Reveal the relationships among student's participation and its outcomes on eLearning environments: Case study

Diego Alonso Gómez Aguilar1, Roberto Therón1, Francisco J. García-Peñalvo1
1Departamento de Informática y Automática, University of Salamanca, Salamanca, Spain
{dialgoag, theron, fgarcia}@usal.

Abstract— Recent research has confirmed the impact of electronic media usage on student outcomes. However, little is known about the impact of measurable reading access to a forum and other resources. We have developed a tool to study this aspect through visual analysis. Results from this study showed that not only participation on electronic discussion fora, such as reading or posting, but also the reading access to the resource, can alert about the student’s performance. This came about after proving the existence of a strong influence relationship between the frequency of reading access to discussion fora and the frequency of access to posting fora, as well as between the frequency of reading access to resources and the frequency of access to posting fora in the outcome of the students. However, we have found that the union of the variables above, i.e., frequency of posts in fora, reading of discussion fora and reading of resources can be visually analyzed, thus obtaining a better visualization of the patterns of students' performance.

Index Terms— eLearning, Educational Data Mining, Feedback, Learning Analytics, Inferring Social Networks, Predictive Systems.

I. INTRODUCTION

The increasing student participation, the ubiquity of information and communication technologies (ICT), and the potential benefit of observing the learning process have prompted the appearance in recent years of a body of research in which technology is used to monitor, analyze, deduce and act upon the deployment of a learning experience. Students interact with numerous resources (documents, fora, chats, specific tools), with the instructors, among themselves, etc. Interaction is identified as a crucial component of the learning process. Research into online learning defines participation as involving different forms and using different analysis units [1]. Participation may be detected from numerical indicators such as the number of accesses to the platform, the number of written messages; content analysis of the quality of the messages; the relationship between reading and writing; and the perception the students have of the activity through interviews, questionnaires and reflective learner reports.

Anderson and Garrison first categorized the different modes of interactions that can take place in a distance-learning environment [2] and later Anderson proposed an “equivocality statement” with two observations: deep and meaningful learning is supported if one of these forms of interactions is at its high level; and high levels of more than one of these modes will likely increase the quality of the experience although at a significantly higher cost [3, 4]. Following this work we may assume that a successful learning experience has a high level of interaction in one of its aspects.

Our work builds upon this idea and we approach the problem from a visual analytics viewpoint: the next section briefly reviews related work, discusses a case study and finally the conclusions of the work are presented.

II. RELATED WORK

The term “Academic Analytics” (AA) became relevant in the research community after being coined by Goldstein and Katz in a seminal work that defines it as the application of business intelligence tools already in use in business and marketing to the area of learning [5]. The main application of AA is to go beyond simple reporting and suggest better decisions in academic institutions by observing analyzing vast amounts of events occurring in their activities. The article characterizes AA as an “engine” working with five steps: capture, report, predict, act and refine. The examples provided in this work refer mostly to the problem of detecting “at risk students,” that is, those students that might drop out of a course or abandon their studies. There is a detailed description of how detecting these situations may benefit the institutions. An example of the use of AA is the Signals project [6].

Educational data mining (EDM) is a field that pursues a similar goal as AA. It proposes the use of various techniques (statistical analysis, machine-learning, data mining, etc.) to resolve educational research issues and understand the setting in which students learn [7]. The use of ICT has prompted the appearance of vast amounts of events recorded during the learning process. Some of these data sets are being collected, labeled and organized to be shared within the research community [8]. Data mining techniques are already used in other contexts [9], and their application to educational contexts has been the target of an increasing number of research groups.

As opposed to the view of AA where the captured data is used to make decisions at the institution level, the objective of Learning Analytics (LA) is to use this data and any other additional observations that can be obtained, and use it to impact directly on the students, the instructors and the details of the learning process, for example, analysing the interaction...
among students and prompting an immediate student participation or change and adaptation of the course material. The aspect that all three areas have in common is the need for a vast amount of observations of the learning process. These observations may come from different sources, in different formats, at different times, and need to be processed so that any of the useful outcomes claimed by each area can be obtained [10, 11].

