



Article

The Use of Educational Process Mining on Dropout and Graduation Data in the Curricula (Re-)Design of Universities

Alexander Karl Ferdinand Loder

Department of Performance and Quality Management, University of Graz, Universitätsplatz 3/I,
8010 Graz, Austria; alexander.loder@uni-graz.at

Abstract: High college dropout rates are not a desired outcome for university management. Efforts have been made to increase student retention via understanding dropouts and building support mechanisms. With the emergence of Big Data, educational process mining came into existence, allowing for new methods of structuring and visualizing data. Previous studies have established an approach to generate process maps from the course sequences students take. This study improves this method by focusing on visualizing students' pathways through a study program dependent on their status as a "dropout" or "graduate" and on the level of every degree program. An interactive framework in a web application dedicated to curriculum designers was created. The data of 53,839 students in 78,495 studies at the University of Graz (Austria) between 2012/13 and 2022/23 were used for process mining. The generated process maps provide information on the exam sequence of students. They have been implemented in discussion forums with stakeholder groups and are part of the curriculum (re)design processes. The maps provide the benefit of being able to compare and monitor successful and non-successful students' maps using real-time data. Despite their use for curriculum development, they are limited in their size and the number of exams that can be displayed, making them a good fit for early dropout evaluation.

Keywords: educational process mining; dropout; curriculum design; academic analytics; student retention



Citation: Loder, A.K.F. The Use of Educational Process Mining on Dropout and Graduation Data in the Curricula (Re-)Design of Universities. *Trends High. Educ.* **2024**, *3*, 50–66. <https://doi.org/10.3390/higheredu3010004>

Academic Editors: Nataša Erceg and Ivica Aviani

Received: 14 December 2023

Revised: 11 January 2024

Accepted: 15 January 2024

Published: 17 January 2024



Copyright: © 2024 by the author. 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

Based on an OECD report, around 12% of students in all OECD regions entering a bachelor's program leave the university during their first year in a degree program [1]. Universities have been concerned with predicting their drop out for a long time [2]. In some countries such as Austria, the university system is publicly funded via student graduation rates, student activity and other indicators that are maximized by students staying in the system until graduation [3]. Keeping students in the system and active, i.e., ensuring they are taking and successfully completing courses, are declared goals of Austrian universities [4]. The potential loss of money, the non-achievement of institutional goals and decreasing enrolment numbers through dropouts [5] can make students' academic performance a crucial aspect of educational institutions [6]. In the case of Austria, the financial loss cannot be compensated via tuition fees, since students can pursue their studies for the minimum required time to finish a degree program plus two additional semesters before universities are allowed to allocate fees to them. In general, enrollment is free, there is no numerus clausus and only some degree programs have entrance exams [4]. Preventing dropouts and increasing student retention will improve the image of an institution and the revenue generation status [7]. Although dropping out from one's studies is a multifactorial process of personal, academic, economic, social and institutional factors [5], the allocation of institutional support can increase student retention [8,9]. In order to deliver the right mechanisms for the right target group, institutional data analysis is either performed on course and departmental level, describing the research field of "learning analytics" [10],

or on institutional, regional and national/international level, which is categorized as “academic analytics” [11,12]. One dimension relevant from both research perspectives is curriculum design, with curricula being an important factor in the academic success of students [13]. However, evidence on how to analyze learning on a higher level to guide curriculum development was lacking and is still growing [14,15]. With the emergence of big databases and information on student behavior the field of educational data mining came into existence and opened new opportunities for evidence-based curriculum design [16,17]. This study presents the development of a descriptive educational data mining method based on student data from both dropouts and graduates, with the goal of aiding curriculum planners in the design processes to increase student retention.

1.1. Data-Driven Curriculum Design

The development or adaptation of a degree program goes through a process of planning, implementation and evaluation. The center of evaluation can be the interactions of people, processes within the curriculum and procedures with it [18]. With curricula being designed toward learning objectives or learning outcomes [19–21], studies have focused on the effects of certain components of conceptual models for a given curriculum or the innovation of teaching strategies [22–24]. Research on data-driven curriculum (re)design is needed [14]. In addition to performing data-based evaluations on course level [14], it was suggested to also gather information at a meso level, addressing the program itself as a whole [25]. When analyzing meso-level data of curricula on the scale of an entire university, methods able to efficiently structure and visualize big datasets need to be applied. One method applicable to these requirements is process mining.

1.2. Educational Data Mining: Process Mining

Data Mining is an umbrella term for a family of methods including process mining. They focus on discovering patterns by processing large amounts of data [26]. These big sources of information are also known as “Big Data” and related research is mainly centered on storing and processing components. Applications using analytical process identification and the analyses of temporal events are not broadly studied [15]. Process mining is a technique that emerged from data mining, being a method for analyzing and visualizing patterns found in large datasets [27,28]. Coming from concerns on how to optimize organizational processes and results, applications include reducing the time to achieve answers and costs, maximizing productivity, balancing resource utilization, improving quality, minimizing risk and improving human well-being [15]. Generally, data mining in education is centered on data dependencies and patterns, does not provide visual representations of the process and does not focus on the process as a whole [29]. Therefore, educational process mining has been proposed [30]. As a subcategory of educational data mining, it concentrates on improving educational processes by obtaining fact-based insights into educational processes [31] with graphs having more relevance and importance [27]. Process mining techniques are based on converting business or educational processes into so-called “event logs”, which are time-coded data of a target variable. They create meaningful process models from temporal series of data retrieved from systems’ databases [32]. Data from university databases has previously been used with success in the context of designing measures for potential dropouts [33], making them a valid source for educational process mining. Previous studies could demonstrate the usefulness of educational process mining in higher education institutions [34,35]. Linking curriculum development to process mining approaches offers opportunities for finding and repairing issues. Previous studies on fictive exams within a six-semester program could visually represent the course-taking sequence of students, underlining the applicability of process discovery methods in program design [36]. Therefore, this study adds another application to the current body of literature, using real educational data on a program level and on the scale of a university.

1.3. Research Aim: Developing an Educational Process Mining Method

This study describes the development of an educational process mining approach, yielding automatically created, interactive and animated process maps that depict student pathways centered on their course-taking sequence. The goal is to create process maps on the program level for all fields of study, including data from all course exams students have taken. Information on the most common pathways students take within every single curriculum of a university is obtained, depicting the sequence of exams prior to dropping out from a program. Previous studies focusing on curriculum design have already established the base of this approach, creating process maps for the course sequences of students [37]. These had, for instance, smaller samples, no automated data curation and user interface, were centered on exams in a single course or did not incorporate animations and real-time data [37]. This study builds upon these existing frameworks, substantially improving the method: (1) It includes data from both dropouts and graduates that will be analyzed to increase the usability of the process maps, enabling comparisons between successful exam sequences with (potentially) less successful exam sequences. (2) The analyses and process mining methods are done on the largest possible scale, i.e., the university level, yielding separate maps for each degree program and curriculum version. (3) Animated process maps are generated interactively in a web application via user input, giving users a set of parameters to explore the data. The app is built upon real-time data, enabling users to monitor curricula in a way static and non-automated approaches do not allow [37]. The implementation of this approach for curriculum designers at a mid-European university is discussed.

2. Materials and Methods

2.1. Data Background

The data used in process mining is retrieved from the student database of the University of Graz. It is Austria's second-biggest university with approximately 30,000 enrolled students per academic year. The database is automatically synchronized every week on Monday, using live data of the last ten academic years. After finishing the development of this process mining approach in the academic year 2022/23, the database included every exam between 2012/13 and 2022/23 up to this point from students that either dropped out of their studies or graduated. A total of 41,755 students in 53,852 studies from dropouts and 19,731 students in 24,643 studies from graduates were included (53,839 students in 78,495 studies total). The dataset for the application of the proposed method consisted of the dates and times of 580,652 exams from dropouts and 1,329,206 exams from graduates (1,909,858 total). No-shows were excluded, meaning dropouts without any exam data.

