On the discovery of educational patterns using biclustering

Rui Henriques*1[0000-0002-3993-0171] , Anna Carolina Finamore*1[0000-0001-7182-2683] and Marco Antonio Casanova2[0000-0003-0765-9636]

[1] INESC-ID and Instituto Superior Técnico, Universidade de Lisboa, Portugal
[2] PUC-Rio, Brasil
[*] Co-first author
rmch@tecnico.ulisboa.pt,anna.couto@ist.utl.pt,casanova@inf.puc-rio.br

Abstract. The world-wide drive for academic excellence is placing new requirements on educational data analysis, triggering the need to find less-trivial educational patterns in non-identically distributed data with noise, missing values and non-constant relations. Biclustering, the discovery of a subset of objects (whether students, teachers, researchers, courses and degrees) correlated on a subset of attributes (performance indicators), has unique properties of interest thus being positioned to satisfy the aforementioned needs. Despite its relevance, the potentialities of applying biclustering in the educational domain remain unexplored. This work proposes a structured view on how to apply biclustering to comprehensively explore educational data, with a focus on how to guarantee actionable, robust and statistically significant results. The gathered results from student performance data confirm the relevance of biclustering educational data.

Keywords: Biclustering, Pattern Mining, Educational Data Mining.

1 Introduction

Large volumes of educational data are increasingly collected due to a closer monitoring of students, teachers, researchers and staff, with the aim of pursuing academic excellence. This context poses new challenges on extracting meaningful and non-trivial educational patterns to support academic decisions.

Current approaches for educational pattern mining are still unable to reveal the true potential underlying educational data [20]. In its simplest form, educational data gather the performance of a set of objects (such as students, teachers, researchers, courses, degrees, among others) along a set of attributes (performance indicators). Although clustering and pattern mining are typically used to explore such educational data, they are unable to fully extract the hidden knowledge. Clustering simply groups objects (attributes) according to all available values, thus being unable to identify local dependencies (associations on subspaces) and guarantee actionable results. Pattern mining shows limitations on handling numeric or non-identically distributed attributes and lacks robustness to noise and missing data. In addition, it is unable to find non-trivial, yet potentially relevant educational patterns with non-constant coherence, i.e., it cannot consider meaningful variations on the values between objects such as coherent variations on grades from students with different academic performance.

To address the aforementioned limitations, this paper proposes the use of biclustering – subsets of objects meaningfully correlated on a subset of attributes – to comprehensively explore educational data. Although biclustering has been largely used in the biomedical field, its full potential in the educational domain remains untapped.

The results presented in this paper confirm the relevance of biclustering to unravel non-trivial yet meaningful, actionable and statistically significant educational patterns. Specifically, we identify patterns of student performance in topics addressed in a course. Such patterns provide a trustworthy context with enough feedback for the teacher to reform the emphasis given to topics addressed in a course. Our proposal can be extended towards curriculum restructuring; personalized support to students, teachers and researchers; among other ends.

The paper is structured as follows. Section 2 provides the background on biclustering and surveys key contributions from related work. Section 3 describes the unique potentialities of biclustering educational data, and places principles on to achieve them. Section 4 presents results that empirically validate our proposal. Finally, Section 5 offers the major concluding remarks.

2 Background

2.1 Biclustering

Definition 1. Given a dataset, $\mathbf{A} = (\mathbf{X}, \mathbf{Y})$, defined by a set of objects $\mathbf{X} = \{\boldsymbol{x}_1, .., \boldsymbol{x}_N\}$, attributes $\mathbf{Y} = \{\boldsymbol{y}_1, .., \boldsymbol{y}_M\}$, and elements $a_{ij} \in \mathbb{R}$ observed in $\boldsymbol{x}_i$ and $\boldsymbol{y}_j$:

- A *bicluster* $\mathbf{B} = (\mathbf{I}, \mathbf{J})$ is a $n \times m$ submatrix of $\mathbf{A}$, where $\mathbf{I} = (\boldsymbol{i_1}, \ldots, i_n) \subset \mathbf{X}$ is a subset of objects and $\mathbf{J} = (\boldsymbol{j_1}, \ldots, \boldsymbol{j_m}) \subset \mathbf{Y}$ is a subset of features;
- The *biclustering task* aims at identifying a set of biclusters $\boldsymbol{\mathcal{B}} = (\mathbf{B_1}, \ldots, \mathbf{B_s})$ such that each bicluster $\mathbf{B_k} = (\mathbf{I_k}, \mathbf{J_k})$ satisfies specific *homogeneity*, *dissimilarity* and *statistical significance* criteria.

