Critical Control Points Course Detection Methodology at University Education

Martín Pratto Burgos, University of the Republic, Uruguay

The Barcelona Conference on Education 2023
Official Conference Proceedings

Abstract
The research presents an innovative study based on adding value to the university educational process. Currently, there is no clear definition to detect at which stage of a career students lag behind after failing a specific subject or activity. An educational innovation proposal, based on the extrapolation of the Hazard Analysis and Critical Control Point System coming from the industrial setting to the university education context, is presented: the Critical Control Point Detection System concerning the curriculum and students' trajectories. The purpose of this work is to generalise a Critical Control Points Detection Methodology, in higher education, through the localization of courses that have greater incidence in the curriculum and identifying the negative result through student performance due to failure. For the curriculum analysis, indicators related to centrality measures were built with a directed network based on the curriculum subjects. Courses were classified as Critical-Control-Point candidates when the numerical values of measures exceed or equal to a minimum threshold value. For student performance, a measure of course failure was used. Those candidate courses exceeding a predefined failure threshold were classified as Critical Control Points. Critical-Control-Points courses implied at least one semester of lagging for any student failing on it. The Critical Control Point Detection Methodology was performed for Chemical Engineering and Law curriculums at the University of the Republic (Uruguay) using a (55 ± 5)% threshold of failure for more than two years during five years. The Critical-Control-Point courses match previous perceptions from university lecturers.

Keywords: Education, Critical Courses, University, Computational Methodology

Introduction

The article is based on a university generalisation research carried out in the Innovation Management Master's degree belonging to the Engineering School of the University of the Republic, Uruguay (Pratto Burgos, 2023).

The University of the Republic is an autonomous public institution that offers higher education opportunities to the general population. Furthermore, it promotes and protects research conducted in both scientific and artistic fields. The University is composed of different Schools (Universidad de la República, 1958).

The increasing popularity of machine learning can be observed across diverse fields according to today's Fourth Industrial Revolution. How well a machine-learning solution works depends on the working type of data and how well the learning algorithms perform (Sarker, 2021).

Machine learning (ML) tools can assess an individual's level of understanding, detect lacking-of-knowledge areas, and offer real-time support. Furthermore, ML is capable of identifying areas where the teacher-student ratio is unbalanced, allowing for the development of educational programs that cater to the larger student population. ML has numerous advantages that demonstrate its transformative impact on the education field (Jagwani, 2019).

Incorporating ML techniques provides an innovative means of investigating data in computing education. Researchers will have the ability to employ algorithms for discovering novel relationships and creating adaptable models (Zahedi, et al. 2020).

At the University, students strategize their trajectories from admission to graduation. In this way, the student trajectory implies the path a student follows during their time in the Faculty.

The theoretical student trajectory refers to the ideal learning journey based on the suggested curriculum. When examining students' trajectories it is important to take into account the date of data query because academic student progress is changing over time. Student's academic performance is determined by the number of courses they have completed by passing or failing them.

Students' real trajectories are different from theoretical ones because students' performance does not match what the curriculum teaches them. This causes them to pursue individualistic (real) academic trajectories. Deviations from curriculums are not always associated with failure. These terms are employed to characterise diverse social and educational settings (Directores que hacen escuela, 2005).

By examining the students' real trajectories, we can identify similar or probable paths that align with the theoretical trajectory. However, alternative paths that differ from theoretical trajectories can also be discovered (Terigi, 2009).

When progressing through university, students might face difficulties that cause them to diverge from their theoretical trajectory. Failing in specific, essential, and critical courses for their educational progression is one of the hurdles that can be faced avoiding to lag for their degree.

Currently, universities lack a meaning to identify when a student is unsuccessful in a crucial course, hindering their advancement in their academic pursuits. Without a distinct definition of critical courses, universities cannot accurately assess instances where students are not progressing in the curriculum. This leads to many guesses that are not proven to be true.

A Critical Control Point (CCP) is a term derived from the industrial and service food sectors. It refers to "the phase at which intervention may be taken to prevent or mitigate a food safety issue or reduce its level of hazard". The procedural method also includes validating the system's ability to accurately identify CCP effectiveness (Food and Agriculture Organization of the United Nations [FAO], 1997).

