


Article

Backpack Process Model (BPPM): A Process Mining Approach for Curricular Analytics

Juan Pablo Salazar-Fernandez ^{1,2,*} , Jorge Munoz-Gama ¹ , Jorge Maldonado-Mahauad ³, Diego Bustamante ¹ and Marcos Sepúlveda ¹ 

¹ Department of Computer Science, School of Engineering, Pontificia Universidad Católica de Chile, Santiago 7820436, Chile; jmun@uc.cl (J.M.-G.); diegobustama@uc.cl (D.B.); marcos@ing.puc.cl (M.S.)

² Institute of Informatics, Universidad Austral de Chile, Valdivia 5110701, Chile

³ Computer Science Department, Universidad de Cuenca, Cuenca 010107, Ecuador; jorge.maldonado@ucuenca.edu.ec

* Correspondence: jpsalaza@uc.cl; Tel.: +56-63-2221427

Featured Application: In this work, Process Mining techniques are used with a curricular analytics approach, to model the educational trajectories of engineering students during their first courses.

Abstract: Curricular analytics is the area of learning analytics that looks for insights and evidence on the relationship between curricular elements and the degree of achievement of curricular outcomes. For higher education institutions, curricular analytics can be useful for identifying the strengths and weaknesses of the curricula and for justifying changes in learning pathways for students. This work presents the study of curricular trajectories as processes (i.e., sequence of events) using process mining techniques. Specifically, the Backpack Process Model (BPPM) is defined as a novel model to unveil student trajectories, not by the courses that they take, but according to the courses that they have failed and have yet to pass. The usefulness of the proposed model is validated through the analysis of the curricular trajectories of $N = 4466$ engineering students considering the first courses in their program. We found differences between backpack trajectories that resulted in retention or in dropout; specific courses in the backpack and a larger initial backpack sizes were associated with a higher proportion of dropout. BPPM can contribute to understanding how students handle failed courses they must retake, providing information that could contribute to designing and implementing timely interventions in higher education institutions.

Keywords: learning analytics; curricular analytics; process mining; curricular trajectories; higher education



Citation: Salazar-Fernandez, J.P.; Munoz-Gama, J.; Maldonado-Mahauad, J.; Bustamante, D.; Sepúlveda, M. Backpack Process Model (BPPM): A Process Mining Approach for Curricular Analytics. *Appl. Sci.* **2021**, *11*, 4265. <https://doi.org/10.3390/app11094265>

Academic Editor: Pentti Nieminen

Received: 31 March 2021

Accepted: 29 April 2021

Published: 8 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the last decade, different techniques have progressively emerged for the analysis of data recorded by information systems, with the purpose of supporting informed decision-making in Higher Education Institutions (HEIs) [1]. In this context, Learning Analytics (LA) is the “measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs” [1]. Many HEIs have high hopes that LA can play an important role in supporting institutional processes, where data analysis can improve teaching [2], curricula [3] and learning outcomes [2] and reduce dropouts [2].

Curricular Analytics (CA) emerged as an area of LA that focuses on analyzing curricula and improving them through continuous improvement processes [4]. For this purpose, analytical tools are used to collect different types of educational data, e.g., data on curricular structure and/or data on course grading [5]. This allows the HEIs to analyze the strengths

and weaknesses of their curricula and to justify the curricular decisions and changes made when designing learning pathways for students [2].

The main objective of CA is to improve the curricular design to provide teaching and learning trajectories according to the needs of the students. For this purpose, it seeks to assess the consistency between course-level outcomes and program-level outcomes, determining whether learning outcomes have been achieved and eventually detecting gaps with respect to the assessments [5]. In addition, it is possible to identify blind spots and bottlenecks of students' trajectories, thus demonstrating the value of the curriculum for different stakeholders [6]. Traditionally, curricular analysis has been a painstaking, laborious, time-consuming, manual process. However, with the emergence of new analytical techniques that take advantage of different data sources (both structured and unstructured), new possibilities have emerged for analyzing the curriculum and students' curricular trajectories [7].

There are often significant differences between the curricular trajectories proposed (ideal trajectories) by academic institutions and the curricular trajectories carried out (actual trajectories) by students [8,9]. Not everyone is able to follow these curricular trajectories in the same way, so it is important to detect when these deviations happen and to understand why they happen. In this way, it is possible to support better decision-making at the curricular level (e.g., what courses students are expected to take in the same semester, prerequisite courses and number of available vacancies to be offered in a course, among other things).

So far, some tools have been developed that integrate different techniques for the analysis of curricular trajectories (e.g., data mining techniques such as clustering and classification) [10,11], with the purpose of obtaining a "snapshot of the student" at a specific moment in time with respect to their progress. Some of these tools are, e.g., academic advising systems, which allow to see the progress status of a student according to their curriculum [4,7,12] or early dropout prediction systems in MOOC courses [13]. However, while these tools are promising, the models implemented to define curricular trajectories as processes (i.e., an ordered sequence of events that describe a student's curricular history) and not just as a snapshot of curricular progress, are scarce. Moreover, several authors consider that the development of LA requires that the selection of data and analysis models be more grounded in learning sciences [1], in order for tool design decisions to be based on them [14]. Therefore, it is believed necessary to develop analysis models that see curricular trajectories as processes and are based on concepts that come from learning sciences.

To broaden the current understanding of how curricula can be improved based on evidence from data, this paper presents a model for systematizing the analysis of curricular trajectories as processes, based on the backpack concept and making use of formal Process Mining (PM) techniques. This broadens previous studies where these techniques have been used directly on curricular records [15–17], without conceptual abstractions. The proposed model is based on preliminary work, presented in [18].

This paper is structured as follows: Section 2 presents the backpack metaphor behind the approach. Section 3 introduces the related work in PM. Section 4 describes the Backpack Process Model (BPPM). Section 5 illustrates the BPPM proposal with an application case. Section 6 discusses the main findings through the application case. Finally, conclusions for BPPM are presented in Section 7.

2. The Backpack Metaphor

Recent research has highlighted the important role that student effort plays in academic success [19]. High school readiness [20], economic disadvantage [20], classroom climate [21] and curriculum design [22] had previously been identified as factors that explain dropout. However, self-efficacy also plays a key role in academic success.

Self-efficacy is defined as a person's belief in their ability to succeed in a specific situation [23] and has recently been linked to the effort that students are willing to make [19]. In a qualitative case study, Meyer and Marx [24] found that loss of confidence due to

poor performance contributed to engineering attrition. Those students who believe that intelligence is fixed and cannot be developed tend to be less tolerant of failure [19]. On the contrary, those who believe that it can be developed, when faced with an adverse situation, try harder, develop active learning strategies and, even when their curricular progress is not optimal, persist [19].

Based on those previous works about the relevance of self-efficacy, this paper defines the backpack metaphor. A metaphor is “a mechanism of analogy in which we conceive a concept that belongs to a certain conceptual domain in terms of another conceptual domain, and in which correspondences between the attributes of both domains are established” [25]. Metaphors are commonly used to communicate ideas in technical disciplines, facilitating interdisciplinary work. The backpack metaphor is defined as follows: The list of failed courses that a student must retake can be represented as stones that the student puts in a backpack. Each time a student fails a course, a new stone is placed in their backpack, which remains there until the course is passed. Carrying many stones in the backpack could awaken in the student the need to empty it as soon as possible, making risky decisions from the curricular point of view [26]. On the other hand, never being able to empty the backpack (even if the failed courses changed), could affect self-efficacy, have serious consequences in the medium term, affect the students’ goals or even result in program dropout.

Traditionally, the analysis of curricular trajectories has been based on the progress of each student in the curriculum [8,27]. Even though this strategy has the advantage that the model is obtained directly from the data and is easy to understand, it does not represent the psychological burden students perceive when they have failed courses that later on they must retake. Understanding how students manage their course backpack could help to better understand their decisions regarding course enrollment, persistence and dropout.

3. Related Work in Process Mining

Process Mining (PM) is a relatively new research discipline that acts as a bridge between data science and process science [28]. It aims to extract knowledge from the event logs obtained from information systems, in order to discover process models, verify conformance, analyze bottlenecks, compare variants of the same process and suggest improvements [28]. This discipline has been applied in multiple domains, with particular success achieved in fields in which processes are insufficiently structured, such as healthcare [29] and education [16].

