IDENTIFYING AND COMPARING COURSE-TAKING PATTERNS IN STEM EDUCATION

LEYLA VILLARROEL SALAS

Thesis submitted to the Office of Research and Graduate Studies in partial fulfillment of the requirements for the degree of Master of Science in Engineering

Advisor:
MARCOS SEPÚLVEDA FERNANDEZ

Santiago de Chile, March 2022

© MMXXII, LEYLA VILLARROEL SALAS
IDENTIFYING AND COMPARING COURSE-TAKING PATTERNS IN STEM EDUCATION

LEYLA VILLARROEL SALAS

Members of the Committee:
MARCOS SEPÚLVEDA FERNANDEZ
JORGE MUÑOZ GAMA
JORGE MALDONADO MAHAUAD
PABLO PASTÉN GONZÁLEZ

Thesis submitted to the Office of Research and Graduate Studies in partial fulfillment of the requirements for the degree of Master of Science in Engineering

Santiago de Chile, March 2022

© MMXXII, LEYLA VILLARROEL SALAS
Gratefully to my family, friends,
and to my advisor Marcos
ACKNOWLEDGEMENTS

First of all I would like to thank my thesis advisor, Marcos Sepúlveda. He has been very supportive from the first minute I decided to start the Master, encouraging me not only in the things related to the research, but also in other projects I have decided to take. His constant smile and light sense of humor in every meeting made my research journey infinitely more pleasant. Also, I am so grateful for his understanding when I was not in my highest point, his patience and all the encouraging words he gave me that helped me to finish this research.

I also want to thank Juan Pablo Salazar, whose knowledge and previous work helped me to improve and polish my paper. He was always willing to contribute and to give clear and assertive feedback.

Thanks to Jorge Muñoz, who encourage me to learn more about Process Mining and to join this program. His support during the first semester of the master was extremely helpful. Also, I would like to thank him for his sponsorship to my internship abroad.

Finally, I want to thank my family, my boyfriend and my friends. They have supported me from the beginning, always being by my side, encouraging me when I needed it and helping me to distract myself when it was necessary. This experience would never have been the same without their company.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS iv  
LIST OF FIGURES vii  
LIST OF TABLES viii  
ABSTRACT ix  
RESUMEN x

## 1. INTRODUCTION 1  
1.1. Context .................................................. 2  
1.2. Research thesis ........................................... 4  
1.3. Background ................................................ 5  
1.3.1. Curricular analytics .................................... 5  
1.3.2. Course-taking Plan ...................................... 5  
1.3.3. Course-taking Pattern ................................... 6  
1.3.4. Clustering ............................................... 7  
1.4. Materials and methods .................................... 10  
1.4.1. Data pre-processing ..................................... 10  
1.4.2. Clustering ............................................... 11  
1.4.3. Analysis ................................................ 12  
1.4.4. Comparing academic performance between course-taking patterns .. 12  
1.5. Research’s paper ......................................... 13

## 2. ARTICLE 14  
2.1. Introduction .................................................. 14  
2.2. Related work .................................................. 16  
2.3. Course-Taking Pattern Analysis Method (CTPA) .............. 18  
2.3.1. Data pre-processing ..................................... 20
2.3.2. Clustering ........................................... 21
2.3.3. Analysis ........................................... 22
2.4. Case study ........................................... 23
  2.4.1. Data pre-processing............................ 24
  2.4.2. Clustering........................................ 25
  2.4.3. Analysis ........................................ 26
2.5. Discussion .......................................... 35
2.6. Conclusions ....................................... 38
Acknowledgment ....................................... 39

3. CONCLUSIONS ....................................... 40
  3.1. Conclusions ....................................... 40
  3.2. Limitations ....................................... 41
  3.3. Future work ...................................... 42
  3.4. Acknowledgement ................................ 43

REFERENCES ........................................... 44
# LIST OF FIGURES

1.1 Clustering stages .................................................. 7

2.1 Stages of the methodology. ........................................ 19

2.2 Silhouette index by number of clusters. ............................ 26

2.3 GPA distribution in courses of the major for each cluster. .... 30

2.4 Distribution and box-plot of the difference between before and during the major in GPA for each cluster. The dotted line indicates where the difference is 0. .... 32

2.5 Cluster n°0: Comparison of the distribution of pass rate before and during the major. ................................................. 34

2.6 Cluster n°1: Comparison of the distribution of pass rate before and during the major. ................................................. 34

2.7 Relative frequency histogram of the completion time in semesters for each cluster. The dotted line represents the mean for each cluster. ................. 35
### LIST OF TABLES

1.1 Students’ Vectors before processing. .......................................................... 11

1.2 Students’ Vectors after processing. .............................................................. 11

2.1 Students’ Vectors Represented by the Semester in Which the Courses Are Passed. ............................................................... 21

2.2 Students’ Vectors Represented by the Normalized Semester in Which the Courses Are Passed. ............................................................... 21

2.3 Students’ Vectors Represented by the Normalized Semester First Semester and the Order in Which the Courses Are Passed ............................................................... 21

2.4 Computer Science Major Courses .............................................................. 24

2.5 Computer Science Major Course-Taking Plan .......................................... 24

2.6 Course-Taking Patterns Identified with the Number of Students That Follows Each and Their Similarity Metric ............................................................... 28

2.7 Comparison Between the Course-Taking Plan and the Course-Taking Patterns ............................................................... 29

2.8 Sub-patterns Found in Course-Taking Patterns .......................................... 30

2.9 GPA Tukey Test Results .............................................................................. 31

2.10 GPA Paired T-Test Results ........................................................................ 32

2.11 Pass Rate Paired T-Test Results .................................................................. 33

2.12 Completion Time Tukey Test Results .......................................................... 35
ABSTRACT

Universities suggest an ideal course-taking plan to their students, but they do not have certainty whether the students follow their suggestions. A new approach is needed to identify the real course-taking patterns followed by the students and understand how these patterns affect their academic performance.

This research presents the Course Taking-Pattern Analysis (CTPA) method, a method to discover the most frequent course-taking patterns followed by students to complete their majors using clustering algorithms and statistical analyses to make comparisons. This method was applied to analyze the course-taking patterns of the computer science major in a Chilean university. Six course-taking patterns were identified for the curriculum analyzed, distinguishing different performances in terms of GPA (Grade Point Average), pass rate, and time required to complete the major.

The CTPA method allowed to determine that, although students do not follow the course-taking plan suggested by the university, there are similarities among the different course-taking patterns identified. It was also identified that some patterns could be detrimental to students’ performance. Therefore, analyzing the course-taking patterns followed by students, and distinguishing which of them are correlated to better results, may allow universities to enhance their curriculum planning based on data.

Keywords: Curricular Analytics, Course-taking Patterns, Clustering, Higher Education, STEM Curriculum
RESUMEN

Las instituciones de educación superior sugieren una malla curricular a sus alumnos, sin embargo, no tienen la certeza de que los alumnos sigan sus recomendaciones. Se necesita una nueva metodología para identificar las verdaderas trayectorias que siguen los alumnos y su impacto en el rendimiento académico.

A partir de esta investigación se propone un nuevo método que permite descubrir las trayectorias más comunes que siguen los alumnos para completar su major. El método utiliza algoritmos de clusterización y análisis estadísticos para comparar las trayectorias encontradas. Se realizó un caso de estudio donde se analizaron las trayectorias del major de Ciencias de la Computación de una universidad chilena utilizando el método propuesto. Como resultado se obtuvieron seis trayectorias representativas, con distinto rendimiento académico considerando el promedio de notas, la tasa de aprobación, y el tiempo que les toma completar el major.

El método propuesto permite identificar que los estudiantes siguen patrones comunes en sus trayectorias, como el curso de inicio y de finalización del major, a pesar de no seguir las recomendaciones de su institución educativa. También se identificaron algunas trayectorias que pueden ser perjudiciales para el rendimiento académico de los alumnos. Por ende, analizar las trayectorias de los alumnos y distinguir cuáles están relacionadas con mejores resultados académicos, en base a métricas como el promedio de notas o tasa de aprobación, puede permitirle a las instituciones educativas mejorar su planificación curricular basándose en los datos.

Palabras Claves: Malla Curricular, Trayectorias, Clusterización, Plan de Estudios, STEM.
1. INTRODUCTION

The main purpose of higher education institutions (HEIs) is to successfully prepare their students with the knowledge and skills they need to join the workforce. To achieve this goal, they need to provide different courses that develop crucial learning outcomes in future professionals. These courses must be structured logically in a curriculum that facilitates the students learning process. Nevertheless, HEIs are not only interested in having students graduating, but they also want that students do it on time (Hickman, 2017).

Nowadays, one of the main concerns of HEIs are students attrition and timely graduation. Therefore, universities invest plenty of resources and time in seeking and implementing solutions to these problems. They offer tutoring programs, financial support, experimenting with classroom structures (online courses, flipped classrooms, etc.), among other interventions (Hickman, 2017).