One "study", i.e., one program, means one full curriculum and one degree program, respectively. In Austria, it is possible that one student enrolls in an unlimited number of programs at once, which explains the higher number of studies/study programs compared to the number of students. The higher number of programs compared to the number of students can be explained by the fact that parallel enrollments are common in Austria due to the absence of tuition fees and other regulations prohibiting simultaneous multiple enrollments. For the purpose of this study, dropouts are defined on the program level, matching the structure of the indicators of the funding system of Austria [3]. This means that one dropout is counted for each study program that is closed, which makes it possible that one student can create multiple dropouts on the university level. However, process mining is performed on a program level, giving insights into the students' course-taking sequence within each curriculum. Including the university aggregation level would make no sense, since the processes depicted in the result of the proposed method are bound to absolute frequencies. This would only produce outcomes for the largest programs. In addition, there is no data in the internal database that allows a control for students switching universities after leaving the University of Graz. There is also no possibility to check for previous institutions students might have come from. This is because the universities in Austria have different campus management systems and their databases are

not connected to each other. For the purpose of obtaining the estimates of student mobility after leaving one university, the project STUDMON (“Student Monitoring”) that included most Austrian universities merged data from their databases and created results on the program level per university. It was found that student transitions exist at the University of Graz, but are generally low. On the curriculum level of Molecular Biology, for instance, less than 20 students out of around 260–300 per observed cohort closed their studies and switched to the Medical University of Graz [38]. The numbers suggest that this happens mainly for the reasons of keeping their student status valid until they can take part in the next entrance exam for medicine. From both the funding and the curriculum design standpoint, these transitions are not crucial, since the goal is to keep students in the system and to keep them active [4].

The programs included in the dataset were exclusively bachelor’s, master’s or diploma degree programs. Bachelor’s programs are intended to be finished within six semesters (with some exceptions in eight semesters), master’s programs within four and diploma programs within eight semesters. The diploma degree is a program type specific to Austria, which is segmented into two or three sections, similar to the transition from bachelor to master. However, this type involves one single curriculum and only one degree once the program is finished. After graduation the degree “Magister (Mag.)” is awarded. Diploma programs were Austria’s original program type before the transition to the international bachelor’s/master’s system. Some fields of study at different institution types have not been changed in the process, which is why diploma programs still exist. Examples are law, at general universities, or acting, as well as scenery, at universities of arts. The versions of the curricula included in the dataset needed to be established in 2010 or later. With the first versions of digital databases and campus management software being introduced in the mid-2000’s, 2010 serves as a reliable cutoff for data stability. The reason for including ten years of data (2012/13–2022/23) is due to runtime, computational resources on the server and the actuality of the results. It is assumed that this timeframe is enough to both make comparisons to a small number of historic curricula and exclude older versions that are of no more interest today, considering several revisions that have likely happened over ten years.

Process mining was performed for each program and status (dropout/graduation) separately, making a process map depict one curriculum version dependent on a given student status (dropout, graduate). Different curricula versions were treated as different programs since intended course sequences can differ depending on this factor.

The analysis of the data and the permission for publication have been approved in a works agreement between the Performance- and Quality Management Department and the works committee of the University of Graz, called “Rahmen-BV IKT 2019”, which includes ethics approval. Furthermore, data analyses on student data by staff departments of the university are regulated by law, defining analyses as a means of increasing the graduation rates of Austrian universities as an obligation [4].

2.2. Event Logs

Exam dates and times were transformed into so-called event logs and used as the main process mining variable. Event logs refer to a combination of a date and time, when an event happens, and the name of the event. This one variable contains the information of when a given student has taken a specific exam.

In this study, the term exams are defined as the major performance review of a given course. In Austria, it is possible to generally repeat an exam three times after the first failed attempt. After failing four times at the same exam or course, students are banned from any additional attempts for the same exam or course at their enrolled institution [4]. In most cases this means that completion of the curriculum is not possible and dropping out is inevitable, except for successfully passing an equivalent at a different institution. However, this involves getting permissions and an extensive validation process by the deanery.

In lectures, the term “exam” refers to a real exam that usually takes place at the end of the semester. However, different from lectures, other course types can have several examinations across a semester or are based on continuous assessment. Most of them do not have one single examination at the end of the semester. This means that students’ grades are also dependent on their behavior in a course. For instance, seminars are types of courses with continuous assessment. In these cases, the point in time when a grade is given, i.e., the “exam” date, is defined as the date and time the grade for the course is officially released in the campus management system. If the successful completion of a course requires a project paper instead of any of the aforementioned criteria, the grade for all students is generally released after a predefined deadline. Partial test dates do not exist and could therefore not be included. Exams were included independent of the grade, no matter if a course has been passed or not. Retries to pass the same exams or courses in immediate succession, i.e., without taking any other exam in between, are handled by the mining algorithm and depicted as a loop on the same box.

Course enrollment alone was not considered a criterion, since there would have been no option to control for student engagement. This means that a student gets a grade for a course once a valid enrollment exists in the system. Prior to a given deadline at the beginning of each semester (around one month), canceling course enrollments is possible, no matter if students have actively taken part in the course or not. These cases are not included in the dataset. If a student drops out of the course during the semester after this time limit, the course is not considered successfully completed and a negative grade is registered. However, there is no data available about the exact point in time when this happens, only the date when the grade was released. For lack of information on the reasons why a given grade was registered, these cases have to be included. In line with this, there is another reason for the exclusion of course enrollment, which is lecture exams. In Austria, students do not have to be registered to take part in a lecture, which is different from other course types. Lectures have no compulsory attendance. Still, they can register to take an exam or a lecture, independently of their registration status. Including course enrollment would include certain gaps, which is not the case for only including cases with a grade.

Additionally, the enrollment and dropout/graduation dates as well as the completion of possible program segments in diploma studies for each student were added to the exam data as starting and ending points of the visualizations. In the Austrian university system, diploma programs are a third type of degree program besides bachelor’s and master’s programs. Before the Bologna process that led to the introduction of the latter two, this was the standard program type at Austrian universities. It is separated into two or three segments, similar to finishing an undergraduate program and continuing with a major, but there is no formal graduation until all requirements of the entire curriculum have been completed [4].

2.3. Process Map Design

By transforming exam data into event logs, the titles of the courses constitute the nodes of the process maps. Interlinks between nodes were calculated by counting the number of traces up to the point of becoming a “dropout” in the maps or by limiting the length of the pathways for graduates via user input. Processes were sorted by absolute frequencies, beginning with the highest number of equal processes. Automatic color coding was used on the generated course boxes to represent higher and lower frequencies. Lower frequencies are depicted as light blue, getting darker and transitioning into the orange and red spectrum the higher the frequency. Each student was depicted as a yellow dot, moving through the process map. The absolute frequencies of the occurrence of a path (how many students did two exams in the same succession) was depicted as a number on the side of each trace between two nodes. These can be switched to relative frequencies using the “Frequencies” button in the user interface, as stated below. Time was represented by an automatically scrolling timeline at the bottom of the plot. The overall framework for interactive calculations and user input was provided via a web server. Each process map

can be adjusted interactively using the following options, which are accessible via buttons and sliders in the web interface:

- Information on the curriculum: field of study, type of degree and curriculum version can be selected to apply the filters to the output.
- Student status: this switch defines the data source dependent on student status, i.e., dropouts or graduates.
- Top-n processes: the number of processes displayed in the maps, from highest frequency to lowest. The animated plots are accompanied by a table, listing the top-n processes per frequency. It is intended as a helper in choosing the optimal settings. The higher the number, the more processes and students are included. The range of the settings is top-1 to top-50 processes.
- Length of processes: the number of exams per process displayed. Every process with a maximum number of exams equal to or lower than the chosen setting will be included. Longer processes for dropouts will not be displayed if the cutoff is set too small. Graduates' processes are shortened by this variable prior to the calculations. The higher the number, the more processes and students are included. The range of settings is one to ten courses.
- Frequencies: a switch to choose between absolute or relative frequencies.
- Line type: the rendering of the map can be influenced by changing the line type (round or straight). In some process maps changes improve the structuring of the plot.

