An approach of collaboration analytics in MOOCs using social network analysis and influence diagrams

Antonio R. Anaya
E.T.S.I. Informática - UNED
C/Juan del Rosal, 16
E-28040 Madrid
34 913986550
arodriguez@dia.uned.es

Jesús G. Boticario
E.T.S.I. Informática - UNED
C/Juan del Rosal, 16
E-28040 Madrid
34 913989387
jgb@dia.uned.es

Emilio Letón
E.T.S.I. Informática - UNED
C/Juan del Rosal, 16
E-28040 Madrid
34 913989473
emilio.leton@dia.uned.es

Félix Hernández-del-Olmo
E.T.S.I. Informática - UNED
C/Juan del Rosal, 16
E-28040 Madrid
34 913988345
felixh@dia.uned.es

ABSTRACT
MOOCs pedagogical strategies assume that students construct their own knowledge and collaborate with their mates. Large-scale learners' interaction figures hinder both proper interpretation of learners' needs and prompt remediation actions. To this we describe a preliminary study of a two-step collaboration analysis, which consists of inferring domain-independent indicators on students' relationships obtained from social network analysis and using an influence diagram to warn teachers on students' problematic circumstances to facilitate prompt remediation actions.

Keywords

Collaboration analytics, SNA, influence diagram, collaborative learning

1. INTRODUCTION
Massive open online courses (MOOCs) are stood out as a new pedagogical methodology since they aimed at large-scale participation and open access via the web [1]. In this situation the teacher loses control over the learning process and students should construct their own learning. The students can use the MOOC's communication means to collaborate with their learning mates [2]. In this respect, although the students are to be provided with the tools and services to collaborate, this thus not suffice and frequent and regular analyses of the team process are needed to know whether the collaboration takes place [3]. Moreover, the special large-scale nature of MOOCs hampers teachers when coming to analyze students' communication acts, which drive the collaboration process.

Some researchers have proposed a well-known analysis method, social network analysis (SNA) to minimize the problems commented above [4, 2]. However, in this collaborative learning context some variables, such as emotion and empathy, are out of control [3]. Under these circumstances, analyzing the collaboration process requires to deal with uncertainty [5], which can be tackle with Influence diagrams (ID) [6].

In our research we propose an approach to automatically warn (or recommend [7] teachers on students' problematic collaboration circumstances so that they can readily provide corrective actions when required. Thus, the objectives of the application are: 1) to analyze the collaboration with a transferable analysis method that provides domain-independent collaborative indicators; 2) to minimize the human intervention.

The rest of the paper is organized as follows. First we describe in Section 2 related research, to both SNA in MOOCs and ID in the educational context. In Section 3 we frame the research and educational context in which this work is being applied and an in-depth description of the proposed methodology. We then comment on our preliminary study in Section 4 and finally briefly provide the main conclusions and further planned research in Section 5.

2. Related research
MOOCs offer more leeway to students and thereof features new challenges [8]. In this more crowded and less constrained learning environment it is advisable to use any available technology to analyze the learning process involved. Here technologies such as SNA are starting to be applied with relative success [2].

SNA has been used to identify students who are actively participating in course discussions and thus are potentially at a risk of dropping out [2]. [4] examined and detected, using SNA, communities of users within a large course so that they can be provided with a personalized and social-oriented recommender system. [9] presented an example of a Social Learning Analytics Tool to visualize real-time discussion activities in a MOOC environment.

SNA has been widely applied to study the social aspect of students learning [10]. This way [11] analyzed networks in order to identify the people from whom an individual learns. Here [12] proposed a methodology to analyze students' interactions in a collaborative learning environment, which consists of using SNA to get meaningful statistical indicators, such as the student reputation. [13] emphasized the use of SNA techniques to discover relevant structures in social networks so that the instructors were able to better assess participation.

As the aforementioned approaches we aimed at improving collaborative settings though SNA outcomes in terms of a technology that has proved its usefulness in tackling problems under uncertainty. Moreover, the educational context has been a traditional suitable field where Bayesian networks (BN) have been applied to deal with the inherent uncertainty involved [14]. [15] proposed a course diagram method, based on an ID framework, which can be used by an instructor to design a course structure. The diagram organizes the instructional material and the tests.