Different techniques have been used to analyze the diverse perspectives of the processes, such as control flow, performance and organizational [30]. For the control flow, which is the focus of this work, petri nets [30,31], causal nets [31] and process trees [32], among other methods, have been used to represent process models and algorithms have been proposed that can generate them (e.g., alpha for petri nets [28] and Inductive Miner for process trees [28]). Directly Follows Graphs (DFGs) that can be obtained by DFG-based algorithms such as heuristic miner [28] are also widely used because they are one of the easiest notations to interpret by non-expert users of process mining [31]. While DFGs show disadvantages over the other mentioned formalisms because they do not manage concurrency properly (i.e., when several events occur simultaneously), they are especially recommended when concurrency representation is not necessary.

To guide the application of PM and analyze the results [33], different methodologies have been developed. Well-known generic methodologies are the L* life cycle [28] and PM² [33]. Both methodologies have a broad scope, covering the entire process management cycle [34]. Different authors have proposed domain-specific methodologies that take into account the particularities of each domain. For education, Maldonado-Mahauad et al. [35] adapted PM², narrowing its scope from data extraction to model analysis. Johnson et al. [36] extended PM² to include domain-specific requirements in healthcare, in terms of ethics and participation of domain experts, among other things. Martin et al. [29] established that a process mining methodology in healthcare should highlight usability, in the building of domain-specific event logs and in the management of unstructured data. In manufacturing,

Lorenz et al. [37] proposed a methodology with a scope that included the improvement of the processes.

In recent years, research has been conducted to visualize event logs through domain models or even theoretical frameworks that can be used to represent and analyze the data [38]. The inclusion of shift work operation to model the organizational perspective of processes [30], the construction of a model representing the added value in service processes [39], the modeling of user behavior in MOOCs to identify self-regulated learning strategies [35] and the analysis of dropout behavior through the investment model [40], are examples of that. In this work, we propose to use the backpack metaphor to conceptualize students' curricular trajectories as a novel approach to understand how they manage their failed courses.

While any general-domain process mining methodology could have been used for this work, the one proposed by Maldonado-Mahauad [35] was chosen because it is based in PM², which is widely used, and its scope goes only from data extraction to model analysis. The main contribution of this work is not the sequence of stages in the methodology, but the approach used to understand the curricular trajectories through the backpack metaphor.

4. The Backpack Process Model (BPPM) Approach

This section proposes the Backpack Process Model (BPPM) approach, in order to systematize the analysis of curricular trajectories using PM techniques, based on the backpack metaphor. This model represents the curricular trajectories of the students as a sequence of failed courses that they must retake; that is to say, as a sequence of backpacks. This sequence of backpacks is represented as a Directly Follows Graph (DFG), one of the most popular and widespread process modeling notations [31]. A DFG is a graph with nodes and transitions (directed edges) that corresponds to directly follows relationships (see Figure 1) [31]. In the BPPM, each node represents the group of failed courses that the student must retake and each edge represents the transition between a given backpack and the next one. Table 1 shows the backpack trajectories for two students, namely 23 and 24. In this example, student 23 failed algebra (A) and chemistry (Q) in the first semester, beginning the following semester with both courses in his/her backpack. This situation is labeled "AQ" in Table 1. The second semester, this student passed chemistry (Q) but failed algebra (A) again, keeping it in his/her backpack. Finally, this student passed algebra (A) and continued studying with an empty backpack. This situation is tagged with "RETENTION" in the event log. On the contrary, student 24 failed chemistry (Q), maintained it in the backpack for one semester and the next semester dropped out. This situation is tagged with "DROPOUT" in the event log. Figure 1a represents the DFG for the backpack trajectories illustrated in Table 1.

Table 1. Example of event log for BPPM.

Student ID	Backpack	Starting Date	Ending Date
23	AQ	1 July 2013	1 December 2013
23	A	1 December 2013	1 February 2014
23	RETENTION	1 February 2014	1 February 2014
24	Q	1 July 2013	1 December 2013
24	DROPOUT	1 December 2013	1 December 2013

Additionally, our analysis considers a derivation of this model, called BPPM-S (BPPM, grouped by size), where curricular trajectories are represented as a sequence of backpack sizes for each academic period. Figure 1b represents the DFG for the backpack trajectories included in Table 1, according to the BPPM-S. In Figure 1b, BP-1 and BP-2 represent the backpack size at the end of a semester; that is, the number of courses students failed and have not yet passed. BP-1 indicates a backpack size of 1, while BP-2 indicates a backpack size of 2. More details on the meaning of nodes and transitions shown in Figure 1 are explained later, in the event log generation subsection.

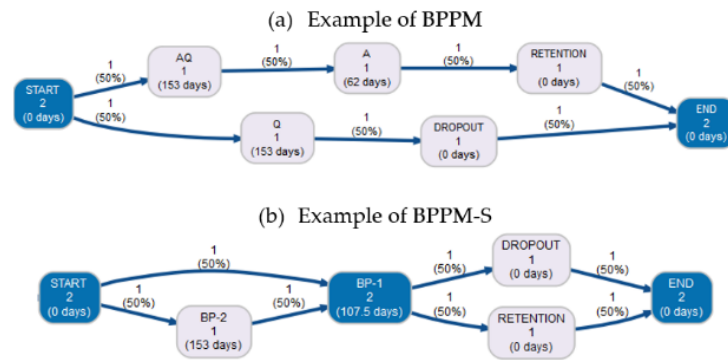


Figure 1. Example of educational trajectories, according to the BPPM and BPPM-S models. (a) Shows an example of BPPM. (b) Shows an example of BPPM-S. The darker color of the nodes represents a higher percentage of students who went through each state. The thickness of the arrows represents the percentage of students who had transitions between both states. All values are percentages in relation to the total number of students included in each model.

In order to develop the previously proposed analysis models from the curricular records of a HEI, it is necessary to apply a PM methodology. An appropriate methodology, such as the one presented below, allows to apply PM techniques and to understand the results [33]. This methodology, as can be seen in Figure 2, defines the following four stages: Data extraction, event log generation, discovery and analysis.

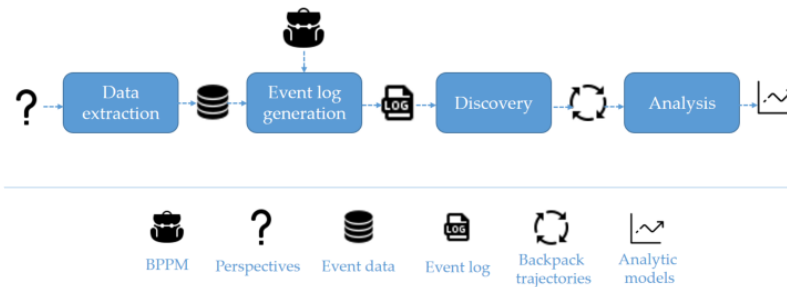


Figure 2. Stages for the generation of the process model, based on an adapted version of PM² methodology [35].

4.1. Data Extraction

In the first stage of the methodology, the minimum necessary data are extracted and used for an exploratory analysis. Since the final objective is the curricular analysis, the BPPM and BPPM-S models define a table, called BPPMt, that contains a record of each course taken by each student, as well as their final grade (see Table 2).

Table 2. Example of table BPPMt.

Student ID (s)	Period (p)	Course (c)	Grade (g)	Ending Date (d)
23	2013-1	Algebra (A)	2.0	1 July 2013
23	2013-1	Chemistry (Q)	3.5	1 July 2013
23	2013-1	Calculus (C)	4.5	1 July 2013
23	2013-1	Innovation (D)	5.5	1 July 2013
23	2013-2	Algebra (A)	3.4	1 December 2013
23	2013-2	Chemistry (Q)	5.0	1 December 2013
23	2013-3	Algebra (A)	6.5	1 February 2014
24	2013-1	Algebra (A)	5.5	1 July 2013
24	2013-1	Chemistry (Q)	3.5	1 July 2013
24	2013-1	Calculus (C)	4.5	1 July 2013
24	2013-1	Innovation (D)	6.0	1 July 2013
24	2013-2	Chemistry (Q)	3.8	1 December 2013

In Table 2, each row of the BPPMt table is represented by the tuple (s, p, c, g, d), where:

- s indicates the ID of the student who took the course
- p the academic period when the course was taken
- c the identifier of the course taken
- g the final grade obtained
- d the end date of the academic period.

