A New Learning Path Model for E-Learning Systems

David Brito Ramos, Federal Institute of Education, Science, and Technology of Amazonas, Manaus, Brazil

Ilmara Monteverde Martins Ramos, Federal Institute of Education, Science, and Technology of Amazonas, Manaus, Brazil

iD https://orcid.org/0000-0003-3916-8081

Isabela Gasparini, Santa Catarina State University, Florianópolis, Brazil

iD https://orcid.org/0000-0002-8094-9261

Elaine Harada Teixeira de Oliveira, Federal University of Amazonas, Manaus, Brazil

iD https://orcid.org/0000-0003-2884-9359

ABSTRACT

This work presents a new approach to the learning path model in e-learning systems. The model uses data from the database records from an e-learning system and uses graphs as representation. In this work, the authors show how the model can be used to represent visually the learning paths, behavior analysis, help to suggest group formation for collaborative activities, and thus assist the teacher in making decisions. To validate the practical utility of the model, the authors created two tools, one to visualize the learning paths and another to suggest groups of students for collaborative activities. Both tools were tested in a real environment, presenting useful results. The authors carried experiments with students from three programs: physics, electrical engineering, and computer science. Experiments show that it is possible to use the proposed learning path to analyze student behavior patterns and recommend group formation with positive results.

KEYWORDS

Grouping, Higher Education, Information Visualization, Learning Management Systems, Learning Trajectories, Moodle, Student Behavior, Student Model

INTRODUCTION

Analytical tools and models that allow understanding students' behavior can infer individual or collective patterns and improve students' experiences (Saito & Watanobe, 2020). These mentioned tools and models can also help teachers monitor the learners' actions in a Learning Management System (LMS) (Baneres et al., 2019; Weiand et al., 2019).

In this context, new information generated from LMS data can facilitate the teaching and learning process. LMS collects data about users that can help define the learner's profile, learner's behavior, and identify their difficulties and needs. One way of accompanying learners is to observe the actions they perform on the system, and these actions can result in paths known as Learning Paths (LP).

The literature works related to LMS consider the LP from two points of view (Ramos et al., 2015):

1. The preferential sequence of access to content (such as educational resources and activities) defined by teachers when planning the course. They made available this sequence in the LMS;
2. The path covered by students during their interactions with the content provided in the LMS.

This work presents a new model of LP based on graphs to assist teachers in their activities. In this work, the term LP refers to the trajectories that students go through. The LP data are obtained based on the student interaction actions with the LMS. The use of stored data in the LMS is already, in many cases, enough to help in decision making, as shown in the experiments. The idea is to propose an alternative for the current user model, which represents the student's domain knowledge state (Huang et al., 2016). So, this paper's main contribution is to demonstrate that data that does not reflect, at least directly, the student's knowledge level, can also be used to create a model of learning paths with the ability to be applied in several different learning environments and to serve as a basis for the creation of new tools.

Monitoring students manually is hard work for teachers, especially when they need to deal with large classes. Thus, based on the proposed model, two tools were created to support the teacher, one to monitor the student's behavior and the other to assist in the groups formation.

Another contribution of this work is to present a more straightforward approach because, in most works, specialists must analyze and classify resources/activities. They also need to indicate attributes such as level of difficulty, expected time of execution, required knowledge, what students learned in that content, among others, and those actions are not necessary for the proposed approach.

The next sections will present the LP model proposed, a review of the literature, materials and methods, the tools and the experiments carried out with them, and, finally, the discussion of the results and conclusions.

LOG-BASED LEARNING PATH MODEL

This section shows the used data source and how the proposed LP model was designed based on that data.

Data for Learning Paths

Our LP model uses the student's actions history from the database records, requiring no more direct data collection, such as filling in forms, surveys, or questionnaires. Most LMSs use resources and activities. Resources are content, in general, static, also called teaching materials. The activities are contents that require the more active participation of the learners, allowing the interaction between them or being a tool of evaluation. Tasks, chats, forums, and quizzes are some examples of activities.

Our approach constructs the paths using cross-data of several tables. The central point is the logs that indicate when and what resources and activities the students accessed. It is possible to chronologically order events that occurred during the student's interaction with the LMS.

Once ordered, logs present the student's browsing sequence by the resources and activities of the LMS. From the observation of this structure and previous studies (Ramos et al., 2015; Ramos & Oliveira, 2015), it was concluded that it was possible to use the graph structure to represent the LP.

Learning Path Model

The proposed LP model uses the directed graph as the representation form. The vertices represent the resources and activities provided on the course, and the navigation (path) of the student between the vertices form the edges. The study obtains the student's path utilizing the logs' time-ordered sequence. Each record indicates a resource accessed, for example, if the records indicate that resources

A, B, C, and D have been accessed in that order, then the edges of the path are AB, BC, and CD. The following are more details about how this work uses the graph representation.

For each resource or activity created, the LMS saves a new record named instance in the database. Also, LMS stores the order in which the instances must appear. This order corresponds to the path suggested by the teacher, the first instance is the starting point, and the last is the end of the proposed learning sequence. The model also uses data from the sections where teachers group resources and activities. If the teacher later reorders the resources and activities provided, the model automatically adapts to the new configuration. The data presented here were selected because they allow the visual representation of learning paths as well as can be used with Machine Learning and Data Mining tools, which is part of another study being conducted.

The model has assigned vertices that store the following information:

- pos: The position of the vertex in the sequence defined for resources and activities.
- cmid: A unique identifier of each instance, this information maintains the relation between vertex and instance of the LMS.
- module: The resource/activity type such as 'forum' or 'assign'.
- module_name: Module name in the translated language, preceded by the value of the pos attribute.
- name: The title of the resource/activity established by the teacher for the instance.
- section: The topic number which the vertex belongs to.
- section_name: The title of the topic. The teacher can define it. Otherwise, the name "Topic" plus the section attribute will be assigned.
- value: Number of interactions (V), indicates the total interactions (visualization, file sending, posting, among other actions). The visual representation uses the value as vertex weight.

The edges contain the following information:

- source: The identifier of the starting instance of the student's interaction. This attribute references the pos attribute of a vertex.
- target: The identifier of the arrival instance of student's interaction, also indicates the pos value of the accessed vertex. Thus, it is possible to identify the origin and destination of the edge, which makes the graph directed.
- value: The number of times the learner goes from source vertex to the target vertex, it indicates the edge weight.
- type: Informs the edge classification. The model has defined three possible taxonomies: return, standard, and advance; these types will be detailed later.
- is_max_value: The boolean attribute that indicates the most used edge among all that originates from source vertex.
- is_min_value: The boolean attribute that indicates the least used of all the edges that start from source vertex.
- is_max_value_advance: Boolean attribute, if the edge is of the type 'advance', it indicates if it, of all the edges of the type 'advance' starting from the source vertex, is the one that has the highest value.
- is_max_value_return: Boolean attribute, if the edge is of the type 'return', it indicates if to it of all the edges of the type 'return' starting from the vertex of origin, is the one that has the higher value.

As stated in the model, the edges are classified into three types: standard, advance, and return, described below. Given the edge (v_i, v_j), where i and j are the positions of the vertex (pos value), the model has:

- standard edge: It represents navigation from a resource/activity to the immediate successor, that is, given a vertex v_i, and a vertex v_j, the edge that part from v_i, and points to v_j is 'standard' type if $j=i+1$.
- advance edge: It indicates navigation from one resource/activity to the other later than the immediate successor, that is, given a vertex v_i, and a vertex v_j, the edge that starts from v_i, and points to v_j is 'advance' type if $j>i+1$.
- return edge: It indicates navigation of a resource/activity to a previous one, in this case, given a vertex v_i, and a vertex v_j, the edge that starts from v_i, and points to v_j is 'return' type if $j<i$.