3. Towards collaboration analytics in MOOCs
In our education context we proposed to combine two different technologies to analyze the collaboration. Firstly, the SNA obtains indicators from students' interactions, which reflects how students connect with their mates. Secondly, an influence diagram

structures students' indicators as a network, which supports a decision on students' problematic circumstances once an expert, who can be the tutor, tunes the probabilities of the network. The software used for SNA was Gephi[1] and for ID was OpenMarkov[2].

3.1 Social network analysis

To date the most common communication service in MOOCs is forums. SNA has been applied to forums in order to infer the social relationships among users [16]. Here SNA metrics support the inference of social relationship indicators.

Figure 1 shows the SNA diagram resulting from the data of the on-line course that we have used in the preliminary study.

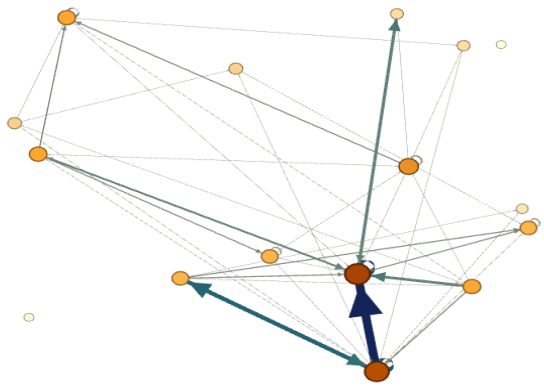

Figure 1. SNA in the preliminary study.

In Figure 1 nodes are students who participated in an online course (see Preliminary study section and communications among students were analyzed through SNA. Within this figure the metric Degree of the nodes is represented as follows: red and big node means high degree, and yellow and small node means low degree. The color and size of the ties mean the weight of the relationship (i.e., number of messages from origin node to destiny node).

We propose the following centrality metrics of the nodes as indicators of the collaboration process:

- **Degree** is the number of ties of one node.
- **In-degree** is the number of ties whose destiny is the node. This indicator is a measure of the node popularity.
- **Out-degree** is the number of ties whose origin is the node. This indicator is a measure of the node sociability.
- **Closeness centrality** is the degree to which an individual is near all other individual in the network. This reflects the ability to access to information by the network members.
- **Betweenness centrality** a measure that quantifies the frequency or number of times that a node acting as a bridge along the shortest path between two other nodes.
- **Eigenvector centrality** is the measure of the importance of a node in the network. Intuitively, the nodes that have a high value of this measure of centrality are connected to many nodes, which are connected also in this sense; therefore, are good candidates to disseminate information.

We use these indictors, because they are well-known in the state-of-the-art research focused on analyzing the position of the students in the network using SNA [16]. These indicators constitute a standard way to measuring network and node features and they can be used in several different context.

3.2 Influence diagrams

IDs provide us with a framework for representing and solving decision problems under uncertainty. As our objective is to maintain a domain independent and general approach of inference IDs include features that are advisable in learning environment as MOOCs, where the collaborative learning is encouraged. The collaboration settings constitute a framework where not all variable are known in advance. In addition, a MOOC is an educational environment where teachers cannot afford the continuous tracking and analysis phases of learners' interactions, which in this case are massive. An ID could help teachers to identify and carried out correction decisions adapted to each student.

We propose an ID where the indicators obtained from the SNA, the centrality metrics commented above, are structured. The network layout of the proposed ID is showed in Figure 2.

In Figure 2 the yellow and round nodes are the variables in the problem. Assessment is the root and hidden variable, which is unknown in future test. The node "Assessment" represents the teacher's assessment of students' collaboration. The ID needs a training dataset with known values of the node "Assessment" to tune the networks probabilities. The other yellow and round nodes are the SNA indicators. The squared node "D" represents the decision, in this case, yes or not. The decision "yes" means a detection of problematic circumstances and the ID supports teacher with a suggestion so that the teacher makes a corrective actions. The node "U" maximizes the decision utility. Notice that the values of the nodes have to be discretized. In order to do the discretization we divided interval values into three groups with equal width. We propose three values: high, medium and low, because these values are easy to understand.