For example, (23, 2013-3, Algebra, 6.5, 1 February 2014) indicates that the student with ID 23 took the algebra course in the period 2013-3, which concluded on 1 February 2014, obtaining a grade of 6.5 out of 7. The data necessary for the creation of the BPPMt table are available in the ERP systems commonly used by HEIs. It must be noted that, although the BPPMt table specifies the minimum fields as a common standardization for the next stage of the methodology, the table can be extended with additional data, such as the student's weighted GPA or current academic status.

4.2. Event Log Generation

In this stage, it is required to model the data in the form of an event log [28] (i.e., a record of the events that have happened in a process). Formally, an event log is defined as a set of cases (executions of the process), where each case is an ordered sequence of events (the actions occurred in that execution) [28]. Therefore, in order to define an event log, it is necessary to define (1) how to identify a case and (2) how to specify a sequence of events.

A classic first event log option for curricular data consists of defining each case as a student and each event as a course taken by that student, in the order in which he/she has taken it. For example, $\langle A, Q, C, D, A, Q, A \rangle$ would define the trajectory of the student with $s = 23$, according to Table 2.

However, as mentioned above, our BPPM model proposes a different perspective on curricular data, where each event in a student's trajectory is the backpack he or she has at the end of each academic term. Formally, (1) each student identifier in the BPPMt table identifies a different case $\langle b_1, b_2, \dots, b_n \rangle$; and (2) a b_i event is defined as the record of the group of failed courses that the student must retake at the end of academic period i . For example, for the student identified with $s = 23$ in Table 2, $\langle AQ, Q, - \rangle$ represents their backpack trajectory and graphically it can be seen in Figure 3. In his/her first academic period (for example, semester), this student had failed algebra (A) and chemistry (Q) and therefore must retake them. In their second period the student must still retake chemistry (either because the student took it and failed it again or because the student decided not to take the course in the second period) and finally, after the third academic period, the student does not have courses that should be retaken.

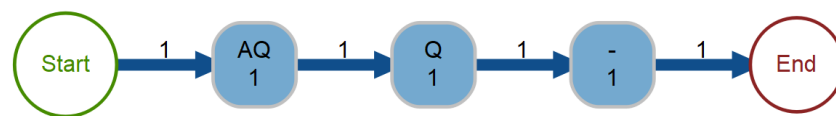


Figure 3. Graphical representation of the sequence of backpacks for the student with $s = 23$, shown as an example in Table 2, according to the BPPM model.

For simplicity, the above makes an abuse of notation, where the set $\{A, Q\}$ is represented by the label AQ and the empty set is presented by the label -. Similarly, an ordering between course identifiers is assumed, such that two equal sets are represented by the same label $\{A, Q\} = \{Q, A\} = AQ$.

This definition of event log allows us to analyze the trajectories in the backpack. However, a student who stays two academic periods with the same backpack will be represented with two consecutive equal events, where the first event ends in the first period and the second in the second period. To analyze how much time the student maintains the same backpack, it needs to be represented as a single event, which begins at the end of the academic period in which this backpack appears and ends in the period in

which the backpack changes. The BPPM model proposes the post-processing of the event log defined above, to merge consecutive events that represent the same backpack. That is, given a case $\langle \dots ; b_i; b_{i+1}; \dots ; b_{i+n}; \dots \rangle$ where $b_i = b_{i+1} = \dots = b_{i+n}$, will result in a case $\langle \dots b_{i:i+n} \dots \rangle$, where backpack $b_{i:i+n}$ ranges from period i to period $i + n$.

To easily distinguish between cases that ended in dropout or retention, for the backpack trajectories that ended with the empty backpack, in the BPPM model, the “-” label was replaced by the label “RETENTION” and in the other cases, a “DROPOUT” event was added at the end of the case. In the example used in the previous paragraph, $\langle AQ, A, - \rangle$, was replaced by $\langle AQ, A, RETENTION \rangle$ in the BPPM event log.

Finally, to obtain the event log for the BPPM-S model, each $b_{i:i+n}$ event was labeled with $BP-j$, where j represents the size of the backpack. For example, the case represented by $\langle AQ, Q, RETENTION \rangle$, was replaced by $\langle BP-2, BP-1, RETENTION \rangle$ in the BPPM-S event log.

4.3. Discovery

In this stage, the event log generated is processed using process mining algorithms. Specifically, process discovery algorithms are used to automatically generate a model of the curricular trajectories.

In this work, the BPPM and BPPM-S models propose the use of the Directly Follows Graphs (DFG), that are built using the DFG algorithm [31]. There is a wide range of technologies that implement the DFG algorithm. Both academic (e.g., PM4Py [41], bupaR [42], ProM [28]) and commercial (e.g., Disco [43], Celonis [44]) alternatives are available. In this case, the model discovery stage was performed using bupaR, an integrated collection of R packages that creates a framework for the reproducible analysis of processes in R [42].

The DFG notation was chosen because it is one of the easiest to interpret by non-expert users in process mining [31]. Moreover, the DFG notation is especially recommended when the concurrency representation is not necessary (i.e., several events occurring simultaneously), as in our case (i.e., each event corresponds to a different, disjointed period).

4.4. Analysis

In this final stage, the model analysis was also performed with bupaR [42], considering different perspectives, which are described in more detail in Table 3, including the selection of node types, the selection of transition types and applied filters.

Table 3. Filters and properties applied to the event logs to perform each analysis.

Model	Perspective	Node Type	Transition Type	Filters	Figure
BPPM	(P1) Final event (DROPOUT or RETENTION)	Number of students	Number of students	Final state: RETENTION; DROPOUT Does not include initial state RETENTION More frequent variants: 80%	Figure 4
		average time	Number of students	Final state: RETENTION; DROPOUT Does not include initial state RETENTION More frequent variants: 80%	Figure 5
	(P2) Most frequent backpacks	Number of students; % students	Number of students	Does include state A; ACQ; Q Final state: RETENTION; DROPOUT	Figure 6
		Number of students	Number of students; % students	Does include state A More frequent variants: 90%	Figure 7a
		Number of students; average time	Number of students; % students	Does include state ACQ More frequent variants: 90%	Figure 7b
		Number of students; average time	Number of students; % students	Does include state Q More frequent variants: 90%	Figure 7c

Table 3. Cont.

Model	Perspective	Node Type	Transition Type	Filters	Figure
BPPM-S	(P3) Size of the backpack	Number of students; average time	Number of students; % students	Final state: RETENTION Does not include initial state RETENTION	Figure 8a
		Number of students; average time	Number of students; % students	Final state: DROPOUT	Figure 8b
		Number of students; % students	Number of students	Initial state: BP-1; BP-2; BP-3; BP-4 Final state: RETENTION; DROPOUT	Figure 9
		Number of students; average time	Number of students; % students	Initial state: BP-1 Final state: DROPOUT	Figure 10a
		Number of students; average time	Number of students; % students	Initial state: BP-2 Final state: DROPOUT	Figure 10b
		Number of students; average time	Number of students; % students	Initial state: BP-3 Final state: DROPOUT	Figure 10c
		Number of students; average time	Number of students; % students	Initial state: BP-4 Final state: DROPOUT	Figure 10d

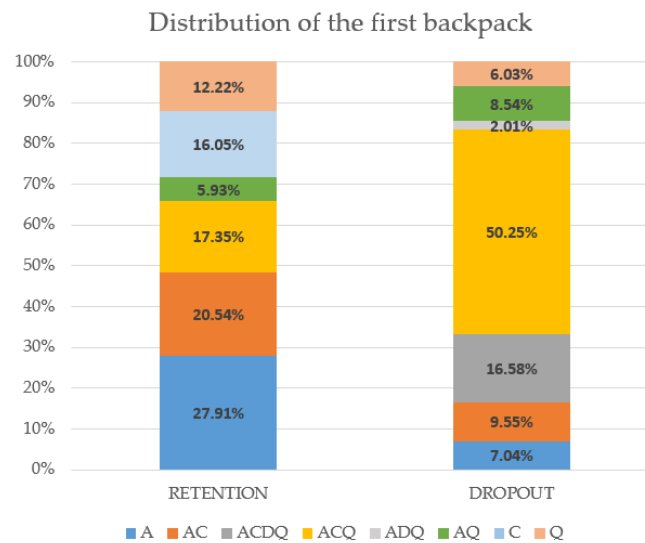


Figure 4. Percentage of educational trajectories, according to the BPPM model, that started with each backpack. The graph considers only the most frequent variants for each model, corresponding to 80% of the students. The RETENTION column considers only students who had a backpack and emptied it (1383 cases, seven more frequent variants). The DROPOUT columns consider only students who had a backpack, were not able to empty it and finally dropped out (199 cases, 13 more frequent variants).

