Visualization Index for Educational Resources by Learning Analytics

Noemí DeCastro-García[0000−0002−5610−0153] and Ángel Luis Muñoz Castañeda[0000−0001−6993−9110]

[1] Department of Mathematics, Universidad de León, Campus de Vegazana s/n 24071, León, Spain `ncasg@unileon.es`

[2] Research Institute of Applied Science in Cybersecurity (RIASC), Universidad de León, León, Spain `amunc@unileon.es`

Abstract. In this paper, we propose an oriented-graph - design of the database generated in a virtual educational platform with the records of students access to the learning resources. This theoretical model lets us compute a visualization index for students and for the available material in the platform in a total and a partial way, by the 1-norm of the adjacency matrix of the graph. These coefficients let us construct a classification system that it could be useful for determining different levels for students and for the material. Then, we can propose different scaffolding for the learning activities and effective learning outcomes for a meaningful learning experience depending on the interaction between the students and the resources.

Keywords: Learning analytics · Learning design · Learning resources

1 Introduction

Current educational context is characterized by the diversity of the instruction, students, and multimedia supports. Moreover, there exist a large number of hybrid courses in which the blended, online and face-to-face learning are present at the same time. In this framework, a thought about the material provided by the teacher is an essential action in order to find learning patterns and improve the learning and teaching processes.

Learning analytics techniques may provide very useful tools that help the teachers to optimize their work (see [1]). The second axiom that is used in [2] to analyze the development of the learning analytics field is based on the idea that learners are agents. This assumption implies that they have *the capability to exercise choice in reference to preferences* (see [3]). And one of the conditions that could influence in their preferences is the course instructional design, that includes the learning material. The analysis of data that are generated from the interactions between the students and the virtual environment can be used to predict the achievement of learning outcomes, take decisions on resource design or analyze the evolution or behavior of the students ([4]).

Although learning analytics give the teachers methods and tools of gathering information on how learners are interacting with learning resources, usually there is a gap between this information and the pedagogical actions that help to the teachers in their learning designs ([5]). Moreover, this gap is bigger in situations in which the teachers have not statistical or technical profile. In this scenario, it is necessary that the learning analytics and learning design come together.

Learning designs describe frameworks that can be used to help the teachers in the design and the choice of learning resources, learning tasks or activities, and learning supports in order to create a meaningful learning experience for the students, especially with the use of Information and Communication Technologies, (see [6],[7], [8] and [9]).

In the conceptual framework that links learning analytics and learning design, we have four dimensions on which we can work: temporal analytics, comparative analytics, cohort dynamics and tool specific analytics . The third one, the cohort dynamics, is a category that helps us to propose different learning outcomes depending on the different interaction patterns manifested in a course between students and resources. In this context, the study of whether a student has accessed or not to a specific resource is one of the most requested questions by different educational agents as desired information about a course (see [5]). The answer to this question could help to identify what learning activities or resources need to be modified in order to adapt them to the needs of students (see [10]). The above question could appear a very simple issue, and frequently, the data that let us answer it are generated in the educational platform as Moodle. However, the usual learning management systems have not available simple tools that provide this information in a visual and easy form (see [5]). The data are structured, but they are registers or logs of activity with a lot of redundant features that mean difficulties to analyze, visualize and discuss, in a global way with effective filters, especially for non-expert teachers in analytical studies.

One of the usual available tools in an educational platform is the flow visualizations of the number of visits that each resource has had trough the time. These graphics are often performed with the absolute frequencies of the visits, so its interpretation has to be done carefully because it could be deceptive and confused. For example, we can have a resource that has the highest number of visits and, however, it has been seen only by a minority number of the students. In Figure 1 we can see a typical flow of the visits of the resources by the students in a course in Moodle. We have information about the number of visits per day but we do not know what resources have been visited or the students that have accessed to the material. Moreover, it is necessary to take into account that the interpretation of the obtained results can imply different learning insights depending on the context of implementation. So, simple tools of an overview of the results, for students and resources, are needed in order to get a good understanding of this type of cohort dynamics.

