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

Miguel-Angel Sicilia, Salvador Sánchez-Alonso and Leonardo Lezcano

Information Engineering Research Unit
Computer Science Dept., University of Alcalá
Ctra. Barcelona km. 33.6 – 28871 Alcalá de Henares (Madrid), Spain
`{msicilia, salvador.sanchez}@uah.es, leonardo.lezcano@alu.uah.es`
http://www.cc.uah.es/ie

Abstract. The on-line 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 the configuration of further learning activities or the filtering of learning resources. This paper explores the use of an affiliation network model for such kind of purposes. Concretely, the use of *blockmodelling* and the examination of *m-slices* are explored as tools to decide on the configuration of topics and/or learner groups.

Keywords: social network analysis, e-learning, affiliation networks

1 Introduction

As learning technologies widespread, an increasing amount of people are getting involved in on-line learning activities. Learning designs nowadays emphasize the organization of learning experiences around *activities*, which in many cases are shared by more than one individual. The interaction of learners takes place through different kinds of *services* including newsgroups and chats. In consequence, people interact with such services, for learning about a particular *objective* (e.g. topic, competency), which might be of different granularities. This kind of relationship *activity-objective-people* (AOP) is the basic material for the empirical analysis of social interaction through technology enhanced learning.

There are several methods for analyzing AOP interaction. Some authors use qualitative analysis of the discourse, while others measure the number of interventions or the users' satisfaction regarding the "social atmosphere" (Kreijns et al., 2007). However, these techniques require intensive effort from the tutors to categorize and examine each of the interventions, and are subject to the subjectivity of the tutors. Since the amount of communication events (e.g. messages) is often large, it seems reasonable to look for indicators of actual social interaction that could be computed with the help of mathematical tools, thus helping tutors in decision making.

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

(i)   Hypothesis testing or exploratory studies aimed at finding correlations,

(ii)   The summative assessment of learners, and

(iii)   Re-configuring the learning environment (e.g. proposing new activities, forming groups, and changing the future course structure) or taking other kind of actions based on the analysis data.

This latter third purpose (iii) includes attempts to act or personalize the learning process or to find data that could be used to direct the learner to more interesting activities or contents from the learner's point of view.

Our study focuses on analyzing AOP data for several purposes related to common interest that can be categorized as (iii) according to the list above. Concretely, we approach AOP data in the form of an *affiliation network*, considering that learners' participation in activities can be used to detect groups of common interest. Obviously, this kind of analysis is only useful when certain conditions hold, including (a) that the participation in activities is not mandatory (b) that the activities are clearly directed towards a concrete, recognizable objective. The former would include a bias to participation, while the latter will make the interpretation of interaction data blurred. The use of an affiliation network directly follows the structure of empirical data found in 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.

Wasserman and Faust (1994) define affiliation networks as two-mode networks in which actors are grouped according to the affiliation (participation) in a finite set of *events*. Co-participation in events determines the sub-group structure for that kind of networks. This paper 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 are the concrete learning activities and, for the sake of simplicity, the case study considers each discussion thread to be a separate *event*. It is important to highlight that the interpretation of affiliation network is not intended to discover *social circles*, since 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.

The rest of this paper is structured as follows. Section 2 describes the general model for affiliation-based similarity, and the SNA methods behind. 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

Many current e-learning activities include some form of observable service for participation. Newsgroups (forums) and chats are among the most common ones. If the conversation activity that takes place in those services is structured around

concrete topics of interest, it is possible to develop techniques for the analysis of common interests in learners through the analysis of their participation in the activities. This idea is essentially the same that originated collaborative filtering (Resnick et al., 1994), but in this case it is not necessary to provide explicit ratings for the items (topics), if we assume that a learner that 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 she/he shows an interest in the topic, thus allowing for a form of quantitative indicator. 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 than other Web-based rating systems such as page ranking (Page et al., 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 modeled that way, it is possible to devise some courses of action, that we label here as different forms of "collaborative filtering"[1]. It should also be noted that the classical model of rating-based collaborative filtering (CF) is essentially based on triples *(item, user, rating)*, and the data described below can be easily mapped to the same kind of data by considering something as *(objective[service], learner, #-communications)*, where *objective[service]* is the concrete learning objective of each service (each thread in a newsgroup, for example) whereas #-communications represents any numerical account of the participation in the concrete service (e.g. number of messages). However, the usefulness of using a pure classical CF approach is limited by the availability of data, since CF performs well for large databases of ratings.

The model

An **affiliation network** is a kind of social network in which the actors are divided in two disjoint sets, and ties are only allowed between elements that belong to different sets. Then, we have a network $G = (A, V)$ where A is a set of g actors of which e actors are considered the subset E of *events* in which the other *(g-e)* actors may be affiliated. The set V is a set of (undirected) ties that connects the events in E with the actors (the rest of the elements in A). The particular interpretation of the affiliation network in our case is the following:

> The set E of events is any of the topical, planned discussion topics in the learning design. 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 e_j.