Although these initiatives may facilitate student progression, there may also exist structural conditions within the curricula itself that limit progress independent of any successful initiatives (Slim, Heileman, Al-Doroubi, & Abdallah, 2016). Many institutions overlook the structure of the curriculum itself when it comes to interventions to improve student attrition and timely graduation.

This research contributes to the analysis and comparison of the course-taking patterns followed by students in STEM degrees. Using clustering models and statistical tests, it was possible to identify the course-taking patterns and their influence on the students’ academic performance. By understanding which course-taking patterns are detrimental to students’ performance, decision-makers can intervene and make adjustments to the curriculum to improve students learning retention, and timely completion, which is the main objective of this research. This analysis could also be useful for program accreditation purposes.
1.1. Context

Recently, a significant amount of work has been done on curricular analytics to show its impact on student success (Molontay, Horváth, Bergmann, Szekrényes, & Szabó, 2020; Slim et al., 2016). These researches show how the course enrollment sequences affects students’ academic performance, mainly affecting retention and timely graduation. On this subject, Klingbeil and Bourne (2015) analyzed the sequence of calculus courses, a common curricular pattern in general engineering programs that generates a bottleneck for students without a solid math background. They noticed that in semester four, most of the programs have as prerequisite Differential Equations to discipline specific courses. Although, the only topic that is required in the discipline specific courses is how to solve linear differential equations. Therefore, they made a curriculum reform to introduce a new course in the first year to teach this topic and remove the Differential Equations prerequisite for discipline specific courses. The authors indicated that with this new structure, students do not need to follow the long chain of prerequisite courses to take discipline specific courses, helping students to graduate on time and without having a negative impact on graduation GPA.

Several approaches have been used to analyze the course-taking patterns followed by the students, one of the first and most used methods is the Markov chain models. Bessent and Bessent (1980) applied for the first time the Markov approach to study the progression of doctoral students from admission to completion. Their method helped university departments to maintain control of faculty load and number of graduates at desired levels.

The Markov approach is also used to estimate the time a student spend in the higher education system (Shah & Burke, 1999) or to predict the performance of students early in their academic careers (Slim, Heileman, Kozlick, & Abdallah, 2014). The research of Shah and Burke (1999) models the movement of the students through the curriculum with Markov chains. Based in the student’s age at the beginning of a course they estimate the probability of them completing the course. Also, it estimates the mean time a student takes
to complete the course, and the mean time they spend in the higher education system. The approach of Slim et al. (2014) uses a Markov Network (MN) to represent the curriculum graphs of a specific degree program and linear regression to make predictions. This model can predict the Grade Point Average (GPA) of the student in subsequent semesters, using the previous semester GPA.

Curricular complexity is another line of research. It analyzes the structure of curriculum and its prerequisite networks with network theory tools. Slim et al. (2014) analyzed university course networks on different levels, quantifying the importance of a course based on its delay and blocking factors. Slim et al. (2016) studied how the courses enrollment sequences impact on student performance and achievement using data mining techniques and graph theory.

Simulation is another widely used approach to analyze course-taking patterns due to its flexibility. Molontay et al. (2020) introduced a framework for restrictive curricula to analyze the prerequisite network and which courses have higher impact on timely graduation. Furthermore, their approach can simulate the effects of policy changes and modifications of the prerequisites network. Saltzman & Roeder (2012) presented a model that helps to analyze changes in curriculum’s policy; the model takes into consideration the student’s flow in a college of business. Saltzman et al. (2019) includes other factors in the model, like pass rates, dropout rates and capacities. Weber (2013) developed a simulation tool that allows policy-makers to analyze the course-taking patterns of the students and analyze how it impacts graduation rates. Also, the simulation identifies bottleneck courses and determines the effects of the alteration of class capacities and offerings.

Other researchers used clustering methods to analyze course-taking patterns. Zeidenberg (2011) used the k-medoids algorithm to group the courses by type and the amount of each type that students take during their degree in Community Colleges. Wang et al. (2019) combined longitudinal multidimensional k-means and multinomial logistic regression to predict whether a student will transfer to a STEM field and when. Sailesh, Lu & Al Aali (2016) used k-means and Expectation-Maximization (EM) algorithms to profile
students whose academic performance in terms of time to degree and cumulative GPA was related to the course-taking pattern of students.

1.2. Research thesis

Due to the context presented above, the following research question arises: Do STEM students follow the course-taking plan suggested by their institutions?

From this question, the following hypothesis arises: Students do not follow the institution’s suggestions, and they follow different course-taking patterns. Some of the course-taking patterns may be detrimental to students’ academic performance, while others may be advantageous.

The main objective of this research is to propose an analytical method to discover the most frequent course-taking patterns followed by students to complete their majors using clustering algorithms and statistical analyses to make comparisons. This method, named Course-Taking Patterns Analysis (CTPA), was applied to analyze the course-taking patterns of the computer science major in a Chilean university. Six course-taking patterns were identified for the program analyzed, distinguishing different performances in terms of GPA, pass rate, and time required to complete the major.

The CTPA method allows to determine that, although students do not follow the course-taking plan suggested by the university, there are similarities among the different course-taking patterns identified. It was also identified that some patterns could be detrimental to students’ performance. Therefore, analyzing the course-taking patterns followed by students, and distinguishing which of them are correlated to better results, may allow universities to enhance their curriculum planning based on data.
1.3. Background

In order to have a better understanding of this research, the most important concepts discussed within it are briefly detailed below.

1.3.1. Curricular analytics

Curricular Analytics (CA) is an area of Learning Analytics (LA) that focuses on supporting curriculum decision-making and program improvement in the long term (Hilliger, Aguirre, Miranda, Celis, & Pérez-Sanagustín, 2020). This sub-field has formally emerged quite recently, its first workshop was held on 2016 in the sixth International Conference on Learning Analytics and Knowledge (LAK), upon the proposal of Greer, Molinaro, Ochoa and McKay (2016). Its purpose is to complement the analysis conducted at course level by LA, adding a wider vision to comprehend the learning process end-to-end. After the workshop, CA was defined as the collection and analysis of educational data to support continuous curriculum improvement and program-level decision-making (Ochoa, 2016).

In order to bring curriculum and program-level insights to stakeholders, it is necessary to have access to educational data. The sources of information that CA consumes can be divided into three groups: intrinsic, extrinsic, and interaction (Ochoa, 2016). The first is the information contained in the curriculum itself. The second is external information that influences the program content or structure, e.g., competencies requested in job advertisements (Sugar, Hoard, Brown, & Daniels, 2012). The last one is the information generated when the students interact with the curriculum, such as course enrollment, student’s grades, among others. It is commonly known as student records, and most educational institutions have it.

1.3.2. Course-taking Plan

A course-taking plan consists of a set of courses that a student must complete in order to earn the degree associated with a given program; and the rules that suggest the order in
which these courses should be taken. When a student successfully complete a course, it means that the student has attained the learning outcome associated to that course. Also, that they have dedicated to the course the requested amount of credit hours. To illustrate, most bachelor’s degree programs require a student to earn a minimum of 120 credit hours.

The set of rules that indicates the course sequence is necessary because some learning outcomes may be linked between two or more courses. In order to capture this, the educational institution can indicate that a course, namely A, is prerequisite for another, namely B. That is, students may not enroll in course B unless they have successfully completed course A. Also, they can specify that two courses are co-requisites, that means that both courses should be enrolled at the same time.

Some programs have more requisites than others. It is known that STEM degrees have relatively restrictive curricula, i.e., the institution determines the courses the student must take to graduate and suggests the order in which they should be taken (Heileman, Hickman, Slim, & Abdallah, 2017; Bordón, Canals, & Rojas, 2015). While liberal arts degrees tends to have more flexibility.

1.3.3. Course-taking Pattern

Every educational term students must enroll in the courses they will be learning during that term. In the long run, this generates a sequence that describes the course enrollment behaviour of each student; this is commonly known as course enrollment sequence. Other authors also refers to it as student flow or course trajectory.

When analyzing students behavior, common practices can be found, and course enrollment is not an exception. In this document, a course-taking pattern is defined as the pattern of courses that a group of students actually enroll to complete their degree (Zeidenberg & Scott, 2011). It is common that universities do not know the course-taking patterns followed by their students, or how these different patterns relate to student performance.
1.3.4. Clustering

Clustering is an unsupervised learning technique that aims to find a structure in a collection of unlabeled data. Therefore, a cluster is a collection of objects which are similar between them and are dissimilar to the objects belonging to other clusters (Madhulatha, 2012). This analytical technique is used in several fields, such as data mining, machine learning, image and pattern recognition, bioanalysis, among others.

Figure 1.1 represents the workflow of a classical clustering application. It has three stages: pattern representation and feature selection/extraction if needed, the definition of a proximity measure, and the clustering or grouping (Jain, Murty, & Flynn, 1999). These steps can be revisited in order to improve the results obtained.