A CCP course in university means that by failing on it, you have lagged for at least one semester in the curriculum. In the article, the most relevant Analysis and Critical Control Point System (HACCP) phases in the food industry are applied in the educational sector, following the two main principles, second and third from FAO (1997): identifying CCP and setting critical limits.

Dueñas (2016) studied CCPs in the education field from a subjective perspective, describing them as highly complex subjects in university. When the average student number exceeds 15, Dueñas affirmed that highly complex subjects should include a failure rate of over 20%.

The research proposes to devise a methodology to locate CCP courses offered in university institutions. Particularly, the following research question is explored based on both principles of HACCP: how do we define a CCP course at university based on the measures obtained from the generalised methodology, and how well it aligns with lecturers' insights?

To organise and gain insight into the methodology presented in this article, based on a systemic point of view, two perspectives are separately analysed: the curriculum and the university students' trajectories performance.

Theoretical Framework

Networks and Centrality Measures

Courses in a curriculum can be visualised as a network, reflecting the prerequisites necessary for enrolment. For example, you might need to pass a course before you can take the next level of that course. As more difficult a course is, the more prerequisite courses are required to be taken before it. We can comprehend and organise this prerequisite system by creating a detailed strategy or visual representation.

According to Kolaczyk and Csárdi (2014), a network can be seen as a visual representation that illustrates the interconnectedness of various components within a complex system. A network is a mathematical concept consisting of vertices, also called nodes, and edges, lines connecting these vertices. In a directed network an edge is represented by an arrow going from one node to another in a specific direction.

Networks can aid in comprehending the functionalities and processes of diverse entities by measures' estimation. "Centrality measure is such an important index because it indicates which node takes up a critical position in one whole network" (Zhang & Luo, 2017).

In the research, the centrality measures used are *degree*, *closeness*, *eigenvector*, and *betweenness*.

Degree centrality quantifies the number of connections with other nodes in the network. Additionally, the significance of a node in a network is quantified through Closeness centrality, which entails the distance summation between one node and other nodes (Zhang & Luo, 2017).

The closeness centrality concept concerns the proximity between a node and its neighbours' elements. The shortest average connection distance with all other nodes in the network is associated with a node having a high numerical value (ArcGIS Pro 3.0.).

Determining the influence of a specific node in a network can be achieved through the eigenvector centrality. The measure expresses how important nodes are connected to others in the network. By computing this centrality measure, we can determine whether specific clusters possess a greater level of influence (ArcGIS Pro 3.0.).

Finally, betweenness centrality indicates how well a node is linked in a network. When a node serves as a crucial pathway for other nodes to communicate, connect, transport, or transact, it gains significance and obtains a high numerical value of betweenness centrality (Zhang & Luo, 2017).

Curriculum

The curriculum outlines the content, purpose, methodology, and timeline for students to acquire knowledge. The overarching aim of a well-designed curriculum is to facilitate students in obtaining knowledge, skills, and values, associated with capabilities and competencies, that will enable them to live meaningful and productive lives. Indicators of a successful curriculum encompass the progress made by students in learning and their subsequent application of acquired knowledge to enhance personal, social, physical, cognitive, moral, psychological, and emotional development (Stabback, 2016).

Methodology

CCP Course Detection

When creating the CCP course detection methodology, we examine the indicators' estimation in the curriculum and in the students' trajectories. We aim to comprehend the significance of both indicators' perspectives occupied in the university context.

According to the second principle of the HACCP plan, the initial step is to identify the CCP (FAO, 1997). To follow the step, courses provided by the curriculum are mapped by using a network. Such courses are represented by nodes and linked by arrows, among the other courses, according to the prerequisite classes required for enrolment.

The curriculum-based network assists in organising and scheduling the necessary courses before enrolment. For example, if a student wishes to enrol in a particular course, the network will furnish information regarding prerequisite courses that the student must finish before becoming eligible for the desired course (represented with arrival arrows to the target course).

The four centrality measures, applied to each network node, help us identify courses as CCP candidates.