---

[1] The usual interpretation of collaborative filtering is that of recommendations or ranking of information. 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 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 it can be related in some known way). This excludes generic "social forums" or "general questions" discussions.
2. Participation in the threads should not be made mandatory to avoid bias.
3. The time planned for each thread must be similar (ideally, with a non-strict limit).

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.

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.

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 et al., 2004). One extreme is identifying groups that are close in their interests. 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 relationships of the participation recorded, for example, in the form of *hypergraphs*. However, it is interesting to go a step beyond and examine structural equivalence, that is, actors that have similar relations to the others. The technique of two-mode *block modelling* provides a way of doing this with the help of automated algorithms. From an examination of block modeling, several filtering strategies for participants can be implemented. Some of them will be discussed as part of the case study below.

Filtering related learning objectives

If some group of participants not to be interested in some topic, it is easy to recommend related topics (or 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.

Further, 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 CF.

Changing course structure

It is possible to derive a (valued) one-mode network from the affiliation network that represents the relationship of interest among topics. This can be used to re-organize structure in some cases, joining or splitting topics. For example, topics that are connected with a high strength 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, concepts that are more peripheral, according to the demonstrated learners' interest, might be removed, separated or re-arranged for future editions of the same learning experience.

The concept of *m-slice* can be used to identify highly related topics to a given intensity, as will be discussed in the case study below.

3 Case study

The case study is based on the second edition of a concrete online summer course on learning object standards and technology. The learning management system used was Dokeos[2], an open source platform that provides standard *Forums* and threaded topics as a form of general-purpose service. The network consisted on 42 actors, 5 of which were tutors (for different parts of the course) and the rest learners. 13 events (discussion threads) covered the syllabus of the course. Actors were numbered for identification purposes, numbers 1, 2, 21, 34 and 35 corresponding to the tutors. Each thread in the course was represented as an event with a concrete topic. It is important to point out that the preconditions stated above (in the description of the model) were all fulfilled.

As a preprocessing step, nodes with degree lower than 2 were removed to filter out either threads with no participation or participants that had no significant interaction, resulting in a net with 48 actors. Also, the tutors were removed from the study since their participation could not be interpreted in terms of topical interest.

Figure 1 provides the resulting partition of random block modeling using the Pajek tool[3]. The number of partitions per mode can be decided on the number of learners and on the number of subjects, even though this is a matter of interpretation and some experimentation could be appropriate. In our case, we used a (6,6) partition, but also experimented with other values. However, the (6,6) partition lead to an interpretation that was meaningful for the tutor of the course.

---

[2] http://www.dokeos.com/
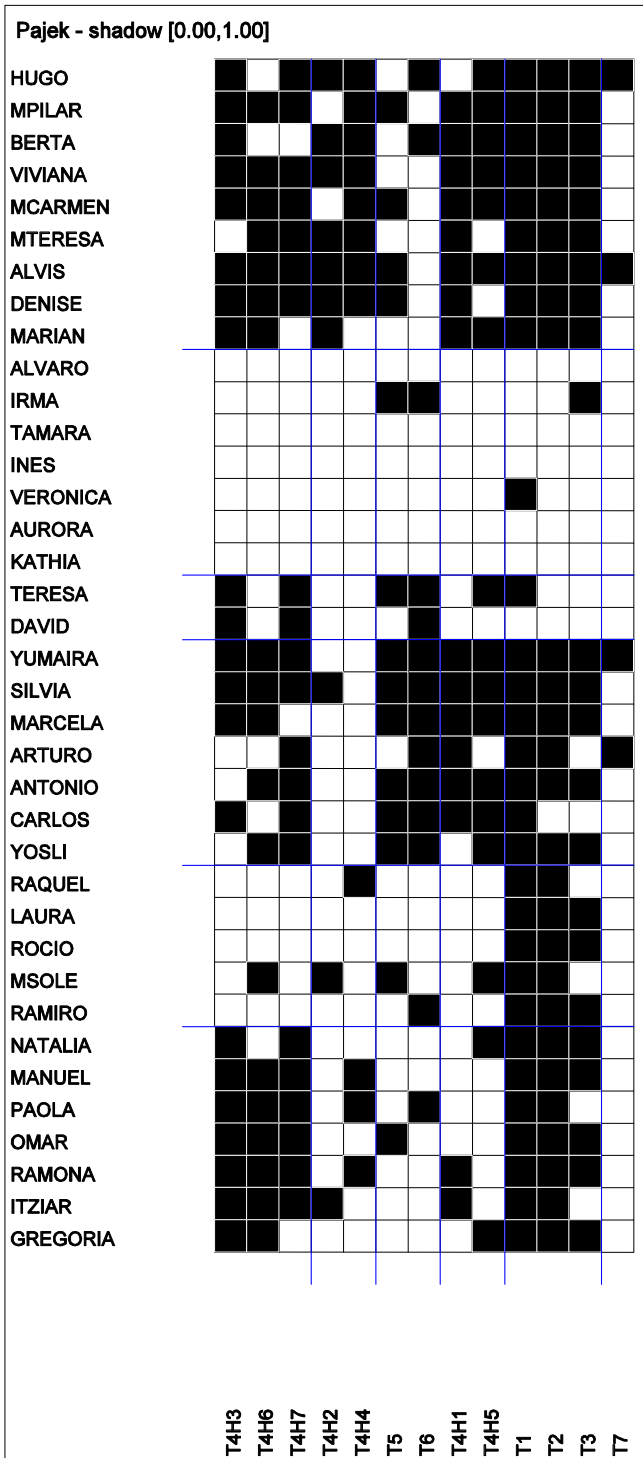[3] http://vlado.fmf.uni-lj.si/pub/networks/pajek/

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.

