



ELSEVIER

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

A structured view on pattern mining-based biclustering

Rui Henriques^{a,b,*}, Cláudia Antunes^a, Sara C. Madeira^{a,b}^a CSE Department, Instituto Superior Técnico, Universidade de Lisboa, Portugal^b INESC-ID, Lisbon, Portugal

ARTICLE INFO

Article history:

Received 1 May 2014

Received in revised form

12 May 2015

Accepted 26 June 2015

Keywords:

Biclustering

Pattern mining

ABSTRACT

Mining matrices to find relevant biclusters, subsets of rows exhibiting a coherent pattern over a subset of columns, is a critical task for a wide-set of biomedical and social applications. Since biclustering is a challenging combinatorial optimization task, existing approaches place restrictions on the allowed structure, coherence and quality of biclusters. Biclustering approaches relying on pattern mining (PM) allow an exhaustive yet efficient space exploration together with the possibility to discover flexible structures of biclusters with parameterizable coherency and noise-tolerance. Still, state-of-the-art contributions are dispersed and the potential of their integration remains unclear.

This work proposes a structured and integrated view of the contributions of state-of-the-art PM-based biclustering approaches, makes available a set of principles for a guided definition of new PM-based biclustering approaches, and discusses their relevance for applications in pattern recognition. Empirical evidence shows that these principles guarantee the robustness, efficiency and flexibility of PM-based biclustering.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

The clustering of data matrices groups rows according to their overall values across columns. However, in real-world contexts, the correlation of a subset of rows is typically only significant and meaningful for a subset of the overall columns [114]. Biclustering seeks to find sub-matrices (biclusters), subsets of rows with a coherent pattern across subsets of columns. Illustrating, given a matrix that captures the expression of a set of genes (rows) across a set of conditions (columns), a bicluster defines a group of genes with coherent expression for a subset of conditions. The biclustering task in this domain is critical for the discovery of putative transcriptional modules of genes that participate in a cellular process that is only active in specific conditions [46,40]. Table 1 provides additional applications in biomedical and social domains, synthesizing the meaning and relevance of discovering biclusters for pattern recognition.

Recent findings from biomedical domains show that exhaustive and flexible approaches to biclustering provide an unprecedented opportunity for an unbiased assessment of the native structure and modular organization of biological networks [14], new insights on the molecular units involved in cellular functions [60,111], and discriminative high-

order combinations of single-nucleotide polymorphisms (SNPs) [42]. However, due to the complexity of the biclustering task¹, most of the existing algorithms are either based on greedy or stochastic approaches, potentially producing sub-optimal and constrained biclustering solutions [82,67]. Illustrative constraints that prevent the flexibility of the biclustering task include the search for a fixed number of biclusters, non-overlapping structures and biclusters with differential-values only (binary settings) or sequential constraints [81,82,119]. In this context, the survey of efficient optimal searches for flexible biclustering scenarios is the target task in this work.

The attempts to perform biclustering based on pattern mining (PM) techniques [86,111,97], referred in this work as PM-based biclustering, show solid results for efficient and flexible exhaustive searches. In fact, since pattern mining research is driven by scalability requirements [54], its integration with biclustering defines a new promising direction. Contributions of PM-based approaches for biclustering include:

- efficient exhaustive searches: PM algorithms as-is allow for the efficient analysis of large matrices (over 10.000×400)

* Correspondence to: DEI, IST, Avenida Rovisco Pais, 1, 1049-001 Lisboa, Portugal. Tel.: +351 21 310 0 300; fax: +351 21 841 7 789.

E-mail addresses: rmch@tecnico.ulisboa.pt (R. Henriques), claudia.antunes@tecnico.ulisboa.pt (C. Antunes), sara.madeira@tecnico.ulisboa.pt (S.C. Madeira).

<http://dx.doi.org/10.1016/j.patcog.2015.06.018>

0031-3203/© 2015 Elsevier Ltd. All rights reserved.

¹ Biclustering involves combinatorial optimization to select and group rows and columns and it is known to be a NP problem (by mapping the task over binary matrices into the problem of finding maximal cliques in weighted bipartite graphs [104]). The problem complexity increases for non-binary settings and when elements are allowed to participate in more than one bicluster (non-exclusive structure) and in no bicluster at all (non-exhaustive structure).

Table 1
Relevance of the biclustering task for pattern recognition applications.