According to the third principle of the HACCP plan, the methodology needs to set up threshold limits (FAO, 1997). The numeric value that must be achieved for each centrality measure is determined by applying the elbow rule to courses taking part in the curriculum.

The elbow rule consists of a graph plotting the approximation sum of squared errors on the y-axis and the values on the x-axis. The presence of distinct clusters will be appreciated in a noticeable decline in the graph (Schubert, 2022).

In the research, for each centrality measure the elbow rule is used with the numerical centrality value plotted in the y-axis, arranging them in diminishing order. The "elbow" is the term used to refer to the point where the slope of the plot changes prominently. This point additionally denotes the numerical boundary. Courses exceeding a minimum threshold limit value (TLV) in three or four centrality metrics are regarded as CCP candidates.

Finally, when evaluating the students' trajectories, the methodology will rely on the measure of course failure ratio in completing a CCP candidate course throughout 5 years. The ratio involves assessing the percentage of students who did not successfully complete the course/exam at a particular course edition. The research includes the period spanning from 2015 to 2019.

The TLV course failure ratio is (55 ± 5)%. This conclusion was reached by examining twelve specific courses in the Chemical Engineering curriculum degree of the Engineering School. As a result, if a CCP candidate course has a course failure ratio exceeding (55 ± 5)% for more than two years within 5 years, it will be concluded as a CCP course (Pratto Burgos, 2023).

In summary, the CCP Course Detection Methodology takes into account the following stages:

1. *Create a network using the curriculum courses* (nodes) and the relationships between them based on the courses before enrolment. Thereby, in order to enrol in a desired course, students must have successfully finished the required prerequisite courses: arrivals edges in the network proceeded from those courses.

2. *Quantify the four centrality measures:* degree, closeness, eigenvector, and betweenness from the curriculum-based network.

3. *Find strong performance courses in three or four centrality measures*, according to the reference threshold value.

Courses complying with the third stage will be identified as CCP candidates.

4. *Determine the course failure ratio*, during 5 years, only for the CCP candidate's courses.

5. *Consider CCP courses* when the course failure ratio exceeds (55±5)% in more than two years within the analysed curriculum.

CCP courses will be finally obtained in the fifth stage and they may be strongly advisable to consider by students during the enrolment procedures.

Implementation

This section explains the CCP Course Detection Methodology introduced in Chemical Engineering and Law curriculums corresponding at the University of the Republic in Uruguay (Consejos de las Facultades de Ingeniería y Química, 1999; Frezelmi, 2021).

To utilise the methodology, it requires accessing essential data from the Chemical Engineering degree curriculum and its prerequisite courses schedule. The University's Administrative Central Office manages course enrolment taking part in the curriculum and their required pre-requisite courses for enrolment (Sistema de Gestión Administrativa de la Enseñanza [SGAE]).

The Statistical Software R is used to process data. The Windows operating system allows users to download the software's latest version at no cost. In the research, R version 4.1.1 was used through RStudio (integrated development environment) version 2021.9.1.372. The R programming environment offers a wide range of tools and resources in the data analysis field to facilitate data processing (Fernández Casal, et al., 2022).

To create the network and quantify centrality measures, the "igraph" package was used in the R environment. To maintain a consistent database format, the "read.excel", or "read.csv", function was applied to convert the spreadsheets from Excel, or CSV, format into data frames. A table-like structure called data frames is utilised to organise data encompassing rows and columns.

The students' trajectories perspective requires them to evaluate the measure of the course failure ratio in completing a CCP candidate course throughout 5 years. Individuals under observation are students who are presently actively pursuing their studies in Chemical Engineering and have completed CCP candidate courses.

SQL queries are used in the Trebol-fuentes platform, managed by the Central Computer Service of the University of the Republic, to retrieve data on students' performance from student-activity databases. The student's ID, course name, grade, course or exam activity, and activity date are crucial pieces of information to be gathered. Academic data up to April 2022 is analysed in this research (Servicio Central de Informática [SeCIU], 2019).

The query code follows the SQL rules by using the "Select", "From", and "Where" parameters in its structure. The "Select" parameter mentions the required variables for the query: student's ID, course name, grade, course or exam activity, and activity date. Information is retrieved from specific tables by using the "From" parameter. The "Where" parameter is associated with the conditions of the query.