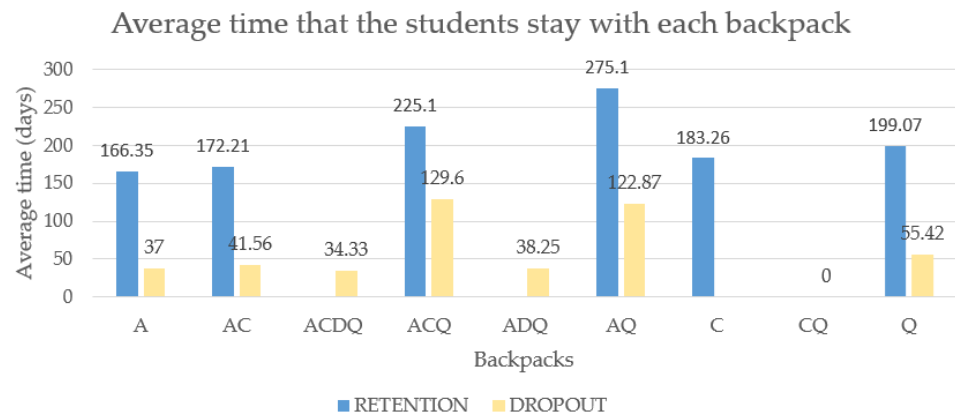


Figure 5. Average time (in days) that the students stay in each backpack. The graph considers only the most frequent variants for each model, corresponding to 80% of the students. The RETENTION columns consider only students who had a backpack and emptied it. The DROPOUT columns consider only students who had a backpack, were not able to empty it and finally dropped out.

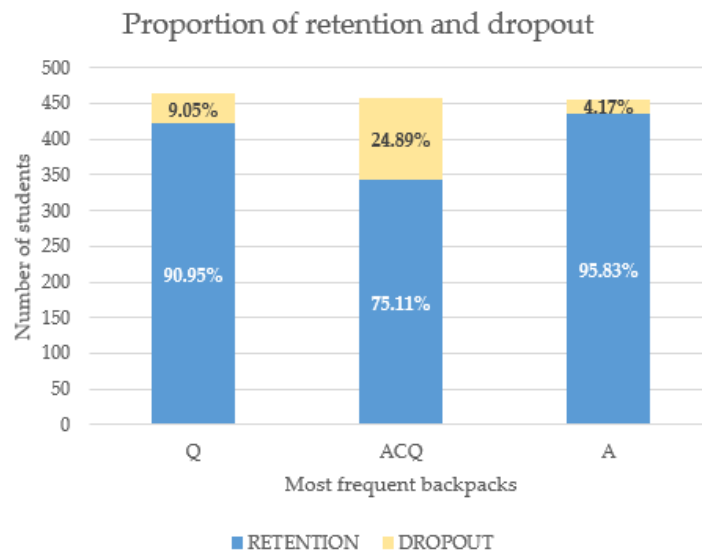


Figure 6. Proportion of educational trajectories, according to the BPPM model, that includes each one of the three most frequent backpacks (Q, ACQ and A) and end either in RETENTION or in DROPOUT.

For the BPPM model, a comparison between the backpack trajectories that ended in DROPOUT or RETENTION was performed, as well as the backpack trajectories that include the three most frequent backpacks were analyzed. For the BPPM-S model, the trajectories in relation to the backpack size that ended in DROPOUT and RETENTION were compared, as well as the backpack trajectories that start with different sizes of backpack.

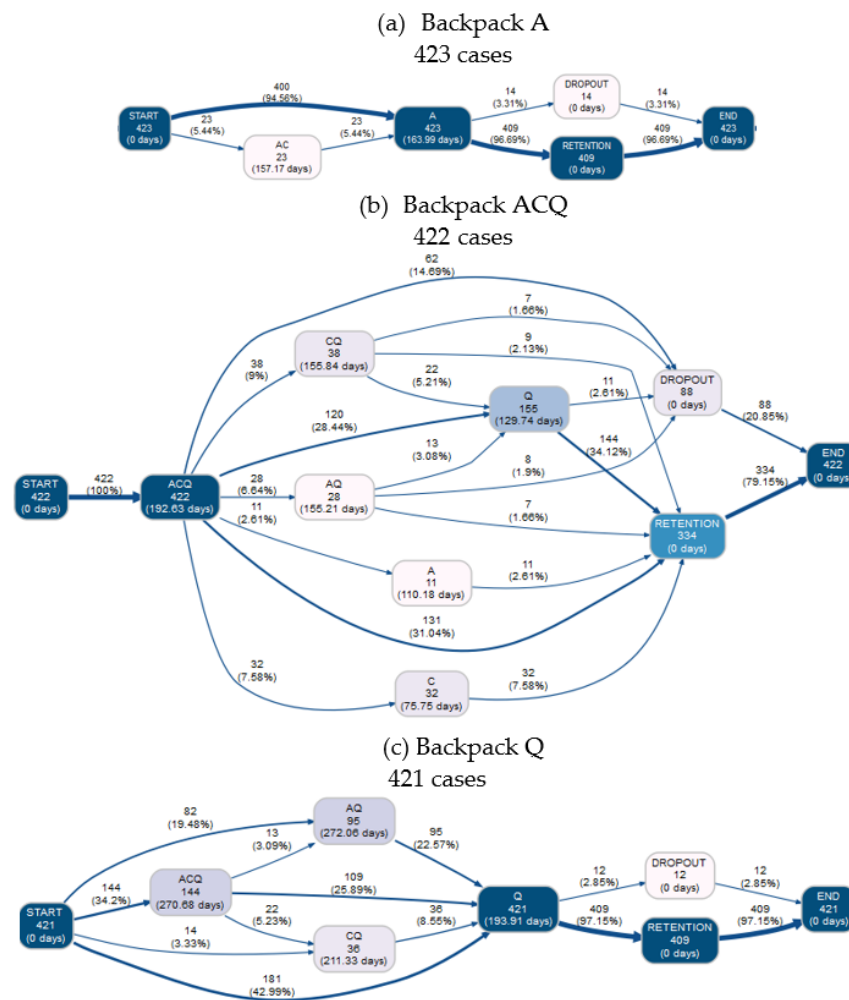


Figure 7. Educational trajectories, according to the BPPM model, which includes the three most frequent backpacks, showing only the most frequent variants, which correspond to 90% of the students in each case. (a) Shows backpack trajectories that include the A backpack. (b) Shows backpack trajectories that include the ACQ backpack. (c) Shows backpack trajectories that include the Q backpack. The darker color of the nodes represents a higher percentage of students who went through a state. The thickness of the arrows represents the percentage of students who had transitions between both states. All values are percentages in relation to the total number of students included in each model.

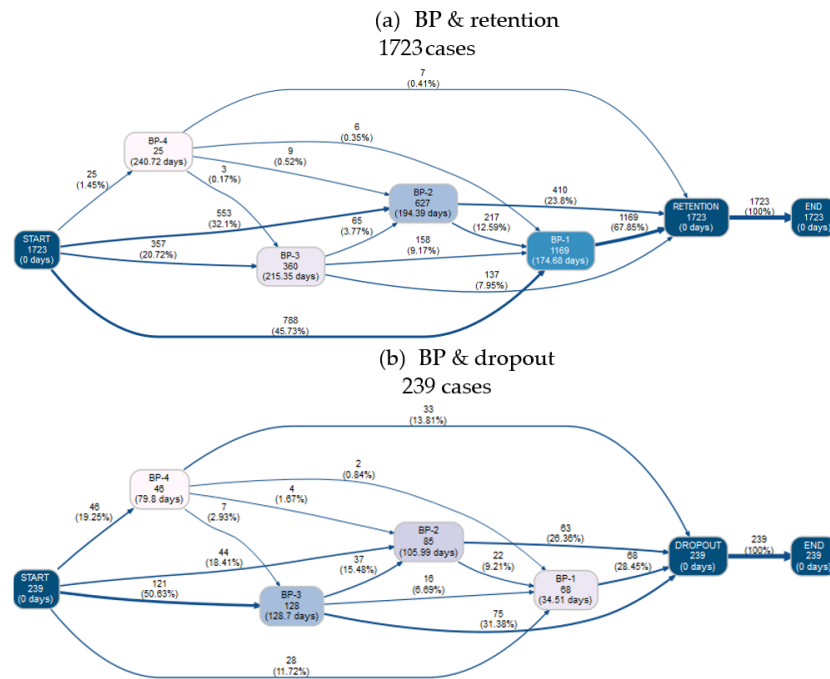


Figure 8. Educational trajectories, according to the BPPM-S model. (a) Shows only students who had a backpack and emptied it. (b) Shows only students who had a backpack, were not able to empty it and finally dropped out. The darker color of the nodes represents a higher percentage of students who went through a state. The thickness of the arrows represents the percentage of students who had transitions between both states. All values are percentages in relation to the total number of students included in each model.