2.4. Apparatus

Data were retrieved via an Oracle[®] SQL Server database. The curation of the data was accomplished using R 4.2.2 [39]: queries were performed using the RODBC package [40], process mining and process mapping were done via bupar and processmapR [41,42] and process maps were animated with the processanimateR package [43]. The shiny package is used as the web interface [44] on a Linux Ubuntu web server.

2.5. Algorithm and Data Model

The packages mentioned are not based on different algorithms that mine and depict the data on their own. They are meant to visualize a series of event logs in a structured way, for instance, so that pathways do not overlap. This makes them an instrument for visual presentation. The algorithm and mining routine used to retrieve the raw data from the database and to transform it into a format that can be read by the packages was programmed specifically for this project and is schematically presented in Figure 1 and explained below.

The project is located in a directory on a web server and contains two major components: (1) data curation and (2) the web app. The web app is constantly running and can be accessed via a webpage. It uses data that has previously been brought in a format that ensures fast processing. An automatic scheduler triggers these data updates once per week. This is done by starting a meta script, which coordinates all the steps until the data is ready to be accessed by the web app. It sets up the data processing environment, based on the current date. This is important, as only data within a target timeframe of ten academic years plus the current academic year should be included. The meta script then loads the SQL files, adapts their filters to this timeframe and runs them. These SQL scripts connect to the database, query the data and feed it back to the meta script. In the next step, the raw data is sent to the main data curation script. The data is then filtered, cleared and aggregated on course and program level per student. This means that each student can have several rows per study program, with each row containing one exam. Using the names of the exams, dates and time values, rows are transformed into an event log format. The results are sent to the meta script, which stores the data and updated settings in the app's directory on the web server. Every time the URL is accessed, the web app loads the data from the directory, making it immediately accessible in a user's session. Upon user input, the web app triggers processing functions, making the data from the app directory accessible. Depending on

the user input, different functions filter the data, e.g., on a specific curriculum level, and generate the process maps. The maps are then displayed on the screen, visible to the user.

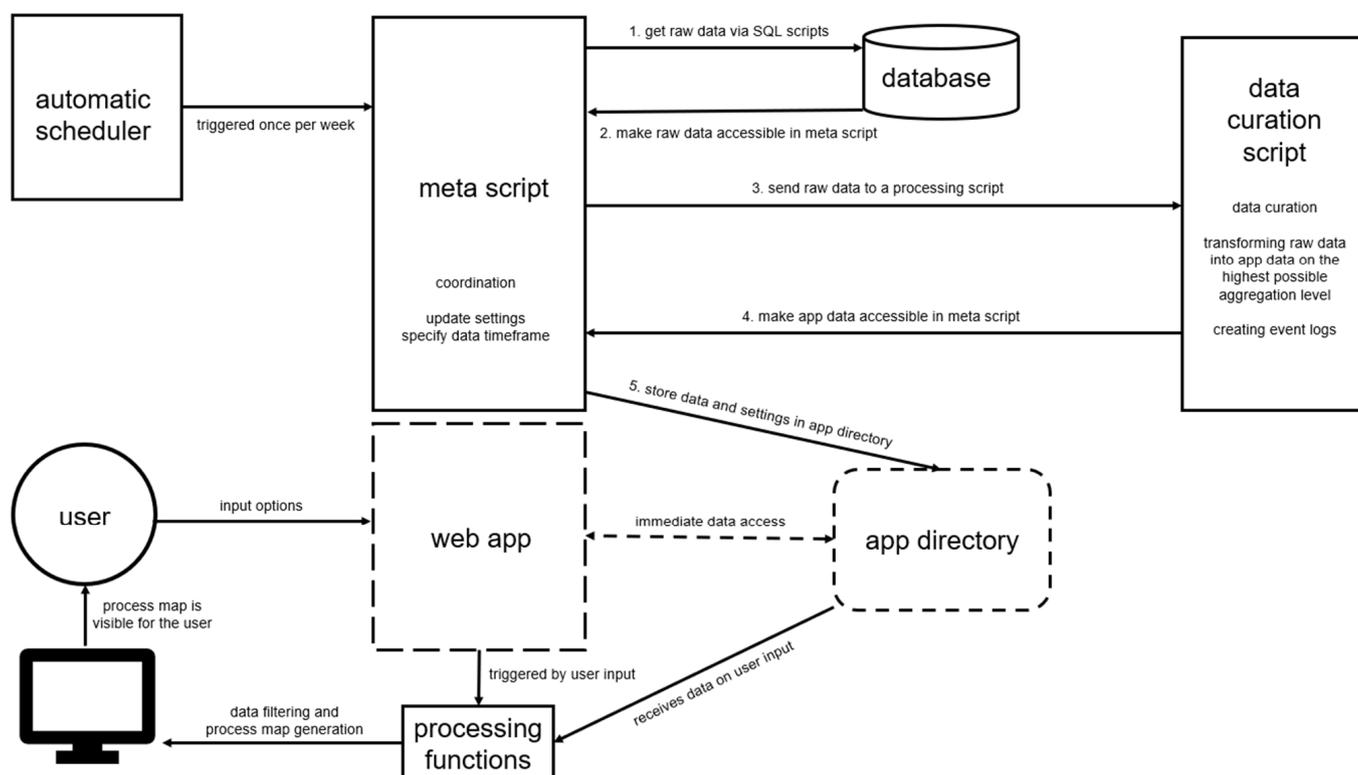


Figure 1. Schematic representation of the process mining algorithm.

3. Results

The figures in the following sections give examples of the processes of dropouts and graduates in the curriculum “law” of the University of Graz using different settings. Each node in the maps depicts one exam, except for the starting and end points (orange), which show the beginning of studies (top) and the end of studies for dropouts or the cutoff of n -exams with continuation for graduates (bottom), respectively. Due to law being a diploma program in Graz, the completion of segments in a graduates’ plot is also shown as orange boxes near the ending point. Each plot contains a timeline, which is synchronized with the animation of the plot, showing the exact point in time linked to the movements of the individual students. Animated versions of the Figures can be found in the Supplementary Material.

Each plot should be read by starting at the “start of the program” node. The number beneath the title represents the number of student pathways included in the plot. From there, the pathways branch out and create different traces, showing the course sequence of students. Each trace also depicts the number of students that transition from one exam to the next. Thicker traces indicate higher frequencies. Red and orange color coding does the same for nodes, while light blue and dark blue is used for lower frequencies. Following the bold black traces, the “main course sequences” with the highest frequencies can be identified. Figure 2 for instance, shows that 541 students fulfilled the criteria defined in the settings as dropping out after taking between one to five courses. Among them, 360 students got a grade in the “Orientation course in Law” as their first course and 181 took “The legal case as an introduction to law” first. Following those 360 students, 276 of them then got a grade in “The legal case as an introduction to law” course, while 84 closed their program without taking another exam. Of the total of 457 students taking the exam in this introductory course, an additional 391 students closed their studies afterwards. Others did some other courses with lower frequency measures, but ended up dropping out after a

maximum of five courses as defined in the settings in the user interface. In comparison, Figure 3 shows successful students with a graduation. Their course sequences are cut off after their first five courses. Via the settings, the cutoff can be increased.