This is the motivation for this work: to obtain more specific information about the visualization of the educational resources by the students, in an easy way
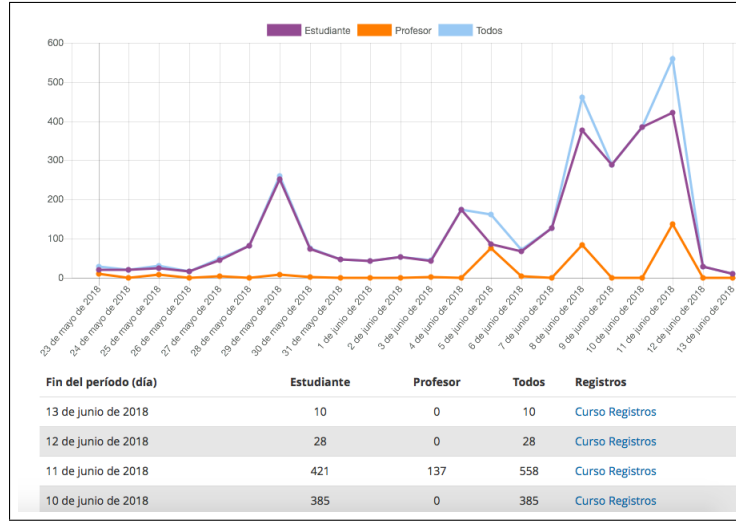
Fig. 1. Flow visualizations by Moodle

that let us create a simple tool. We propose a simple mathematical model that allows us to analyze the interaction between them and, finally, to obtain a classification system for both of them. This approach is based on the consideration of the visualization index. Its computation and the subsequent classification let us provide an overview of the results of the analyses that it is easily understandable for non- experts in learning analytics. In addition, this index can be applied for a global course, and in a partial form (for students or for resources).The proposed model is based on an oriented graph having as nodes both, the students and the resources. Since the most of the learning management systems for education have implemented models with social network analysis, it would make easier its integration in the usual educational platforms.

This work is organized as follows: In Section 2, we list the goals of this article. The proposed model to analyze the data is developed in Section 3, together with a simple example that helps to understand the essential ideas. Finally, the conclusions and references are given.

2 Goals

The main goals of this work are:

1. To define a measurement index of visualization that let us analyze what happens or what happened in a course regarding the interaction between the students and the provided resources.
2. To define a classification system for resources and students that depends on the developed metric.
3. To obtain a simple procedure that gives us a general overview of the situation.

3 Main results

This section describes the design of the mathematical model that let us compute the visualization indices of a course.

3.1 Model of the database

We have used a graph approach to model the data. A graph database is a database that can be structured in graph form so that the nodes of the graph contain the information, and the edges contain properties and/or define relations between the information contained in the nodes. One of the main strengths of this kind of databases is the capability to give answers in short time for questions regarding relations (see [11]).

Remark 1. We will state the following notation

1. The set $X = \{x_1, \ldots, x_l\}$ is going to be the set of nodes that represent the students.
2. The set $Y = \{y_1, \ldots, y_s\}$ contains the resources that the teacher provides through the virtual platform
3. The database D that we can download from the virtual platform, usually in .csv format. This database contains the set of all registers R_k. Each register represents a case in which each student visits one specific resource.

We can now attach a graph structure to D. In order to do so, we have to define the set of nodes, N, and the set of arrows, A. The graph, $G = (N, A)$, is going to be composed by two layers of nodes in such a way that all the arrows have their source in one layer and target in the other one.
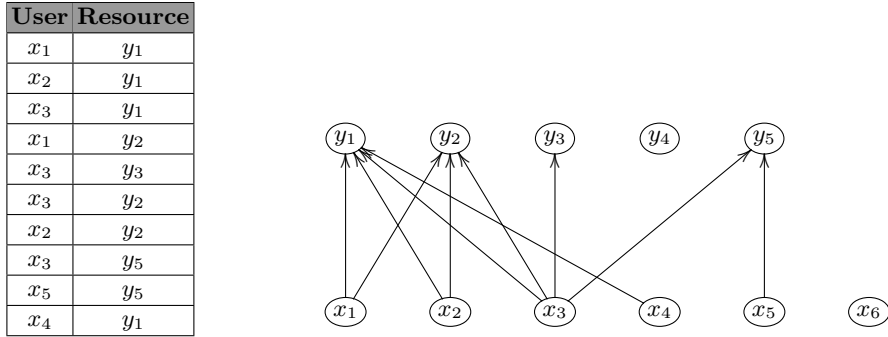
1. Layer 1: the nodes are the elements of the set X.
2. Layer 2: the nodes are the elements of the set Y.
3. We have an edge $x_i \to y_j$ if and only if the student x_i has visited the resource y_j. From now on this relation will be expressed as $y_j \in x_i$.

