cooperation between learners, systems would be of great help if they were capable of suggesting groups or simply providing hints for intelligent group formation. In a similar vein, although not using SNA, Muehlenbrock (2005) proposed an approach for group support based on both learner profiles and context information.

In relation to the reconfiguration of both *learning environments and activities*, especially in the area of open learner models (Kay, 2001), SNA is just another technique that can be used for adapting educational systems (Bull, Dimitrova, & McCalla, 2007; Dimitrova, McCalla, & Bull, 2007). Therefore, SNA needs to be compared and analyzed on how it can complement the information gather with others. For example, Guo and Greer (2006) have used portfolios for gathering the initial information for learner models.

In SNA, two-mode networks are defined by two sets of social units (constituting the nodes of the network) and the relations which connect the two sets (arcs or edges in the network). One of the most common examples is that of networks whose sets of nodes are people and events, and the arcs connecting these subsets denote the participation (e.g. attendance) of people to each of the events. This particular case of two-mode network was named as "affiliation network" by Wasserman and Faust (1994), as the subsets of actors are grouped according to the participation, i.e. *affiliation*, in a finite set of events. This study focuses on analyzing AOP data for several purposes related to common interest that can be categorized as "reconfiguring learning environments". Concretely, we approach AOP data in the form of an affiliation network, considering that learners' participation in activities can be used to detect groupings of common interest. This kind of analysis can be better applied under certain conditions including: (a) the participation of learners in activities is not mandatory, (b) groups are not enforced, and (c) the activities are clearly directed towards a concrete, recognizable objective. If the former does not hold, the mandatory character of those activities would probably add some bias to participation. If the latter does not hold, the interpretation of interaction data becomes blurred. The use of an affiliation network directly follows the structure of empirical data found in typical AOP interaction. However, the models and the case study described herein represent just an exploration in possible applications that exploit the automated computing of network measures, and other analysis techniques could be considered in further studies.

This article describes the initial model and provides a first case study for the use of affiliation network analysis as a collaborative filtering technique in e-learning settings. In our approach, the events of the affiliation networks are the concrete e-learning activities that include some form of observable service for participation. Newsgroups (forums) and chats are among the most common ones. It is important to highlight that the interpretation of affiliation network is not intended to discover *social circles* (although that is a frequent analysis in affiliation network studies in general), because typical AOP interaction is relatively short in the time span and it is not the norm that the same learners share activities in a continued manner. In contrast, the emphasis is put in finding techniques that allow to find structural patterns of interest, which could be used to filter or rearrange activities.

The rest of the article is structured as follows. Section 2 describes the general model for affiliation-based similarity, and the SNA methods behind it. Section 3 describes a concrete case study. Finally, conclusions and outlook are provided in Section 4.

## 2. An affiliation model for participation in targeted learning activities

It is possible to develop techniques for the analysis of common interests in learner groups through the analysis of their participation in activities organized around communication forums, which are common in e-learning environments. This idea is essentially the same that originated collaborative filtering (Resnick, Iacovou, Sushak, Bergstrom, & Riedl, 1994), but in this case it is not necessary to provide explicit ratings for the items (topics), if we assume that a learner who participates in the discussion of a topic is (to some extent) interested in it. Going further, we can hypothesize that the more the learner contributes to discussing a topic, the more he or she shows interest in the topic, thus allowing for a form of quantitative indicator to be explored. Of course, these assumptions can be questioned by different forms of noise or spurious motivations to participate, but they provide an empirical source of data that follows similar assumptions as other Web-based rating systems such as page ranking (Page, Brin, Motwani, & Winograd, 1998). It should be noted that not all of the uses of these services can be analyzed in this manner. In fact, some preconditions are required for the analysis to be meaningful, as described below.

Affiliation networks can be used to model AOP data, even though the kind of relation usually analyzed in those networks is of a more long-lasting nature than the typical course-based scenario in e-learning. From the analysis of network data modelled that way, it is possible to devise some courses of action, that we label here as different forms of "collaborative filtering". Note that the usual interpretation of collaborative filtering consists of recommendations or ranking of information, but here we adopt a more general position, considering collaborative filtering as any course of action taken on the basis of the analysis of the social network structure.

The classical model initially defined by Resnick et al. (1994) of rating-based collaborative filtering to select articles of interest from bulletin boards (newsgroups) is based on triples (*item, user, rating*) where posts (*item*s) are subjectively evaluated by users. For example, a user A puts "four stars" to a concrete message (or to a book in an online bookstore as can be done in amazon[1]), thus explicitly indicating his or her assessment or preference on that item. Following that modelling approach in our e-learning context, we could map the data model described in the next subsection as the same kind of data structure (*objective*[*service*], *learner, #-communications*), where:

- *objective* [*service*] is a concrete learning or knowledge goal for each service (e.g. thread in a newsgroup).
- *#-communications* represents a numerical account of the participation in the concrete service (e.g. the counting the number of messages of the learner in the particular objective[service])
- The learner is a concrete kind of user.