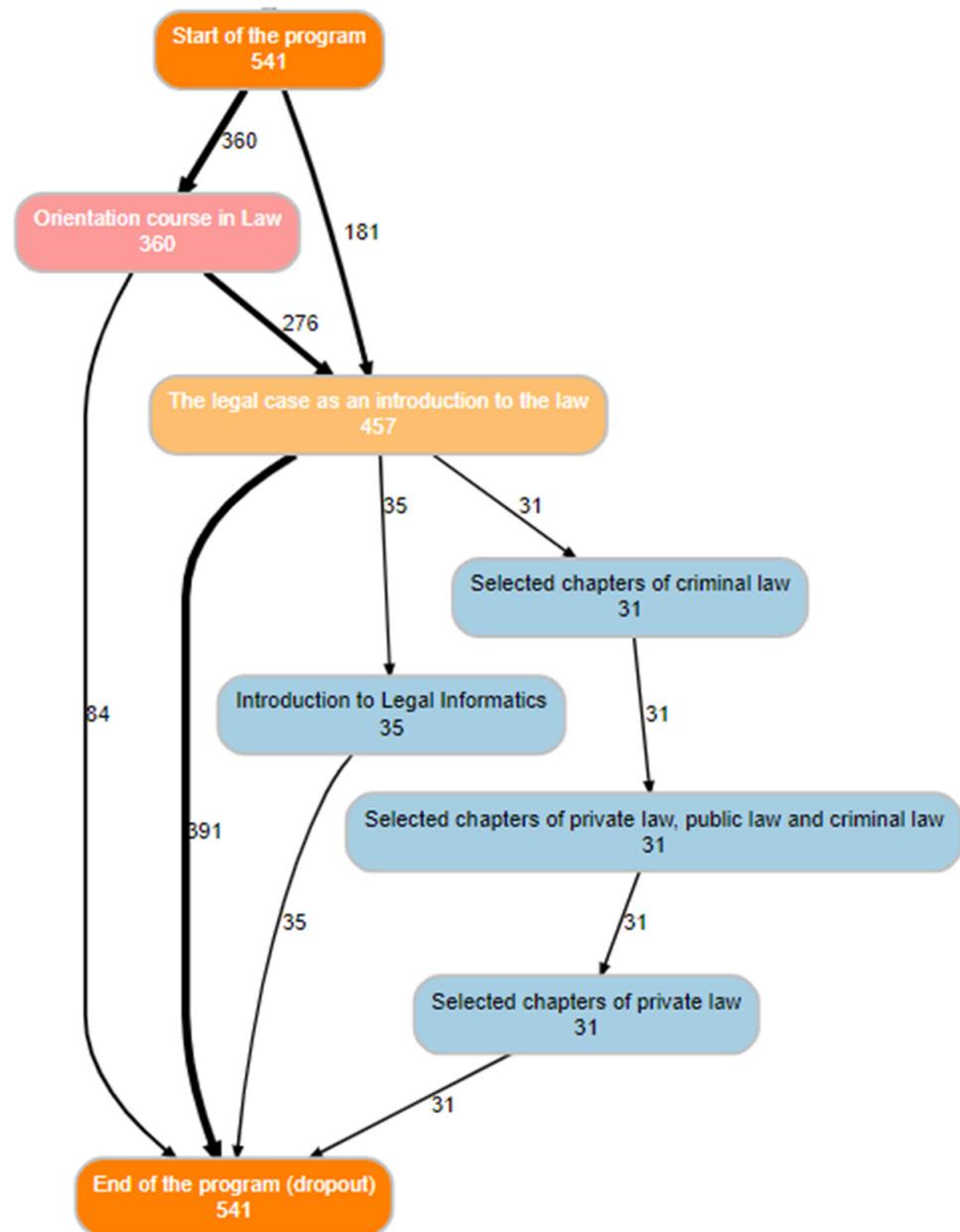


Figure 2. Depicting a process map at standard settings for dropouts, using the top-5 processes and including processes with 5 courses or less.

Comparing dropouts and graduates can be important in curriculum design. For instance, in both Figures 2 and 3, it becomes evident that graduates exclusively started with “Orientation course in Law”, while dropouts have a second branch, taking “The legal case as an introduction to law” first. The reasons need to be brought up in curriculum design meetings and discussed with the curriculum managers.

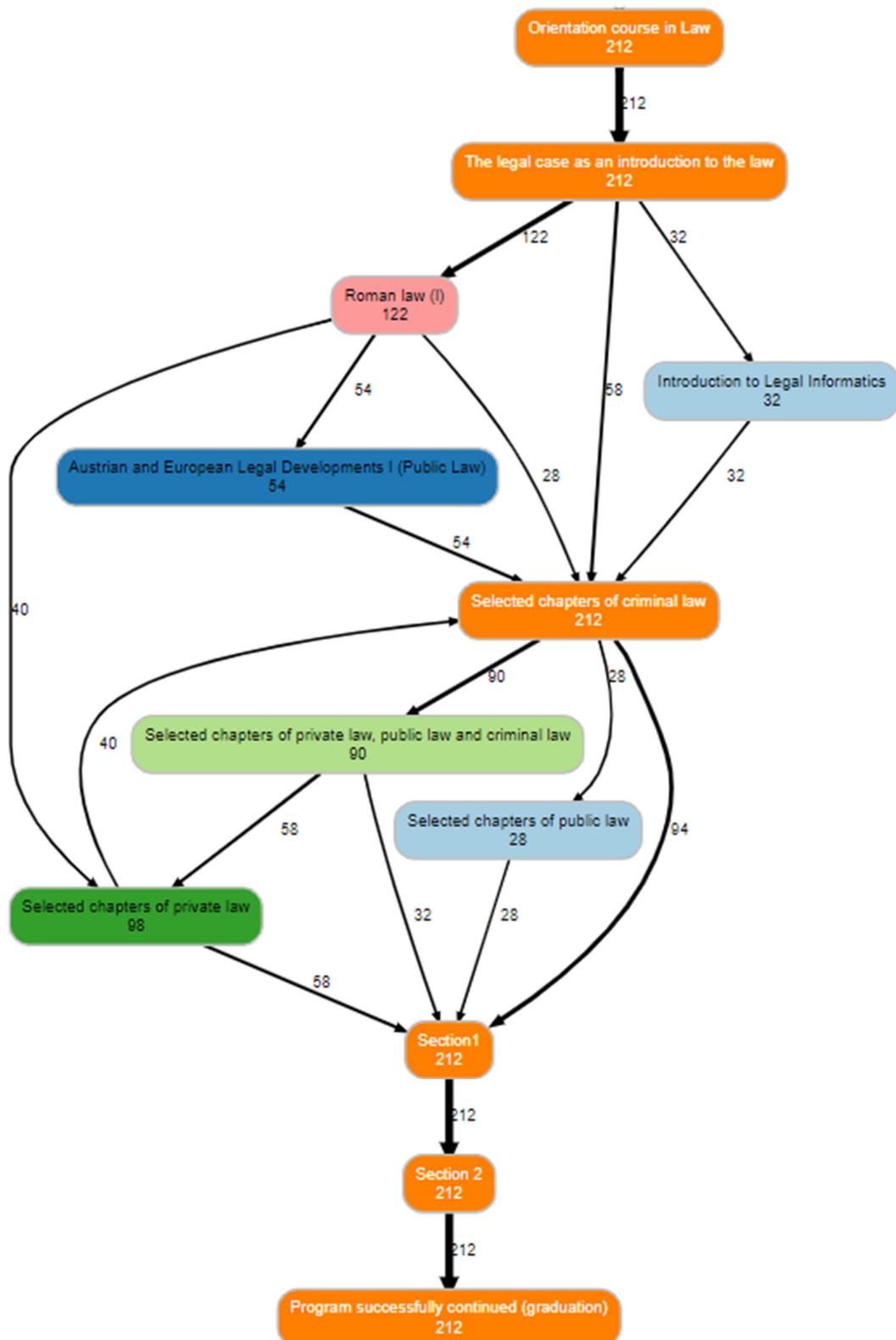


Figure 3. Depicting a process map at standard settings for graduates, using the top-5 processes and including processes with 5 courses or less.