Pattern representation refers to the structure of the features vectors that will be used, i.e., number, type and the scale of features; and the configuration of the algorithm, namely the number of groups that will be obtained (Jain et al., 1999). Feature selection is the process of selecting the subset of features that describes best the problem, while feature extraction is the process of transformation of existing features to create new ones. The proximity measure refers to the distance function used to identify the similarity between the observations. The grouping or clustering step is the final one and can be performed in different ways. For example, it can be a hard division, i.e., the algorithm delivers the class
label for each object, or fuzzy, i.e., the algorithm delivers the probability that the object belongs to each class.

Clustering algorithms can be classified depending on the model used to create the groups:

(i) **Hierarchical clustering**: It calculates the proximity measure between pair of items, and groups the ones which are most similar, then it calculates again the proximity measure between the group of items, and keeps doing it until it gets only one group. This is known as agglomerative clustering. Another option is to start with only one group and divided it until you only have separate objects. This is known as divisive clustering.

(ii) **Partitioning clustering**: It represents all of its items with a vector, that can be part of the data set or not. It iteratively relocates items by moving them from one cluster to another, starting from an initial partitioning (Rokach & Maimon, 2005). Most of these methods require that the user pre-set the number of cluster wanted.

(iii) **Density-based clustering**: It works by assuming that the items of each cluster are drawn from a specific probability distribution. While the overall distribution of the data is assumed to be a mixture of several distributions. These methods are designed for discovering clusters of arbitrary shape which are not necessarily convex (Rokach & Maimon, 2005).

In order that clustering algorithms work correctly, it is necessary to accurately define the proximity measure. In the literature, two main type of measures are used to estimate this relation: distance measures and similarity measures. Some of the most used distances and similarity measures are:

(i) **Euclidean distance**: \[ d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \]

(ii) **Manhattan distance**: \[ d(x, y) = \sum_{i=1}^{n} |x_i - y_i| \]
(iii) **Cosine similarity:** \( s(x, y) = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2 \sqrt{\sum_{i=1}^{n} y_i^2}}} \)

(iv) **Pearson similarity:** \( s(x, y) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}}} \)

Where:

\( n \): number of attributes of the vectors; \( x, y \): vectors of the data set; \( \bar{x}, \bar{y} \): mean of the attributes, i.e. \( \frac{1}{n} \sum_{1}^{n} x_i \).

The proximity measure is also used as a way to measure the effectiveness of the clustering, however, more complex indicators can be used. Nevertheless, evaluating the correctness of a clustering result is a controversial issue, because there is no universal definition for what is a good clustering (Bonner, 1964). Lui et al. (2010), describe some criteria to evaluate the quality of the clusters obtained as follow:

(i) **Root-mean-square std dev:** It is calculated as the square root of the pooled sample variance of all the attributes. It only measures the homogeneity of the formed clusters. In addition, to calculate the optimal number of clusters it is necessary to use the elbow method.

\[ \sqrt{\sum_{i} \sum_{x \in C_i} \frac{||x - c_i||^2}{P \sum_{i}(n_i - 1)}} \]

(ii) **Dunn’s index:** It uses the minimum pairwise distance between objects in different clusters as the inter-cluster separation and the maximum diameter among all clusters as the intra-cluster compactness. The optimal clusters number can be obtained by maximizing this index.

\[ \min_{1 \leq i \leq NC} \left( \min_{1 \leq j \leq NC, j \neq i} \left( \frac{\min_{x \in C_i, y \in C_j} d(x, y)}{\max_{1 \leq k \leq NC} (\max_{x, y \in C_k} d(x, y))} \right) \right) \]

(iii) **Silhouette index:** It measures how similar an object is to its own cluster, and compares it to other clusters. Silhouette values ranges between \(-1\) and \(+1\), where a high value indicates that the object is correctly grouped. To get the optimal cluster number it is necessary to maximize the value of this index. In
addition, the silhouette index can be calculated with any distance metric, and it is defined as:

$$\frac{1}{NC} \sum_{i=1}^{n} \left( \frac{1}{n_i} \sum_{x \in C_i} \frac{b(x) - a(x)}{\max(a(x), b(x))} \right)$$

where:

$$a(x) = \frac{1}{n_i - 1} \sum_{y \in C_i, y \neq x} d(x, y)$$

$$b(x) = \min_{j, j \neq y} \left( \frac{1}{n_j} \sum_{y \in C_j} d(x, y) \right)$$

Where:

- $x, y$: objects of the data set;
- $P$: number of attributes of the data set;
- $NC$: number of clusters;
- $C_i$: the $i$-th cluster,
- $n_i$: number of objects in $C_i$;
- $c_i$: center of $C_i$;
- $d(x, y)$: distance measure between $x$ and $y$.

### 1.4. Materials and methods

In this section the CTPA method is described in detail. It has three main stages that are described below: Data pre-processing, Clustering, and Analysis.

#### 1.4.1. Data pre-processing

In order to have a standard type of data to apply the method, some data pre-processing is performed. The first step is to select the curricular program that will be analyzed. Afterwards, it is necessary to only select the students that have passed all the courses of the chosen program section.

The second step is to represent each student by a vector that describes in which order they passed the courses, regardless of whether they passed the courses in consecutive semesters or not. Table 1.1 shows how two students, named 100 and 101, took four
Table 1.1. Students’ Vectors before processing.

<table>
<thead>
<tr>
<th>Student ID</th>
<th>Course A</th>
<th>Course B</th>
<th>Course C</th>
<th>Course D</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>3</td>
<td>6</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>101</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 1.2. Students’ Vectors after processing.

<table>
<thead>
<tr>
<th>Student ID</th>
<th>Course A</th>
<th>Course B</th>
<th>Course C</th>
<th>Course D</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>101</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

courses: A, B, C, and D. To exemplify the representation, Table 1.2 shows students’ vectors after the processing.

The third and last step is to filter out courses that can be taken at any time during the program. To identify these courses, it is necessary to calculate the standard deviation for the semester in which every course is passed, and if it is over a given threshold, it is filtered out.

1.4.2. Clustering

In order to group students that have similar course enrollment sequences, clustering algorithms were used. In particular, iterative partitioning clustering methods and hierarchical agglomerative clustering methods. The algorithms tested for the first kind were: k-means (MacQueen, 1967) and k-medoids (Vinod, 1969). While the only algorithm tested for the second kind was agglomerative clustering (El-Hamdouchi & Willett, 1989), with four linkage criteria (ward, complete, average, and single). These algorithms are distance-based, three different distance measures were tested: Euclidean, Manhattan (Krause, 1986), and Hamming (Bookstein, Kulyukin, & Raita, 2002).
To compare all these algorithms and select the number of clusters, when need it, the silhouette index was used as the main criteria. In addition, the number of students in each cluster was taken into consideration; no cluster can have less than 10 students.

Due to the interpretability of the results and the good outcomes obtained in various cases, the k-medoids algorithm with Manhattan distance was selected. In this case, the representative course-taking pattern of each cluster is the medoid, the data item that is closest, on average, to all of the other items in the cluster.

1.4.3. Analysis

The analysis can be divided into three steps: compliance with course-taking plan, identifying common course sub-patterns among course-taking patterns, and comparisons of academic performance between course-taking patterns.

1.4.3.1. Compliance with course-taking plan

It is necessary to know if students are following the recommendations of the university. In order to do this, the distance between the course-taking plan and each course-taking pattern is calculated. In addition, a tabular visualization with color codes (green for similarities, and red for differences) is made.

1.4.3.2. Identifying common course sub-patterns among course-taking patterns

A comparison between the course-taking patterns is done. Moreover, it is relevant to know if common course sub-patterns can be identified. This is achieved by using a visualization with color codes to identify the differences and similarities.

1.4.4. Comparing academic performance between course-taking patterns

The academic performance of the students can be measured in different ways. In this case, the following metrics are used:
• GPA before and during the program
• Pass rate before and during the program
• Program completion time

Three statistical tests are used to compare these metrics: the One-way ANOVA, the Tukey test, and the paired t-test. The first identifies if a statistically significant mean difference exists among course-taking patterns. The second identifies which of the course-taking patterns has significant differences in their mean. And, the latter compares the means before versus during the program.

1.5. Research’s paper

The results and conclusions of this research are reflected in the paper “Course-Taking Pattern Analysis (CTPA): Identifying and Comparing Course-Taking Patterns in STEM Education” that will be presented in the next chapter. The comparison between the course-taking plan and the course-taking patterns is presented in detail, as the sub-patterns found among the course-taking patterns. Also, a comparison between the academic performance, e.g. GPA, pass rate, and time to completion, of the course-taking patterns was made. This paper was submitted to the WoS “IEEE Transactions on Learning Technologies” journal.
2. ARTICLE: COURSE-TAKING PATTERN ANALYSIS (CTPA): IDENTIFYING AND COMPARING COURSE-TAKING PATTERNS IN STEM EDUCATION

2.1. Introduction

In recent years, there have been concerns regarding undergraduate education in STEM. Despite the increasing demand for workers with Science, Technology, Engineering and Mathematics (STEM) skills (Hasanah, 2020), a high proportion of students drop out, change their major, or graduate in a much longer time than expected (Sithole et al., 2017). Students are delayed in their insertion into the working world, which often affects students psychologically and financially (Brodaty, Gary-Bobo, & Prieto, 2008; Aina, Baici, & Casalone, 2011). On the other hand, universities must divide their resources among more students, and inefficiencies are generated (Agasisti & Salerno, 2007). Moreover, society also loses new human resources with up-to-date skills (Aina et al., 2011).