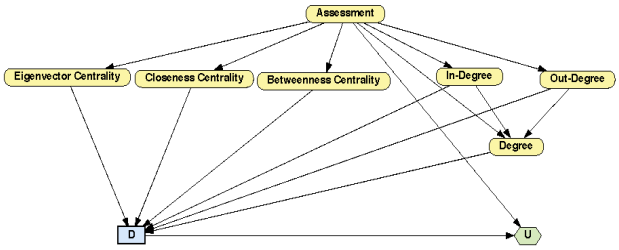

Figure 2. Network of the influence diagram

4. Preliminary study

In the preliminary study we have used data from an on-line course to fine-tune the ID's network. The experience was done with students of the subject "Complexity and Computability" in the forth course of the degree of Computer Systems Engineering at UNED (Spanish National University for Distance Education). In this subject we have mimicked the characteristics of MOOCs, with particular emphasis on the participation on the forum. For that reason we have undertaken a continuous assessment process on the Learning Management System (LMS) forum's interactions

[1] http://gephi.github.io

[2] http://www.openmarkov.org

in order to detect the student level of participation and the recording of a special type of video podcast [17].

We have tested our approach in an online course, which let us make a preliminary proof of concept on the main issues involved, namely tracking and assessing students (16 students). This curse has been designed following the large-scale MOOC's course settings, meaning that it consists of the same video lectures, individual tasks and a communication services that will be ultimately provided [17].

Node Potential: Degree

Relation Type: Table

Assessment	low	low	low	low	low
Out-Degree	low	low	low	middle	middle
In-Degree	low	middle	high	low	middle
high	0	0.333333	0.333333	0	0.333333
middle	0	0.333333	0.333333	0	0.333333
low	1	0.333333	0.333333	1	0.333333

Figure 3: An example of node "Degree" probabilities.

In the fine-tuning process experts can insert knowledge into the ID's network, that is, in the automatic inferring process. Firstly the students should be assessed according to their interactions. It is fairly common that experts decide which students' features, that their interactions have revealed, are the most relevant to be assessed. This knowledge is showed when the assessments are compared with the analysis of students´ interactions, which is independent of expert's assessments. We made the SNA of the students´ interactions and independently an expert assessed the students.

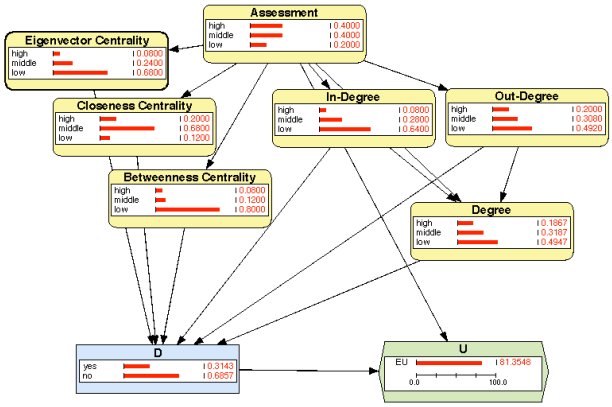

Figure 4: A general perspective of the ID's network results.

Once the students were assessed and we obtained the SNA centrality attributes of each student, we then discretized the data. Then, we were able to measure the probabilities for each case. An example is show in Figure 3. According to the Figure 2, the node "Degree" has three fathers, the nodes "In-Degree", "Out-Degree" and "Assessment". For each possibility of "Degree" value (low, medium or high) we measured the probability according to the values of the father nodes. Figure 3 shows some cases. For instance, when the father node have "low" value, the node "Degree" have "low" value. Because node "Degree" has three father nodes, there are 27 possible cases (the Cartesian product of three variables with three possible values). We made the fine-tuning process with 16 students, thus, we did not have enough data to fine-tune the network completely. We could solve this lack with data from the next experience.

