Academic Analytics: Mapping the Genome of the University

Sérgio André Ferreira and António Andrade

Abstract—Higher education institutions have multiple technologic subsystems for administrative, pedagogical management, and quality purposes, which gather an immense volume of data from various sources and which do not communicate with each other. The domain of the analytic performances in education emerges from the need to aggregate multiple sources of data, which the complexity of treatment associated with the ease of mobilizing selected information will make it possible to understand reality and optimize management actions. In this paper, we present the architecture and results achieved in the development of an academic analytics aggregator of multiple sources of data on the educational activity.

Index Terms—Academic analytics, data-driven decision making, higher education, learning analytics, planning.

I. INTRODUCTION

A GENOME is a set of genes, where all the information necessary to construct an organism is. The mapping of the genome allows us to understand the functioning of the organism and its decryption enables knowledge of the causes of many diseases, paving the way so they can be prevented or tackled [1]. Applying this concept to organizations, the genome will match the huge amount of data about the most diverse activities which, if properly structured into information, can produce knowledge about the functioning of the organization, for example, by identifying potential and problems and promoting a better informed decision making, which is fundamental to the effectiveness of management. This explains the title of this paper.

By its nature, the data related to the educational process are of particular interest to the higher education institutions (HEIs). The integration of technology in teaching and learning through the use of hardware such as computers and mobile devices, and the investment of HEIs in technological platforms, with particular emphasis on the Learning Content Management Systems (LCMS) activities, are associated with an increased volume of recorded data: each click, each task performed online is stored somewhere in a database. The idea of Analytics in Education is simple and is associated with a very high transformative potential: the use of this data allows a more informed process of decision making, opening doors to new models in the management of HEIs in the organizational efficiency and pedagogical fields.

The use of Analytics – Decision making based on data, where information is used to support and justify decisions at all levels of the company [2], [3] – in education can occur on two levels (Table I):

1) Learning Analytics (LA), more directed to the process of teaching and learning at an institutional or infra-institutional scale (e.g. student, curricular unit, course, faculty). The Society for Learning Analytics Research (SOLAR) defines LA as the measurement, collection, analysis and reporting of data about students and their contexts, with the purpose of understanding and optimizing learning and the environments in which it occurs [4];

2) Academic Analytics is focused on the institutional or supra-institutional scale and, in terms of the goals it sets, is a parallel concept to Business Analytics. At this level, Analytics enables managers/executives to have access to indicators – historical or in real-time – on the business (high education institution) and its units (faculties, schools or departments) [2].

In Marketing, Management and related areas, Analytics is already in a mature stage; however, its use in education is still in its infancy. Two major factors contribute for this
embryonic stage: i) a later manifestation of interest in the potential of Analytics by the actors of the education sector; ii) the need for the conjunction of requirements of technological, educational and organizational order magnifies the complexity – and difficulty – in the development of Analytics systems [6].

Analytics in Education is clearly an emerging theme, fore-shadowing its adoption by HEIs in the near future – according to the Horizon Reports, within two or three years [7]. The interest shown by national and international organizations, the organization of conferences specifically for the discussion of this subject (e.g. three editions of Conference Learning Analytics and Knowledge, in 2011, 2012 and 2013 [8]) and the increasing number of scientific publications prove the relevance of the source.

Despite this momentum, the practical application of Analytics is still very focused on the LMS. Several researchers believe that this approach is a necessary, although not sufficient, part of a comprehensive solution, arguing that LA efficient systems should include more data than the LMS [9].

Although the functionality of monitoring the student’s activity is usually included as a resource of generic software, Analytics systems with the possibility of extracting and aggregating data and which offer versatile tools for the reporting and data visualization were, until recently, basic or non-existent [6]. However, several projects have begun to emerge in the area of Analytics, as, for example, the Horizon Projects [10].

This study falls within this context. Starting from a real HEI, the Universidade Católica Portuguesa - Centro Regional do Porto (Católica-Porto), the goals of this study are:

1) Present the study in terms of building a system of LA, fueled by LCMS data in use at the institution – the Blackboard – and by the opinion of students on the use and integration of this platform in the teaching and learning process.