Research has been conducted on institutional factors as well as student-related factors that affect students’ late graduation (Sithole et al., 2017; Geisinger & Raman, 2013). Regarding the way in which the curriculum is organized, aspects such as course load (Sithole et al., 2017) and the sequential nature of curricula (Reeping et al., 2019) are relevant in STEM, due to their influence on satisfaction, dropout and long graduation time (Molontay et al., 2020). There is a growing interest in research on STEM curriculum organization, particularly on structural complexity (Heileman et al., 2017) and efficiency (Wigdahl, Heileman, Slim, & Abdallah, 2014).

Curricular Analytics (CA) is an area of Learning Analytics (LA) that focuses on analyzing and improving curriculum design, providing teaching and learning trajectories according to the needs of the students (Hilliger, Ortiz-Rojas, et al., 2020). Traditionally, curricular analysis has been a time-consuming and manual process, but with the advent of new analytical techniques, new possibilities have emerged for analyzing the course enrollment sequences of the students (Simanca, Gonzalez Crespo, Rodríguez-Baena, & Burgos, 2019). The course enrollment sequence that each student follows to complete
his/her degree, is also known as student flow (Bessent & Bessent, 1980) or course trajectory (Almatrafi et al., 2016).

There are usually differences between the course-taking plan that are proposed by academic institutions and the course enrollment sequences followed by students (Mabel & Britton, 2018; Rawatlal, 2016). In Chile, as in other countries, the STEM university programs have relatively restrictive curricula, i.e., the institution determines the courses the student must take to graduate and suggests the order in which they should be taken (Heileman et al., 2017; Bordón et al., 2015). To ensure that students follow their recommendations, universities use prerequisites networks and restrict the students’ course enrollment sequence options (Lattuca & Stark, 2011). However, the students do not always follow the suggested course-taking plan, for a variety of reasons, e.g., because they fall behind (Mabel & Britton, 2018) or because they decide to bring forward or delay a course (Almatrafi et al., 2016).

Currently, it is common for universities not to know the real course-taking patterns followed by their students, or how these different patterns relate to student performance (Reeping et al., 2019; Molontay et al., 2020; Almatrafi et al., 2016). A course-taking pattern can be defined as the pattern of courses that a group of students actually takes to complete their degree (Zeidenberg & Scott, 2011). Among various factors that have been studied to affect academic performance, graduation time, and students’ dropout, students’ course-taking patterns have been shown to have a direct influence (Heileman, Abdallah, Slim, & Hickman, 2018; Jansen, 2004). Therefore, it is in the universities best interests to analyze the course enrollment sequences that are followed by students, and execute improvements to deliver the best possible education and do not waste resources.

In this article, we propose the Course-Taking Patterns Analysis (CTPA) method to identify the most frequent course-taking patterns followed by students to complete their degree, using clustering algorithms, and statistical analyses to make comparisons. This method can be useful for the institutions to understand which course-taking patterns the students are following and how these patterns affect the students’ performance. These
insights can help stakeholders to better design and evaluate the curriculum, and it can also help students to make informed decisions.

The novelty of the method is the use of cluster analysis with a process perspective. In addition, the method works for more restrictive curricula, unlike most of the previous studies (Zeidenberg & Scott, 2011; Wang et al., 2019). Furthermore, the method is not restrained to using a single academic performance metric as in other cases (Molontay et al., 2020; Ezhilarasi, Marudhachalam, Selvanayaki, Tamilselvi, & Devaki, 2020), but rather it uses various metrics (GPA, pass rate, and completion time) to have a broader view.

This article is structured as follows: Section II presents the related work. Section III introduces the proposed method. Section IV illustrates the application of the method to a case study. Section V discusses the main findings through the case study. Finally, conclusions and future work are described in Section VI.

2.2. Related work

Researchers have used different approaches to analyze the course-taking patterns followed by the students. A widely used method is Markov chain models. Bessent & Bessent (1980) were one of the first to apply a Markov approach to study the progression of doctoral students from admission to completion. It is also used to predict if a student will graduate from the institution or not (Ezhilarasi et al., 2020). Nevertheless, Markov’s assumptions might be quite restrictive for this problem because the system under study needs to possess the Markovian property and stationary transition probabilities (Hillier & Lieberman, 1990). A lot of universities do not meet these conditions, for example, the number of vacancies for the courses varies semester to semester, as does the number of students who want to take the courses.

Another method used is graph theory, Slim et al. (2014) analyzed university course networks on different levels, quantifying the importance of a course based on its delay and
blocking factors. Slim et al. (2016) studied how courses’ enrollment sequences impact on student performance and achievement using data mining techniques and graph theory.

Other authors prefer a computer simulation approach due to its flexibility. Saltzman & Roeder (2012) presented a model that helps to analyze changes in curricular policy; the model takes into consideration the student’s flow in a business college. Saltzman et al. (2019) included other factors in the model, like pass rates, dropout rates and capacities. Additionally, Molontay et al. (2020) introduced a framework for restrictive curricula to analyze the prerequisite network and which courses have higher impact on completion time.

Clustering methods also are used to analyze course-taking patterns. Zeidenberg (2011) used k-medoids algorithm to group the type and number of courses students take during their degrees at Community Colleges. Wang et al. (2019) combined longitudinal multidimensional k-means and multinomial logistic regression to predict whether a student will transfer to a STEM field and when.

Most of the related articles working with course-taking patterns focus on flexible curriculum programs where students can take a variety of courses in different areas and adapt the order in which they do them. This article, instead, focuses on a restrictive curriculum where students have to declare a major and a minor on their first and second year respectively; and thereby have to take a list of courses in a semi-established order. Another difference is that we identify several course-taking patterns, not only one. This way, personalized recommendations for students can be made, and comparisons between the different patterns can be established.

Although the proposed method was intended for restrictive curricula, it could be applied to STEM degrees due to the sequential structure of these programs. STEM curricula require that some courses be done in a sequential order, since they become increasingly difficult and depend on the content of previous courses. (Lattuca & Stark, 2011). Specifically, engineering curricula tend to have a sequential structure in the first years, where
general math and science courses are taught, and they become more flexible towards the end (Reeping, Grote, & Knight, 2020).

It is important to analyze STEM curricula because of the longer completion time and higher dropouts rates that this area is facing (Geisinger & Raman, 2013). Furthermore, due to the increasing demand of human resources in STEM degree programs, and the under-representation of women, racial/ethnic minorities, and students from low socioeconomic backgrounds in STEM fields (Mau, Perkins, & Mau, 2016).

The outlook is even worse in Latin America; dropout rates are higher, and minority groups have lower enrollment rates in STEM degree programs (Oliveros Ruiz, 2021; Oliveira, Unbehaum, & Gava, 2019; García-Holgado et al., 2020). Moreover, Latin America countries require human resources in these fields to contribute to the development and growth of the region (Silva, Peña, & Saucedo, 2020). Mainly, because this region has an extractive economy and the trend of this century is the use, understanding, and creation of new technologies. Therefore, this region need more human resources with STEM skills to promote the use and application of technologies in society, helping the development of their communities.

2.3. Course-Taking Pattern Analysis Method (CTPA)

This section proposes the CTPA method, an approach to understanding the course-taking patterns followed by students who are taking a given program and their influence on the students’ performance. This method lets stakeholders and decision-makers discover different course-taking patterns, compare them with the course-taking plan, and make comparisons between them in order to identify similarities and differences. Furthermore, this approach enables stakeholders to analyze the influence of the course-taking patterns in the students’ performance: GPA (Grade Point Average), pass rate and completion time.
The method defines an n-dimensional space, with n equal to the number of courses considered in the analysis. Each point in the space represents the course enrollment sequence of a student. We cluster the course enrollment sequence using k-medoids as a clustering algorithm. The medoid is selected as the representative course-taking pattern of each cluster. The students in each cluster and the representative of the cluster will have a similar, but not necessarily the same, course enrollment sequence. In order to compare the course-taking patterns with the course-taking plan, we calculate the average Manhattan distance from a cluster to the course-taking plan. Also, we use visualizations to make comparisons.

To analyze the influence of the course-taking patterns in the students’ performance, we select three academic Key Performance Indicators (KPIs) to measure the students’ performance. They are GPA, pass rate, and completion time. We visualize these KPIs in box-plots, histograms, and KDE (Kernel Density Estimation) (Y.-C. Chen, 2017) visualizations to make comparisons. Additionally, we use the one-way ANOVA test (Ross & Willson, 2017), the Tukey’s test (Ramachandran, 1956), and the paired t-test (Hsu & Lachenbruch, 2014) to compare the means between clusters.