In other words, we keep the information in triples that have a similar interpretation to those in classical collaborative filtering, For example, let us consider the case of an online course on genetics in which the tutor has opened a discussion forum with several threads inside, covering different learning objectives. Then a triple as (*mitochondrial translation* [*regulation of mitochondrial translation*], *John Smith, 4*) represents the fact that the learner John Smith has contributed four times to that thread. However, the usefulness of using a pure classical collaborative filtering approach is limited by the availability of data, as collaborative filtering only provides meaningful results for large databases of ratings.

### *2.1. The model*

As previously said, an *affiliation network* is a kind of social network in which the actors are divided into two disjoint sets, and ties are only allowed between elements that belong to different sets. So, one affiliation network $N$ can be expressed as a function of three subsets $A$, $E$ and $R$; $N = (A, E, R)$, where, $A$ is defined as the set including a number $a$ of actors who participate in the $e$ events in the subset $E$. Participations of actors in events are the element members of the affiliations set $R$, formed by all (undirected) ties that connect the actors in $A$ with the events in $E$. An example of an affiliation network follows. The events in $E$ can be considered to be classical music concerts, and the relationship of actors (attendants in this case) to concerts will conform relation $R$. In this example, people attending the same concerts (even when they do not know each other) implicitly define sub-groups of preference, which when examined may be related to musical style or other issues.

For our particular case study, each event in the set of events $E$ is any of the planned discussion topics in the particular learning design of the online course or learning event under consideration. The actors are both the tutors and the learners in the course. Each tie $(a_i, e_j)$ in the network represents a distinct, significant message of an individual $a_i$ in a thread of discussion corresponding to the topic $e_j$. The following preconditions are required for the analysis to be meaningful:

(1) Each thread must have a clear topic or objective, distinguishable from the rest (even though they can be related in some known way). This excludes generic "social forums" or "general questions" discussions from the analysis. Although this forces the distribution of topics in concrete "threads" of discussion, if they were merged it could not be possible to distinguish which learners are interested in which topic, or at least, a more complex and not fully automated pre-processing would need to be performed.
(2) Participation in the threads should not be made mandatory to avoid bias. Forcing people to participate in certain topics (or in all of them) will artificially increase the communication in specific topics, but that participation will in some cases not be related to the learner's interest in the topic, thus producing noise, e.g. as when learners participate just to pass the evaluation, but are personally not concerned with the activity.
(3) The time planned for each thread must be similar (ideally, with a non-strict limit). Although hard to implement and control, we hypothesize that the span period of a thread remains active and influences the number of communication to the forums. Allowing similar time constraints for each topic (measured for example in days or weeks the activity is open), the importance given to each "part" will be similar. Of course, this is a strong precondition but in the future works some weighting or pre-processing techniques could be used to weaken this precondition.

The above preconditions are aimed at guaranteeing, to the extent possible, that participation in discussion is a function of interest in the topic, and not a result of any other extrinsic constraint, with the drawback of forcing the design of the course.

Here we will concentrate on filtering, understood as selecting the most appropriate objects of interest from a given set. In e-learning settings, the elements to be filtered may be any of the constituents of the e-learning setting, such as participants (when forming groups), learning objects or activities. In what follows, we will deliberately focus on just a few of the possible cases, although others might be proposed as well.

### 2.2. *Filtering participants*

The network can be used to implement different strategies for the definition of subgroups, including those that are recommended by problem-based learning methods (Oakley, Felder, Brent, & Elhajj, 2004). In general, courses are organized in several modules (or topics) having different assignments for each module. One of the possible applications is that of identifying groups that are close to the common interests of groups of learners. As an example from our case study, SNA tools allowed us to clearly identify two different groups, one interested in the learning tools used during the course and the other group more interested in the theoretical aspects of the course. This can be accomplished in several ways. A straightforward technique is computing the participation of actors in each of the topics, and then examining the relations of the participation recorded, for example, in the form of *hypergraphs*. However, it is interesting to go a step beyond and examine structural equivalence, i.e. actors that have similar relations to others. The technique of two-mode *blockmodeling* (Doreian, Batagelj, & Ferligoj, 2004) provides a way of doing this with the help of automated algorithms.