When students follow the trajectory defined by the teacher, they walk from vertex v_i to v_j where always $j=i+1$. Given this, there can only be at most one standard edge for each vertex v_i. When the vertex represents the last instance, it has no standard edge and no advance edge, and when the vertex represents the first instance, it has no return edge. The proposed model is dynamic, whenever a student starts from vertex v_i to v_j, the algorithm adds 1 to the value attribute of the edge (v_i,v_j).

With information about the edges, the model can calculate the proportion of the amount of each type of edge to the total of edges, as well as the dispersion measure, which indicates how dispersed the student's navigation is.

The calculation of edge ratio (prop) is given by the ratio of the number of edges of a given type x by the total of all edges n, according to, where $x \in \{advance, standard, return\}$ and $n>0$:

$$prop_x = \frac{edges_x}{n}$$

The average of the edges' length gives the dispersion measure (*disp*). The length of an edge e is the difference between origin and destination vertices positions and is given by $length_e=|i-j|$, where i and j are the position of the vertices bounded by the edge (v_i,v_j), and n is the total number of edges of the graph. The dispersion measure can be calculated by considering all edges of the graph or by type of edge. Equation represents the dispersion calculation, considering all the edges of the graph, whereas represents the calculation of the dispersion by edge type. In $x \in \{advance, standard, return\}$ and $edges_x > 0$:

$$disp = \frac{\sum_{e=1}^{n} length_e}{n}$$

$$disp_x = \frac{\sum_{e=1}^{edges_x} length_e}{edges_x}$$

From the model information, some metrics were created that could be used in analysis with data mining and machine learning tools, as well as statistical analysis. Thus, the following metrics were created:

- total dispersion: it is the measure of dispersion given by (2), that is, the calculation considers all edges of the LP.
- advance dispersion: dispersion measure given by (3), considering for the calculation only the advance edges.
- return dispersion: dispersion measure given by (3), considering for the calculation only the return edges.

**Figure 1. A representation of the Learning Path Model**
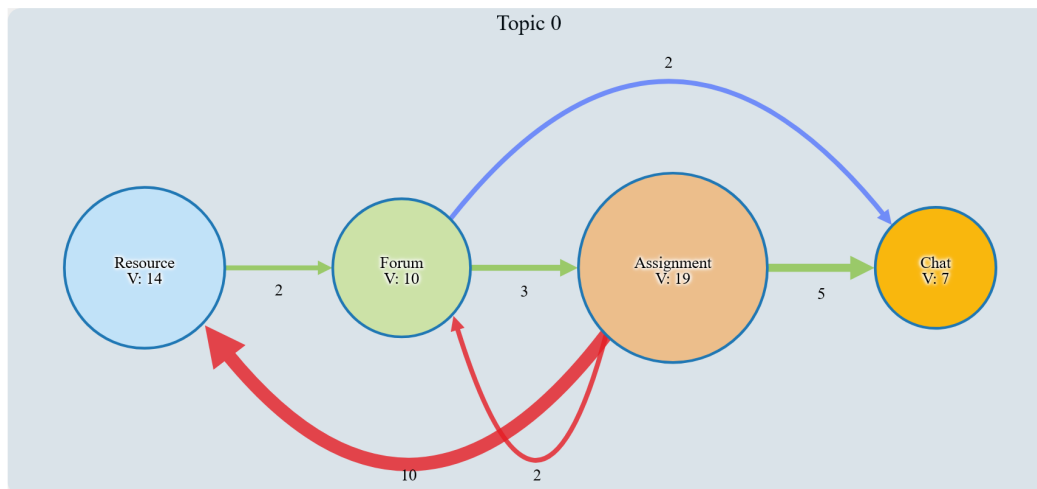
- standard dispersion: it will always have a value of 1 if the number of edges is greater than 0, or 0 if there are no edges in the LP.
- advance standard deviation: the standard deviation of advance edges dispersion.
- advance variance: the variance of advance edge dispersion.
- average access to vertices: average access/interaction with all resources and activities available in the learning environment.
- return proportion: the ratio between the number of edges of the return type and the total number of edges of the graph. This value was defined as a percentage. Therefore, proportion metrics always vary in value from 0 to 100.
- return standard deviation: the standard deviation of the return edges dispersion.
- return variance: the variance of the return edges dispersion.
- advance proportion: the ratio between the number of advance edges and the total number of graph edges.
- standard proportion: the ratio between the number of standard edges and the total number of graph edges.

Figure 1 shows the visual representation of the LP model proposed in this work. Each vertex has a color that represents a type of resource/activity, and the diameter is proportional to the number of students' interactions (V), so it is possible to observe more clearly the access number to each instance. The values in the edges are from value attribute, they specify the edge thickness, and the arrow indicates the direction of the path. The green edges are the 'standard' edges and indicate the teacher's LP. The blue edge is the 'advance' edge. The red edges represent the 'return' edges. In the graph, it is possible to verify that the student is possibly finding it difficult to answer the assignment (third vertex).

As the model usages the data stored by the LMS, it is possible to follow the evolution of the student's LP over time. The model can be used to analyze data from a finished class or during classes because the model is flexible and allows the representation of data that can be delimited by a time interval.

For the feasibility study of this research, Moodle was selected since it is one of the LMSs most used in several educational institutions (Cerezo et al., 2014). The model is also being applied to an online judge, which will be presented in a future work.

Figure 2. Work architecture
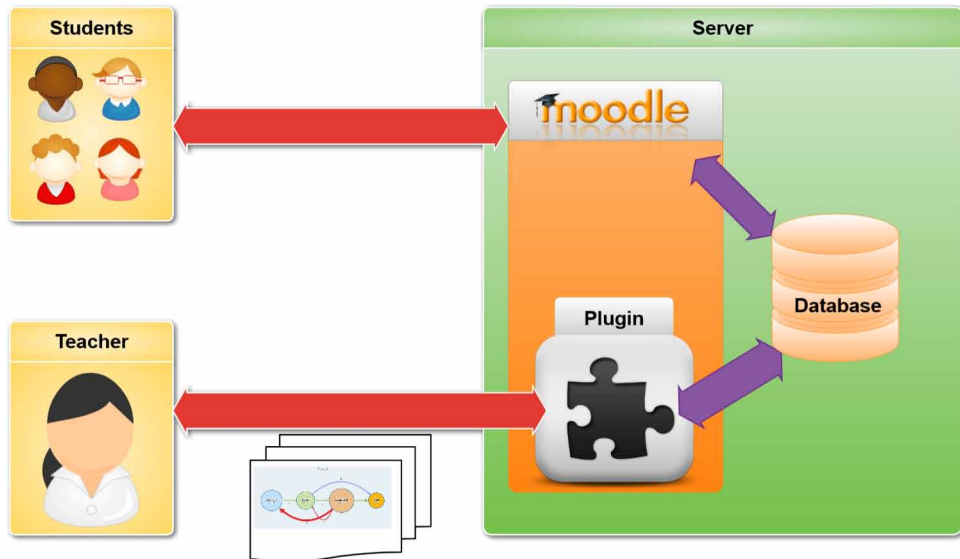
It was considered the implementation of plugins for Moodle so that the teacher will benefit from the tools developed directly in the LMS. Figure 2 represents the work architecture used. Students interact with the LMS, which records actions. In sequence, a plugin extracts the LPs from the data of the database. Finally, plugins use the LPs and present the result to the teacher. In this work, two plugins are presented, the first shows the LPs visually, and the second recommends groups for collaborative activities.

Given this, this paper has the following research question: How can Learning Paths help to improve the teaching-learning process in LMS?