As shown in Figure 2.1, this method consists of three stages: data pre-processing, clustering, and analysis. Each stage is described in detail below.
2.3.1. Data pre-processing

First of all, we need to select the curricular program that will be analyzed. We suggest choosing only a section of the program, such as science courses, common engineering courses, major courses, minor courses, among others. Then, we select the students that passed all the courses that we previously chose.

Following that, we represent each student by a vector that describes their course enrollment sequence. We considered several alternatives. The first one was selecting the semester the student passes the courses, as shown in Table 2.1 for two students, whose IDs are 100 and 101. We discarded this alternative because not all the students start the chosen section of the program in the same semester, e.g., because they failed courses before and are behind in their curricular progress. The second alternative, shown in Table 2.2, was to normalize the starting semester and select the semester the student passes the courses. We discarded this option because in some semesters students do not take/pass any course of the program, and these semesters were considered irrelevant for the expected analysis. The third alternative was representing each student by a vector that describes in which order they passed the courses regardless of whether they passed the courses in consecutive semesters or not. Table 2.3 shows this approach for students 100 and 101. We chose this third alternative to represent the course enrollment sequence. Another consideration is that we select the semester when a student passes a course instead of the first time they take it. The main reason for this is that we are interested in when it is advisable for the student to take the course and pass it.

Finally, we filter out courses that can be taken at any time during the program. To identify these courses, we calculate the standard deviation for the semester in which every course is passed, and if it is over a given threshold, it is filtered out.
Table 2.1. Students’ Vectors Represented by the Semester in Which the Courses Are Passed.

<table>
<thead>
<tr>
<th>Student ID</th>
<th>Course A</th>
<th>Course B</th>
<th>Course C</th>
<th>Course D</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>3</td>
<td>6</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>101</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 2.2. Students’ Vectors Represented by the Normalized Semester in Which the Courses Are Passed.

<table>
<thead>
<tr>
<th>Student ID</th>
<th>Course A</th>
<th>Course B</th>
<th>Course C</th>
<th>Course D</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>101</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 2.3. Students’ Vectors Represented by the Normalized Semester First Semester and the Order in Which the Courses Are Passed.

<table>
<thead>
<tr>
<th>Student ID</th>
<th>Course A</th>
<th>Course B</th>
<th>Course C</th>
<th>Course D</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>101</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

2.3.2. Clustering

To cluster the students’ course-taking patterns, we considered iterative partitioning clustering methods and hierarchical agglomerative clustering methods. Specifically, we tested two algorithms of the first kind: k-means (MacQueen, 1967) and k-medoids (Vinod, 1969), and one algorithm of the second kind, agglomerative clustering (El-Hamdouchi & Willett, 1989), with four linkage criteria (ward, complete, average, and single). The algorithms considered are distance-based, so we tested three different distance measures: Euclidean, Manhattan (Krause, 1986), and Hamming (Bookstein et al., 2002). We selected Manhattan distance because we are working with discrete variables, and the result can be interpreted as the number of semesters of difference with which the courses were taken.
Iterative partitioning methods need the user to specify the number of clusters. To find the best number of clusters and the best algorithm we used the silhouette index (Rousseeuw, 1987) as the main selection criteria. This index measures how similar an object is to its own cluster compared to other clusters; its values go from -1 to 1. We also had to take into consideration the number of students in each cluster, no cluster can have less than 10 students.

Due to the interpretability of the results and the good outcomes obtained in various cases, we suggest the use of the k-medoids algorithm. In this case, the representative course-taking pattern of each cluster is the medoid, the data item that is closest, on average, to all of the other items in the cluster.

2.3.3. Analysis

The analysis we want to make to our data can be divided into two types: comparisons of the course-taking patterns and comparisons of the clusters’ performance.

2.3.3.1. Compliance with course-taking plan

First, we want to compare the course-taking patterns obtained with the ideal course-taking plan to analyze whether the students are following the university’s recommendations or not.

2.3.3.2. Identifying common course sub-patterns among course-taking patterns

Second, we want to compare the course-taking patterns among them, to identify similarities and differences. Furthermore, we want to know if sub-patterns can be found.

In order to do this, we use a tabular visualization with color codes to identify the differences and similarities.
2.3.3.3. Comparing academic performance between course-taking patterns

The metrics we use to analyze the students’ performance are:

- GPA before and during the program
- Pass rate before and during the program
- Program completion time

To compare the metrics named before, we use three statistical tests; the one-way ANOVA to identify if there is a statistical mean difference among clusters and the Tukey test to know which clusters have differences in their mean. Besides, we use the paired t-test to compare the means before versus during the program.

2.4. Case study

This section illustrates the method’s usefulness through a case study of the course-taking patterns of engineering students from a Latin American university. Specifically, the study was carried out on data of about 193 students who have completed the Computer Science major between the first semester of 2016 and the first semester of 2020. These students began their studies in the Pontificia Universidad Católica de Chile Engineering program between 2013 and 2016. This program allows the students to declare a major and a minor in the first and second year respectively. We selected only the major of the degree program because it has more enrolled students and has a properly number of courses to do the analysis.

The Computer Science major has ten fixed courses listed in Table 2.4 with their corresponding prerequisites. From the list of prerequisites, we can observe that four out of ten courses do not have prerequisites; the most common prerequisite is “IIC2233 - Advanced Programming”, required for four courses; and, that “IIC2154 - Capstone: Specialty Project” is the course with more prerequisites. Moreover, we can highlight that
Table 2.4. Computer Science Major Courses

<table>
<thead>
<tr>
<th>Course ID</th>
<th>Course Name</th>
<th>Prerequisites</th>
</tr>
</thead>
<tbody>
<tr>
<td>IIC1253</td>
<td>Discrete Mathematics</td>
<td>-</td>
</tr>
<tr>
<td>IIC2133</td>
<td>Data Structure and Algorithms</td>
<td>IIC1253, IIC2233</td>
</tr>
<tr>
<td>IIC2143</td>
<td>Software Engineering</td>
<td>IIC2233</td>
</tr>
<tr>
<td>IIC2154</td>
<td>Capstone: Specialty Project</td>
<td>IIC2143, IIC2333, IIC2413, IIC2613</td>
</tr>
<tr>
<td>IIC2233</td>
<td>Advanced Programming</td>
<td>-</td>
</tr>
<tr>
<td>IIC2333</td>
<td>Operating Systems and Networks</td>
<td>IIC2343</td>
</tr>
<tr>
<td>IIC2343</td>
<td>Computer Architecture</td>
<td>-</td>
</tr>
<tr>
<td>IIC2413</td>
<td>Databases</td>
<td>IIC2233</td>
</tr>
<tr>
<td>IIC2613</td>
<td>Artificial Intelligence</td>
<td>IIC2233</td>
</tr>
<tr>
<td>IIC2713</td>
<td>Information Systems</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 2.5. Computer Science Major Course-Taking Plan

<table>
<thead>
<tr>
<th>Semester 4</th>
<th>Semester 5</th>
<th>Semester 6</th>
<th>Semester 7</th>
<th>Semester 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>IIC2233</td>
<td>IIC2343</td>
<td>IIC2713</td>
<td>IIC2133</td>
<td>IIC2154</td>
</tr>
<tr>
<td>IIC1253</td>
<td>IIC2413</td>
<td>IIC2143</td>
<td>IIC2613</td>
<td>IIC2333</td>
</tr>
</tbody>
</table>

The curriculum despite being restrictive and suggesting a fixed course-taking plan to students (in Table 2.5), allows certain freedoms that gives students flexibility in their course enrollment.

2.4.1. Data pre-processing

To properly analyze the data, some pre-processing was done. First, only the major courses were considered, general and minor courses were excluded from the analysis. Second, if a student failed a course, the semester considered in the analysis was the one in which the course was passed. Third, courses that had high variability in the semester in which they are taken, i.e., courses that could be taken in any semester, were filtered out of the analysis.
The courses filtered out of the analysis were “IIC2713 – Information Systems”, “IIC2133 - Data Structure and Algorithms”, “IIC2613 - Artificial Intelligence”. All of them had a standard deviation greater than 1.1, while the average standard deviation among the other courses was 0.7. The first two courses, “IIC2713 – Information Systems”, and “IIC2133 - Data Structure and Algorithms”, were the only prerequisites of the last course of the major, “IIC2154 - Capstone: Specialty Project”. The last course, “IIC2613 - Artificial Intelligence”, was not a prerequisite of any other course of the major. The prerequisite of “IIC2133 - Data Structure and Algorithms” was “IIC1253 – Discrete Mathematics”, a first semester course. Meanwhile, the other two courses did not have prerequisites. This means that all of these courses have could be taken in at least three semesters.

2.4.2. Clustering

This section shows the results obtained from the clustering. One of the seven findings is presented below.