*Blockmodeling* is aimed at transforming an apparently non-coherent network into a more easily comprehensible arrangement, according to a previously defined function of equivalence. The identification of classes of units (called clusters) sharing structural characteristics is one of the main goals of this procedure, a method not only flexible but also able to detect different kind of structures (e.g. cohesion and centrality). As the data gathered in our study are empirical, a kind of data which is rarely "perfect", we needed a tool for not only checking the structural features of affiliation networks but also capable of allowing for exceptions or errors in the dataset. After studying other possibilities we chose *blockmodeling* as the best available tool. From an examination of *blockmodeling*, several filtering strategies for participants can be implemented. Some of them will be discussed as part of the case study later. Other existing approaches use formal similarity models are based on computation on the skill and knowledge profiles of users (Pollalis & Mavrommatis, 2009), but ours uses only learner interaction within activities, thus being complementary and being operational in cases in which the profile of learners is not available or not formalized.



### 2.3. *Filtering-related learning objectives*

If it is found that some group of participants is not interested in a topic, it is easy to recommend alternate topics or even to hide them. For doing that, some kind of representation on the relationship among topics is needed. For example, if we have a model of topics and subtopics with similarity relations, different additional topics can be provided to the different groups of interest. *A priori* relationships between topics have been the focus of a large number of research initiatives, including general domain ontologies of a diverse kind. For example, Gašević, Jovanović, and Devedžić (2007) propose a framework for annotating learning object content with ontologies as a way to improve the retrieval of learning objects from repositories. Both formal ontologies (as the Gene Ontology[2] (GO)) or non-formal classifications as for example the one developed by the ACM[3] can be used to describe the topics of learning resources and eventually their relationship with learning activities. In the case reported here, an informal conceptualization is used, but other formal accounts could be used to allow that kind of filtering. Filtering learning objectives can be connected to filtering contents, as learning resources inside repositories are usually marked by these objectives, and ontologies can be used to find related objectives (Lee, Tsai, & Wang, 2008). These ontologies provide objective subject classifications as those used by sequencing methods in instructional theories, e.g. in *elaboration theory* (Reigeluth, 1999). For example, the GO provides a complete model of biological processes at the cellular level that could be used for that, e.g. GO term code 00008380 is the process "RNA splicing", and from the GO a software program can obtain more specific related terms (which can be used as related objectives) as "Regulation of RNA splicing" (code 0043484) or broader process categories as "RNA processing" (code 0006396).

Furthermore, if we have a historical database on the interest of individuals on certain topics, it is possible to choose topics that were of interest to similar people as new course offerings, thus implementing a specialized form of collaborative filtering. In doing so, it is necessary to take into account the type of learners, the courses in which they are/were enrolled and their difficulty. To avoid a bias for easy learning paths, choices must be made among alternative forms of the same instruction with equal degrees of difficulty so that teachers can be guided in the mode of presentation rather than being substituted by irrelevant ones.

### 2.4. Changing course structure

The use of specific SNA techniques provides support to changes in the course structure according to the students' profiles or preferences aiming at improving the way students assimilate the learning outcomes of the course providing or rearranging alternatives for of the same instruction rather than diluting the course content. For example, strongly connected topics can be candidates to be joined together, or even to be separated in another course, to provide enhanced modularity in some learning offering. In a similar vein, more peripheral concepts, according to the interest demonstrated by learners, might be removed, separated or re-arranged for future editions of the same learning experience.

As long as techniques for analyzing one-mode networks cannot always be applied to two-mode networks, it is possible to derive two (valued) one-mode networks from the original affiliation network, one representing the relations of interest among topics, and the other one representing common interests between pupils. Analyzing the data in the former, tutors will be capable of re-organizing the course structure in some cases, joining or splitting topics.

A helpful technique for the analysis of this kind of data is that of *m-slices*. An *m-slice* (also known as an *m-core*) is a maximal sub-network containing the lines with a multiplicity equal or higher to *m* and the vertices incident with these lines (de Nooy, Mrvar, & Batagelj, 2005). In an affiliation network of Hollywood female stars and films, for instance, a 3-slice would include all the films which share three actresses so, if the three-slice contains seven films, these three actresses must have acted in all the seven films in the three-slice. It should be noted that the *m-slices* of a network are nested – films sharing three actresses also share two actresses, so they belong to a two-slice as well –, and they represent cohesion in relation to the lines weights. This concept of *m-slice* can be used to identify highly related topics to a given intensity, and will be discussed in the case study later.

## 3. Case study

The case study is based on the second edition of an online course entitled "Design and Evaluation of Learning Objects and Reusable Learning Activities". The course content is composed of four modules which, in turn, are composed of several topics or themes as shown in Table 1.

The course was open to final year undergraduate and postgraduate students so we could classify them as mature students. There were a total of 47 students enrolled, 28 being females and 19 males with an age range from middle 20 to late 50. It is worth noting that 50% of them were lecturers from Spain or South American universities interested in education, which indicates a high level of interest in the course for a large number of students.