This work deal with the application of the model in two activities: behavior analysis and group formation. Thus, this study has two questions derived from the main one:

1. How can Learning Paths help in student behavior analysis?
2. How can Learning Paths help in the group formation for collaborative activities?

LITERATURE REVIEW

A Systematic Review of Literature (SRL) was conducted to investigate how the LMSs organize and use the LP (Ramos & Oliveira, 2015). The findings are summarized as follows.

Visual Representation of Learning Paths

The visualization of the LP is a way for teachers to have a simplified overview of their class (Adesina & Molloy, 2011; Cerezo et al., 2014; Schröck et al., 2010; Teutsch & Bourdet, 2010). Some works (Marquez et al., 2008; Pires & Cota, 2010; Saito & Watanobe, 2020; Sitthisak et al., 2013; Vantroys & Peter, 2003; Wisuttikul & Boonmee, 2004) show that LP information helps in LMS environment, like choose dynamically next lesson for learners. Thus, in analyzing the LP, the teacher can monitor the students' behavior, their contributions in the LMS (Teutsch & Bourdet, 2010), learning process (Cerezo et al., 2014), or trying to predict students' performance (Weiand et al., 2019).

Table 1. Comparison between works on the visualization of learning paths

Reference	Main Types of Representation	LMS Independent	Multilevel View	Used With Moodle	Available for Teacher	Available to the Student	Path Viewpoint
(Schröck et al., 2010)	Cartography	-	Yes	No	Yes	Yes	Teacher and Students
(Teutsch & Bourdet, 2010)	Gantt Chart	-	No	No	Yes	No	Students
(Adesina & Molloy, 2011)	BPMN	No	No	No	Yes	No	Teacher and Students
(Cerezo et al., 2014)	Graph	Yes	No	Yes	Yes	No	Students
(Li et al., 2019)	Concept Map	-	No	No	Yes	-	Teacher
(Weiand et al., 2019)	Bar chart, table lens	Yes	No	Yes	Yes	No	Students
(Saito & Watanobe, 2020)	Sequence	No	No	No	No	No	Students
This work proposal	**Graph**	**Yes**	**Yes**	**Yes**	**Yes**	**No**	**Teacher and Students**

It is possible to verify, through comparative Table 1, the relation between this work proposal with the strongly related works. Columns with trace mean it could not possibly identify the information in publications.

Two works represent the teacher and student paths at the same graph. However, Adesina and Molloy (2011) can show only one student LP per time, whereas Schröck et al. (2010) had manually created the map. In the study approach, the teacher also can observe the entire class or groups, and the tool automatically created the visual representation. Graphs structuring the LP data can generate possibilities for the creation of new tools, such as eGraph (Cerezo et al., 2014), LPGraph (Ramos et al., 2016), M-Cluster (Ramos et al., 2017), the application of CbKST in Moodle (Sitthisak et al., 2013) or the selection of the shortest LP (Nurjanah & Fiqri, 2017). As this work will show group formation as one of the applications of the LP model proposed, next, some works about groups in LMS are presented.

Groups in Learning Management Systems

The following works present an overview of the different forms of student grouping in LMS. Jagadish (2014) used the KNN (K-Nearest Neighbor) to perform grouping of students in Moodle. Abnar et al. (2012) proposed applied genetic algorithms and Likert scale, in their approach the teacher can choose a set of different attributes and classify them based on their impact on the groups' formation. Yathongchai et al. (2013) used the decision tree, data mining, and Hartigan indexing techniques to form student groups in Moodle based on their grades and behaviors. Montazer and Rezaei (2012) approached the creation of a called Hybrid Clustering Method (HCM) of grouping that consists of the junction of the methods K-Means and Fuzzy C-Means (Table 2).

The approach of Jagadish (2014) performs group formation explicitly, it collects data through forms/questionnaires. The works of Abnar et al. (2012), Montazer and Rezaei (2012), and Yathongchai et al. (2013) form the groups implicitly, the data sources used to characterize the profiles of the students were the logs extracted from the Moodle platform.

Table 2. Comparison between works on the student grouping. Source: (Ramos et al., 2017)

Papers	Type of Data Collection	Used With Moodle	Data Source	Techniques Used to Form Groups
(Jagadish, 2014)	Explicit	Yes	Questionnaire	KNN
(Abnar et al., 2012)	Implicit	Yes	Logs (-)	Genetic algorithm
(Yathongchai et al., 2013)	Implicit	Yes	Logs (Login frequency, download time, and others)	K-Means and C4.5
(Montazer & Rezaei, 2012)	Implicit	-	Logs (-)	K-Means and C-Means Fuzzy

MATERIALS AND METHODS

An SRL of LP in LMS (Ramos et al., 2015) and group formation (Ramos et al., 2016) and two field studies are conducted. Data from an already closed class of an LMS were collected, a distance course in Applied Informatics from a program of Degree in Physical Education, with five classes and 124 students, during the second academic semester of 2013, where the interactions recorded in the LMS database were studied.

From analyzes performed, the representation of the model in the directed graph format was constructed. The number of interactions weights the vertices, and the edges have as weight the number of times the student has covered them. Next, two tools for Moodle in the plugin format to analyze the proposed model were developed, to validate the model and demonstrate its possible applications.

The first tool, called Learning Path Graph (LPGraph) (Ramos et al., 2016), generates the LP model and creates a visual representation of the LP. The second tool is called Moodle Cluster or M-Cluster (Ramos et al., 2017). It uses the model data in conjunction with the K-Means algorithm to suggest groupings for collaborative activities. The tools created particularly seek to help the teachers in their activities of behavior analysis and grouping of students. Data from 82 students were used to test the tool during development (training data).

Subsequently, experiments with three programs were conducted: Physics, Electrical Engineering (EE), Computer Science (CS). Each experiment lasted a semester. In 2015/2, the LPGraph was tested with 113 students, and in 2016/2, we tested the M-Cluster with one class of 40 students. The 153 students signed a consent form accepting to participate in the experiments. In all, 206 students participated from tool development tests and 153 from experiments.

LPGRAPH

This section presents a Moodle plugin called LPGraph developed using the LP model. The experiments showed the analysis and how the teacher can select students and the time interval to generate a visual representation.

Figure 3 shows the LPGraph plugin. LPGraph has data selection, selected data identification, graph options, the graph representing the LP, percentage of edges, and, finally, a list of resources and activities.

Teachers can select which data they want to use to generate the graph. Below the graph, there are the edges proportions. The proportions are calculated dynamically. When the teacher modifies the "Paths" option of the graph, the bar graph will adjust to the number of edges displayed. Within the bars have percentages of each type of edge, and the axis presents the absolute value. Below the bar chart is the list of resources and activities.