The research agenda in [12] provides a comprehensive list of key aspects that influence visual analytics (VA,) the process by which users gain insight into complex data. For discussion here, the most relevant aspects are: science of analytical reasoning and interaction techniques. Visual analytics deals with the capabilities of visualization tools that help users make judgments about the data. It is important to create visualization tools that maximize human capabilities to perceive and understand complex and dynamic data. Though visual representations provide an initial direction to data analysis, they will not be effective without interaction mechanisms that let users explore data [13].

The level of supervision of the media has recently become one important element. The importance for providing teachers with “real-time” data related to relationship establishment is critical to inform the professor of any necessary modifications of activities to pursue better eLearning. There is strong evidence to suggest that within well-structured activities, knowledge construction processes reach higher levels of critical thinking and that students are able to establish and sustain cohesive groups. This evidence serves to substantiate the need for automated Social Network Analysis (SNA) tools that are able to deliver “real-time” analytics for tutors and instructors [14]. Some studies use the number of event upload operations and the number of file update operations, and other measurable events to investigate the students performance [11]. According to Barry and Smithey Fulmer [15], the degree to which other people can retrieve and read messages from an individual can affect their use. In this sense, the fact that everyone could access on-line discussion forum messages, both by teachers and by students, at any time could affect the use of the medium. For example, a comment not suitable, a stupid question, or even poor spelling by a student could change the opinion of classmates about him/her.

This phenomenon could influence students when using the online discussion forum. On the other hand, lecturers included the participation of on-line discussion fora a small part of the final grade. So student participation in discussion fora could also be influenced by this kind of oversight.

In order to assess the degree of participation in online courses, various methodological approaches have been adopted. As for quantitative methods, counting blackboard hits [16] and message postings [17, 18] has been often used [19]. The Moodle (learning management system, LMS,) the iGraph [20] out-of-the-box system is ready to be used and is coupled with a simple content analysis to reject irrelevant messages and to collect social network metrics. The metrics help to give a better measurement of the characteristics of the network (usually important when comparing different networks).

The students’ participation in the activity in terms of quantity of messages posted in the forum seems to be influenced by the quality of infrastructures put at the students’ disposal [21, 22]. High numbers of student contributions are not always correlated to their learning performance. In fact, Hwang and Arbaught [23] found that those who write fewer messages, but participate in the discussion regarding a wider number of topics, achieve better results in the final knowledge test in comparison to those who write more but are limited to only one topic of discussion.

As with most research areas, terminology is not as homogeneous as it would be desirable, but we could say that the terms EDM, AA, LA, VA and the studies discussed above, cover most of this work.

III. PROPOSAL AND CASE STUDY

We have previously approached the analysis by means of Web Services on a Virtual Learning Environment (VLE) [24, 25]. For our case study we used our visual analysis tool (based on a social network visualization), deployed for a Moodle-based course in order to analyse a second year engineering course during the Sep 2009 - Sep 2010 scholar year. Almost 46,500 events were analysed for a total of 115 students.

We develop a personalized Social Graph visualization from our previous work [26] in order to improve it and carry out this study. Taking advantage of its benefits, i.e., it can be used at any given time in the course, it is independent of the VLE version, can be used offline, and so on. This developing expands the possibilities of the visual analytic tool in order to reveal other kind of patterns. The social graph visualization for this study is a force-directed graph. It interprets a graph as a physical system with forces between the nodes and then tries to minimize the energy of the system to obtain a nice drawing. Such algorithms are used for drawing arbitrary (sparse) networks such as flow charts, program planning graphs, telephone call graphs, etc. They can also be applied to clustered layouts [27]. For the purpose of this study we apply internal-spring and external-spring to create the clusters.
The analysis consists in the selection, unification, standardization and filter from the VLE log events. That filter was made to eliminate the forum posts that are not relevant (like "smiles", "hugs", "hello", etc.) [28].