The learning management system used was Dokeos,[4] an open source platform that provides standard e-learning functionality including forums and threaded topics. The social network obtained from the discussion forums at the end of the course consisted of 47 actors, five of whom were tutors for different sections of the course and the rest, learners. At the end of the course, there were 16 events (discussion threads) stored in the system that covered all the syllabi of the course. Each thread in the course was represented as an event with a specific topic. It is important to point out that the preconditions stated in the description of the model – each thread having a clear topic, non-mandatory participation, and unlimited time for each thread – were all fulfilled.

The network representation followed the model presented above, where the arcs from actors to topics (having a different thread for discussing each topic) represented contributions. As a pre-processing step, topic nodes with a degree lower than two were removed. In this case, the degree represents the number of links to other nodes, i.e. threads without activity or just a single question–answer or learners who contributed less than two messages to the forums during the course (these could be considered drop-out students or with no interaction) Therefore, such removal allowed us to filter out both threads with no participation or participants who had no significant interaction. Also, the tutors were removed from the study because their participation followed very different objectives than that of the learners.

Table 1. Course contents.

| Course content | Hours (90 h) |
|---|---|
| MODULE I – Teaching and learning through the internet | |
| T1. Learning online | 5 |
| T2. Design of online learning materials | 5 |
| T3. Learning management systems and on-line tutoring | 5 |
| MODULE II. The concept of learning objects | |
| T4. Concepts of learning objects and learning designs | 10 |
| T5. Introduction to learning objects standards | 10 |
| MODULE III. Elaboration of learning objects | |
| T6. Metadata for learning objects | 15 |
| T7. The SCORM reference model | 15 |
| T8. IMS LD specification | 15 |
| MODULE IV. Use and evaluation of learning objects | |
| T9. Learning objects repositories | 5 |
| T10. Evaluation criteria and tools | 5 |

As a result, the final network was composed of 42 actors (nodes). Figure 1 provides the resulting partition of random blockmodeling using the Pajek[5] (Škerlavaj, Dimovski, Mrvar, & Pahor, 2009). This software tool implements SNA to derive empirically meaningful patterns of relationships, but it needs to be setup beforehand by establishing the number of partitions per network mode as an entry parameter for the blockmodeling algorithm. In this study, we tried to establish a number of partitions per mode according to both the number of learners and the number of subjects, even though this is a matter of interpretation and some experimentation could be appropriate. In this case, the (6, 6) partition was the one providing the most meaningful interpretation for the tutors of the course. Further experimentation with other values was performed but no relevant results were obtained.

From the data in Figure 1, a tutor can easily interpret that the partition commencing with MARCOS is that of learners with no significant activity, while the partition starting with BERTA ELENA includes the most active ones, except for the partitions with T5 – T6, and T7 and T10 in the columns, where T*i* corresponds to the *i* theme of course. Also, note that fictitious nicknames are used in the report that follows to guarantee anonymity of the data. Partitions T5, T6 and T7 deal with practical topics that require the use of computer tools, while the other activities do not. This clearly differentiates the group of RAMONA from the group commencing with MARCELA CRISTINA. On the other hand, this group shows less interest in threads T4H2 and T4H4, linked to topics dealing with theoretical issues on IMS Learning Design[6] which include concepts about learning objects and learning designs (definitions, justification of the paradigm, differences between learning objects and learning designs, etc.) a complex conceptualization model where a course is defined in terms of the nature of learning activities and interactions, without specific reference to discipline, content, or context. This discovered difference not only helps in differentiating learners more inclined to working with computer tools for creating learning object metadata, but it also speaks about the learners' background, as Learning Design concepts are particularly difficult for teaching staff – the case of most people in the group starting with MARCELA CRISTINA –, people who are naturally and deeply concerned with discipline issues and course topics (McAlpine & Allen, 2007).

Another straightforward interpretation is that the group starting with RAQUEL only showed interest on topics T1–T2 which were introductory issues on e-learning, before the key concepts of the learning objects paradigm were introduced.

Such kind of analysis could lead to instructional design decisions as varied as:

(1) Combining people with different interests to foster discussion in forthcoming activities, or combining people with the same interest to better focus those discussions.
(2) Combining people from more active and more passive groups, or filtering out the latter.
(3) Group formation inside the system, as previously commented in the way described by Wessner and Pfister (2001), for those distributed groups where learners and tutors are separated geographically and communicate solely through the Internet.

These options represent some possible forms of instructor-led on-the-fly filtering, which can use network data for the decision. They may also be used for assessment purposes or as a source of information for offering additional learning activities.