2) To suggest an architecture for an Academic Analytics system that can aggregate multiple sources of data from the computer systems in use at the educational institution, allowing it to quickly display indicators on the pedagogical activity at the different levels of the organization (student, teacher, CU, course, faculty, Católica-Porto), having the use and integration of the LCMS in the teaching and learning process as the associated scale (LA developed in the previous section).

3) To equate a future scenario of an Academic Analytics that aggregates all sources of data relevant to the management of HEIs, including sources of teaching, research, employment, financial and others.

The theoretical relevance of this work is justified by the questioning of some of the main issues discussed in the areas of LA and Academic Analytics. The practical contribution is fulfilled through the creation of a LA system based on the use and integration of the LCMS in the training process, and in the opening of paths for future operationalization of LA systems and Academic Analytics in HEIs.

This article is organized into five more sections, besides this introduction: in section II we demonstrate the complementarity of LA and Academic Analytics; in sections III to V we sequentially present three concrete cases of application of Analytics in Católica-Porto, with different objectives and different scales of application, but complementary to each other; in the last section we summarize the main conclusions and present prospects for future work.

II. THE COMPLEMENTARITY OF LEARNING ANALYTICS AND ACADEMIC ANALYTICS IN HEIS

HEIs are organized according to a hierarchical structure; a typical organization can be: University > Faculty > Department > Course > CU. These structures and the different actors of the HEIs are in need of Analytics systems with different scales and granularities. Figure 1, adapted from the work of George Siemens [11], summarizes the use of Analytics at the macro and micro scales, emphasizing the complementary relationship between them.

The upper part of Figure 1 reflects the role of Analytics at a macro scale – the HEI as a whole – and lists some
important ideas. The Analytics systems have at their base the Big Data, a term that portrays this context of abundance of data [5], which is the raw material of Academic Analytics. At this scale, Analytics reflects a top-down view, in the sense that the analysis is done on a global level through the integration of data sources from different technological subsystems of the HEI in an analysis tool, which should allow access to indicators, historical or in real-time in the different dimensions of the HEI and its units (e.g. faculties, departments, courses, CU...).

The data sources used to power the Academic Analytics can be limited to aspects related to the educational aspect, being in this case nearer to the concept of Learning Analytics. However, at the scale of Academic Analytics, data analysis is performed in a more institutional level and less at the individual student’s level (as in LA), taking into account data from all the technological subsystems in the process of learning (which includes examining the relationship between student, content, institution and teacher) [5]; or to integrate data from a variety of sources (e.g. data from pedagogical activity, but also administrative, financial and other sources, enabling the HEIs to manage the different dimensions behind its functioning), becoming a parallel concept to Business Analytics. In this scenario, data is the foundation of the entire analysis. Academic Analytics can be based on data sources of different kinds – such as administrative services, LCMS or financial systems – and in multiple formats – such as spreadsheets or reports of system [12].

Institutionally, these systems are used by administrators, quality control services, marketing and financing of HEIs to support decision making and guide to action [5], [12], [13], and it is possible to identify five stages in their use: capture data > production of reports based on the data collected > identification of relevant information/patterns/predictions > intervention > redefinition of actions [12].

The lower part of Figure 1 focuses on a micro scale level and the approach is the bottom-up type, where the smallest elements of the system are specified in great detail. The data (small data) is used by a user or small group of users (e.g. course coordinator, teachers who teach a certain CU), who are looking for more filtered information to the level of a course, CU or student.

LA may have a micro level that allows the teacher to define rules to monitor the activity of a specific CU or any particular student. The LCMS offers a range of skills at that level. Figs. 2–4 show some practical examples of the operationalization of LA at that level.

Figure 2 shows the network of interactions of a forum through a graph system in order to evaluate the dynamics of the forum, identify the students most important for the discussion, the most absent and the number of interactions between individual students. Figure 3 shows the mode of operation of a detection system of students at risk in a LCMS, where you can customize the value of each item for which the system sends the alert. In Figure 4 one can see the report of the system when a student at risk is detected (in this case by registering low activity in the LCMS and poor access). The interpretation of such data allows the teacher to assess and predict the academic progress of students and intervene to give students more opportunities for success.