3.1. Standard Settings

Standard settings for process generation are pre-defined at the top-5 processes and five courses. This setting produces animated maps for a chosen curriculum and includes

the five most frequent pathways showing absolute frequencies. Figure 2 shows a static example of the animated process map for dropouts and Figure 3 shows a map with the same settings for graduates.

3.2. Increased Top-n Processes

Increasing the number of processes to the top-20 processes increases the size of the process map. The number of exams is kept at five. Figure 4 shows a static example of the animated process map for dropouts.

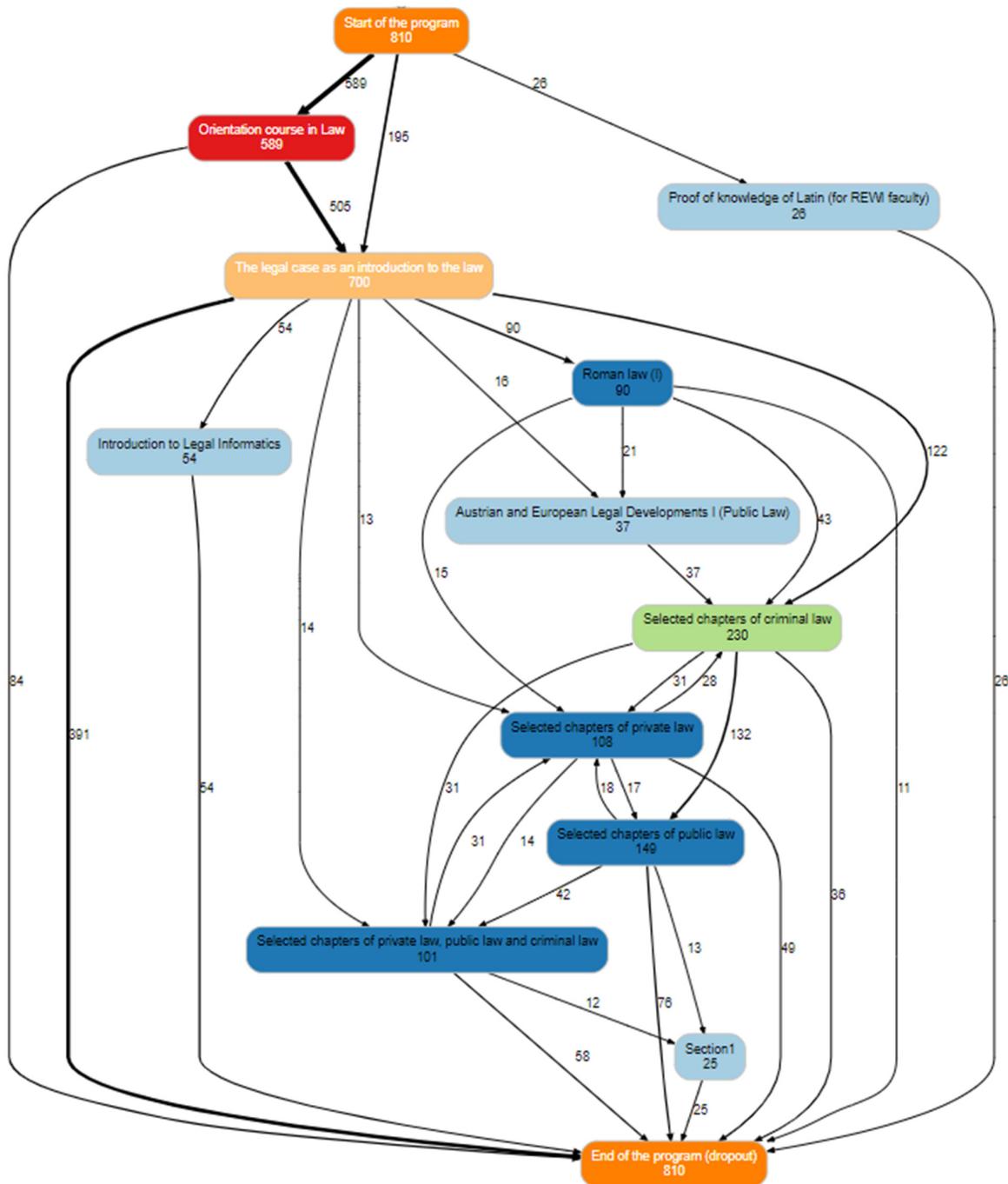


Figure 4. Depicting a process map with adjusted settings for dropouts, using the top-20 processes and including processes with 5 courses or less.

3.3. An Increased Number of Courses

Slightly decreasing the number of top-n processes to 20 (for reasons of clarity and comprehensibility) and increasing the number of exams within the processes to six yields a larger process map. This increases the number of processes depicted in the map as well as the number of dropouts being shown. Numbers were switched to relative frequencies. Figure 5 shows a static example of the animated process map for dropouts.