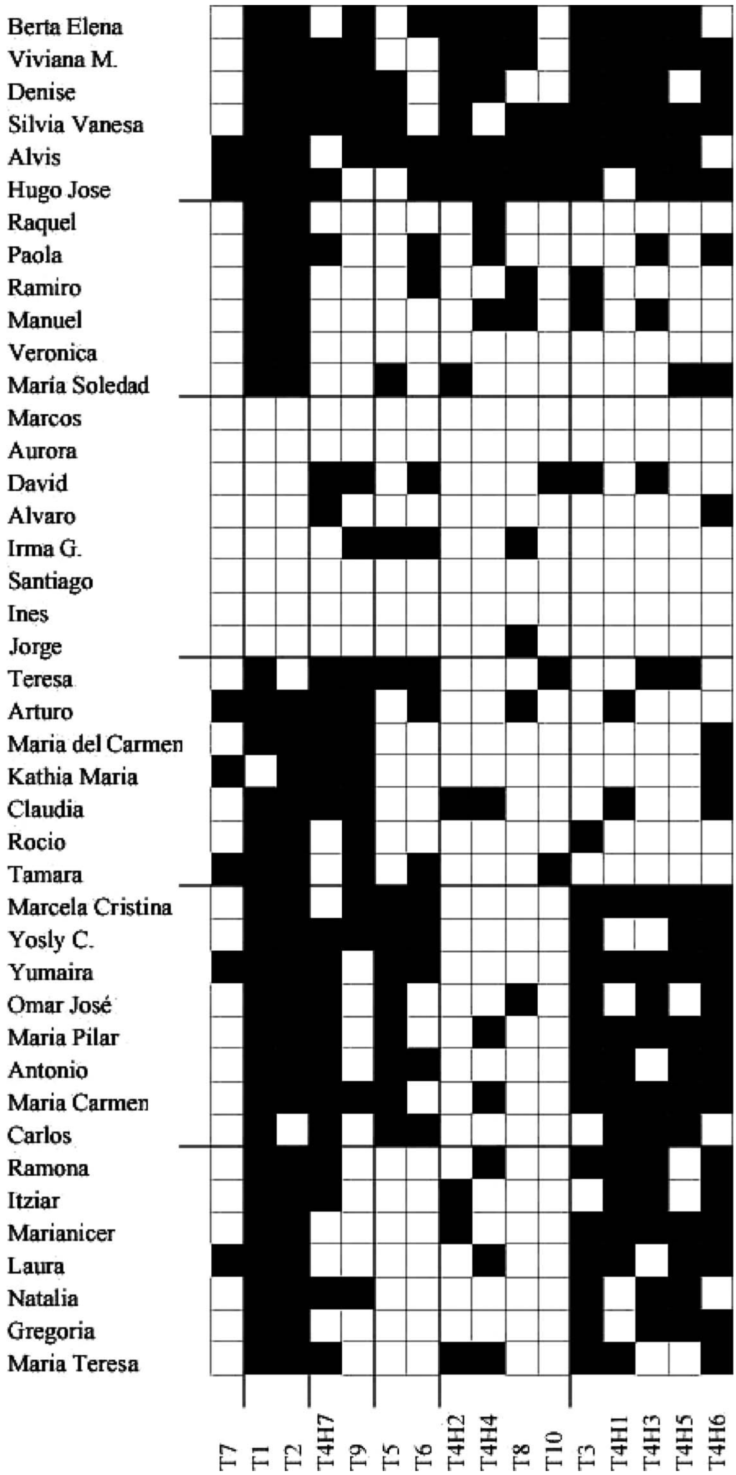


Figure 1. A (6,6) partition of the interaction data, with learners as rows and topical discussion threads as columns. The lines forming an irregular grid superimposed in the matrix separate the partitions on each of the two models.