*F1 - Six different course-taking patterns were identified using clustering.*

Six different course-taking patterns were identified using clustering. Nevertheless, we tested different numbers of clusters, precisely, between 2 and 10. Figure 2.2 shows the silhouette index obtained for different numbers of clusters. We selected six clusters as the n of the algorithm because it has the highest silhouette index. Moreover, having a larger number of clusters helped us to have a better representation of students’ behavior in each cluster, and helped us make a more accurate analysis.

Table 2.6 shows the course-taking patterns identified. The largest course-taking pattern was nº1, which has 73 students, while the smallest one was nº5, which has 12 students. The shortest course-taking pattern was nº3, which has four semesters, while the longest ones were nº4 and nº5, which have six semesters. Clusters nº0, nº1 and nº2 had five semesters.
For each cluster, we calculated a similarity metric. This metric is calculated as the average Manhattan distance of each element of a cluster to its medoid. The cluster with students that had the most similar behavior was nº5, while the cluster with students that had the least similar behavior was nº4.

### 2.4.3. Analysis

This section shows the results obtained from the analysis. Specifically, the analysis is divided into three parts: analysis of compliance with the course-taking plan, identification of common course sub-patterns among course-taking patterns, and identification of which course-taking patterns are associated with a better/worse academic performance.

#### 2.4.3.1. Compliance with course-taking plan

A comparison was made between the identified course-taking patterns and the course-taking plan to detect whether students followed the recommendations made by the university or not.
None of the course-taking patterns comply with the course-taking plan suggested by the university.

In fact, no student followed the course-taking plan. A similarity metric was calculated as the average Manhattan distance of each point of a cluster to the course-taking plan. Using this metric the most similar cluster to the course-taking plan was cluster nº2, while the most different one was cluster nº5.

Table 2.7 shows the courses that were taken in the same semester as the course-taking plan in green, while the ones taken in another semester in red. It can be observed that all clusters had one of the starting courses that the course-taking plan suggested, but none of them had both. Conversely, the course suggested to be taken at the end was always taken at the end, although not necessarily in semester 5. In addition, the more red cells in a course-taking pattern, the larger the similarity metric. Furthermore, there were more red cells than green ones, which indicates that the course-taking patterns did not comply with the course-taking plan.

2.4.3.2. Identifying common course sub-patterns among course-taking patterns

Even though all the course-taking patterns were different, we could find some similarities among them.

Four course sub-patterns were identified among the course-taking patterns.

Table 2.8 shows the recurring sub-patterns between the course-taking patterns. First, as shown in green, the first course was almost always “IIC2233 – Advanced Programming”. Second, shown in yellow, the courses “IIC1253 - Discrete Mathematics” and “IIC2413 – Databases” were taken in block. The same happened with the courses “IIC2343 – Computer Architecture” and “IIC2143 – Software Engineering”, as shown in purple. Finally, as shown in red, there were two courses that were almost always taken sequentially:
Table 2.6. Course-Taking Patterns Identified with the Number of Students That Follows Each and Their Similarity Metric

<table>
<thead>
<tr>
<th>Cluster</th>
<th>N</th>
<th>Similarity</th>
<th>Sem 1</th>
<th>Sem 2</th>
<th>Sem 3</th>
<th>Sem 4</th>
<th>Sem 5</th>
<th>Sem 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>12</td>
<td>1.58</td>
<td>IIC1253</td>
<td>IIC2233</td>
<td></td>
<td>IIC2143</td>
<td>IIC2333</td>
<td>IIC2154</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>IIC2343</td>
<td>IIC2143</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>IIC2413</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>73</td>
<td>1.67</td>
<td>IIC2233</td>
<td>IIC1253</td>
<td>IIC1253</td>
<td>IIC2143</td>
<td>IIC2333</td>
<td>IIC2154</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>IIC2413</td>
<td></td>
<td>IIC2343</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>IIC2343</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>24</td>
<td>1.79</td>
<td>IIC2233</td>
<td>IIC1253</td>
<td>IIC1253</td>
<td>IIC2143</td>
<td>IIC2333</td>
<td>IIC2154</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>IIC2343</td>
<td></td>
<td>IIC2343</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>IIC2413</td>
<td></td>
<td>IIC2143</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>49</td>
<td>1.53</td>
<td>IIC2233</td>
<td>IIC1253</td>
<td>IIC1253</td>
<td>IIC2143</td>
<td>IIC2333</td>
<td>IIC2154</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>IIC2343</td>
<td></td>
<td>IIC2343</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>IIC2413</td>
<td></td>
<td>IIC2333</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>19</td>
<td>2.21</td>
<td>IIC2233</td>
<td>IIC2413</td>
<td>IIC1253</td>
<td>IIC2143</td>
<td>IIC2333</td>
<td>IIC2154</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>IIC2343</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>IIC2143</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>16</td>
<td>1.5</td>
<td>IIC2233</td>
<td>IIC1253</td>
<td>IIC1253</td>
<td>IIC2143</td>
<td>IIC2333</td>
<td>IIC2154</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>IIC2413</td>
<td></td>
<td>IIC2343</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>IIC2333</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

“IIC2333 – Operative Systems and Networks” followed by “IIC2154 – Capstone: Specialty Project”. It should be noted that every course-taking pattern contained at least two of these sub-patterns.

2.4.3.3. Comparing academic performance between course-taking patterns

In order to determine if some course-taking patterns were related to a particular academic performance, we characterized students in each cluster using three metrics of their academic performance: GPA, pass rate, and completion time. This way, we could analyze if there was a correlation between these metrics and the course-taking patterns. The analysis of each metric is developed below.
Table 2.7. Comparison Between the Course-Taking Plan and the Course-Taking Patterns

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Similarity</th>
<th>Sem 1</th>
<th>Sem 2</th>
<th>Sem 3</th>
<th>Sem 4</th>
<th>Sem 5</th>
<th>Sem 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plan</td>
<td></td>
<td>IIC2233</td>
<td>IIC2343</td>
<td>IIC2333</td>
<td>IIC2413</td>
<td>IIC2143</td>
<td>IIC2154</td>
</tr>
<tr>
<td>0</td>
<td>3.57</td>
<td>IIC1253</td>
<td>IIC2233</td>
<td>IIC2343</td>
<td>IIC2143</td>
<td>IIC2333</td>
<td>IIC2154</td>
</tr>
<tr>
<td>1</td>
<td>5.19</td>
<td>IIC2233</td>
<td>IIC1253</td>
<td>IIC2413</td>
<td>IIC2143</td>
<td>IIC2333</td>
<td>IIC2154</td>
</tr>
<tr>
<td>2</td>
<td>2.54</td>
<td>IIC2233</td>
<td>IIC1253</td>
<td>IIC2343</td>
<td>IIC2333</td>
<td>IIC2143</td>
<td>IIC2154</td>
</tr>
<tr>
<td>3</td>
<td>4.47</td>
<td>IIC2233</td>
<td>IIC2343</td>
<td>IIC1253</td>
<td>IIC2143</td>
<td>IIC2333</td>
<td>IIC2154</td>
</tr>
<tr>
<td>4</td>
<td>7.53</td>
<td>IIC2233</td>
<td>IIC2413</td>
<td>IIC1253</td>
<td>IIC2343</td>
<td>IIC2143</td>
<td>IIC2333</td>
</tr>
<tr>
<td>5</td>
<td>7.75</td>
<td>IIC2233</td>
<td>IIC1253</td>
<td>IIC2143</td>
<td>IIC2343</td>
<td>IIC2333</td>
<td>IIC2154</td>
</tr>
</tbody>
</table>

The row in gray shows the course-taking plan. The similarity column shows the Manhattan distance between each course-taking pattern and the course-taking plan. The courses with green background are the ones that are taken in the same semester the course-taking plan suggests. The courses with red background are the ones that are taken in other semesters.

GPA in major: Grades in the courses range between 1 and 7; the students pass a course with grades greater or equal to 4. Figure 2.3 shows the GPA distribution in courses of the major for each cluster. The clusters n°0 and n°4 had a more compact GPA distribution and their medians were lower than those of the other clusters.
Table 2.8. Sub-patterns Found in Course-Taking Patterns

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Sem 1</th>
<th>Sem 2</th>
<th>Sem 3</th>
<th>Sem 4</th>
<th>Sem 5</th>
<th>Sem 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>IIC1253</td>
<td>IIC2233</td>
<td></td>
<td>IIC2143</td>
<td>IIC2333</td>
<td>IIC2154</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>IIC2343</td>
<td>IIC2413</td>
<td>IIC2333</td>
<td>IIC2154</td>
</tr>
<tr>
<td>1</td>
<td>IIC2233</td>
<td></td>
<td>IIC1253</td>
<td>IIC2143</td>
<td>IIC2333</td>
<td>IIC2154</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IIC2413</td>
<td></td>
<td>IIC2343</td>
<td>IIC2333</td>
<td>IIC2154</td>
</tr>
<tr>
<td>2</td>
<td>IIC2233</td>
<td></td>
<td>IIC1253</td>
<td></td>
<td>IIC2143</td>
<td>IIC2154</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>IIC2343</td>
<td>IIC2413</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>IIC2233</td>
<td></td>
<td>IIC1253</td>
<td></td>
<td>IIC2143</td>
<td>IIC2154</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>IIC2343</td>
<td></td>
<td>IIC2333</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>IIC2233</td>
<td></td>
<td>IIC2413</td>
<td></td>
<td>IIC1253</td>
<td>IIC2143</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>IIC2343</td>
<td></td>
<td>IIC2333</td>
</tr>
<tr>
<td>5</td>
<td>IIC2233</td>
<td></td>
<td>IIC1253</td>
<td></td>
<td>IIC2143</td>
<td>IIC2154</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>IIC2413</td>
<td></td>
<td>IIC2343</td>
<td>IIC2333</td>
</tr>
</tbody>
</table>

There are four sub-patterns highlighted in different colors. The light blue sub-pattern shows that the first course of the major is IIC2333. The yellow sub-pattern represents that the courses IIC1253 and IIC2413 are taken in block, i.e., in the same semester or consecutively. The purple sub-pattern shows that the courses IIC2143 and IIC2343 are taken in block. The pink sub-pattern shows that the major ends with the course IIC2333 followed by IIC2154.