Figure 3. LPGraph Screen
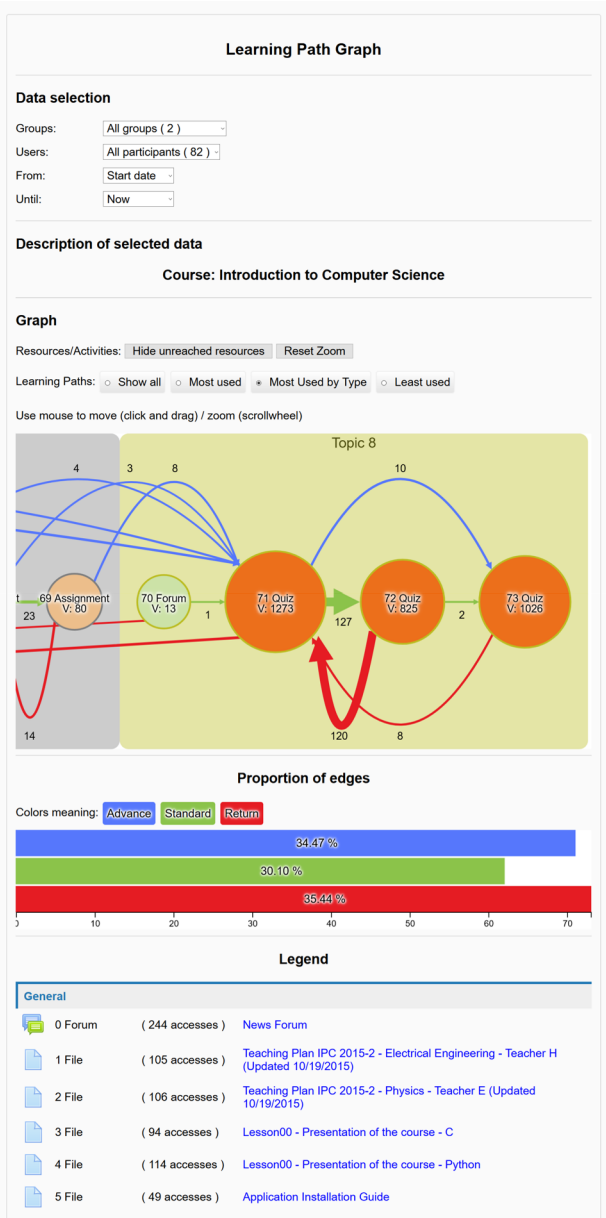
Implementation of Visual Representation

The representation has two levels. Level 1 represents course topics or units (Figure 4), and Level 2 displays the graph representing the LP (Figure 5).

Colored circles represent the vertices. Inside the circles appear three information:

- Order: the position of the vertex within the sequence (pos attribute).
- Module name: type of resource or activity such as 'Forum' or 'Assignment.'
- V: number of instances views. Indicates how many times students access the vertex.

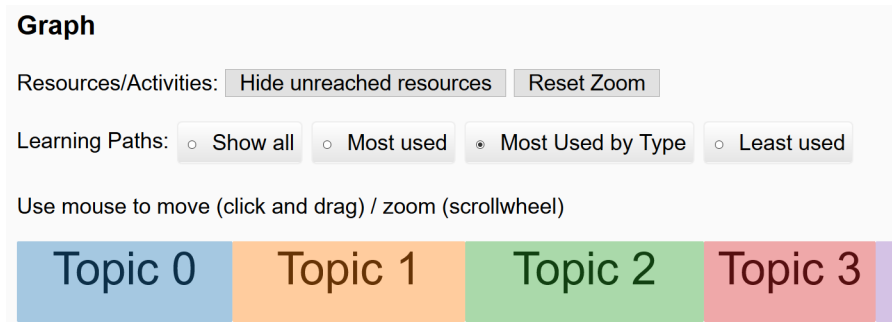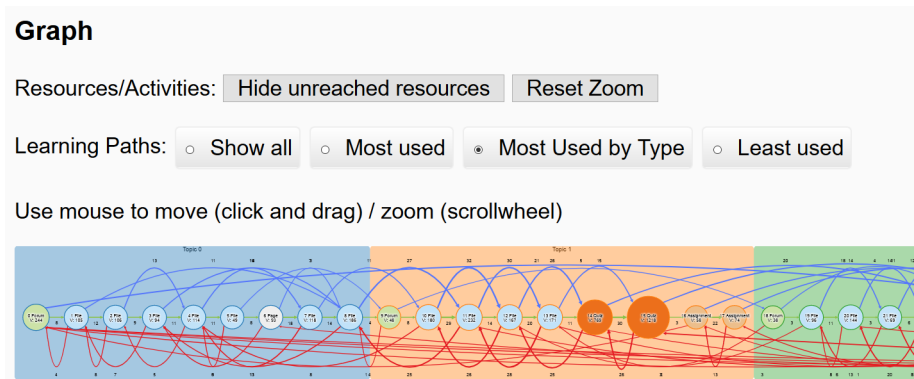Figure 4. Level 1 Visualization of Learning Paths

Figure 5. Level 2 Visualization of Learning Paths

The edges are colorful and follow the idea of the proposed model. The thickness of the edge is proportional to this value.

Regarding the options for viewing the paths, the teacher can show or hide resources and activities that are not accessed. Figure 6 shows all resources and activities, including vertices not accessed by students. Figure 7 shows only vertices accessed from the same course.

As many edges were generated, it was necessary to create some options to limit the number of edges displayed and thus, to facilitate the understanding of the graph. The options for displaying the LP are:

- Show all: Displays all edges of the graph.
- Most used: for each vertex, only the most used edge. This option presents the main LP.
- Most used by Type: for each vertex, the graph shows only the highest value edge of each type that leaves it. In this case, there are at most three edges per vertex.
- Least used: for each vertex, only the lower weighted edge of the vertex is displayed.

Experiment With LPGraph

The LPGraph plugin was used in two Moodle courses: Introduction to Computer Science (CS1), with one class of Physics (28 students), and one of EE (42 students), and Discrete Mathematics (DM) (43 students) from a class of CS.

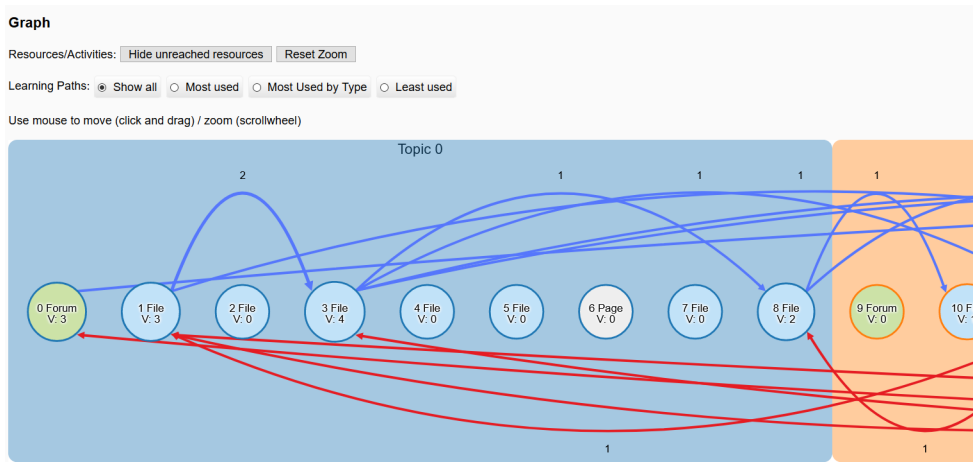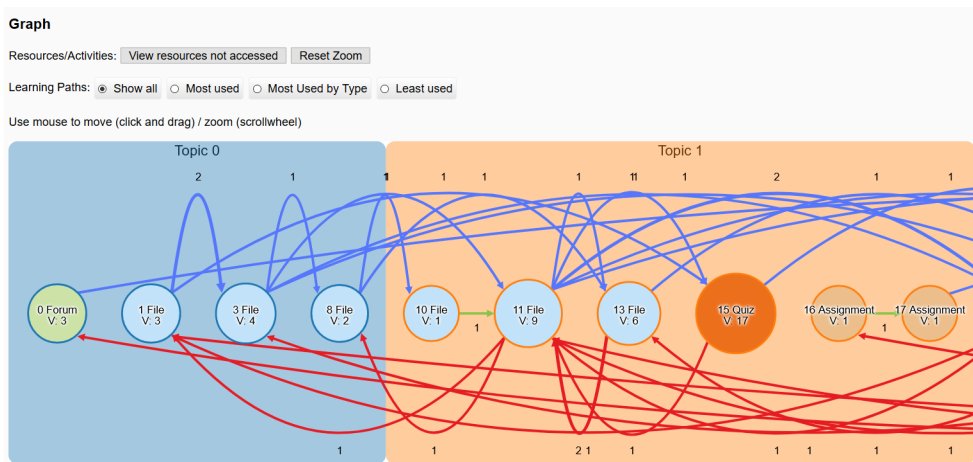**Figure 6. Visualization of all resources and activities**

Figure 7. Visualization of only the resources and activities accessed from the same course

Courses are in the topic format, which divided their content into modules according to the topic issue, ordered by the teachers. The objective of the first module, both CS1, and DM, was to address general guidelines such as teaching plan, presentation of the course, and use of a mobile application that was made available to students. Also, by default, the News forum was presented in the first module, where the teacher can communicate information about the course, such as the disclosure of notes, change of place/time of classes, or a new evaluation date.