Example 1. In this example, we suppose that we have a face-to-face course with six students and five resources (four of them about contents of the course, and y_5 that is a learning task). So, the layer 1 has six nodes and the layer 2 has five nodes. We suppose that we have obtained the database D that is shown in Table 1.

Table 1. Example of D of registers.

Date & time	User	Resource	Component	Source	IP
1/11/2018 11:25	x_1	y_1	System	web	79.109.36.207
1/11/2018 11:40	x_2	y_1	System	web	188.76.8.106
1/11/2018 18:36	x_2	y_1	System	web	188.76.8.106
2/11/2018 12:10	x_1	y_1	System	web	79.109.36.207
2/11/2018 12:20	x_3	y_1	System	web	80.221.98.45
2/11/2018 12:55	x_1	y_2	System	web	79.109.36.207
2/11/2018 15:36	x_3	y_3	System	web	80.221.98.45
3/11/2018 19:46	x_3	y_1	System	web	80.221.98.45
4/11/2018 20:54	x_4	y_1	System	web	192.220.166.45
6/11/2018 17:36	x_3	y_2	System	web	88.13.134.88
6/11/2018 17:52	x_2	y_2	System	web	188.76.8.106
6/11/2018 21:52	x_3	y_2	System	web	88.13.134.88
7/11/2018 21:52	x_3	y_5	System	web	88.13.134.88
7/11/2018 21:52	x_5	y_5	System	web	188.76.8.108

The look of the corresponding graph of D and the filtered data are in Figure 2.

User	Resource
x_1	y_1
x_2	y_1
x_3	y_1
x_1	y_2
x_3	y_3
x_3	y_2
x_2	y_2
x_3	y_5
x_5	y_5
x_4	y_1

Fig. 2. Graph and filtered database associated to D

In the sequel, B will denote the set $\{0, 1\}$, and the set of matrices with entries in B with l rows and s columns will be denoted by $B^{l \times s}$.

Definition 1. *The adjacency matrix of D is defined as the adjacency matrix, $A \in B^{(l+s) \times (l+s)}$, of the associated graph:*

$$\begin{array}{c} \begin{array}{cc} x_1 \ldots x_l & y_1 \ldots y_s \end{array} \\ \begin{array}{c} x_1 \\ \vdots \\ x_l \\ y_1 \\ \vdots \\ y_s \end{array} \left(\begin{array}{c|c} \mathbf{0} & \mathcal{C} = (c_{ij}) \\ \hline \mathbf{0} & \mathbf{0} \end{array} \right) \end{array}$$

where

$$c_{ij} = \begin{cases} 1 & \text{if } y_j \in x_i \\ 0 & \text{if } y_j \notin x_i \end{cases} \tag{1}$$

for $i = 1, \ldots, l$, $j = 1, \ldots, s$

Remark 2. Note that we are only interested in the block $\mathcal{C}$ that have been defined. So, the study of the adjacency matrix is reduced to the study of the matrix $\mathcal{C}$. For this reason, in this paper, we will use both, the letter A and the letter $\mathcal{C}$, to make reference to such adjacency matrix if there is no possible confussion.

Example 2. The adjacency matrix associated to the vector database of Example 1 is

$$A = \begin{pmatrix} 0\,0\,0\,0\,0\,0\,\mathbf{1}\,\mathbf{1}\,\mathbf{0}\,\mathbf{0}\,\mathbf{0} \\ 0\,0\,0\,0\,0\,0\,\mathbf{1}\,\mathbf{1}\,\mathbf{0}\,\mathbf{0}\,\mathbf{0} \\ 0\,0\,0\,0\,0\,0\,\mathbf{1}\,\mathbf{1}\,\mathbf{1}\,\mathbf{0}\,\mathbf{1} \\ 0\,0\,0\,0\,0\,0\,\mathbf{1}\,\mathbf{0}\,\mathbf{0}\,\mathbf{0}\,\mathbf{0} \\ 0\,0\,0\,0\,0\,0\,\mathbf{0}\,\mathbf{0}\,\mathbf{0}\,\mathbf{0}\,\mathbf{1} \\ 0\,0\,0\,0\,0\,0\,\mathbf{0}\,\mathbf{0}\,\mathbf{0}\,\mathbf{0}\,\mathbf{0} \\ 0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0 \end{pmatrix} \tag{2}$$

where the matrix $\mathcal{C}$ is highlighted in bold characters.