At the University of the Republic Engineering School courses are graded on a scale from 0 to 12. Failing the course and being unable to take the exam is the consequence for students who receive a grade ranging from 0 to 2. Achieving a grade between 3 and 5 will enable students to pass the course, granting them the opportunity to take the exam. Provided students' grade falls between 6 to 12, they will successfully complete the course and be exempted from the exam (Peláez & Collazo, 2017).

The higher-education-generalised methodology involves contrasting the Law Degree, at the University of the Republic, with the Chemical Engineering Degree. For the 2016-approved

Law curriculum, the CCP Course Detection Methodology was applied to determine whether the centrality measures align with those employed in the Chemical Engineering curriculum.

Findings and Discussion

To assess the importance of a node in a network, its distinctive qualities are taken into consideration using the centrality measures. The characteristics can be given by the use of local information, e.g., degree, or global network information, e.g., closeness, betweenness, and eigenvector centrality, requiring the entire structure of the network (Akrati Saxena &, Sudarshan Iyengar, 2020).

The TLVs for each centrality metric, programmed in the R environment through the network, are as follows: degree has a threshold of 4, closeness has a value of 0.63, 0.73 for eigenvector, and 22 for betweenness. These measurements apply to both the Chemical Engineering and Law curriculum.

The centrality-measures-based distinctions among the CCP candidate courses are considered significant in the pre-requisite courses schedule, characterised in the network, only if they surpass the TLV. The determination of the CCP candidate course power relies on the positioning of the courses in the network, as determined by centrality measures interpretation.

Four courses in the Chemical Engineering curriculum surpassed the TLV of at least three different centrality measures. About the Law curriculum, there exists a single course that has exceeded the threshold limit value in four centrality measures. Consequently, the methodology's curriculum perspective identifies specific courses as CCP candidates.

Through a successful performance in CCP candidates' courses, students can thrive academically without lagging for in their studies and advancing in the curriculum at a suitable rate. Therefore, it is essential to investigate the students' lack of success in those CCP candidate courses through the course failure ratio.

The measure of course failure ratio in completing a CCP candidate course, throughout 5 years in the Chemical Engineering curriculum, is illustrated in table 1. The table also includes students who were unsuccessful in passing the CCP-candidate-course exam.

	Course edition year									
	2015		**2016**		**2017**		**2018**		**2019**	
	n	CFR (%)	n	CFR (%)	n	CFR (%)	n	CFR (%)	n	CFR (%)
Course 1	10	**77**	14	**67**	27	**63**	31	**74**	33	**55**
Course 2	16	**53**	25	**64**	37	**49**	48	**63**	42	**44**
Course 3	32	**89**	49	**86**	69	**77**	74	**63**	79	**56**
Course 4	10	**44**	10	**44**	8	**30**	9	**22**	16	**25**

Table 1: CCP-candidates-courses failure ratio (CFR) and student numbers who fail each CCP-candidates' courses (n) in the Chemical Engineering curriculum.

Database academic information lower than 5 years, from the closing-date database query, might come across a reduced number of current students. It can be explained that when graduating as an Engineer, students are required to dedicate 5 years to their studies at the Engineering School. Consequently, a 5-year period was selected to simplify the process of locating current students within the database avoiding their graduation. Due to a health emergency, the university had to suspend on-site activities, thus excluding the years 2020 and 2021 from the analysis.

From 2015 to 2018, table 1 indicates that courses 1 and 3 have maintained scores surpassing the TLV course failure ratio (55 ± 5)% in four editions. In contrast, course 2's TLV has not been exceeded for over two years, and for course 1, the TLV has not been surpassed in any year. Therefore, courses 1 and 3 are CCPs that play a significant role in the Chemical Engineering curriculum. The methodology's results indicate that failing to pass both courses will result in a student being lagged for at least one semester and unable to continue with the curriculum.

Confirmed by Chemical Engineering lecturers from Engineering School, courses 1 and 3 are identified as the crucial courses in the Chemical Engineering curriculum. They accomplished this conclusion by comparing the methodology's findings with the lecturers' insights. The methodology's adjustment was changed based on the lecturers' expertise. The iteration process consisted of two phases: modifying the centrality measures TLV and introducing the TLV on the course failure ratio (Pratto Burgos, 2023).