Homogeneity criteria are commonly guaranteed through the use of a merit function, such as the variance of the values in a bicluster [16]. Merit functions are typically applied to guide the formation of biclusters in greedy and exhaustive searches. In stochastic approaches, a set of parameters that describe the biclustering solution are learned by optimizing a merit (likelihood) function.

The homogeneity criteria determine the structure, coherency and quality of a biclustering solution. The *structure* of a biclustering solution is defined by the number, size, shape and positioning of biclusters. A flexible structure is characterized by an arbitrary number of (possibly overlapping) biclusters. The *coherence* of a bicluster is determined by the observed form of correlation among its elements (coherence assumption) and by the allowed deviations per element against the perfect correlation (coherence strength). The *quality* of a bicluster is defined by the type and amount of accommodated noise. Defs.2-3 formalize these concepts, and Fig.1 shows biclusters with different coherence assumptions.

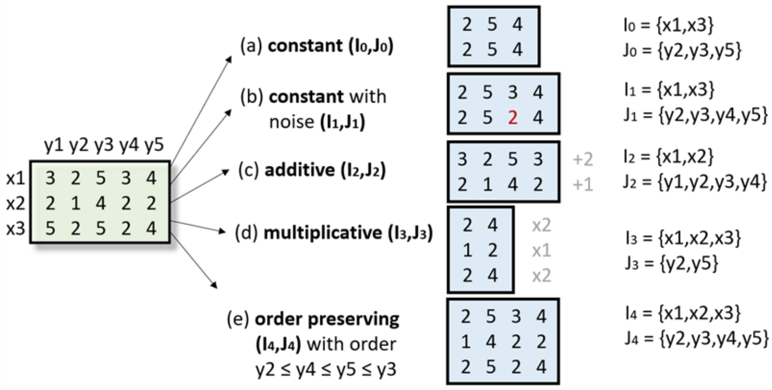

Fig. 1. Discrete biclusters with varying coherence.

Definition 2. Given a numeric dataset **A**, elements in a bicluster $a_{ij} \in (\mathbf{I},\mathbf{J})$ have *coherence* across objects iff $a_{ij} = c_j + \gamma_i + \eta_{ij}$ (or attributes iff $a_{ij} = c_i + \gamma_j + \eta_{ij}$), where c_j (or c_i) is the value of attribute $\boldsymbol{y_j}$ (or object $\boldsymbol{x_i}$), γ_i (or γ_j) is the adjustment for object $\boldsymbol{x_i}$ (or attribute $\boldsymbol{y_j}$), and η_{ij} is the noise factor of a_{ij}.

Let $\bar{\mathbf{A}}$ be the amplitude of values in **A**, *coherence strength* is a value $\delta \in [0, \bar{\mathbf{A}}]$ such that $a_{ij} = c_j + \gamma_i + \eta_{ij}$ where $\eta_{ij} \in [-\delta/2, \delta/2]$.

Given non-iid finite data where $\boldsymbol{y}_j \in \mathcal{Y}_j$, then $a_{ij} = c_j + \eta_{ij}$ where $a_{ij} \in \mathcal{Y}_j$ and $\delta_j \in [0, \bar{\mathcal{Y}}_j]$ for continuous attributes and $\delta_j < |\mathcal{Y}_j|$ for integer attributes.

The γ_i factors define the *coherence assumption*. A bicluster satisfies a *constant* when $\gamma_i = 0$ (or $\gamma_j = 0$), *additive* assumption when $\gamma_i \neq 0$ (or $\gamma_j \neq 0$), and *multiplicative* assumption if a_{ij} is better described by $c_j\gamma_i + \eta_{ij}$ (or $c_i\gamma_j + \eta_{ij}$).