3.2 Visualization indices

Once we have obtained the adjacency matrix associated to D, we can compute the total visualization index of the course by the following formula

Definition 2. *The total visualization index, ι_D, is the rate of resources that have been visited, at least, one time.*

$$\iota_D := \frac{\| A \|_1}{l \cdot s}, \tag{3}$$

$\| A \|_1$ *being the 1-norm of the matrix A, that it is, the number of $1's$ in the matrix A.*

Ths visualization index is a real positive number less or equal than one, $\iota_D \in [0,1]$.

Definition 3. *The partial row-visualization index, ι_{x_i}, is the rate of resources that have been visited by the student x_i at least one time,*

$$\iota_{x_i} := \frac{\| a_{i\bullet} \|_1}{s}, \tag{4}$$

$\| a_{i\bullet} \|_1$ *being the 1-norm of the i-th row of the matrix A, that it is, the number of $1's$ in the i-th row of the matrix A. As more resources the student visit, as higher the index is.*

Definition 4. *The partial column-visualization index, ι_{y_j}, is the rate of students that have visited the resource y_j at least one time,*

$$\iota_{y_j} := \frac{\| a_{\bullet j} \|_1}{l}, \tag{5}$$

$\| a_{\bullet j} \|_1$ *being the 1-norm of the j-th column of the matrix A, that it is, the number of $1's$ in the j-th column of the matrix A. As more students visit the resource, as higher the index is.*

Like the visualization index, the partial visualization indices are positive real numbers less or equal than one.
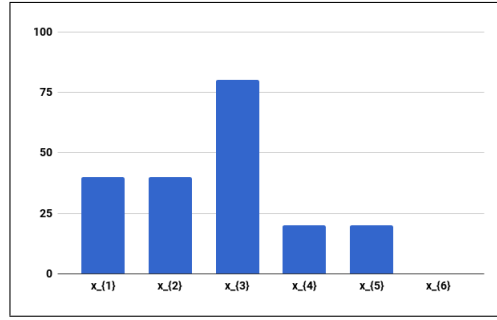
We highlight that the value of the visualization index has a different interpretation depending on the type of the course we are teaching. For instance, we have to take into account that the index has different consequences if we deal with an in-person course or an online course. This fact is very important in the prescriptive design stage. The ideal scenario would get the indices in different checkpoints that let us analyze the dynamic of the course.

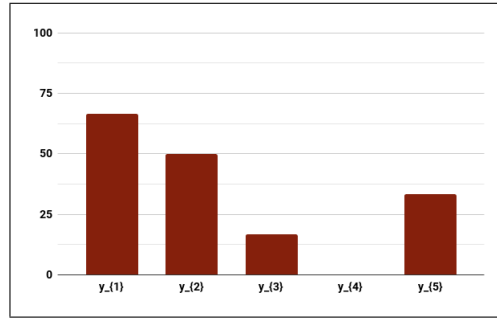*Example 3.* The total visualization index for D given in Example 1 is

$$\iota_D = \frac{10}{30} = 0,\widehat{3}, \tag{6}$$

As we can see in the partial row-visualization indices shown in Figure 3, the student x_6 has not visited any resource. On the other hand, the student x_3 is the one that more different resources has visited.

The partial column-visualization indices are included in Figure 4. In this case, the resource y_1 has a high visualization index. It could mean that the content of the resource has not been well understood because of most of the students have revisited it. Another possibility is that this resource is directly related to the assessment. In the case of y_4, the interpretation is the inverse. Finally, at the stage computation, the task has been visited by less than half of the students.

Fig. 3. Students visualization indices of D (%).

Fig. 4. Resources visualization indices of D (%).

3.3 Classification system

Although we can have a global insight of the use of the resources with the values of the indices mentioned above, it would be interesting to have a visualization tool that let us label the resources or the students depending on the value that the index has. This system provides us with a ranking of the most visited resources and the most visiting students. The system is done by constructing classification intervals, each of them carrying with a label, in such a way that we assign to every student or resource the label corresponding to the interval to which the visualization index belongs.