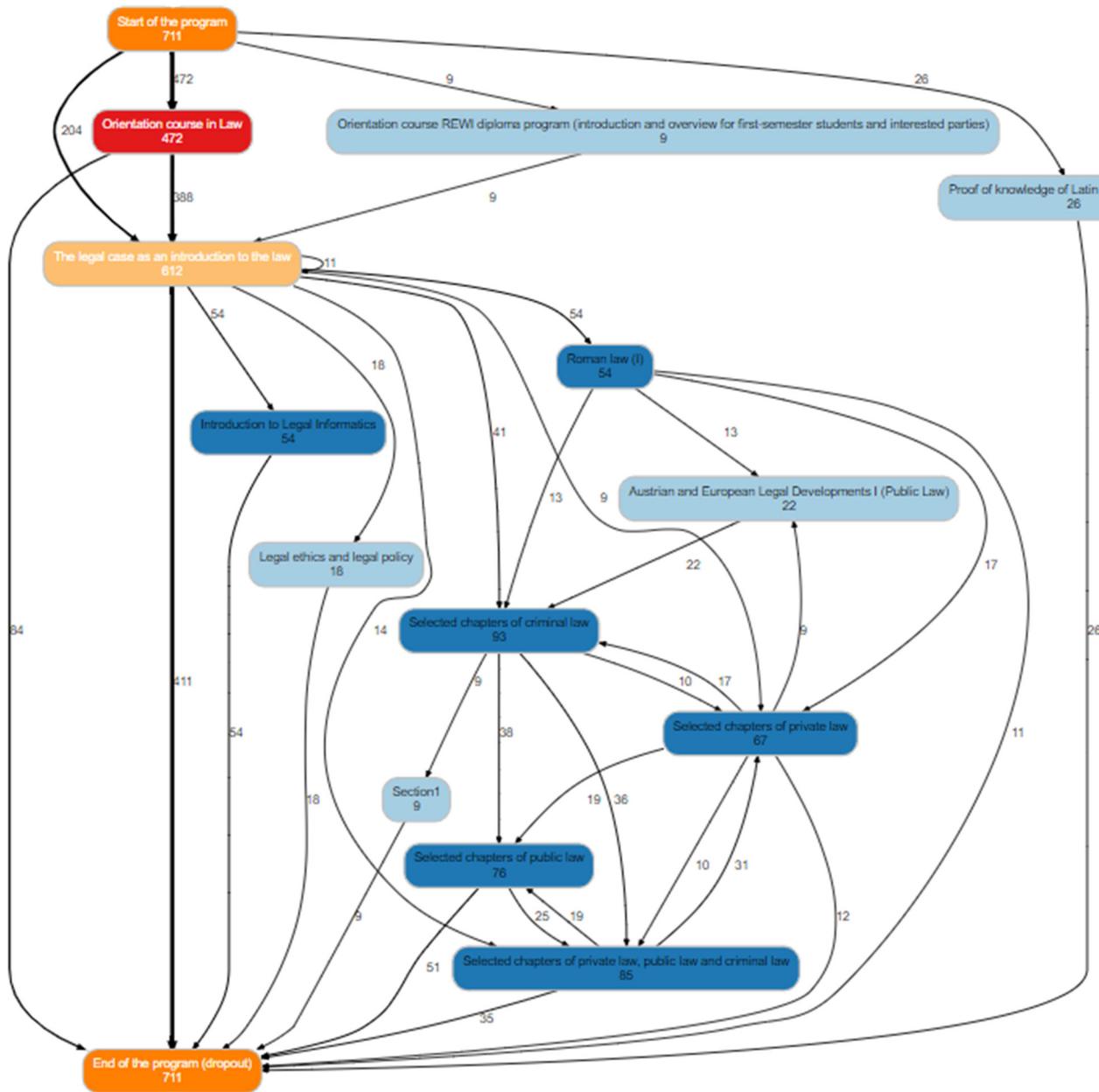


Figure 5. Depicting a process map with adjusted settings for dropouts, using the top-20 processes and including processes with 6 exams or less and round edges.

4. Discussion

4.1. Usability and Implementation

This study substantially improved upon a method for the educational process mining and visualization of students' course sequences [37]. A way of automating interactive and animated process maps was introduced. The maps could be set up on the scale of a university, focusing on the program level and being able to show results for each individual degree program or on aggregated stages. They are based on the most frequent pathways that students take within study programs before they drop out of their studies or on how successful students start their studies. It needs to be considered that this method is useful for visualizing dropout processes and that comparison with successful students improves its usability. Since process maps give no information on the real reasons why students ended a certain study program, only assumptions can be made. However, adding the visualizations of processes for graduates offers the possibility to identify ideal and non-ideal exam sequences and to gain more insight into possible reasons for leaving a program. Some maps showed that dropouts neglected or postponed certain elective courses (sequence comparison), concentrated on the "wrong" courses at the wrong time (sequence comparison), course requirements hindered them in their progress (sequence and time comparison) or their expectations of the discipline of study could have been wrong (as indicated by dropping out after the first introductory exam).

One way of implementing this method and using it as a managerial tool is stakeholder integration. People with knowledge about the reasons why certain process maps are shaped in a specific way are the addressed target group. At the University of Graz, the project "study forums" has been implemented as a way to communicate the results of this process mining technique with the goal to increase student retention via data-driven curriculum improvements. These are meetings between the staff of the administrative departments concerned with quality management and services for students with the curriculum designers, academic and teaching staff of each field of study and the student representatives at the university. In these forums, the interactive web application is used to present and explore the process maps. Curriculum planners also receive access to the app, for further data exploration. Discussions on the results are a central part of the study forums, leading to adaptations and new developments in both curricula and services for students. At the University of Graz, bringing together these stakeholder groups with curriculum designers is considered an important management process to reduce dropout. The process maps visualize common paths of dropouts and graduates and highlight possible sticking points for students but cannot give insights into the reasons why the paths show up the way they do. In conjunction with other analyses provided by the quality management staff, student representatives as well as the academic and teaching staff may be able to identify and elaborate on possible problems and improvement potentials, which can then be incorporated by curriculum designers.

Since the implementation of this method, 28 forums have been carried out with ten follow-up meetings, leading to evidence-based changes in more than 40 curricula up until the beginning of the academic year 2023/24. Prior to the development of the study forums and this process mining approach, no such method was used for curriculum design, relying on evidence from general key figures (e.g., the number of enrollments in a program over time) and anecdotal evidence from faculty staff and students only.

From an international standpoint, the method is transferrable to university systems different from Austria. The technical aspect of depicting course-taking sequences of dropouts and graduates does not change in other systems. What changes are the results and their interpretation. In this respect, it needs to be noted that the stricter curriculums and rules for course taking are designed, the less variation will be found in the process maps. The benefit of this process mining method should be higher, the more freedom students have in their exam-taking sequences. Additional management measures such as stakeholder integration and discussion forums may also be a good approach in international settings to better understand the results of the process maps and to transfer them into curriculum changes.

4.2. Challenges and Limitations

During the development of this educational process mining method, several challenges and limitations could be identified: dropout sequences can be included as a whole, i.e., from the first to the last exam before ending, but graduate pathways need to be limited by a cutoff defined by users to be comparable to each other. As graduate processes must include all exams in a curriculum and dropouts' pathways can have any given length, a definable cutoff was implemented. The maximum cutoff value was set to ten exams. Most curricula allow students to choose the sequence of exams freely, which means that the number of possible combinations of exams resulting in the displayed processes is rather large, even when only considering the pool of first-year courses. The process maps are not limited in how often an exam is taken since a node is created on the first try and traces lead back to it on repeat. This defines the number of possible exams in a process as the number of unique course IDs of a given student. In some curricula, the freedom of sequencing leads to a high rate of individuality, decreasing the likelihood of the occurrence of the exact same processes. Given the maximum number of exams displayed per student is defined by the unique names of the courses, the number of possible combinations C can be calculated via binominal coefficients [45]:

$$C(n;k) = \binom{n}{k} = \frac{n!}{k! \times (n-k)!} \quad k \leq n$$

For instance, displaying a plot with a potential pool of $n = 10$ first-year exams and a process with $k = 5$ exams in each path, per student 252 different pathways are possible. The size of a process map will increase further, considering the maximum number of the top 50 processes within a given curriculum and the higher number of exams within a real curriculum compared to this example. The bigger a process map gets, the higher the number of single-case processes due to the uniqueness of each student's sequence. Additionally, the R package used for animation has a limited capacity in how large maps can get. Above a certain point of complexity, animated process maps cannot be generated any longer [43]. In such a case, only static versions can be produced, leaving out the time dimension [41,42]. Including artificial and adjustable limits made visualizations clearer and narrowed processes down to a point, where they are still similar to each other by focusing on the first ten unique exams. This necessary step suggests that certain educational process mining methods performed on the program level are likely to perform well only within predefined boundaries. In this case, the limits for the method at hand are the beginning phase of study programs, useful for gaining insight into the course sequences of early dropouts. Due to the diversity of exams in later dropouts, applicability cannot be recommended without adjustments.

Implementing failed exams and passed exams as another variable can add information to the proposed approach. However, at the current state of the implemented R packages, no distinctive visual highlights can be defined to visualize these differences (e.g., by color coding). It is possible to include a switch that either only displays successfully passed and failed exams or to include the outcome in the name of the courses, e.g., "Orientation course in Law (passed)" and "Orientation course in Law (failed)". By doing so, the process map layouts would change. On the one hand, by implementing a switch only failed sequences could be displayed, leaving out the rest of the (successful) pathway a student takes. On the other hand, including exam outcomes in the course names would double the size of the process maps, which is not optimal, considering the size increases that already happen when adjusting the settings to display all of the pathways of students with five courses to six or seven courses. At the University of Graz, other tools and analyses have been created that can be used in conjunction with the process maps, allowing for a better presentation of individual course metrics.

Using exam data as the only reference in this process mining approach does not give concrete information about why students became dropouts, only when and where this happened. Qualitative data or other variables are not involved. Previous research shows

that leaving university or a study program is a complex and multi-faceted process [46–49]. Several factors, not only related to curriculum design, can accumulate over time. This creates different problems and increases the likelihood of dropping out until a student decides to leave. Demographic variables, academic readiness and financial aspects and other variables can lead to this outcome [50]. Besides academic dimensions, personal, economic, social and institutional aspects need to be taken into consideration [5]. This means that interpretations of the process maps always need to be handled with care, also incorporating the influence of outside factors. Assuming that a curriculum is not always the only reason why students end a program, this educational process mining method should be used in accompaniment with other evaluation methods. They can give clues about potential problem segments of curricula but cannot provide definitive answers [51,52].