About resources and activities, Moodle courses used: Forum, File, Page, Quiz, Assignment, and URL. Each module contained an exclusive question forum for the topic material. The resource File provides lesson content, exercise list, and guidelines for practical activities. The resource Page and the URL served to deliver links to external features. The Quiz and the Assignment are activities used to evaluate students. Figure 3 presents the options available to the teacher to specify the source of the data to be analyzed (Data selection). It can choose by group (class), users, and the range of data.

Next, the results of the experiment are presented. Figures show the "Most used" LP option to facilitate the visualization of the graphs.

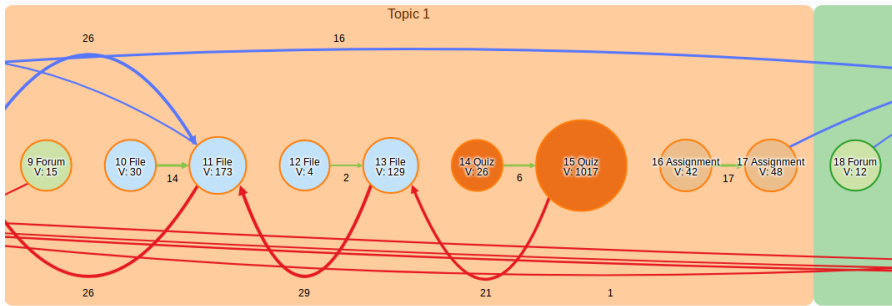**Figure 8. Part of the LP of the EE class in the CS1 course**

Figure 9. Part of the LP of the Physics class in the CS1 course
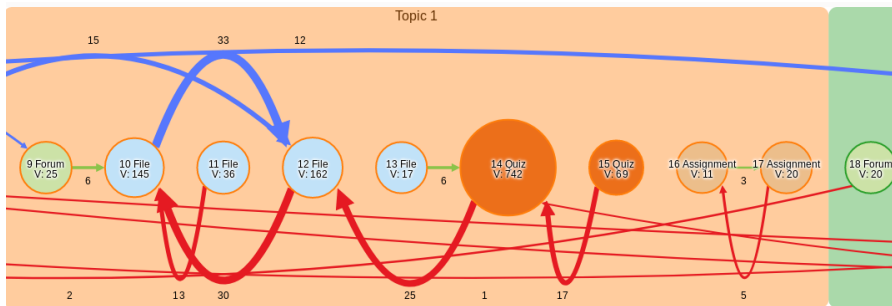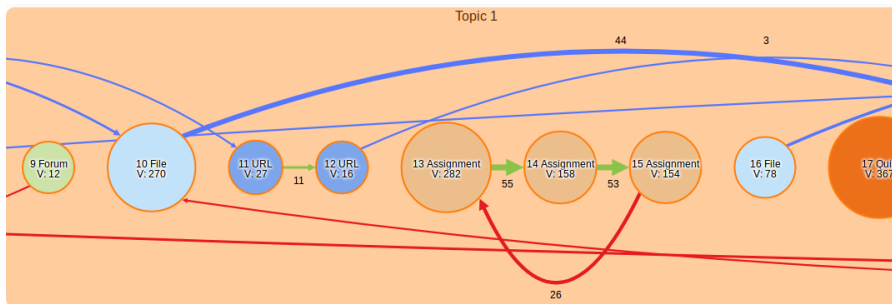

Figure 10. Part of the LP of the CS class in the DM course

Results Observed During the Experiment

This section presents an analysis of the visual representation of LPs. The teacher can visualize the entire LP, but this analysis was based on students' actions taken only on 'Topic 1' because it is impossible to show all topics here.

Figure 8, 9, and 10 present the LPs of the courses CS1 and DM partially, according to the available settings, with the selection of the class in the Group option for CS1 courses and showing only the most used paths. Figures present only a piece related to Topic 1 of each course. It is interesting to realize that Figures 8 and 9 are from the same course (CS1), analyzed during the same time interval, but each class presents a different behavior. The class of Physics has navigation more concentrated, which can be observed by the wider edges. It is also possible to conclude that vertices 10 (Lesson01

- Variables, and Sequential Structure - Python) and 12 (LabCod01 - Python) are essential for learners who access the questionnaire, indicated by vertex 14 (Quiz 1 - Python). In this case, it became clear that the studies of the class of Physics (Figure 9) focus on Python programming language.

Differently, the class of EE (Figure 8) studied the C programming language. So, these students access more the vertex 11 (File), 13 (File), and 15 (Quiz). In contrast, Physics students access vertex 10 (File), 12 (File), and 14 (Quiz).

In both classes, the learners interact much more with the questionnaires. Although the Physics program has fewer students than the EE program, the number of interactions with vertices of type File is similar. In the end, for the activity of the questionnaire, the class grade average is slightly higher for the Physics class, which in general interacted more, considering the number of students, they are 28 students against 42 of the EE class.

However, comparing the grade of Topic 1 between the two classes, the lowest grades were of the Physics class students. They returned to previous resources more frequently, which is indicated by return edges, comparing to the other edges' types. This scenario seems to suggest that the students had difficulties in understanding the topic concept and, therefore, had to go back to the previous content more times. The tendency to return to earlier resources remains higher during the course for Physics class. At the end of the course, analyzing the final grades, the uppermost grades are in the EE program.

Thus, for the analyzed classes, a higher proportion of the number of red edges indicates that there is a high chance of the students be experiencing learning difficulties.

Figure 10 presents a partial view of the LP of the Discrete Mathematics class. It shows that the class usually accesses resources directly but little access to external links (URL). In Topic 1 were 27 access (vertex 11) and 16 accesses (vertex 12), although the class had 43 students enrolled. The teacher, in possession of this information, could assess why students do not use these resources effectively. A student can view the same resource multiple times, that is, at most, 27 students accessed the vertex 11 (URL) if each of them accessed only once. Therefore, the teacher can also identify if students are using an available resource/activity.

LPGraph was also used to analyze student LP individually. In this case, The LPs of some students with high and low grades at the 'Topic 1' were observed. It was noticed that for the first case, students' paths contained few edges, and therefore were visually cleaner. Also, the most used edges are predominantly the 'advance' edges. When analyzing the LP with students with lower grades, it was observed that there are more edges than LP with students with higher grades. Also, they are more dispersed and that when selecting the option to visualize only the most used tracks, the amount of return edges is higher. Together with this information, a small number of edges may also indicate a tendency to evasion.

Figure 11 shows a student LPs, from the CS1 course, with grade close to ten (max grade) at the end of Topic 1 and the proportion of the number of edges per type.

Figure 12 shows the LPs of a student with a low grade for Topic 1. It also shows a more significant number of dispersed edges, and when viewing only the most used edges, the proportion of edges of type 'return' tends to increase.