Definition 3. Given a numeric dataset **A**, a bicluster $(\mathbf{I},\mathbf{J})$ satisfies the *order-preserving* assumption iff the values for each object in **I** (attribute in **J**) induce the same linear ordering π along the subset of attributes **J** (objects **I**).

Statistical significance criteria, in addition to homogeneity criteria, guarantees that the probability of a bicluster's occurrence (against a null data model) deviates from expectations. **Dissimilarity** criteria can be further placed to comprehensively cover the search space with non-redundant biclusters.

Following Madeira and Oliveira's taxonomy [16], biclustering algorithms can be categorized according to the pursued homogeneity and type of search. Hundreds of biclustering algorithms were proposed in the last decade, as shown by recent surveys [6,9].

In recent years, a clearer understanding of the synergies between biclustering and pattern mining paved the rise for a new class of algorithms, referred to as pattern-based biclustering algorithms [11]. Pattern-based biclustering algorithms are inherently prepared to efficiently find exhaustive solutions of biclusters and offer the unprecedented possibility to affect their structure, coherency and quality [12,13]. This behavior explains why this class of biclustering algorithms are receiving an increasing attention in recent years [11]. BicPAMS (Biclustering based on PAttern Mining Software) [12] consistently combines such state-of-the-art contributions on pattern-based biclustering.

2.2 Related Work

Despite the diversity of research contributions on unsupervised educational data mining [2,8], real-world decisions are still primarily led by data summarization, visualization and statistics. Such approaches are centered on efforts to test simple hypotheses, facilitate searches and support data navigation, whether data is tabular, event-based, relational, multi-dimensional, or semi-structured [8]. In an attempt to automatize educational data analysis and guarantee a focus on less-trivial data relations, contributions in the fields of clustering and pattern mining have been also proposed [2,7]. In the context of pattern mining, Buldu and Üçgün [4], Chandra and Nandhini [5], Gottin et al. [10], and Olaniyi et al. [17] pursued association rules pertaining to student performance and topic agreement to support curriculum redesign. Sequential pattern mining has been alternatively applied for topic data analysis to model students' behaviors along an educational programme [1,3]. Results suggest that sequential patterns can be used to enrich training data for improving predictions of students' performance.

Biclustering has been firstly suggested for educational data exploration by Trivedi et al. [18,19] to understand the impact of tutor interaction in students' performance. To this end, the authors partitioned students and interaction features to produce biclusters for predicting out-of-tutor performance of students. Results show a moderately reduced error (against baseline predictors). Despite its merits, the applied algorithm imposes biclusters to follow a checkboard structure, a severe restriction, which possibly explains the modest results.

Vale et al. [20] offered a comprehensive roadmap on the relevance of biclustering for two distinct sources of educational data: 1) matrices relating students and subjects through achieved marks, where the interest is placed on students showing coherent grades in a particular subset of subjects, and 2) matrices collecting performance indicators of subjects along time with the aim of finding temporal patterns. The goal of the work was to find biclusters not trivially retrieved using alternative pattern mining methods. To this end, xMOTIFs, ISA and OPSM biclustering algorithms are considered. Despite its relevance, the obtained patterns are approximate and not statistically tested.

3 Solution: biclustering in educational data

As surveyed in previous section, pattern-based biclustering approaches provide the unprecedented possibility to comprehensively find patterns in non-iid data with parameterizable homogeneity and guarantees of statistical significance. Despite their relevance, their use to explore educational data remains unassessed. This section provides a structured view on how to bicluster educational data, identifying its unique potentialities.

Real-valued educational patterns. Biclustering seeks patterns in real-valued data with *coherence orientation* along objects or attributes (Def.2). Illustrating, in student performance analysis, biclusters with patterns on objects reveal coherent grades on specific topics for a subset of students.

Biclustering also allows the calibration of *coherence strength* (Def.2) – e.g. how much two academic indicators need to differ to be considered dissimilar. Allowing deviations from pattern expectations in real-valued educational data is key to prevent the item-boundaries problem, thus tackling discretization problems faced by classic pattern mining methods. Patterns are inferred from similar (yet non-strictly identical) performance indicators, whether numerical or ordinal.