According to the CCP Course Detection Methodology and the lecturers' insights, a CCP course definition can be inferred at the university.

For a CCP course to be effectively detected it is necessary to exceed, or equal to, centrality measures (degree, closeness, eigenvector, and betweenness) ideally at 4, 0.63, 0.73, and 22 respectively, based on the pre-requisite course curriculum network. Furthermore, the CCP course must demonstrate a course failure ratio over (55±5)% for at least three years in a five-year term.

The CCP Course Detection Methodology and its definition are subject to certain limitations. The TLV course failure ratio fluctuates according to the specific higher-education institution. Nonetheless, the TLV proposed in this article serves as a basis for authorities to strategize improvements for every institution. Additionally, people responsible for using the methodology should consistently access students' performance databases. It allows them to continuously implement it and effectively monitor every aspect.

Conclusion

The CCP Course Detection Methodology was created to help university institutions detect crucial courses (CCP). They were based on the second and third principles of the HACCP plan, which is used in the food industry. The purpose of the methodology was to identify CCP courses in which students were lagging for at least one semester at university.

By constructing a curriculum-based network, the CCP Course Detection Methodology makes it possible to locate nodes (courses) that possess substantial significance using local and global centrality measures like degree, closeness, eigenvector, and betweenness within the network. Courses will be deemed significant if their centrality measures are equal to or surpass the TLV. Only those courses achieving high scores in three or four centrality measures will be eligible for consideration as CCP candidate courses.

According to the students' trajectories perspective, the TLV course failure ratio represents the percentage of students who did not achieve a passing grade in their courses. Only courses that consistently score in a percentage above (55±5)% for at least three years, in a five-year term, will be considered as CCPs.

The curriculum-based network to locate CCP candidate courses was implemented for the Chemical Engineering and Law curriculum, both from the University of the Republic, Uruguay. Following that, the significant interconnectedness of each course in the network was evaluated through the centrality measures. The implementation evidenced that five courses, four from the Chemical Engineering curriculum and one from the Law curriculum, are strong contenders for being regarded as highly significant in each curriculum due to their strong performance in more than three centrality measures.

Only two courses, of the four CCP candidates courses from the Chemical Engineering curriculum, exceed the TLV course failure ratio of over (55 ± 5)%. Hence, both courses are identified as CCP in the curriculum assessed between 2015 and 2019. The methodology's result aligns with the Chemical Engineering lecturers' insights. CCP courses are considered important components of the curriculum according to their lecturers' experience.

The methodology's result and the lecturers' insights lead to a CCP definition. It adjusts a course exceeding, or equalling to, centrality measures values (degree, closeness, eigenvector, and betweenness) at 4, 0.63, 0.73, and 22 respectively, through the curriculum-based network, and it must demonstrate a course failure ratio over (55±5)% for at least three years in a five-year term.

The CCP Course Detection Methodology limitations determine its functionality and implications. Different higher-education institutions have different TLV ratios of course failure. Nevertheless, the (55±5)% TLV might be utilised as a basis to begin planning curriculum improvements.

References

ArcGIS Pro 3.0. Usar análisis de centralidad [Online] https://pro.arcgis.com/es/pro-app/latest/help/analysis/link-charts/centrality.htm Consultado 9 de diciembre de 2022.

Akrati Saxena &, Sudarshan Iyengar (2020). *Centrality Measures in Complex Networks: A Survey*. https://doi.org/10.48550/arXiv.2011.07190

Consejos de las Facultades de Ingeniería y Química (1999). *Plan de estudios 2000 para la carrera Ingeniería Química*. Universidad de la República. Montevideo, Uruguay.

Csardi, G., Nepusz T. (2006). The igraph software package for complex network research. Inter-Journal, Complex Systems 1695. https://igraph.org

Directores que hacen escuela, en colaboración con Joana López (2015). *De la trayectoria en singular a las trayectorias en plural*. Organización de los estados Iberoamericanos para la educación, la ciencia y la cultura. Buenos Aires, Argentina.