Students with the lowest grades have more 'return' edges, and have limited access to activities, for example, the instance '15 Quiz'. Students whose grade is high have more 'advance' edges and interact more with activities. Both kinds of students accessed instances of file or forum occasionally, and their dispersion is opposite, students with higher grades often access instances close to each other, whereas students with the lowest grades navigate between instances that are far each other.

After the end of the course, a correlational analysis was performed between the attributes created from the model and the students' grades in Topic 1 and in the final grade. The Table 3 shows only the significant correlations ($\rho > 0.3$) found. The analysis was performed considering only the students with the 10 highest and 10 lowest grades in each class, as well as in the visual analysis, this allows to better analyze the differences between students with opposite performances. Thus, the relationship between the attributes and the grades can be better perceived.

Figure 11. Student LP of the Physics class, whose grade is high. This student uses more resources/activities.
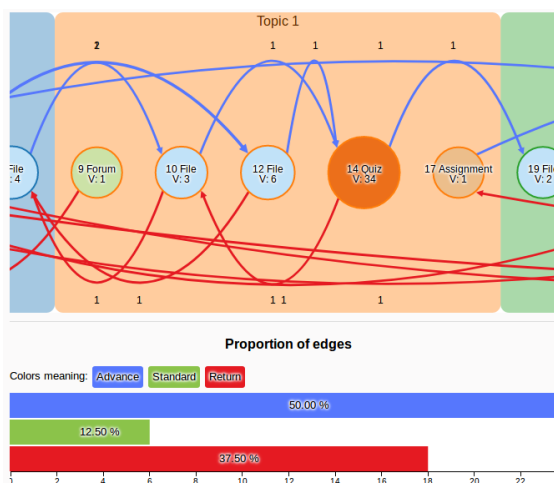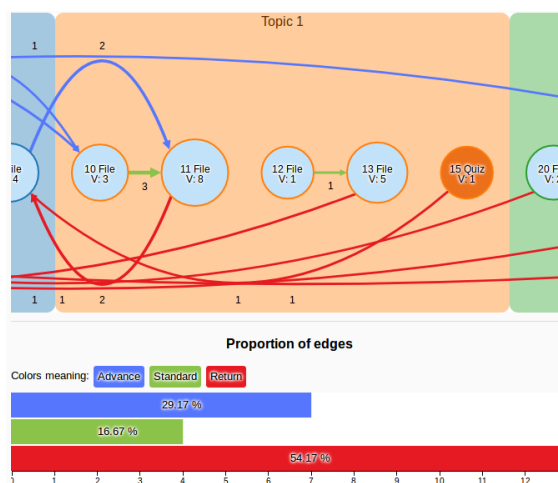

Figure 12. Student LP of the EE class, whose grade is low. This student accessed the quiz just only once.

Students with better grades tended to use the advance edges more. The metrics related to the advance edges confirm this observation, although they are more evident in the DM course. It is possible that the fact of the EE and Physics programs share the same environment may have generated lower correlation values. For example, while the EE class answered the C language quiz, the Physics class answered the Python language quiz, but both activities were visible for both programs. The average access to resources and activities is positively related to the grades, contributing to the results found in other works in the literature. When observing the negative correlations of the total and return dispersions, a more dispersed student may not have good grades depending on the type of dispersion. From an educational point of view, the advance dispersion can be understood as a student's quest to satisfy his/her curiosity about what subjects will come next and this dispersion is positively related to the grades. The metrics related to the return edges are positively correlated to the grades. Although the access to previous content is related to possible difficulties of the student, that such difficulties can be overcome by means of a content review before carrying out an evaluative activity.

Table 3. Pearson's correlation between attributes derived from LPs and students' grades

Attributes	EE - CS1		Physics - CS1		CS - DM	
	Grade at Topic 1	Final Grade	Grade at Topic 1	Final Grade	Grade at Topic 1	Final Grade
advance standard deviation	0.373	-	-	0.434	0.644	0.599
advance variance	0.307	-	-	0.378	0.651	0.613
average access to vertices	0.693	0.432	0.329	0.479	0.432	0.747
advance dispersion	0.410	-	-	0.413	0.505	-
total dispersion	-0.491	-	-0.535	-0.384	-	-
return dispersion	-0.470	-	-	-0.515	-	-0.519
return proportion	-	-	0.307	-	0.539	-
return standard deviation	0.628	-	-	0.560	-	0.577
return variance	0.508	-	-	0.458	-	0.561

M-CLUSTER

Groups are a set of individuals that forms and change themselves for multiple purposes. The model was used to develop the Moodle Cluster (M-Cluster) (Ramos et al., 2017) for group recommendation. The learning path model is the user model of M-Cluster. It uses the K-Means algorithm and three similarity metrics, which are the distances: Euclidian, Manhattan, and Cosine, along with attributes (vertex access average, amount, dispersion, and variance of standard, advance and return edges), obtained from the proposed LP model. The Euclidean distance was used because it is one of the most used metrics for calculating the center of the clusters, whereas the Manhattan distance is a simple version of the Euclidean distance. The cosine distance is a popular metric in recommendation systems in general. The tool used these similarity metrics because vectors represent the attributes and because they are metrics used in some works related to this research.

M-Cluster generates homogeneous groupings (groups with a similarity between LPs). The tool suggests to the teacher three results of group formation, and he/she can choose which one is more suitable for his/her students in a specific activity. The tool represents this grouping in two ways: the descriptive list (Figure 13), and the bubble chart (Figure 14).

The descriptive list is an unordered list in which a group identification followed by a list of names of the students belonging to the group. The bubble chart displays the bubble groups, and each bubble shows the names of students belonging to the same group.

Experiment With M-Cluster

It was conducted field research to collect training data from an LMS for two semesters, from 82 students of two classes already closed, one from the CS program and the other from the Computer Engineering program, where it was analyzed the interaction of group activities. From analyzes performed, the degree of similarity grouped students, calculated using the attributes extracted from LP.
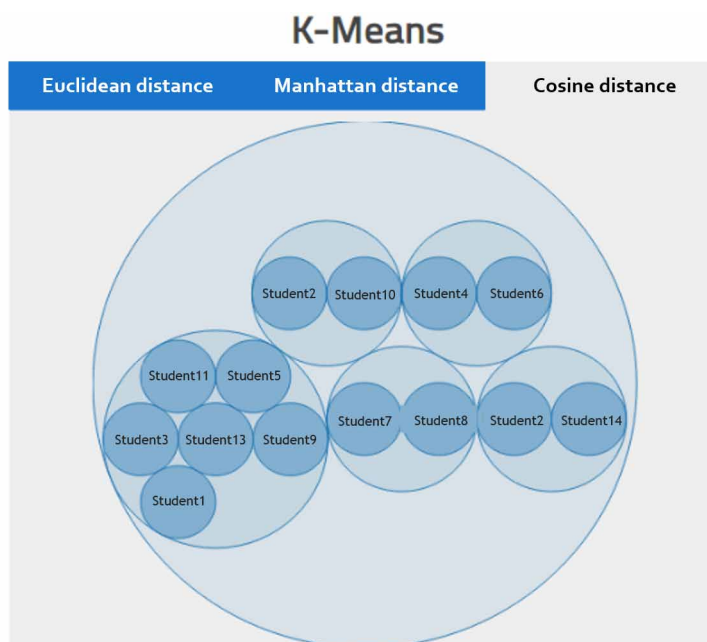
Thus, the formations of the students' group were analyzed using many cluster algorithms of data mining and machine learning techniques to generate the suggestions of groups of students. Then the best attributes (mentioned before) from LP for grouping were selected. Subsequently, with new data, it was performed validation in a Discrete Mathematics course, of a class of 40 students in progress in blended mode.