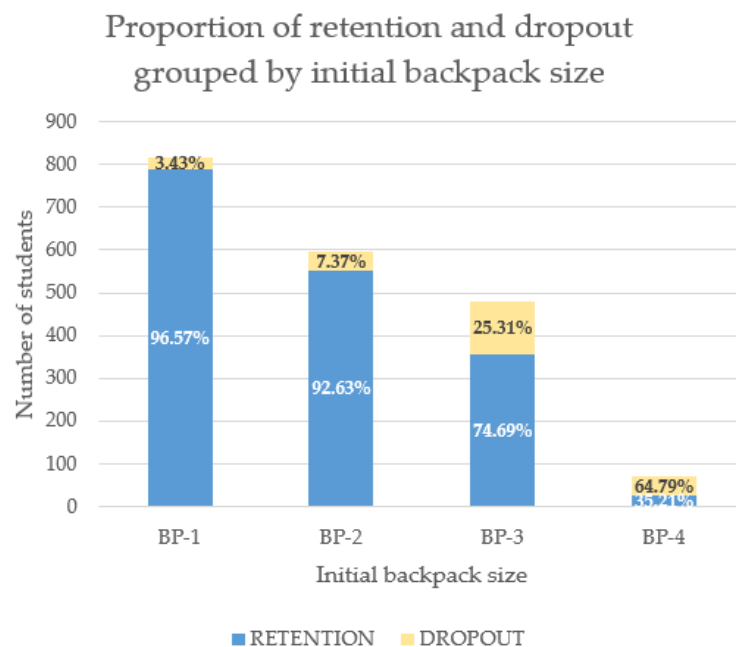


Figure 9. Number and proportion of students who, according to the BPPM-S model, either finished in RETENTION or in DROPOUT, grouped according to the initial backpack size.

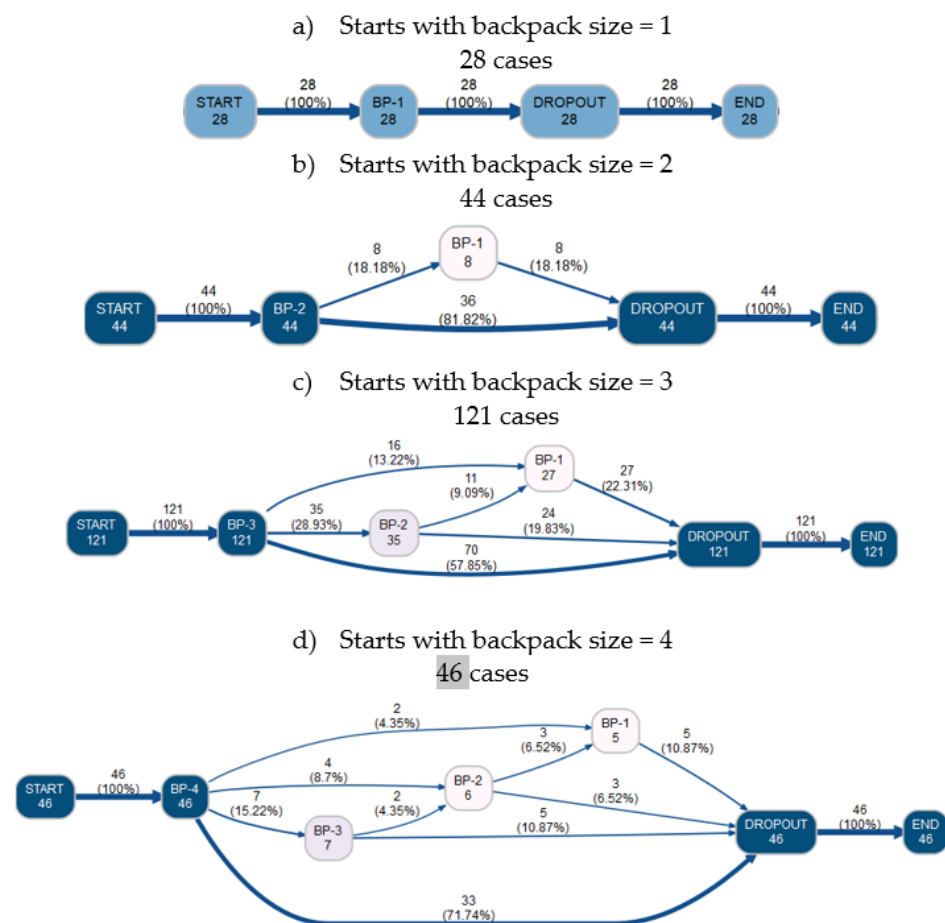


Figure 10. Educational trajectories that ended in DROPOUT, according to the BPPM-S model. (a) Shows only students who started with only one course in the backpack. (b) Shows only students who started with two courses in the backpack. (c) Shows only students who started with three courses in the backpack. (d) Shows only students who started with four courses in the backpack. The darker color of the nodes represents a higher percentage of students who went through a state. The thickness of the arrows represents the percentage of students who had transitions between both states. All values are percentages in relation to the total number of students included in each model.

5. Application Case: First Engineering Courses

This section illustrates the usefulness of the BPPM and BPPM-S models through a real application case that analyzes the backpack trajectories for $N = 4466$ engineering students from a Latin American university. Specifically, the trajectories they followed to take the first four courses of the curriculum were analyzed. The courses are calculus (C), algebra (A), chemistry (Q) and innovation (D). All four courses are automatically enrolled at the beginning of the first semester. The $N = 4466$ correspond to the students of the 2013 to 2019 cohorts, who passed the four courses or dropped out after having failed any of these courses. Specifically, the following three perspectives are analyzed: (P1) BPPM trajectories, ending either in retention (at the undergraduate program) or in dropout; (P2) most frequent backpacks; and (P3) the size of the backpack.

(P1) BPPM trajectories ending either in retention or in dropout

The BPPM model allows us to compare the backpack trajectories between dropout students and those who remained. In particular, the differences can be seen in the distribution of the variants, the relative frequency of each backpack, the elapsed time in the entire trajectory and the average time students spend with each backpack.

Table 4 shows three groups of trajectories. The first of these corresponds to those that include only the empty backpack (No BP). This is the most frequent variant (2504 cases)

and also the shortest (a single event). The second group corresponds to those students who, having failed one or more courses, manage to empty the backpack and remain in the study program. The third group corresponds to those students who, having failed one or more courses, drop out of the study program without having managed to empty the backpack.

Table 4 also shows that, for the trajectories that include backpacks, both those that drop out and those who remain show a high variability (51 and 40 variants, respectively). In the same way, the number of backpack events is similar between students who remain and those who drop out. However, the average time that students who drop out stay with a non-empty backpack is significantly lower ($p < 0.01$).

Table 4. Statistics for BPPM trajectories.

Statistics	No BP	BP & Retention	BP & Dropout
Number of cases	2504	1723	239
Number of variants	1	51	40
Average number of BP events	0	1.27	1.37
Std. dev number of BP events	0	0.52	0.62
Mean time BP (days)	0	237.74	131.80
Std. dev time BP (days)	0	183.71	178.48

Figure 4 shows that the first backpack of students who remain in their undergraduate programs are more evenly distributed, compared to the first backpack of students who dropped out, where close to half (50.25% of cases) began with backpack ACQ. An institution seeking to reduce the early dropout risk could use this information, for example, to implement support mechanisms for students who have failed certain courses simultaneously or change the design of its curricula to prevent certain combinations of failed courses from occurring at high frequencies.

When the average time that the students stay with each backpack is compared, it is possible to see that the average time that the dropout students stay with each backpack is less than the average time for students who remain. Figure 5 shows that the average time that students who ended in RETENTION stay with each backpack, varies from 166.35 to 275.1 days. In contrast, Figure 5 also shows that the average time that students who ended in DROPOUT stay with each backpack, varies from 0 to 129.6 days. In particular, those backpacks with an average duration of much less than one semester (A, AC, ACDQ, ADQ, CQ, Q), show that a significant proportion of students who fall into this situation dropout without even retaking such courses or attempting to pass them.