After the probabilities were established for each possible case of each node, the ID was able to infer a decision and the decision utility for each case. Figure 4 shows the general perspective ID's results. It can be seen that the ID advises to recommend only in around one third of cases (In node "D", "yes" is 0.3143 and "no" 0.6857).

We can observe what happens when the ID advises to recommend, i.e., identifies a possible collaboration problematic circumstance. Figure 5 shows the case when the ID advises to recommend. When the ID advises to recommend, the student has low value in the nodes "Degree" and "In-Degree". This informs us that when a recommendation is advisable, the student is not active and her/his classmates ignore her/him. Thus, the ID has identified a problematic collaboration scene, which can be happened over the course. With this information the teacher could make a corrective activity to improve the collaboration process.

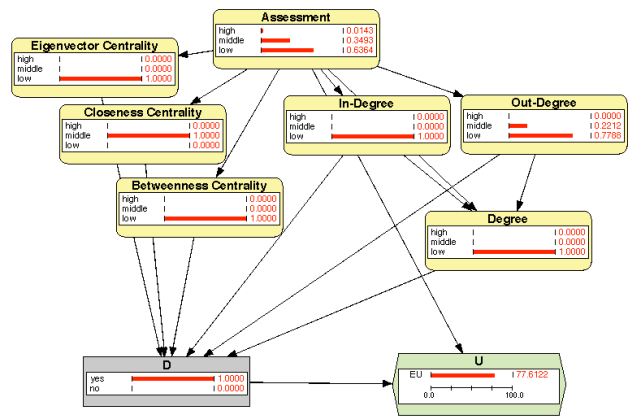

Figure 5. Example: ID advises to recommend.

In addition to the previous analysis, it is possible to calculate the optimal policy of the ID (see Figure 6). Thus, the optimal policy informs about the decision (yes or not) for each combination of nodes values.

Optimal policy: D

Relation Type: Table Reorder variables

Degree	low	low	low	low	low	low	low	low
Eigenvector Centrality	low	low	low	low	low	low	low	low
Betweenness Centrality	low	low	low	low	low	low	low	low
Out-Degree	low	low	low	low	low	low	low	low
Closeness Centrality	low	low	low	middle	middle	middle	high	high
In-Degree	low	middle	high	low	middle	high	low	middle
yes	0	0.5	0.5	1	0	0	0	0
no	1	0.5	0.5	0	1	1	1	1

Figure 6. Optimal policy: all possible decisions of ID.

Figure 6 shows an example of the decisions, yes or not, according to the values (high, middle or low) of the centrality attributes obtained from the SNA. In the preliminary study we had 16 assessed students. The possible cases that the ID can consider mathematically are the Cartesian product of network nodes and the student indicators (a total of 729 cases). Thus, not all the possible cases of the nodes values combination have to be considered by the ID. However, the results (see Figure 6) show that the ID is capable to support with different decisions according to the students SNA centrality attributes values. However, more

interaction data are needed to continue with the ID tuning process. When tuning process is finished, a new student's attributes values from the SNA feed the ID that, in turn, can offer accordingly a new decision (i.e., "yes", suggestion of a corrective action due the possible student's problematic circumstance in the collaboration).

The approach labels students with "yes" (the student needs a recommendation) or "not" (the student does not) and this way guides teachers to identify the student's collaboration problem. Based on this the teacher can create the appropriate recommendation to the student.

5. Conclusions and future work

To facilitate collaborative learning management within MOOCs in this paper we propose a domain independent and transferable approach, which is based on two different technologies: 1) Inferring domain-independent indicators on students' relationships obtained from social network analysis (SNA) in their interactions; 2) From these indicators an ID is used to warn teachers on students' problematic circumstances so they can provide them with prompt remediation actions. Here teachers cannot afford the continuous tracking and analysis phases of learners' interactions, which in this case are massive.

The preliminary results described in this paper confirm that the approach can identify problematic collaboration scenes, although it should be further investigated. Thus, data from more students will be considered, which will be used to tune the ID's network probabilities. Thanks to the approach, the tuning process can be made while the students are participating in the MOOC. Moreover, the final suggestion that is offered to the teacher can also be improved. The suggestion should be easily understandable by any non-expert user so that the analysis process involved won't prevent them from its usage.