We develop the procedure only for the resources since for the other case the procedure can be developed in a similar way.

Definition 5. *Let $\mathcal{I}$ be the set of partial column-visualization indices of all resources. Let $\mathcal{C}$ be the set $\{v_{min}, I_1, I_2, I_3, I_4, v_{max}\}$ where v_{min} and v_{max} are the minimal and maximal values that the visualization indices take in the database D, and I_1, I_2, I_3, I_4 are the intervals between Q_1, Q_2, Q_3, the quartiles of the set $\mathcal{I}$. Then, we define the following map*

$$\zeta : Y \longrightarrow \mathcal{C}$$

$$y_j \mapsto \zeta(y_j) = \begin{cases} v_{min} & if & \iota_{y_j} \leq v_{min} \\ I_1 & if & v_{min} < \iota_{y_j} & \leq Q_1 \\ I_2 & if & Q_1 < \iota_{y_j} & \leq Q_2 \\ I_3 & if & Q_2 < \iota_{y_j} & \leq Q_3 \\ I_4 & if & Q_3 < \iota_{y_j} & < v_{max} \\ v_{max} & if & \iota_{y_j} \geq v_{max} \end{cases}$$

Example 4. In Table 2 we can observe the obtained ranking of D.

Table 2. Classification system of D. Ranking of students and resources

Interval	Resource	Student
Max	y_1	x_3
I_4		
I_3	y_2	x_1, x_2
I_2	y_5	
I_1	y_3	x_4, x_5
Min	y_4	x_6

4   Conclusions

Learning analytic techniques, together with strategies for learning design, may provide very useful tools that let the teachers optimize their sequences of the didactical resources.One of the current challenges that the techniques and systems of learning analytics face are to help to assess the success and adequacy of a concrete educational resource, taking into account the pedagogical and local context of the course in which these procedures are being implemented. In this work, we propose a mathematical model that can be useful to solve this need.

Our future work is related to the use of the visualization index to obtain information about an effective instructional course by scaffolding. In addition, we are working in a visualization tool, based on the classification system developed in this work, that allows the users to optimize the learning resource design of a course, and purpose prescriptive actions based on learning design for the students.

References

1. Siemens, G., Gašević, D. , Special Issue on Learning and Knowledge Analytics. Educational Technology & Society, **15**(3), 1–163, (2012).

2. Gaŝević, D., Dawson, S. & Siemens, G., Let's not forget: Learning analytics are about learning. TechTrends, **59**(1), 64–71, (2015).

3. Winne, P. H., How Software Technologies Can Improve Research on Learning and Bolster School Reform. Educational Psychologist, **41**(1), 5–17. (2006).

4. Gaŝević, D., Dawson, S., Rogers,T. & Gasevic, D. , Learning analytics should not promote one size fits all: The effects of instructional conditions in predicating academic success. Internet and Higher Education, 28, 68–84, (2016).

5. Bakharia, A., Corrin, L., de Barba, P., Kennedy, G., Gasevic, D., Mulder, R., Williams, D., Dawson, S. & Lockyer, L., A conceptual framework linking learning design with learning analytics. In Proceedings of the Sixth International Conference on Learning Analytics & Knowledge, 329–338. (2016)

6. Goodyear, P., Teaching, technology and educational design: The architecture of productive learning environments. The Australian Learning and Teaching Council, (2009).

7. Lockyer L., Heathcote, E. & Dawson, S., Informing pedagogical action: Aligning learning analytics with learning design. American Behavioral Scientist, **57**(10), 1439–1459, (2013).

8. Lockyer, L., Bennett, S., Agostinho, S., & Harper, B., Handbook of research on learning design and learning objects: issues, applications, and technologies (2 volumes). IGI Global, Hershey, PA, (2009).

9. Oliver, R. Exploring strategies for online teaching and learning. Distance Education, **20**(2), 240–254, (1999).

10. Persico, D., Pozzi, F., Informing learning design with learning analytics to improve teacher inquiry. British Journal of Educational Technology, **46**(2), 230–248, (2015).

11. Robinson,I. , Webber,J., Eifrem,E., Graph Databases.New Opportunities for Connected Data, 2nd Edition, O'Reilly Media, (2015).