(P2) most frequent backpacks.

The BPPM model also allows us to compare the backpack trajectories that include specific backpacks. For the three most frequent backpacks: Q (464 cases), ACQ (458 cases) and A (456 cases), Figure 6 shows the proportion of educational trajectories, according to the BPPM model, that ended either in RETENTION or in DROPOUT. While there are differences in the proportion of students who dropped out for each backpack, in all cases the majority of students remained. Following, a more fine-grained analysis is presented, to illustrate the differences in the educational trajectories that include each backpack.

Figure 7 shows the 90% most frequent variants of the students who had backpack A, ACQ and Q. Figure 7a shows that the vast majority (94.56%) of the students had backpack A as the first and the only backpack in their trajectories.

Figure 7b shows that the backpack ACQ was the first backpack for all of these students. Most of those who dropped out (62 over 88) had direct transitions from ACQ to DROPOUT. In contrast, the majority of students who remained emptied their backpack after several stages. The institution could then encourage those students who have the ACQ backpack not to take such failed courses simultaneously by suggesting a certain sequence. In this case, the vast majority of students who defer the Q course remain.

Figure 7c shows that most students did not have the backpack Q as the first backpack, but they had it after passing one or more courses they had previously failed. This behavior could show a sort of prioritization of students who have multiple courses in their backpacks, postponing taking course Q. Furthermore, only a small minority of the students who had this backpack, ended in DROPOUT. This reinforces the idea that students who remain, and have failed a course, have given higher priority to courses other than Q. The institution should then analyze the possibility of postponing this course in the study plan.

(P3) size of the backpack.

The BPPM-S model allows us to compare the educational trajectories across different backpack sizes. This study illustrates the comparison between backpack trajectories for students who dropout and students who remain, as well as the comparison between those that start with different backpack sizes.

Figure 8a shows that backpack trajectories for students who manage to empty it. They mostly start with a backpack size of 1 or 2 and most students who start with a larger backpack, reduce its size before emptying it, going through BP-1. In addition, the average time that students spend with a given backpack size is longer than one semester (150 days, approximately).

Figure 8b shows that most backpack trajectories for dropout students start with a larger backpack size. Moreover, for backpack sizes larger than 1, there are mainly direct transitions from the nodes to dropout.

Figure 9 shows the proportion of students who dropped out or remained, grouping them according to the initial backpack size. Most students whose initial backpack size is less than 4 (BP-1, BP-2 and BP-3), emptied their backpack and remained in their undergraduate programs. 96.57% of students who started in BP-1, 92.63% of students who started in BP-2 and 74.69% of students who started in BP-3, ended in RETENTION. In contrast, only 35.21% of students who started in BP-4, ended in RETENTION.

Figure 10 shows that, in all cases, the majority of students who dropped out, have direct transitions from the initial backpack to dropout. The above shows the importance of defining support strategies for students who simultaneously fail several courses, as well as the need to review the curriculum, evaluating the placement of several high-failure rate courses in the same semester.

6. Discussion

In this paper, the Backpack Process Model (BPPM) was presented, a model that allows systematizing the analysis of curricular trajectories using the backpack metaphor through PM techniques. Its purpose is to represent the psychological burden that students perceive while they have failed courses that they must retake. This model offers a new alternative to analyze curricular trajectories and contribute to understanding why a student remains or drops out from their undergraduate program. This model will help to carry out timely interventions that allow the retention of students at risk of dropping out. We believe that this approach is relevant for an international audience given that global participation in higher education has grown in many countries [45], with HEIs receiving more heterogeneous students, in terms of prior preparation, socioeconomic background and beliefs about learning. Actually, students show more complex enrollment patterns [46] and more variability in their academic results [8]. In these contexts, curricular analytics models that are based on recent research and go beyond the analysis of curricular records can be useful.

We highlight the main findings in the application case:

First, BPPM shows differences between the backpack trajectories that ended in retention or in dropout. Almost half of the students who ended up dropping out start with the ACQ backpack and most of them did not empty their backpacks before dropping out. On the other hand, most of the students who partially emptied their backpacks in more than one stage, even in several semesters, remained. According with Stump, Husman & Corby [47], this difference in their behaviors could be explained by their beliefs about

the nature of intelligence and whether it can be developed or not. Students with incremental beliefs about intelligence and self-efficacy may try harder in higher education [19]. Those with less successful initial trajectories but who nevertheless remain and eventually finish their undergraduate programs are termed struggling persisters [48] and they have received more attention while the proportion of less prepared student has increased in higher education [45].

Second, BPPM-S shows differences in the proportion of students who dropped out or remained, depending on the initial size of the backpack. The larger initial backpack size was associated with a higher proportion of dropouts and also with a higher proportion of direct transitions to dropout. In the case of engineering, the competitive culture and the nature of the first courses as contributors to a process of “natural selection” [49] have been used before to explain this phenomenon. The HEI in which this application case was made is highly selective [50] and previous self-efficacy beliefs are expected to influence the decision to stay, although the initial results are not satisfactory. It could explain that only students with the largest backpack size ended up dropping out at a higher rate. According to Snyder et al. [19], positive beliefs about effort were moderately associated with the success of the first semester in engineering, but the association between their beliefs about effort and the decision to stay was found to be stronger.

Descriptive statistics on BPPM have shown interesting findings, related with backpack size and frequency of each backpack type, as well as their relationship with the student decision on drop out or remain. Nevertheless, the expressive power of BPPM goes further, providing insights on the dynamic behavior of students when managing their backpacks. The sequence followed by students who drop out or remain to empty their backpacks, as well as the time it takes to do so, are good examples. PM tools, combined with domain models, provide a powerful instrument to obtain a deeper understanding of the dynamic behavior of students [30,35].

The findings in the application case should not be taken as general conclusions, but as examples of the expressive power of the BPPM and BPPM-S models.

This application case has two main limitations and the conclusions drawn from it should take them into account. First, the conclusions derived from the BPPM and from process mining in general depend largely on the accuracy and completeness of the information used [51]. The BPPM uses data extracted from the curriculum dimension, so to obtain a deep understanding of the student’s decisions, it should be used as a complement to other information sources. Second, the conclusions derived from the application case study should not be considered as general findings because this study was carried out in a specific institution and time window and only first-semester courses were considered.

7. Conclusions

From the PM perspective, the contribution of BPPM to CA is twofold: Firstly, it systematizes the analysis of curricular trajectories based on the backpack metaphor, characterizing students with similar behaviors in similar contexts. BPPM can improve the understanding of how students handle failed courses they must retake in each study program and in which sequences students stay or drop out most often. Furthermore, the BPPM shows how it is possible to integrate methodologies for sequence analysis and give them a specific meaning in study contexts such as higher education. It is a way of complementing other study metrics to seek an understanding of the educational trajectories that lead to dropout [52].

Secondly, BPPM can help managers and policymakers because the analysis of educational data can help to design and implement timely interventions. Backpack monitoring could be implemented in HEIs, to support counseling. While academic performance is a strong predictor of retention [53], student’s beliefs about the usefulness of effort have a significant influence on academic performance [19]. Good counseling services use technology to identify students at risk [6,27] and BPPM could be used as a complement to identify these students. Additionally, understanding how students handle their backpacks could be used to redesign the curriculum, to reduce the risk of students getting a very large

backpack size and to improve student satisfaction with the curriculum. These kinds of decisions could help to reduce early dropout.

We believe that the BPPM and BPPM-S models could be used to analyze longer curricular trajectories that include the entire study plan. This analysis could help to understand the impact of failed courses that students must retake on late dropout and stop out decisions. The application of trace clustering techniques, in combination with DFGs, could be useful to reduce the complexity of longer BPPM educational trajectories, decomposing traces into smaller and more understandable backpack trajectories. In this context, hierarchical clustering [54] looks promising for future works. Furthermore, qualitative analysis could expand the understanding of students' beliefs about effort and the nature of intelligence [19] on decisions about course taking, stop out and dropout.