The research described in this paper will be further applied within the MAMIPEC project, which aimed to infer and provide affective personalized support to learners in educational contexts [18].

6. ACKNOWLEDGMENTS

Authors would like to thank the MAMIPEC project (TIN2011-29221-C03-01), which has been funded by the Spanish Ministry of Economy and Competence.

7. REFERENCES

[1] Masters, K. (2011). A brief guide to understanding MOOCs. The Internet Journal of Medical Education, 1(2).

[2] Sinha, T. (2014, February). Together we stand, Together we fall, Together we win: Dynamic team formation in massive open online courses. In Applications of Digital Information and Web Technologies (ICADIWT), 2014 Fifth International Conference on the (pp. 107-112). IEEE.

[3] Johnson, D., Johnson, R. (2004). Cooperation and the use of technology. In: Handbook of research on educational communications and technology. Taylor and Francis Group, pp. 401–424.

[4] Zhuhadar, L. and Butterfield, J. (2014). Analyzing Students Logs in Open Online Courses Using SNA Techniques, Twentieth Americas Conference on Information Systems, Savannah, 2014.

[5] Anaya, A. R., Luque, M., García-Saiz, T. (2013). Recommender system in collaborative learning environment using an influence diagram. Expert Systems with Applications 40, 7193–7202.

[6] Howard, R. A., Matheson, J. E. (1984). Influence diagrams. In: Howard, R. A., Matheson, J. E. (Eds.), Readings on the Principles and Applications of Decision Analysis. Strategic Decisions Group, Menlo Park, CA, pp. 719–762.

[7] Hernández del Olmo, F., & Gaudioso, E. (2008). Evaluation of recommender systems: A new approach. Expert Systems with Applications, 35(3), 790-804.

[8] Duque, R., Bravo, C., Ortega, M. (2013). An ontological approach to automating collaboration and interaction analysis in groupware systems. Knowledge-Based Systems 37, 211 – 229.

[9] Schreurs , B., Teplovs, C., Voogd, S. (2014). Social Learning Analytics applied in a MOOC-environment, eLearning Papers, 36, January 2014

[10] Shum, S. B. and Ferguson, R. (2012). Social Learning Analytics. Educational Technology & Society, 15, 3—26

[11] Haythornthwaite, C., & De Laat, M. (2010). Social networks and learning networks: using social network perspectives to understand social learning. Paper presented at the 7th International Conference on Networked Learning, Aalborg, Denmark.

[12] Bratitsis, T., Dimitracopoulou, A., Martínez-Monés, A., Marcos-García, J., Dimitriadis, Y. (2008). Supporting members of a learning community using interaction analysis tools: the example of the Kaleidoscope noe scientific network. In: Proceedings of the IEEE International Conference on Advanced Learning Technologies, ICALT 2008. pp. 809–813.

[13] Rabbany, R., Takaffoli, M., & Zaïane, O. R. (2011). Analyzing participation of students in online courses using social network analysis techniques. In Proceedings of educational data mining.

[14] Millán, E., Loboda, T., de-la Cruz, J. P. (2010). Bayesian networks for student model engineering. Computers & Education 55, 1663–1683.

[15] Chang, F.-I. (2003). Quantitative analysis on distance learning courseware. Multimedia Tools and Applications 20, 51–65.

[16] Dawson, S. (2008). A Study of the Relationship between Student Social Networks and Sense of Community, Educational Technology & Society, 11, 224—238

[17] Letón, E., Molanes-López, E.M. (2014). Two New Concepts in Video Podcasts: Minimalist Slides and Modular Teaching Mini-videos. 6th International Conference on Computer Supported Education.

[18] Santos, O.C., Salmeron-Majadas, S., Boticario, J.G. (2013). Emotions Detection from Math Exercises by Combining Several Data Sources. AIED 2013, Lecture Notes in Computer Science LNCS/LNAI 7926, 742–745.