Dueñas, M. (2016). *Estilos de aprendizaje, estrategias didácticas y uso de tic en la atención a las asignaturas de alta complejidad de la Universidad de Tarapacá*. MSc. Tesis. Universidad Andrés Bello.

Fernández Casal, R., Roca Pardiñas, J., Costa Bouzas, J. y Oviedo de la Fuente, M. (2022). *Introducción al Análisis de Datos con R*. ISBN: 978-84-09-41823-7.

Food and Agriculture Organization of the United Nations (1997). *Sistema de análisis de peligros y puntos críticos de control (HACCP) y directrices para su aplicación.* [Online] https://www.fao.org/3/y1579s/y1579s03.htm. Consultado 11 de noviembre de 2022.

Frezelmi (2021). *Plan 2016 Abogacía-Notariado* [Online] https://frezelmi.com/abogacia-notariado/ Consultado: 10 de marzo de 2023.

Hadley Wickham and Jennifer Bryan (2019). readxl: Read Excel Files. R package version 1.3.1. https://CRAN.R-project.org/package=readxl

Jagwani, Anjali (2019). A review of machine learning in education. J*ournal of Emerging Technologies and Innovative Research (JETIR)*, 6(5), 384-386. ISSN-2349-5162.

Kolaczyk, E. and Csárdi, G. (2014). *Statistical Analysis of Network Data with R*. Springer, New York, 2a edición.

Peláez F., Collazo M. (2017). *Propuesta de actualización de la escala de calificaciones de la Udelar*. Unidad Académica de la Comisión Sectorial de Enseñanza, Pro Rectorado de Enseñanza. Universidad de la República.

Pratto Burgos, M. (2023). *Sistema de detección de puntos críticos de control para una carrera de ingeniería* [Tesis de Maestría, Facultad de Ingeniería, Universidad de la República]. Repositorio Institucional Colibrí, UdelaR.

R Core Team (2021). R: *A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Sampieri, R., Collado, C., and Lucio, P. (2014). *Metodología de la investigación*. McGraw-Hill, México D.F, 6a edición.

Sarker, I.H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN COMPUT. SCI,* 2, 160. https://doi.org/10.1007/s42979-021-00592-x

Schubert, E. (2022). Stop using the elbow criterion for k-means and how to choose the number of clusters instead. *ACM SIGKDD Explorations Newsletter,* 25(1), 36–42 https://doi.org/10.1145/3606274.3606278

Servicio Central de Informática (2019). Proyecto Trébol. Recuperado de la base de datos de trebol_fuentes v1.0.1.

Sistema de Gestión Administrativa de la Enseñanza. Universidad de la República. Servicio Central de Informática, versión 15.3.1. [Online] https://bedelias.udelar.edu.uy/ Consultado 2 de marzo de 2022.

Stabback, P. (2016). What Makes a Quality Curriculum? In-Progress Reflection on Current and Critical Issues in Curriculum and Learning. International Bureau of Education. *UNESCO*, (2), 8.

Terigi F. (2009). *Las trayectorias escolares, del problema individual al desafío de política educativa.* Ministerio de Educación de la Nación. Buenos Aires, Argentina.

Universidad de la República (1958). *Carta Orgánica de la Universidad de la República Ley 12.549* de 16/X/1958 - D.O. 29/X/1958.

Zahedi, L.; Lunn, S. J.; Pouyanfar, S.; Ross, M. S.; Ohland, M. W. (2020). *Leveraging Machine Learning Techniques to Analyze Computing Persistence in Undergraduate Programs*. American Society for Engineering Education. Paper presented at 2020 ASEE Virtual Annual Conference Content Access, Virtual Online. 10.18260/1-2—34921.

Zhang, Junlong & Luo, Yu (2017). *Degree Centrality, Betweenness Centrality, and Closeness Centrality in Social Network*. Advances in Intelligent Systems Research, 2nd International Conference on Modelling, Simulation and Applied Mathematics (MSAM 2017), 132, 300-303. Published by Atlantis Press. 10.2991/msam-17.2017.68.

Contact email: martinpburgos@gmail.com