Figure 13. An Example of Descriptive Representation

Figure 14. An example of bubble chart visualization

Results Observed During the Experiment

Two activities were performed in the course to corroborate the indicated groups. In the first activity, students chose their group partners, and in the second, the groups' formation followed according to the tool suggestions. The teacher validated and visualized the groups suggested by the tool.

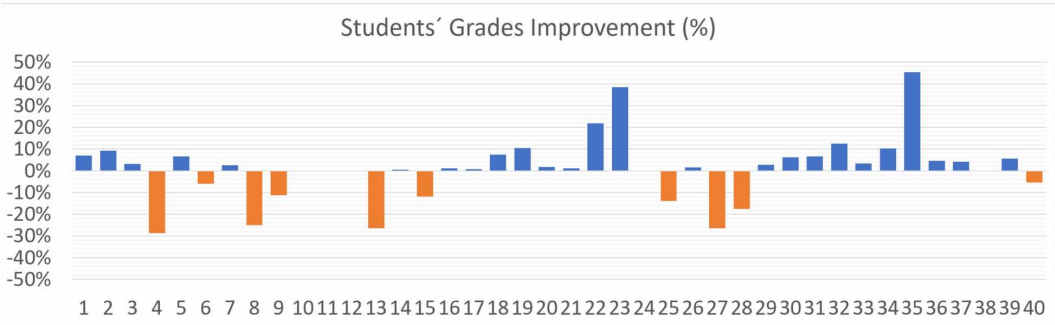**Figure 15. Discrete Mathematics students' grades improvement**

Figure 15 shows the individual improvement between activities of the analyzed students (n=40). Dropouts do not appear in the chart. The experiments showed that 75% of the students matched or exceeded their grades in the second activity concerning the grades achieved in the first activity. When analyzing groups, from the total number of formed groups, 30% were identical pairs to those of the first activity (these same students obtained good results in both activities). These results are not conclusive yet, but they are a starting point for deeper future analyzes.

RESULTS AND DISCUSSION

The studies were empirical, in which the results show that it is possible to use the proposed model to perform actions such as group formation and behavior analysis. The visual tool developed from the proposed model can show the student's and class' behaviors. The model was applied in group formation for collaborative activities, and the teacher of the course analyzed the suggested groupings and evaluated the results positively.

With LPGraph, the teacher can visually observe which resources and activities the student, group, or entire class are using more. Teachers can also track students' paths along time. The LP are dynamic, and LPGraph can generate visualizations for different configurations. In this way, our approach can make real-time analysis too. In the experiments it was possible to check some observations on student behavior using the proposed model, which confirms the first question. It was analyzed Topic 1 of two courses, CS1 and DM, and compare the different characteristics of the LPs such as access, the proportion between the type of edges, and dispersions. The tool can analyze a class in progress or an already closed class.

M-Cluster was able to suggest groups of students who performed the same or above the activity where students chose their partners. This observation confirms our second question. Nevertheless, it is important to conduct other experiments in different situations to verify the completeness of the approach. The study performed groups' formation of students with the similarity of LPs. So, groups formed are homogeneous. The tool, in some cases, grouped, for the second activity, the same students who chose themselves for the first activity.

CONCLUSION

This work presented a new representation model of LP and its application in group formation and behavior analysis. A tool called LPGraph that identifies and represents the LPs of students who use Moodle was created. Another developed tool from the model was the M-Cluster, which makes grouping suggestions by applying the K-Means algorithm with attributes generated from the proposed model.

LPGraph deals with the visual representation of the proposed LP model to assist the teachers in monitoring the learning process of their students. M-Cluster was implemented considering future improvements to allow more types of groupings, such as homogeneous, heterogeneous, and hybrid. Thus, the teacher can choose which type of grouping to use for a given collaborative activity. It is possible to cite as contributions of this work:

- Grouping of the students according to their LP.
- Identification of LP-based attributes in graph format since until then, they were not observed in the literature on the group recommendation of students in LMS.

New researches can use the presented model for many applications besides those applied in this work. Among the most relevant future works, it is possible to mention:

- Formation of both homogeneous and heterogeneous groups, giving the teacher the option to group individuals with complementary LP.
- Integration with the techniques of collaborative learning.
- Checking of which types of paths are most likely to improve student performance, and
- Contribution to the creation of adaptive LMS.

The tools have some limitations. LPGraph does not store LPs, it only generates them in real-time. The tool was tested in courses with topic format only. The M-Cluster still does not recommend heterogeneous groups. The study analyzed only the data of students who agreed to participate in the research, limiting the number of participants. Despite the above limitations, this work was concerned with developing an LP model in a way that could serve as a basis for other research. Based on what was observed, it is intended for future work to automate the analysis of LPs, reducing the effort required by the teachers to understand the behavior of their students. With this, it will be possible to provide diagnoses and suggestions for decision making. Currently, a study on the relationship between LPs and the achievement goals orientations theory is being conducted.

ACKNOWLEDGMENT

REFERENCES

Abnar, S., Orooji, F., & Taghiyareh, F. (2012). An evolutionary algorithm for forming mixed groups of learners in web based collaborative learning environments. In *Proceedings of 2012 IEEE International Conference on Technology Enhanced Education (ICTEE)*. IEEE. doi:10.1109/ICTEE.2012.6208612

Adesina, A., & Molloy, D. (2011). A Business Process Management Based Virtual Learning Environment: Customised Learning Paths. In *Proceedings of International Conference on Computer Supported Education (CSEDU 2011)*. SciTePress.

Baneres, D., Rodríguez-Gonzalez, M. E., & Serra, M. (2019). An Early Feedback Prediction System for Learners At-Risk within a First-Year Higher Education Course. *IEEE Transactions on Learning Technologies*, *12*(2), 249–263. doi:10.1109/TLT.2019.2912167

Cerezo, R., Suarez, N., Núñez, J. C., & Sánchez-Santillán, M. (2014). eGraph tool: graphing the learning process in LMSs. In *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge (LAK '14)*. Association for Computing Machinery. doi:10.1145/2567574.2567596

Huang, Y., Guerra-Hollstein, J. D., & Brusilovsky, P. (2016). Modeling skill combination patterns for deeper knowledge tracing. UMAP 2016 Extended Proceedings.

Jagadish, D. (2014). Grouping in collaborative e-learning environment based on interaction among students. In *Proceedings of 2014 International Conference on Recent Trends in Information Technology*. IEEE. doi:10.1109/ICRTIT.2014.6996170

Li, Y., Shao, Z., Wang, X., Zhao, X., & Guo, Y. (2019). A Concept Map-Based Learning Paths Automatic Generation Algorithm for Adaptive Learning Systems. *IEEE Access: Practical Innovations, Open Solutions*, 7, 245–255. doi:10.1109/ACCESS.2018.2885339

Marquez, J. M., Ortega, J. A., Gonzalez-Abril, L., & Velasco, F. (2008). Creating adaptive learning paths using Ant Colony Optimization and Bayesian Networks. In *Proceedings of IEEE International Joint Conference on Neural Networks*. IEEE. doi:10.1109/IJCNN.2008.4634349

Montazer, G. A., & Rezaei, M. S. (2012). A new approach in e-learners grouping using Hybrid Clustering Method. In *Proceedings of International Conference on Education and E-Learning Innovations*. IEEE. doi:10.1109/ICEELI.2012.6360671

Nurjanah, D., & Fiqri, M. (2017). Graph-based domain model for adaptive learning path recommendation. In *Proceedings of IEEE Global Engineering Education Conference (EDUCON)*. IEEE. doi:10.1109/EDUCON.2017.7942875

Pires, J. M., & Cota, M. P. (2010). Evolutive mechanism for E-Learning platforms: A new approach for old methods. In *Proceedings of IEEE EDUCON 2010 Conference*. IEEE. doi:10.1109/EDUCON.2010.5492482