Author Contributions: The contributions to this paper are as follows: Conceptualization, J.P.S.-F., J.M.-G. and J.M.-M.; Data curation, J.P.S.-F., J.M.-G. and D.B.; Formal analysis, J.P.S.-F., J.M.-G. and J.M.-M.; Funding acquisition, J.P.S.-F., J.M.-G. and M.S.; Investigation, J.P.S.-F. and J.M.-G.; Methodology, J.P.S.-F. and J.M.-G.; Project administration, J.M.-G. and M.S.; Resources, J.P.S.-F., J.M.-G. and M.S.; Software, J.P.S.-F., J.M.-G. and D.B.; Supervision, J.M.-G. and M.S.; Validation, J.P.S.-F., J.M.-G. and J.M.-M.; Visualization, J.M.-G.; Writing—original draft, J.P.S.-F., J.M.-G. and J.M.-M.; Writing—review & editing, J.M.-G. and M.S. All authors have read and agreed to the published version of the manuscript.

Funding: 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, Escuela de Ingeniería UC/Investigación en Pregrado (IPRE) 2019-1546 and Dirección de Investigación de la Universidad de Cuenca (DIUC) under the project “Analítica de aprendizaje para el estudio de estrategias de aprendizaje autorregulado en un contexto de aprendizaje híbrido” (DIUC_XVIII_2019_54).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: We've decided not to share the data used in this research.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Viberg, O.; Hatakka, M.; Bälter, O.; Mavroudi, A. The current landscape of learning analytics in higher education. *Comput. Hum. Behav.* **2018**, *89*, 98–110. [CrossRef]
- Banihashem, S.K.; Aliabadi, K.; Ardakani, S.P.; Delavar, A.; Ahmadabadi, M.N.; Ardakani, S.P.; Delavar, A. Learning Analytics: A Systematic Literature Review. *Interdiscip. J. Virtual Learn. Med. Sci.* **2018**, *9*. [CrossRef]
- Gottipati, S.; Shankararaman, V. Competency analytics tool: Analyzing curriculum using course competencies. *Educ. Inf. Technol.* **2018**, *23*, 41–60. [CrossRef]
- Hilliger, I.; De Laet, T.; Henríquez, V.; Guerra, J.; Ortiz-Rojas, M.; Zuñiga, M.Á.; Baier, J.; Pérez-Sanagustín, M. For Learners, with Learners: Identifying Indicators for an Academic Advising Dashboard for Students. In Proceedings of the European Conference on Technology Enhanced Learning. EC-TEL 2020: Addressing Global Challenges and Quality Education, Heidelberg, Germany, 14–18 September 2020; Lecture Notes in Computer Science. Alario-Hoyos, C., Rodríguez-Triana, M.J., Scheffel, M., Arnedillo-Sánchez, I., Dennerlein, S.M., Eds.; Springer: Cham, Switzerland, 2020; Volume 12315, pp. 117–130. [CrossRef]
- Ochoa, X. Simple Metrics for Curricular Analytics. In Proceedings of the 1st Learning Analytics for Curriculum and Program Quality Improvement Workshop, Sixth International Learning Analytics and Knowledge Conference (LAK), 25–29 April 2016; Available online: <http://ceur-ws.org/Vol-1590/paper-04.pdf> (accessed on 25 March 2021).
- Levander, L.M.; Mikkola, M. Core Curriculum Analysis: A Tool for Educational Design. *J. Agric. Educ. Ext.* **2009**, *15*, 275–286. [CrossRef]
- Simanca, F.; Crespo, R.G.; Rodríguez-Baena, L.; Burgos, D. Identifying Students at Risk of Failing a Subject by Using Learning Analytics for Subsequent Customised Tutoring. *Appl. Sci.* **2019**, *9*, 448. [CrossRef]
- Mabel, Z.; Britton, T.A. Leaving late: Understanding the extent and predictors of college late departure. *Soc. Sci. Res.* **2018**, *69*, 34–51. [CrossRef]
- Rawatlal, R. Application of Graph Theory to Analysing Student Success Through Development of Progression Maps. In *Engineering Education for a Smart Society. GEDC 2016, WEEF 2016. Advances in Intelligent Systems and Computing*; Auer, M.E., Kim, K.-S., Eds.; Springer: Cham, Switzerland, 2018; Volume 627, pp. 295–307. [CrossRef]

10. Campbell, C.M.; Mislevy, J.L. Student Perceptions Matter: Early Signs of Undergraduate Student Retention/Attrition. *J. Coll. Stud. Retention: Res. Theory Pr.* **2013**, *14*, 467–493. [[CrossRef](#)]
11. Mason, C.; Twomey, J.; Wright, D.; Whitman, L. Predicting Engineering Student Attrition Risk Using a Probabilistic Neural Network and Comparing Results with a Backpropagation Neural Network and Logistic Regression. *Res. High. Educ.* **2018**, *59*, 382–400. [[CrossRef](#)]
12. Zúniga-Prieto, M.A.; Ortiz, M.; Ulloa, M.; Jiménez, A. Applying the LALA Framework for the adoption of a Learning Analytics tool in Latin America: Two case studies in Ecuador. In Proceedings of the Third Latin American Conference on Learning Analytics, Cuenca, Ecuador, 1–2 October 2020. In press.
13. Moreno-Marcos, P.M.; Muñoz-Merino, P.J.; Maldonado-Mahauad, J.; Pérez-Sanagustín, M.; Alario-Hoyos, C.; Kloos, C.D. Temporal analysis for dropout prediction using self-regulated learning strategies in self-paced MOOCs. *Comput. Educ.* **2020**, *145*, 103728. [[CrossRef](#)]
14. Jivet, I.; Scheffel, M.; Specht, M.; Drachler, H. License to evaluate: Preparing learning analytics dashboards for educational practice. In Proceedings of the 8th International Conference on Learning Analytics and Knowledge, Sydney, NSW, Australia, 5–9 March 2018. [[CrossRef](#)]
15. Cairns, A.H.; Gueni, B.; Fhima, M.; Cairns, A.; David, S.; Khelifa, N. Process Mining in the Education Domain. *Int. J. Adv. Intell. Syst.* **2015**, *8*, 219–232. Available online: http://www.ariajournals.org/intelligent_systems/tocv8n12.html (accessed on 25 March 2021).
16. Bogarín, A.; Cerezo, R.; Romero, C. A survey on educational process mining. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2017**, *8*, e1230. [[CrossRef](#)]
17. Salazar-Fernández, J.P.; Sepúlveda, M.; Muñoz-Gama, J. Describing Educational Trajectories of Engineering Students in Individual High-Failure Rate Courses that Lead to Late Dropout. In Proceedings of the Second Latin American Conference on Learning Analytics, Valdivia, Chile, 18–19 March 2019. Available online: <http://ceur-ws.org/Vol-2425/paper03.pdf> (accessed on 25 March 2021).
18. Muñoz-Gama, J.; Maldonado-Mahauad, J.; Salazar-Fernández, J.P.; Bustamante, D.; Sepúlveda, M. Backpack Process Model (BPPM): Curricular Analytics through Process Mining. In Proceedings of the III Conferencia Latinoamericana de Analíticas de Aprendizaje, Cuenca, Ecuador, 1–2 October 2020. In press.
19. Snyder, K.E.; Barr, S.M.; Honken, N.B.; Pittard, C.M.; Ralston, P.A.S. Navigating the First Semester: An Exploration of Short-Term Changes in Motivational Beliefs Among Engineering Undergraduates. *J. Eng. Educ.* **2018**, *107*, 11–29. [[CrossRef](#)]
20. Choi, Y. Student Employment and Persistence: Evidence of Effect Heterogeneity of Student Employment on College Dropout. *Res. High. Educ.* **2018**, *59*, 88–107. [[CrossRef](#)]
21. Geisinger, B.N.; Raman, D.R. Why They Leave: Understanding Student Attrition from Engineering Majors? *Int. J. Eng. Educ.* **2013**, *29*, 914–925. Available online: https://lib.dr.iastate.edu/abe_eng_pubs/607 (accessed on 25 March 2021).
22. Kuley, E.A.; Maw, S.; Fonstad, T. Engineering Student Retention and Attrition Literature Review. In Proceedings of the Canadian Engineering Education Association, Hamilton, ON, Canada, 31 May–3 June 2015. Available online: <https://ojs.library.queensu.ca/index.php/PCEEA/article/view/5813/pdf> (accessed on 25 March 2021).
23. Bandura, A. *Self-Efficacy: The Exercise of Control*; W.H. Freeman: New York, NY, USA, 1997.
24. Meyer, M.; Marx, S. Engineering Dropouts: A Qualitative Examination of Why Undergraduates Leave Engineering. *J. Eng. Educ.* **2014**, *103*, 525–548. [[CrossRef](#)]
25. Roldán-Riejós, A.M.; Úbeda-Mansilla, P. Metaphor use in a specific genre of engineering discourse. *Eur. J. Eng. Educ.* **2006**, *31*, 531–541. [[CrossRef](#)]
26. Lopes, L.L. Between Hope and Fear: The Psychology of Risk. In *Advances in Experimental Social Psychology*; Berkowitz, L., Ed.; Academic Press: Cambridge, MA, USA, 1987; Volume 20, pp. 255–295. [[CrossRef](#)]
27. Arnold, K.E.; Pistilli, M.D. Course signals at Purdue: Using learning analytics to increase student success. In *LAK '12: Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*; ACM: New York, NY, USA, 2012; pp. 267–270. [[CrossRef](#)]
28. van der Aalst, W. Process Mining: The Missing Link. In *Process Mining: Data Science in Action*; van der Aalst, W., Ed.; Springer: Berlin/Heidelberg, Germany, 2016. [[CrossRef](#)]
29. Martin, N.; De Weerd, J.; Fernández-Llatas, C.; Gal, A.; Gatta, R.; Ibáñez, G.; Johnson, O.; Mannhardt, F.; Marco-Ruiz, L.; Mertens, S. Recommendations for enhancing the usability and understandability of process mining in healthcare. *Artif. Intell. Med.* **2020**, *109*, 101962. [[CrossRef](#)]
30. Utama, N.I.; Sutrisnowati, R.A.; Kamal, I.M.; Bae, H.; Park, Y.-J. Mining Shift Work Operation from Event Logs. *Appl. Sci.* **2020**, *10*, 7202. [[CrossRef](#)]
31. Van Der Aalst, W.M. A practitioner's guide to process mining: Limitations of the directly-follows graph. *Procedia Comput. Sci.* **2019**, *164*, 321–328. [[CrossRef](#)]
32. Bin Ahmadon, M.A.; Yamaguchi, S. Verification Method for Accumulative Event Relation of Message Passing Behavior with Process Tree for IoT Systems. *Information* **2020**, *11*, 232. [[CrossRef](#)]
33. van Eck, M.L.; Lu, X.; Leemans, S.J.J.; van der Aalst, W.M.P. PM²: A Process Mining Project Methodology. In *Advanced Information Systems Engineering*; Zdravkovic, J., Kirikova, M., Johannesson, P., Eds.; Springer: Cham, Switzerland, 2015; pp. 297–313. [[CrossRef](#)]