Comprehensive educational data exploration. Pattern-based biclustering offers principles to find complete solutions of educational patterns by: 1) pursuing multiple homogeneity criteria, including multiple coherence strength thresholds, coherence assumptions and quality thresholds, and 2) exhaustively yet efficiently exploring different regions of the search space, preventing that regions with large patterns jeopardize the search. As a result, less-trivially correlated indicators of academic performance are not neglected. By contrast, classic pattern mining procedures uniquely focus on educational patterns with constant coherence and the underlying searches have efficiency bottlenecks in the presence of lengthy patterns. Furthermore, pattern-based biclustering does not require the input of support thresholds as it explores the search space at different supports, i.e. we do need to place expectations on the minimum number of students/teachers/researchers per pattern. Still, the minimum number of (dissimilar) patterns, minimum percentage of covered data elements, and minimum number of objects and/or performance indicators in a bicluster can be optionally inputted to guide the search. Parameterizable dissimilarity criteria and condensed representations can be placed [12] to prevent redundant educational patterns.

Non-constant educational patterns. Depending on the goal, one or more *coherence assumptions* (Def.2-3) can be pursued. Let us illustrate paradigmatic cases in student performance analysis. The classic constant assumption can be placed to unravel groups of students with similar grades on a subset of topics/courses. However, it is unable to correlate grades from students with different performance profiles. In this context, non-constant patterns can be pursued:

- *additive* pattern: set of students with different average of performance yet coherent grades on a subset of topics explained by shifting factors (Fig.1c);
- *multiplicative* pattern: set of students with linearly correlated grades on a subset of topics/courses explained by scaling factors (Fig.1d);
- *order-preserving* pattern: set of students with preserved orderings of grades on a subset of topics/courses (Fig.1e).

As a result, pattern-based biclustering allows the discovery of less-trivial yet coherent, meaningful and potentially relevant educational relations.

Robustness to noise and missing values. With pattern-based biclustering, and by co trast with classic pattern mining, the user can find biclusters with a parameterizable tolerance to noise. This possibility ensures, for instance, robustness to the inherent subjectivity of Likert scale evaluations in questionnaires.

Similarly, pattern-based biclustering is robust to missing data by permitting the dis covery of biclusters with an upper bound on the allowed amount of missings. This is particularly relevant to handle missing ranks in questionnaire data or missing grades due to unassessed topics or unattended exams.

In turn, this ability to handle missing data allows the discovery of coherent modules (biclusters) in network data (sparse adjacency data) such as student community data or research collaboration data.

Statistical significance. A sound statistical testing of educational patterns is key to guarantee the absence of spurious relations, validate conclusions inferred from educational patterns, and ensure pattern relevance when building academic reforms and making other decisions. To this end, the statistical tests proposed in BSig [15] are suggested to minimize the number of false positives (output patterns yet not statistically significant) without incurring on false negatives. This is done by approximating a null model of the target educational data and appropriately testing each bicluster in accordance with its underlying coherence.

Other opportunities. Additional benefits of pattern-based biclustering can be carried towards educational data analysis, including: **1)** the removal of uninformative elements in data to guarantee a focus, for instance, on lower student grades or assessments of faculty members suggesting problematic performance; **2)** incorporation of domain knowledge to guide the biclustering task, useful in the presence of background data on courses, students or faculty members [14]; and **3)** support to classification and regression problems in education in the presence of annotations by guaranteeing the discriminative power of biclusters [11].

4 Results on student performance data

To illustrate the enumerated potentialities of biclustering educational data, we discuss results from student-topic performance data in four major steps. First, we empirically delineate general advantages of biclustering student-topic data. Second, we show that biclustering finds educational patterns robust to noise and missings. Third, we provide evidence for the relevance of finding non-trivial (yet meaningful) educational patterns with non-constant coherence. Finally, we show that biclustering guarantees the statistical significance of relations, providing a trustworthy means for academic reforms.