From the data in Figure 1, the tutor can easily interpret that the partition commencing with ALVARO is that of learners with no significant activity, while the partition starting with HUGO includes the most active ones, except for the partitions with T5-T6 and T7 in the columns. 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 HUGO from that of the group starting with YUMAIRA, which in turn has less interest in T4H2 and T4H4, whose topic deals with theoretical issues on IMS LD[4]. This discovered difference helps in differentiating learners more inclined to working with computer tools for creating learning object metadata.

Another straightforward interpretation is that the group starting with RAQUEL only showed interest on topics T1-T3 which were introductory issues on e-learning, before the concepts of learning object 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.

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

In addition to considering participants, it is possible to take some decisions on the basis of topics. For example, according to the model, it looks like the participants in general have some less interest from topics T5 onwards, which suggests introducing reinforcement activities on topic T5, which started the study of e-learning standards. However, the group of YUMAIRA seems to be more interested on the tool side, so a possible strategy is that of personalizing their part in that direction, or suggesting some more advanced activities on T5-T6.

Besides block models, the result of converting the network to one-mode and the subsequent identification of *m-slices* is shown in Figure 2. An *m-slice* is a maximal sub-network containing the lines with a multiplicity equal or higher to m and the vertices incident with these lines. The m-slices of a network are nested, and they represent cohesion in relation to the lines weights. 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.

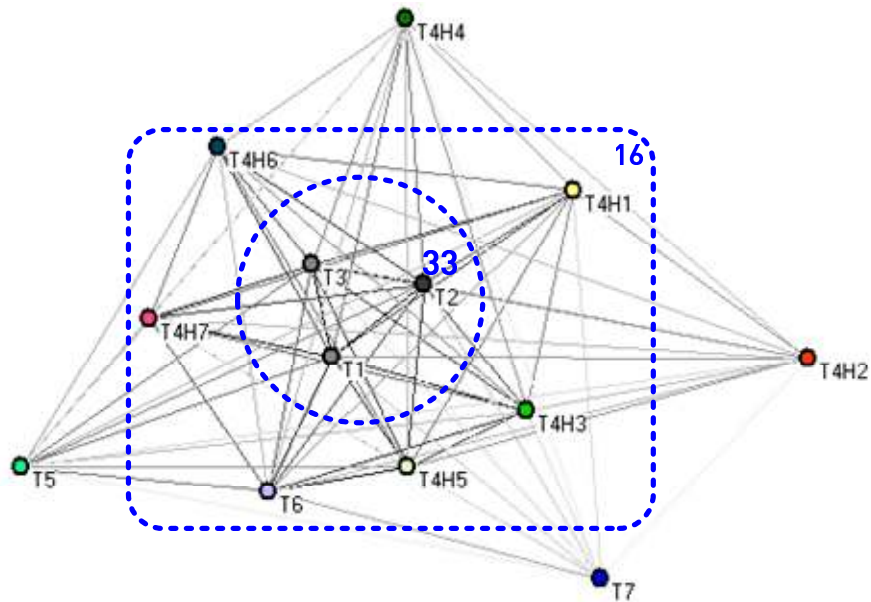[4] http://www.imsglobal.org/learningdesign/

Figure 2 – m-slices for the threads in the network.

The m-slices depicted in Figure 2 identify a 33-slice including the three introduction topics, which seems a reasonable cohesive group of interest. Then, there is another layer of 16-slice that only excludes T7, T4H2, T4H4 and T5. The two topics in T4 (T4H2 and T4H2) that are excluded are those related to IMS LD, which may suggest that it could be reasonable to separate LD contents to a second part of the course. T6 is about IEEE LOM and it is in a 27-slice, so it is closely related to the rest, while T5 about SCORM is in a 4-slice.

4 Conclusions and outlook

The use of affiliation models for exploring on-line interaction in e-learning allows for the development of mathematical, quantitative techniques that are useful for filtering and personalization of the environment. This paper has explored several techniques that could be useful for different kinds of filtering that are useful in the context of e-learning settings.

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 not 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 social net analysis. 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 behavior 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. 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.

5 Acknowledgements

References

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

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

Oakley, B., Felder, R.M., Brent, R., & and 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.

Resnick, P., Iacovou, N., Sushak, M., Bergstrom, P., and Riedl, J. GroupLens: An open architecture for collaborative filtering of netnews. Proceedings of the 1994 Computer Supported Collaborative Work Conference. (1994).

Wasserman, S., & Faust, K. (1994) Social *Network Analysis: Methods and Applications*. New York and Cambridge, ENG: Cambridge University Press.

.