The data for this study were collected throughout the course. First, the number of hits, reading access and messages for each discussion, the number of reading accesses to a resource was counted to produce quantitative data for the analysis.

To represent the information we used the visualization of social network. Visualizing social networks is more than simply creating intriguing pictures, it is about generating learning situations: “Images of social networks have provided researchers with new insights about network structure and have helped them communicate those insights to others” [29-31]. Such network images are created by drawing graphs made up of nodes and connecting lines.

This study tries to prove the existing correlation between two and three quantitative variables of the student's participation, but also, as complementary contribution to the visual analytic tool which already can be used with other existing metrics (e.g., degree of separation, centrality, etc.) [14, 25]; and all of it can be represented on the same intuitive interactive visualization. Therefore, we have not, on one hand, made changes to the current functionality of the visualization and, on the other hand, we made the representation of the above relationship possible. That is the reason to use the Social Graph representation.

We created a graph that shows all activity types among all students. This graph is considerably dense (Fig. 1) but we can see the overview of the structure's course. The red and grey shapes forming a circle are the people, the blue triangles are the resources, the green ones are discussions, the orange are fora and the green and gray lines are conversations. Also, edge assessment makes possible the illustration of the relationship intensity among students with the access reading of a resource or discussion and the frequency of post fora. The weakening of relationship edge intensity leads to sequential distancing of edge.

For this reason, the behaviour of the graph is depicted differently depending of the visible elements (Fig. 2, Fig. 3, Fig. 4). This means that the smaller the difference between frequency of activities among students, the closer they are placed between them and also closer to the group of discussions or resources or both, depending on the case. Following the idea that the farther the students are from the group discussions or group of resources, depending on the case, the activity frequency is lower. This is because the size of edge lines is calculated so as to represent a dispersion of the students in relation to the correlation among the quantitative variables visualized. Note that this type of calculation through the edges of the graph is not currently done thus, it is the contribution of this work.

With this study we try to answer whether it is possible to find a relationship between the student performance and the participation explained in the introduction, by means of visible changes within our visual analysis tool [26]. In order to confirm this hypothesis, we analysed three cases:
The relationship among the frequency of access to posting fora and the frequency of access to reading resources with the grades of the student Fig. 2;
- The relationship among the frequency of access to posting fora, the frequency of access to reading discussions and the grades of the student Fig. 3; and
- The relationship among the frequency of access to posting fora, the frequency of access to reading resources, the frequency of access to reading discussions and the grades of the student. Fig. 4.

Thus, we hid the elements of the visualization that we did not want to correlate (in this case posts and fora) and left visible the discussions icon (green ones on Fig. 1, Fig. 3 and Fig. 4) or resources (blue ones on Fig. 1, Fig. 2 and Fig. 4) or both depending on the case; furthermore, for the purpose of this work we divided the group of people into three parts for each case above. We selected 3 groups through the visualization interaction, it creates a red frame (as you can see on Fig. 3, Fig. 2, Fig. 4). Thus, when selecting three times, we have 1, 2, 3 groups of people with a specific visual analytics relationship with variables. Finally we assigned the students’ grades to each person in the groups (see Fig. 5) and the average:

<table>
<thead>
<tr>
<th>Group</th>
<th>Grade</th>
<th>Group</th>
<th>Grade</th>
<th>Group</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>D: 1</td>
<td>5.7</td>
<td>D: 2</td>
<td>5.8</td>
<td>D: 3</td>
<td>5.0</td>
</tr>
<tr>
<td>R: 1</td>
<td>5.7</td>
<td>R: 2</td>
<td>6.0</td>
<td>R: 3</td>
<td>4.5</td>
</tr>
<tr>
<td>DR: 1</td>
<td>6.0</td>
<td>DR: 2</td>
<td>5.6</td>
<td>DR: 3</td>
<td>5.4</td>
</tr>
</tbody>
</table>