ADS dataset The ADS dataset[1] captures the performance of students along the topics of the Advanced Data Structures (ADS) course offered every academic term by the Department of Informatics of the Pontifical Catholic University of Rio de Janeiro (PUC-Rio). The dataset combines the results of exams, covering 10 academic terms in which the course was under the responsibility of the same teacher, amounting a total of 229 students and 325 enrollments.

Experimental setting The BicPAMS algorithm[1] [12] is applied since it consistently integrates the state-of-the-art algorithms on pattern-based biclustering and guarantees the efficiency of the underlying searches. BicPAMS is below used with default parameters: varying coherence strength ($\delta = \overline{\mathrm{A}}/|\mathcal{L}|$ where ($|\mathcal{L}| \in \{3,4,5\}$), decreasing support until at least 50 dissimilar biclusters are found, up to 30% noisy elements, 0.05 significance level, and a single coherence assumption at a time (constant, additive, multiplicative and order-preserving). Two search iterations were considered by masking the biclusters discovered after the first iteration to ensure a more comprehensive exploration of the data space and a focus on less-trivial educational patterns. Topic-based frequency distributions were approximated, and the statistical tests proposed in [15] were applied to compute the statistical significance of each found bicluster.

[1] BicPAMS available at https://web.ist.utl.pt/rmch/bicpams/. ADS data available upon request.

4.1 Real-valued educational patterns

Table 1 synthesizes the results produced by biclustering student-topic data with BicPAMS [12]. Confirming the potentialities listed in Section 3, BicPAMS was able to efficiently and comprehensively find a large number of homogeneous, dissimilar and statistically significant biclusters – subsets of students with coherent performance on a subset of topics. One can check, for instance, in the first row of Table 1, that among the total number of discovered biclusters (135), we found that 120 of them are statistically significant with a *p*-value lower that 0.1%. Given these 135 biclusters, there are approximately $u(|\mathbf{I}_1|,\dots,|\mathbf{I}_{135}|)$=16 students per bicluster on average and $u(|\mathbf{J}_1|,\dots,|\mathbf{J}_{135}|)$=3 topics per bicluster on average considering a constant assumption, three bins ($|\mathcal{L}|$=3 and $\delta=\bar{\mathbf{A}}/|\mathcal{L}|$), and a perfect quality (100% / no noise).

Assumption	$\|L\|$	quality	#bics	$\mu(\|I\|)$ $\pm\sigma(\|I\|)$	$\mu(\|J\|)$ $\pm\sigma(\|J\|)$	p-value >0.01	p-value $\in$ [0.1,1E-3]	p−value <1E-3
Constant	3	100%	135	15.6±5.8	2.9±0.4	10	15	120
Constant	3	70%	123	20.3±5.1	3.1±0.3	2	10	111
Constant	4	70%	168	15.1±4.8	3.0±0.1	10	26	132
Constant	5	70%	241	10.8±3.7	3.1±0.2	9	65	167
Additive	4	70%	310	15.3±8.2	3.1±0.4	17	23	270
Multiplicative	4	70%	195	14.3±5.3	3.1±0.4	9	13	173
Order-preserving	–	70%	91	27.4±4.4	3.4±0.5	11	20	60

Table 1: Properties of the biclustering solutions found in ADS data with BicPAMS when varying the homogeneity criteria.

These initial results further show the impact of tolerating noise by placing different coherence assumptions (such as the order-preserving assumption), and parameterizing coherence strength ($\delta \propto 1/|\mathcal{L}|$) on the biclustering solution.

4.2 Constant educational patterns

Table 2 provides the details of an illustrative set of four constant biclusters (and the respective performance pattern, subset of topics, coherence strength and statistical significance) using BicPAMS with default parameters. Each bicluster shows a unique pattern of performance. For instance, bicluster $\mathbf{B}_5$ reveals a group of 13 students who coherently encountered moderate, delineate and no difficulties (corresponding to the pattern {1,0,4} using 5 bins where 0 denotes a low grade and 4 an excelling grade) in 3 topics (binary searches, bit-vectors and complexity).