34. Dumas, M.; la Rosa, M.; Mendling, J.; Reijers, H.A. Introduction to Business Process Management. In *Fundamentals of Business Process Management*; Springer: Berlin/Heidelberg, Germany, 2013. [CrossRef]
35. Maldonado-Mahauad, J.; Pérez-Sanagustín, M.; Kizilcec, R.F.; Morales, N.; Muñoz-Gama, J. Mining theory-based patterns from Big data: Identifying self-regulated learning strategies in Massive Open Online Courses. *Comput. Hum. Behav.* **2018**, *80*, 179–196. [CrossRef]
36. Johnson, O.A.; Dhafari, T.B.; Kurniati, A.; Fox, F.; Rojas, E. The ClearPath Method for Care Pathway Process Mining and Simulation. In *Business Process Management Workshops*; Springer International Publishing: Cham, Switzerland, 2019; Volume 342, pp. 239–250. [CrossRef]
37. Lorenz, R.; Senoner, J.; Sihm, W.; Netland, T. Using process mining to improve productivity in make-to-stock manufacturing. *Int. J. Prod. Res.* **2021**, 1–12. [CrossRef]
38. Reimann, P.; Markauskaite, L.; Bannert, M. e-Research and learning theory: What do sequence and process mining methods contribute? *Br. J. Educ. Technol.* **2014**, *45*, 528–540. [CrossRef]
39. Zhou, X.; Zacharewicz, G.; Chen, D.; Chu, D. A Method for Building Service Process Value Model Based on Process Mining. *Appl. Sci.* **2020**, *10*, 7311. [CrossRef]
40. Salazar-Fernandez, J.P.; Sepúlveda, M.; Muñoz-Gama, J. Influence of student diversity on educational trajectories in engineering high-failure rate courses that lead to late dropout. In Proceedings of the 2019 IEEE Global Engineering Education Conference (EDUCON), Dubai, United Arab Emirates, 9–11 April 2019; pp. 607–616. [CrossRef]
41. Berti, A.; van Zelst, S.J.; van der Aalst, W. Process mining for python (PM4Py): Bridging the gap between process-and data science. *arXiv* **2019**, arXiv:1905.06169.
42. Janssenswillen, G.; Depaire, B.; Swennen, M.; Jans, M.; Vanhoof, K. bupaR: Enabling reproducible business process analysis. *Knowl. Based Syst.* **2019**, *163*, 927–930. [CrossRef]
43. Günther, C.W.; Rozinat, A. *Disco: Discover Your Processes*; BPM (Demos): Tallin, Estonia, 2012; Volume 940, pp. 40–44. Available online: <http://ceur-ws.org/Vol-940/paper8.pdf> (accessed on 25 March 2021).
44. Geyer-Klingeberg, J.; Nakladal, J.; Baldauf, F.; Veit, F. Process Mining and Robotic Process Automation: A Perfect Match. In Proceedings of the International Conference on Business Process Management, Sydney, Australia, 9–14 September 2018; pp. 1–8, 124–131. Available online: http://ceur-ws.org/Vol-2196/BPM_2018_paper_28.pdf (accessed on 25 March 2021).
45. Kember, D.; Leung, D.; Prosser, M. Has the open door become a revolving door? The impact on attrition of moving from elite to mass higher education. *Stud. High. Educ.* **2021**, *46*, 258–269. [CrossRef]
46. Rodríguez-Gómez, D.; Meneses, J.; Gairín, J.; Feixas, M.; Muñoz, J.L. They have gone, and now what? Understanding re-enrolment patterns in the Catalan public higher education system. *High. Educ. Res. Dev.* **2016**, *35*, 815–828. [CrossRef]
47. Stump, G.S.; Husman, J.; Corby, M. Engineering Students' Intelligence Beliefs and Learning. *J. Eng. Educ.* **2014**, *103*, 369–387. [CrossRef]
48. Suresh, R. The Relationship between Barrier Courses and Persistence in Engineering. *J. Coll. Stud. Retent. Res. Theory Pr.* **2006**, *8*, 215–239. [CrossRef]
49. Haag, S.; Collofello, J. Engineering undergraduate persistence and contributing factors. In Proceedings of the 38th Annual Frontiers in Education Conference, Saratoga Springs, NY, USA, 22–25 October 2008; pp. T4D-8–T4D-14. [CrossRef]
50. Bernasconi, A. Inclusion Programs at Elite Universities: The Case of Chile. In *Mitigating Inequality: Higher Education Research, Policy, and Practice in an Era of Massification and Stratification*; Advances in Education in Diverse Communities; Emerald Publishing Limited: Bingley, UK, 2015; Volume 11, pp. 303–310. [CrossRef]
51. Bose, R.J.C.; Mans, R.S.; Van Der Aalst, W.M. Wanna improve process mining results? In Proceedings of the 2013 IEEE Symposium on Computational Intelligence and Data Mining (CIDM), Singapore, 16–19 April 2013; IEEE: New York, NY, USA, 2013; pp. 127–134. [CrossRef]
52. Salazar-Fernandez, J.; Sepúlveda, M.; Muñoz-Gama, J.; Nussbaum, M. Curricular Analytics to Characterize Educational Trajectories in High-Failure Rate Courses That Lead to Late Dropout. *Appl. Sci.* **2021**, *11*, 1436. [CrossRef]
53. Sage, A.J.; Cervato, C.; Genschel, U.; Ogilvie, C.A. Combining Academics and Social Engagement: A Major-Specific Early Alert Method to Counter Student Attrition in Science, Technology, Engineering, and Mathematics. *J. Coll. Stud. Retent. Res. Theory Pract.* **2021**, *22*, 611–626. [CrossRef]
54. Tariq, Z.; Khan, N.; Charles, D.; McClean, S.; McChesney, I.; Taylor, P. Understanding Contrail Business Processes through Hierarchical Clustering: A Multi-Stage Framework. *Algorithms* **2020**, *13*, 244. [CrossRef]