Figure 2.3. GPA distribution in courses of the major for each cluster.
Table 2.9. GPA Tukey Test Results

<table>
<thead>
<tr>
<th>Cluster</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-</td>
<td>0.4484</td>
<td>0.6172</td>
<td>0.7894</td>
<td>0.9</td>
<td>0.799</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
<td>0.9</td>
<td>0.9</td>
<td><strong>0.0623</strong></td>
<td>0.9</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.9</td>
<td>0.2135</td>
<td>0.9</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.3258</td>
<td>0.9</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.4556</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

**F4** - There is a statistically significant difference in the GPA in major between cluster n°1 and cluster n°4.

The one-way ANOVA test allows us to know if at least one cluster has a mean with a statistically significant difference compared to the other clusters. The one-way ANOVA results were an F-statistic equal to 2.0732 and a p-value of 0.0706, with an $\alpha$ equal to 0.1; therefore, we could establish that at least one cluster had a different mean compared to other clusters. To know which clusters have a statistically significant difference in their mean we run the Tukey test. Table 2.9 shows that the only pair that had a statistically significant difference are cluster n°1 and n°4, with cluster n°1 having a better GPA compared to cluster n°4.

**F5** - Students belonging to clusters n°0, n°3 and n°4 decreased their GPA during the major compared to their previous GPA.

Figure 2.4 shows the distribution of the difference between the students’ GPA in the courses taken before or during the major for all clusters. A dotted line is also shown at 0 to differentiate the positive differences, i.e., students who improved their GPA, from the negative ones, i.e., students who decreased their GPA. As shown, the median of the clusters n°1, n°2 and n°5 were close to 0, and the median of the clusters n°0, n°3 and n°4 were below 0. We used the paired t-test to determine whether these differences were
Figure 2.4. Distribution and box-plot of the difference between before and during the major in GPA for each cluster. The dotted line indicates where the difference is 0.

Table 2.10. GPA Paired T-Test Results

<table>
<thead>
<tr>
<th>Cluster</th>
<th>F-Statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2.1732</td>
<td>0.0525</td>
</tr>
<tr>
<td>1</td>
<td>1.238</td>
<td>0.2197</td>
</tr>
<tr>
<td>2</td>
<td>0.1812</td>
<td>0.8577</td>
</tr>
<tr>
<td>3</td>
<td>2.7673</td>
<td>0.008</td>
</tr>
<tr>
<td>4</td>
<td>2.2489</td>
<td>0.0373</td>
</tr>
<tr>
<td>5</td>
<td>0.1839</td>
<td>0.8566</td>
</tr>
</tbody>
</table>

statistically significant. As shown in Table 2.10, clusters that had statistically significant differences in their GPA were clusters n°0, n°3, and n°4. The students belonging to these clusters decreased their GPA during the major compared to their previous GPA.

While causality cannot be established, only correlation, one possible reason for the GPA decrease in cluster n°3 is that they decided to take the courses faster than suggested. Table 2.7 shows that the course-taking pattern of these students spans for 4 semesters, unlike the course-taking plan that suggests 5 semesters. So, the hypothesis is that these students, by rushing, lowered their GPA.
Table 2.11. Pass Rate Paired T-Test Results

<table>
<thead>
<tr>
<th>Cluster</th>
<th>F-Statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-2.0764</td>
<td><strong>0.0621</strong></td>
</tr>
<tr>
<td>1</td>
<td>2.9604</td>
<td><strong>0.0042</strong></td>
</tr>
<tr>
<td>2</td>
<td>0.3419</td>
<td>0.7355</td>
</tr>
<tr>
<td>3</td>
<td>0.2268</td>
<td>0.8215</td>
</tr>
<tr>
<td>4</td>
<td>0.9394</td>
<td>0.3599</td>
</tr>
<tr>
<td>5</td>
<td>1.0729</td>
<td>0.3003</td>
</tr>
</tbody>
</table>

**Pass rate:** The pass rate is calculated as the number of courses passed divided by the number of courses taken; this metric can take values between 0 and 1.

**F6 - There is a statistically significant difference in the pass rate before and during the major for cluster n°0 and for cluster n°1.**

Table 2.11 shows the results of the paired t-test of this ratio before and during the major. Clusters n°0 and n°1 had statistically significant differences. In the case of cluster n°0, the pass rate before the major was closer to 1 than the pass rate during the major. Meanwhile, in the case of cluster n°1, students had a pass rate during the major closer to 1 than before the major. These differences are shown in Figures 2.5 and 2.6, respectively.

**Completion time:** Students should complete their major in 5 semesters. Figure 2.7 shows that the mode of clusters n°1, n°2 and n°3 was to complete the major on time. The mean of these clusters was also to complete the major in almost 5 semesters. Clusters n°0 and n°5 had a mean close to 6 semesters, and cluster n°4 was the one with the largest mean, close to 7 semesters.

**F7 - There is a statistically significant difference in the completion time between different clusters.**
Since the one-way ANOVA test had a result of F-statistic of 12.1658 and a p-value of 3.2443e-10, with an $\alpha$ equal to 0.1, we can say that at least one of the means of these clusters was different. The results of the Tukey test are shown in Table 2.12. They show that the differences in means were statistically significant between cluster n°0 and clusters n°1 and n°3. This is because cluster n°0 was one of the clusters that groups students who take more semesters to complete the major, while clusters n°1 and n°3 group students who take fewer semesters to complete the major. On the other hand, cluster n°4 had a statistically significant difference with all clusters, except cluster n°0. This is because cluster n°4 was the cluster that groups students who take longer to complete the major.
Figure 2.7. Relative frequency histogram of the completion time in semesters for each cluster. The dotted line represents the mean for each cluster.

Table 2.12. Completion Time Tukey Test Results

<table>
<thead>
<tr>
<th>Cluster</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-</td>
<td>0.0854</td>
<td>0.3032</td>
<td>0.0015</td>
<td>0.5204</td>
<td>0.9</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
<td>0.9</td>
<td>0.169</td>
<td>0.001</td>
<td>0.2133</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.2622</td>
<td>0.001</td>
<td>0.567</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.001</td>
<td>0.0038</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.1366</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

2.5. Discussion

In this paper, we presented the CTPA method. This method allows Higher Education Institutions (HEI) to identify and analyze the course-taking patterns that students follow, and how they impact on the academic performance of the students, e.g., their GPA, pass rate, and completion time. The method has several novel features: it uses cluster analysis of the curriculum with a process perspective; it works for more restrictive curricula; and, it is not restrained to using a single academic performance metric as in other cases, but rather it uses various metrics to have a broader view of the students. We believe that this
approach is relevant for an international audience, because HEIs are receiving more heterogeneous students in terms of prior preparation, socioeconomic background and beliefs about learning (Salazar-Fernandez, Munoz-Gama, Maldonado-Mahauad, Bustamante, & Sepúlveda, 2021). They usually show more variability in their academic results (Mabel & Britton, 2018) and, therefore, it is expected a high variability in the course enrollment sequences followed by the students.

The main findings presented in the Case Study section should not be taken as general conclusions, but as examples of the expressive power of the CTPA method. Although it is possible to establish hypotheses about why different clusters form and why differences occur, further research is needed to delve into the causes of these differences. These main findings are discussed below.

First, it was possible to identify different course-taking patterns through clustering (finding F1). This finding is the basis for analyzing student behavior concerning the curriculum. The CTPA method approaches the problem from a process perspective, unlike other methods that also use clustering, which allows a more comprehensive understanding of how students progress through their academic degrees and the impact on their performance (Bahr, 2013). This method is especially useful for analyzing curricula of a highly sequential nature, which are common in STEM programs (Reeping et al., 2019).