III. ANALYTICS IN PRACTICE: USE AND INTEGRATION OF LCMS IN THE TEACHING AND LEARNING PROCESS

A. Goals

Three main factors contributed to the idea of designing and implementing a LA system in Católica-Porto:

1) The LCMS Blackboard is a technological platform that supports the teaching and learning activity in general, and which is widely used in Católica-Porto [14],
making it important to understand the effective inte-
gregation in the different dimensions of the teaching and
learning process. The goal is not only to manage better,
but also to use information to promote rapid adaptation
to new situations.

2) Católica-Porto has in operation a System of Assurance
and Evaluation of Internal Quality (SAEIQ). However,
its action does not include a specific assessment of
aspects related to the use of technology in the teaching
and learning process, in particular the LCMS, which is
a centerpiece in the institution [14]. Previous studies
recommend that HEIs should define criteria for assessing
the quality of Technology Enhanced Learning Envi-
nvironments, which support a proactive management that
maximizes the benefits and avoids chaotic situations
that are often associated with the introduction of these
environments [15]. This project falls within this context,
with plans to link the LA with the work of SAEIQ, con-
tributing to the construction of an Academic Analytics
system.

3) LA is referred to in published studies as an emerging
trend with high potential in the transformation of higher
education. Indeed, basing decisions on data and evidence
improves the quality of the decisions, productivity and
performance of organizations [5]. The investment and
the commitment of Católica-Porto in the use of tech-
nology to enhance teaching practice, shown throughout
its history, is an incentive and justifies going down this
path [14].

B. Methodological Notes

These factors contain issues of the organizational, educa-
tional, and technological kind. The genome of the LA’s success
depends on the response to the dimensions of this triangle:

1) Organizational Dimension – It is expected that the
LA system responds to the requirements of the institu-
tion’s information, providing useful information on the
integration of the LCMS in the teaching and learning
process, which can be aggregated by multiple levels of
analysis that reflect the organization of the institution:
university, faculty/school and CU. The organization of
information according to level of detail allows condi-
tional access to information, if that is the objective of
the institution. In this case, each teacher will only have
access to information about the CUs he or she teaches,
the director of the faculty/school to all CUs of the
institution he or she manages, the SAEIQ and the board
of Católica-Porto to all information.

2) Educational Dimension – In published studies there
are multiple models that characterize the evolution of
technology integration in online teaching activity, with
descriptors to make this placement (e.g. Technology
Florida Center for Instructional [16]). Based on the
literature review and the features offered by LCMS, the
dimensions of the analysis model were defined (Fig. 6).

3) Technological dimension – The design of all the back-
office of the system of the extraction and analysis of
data was performed to meet the requirements of orga-
nizational and educational dimensions but requiring a
constant reconciliation with the technological feasibility.
In a dialectical process, progressive approximations were
made in order to ensure that the LA reflects a balanced
management of the three dimensions of the triangle.

Based on the Technology Integration Matrix, of the Florida
Center for Instructional [16], an instrument was developed for
the positioning of the CUs and Schools/Faculties in five levels:
entry, adoption, adaptation, immersion and transformation.
Fig. 5 shows the generic descriptions of the five stages [17].

Figure 6 depicts the dimensions considered in the model,
which roam the main valences offered by the LCMS, and
cumulatively consider dimensions identified in published
studies as fundamental to the development of teaching activities in online environment, namely: access, collaboration, content, multimedia content, delivery of papers and work assessment. These dimensions have different weights in calculating the overall positioning of the CU, according to the importance they assume in the teaching and learning process [17].

In addition to the data from the Blackboard reports, it was necessary to measure the perceptions of students on the degree of integration of the LCMS in the learning process. Therefore a scale was created [18] with the same six dimensions present in the LCMS reports, and to the dimension communication which was added, which does not appear in the Blackboard reports due to issues related to technological feasibility, as the system does not allow the counting of items on this scale. Thus, it was not possible to integrate this dimension in the backoffice of the LCMS reports.