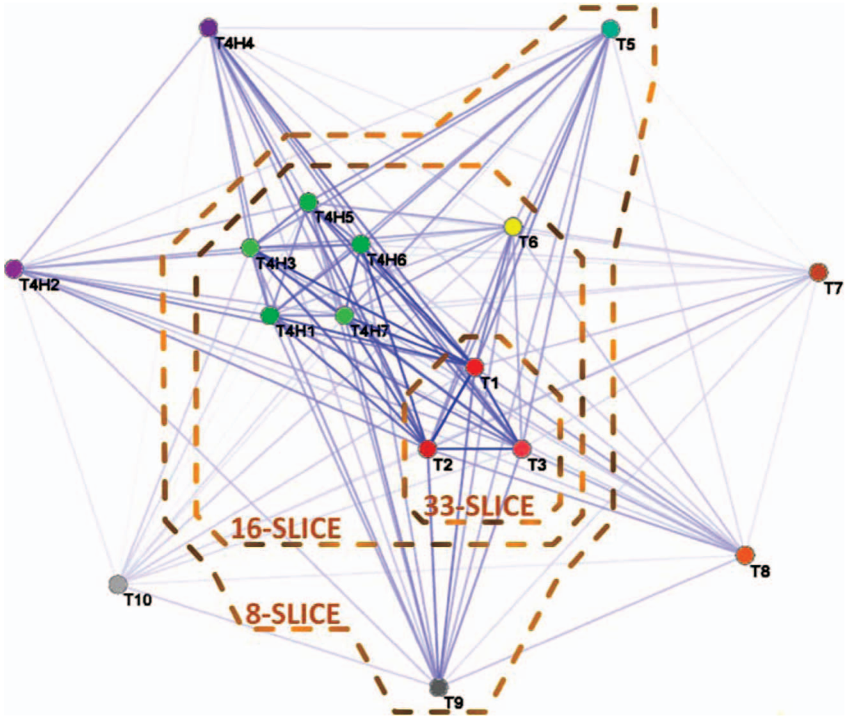


Figure 2. The *m-slices* for the threads in the network.

According to Murphy, Blaha, VanDeGrift, Wolfman, and Zander (2002), tutors should encourage and capitalize on student diversity to foster discussion. These authors state that "*real problems* [...] *rarely have a single answer; instead*, [*they*] *define a rich design and implementation space encompassing fundamental tradeoffs and room for creative difference*". So, leaving room for variance, as well as recognizing different students' values and backgrounds that lead to different choices, could be a good pattern to follow.

In addition to considering participants, it is possible to make some decisions on the basis of topics. For example, according to the model, it gives the impression that the participants in general were less interested from topic T5 onwards, which suggests the introduction of reinforcement activities on topic T5, which started the study of e-learning standards. However, the group of MARCELA CRISTINA seemed to be more interested in the topics related to software tools, pointing out a possible strategy for personalizing their participation in that direction, or suggesting some more advanced activities for topics T5 and T6.

In addition to *blockmodels*, the result of converting the network to one-mode and the subsequent identification of *m-slices* is shown in Figure 2. As previously said, the *m-slices* of a network represent cohesion in relation to the line weights, the higher the weight the more cohesive the sub-network. In our case, the slices studied come from the conversion of the original bimodal network to one-mode, high edge values representing high common interest as evidenced in activity in discussion threads.

The *m-slices* depicted in Figure 2 identify a 33-slice comprised of three introductory topics (Module 1), which seems to be a fairly cohesive group of interest. As expected, most students (around 80%) were very interested in the basic topics of the course but perhaps some of the students could have some spurious goals, e.g. they selected this course as a means to obtain optional credits needed to finish their degree. As the course progresses, their interest is groups are less cohesive and students' interest is divided between those that prefer more theoretical topics and those that prefer more practical topics. The 16-*slice* layer includes all T4 threads with the exception of the T4H2 and T4H4. These last two threads (T4H2 and T4H4) related to IMS Learning Design showed low levels of participation, which may suggest that it could be reasonable to separate IMS Learning Design contents to a second part of the course (optional or more advanced topics for only those that are really interested in such topics).

T6 topic is a 27-*slice*. This is a quite generic topic about Learning Objects metadata in general, later dealing with the IEEE LOM standard in particular, so it is closely related to the rest. T5 and T9 topics are both in a 8-*slice* and it is likely that lecturers attending the course found that these topics could be useful for the students in their everyday practice. T5 covers the foundation of SCORM as the basic reference in the field, whereas T9 is related to learning object repositories, where freely accessible educational materials can be obtained.

As evidenced by the use of the technique of slicing, the affiliation network structure is helpful in identifying differences in topics inside the same module (the case of T4), which could be used to break the module in two parts and perhaps offering it separately. Also, an account of what is "basic", i.e. of a more generalized interest is also exposed (as in the case of T6 which irrespective of the temporal sequence of the modules gathers significant general interest).

It should be noted that the differences found as in the case of T4 may be attributed to a difference in the difficulty posed by the topic, as IMS LD is arguably harder to master than the rest of topics in T4. Consequently, it is important to analyze other elements in the resources and activities used that complement the interaction data used to build the social networks models. For example, in this case, the *difficulty level* of resources and activities (that difficulty is an standardized metadata element in the IEEE LOM standard for learning object metadata) could be used to contrast with *m-slices*. If some correlation is found (as in the case of T4H2 and T4H4) then preferences can be hypothesized to be related to difficulty level. In any case, this does not invalidate the network model as a tool but adds an additional dimension that need to be considered when interpreting the output of the model analysis.

## 4. Conclusions and future work

The use of affiliation models for exploring online interaction in e-learning allows for the development of mathematical, quantitative techniques that are useful for filtering and personalization of the environment, and have the potential to help instructors, tutors and managers in their decisions both during a course or when re-designing or re-planning course offerings. This article has explored several techniques that could be helpful for different kinds of filtering that are useful in the context of e-learning settings. To the best of our knowledge, the results presented here open a new research direction in the intersection of SNA and the empirical analysis of educational data. The techniques presented and other similar ones will be subject to further study, as the work reported here has focused on the methodological issues for devising new analysis instruments, as there were no previous reports dealing with these issues so the approach was practically unexplored.