Fig.2a visually depicts an additional constant bicluster. Each line in the chart represents a student and her/his grades on along the 3 topics in the bicluster.

These results motivate the relevance of finding constant biclusters to understand coherent patterns of difficulty between topics for a statistically significant population of students. One can check that a bicluster considers both identical grades (where lines converge) and more loosely similar values (where lines diverge). The profile of the students in a specific bicluster can be further analyzed to further understand its influence on the resulting performance.

A closer analysis of the found biclusters shows their robustness to the item-boundaries problem: students with slightly deviating grades from pattern expectations are not excluded from the bicluster. This allows the analysis of real-valued or/and integer data without the drawbacks of aggregated/discrete views on students' performance.

Bicluster properties	Pattern (studentID⇒values)
B$_1$ topics=[Bitvector,Heap, Dijkstra,SocialNet] #students=16, $\|L\|$=3 p-value=**1.06E-18**	x_{14} 0 0 0 0 x_{21} 0 0 0 0 ... x_{267} 0 1 0 0 x_{317} 0 0 0 0
B$_2$ topics =[B-TreesRemoval, Bitvector,Complexity] #students=16, $\|L\|$=4 p-value=**8.7E-4**	x_{11} 0 0 3 x_{55} 0 0 3 ... x_{269} 0 0 3 x_{317} 0 0 3
B$_3$ topics=[B-TreesRemoval, Heap,Morton] #students=14, $\|L\|$=4 p-value=**1.6E-8**	x_{15} 0 0 1 x_{23} 0 0 1 ... x_{280} 0 0 1 x_{320} 0 0 1
B$_4$ topics=[Hash,Union-Find, Complexity] #students=15, $\|L\|$=4 p-value=**1.3E-6**	x_{11} 1 3 3 x_{16} 1 3 3 ... x_{292} 1 3 3 x_{321} 1 3 3
B$_5$ topics=[BinarySearchT, Bitvector,Complexity] #students=13, $\|L\|$=5 p-value=**6E-7**	x_{11} 1 0 4 x_{46} 1 0 4 ... x_{292} 1 0 4 x_{299} 1 0 4

Table 2: Constant biclusters found in ADS data.

Bicluster properties	Pattern (studentID⇒values)
B$_6$ topics=[B-TreesRemoval, BinarySearchTrees,Heap] Order of difficulty: Heap<BinarySearchT< <B-TreesRemoval #students=35	x_{42} 9 7 0 x_{103} 7 4 3 x_{181} 5 3 2 x_{218} 4 3 2 x_{247} 7 6 4 x_{256} 7 4 1 x_{260} 9 8 4 x_{265} 9 3 2 x_{281} 9 8 4 x_{286} 7 6 2 x_{292} 8 3 2 x_{307} 9 8 7 x_{308} 9 3 0 ...
B$_7$ topics=[B-TreesInsertion, Bitvector,Hash,Heap] Order of difficulty: Heap<B-TreesInsertion= =Bitvector<Hash #students=17	x_{12} 8 8 3 9 x_{30} 8 8 0 9 x_{32} 1 1 0 3 x_{94} 8 8 5 9 x_{107} 1 1 0 2 x_{151} 5 5 4 8 x_{267} 1 1 0 4 x_{268} 1 1 0 9 x_{269} 1 1 0 7 x_{289} 8 8 0 9 x_{310} 1 1 0 9 x_{313} 8 8 7 9 ...

Table 3: Order-preserving biclusters in ADS data.

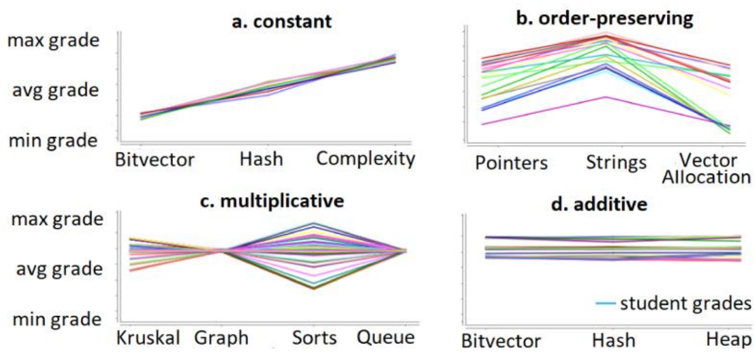

Fig. 2: Set of (a) constant, (b) order-preserving, (c) multiplicative, and (d) additive biclusters found in ADS data (subsets of students with coherent grades on subsets of topics in the absence and presence of ordering, scaling and shifting factors).