Another issue encompasses the data not present in the process maps: students are not only influenced in their decisions by exams they take, but also by exams they do not take. Since there is no information on partial tests, instances of skipping specific exams and courses due to their difficulty levels can hardly be identified as they will not be visible. Due to procrastination being associated with students' mental health dimensions such as test anxiety [53], underlying factors that lead to burnout cannot be spotted in the outcomes [54]. The data mining approach at hand cannot account for the courses students are enrolled in, but never take exams or receive a grade for since exam outcome was defined as the major variable for generating event logs.

4.3. Future Advancements

Studies building upon the methodology of this paper should have a closer look at the suitability of exam outcomes as a major event log variable. Differentiating between courses with enrollment but no exam outcomes and exams with results may improve the insights into students' dropout behavior from a program perspective. This will increase the size of the process maps and compromise both the visibility and clarity of the depicted process. Therefore, a dynamic function for the clustering of course bundles can be implemented to display more courses within one node. By creating chunks of courses that belong together, the length of the processes can be improved for both dropouts and graduates. At the University of Graz, curricula are structured in modules, providing overall structures for different subject areas. However, curricula do not follow a strict module sequence, which means that the recommended course sequence is only loosely related to the numbering of the modules. A function is needed for creating unequally sized chunks of courses with control over the sizing parameters. Approaches such as unsupervised classification algorithms may offer benefits in this regard [55].

5. Conclusions

This study demonstrated the implementation of an automated educational process mining approach focused on visualizing students' pathways through a study program dependent on their outcome status as a "dropout" or "graduate". The outcome process maps provide information on the exam sequence of students and can be used in curriculum (re)design as the comparison of successful and non-successful students' maps allows for drawing certain conclusions. Despite their benefits, they are limited in their size and the number of exams that can be displayed, making them a good fit for early dropout evaluation. In order to increase the size and usability of the maps, dynamic clustering approaches may be a solution, using chunks of process parts that belong together in this process mining method.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/higheredu3010004/s1>, Figures S1–S4: Animated HTML versions of the figures of this study.

Funding: This research was funded by the University of Graz.

Institutional Review Board Statement: The analysis of the data and the permission for publication have been approved in a works agreement between the Performance- and Quality Management Department and the works committee of the University of Graz, called “Rahmen-BV IKT 2019”, which includes ethics approval. Furthermore, data analyses on student data by staff departments of the university is regulated by law, defining analyses as a means of increasing graduation rates of the Austrian universities as an obligation (Federal Ministry of Education, Science and Research, 2002; see references in manuscript).

Informed Consent Statement: Not applicable.

Data Availability Statement: Data cannot be made available due to being part of the internal university database.

Acknowledgments: The author acknowledges the financial support by the University of Graz (Open Access Funding).

Conflicts of Interest: The author declares no conflicts of interest.

References

1. OECD. Education at a Glance 2019. Available online: https://www.oecd-ilibrary.org/education/education-at-a-glance-2019/summary/spanish_f6dc8198-es (accessed on 14 December 2023).
2. Marsh, L.M. College dropouts—A review. *Pers. Guid. J.* **1966**, *44*, 475–481. [CrossRef]
3. Federal Ministry of Education, Science and Research, Federal Act on the Capacity Orientated, Student-Centered Financing of Universities (Universities’ Financing Act—UniFinV). 2018. Available online: <https://eurydice.eacea.ec.europa.eu/national-education-systems/austria/legislation-and-official-policy-documents> (accessed on 14 December 2023).
4. Federal Ministry of Education, Science and Research, Federal Act on the Organisation of Universities and their Studies (Universities Act 2002—UG). 2002. Available online: <https://eurydice.eacea.ec.europa.eu/national-education-systems/austria/legislation-and-official-policy-documents> (accessed on 14 December 2023).
5. Alban, M.; Mauricio, D. Predicting university dropout through data mining: A systematic literature. *Indian J. Sci. Technol.* **2019**, *12*, 1–12. [CrossRef]
6. Solomon, D.; Patil, S.; Agrawal, P. Predicting performance and potential difficulties of university student using classification: Survey paper. *Int. J. Pure Appl. Math.* **2018**, *118*, 2703–2707.
7. Márquez-Vera, C.; Cano, A.; Romero, C.; Noaman, A.Y.M.; Mousa Fardoun, H.; Ventura, S. Early dropout prediction using data mining: A case study with high school students. *Expert Syst.* **2016**, *33*, 107–124. [CrossRef]
8. Suhlmann, M.; Sassenberg, K.; Nagengast, B.; Trautwein, U. Belonging mediates effects of student-university fit on well-being, motivation, and dropout intention. *Soc. Psychol.* **2018**, *49*, 16–28. [CrossRef]
9. Zajac, T.Z.; Komendant-Brodowska, A. Premeditated, dismissed and disenchanting: Higher education dropouts in Poland. *Tert. Educ. Manag.* **2019**, *25*, 1–16. [CrossRef]
10. Ferguson, R. Learning analytics: Drivers, developments and challenges. *Technol. Enhanc. Learn.* **2012**, *4*, 304–317. [CrossRef]
11. Campbell, J.P.; DeBlois, P.B.; Oblinger, D.G. Academic analytics: A new tool for a new era. *Educ. Rev.* **2007**, *42*, 40.
12. Siemens, G.; Long, P. Penetrating the fog: Analytics in learning and education. *Educ. Rev.* **2011**, *46*, 30–32.
13. Enarson, C.; Cariaga-Lo, L. Influence of curriculum type on student performance in the United States Medical Licensing Examination Step 1 and Step 2 exams: Problem-based learning vs. lecture-based curriculum. *Med. Educ.* **2001**, *35*, 1050–1055. [CrossRef]
14. Mendez, G.; Ochoa, X.; Chiluiza, K.; De Wever, B. Curricular design analysis: A data-driven perspective. *J. Learn. Anal.* **2014**, *1*, 84–119. [CrossRef]
15. dos Santos Garcia, C.; Meinheim, A.; Junior, E.R.F.; Dallagassa, M.R.; Sato, D.M.V.; Carvalho, D.R.; Santos, E.A.P.; Scalabrin, E.E. Process mining techniques and applications—A systematic mapping study. *Expert Syst. Appl.* **2019**, *133*, 260–295. [CrossRef]
16. Ameen, A.O.; Alarape, M.A.; Adewole, K.S. Students’ academic performance and dropout predictions: A review. *Malays. J. Comput.* **2019**, *4*, 278–303. [CrossRef]
17. Moretti, A.; Gonzalez-Brenes, J.; McKnight, K. Data-driven curriculum design: Mining the web to make better teaching decisions. In Proceedings of the 7th International Conference on Educational Data Mining, London, UK, 4–7 July 2014.
18. Ornstein, A.C.; Hunkins, F.P. *Curriculum Foundations, Principles and Issues*, 5th ed.; Allyn & Bacon: Boston, MA, USA, 2009.
19. DaRosa, D.A.; Bell, R.H. Graduate surgical education redesign: Reflections on curriculum theory and practice. *Surgery* **2004**, *136*, 974–996. [CrossRef] [PubMed]
20. Neary, M. Curriculum concepts and research. In *Curriculum Studies in Post-Compulsory and Adult Education: A Teacher’s and Student Teacher’s Study Guide*; Nelson Thornes Ltd.: Cheltenham, UK, 2003; pp. 33–56.
21. González, J.; Wagenaar, R.; Beneitone, P. Tuning—América Latina: Un proyecto de las universidades [Tuning—Latin America: A project of the universities]. *Rev. Iberoam. Educ.* **2004**, *35*, 151–164.