Ramos, D. B., & Oliveira, E. H. T. (2015). *Relatório Técnico da Revisão Sistemática de Literatura sobre Trilhas de Aprendizagem em Ambientes Virtuais de Aprendizagem* [Technical Report of the Systematic Literature Review on Learning Paths in Virtual Learning Environments]. http://mobmoodle.icomp.ufam.edu.br/projeto/wp-content/uploads/2014/09/RT-da-RSL-sobre-LMS-e-LP-impressao-v1.pdf

Ramos, D. B., Oliveira, E. H. T., Monteverde, I. M. M., & Oliveira, K. M. T. (2015). Trilhas de Aprendizagem em Ambientes Virtuais de Ensino-aprendizagem: Uma Revisão Sistemática da Literatura [Learning Paths in Virtual Teaching and Learning Environments: A Systematic Literature Review]. In *Proceedings of the XXVI Brazilian Symposium on Computers in Education (SBIE 2015)*. Sociedade Brasileira de Computação. doi:10.5753/cbie.sbie.2015.338

Ramos, D. B., Ramos, I. M. M., Nascimento, P. B., & Oliveira, E. H. T. (2016). Uma Ferramenta Baseada em Grafo para Identificação e Visualização de Trilhas de Aprendizagem [A Graph-Based Tool for Identification and Visualization of Learning Paths]. In J. Sánchez (Ed.), *Nuevas Ideas en Informática Educativa* (pp. 237–243). Universidad de Chile.

Ramos, I. M. M., Ramos, D. B., Amaral, G. S., Nascimento, P. B., Gadelha, B., & Oliveira, E. H. T. (2017). M-Cluster: Uma ferramenta de Recomendação para Formação de Grupos em Ambientes Virtuais de Aprendizagem [M-Cluster: A Recommendation Tool for Group Formation in Virtual Learning Environments]. In *Proceedings of the XXVIII Brazilian Symposium on Computers in Education (SBIE 2017)*. Sociedade Brasileira de Computação. doi:10.5753/cbie.sbie.2017.1657

Ramos, I. M. M., Ramos, D. B., Amaral, G. S., Gadelha, B. F., & Oliveira, E. H. T. (2018). Framework Conceitual para Formação de Grupos de Alunos utilizando Trilhas de Aprendizagem em um Ambiente Virtual de Aprendizagem [Conceptual Framework for Student Groups Formation using Learning Paths in a Virtual Learning Environment]. In *Proceedings of the XXIX Brazilian Symposium on Computers in Education (SBIE 2018).* Sociedade Brasileira de Computação. doi:10.5753/cbie.sbie.2018.1673

Ramos, I. M. M., Ramos, D. B., Nascimento, P. B., Gadelha, B. F., & Oliveira, E. H. T. (2016). Formação de Grupos em Ambientes Virtuais de Aprendizagem: Uma Revisão Sistemática da Literatura [Groups Formation in Virtual Learning Environments: A Systematic Literature Review]. In *Proceedings of the XXVII Brazilian Symposium on Computers in Education (SBIE 2016)*. Sociedade Brasileira de Computação. doi:10.5753/cbie.sbie.2016.310

Saito, T., & Watanobe, Y. (2020). Learning path recommendation system for programming education based on neural networks. *International Journal of Distance Education Technologies*, *18*(1), 36–64. doi:10.4018/IJDET.2020010103

Schröck, J., Bargel, B. A., Roller, W., & Rauner, A. (2010). Learning maps usage of cartographic metaphors for orientation in e-learning courses. In *Proceedings of IADIS International Conference on E-Learning*. IADIS.

Sitthisak, O., Gilbert, L., & Albert, D. (2013). Learning in Moodle using Competence-Based Knowledge Space Theory and IMS QTI. In *Proceedings of 2013 International Computer Science and Engineering Conference (ICSEC)*. IEEE. doi:10.1109/ICSEC.2013.6694752

Teutsch, P., & Bourdet, J.-F. (2010). How to See Training Paths in Learning Management Systems? In *Proceedings of 2010 10th IEEE International Conference on Advanced Learning Technologies (ICALT)*. IEEE. doi:10.1109/ICALT.2010.98

Vantroys, T., & Peter, Y. (2003). COW, a Flexible Platform for the Enactment of Learning Scenarios. In J. Favela & D. Decouchant (Eds.), *Groupware: Design, Implementation, and Use* (pp. 168–182). Springer Berlin Heidelberg. doi:10.1007/978-3-540-39850-9_15

Weiand, A., Manssour, I. H., & Silveira, M. S. (2019). Visual analysis for monitoring students in distance courses. *International Journal of Distance Education Technologies*, *17*(2), 18–44. doi:10.4018/IJDET.2019040102

Wisuttikul, T., & Boonmee, C. (2004). A study of Coaching Assisted System using Learning Object value model. In *Proceedings of 2004 IEEE Region 10 Conference TENCON 2004*. IEEE. doi:10.1109/TENCON.2004.1414589

Yathongchai, C., Angskun, T., Yathongchai, W., & Angskun, J. (2013). Learner classification based on learning behavior and performance. In *Proceedings of 2013 IEEE Conference on Open Systems (ICOS)*. IEEE. doi:10.1109/ICOS.2013.6735050

David Brito Ramos received the technologist's degree in Technology in Data Processing from the State University of Amazonas (2007), a bachelor's degree in Computer Science (2010), and a master's degree in Informatics (2016) from the Federal University of Amazonas (Ufam). He is currently a Ph.D. student in Informatics at Ufam, a member of the Innovation Network for Brazilian Education, and an active professor at the Federal Institute of Education, Science, and Technology of Amazonas (IFAM). His research interests include Informatics in Education, Collaborative Systems, Artificial Intelligence, Programming Languages, Computing Education, and Mobile Learning. More details in his CV: http://lattes.cnpq.br/5494652135358138.

Ilmara Monteverde Martins Ramos has got graduation in Technology in Data Processing from the State University of Amazonas (2007), specialization in Residency in Web Software Development from Center for Analysis, Research and Technological Innovation Foundation (2009) and a master's degree in Informatics from Federal University of Amazonas (2018), all in Manaus, Brazil. She is currently an active professor at the Federal Institute of Education, Science, and Technology of Amazonas - IFAM Campus Parintins. Her research interests include Informatics in Education, Collaborative Systems, Artificial Intelligence, Software Engineering, and Mobile Learning. More details in her CV: http://lattes.cnpq.br/0006832726895462.

Isabela Gasparini received the bachelor's degree from Londrina State University, in 1999, and the master's and Ph.D. degrees from the Federal University of Rio Grande do Sul, in 2003 and 2013, respectively, all in computer science. She is currently an Associate Professor at the Santa Catarina State University, where she is involved in human-computer interaction and technology-enhanced learning fields, with special interest in adaptive e-learning systems, gamification, learning analytics, context-awareness, and cultural issues. She is the editor-in-chief of the Brazilian Journal of Computers in Education (2019-2021). More details in her CV: http://lattes.cnpq.br/3262681213088048.

Elaine Oliveira has got a bachelor's degree in Computer Science (1997) and a master's in Applied Physics, both from São Paulo University (2001), and a Ph.D. in Computers and Education from Federal University of Rio Grande do Sul (2011), all in Brazil. Since 2002, she is a professor at the Institute of Computing at the Federal University of Amazonas. Her main research interests are in Computers and Education and her work currently focuses on educational software, adaptive learning, educational data mining, and learning analytics. She is an associate editor of the Brazilian Journal of Computers in Education (2019-2021). More details in her CV: http://lattes.cnpq.br/6553721651836761.