C. Results Achieved

The results achieved were reflected in the implementation of a system for analyzing and extracting data from the LCMS – consequence of the design made in backoffice – and in the development and validation of a scale on the integration of the campus in the teaching and learning process [18]. In Figure 7 one can see the report of the CU “Information and Technological Systems, of the Faculty of Economics and Management”, at Católica-Porto, in which Figs. 5 and 6 are based. Following this, the stage of the positioning is found for each dimension, which is calculated by taking into consideration the following points of maturation: introduction [0-20%]; adoption [20%–40%]; adaptation [40%–60%]; immersion [60%–80%]; transformation [80%–100%]. Similarly, from the results of the points of maturation of the different dimensions, the overall stage of the CU is calculated. The positioning of the CU in the

\[
\sum = (\text{access} \times 0.1) + (\text{collaboration} \times 0.25) + (\text{contents} \times 0.1) + (\text{multimedia contents} \times 0.25) + (\text{Delivery of papers} \times 0.15) + (\text{evaluation} \times 0.15). \]

Figure 8 shows a radar chart of the system – also a result of the backoffice’s design – which allows a detailed reading of the integration of the LCMS in the CU on the different dimensions. In the case of the CU in analysis, it can be concluded that its global positioning in the stage immersion results from the high level of integration of the dimensions collaboration, content and evaluation (maturity points = 100%) in the teaching and learning process. In contrast, the contribution of dimensions delivery of papers and multimedia content is more modest, which shows that there is still a long way to go in these fields (maturation = 50%). Regarding the dimension dynamics of access it can be said that it is at a low level, revealing a very low number of hits/students/week.

At the present moment the LA program is still being developed and implemented; however, it is possible to give a general perspective of the integration of the LCMS in the process of teaching and learning at Católica-Porto (Fig. 9), where one can see (from the reports of the LCMS) that all Faculties have most CUs positioned in the introduction stage, and that the number of CUs positioned on higher levels is of little significance. Currently, a form of aggregation of the system’s data is under study with the students’ view (collected by applying a scale to the integration and to the importance of the different dimensions of the LCMS in the teaching and learning process [18]).

IV. ANALYTICS IN PRACTICE: ACADEMIC ANALYTICS OF PEDAGOGICAL ACTIVITY

A. Objective

The initial focus on the use of Analytics in education is usually on the use of the LCMS. Several researchers believe this approach is relevant and part of a solution, a path that is important to tread, but insufficient [9]. The penetration of technology in different areas of HEIs gives rise to various technological subsystems, each with its information cell (data source), containing a multitude of genes (data), which can be organized into chromosomes (organized information) within the cell. This challenge of data organization within a single
The technological system – in this case the LCMS – was the procedure shown in the previous section.

The next step is to integrate these data sources in a technological system – the Academic Analytics service – so as to obtain a single large database – Big data base – for the integrated processing of data from different sources. Recovering the metaphor, it is important that the different cells are not reduced to unicellular beings, each with its island of information, but to move towards the construction of multicellular organisms of information which are more complex and consistent.

The aim of this section is to outline the methodology and describe the design of the architecture of the Academic Analytics of Católica-Porto, which is intended to be aggregator of multiple data sources of computer subsystems in use in the institution, allowing to quickly display indicators on pedagogical activity at different levels of the organization, and having the dimension use and integration of the LCMS in the teaching learning process associated.

B. Notes on Methodology and Architecture Design

Figure 10 shows the overall architecture of what may be the Academic Analytics of pedagogical activity at Católica-Porto. This system has as input all data sources with information on the educational activities at different levels of the organization (Católica-Porto > Faculty > Department > Course > CU) and on the actors most directly involved in the teaching and learning process (teachers and students).

After the overall goal of the Academic Analytics was defined, a survey took place on the institution’s technological subsystems that contain data related to the pedagogical activity. The procedure started by using existing data and checking the type of information that one can obtain from them, and then we materialized the more specific goals [19], i.e., the information intended to be withdrawn from the Academic Analytics.

Thus, the following data sources were identified:

- **Blackboard** that has data from the reports on the access and use of the platform, and which are embedded in the platform and in the LA’s data, presented in Section III (matrix of the LCMS’ integration in the teaching and learning process);
- **Google Analytics** with data on the access and activity in the LCMS;
- **The Mobile Analytics**, which provides data on the mobile access to the platform;
- **The SAEIQ**, with data about the CUs and teachers’ assessment obtained through a questionnaire given to students;
- **School Administrative Management System (SOPHIA)**, with demographics, academic activity and student outcomes.