Second, when comparing the course-taking patterns with the course-taking plan, we can identify that students are not following the university’s suggestions (finding F2). Therefore, we can infer that the course-taking plan is not sufficient as a guide for students. Furthermore, curriculum suggestions made by HEIs should consider students’ diversity and their contexts, such as cultural background, motivation, previous education, personal and professional experiences, study approach, among others (Jonker, März, & Voogt, 2020). HEIs should give flexible learning options to students, taking into account their personal characteristics and learning styles. This way, students could regulate their learning process (Hill, 2006).
Third, it was possible to identify four course sub-patterns (finding F3) that are not directly related to prerequisites or co-requisites. These behaviors are observed for most students, regardless of their course-taking pattern. Apparently, students are using in their decisions, prerequisites, and co-requisites that are not defined in the suggested curriculum. According to Biggs (Biggs & Tang, 2011), requisites should be defined not only in terms of content but also in terms of learning outcomes, and students could be defining their course-taking patterns based on the experiences of peers about which are the best paths (Demetriou & Schmitz-Sciborski, 2011). Managers should analyze these sub-patterns to improve the suggested curriculum. Additionally, social influence can affect student enrollment decisions in courses. According to Tinto (Tinto, 2017), the sense of belonging is important in students’ decisions, and the decisions to take courses are not only individual decisions, but also peer decisions.

Fourth, when comparing the performance of the different clusters (findings F4, F5, F6 and F7), we could identify that some course-taking patterns are detrimental to students’ performance. For instance, completing the major in less time (cluster n°3) or more time than suggested (cluster n°4) were associated with a lower GPA. These results could be useful information for policymakers to design strategies to prevent students from following detrimental course-taking patterns. There is evidence that assesses the impact of financial restrictions on student decisions, specifically those related to financial aid (R. Chen, 2008) and part-time work (Almenberg, Lusardi, Säve-Söderbergh, & Vestman, 2016). Students may be choosing detrimental course-taking patterns to manage financial constraints, slowing down their curricular progress, or trying to finish the major faster to meet some scholarship requirements, affecting their academic results.

The CTPA method has four main limitations that should be taken into account when using it. First, the curriculum has to be regulated, or at least suggested, i.e., there needs to be a defined course-taking plan and a list of possible courses. This makes it possible to compare the patterns obtained with a base plan. Second, there is a trade-off between the number of courses and the prerequisites between them. The more courses we include
in the analysis, the more prerequisites are needed. To ensure better interpretability of the method’s results, we suggest not analyzing the complete curriculum of a degree, rather a section of it, such as the basic science courses, or courses in a particular major or minor. Third, the curriculum to be analyzed must be stable throughout the selected period, i.e., there must be no changes in the course-taking plan nor a context that generates noise for the analysis, such as the current pandemic, because many HEIs have implemented temporary flexibility measures regarding COVID-19 (Daniel, 2020). Finally, the method only allows the analysis of the course-taking pattern of students that have completed all the courses in the program, therefore it is not useful to understand the dropout rate, but mainly the timely or late graduation, and the academic performance of those who graduate.

2.6. Conclusions

In this article, we presented the CTPA method that allows HEIs institutions and policymakers to analyze the course-taking patterns followed by students and compare their academic performance. The novelty of the method is that it works for more restrictive curricula, and is not restrained to using a single academic performance metric as in other cases. Besides, we presented a case study to illustrate the use of the method in a Computer Science major of a Latin American university. From the case study, we extracted seven findings that exemplify the value of the method.

The CTPA method can identify course-taking patterns and sub-patterns of a curriculum, allowing policymakers to make data-driven decisions about which course-taking plan to suggest to students in the future. Moreover, as the method allows the identification of various patterns, it gives the possibility of delivering more personalized suggestions to the students, taking into account their personal characteristics and their learning process.

A future line of research is to analyze the patterns in a multidimensional way, including qualitative information from students to help us understand why students choose to follow a given pattern. Another line for further research is to analyze how each sub-pattern affects
students’ academic performance, i.e., how the sequence of a specific group of courses impacts the learning outcomes of students. This analysis would provide valuable information to policymakers that will help modify the curriculum prerequisites.

Acknowledgment

This study was funded by National Agency for Research and Development (ANID) – Scholarship Program / Doctorado Nacional 2015 - 21150985, and supported by National Agency for Research and Development (ANID) / FONDECYT - Chile Regular Project – 1200206. The application case was conducted using anonymized data, provided by Pontificia Universidad Católica de Chile.
3. CONCLUSIONS AND FUTURE WORK

In this last chapter, the main conclusions derived from the research are detailed. In addition, the limitations of the proposed method are described. Finally, some possible future lines of work are discussed.

3.1. Conclusions

This research was guided by the following question: "Do STEM students follow the course-taking plan suggested by their institutions?". To answer this question, the CTPA method was introduced; a method that allows HEIs institutions to analyse students’ compliance with the course-taking plan. Furthermore, it allows policymakers to analyze the course-taking patterns followed by students and compare their academic performance.

In detail, this method can identify course-taking patterns and sub-patterns of a curriculum, allowing policymakers to make data-driven decisions about which course-taking plan suggest to students in the future. Moreover, as the method allows the identification of various patterns, it gives the possibility of delivering more personalized suggestions to the students, taking into account their personal characteristics and their learning process.

The main hypothesis of this research was that students do not always follow the suggestions made by their institutions. From this hypothesis, two objectives arise: to prove that students do have defined course-taking patterns, and that some patterns might be detrimental to their academic performance. This was proved in the case study presented in the article included in Chapter 2, and from the case study, seven findings arise. These findings are not general conclusions, but an example of the expressive power of the CTPA method.

First, it was possible to identify different course-taking patterns through clustering (finding F1), and from these it was possible to analyze student behavior concerning the curriculum. Second, when analyzing the compliance with the course-taking plan, it was identified that students are not following university’s suggestions (finding F2). Therefore,
it can be inferred that the course-taking plan is not sufficient as a guide for students. Students need to have more information in case they decide to enroll their courses in a different way. Third, it was possible to detect four course sub-patterns (finding F3) that are not directly related to prerequisites or co-requisites. F3 indicates that students could be defining their course enrollment sequence based on the experiences of peers rather than on the course-taking plan. When comparing the performance of the different course-taking patterns (findings F4, F5, F6 and F7), some of them were recognize as detrimental to students’ performance. These results could be useful information for policymakers to design strategies to prevent students from following detrimental course-taking patterns.

In light of the conclusions described above, and the results presented in the article included in Chapter 2, it is possible to confirm the research hypothesis that “Students do not follow the institution’s suggestions, and they follow different course-taking patterns. Some of the course-taking patterns may be detrimental to students’ academic performance, while others may be advantageous”.

3.2. Limitations

The method proposed has four main limitations. The first limitation is that the program degree to analyze must have a suggested course-taking plan, in order to compare the course-taking patterns obtained with a base course-taking plan. For this reason the CTPA method is suggested to STEM degree programs.

Secondly, it is not possible to analyze a full degree program because of the number of courses. There is a trade-off between the number of courses and the prerequisites between them. If more courses are included into the analysis, then more prerequisites are needed in order to have accurate results. To ensure better interpretability of the method’s results, it is suggested to analyze only a section of the curriculum, such as basic science courses, or courses in a particular major or minor.
Thirdly, the curriculum to be analyzed can not have changes during the period of analysis, i.e. the course-taking plan must be the same during the analysis period. In addition, the context needs to be stable, i.e., it can not generate noise for the analysis. An example of this is the current pandemic, where many HEIs have implemented temporary flexibility measures due to COVID-19 (Daniel, 2020).

Finally, the method only allows the analysis of students that have passed all the courses of the selected portion of the program. Therefore, this method is not useful to understand the dropout rate, but mainly the timely or late graduation, and the academic performance of those who finish the selected part of the program.

3.3. Future work

From this research, different future lines of work can be followed. First, it is possible to analyze the patterns in a multidimensional way, including qualitative information from students to help us understand why students choose to follow a given pattern. In the same line, it is necessary to conduct a student survey in order to know which factors mainly influence student course enrollment and analyze the course-taking patterns obtained in the lights of these factors.

Secondly, it could be useful to conduct a focus group with students to discuss the current course-taking plan and the reasons why they do not follow it. Also, some findings obtained from this research could be discussed to verify some conjectures.

A third line for further research is to analyze how each sub-pattern affects students’ academic performance, i.e., how the sequence of a specific group of courses impacts the learning outcomes of students. This analysis would provide valuable information to policymakers that will help modify the curriculum prerequisites.
Fourthly, the method can be applied when the students take a course for the first time, rather than when they pass it. In this case, it could be used to understand the course-taking patterns that ends in dropout.

### 3.4. Acknowledgement

This thesis was supported by the National Agency for Research and Development (ANID) / FONDECYT - Chile Regular Project – 1200206.
REFERENCES


Reeping, D., Grote, D. M., & Knight, D. B. (2020). Effects of large-scale programmatic
change on electrical and computer engineering transfer student pathways. *IEEE Transactions on Education*.