In Fig. 5 lines represent the cluster formed through the visual selection of people from the visualization of the social graph. The X-axis represents the student’s grades and the Y-axis the number of people. Each line represents a group of people that was selected from a relationship in a visual representation. Thus, 1, 2, 3 are the groups of people of a specific visual analytics relationship of variables. The one represents the most participative group (blue line), the two the medium (green line) and the three is the lowest participative group (orange line). The letters D and R mean the relation of each person with the number of read access to a discussion (D) or a resource (R). DR is the combination of both.

The group of people nearest to the group of discussions or resources, depending on the case, are the most active people of the courses (for example D:1 are the group of people who accessed the reading of discussions more frequently and made more posts in the fora). The more distant people in the group of discussions or resources are the lower activity people (group 3 with orange lines in Fig. 5).

In general, and taking into account the data in Fig. 5, we can see the improvement in the students’ grades in the three cases from 3 to 1, i.e., the orange lines have the lowest students’ grades and the blue lines have the highest. Nevertheless, if we see the average of the D's groups or the R's groups, the above pattern seems to remain hidden. This is, for the case of R's groups (Fig. 2) the R case (correlation between frequency of reading access to resources and frequency of post fora and the student's grades) group 1 (in theory the students with high outcome) should be the best average, but it is the R:2. The results of D's groups are similar. This probably looks like bad news, but notice that differences between the D:1 case (that should be the highs) and the D:2 case (0.1 of difference) are lower than the same relation on R's groups (0.3). Now look at the same difference, but this time between the spaces among the students (Fig. 3, Fig. 2). Notice that the students are closer among them in Fig. 3 in comparison with Fig. 2. As a conclusion, this visualization helps to reveal a tendency easily and in an interactive way, thus quickly providing hidden information.

It is clear that the future trend will be toward greater adoption of online learning. Thus, there will be an increased opportunity for instructors and administrators to monitor student activity and interaction with the course content and their peers. The analyses of these data sets are directly relevant to student engagement as well as evaluating implemented learning activities. Questions related to how and when students are engaged, and what activities promote student engagement, can be answered through the monitoring and analysis of student online behavior. While the data examined in this paper are indicative at this stage, and the interpretation of results discussed here may be influenced to some degree by a number of exogenous variables, the findings nonetheless provide new insights into student learning that complement the existing array of evaluative methodologies (e.g. evaluations of teaching). In this regard, the study provides a platform for further investigation into new suites of diagnostic tools that can, in turn, provide new opportunities and new data sets to inform instructor reflection for continuous and incremental improvement of pedagogical practice.

IV. Conclusion

In this article we used the Social Graph representation of a visual analytic system which is able to depict interactions in online environments (particularly in LMSs). A first contribution of this work is to prove, with the experiment, that there exists a pattern of behavior between this kind of participation and student grades.

This visualization brings up new possibilities to e-learning.
as a tool capable of not only helping the teacher assorting and illustrating the degree of participation of students, identify key students in information passing, and find the implicit relations between fora participants, but also will help to find hidden patterns.

In many ways, graphs are the obvious solution for trying to visually represent a network structure. No other graphical system functions as much as a full time symbol of connections between nodes. Nonetheless, given the limitations of the resolution and screen size of computers, social network graphs usually end up cluttered and rather illegible. As a contribution of this article we suggest that graphs should be designed with interaction and some other techniques following the mantra: “Analyze First -Show the Important -Zoom, Filter and Analyze Further - Details on Demand” [32, 33].

ACKNOWLEDGES

This work is partially supported by Spanish Science and Innovation Ministry through the oPLe project (ref. TIN2010-21695-C02) and Vidovic project (ref. FFI2010-16234, ) also by the Regional Council of Education of Junta de Castilla y León through the projects GRA7 and MPLE (ref. SA294A12-2).

REFERENCES