All this data is crisscrossed in the Academic Analytics – it is important to choose a robust, reliable technological platform that ensures data integrity and interoperability of sources – and there is an important challenge we must face: integrating multiple databases with different formats and organized in different ways. This challenge involves work on two levels:

- **Technological level** – to ensure the interoperability of autonomous and heterogeneous data sources located in different hardware and software environments;
- **Organizational level** – to establish correspondences between codes and classifications of the elements of the different data sources, so that the system realizes that different codes may refer to the same element. An example: CUs have different codes in the Blackboard, SOPHIA and SAEIQ and it is important that the system treats them as being the same.

The right corner of Figure 10 shows the synthesized output of the system and its role in supporting decision making, which can be translated into reports with tables that can be used in planning, process control and in the definition of intervention actions in the different scales of the HEI.
same reasons for adopting Analytics Systems: to increase operational efficiency and financial dividends; to expand its strategy; and to establish new models more quickly during periods requiring changes [2].

This parallelism between the business world and the world of HEIs, between Business Analytics and Academic Analytics, is embodied in the statement: “Luan [20] showed how many critical issues of the area of business are present, in a parallel way, in higher education. For example, while the business may ask ‘Which are my most loyal consumers?’ or ‘Who is able to increase the volume of purchases?'; a HEI may ask ‘Which are the most persistent students in the institution throughout their training?’ or ‘who are the alumni who can potentially make larger donations?’” (free translation done by the authors) [2].

Fig. 11 depicts the data sources that can be aggregated in the Academic Analytics program, showing the parallelism with Business Analytics.

Some possible examples of the Academic Analytics operationalization in a HEI similar to Católica.Porto:

- Crossing the sources of investigation – which provide information on scientific production in the different research centers of the HEI – with financial sources, it may allow indicators to be extracted such as: number of applications for projects/amounts financed; financial amounts made available by the institution to research centers/scientific production indicators.
- Crossing data from administrative sources with financial sources may allow calculating the ratio cost of a particular course and tuition revenue.
- Information extracted from alumni sources (former students) and employability can be used for marketing purposes by the institution.

VI. CONCLUSIONS

The use of Analytics in Education is an emerging issue, which has high potential for the management of HEIs at different scales and in a multitude of dimensions. In this paper we have used the genome metaphor to put into the equation the importance and implications of Analytics: genome, originating from the Greek (geo = which forms and ma = action), is the result of thousands of genes (DNA) containing the necessary information for the expression and adaptation of the organism to the environment. By analogy, the data (Big Data = genes of the organization), featured increasingly in a larger volume of the different technological subsystems of the HEIs, if properly mapped may help to understand the activity within
the organization (geo = which forms), enabling a more informed decision making when defining policies and actions (ma = action).

In this paper, the application of Analytics in HEIs was analyzed, from the micro level (student/CU) to the scale of global business management (macro scale). Starting from the reality of the Católica-Porto, the use of Analytics was shown with increasing levels of integration and scope: i) LA centered in the use and integration of the LCMS in the teaching and learning process. This LA project is listed in the Horizons Projects - The New Media Consortium [21]; ii) the presentation of an architecture with a focus on teaching and learning aspects which provides for the integration of different sources with information about this activity; iii) prediction of the future Academic Analytics with all information relevant to the management of HEIs (integration of teaching supplies, but also research, administrative, financial, i.e. all sources of interest to the organization’s management).

The implementation of an effective Analytics system that enables mapping the genome of Católica-Porto, addressing the real information needs of the institution, involves: i) the definition – by decision makers and other stakeholders of HEIs – of the dimensions and data that should integrate the Academic Analytics, ii) the development of a technology-based system that allows aggregating a variety of sources.

References


Sérgio André Ferreira received the M.Sc. degree in educational sciences with a specialization in educational computing. He is currently pursuing the Ph.D. degree in educational sciences. His research interests focus on the technology-enhanced learning environments, with a focus on the theme of learning analytics and academic analytics.

António Andrade is a Professor with the Faculdade de Economia e Gestão, Universidade Católica Portuguesa-Porto, Porto, Portugal. He received the Ph.D. degree in technologies and information systems and the M.Sc. degree in computer management. He is the Director of the Masters in Information and Documentation. His research interests focus on the technology-enhanced learning environments.