The approach presented herein has certain limitations regarding the organization of the interactions – which must be topical –, and is only intended to provide indicators. It is unclear whether the indicators can be directly used for automatic personalization or not, because there are no clear-cut thresholds or mathematical models for automated decision making – rather, it is the tutor or facilitator who should decide on the basis of the data coming from the SNA. Further, the results of the techniques described should be taken as an indication to aid in the decision making process of the tutors or facilitators that guide the learning process, since different types of noise and diverse variations in learner behaviour make the results reliable only as a confirmation or guide, and not as an straight automated decision.

Further work should go in the direction of evaluating indicators and metrics regarding AOP data and their potential usages as there are other SNA techniques and statistics in addition to the ones used in this article that might be applicable. Eventually, when enough evidence will be available, they could evolve into standard facilities in e-learning platforms, providing an advanced tool for the analysis of social interaction available for tutors and instructors in general.

Future work must also go in the direction of performing experiments with such new facilities, for instance, applying different strategies to rearrange some of the groups during the execution of the course and compare the learning outcomes of the groups within a course and across several editions of the course as well as and subjective evaluations.

## Acknowledgements

The results presented in this project have been partially funded by the University of Alcalá and the CAM (Comunidad de Madrid), as part of project MARIA "Medición y análisis de recursos para el acceso a la información y el aprendizaje", code CCG08-UAH/TIC-4178.

## Notes

1. http://www.amazon.com/
2. http://www.geneontology.org/
3. http://www.acm.org/about/class/
4. http://www.dokeos.com/
5. http://vlado.fmf.uni-lj.si/pub/networks/pajek/
6. http://www.imsglobal.org/learningdesign/

## Notes on contributors

Daniel Rodríguez is an associate professor at the Computer Science Department, University of Alcalá, Madrid, Spain. He obtained his degree in Computer Science at the University of the Basque Country, Spain and PhD degree at the University of Reading, UK in 2003. Daniel has also worked as software engineer and consultant between Spain and Sweden. His research interests include the application of data mining techniques to learning and software engineering. He is also a member of IEEE and ACM associations.

Miguel Ángel Sicilia obtained an MSc degree in Computer Science from the Pontifical University of Salamanca, Madrid (Spain) in 1996 and a PhD degree from the Carlos III University in 2003. He worked as a software architect in e-commerce consulting firms, being part of the development team of a Web personalization framework at Intelligent Software Components (iSOCO). Currently, he is an associate professor and leads the Information Engineering Unit at the Computer Science Department, University of Alcalá. His research interests are primarily in the areas of adaptive hypermedia, learning technology and human–computer interaction. He leads the AIS SIG on reusable learning objects (www.sigrlo.org).

Salvador Sánchez-Alonso obtained a university degree in Computer Science from the Pontifical University of Salamanca in 1997, and a PhD in Computing from the Polytechnic University of Madrid in 2005. He worked as an assistant professor at the Pontifical University of Salamanca for 7 years during different periods, and also as a software engineer at a software solutions company during 2000 and 2001. From 2005, he is an associate professor at the Computer Science Department of the University of Alcalá. His current research interests include learning objects reusability, metadata, Object-Oriented technologies and Software and Web Engineering.

Leonardo Lezcano obtained his degree in Computer Science from the University of Alcala (UAH) where he is currently a PhD candidate. He is also a lecturer and researcher at the Open University of Madrid (UDIMA) where he reads Information and Communication Technologies (ICT). His research interests include the Semantic Web and Semantic Interoperability both oriented to the Healthcare domain. He is a member of the Information Engineering Research Unit at the UAH where he has collaborated in several projects. He joined the openEHR development team that is in charge of the openEHR Java Implementation Project (www.openehr.org).

Elena García-Barriocanal obtained a university degree in Computer Science from the Pontifical University of Salamanca in Madrid (1998) and a PhD from the Computer Science Department of the University of Alcalá. In 1999, she joined the Computer Science Department of University of Alcalá and from 2001 she is an associate professor. She is a member of the Information Engineering Research Group. Her research interests focus on topics related to the role of knowledge representation in human–computer interaction and learning technologies where she actively works on ontological aspects both in e-learning and in usability and accessibility.



## References

Allert, H. (2004). Coherent social systems for learning: An approach for contextualized and community-centred metadata. *Journal of Interactive Media in Education*, *2*. Retrieved June 22, 2009, from http://www-jime.open.ac.uk/2004/2

Bull, S., Dimitrova, V., & McCalla, G. (2007). Open learner models: Research questions (Special Issue Part 1). *International Journal of Artificial Intelligence in Education*, *17*(2), 83–87.