22. Pukkila, P.J.; DeCosmo, J.; Swick, D.C.; Arnold, M.S. How to engage in collaborative curriculum design to foster undergraduate inquiry and research in all disciplines. In *How to Design, Implement, and Sustain a Research-Supportive Undergraduate Curriculum: A Compendium of Successful Curricular Practices for Faculty and Institutions Engaged in Undergraduate Research*; Council on Undergraduate Research: Washington, DC, USA, 2007; pp. 341–357.
23. Denton, J.W.; Franke, V.; Surendra, K.N. Curriculum and course design: A new approach using quality function deployment. *J. Educ. Bus.* **2005**, *81*, 111–117. [[CrossRef](#)]
24. Wolf, P. A model for facilitating curriculum development in higher education: A faculty-driven, data-informed, and educational developer supported approach. *New Dir. Teach. Learn.* **2007**, *112*, 15–20. [[CrossRef](#)]
25. Van den Akker, J.; De Boer, W.; Folmer, E.; Kuiper, W.; Letschert, J.; Nieveen, N.; Thijs, A. *Curriculum in Development*; Netherlands Institute for Curriculum Development (Slo): Enschede, The Netherlands, 2009.
26. Van der Aalst, W.; Damiani, E. Processes meet big data: Connecting data science with process science. *IEEE Trans. Serv. Comput.* **2015**, *8*, 810–819. [[CrossRef](#)]
27. Bogarín, A.; Cerezo, R.; Romero, C. A survey on educational process mining. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2018**, *8*, e1230. [[CrossRef](#)]
28. Romero, C.; Ventura, S. Educational data science in massive open online courses. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2017**, *7*, e1187. [[CrossRef](#)]
29. Weijters, A.J.M.M.; van Der Aalst, W.M.; De Medeiros, A.A. Process mining with the heuristics miner-algorithm. *Tech. Univ. Eindh. Tech Rep. WP* **2006**, *166*, 1–34.
30. Trcka, N.; Pechenizkiy, M.; van der Aalst, W. Process mining from educational data. In *Handbook of Educational Data Mining*; CRC Press: Boca Raton, FL, USA, 2010; pp. 123–142.
31. Ghazal, M.A.; Ibrahim, O.; Salama, M.A. Educational process mining: A systematic literature review. In Proceedings of the 2017 European Conference on Electrical Engineering and Computer Science (EECS), Bern, Switzerland, 17–19 November 2017.
32. Van Der Aalst, W. *Process Mining: Data Science in Action*; Springer: Berlin/Heidelberg, Germany, 2016.
33. Aulck, L.; Velagapudi, N.; Blumenstock, J.; West, J. Predicting student dropout in higher education. *arXiv* **2017**, arXiv:1606.06364.
34. Bogarín, A.; Cerezo, R.; Romero, C. Discovering learning processes using inductive miner: A case study with learning management systems (LMSs). *Psicothema* **2018**, *30*, 322–329. [[PubMed](#)]
35. Mukala, P.; Buijs, J.; Leemans, M.; Van der Aalst, W. Learning analytics on coursera event data: A process mining approach. *CEUR Workshop Proc.* **2015**, *1527*, 18–32.
36. Buck-Emden, R.; Dahmann, F.D. Analyse von Studienverläufen mit Process-Mining-Techniken. *HMD Prax. Wirtsch.* **2018**, *55*, 846–865. [[CrossRef](#)]
37. Buck-Emden, R.; Dahmann, F.D. *Zur Auswertung von Studienverläufen Mit Process-Mining-Techniken*; Technical Report 07-2017; Hochschule Bonn-Rhein-Sieg: Sankt Augustin, Germany, 2017.
38. Thaler, B.; Haag, N.; Binder, D.; Unger, M. *Studierenden-Monitoring (STUDMON), Begleitender Projektbericht, Version 2, 24.09. 2019*; HIS: Vienna, Austria, 2019.
39. R Core Team. *R: A Language an Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2022.
40. Ripley, B.; Lapsley, M. RODBC. 2022. Available online: <https://CRAN.R-project.org/package=shiny> (accessed on 14 December 2023).
41. Janssenswillen, G.; van Hulzen, G.; Mannhardt, F.; Martin, N.; Van Houdt, G. bupaR. 2023. Available online: <https://CRAN.R-project.org/package=shiny> (accessed on 14 December 2023).
42. Janssenswillen, G.; van Hulzen, G.; Depaire, B.; Mannhardt, F.; Beuving, T. processmapR. 2023. Available online: <https://CRAN.R-project.org/package=shiny> (accessed on 14 December 2023).
43. Mannhardt, F. processanimateR. 2023. Available online: <https://CRAN.R-project.org/package=shiny> (accessed on 14 December 2023).
44. Chang, W.; Cheng, J.; Allaire, J.; Sievert, C.; Schloerke, B.; Xie, Y.; Allen, J.; McPherson, J.; Dipert, A.; Borges, B. shiny. 2023. Available online: <https://CRAN.R-project.org/package=shiny> (accessed on 14 December 2023).
45. Mol, M. Rosetta Code. 2023. Available online: https://rosettacode.org/wiki/Evaluate_binomial_coefficients (accessed on 14 December 2023).
46. Heublein, U. Student drop-out from German higher education institutions. *Eur. J. Educ.* **2014**, *49*, 497–513. [[CrossRef](#)]
47. Ozga, J.; Sukhnandan, L. Undergraduate non-completion: Developing an explanatory model. *High. Educ. Q.* **1998**, *52*, 316–333. [[CrossRef](#)]
48. Wilcox, P.; Winn, S.; Fyvie-Gauld, M. ‘It was nothing to do with the university, it was just the people’: The role of social support in the first-year experience of higher education. *Stud. High. Educ.* **2005**, *30*, 707–722. [[CrossRef](#)]
49. Bardach, L.; Lüftenegger, M.; Oczlon, S.; Spiel, C.; Schober, B. Context-related problems and university students’ dropout intentions—The buffering effect of personal best goals. *Eur. J. Psychol. Educ.* **2020**, *35*, 477–493. [[CrossRef](#)]
50. Barbera, S.A.; Berkshire, S.D.; Boronat, C.B.; Kennedy, M.H. Review of undergraduate student retention and graduation since 2010: Patterns, predictions, and recommendations for 2020. *J. Coll. Stud. Retent. Res. Theory Pract.* **2020**, *22*, 227–250. [[CrossRef](#)]
51. Chan, Z.C.; Cheng, W.Y.; Fong, M.K.; Fung, Y.S.; Ki, Y.M.; Li, Y.L.; Wong, H.T.; Wong, L.T.; Tsoi, W.F. Curriculum design and attrition among undergraduate nursing students: A systematic review. *Nurse Educ. Today* **2019**, *74*, 41–53. [[CrossRef](#)]

52. Devadas, B. A Critical Review of Qualitative Research Methods in Evaluating Nursing Curriculum Models: Implication for Nursing Education in the Arab World. *J. Educ. Pract.* **2016**, *7*, 119–126.
53. Ariani, D.W.; Susilo, Y.S. Why do it later? Goal orientation, self-efficacy, test anxiety, on procrastination. *J. Educ. Cult. Psychol. Stud. (ECPS J.)* **2018**, *17*, 45–73. [[CrossRef](#)]
54. Krispenz, A.; Gort, C.; Schülke, L.; Dickhäuser, O. How to reduce test anxiety and academic procrastination through inquiry of cognitive appraisals: A pilot study investigating the role of academic self-efficacy. *Front. Psychol.* **2019**, *10*, 1917. [[CrossRef](#)]
55. Bey, A.; Champagnat, R. Analyzing Student Programming Paths using Clustering and Process Mining. In Proceedings of the CSEDU 2022—14th International Conference on Computer Supported Education, Virtual Event, 22–24 April 2022.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.