4.3 Non-constant patterns

Non-constant patterns are suggested if the focus is not on determining levels of performance but to assess the relative difficulty among topics. BicPAMS [12] was applied to find such less-trivial yet relevant topic-student patterns, including patterns with order-preserving, additive, and multiplicative coherence assumptions (Table 1).

Table 3 provides the details of two statistically significant order-preserving biclusters, including the subset of students and topics, and the permutation of topic grades (the pattern). For instance, bicluster $\mathbf{B_6}$ reveals an unexpectedly large group of students with arbitrarily-different grades yet coherently facing more difficulties in heaps, then binary searches and, finally, B-tree removals.

Fig.2 depicts 3 additional biclusters with order-preserving (2b), multiplicative (2c) and additive (2d) coherence. These coherence assumptions are useful to accommodate coherent orders, shifts and scales in student performance, thus being able to account for differences in students' aptitude.

4.4 Noise-missing robustness

Tolerance to noise can be customized (see Table 1) in order to comprehensively find patterns with parameterizable degree of quality. In addition to noise-tolerance, η_{ij}, coherence strength $\delta = \bar{\mathbf{A}}/|\mathcal{L}|$ can be explored (Table 1) to comprehensively model relations between students and topics with slight-to-moderate deviations from expectations.

The analysis of the found biclusters further confirms their ability to tolerate missing educational data. ADS data have two major types of missing grades caused by: 1) students not showing up to an exam (not evaluated), and 2) not all topics being covered in the context of an exam applied in a given semester.

4.5 Statistical Significance

Table 1 shows the biclustering ability to find statistically significant relations in student-topic data. A bicluster is statistically significant if the number of students with a given pattern or permutation of topic grades is unexpectedly low [15. Fig.3 provides a scatter plot of the statistical significance (horizontal axis) and area $|\mathrm{I}| \times |\mathrm{J}|$ (vertical axis) of constant biclusters with $|\mathcal{L}|$=3 and >70% quality. This analysis suggests the presence of a soft correlation between size and statistical significance. We observe that few biclusters have low statistical significance (right bottom dots) and therefore should be discarded to not incorrectly bias decisions in educational contexts.

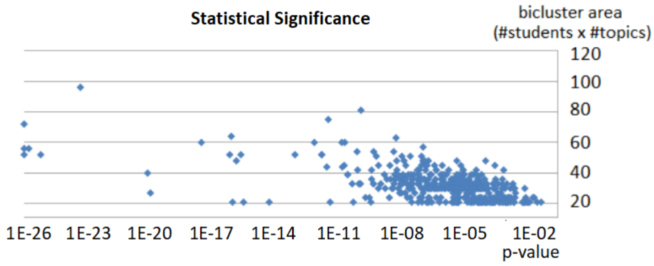

Fig. 3: Statistical significance vs. size of collected constant biclusters ($|\mathcal{L}|$=3).

5 Conclusions

This work proposed comprehensive principles on how to apply biclustering for the exploration of educational data in order to tackle the limitations of peer unsupervised tasks, such as clustering and pattern mining, and untap the hidden potential underlying educational data by focusing on non-trivial, yet meaningful and statistically significant relations. Pattern-based biclustering searches are suggested to find actionable educational patterns since they hold unique advantages: efficient exploration; optimality guarantees; discovery of non-constant patterns with parameterizable coherence; tolerance to noise and missing data; incorporation of domain knowledge; complete biclustering structures without positioning restrictions; and sound statistical testing.

Results from student-topic data confirm the unique role of biclustering in finding relevant patterns of student performance, such as similar topic difficulties experienced by students with a specific profile (given by constant or additive biclusters) and orders of topic difficulties (given by order-preserving biclusters).