Cho, H., Gay, G., Davidson, B., & Ingraffea, A. (2007). Social networks, communication styles and learning performance in a CSCL community. *Computers and Education*, *49*, 309–329.

de Nooy, W., Mrvar, A., & Batagelj, V. (2005). *Exploratory social network analysis with Pajek*. New York: Cambridge University Press.

Dillenbourg, P. (1999). Introduction: What do you mean by collaborative learning? In Dillenbourg (Ed.), *Collaborative learning. Cognitive and computations approaches* (pp. 1–19). Amsterdam: Elsevier.

Dimitrova, V., McCalla, G., & Bull, S. (2007). Open learner models: Future research directions (special issue part 2). *International Journal of Artificial Intelligence in Education*, *17*(3), 217–226.

Doreian, P., Batagelj, V., & Ferligoj, A. (2004). Generalized blockmodeling of two-mode network data. *Social Networks*, *26*, 29–53.

Gašević, D., Jovanović, J., & Devedžić, V. (2007). Ontology-based annotation of learning object content. *Interactive Learning Environments*, *15*(1), 1–26.

Guo, Z., & Greer, J.E. (2006). Electronic portfolios as a means for initializing learner models for adaptive tutorials. First European Conference on Technology Enhanced Learning, (EC-TEL 2006), Crete, Greece, October 1–4, 2006, Proceedings. In *Lecture Notes in Computer Science* 4227 (pp. 482–487).

Kay, J. (2001). Learner control. *User Modeling and User-Adapted Interaction*, *11*, 111–127.

Kreinjs, K., Kirschner, P.A., Jochems, W., & van Buuren, H. (2007). Measuring perceived sociability of computer-supported collaborative learning environments. *Computers and Education*, *49*, 176–192.

Lee, M.-C., Tsai, K.H., & Wang, T.I. (2008). A practical ontology query expansion algorithm for semantic-aware learning objects retrieval. *Computers and Education*, *50*, 1240–1257.

McAlpine, I., & Allen, B. (2007). Designing for active learning online with learning design templates. In *ICT: Providing choices for learners and learning* (pp. 639–651). Paper presented at the Proceedings of Ascilite Singapore 2007. Retrieved June 22, 2009, from http://www.ascilite.org.au/conferences/singapore07/procs/mcalpine.pdf

Muehlenbrock, M. (2005). Formation of learning groups by using learner profiles and context information. In C. Looi, G. McCalla, B. Bredeweg, & J. Breuker (Eds.), *Proceedings of the 12th International Conference on Artificial Intelligence in Education, AIED 2005* (pp. 507–514). Amsterdam: IOS Press.

Murphy, L., Blaha, K., VanDeGrift, T., Wolfman, S., & Zander, C. (2002). Active and cooperative learning techniques for the computer science classroom. *Journal of Computing Sciences in Colleges*, *18*(2), 92–94.

Nussbaum, E.M. (2008). Collaborative discourse, argumentation, and learning: Preface and literature review [Special issue]. *Contemporary Educational Psychology*, *33*(3), 345–359,

Oakley, B., Felder, R.M., Brent, R., & Elhajj, I. (2004). Turning student groups into effective teams. *Journal of Student Centered Learning*, *2*(1), 9–34.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1998). The PageRank citation ranking: Bringing order to the web. Technical Report. Stanford Digital Library Technologies Project.

Paz Dennen, V. (2008). Looking for evidence of learning: Assessment and analysis methods for online discourse. *Computers in Human Behavior*, *24*(2), 205–219.

Pollalis, Y., & Mavrommatis, G. (2009). Using similarity measures for collaborating groups formation: A model for distance learning environments. *European Journal of Operational Research*, *193*(2), 626–636.

Reigeluth, C.M., (Ed.). (1999). *Instructional-design theories and models Volume II: A new paradigm of instructional theory*. Mahwah, NJ: Lawrence Erlbaum.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994). GroupLens: An open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 Computer Supported Collaborative Work Conference*, Chapel Hill, North Carolina, USA, October 22–26, April 1994. CSCW '94 (pp. 175–186). New York, NY: ACM.

Škerlavaj, M., Dimovski, V., Mrvar, A., & Pahor, M. (2008). Intra-organizational learning networks within knowledge-intensive learning environments. *Interactive Learning Environments*. DOI: 10.1080/10494820802190374.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. New York and Cambridge, ENG: Cambridge University Press.

Wessner, M., & Pfister, H. (2001). Group formation in computer-supported collaborative learning. In C. Ellis & I. Zigurs (Eds.), *Proceedings of the 2001 international ACM SIGGROUP Conference on Supporting Group Work* (pp. 24–31). New York: GROUP '01. ACM.