Results further evidence the ability to unveil interpretable patterns with guarantees of statistical significance and robustness, thus providing a trustworthy context with enough feedback for the teacher to reform the emphasis given to topics addressed in a course. A similar analysis can be conducted in alternative educational data domains, including monitored lecturing and research activities.

Acknowledgement. This work is supported by national funds through FCT under project iLU DSAIPA/DS/0111/2018 and INESC-ID pluriannual UID/CEC/50021/2019.

References

1. Antunes, C.: Acquiring background knowledge for intelligent tutoring systems. In: EDM (2008)
2. Baker, R.S., Inventado, P.S.: Educational data mining and learning analytics. In: Learning analytics, pp. 61–75. Springer (2014)
3. Barracosa, J., Antunes, C.: Anticipating teachers performance. In: KDD IW on Knowledge Discovery in Educational Data. pp. 77–82 (2011)
4. Buldu, A., Üçgün, K.: Data mining application on students data. Procedia-Social and Behavioral Sciences **2**(2):5251–5259 (2010)
5. Chandra, E., Nandhini, K.: Knowledge mining from student data. European journal of scientific research **47**(1), 156–163 (2010)
6. Charrad, M., Ahmed, M.B.: Simultaneous clustering: A survey. In: Pattern Recognition and Machine Intelligence, pp. 370–375. Springer (2011)
7. Dutt, A., Aghabozrgi, S., Ismail, M.B., Mahroeian, H.: Clustering algorithms applied in educational data mining. IJ Info. and Electronics Eng. **5**(2), 112 (2015)
8. Dutt, A., Ismail, M.A., Herawan, T.: A systematic review on educational data mining. IEEE Access **5**, 15991–16005 (2017)
9. Eren, K., Deveci, Küçüktunç, Çatalyürek, Ü.: A comparative analysis of biclustering algorithms for gene expression data. Brief. Bioinf., **14**(3):279–292 (2013).
10. Gottin, V., Jiménez, H., Finamore, A.C., Casanova, M.A., Furtado, A.L., Nunes, B.P.: An analysis of degree curricula through mining student records. In ICALT, pages 276–280. IEEE, 2017.
11. Henriques, R., Antunes, C., Madeira, S.C.: A structured view on pattern mining-based biclustering. Pattern Recognition **48**(12):3941-3958. Elsevier (2015)
12. Henriques, R., Ferreira, F.L., Madeira, S.C.: Bicpams: software for biological data analysis with pattern-based biclustering. BMC bioinformatics **18**(1):82 (2017)
13. Henriques, R., Madeira, S.C.: Bicpam: Pattern-based biclustering for biomedical data analysis. Alg. for Molecular Biology **9**(1):27 (2014)
14. Henriques, R., Madeira, S.C.: Bic2pam: constraint-guided biclustering for biological data analysis with domain knowledge. Algo. for Molecular Biology **11**(1):23 (2016)
15. Henriques, R., Madeira, S.C.: Bsig: evaluating the statistical significance of biclustering solutions. Data Mining and Knowledge Discovery **32**(1):124–161 (2018)
16. Madeira, S.C., Oliveira, A.L.: Biclustering algorithms for biological data analysis: A survey. IEEE/ACM Trans. Comput. Biol. Bioinformatics **1**(1):24–45 (2004)
17. Olaniyi, A.S., Abiola, Taofeekat Tosin, Kayode, Babatunde, A.N.: Knowledge discovery from educational database using apriori algorithm. CS&Telec., **51**(1) (2017)
18. 19. Trivedi, S., Pardos, Z., Sárkozy, G., Heffernan, N.: Co-clustering by bipartite spectral graph partitioning for out-of-tutor prediction. Int. EDM Society (2012)
19. Trivedi, S., Pardos, Z., Sárkozy, G., Heffernan, N.: Spectral clustering in educational data mining. In EDM (2010)
20. Vale, A., Madeira, S.C., Antunes, C.: Mining coherent evolution patterns in education through biclustering. In Educational Data Mining (2014)