	Data	Biclustering solutions
Biomedical	Physiological [23,39,44]	Modules of sliding features and partitions of the signal across a subset of case or stimuli-elicited responses; groups of patients with shared local patterns; markers for phenotype characterization.
	Clinical [59,27]	Groups of patients with correlated clinical features or health records (shared treatments, diagnoses, prescriptions and clinical tests); class-conditional profiles for computer-aided diagnosis.
	Genomic structural variations [42,67]	Correlated groups of mutations and copy number variations, such as genetic similarities and dissimilarities of different populations.
	Biological networks [12]	Modules of genes, proteins or metabolites with cohesive local interaction using matrices that capture the pairwise connections between all molecular units.
	Gene expression [62,82]	Groups of genes involved in functional processes and pathways (cellular responses to growth, development, drugs and disease progression) only active under certain conditions.
	Genome-wide [127,124] Other [37,78,73]	Conserved functional subsequences (alignments), factor binding sites and insertion mutagenesis. Local regularities in translational, chemical or nutritional data;
Social	Social networks [50]	Groups of individuals with shared interests, correlated activity and/or coherent intercommunication; aggregation of contents based on correlated accessors' profile, comments and tags.
	Text [82]	Groups of content-related documents to support searches, suggestions and tagging (rows in the input matrix denote documents and columns denote the words), among others.
	(e-)commerce [9]	Hidden browsing patterns containing relationships between sets of (web) users and (web) pages and acquisitions which are useful for (web) advertising and marketing.
	Financial trading [68]	Subsets of indicators producing similar profitability for subsets of trading points (buy and sell signals) in the stock market in order to support buy-and-hold decisions.
	Collaborative filtering [33]	Groups of users who share the same rating patterns and behavioral patterns for a subset of all available actions for recommendation and quality studies.

elements). Additional PM principles can be used to foster scalability, including searches in distributed/partitioned data settings or targeting approximate patterns [54,52].

- dealing with missing and noisy values [62,63]: PM methods can mine transactions with varying length, and therefore a specific element from the input matrix can be associated with zero or multiple values, allowing the removal or bounded estimations of a missing or noisy value.
- inherent orientation to learn constant models, yet recently extended to also learn additive, multiplicative, symmetric, order-preserving and plaid models [62,60,63];
- capturing biclusters from patterns with multiple levels of expression [96,101]. This contrasts with the majority of existing approaches that rely on differential values or fixed coherency strength [119];
- flexible structures of biclusters (arbitrary positioning of biclusters) and searches (no need to fix the number of biclusters a priori) [96,111];
- annotating the significance of biclusters with PM principles to assess the relevance of patterns [72];
- easy extension for multi-class settings using discriminative PM or classification rules [43,95];
- easy incorporation of PM-based constraints that can be effectively used to guide the search, promoting both efficiency, by pruning the search space, and a focus on non-trivial biclusters [116].

These properties of PM-based biclustering approaches are critical to tackle the problems highlighted in Table 1. Although the latest biclustering advances for pattern recognition are increasingly deterministic [89,110,128,137,47,35,131], they fail to meet several of the enumerated properties of PM-based biclustering. Table 2 pinpoints the benefits of using PM-based biclustering for pattern recognition.

Despite these listed potentialities, recent surveys on biclustering [46,40,28,114] fail to explore the opportunities associated with PM-based biclustering. Additionally, the existing efforts towards PM-based biclustering provide critical principles that are not yet integrated [14,86,111]. As such, there is still space for new approaches that benefit from the integration of principles provided by these existing contributions as well as from other fields of

research. In this context, this work provides three major contributions:

- motivates, formalizes and provides a qualitative and quantitative assessment of the state-of-the-art algorithms for PM-based biclustering;
- offers a structured view on how to define, parameterize and extend PM-based biclustering by coherently integrating the available yet dispersed contributions;
- further surveys PM principles as well as adequate preprocessing and postprocessing criteria to guarantee the robustness, flexibility and scalability of PM-based biclustering across domains.

The paper is organized as follows. The remainder of this section provides background on pattern mining and biclustering, and surveys the contributions from existing PM-based biclustering approaches. Section 2 introduces a consistent set of principles to guide the definition of PM-based biclustering approaches. In particular, Sections 2.1–2.3 cover principles according to three major decision dimensions (mining, mapping and closing), and Section 2.4 compares the behavior of state-of-the-art PM-based biclustering approaches and proposes a set of principles to address their current challenges. Section 3 provides initial empirical evidence of the relevance of the proposed principles. Finally, the implications of this work are synthesized.

1.1. Background on PM-based biclustering

Pattern mining: Frequent patterns are itemsets, rules, subsequences, or substructures that appear in a dataset with frequency no less than a user-specified threshold. Let \mathcal{L} be a finite set of items, and P be an itemset $P \subseteq \mathcal{L}$. A transaction t is a pair (t_{id}, P) with $id \in \mathbb{N}$. An itemset database D over \mathcal{L} is a finite set of transactions $\{t_1, \dots, t_n\}$. A transaction (id, P) contains P' , denoted $P' \subseteq (t_{id}, P)$, if $P' \subseteq P$. The coverage Φ_P of an itemset P is the set of all transactions in D in which the itemset P occurs: $\Phi_P = \{t \in D | P \subseteq t\}$. The support of an itemset P in D , denoted sup_P , can either be absolute, being its coverage size $|\Phi_P|$, or a relative threshold given by $|\Phi_P| / |D|$.

An association rule is defined as an implication of the form $P \rightarrow P'$, where $P, P' \subseteq \mathcal{L}$ and $P \cap P' = \emptyset$. The left-hand side of the rule

Table 2
Benefits of PM-based biclustering for pattern recognition.

Property	Benefit
Exhaustive scalable searches	Delivery of optimality guarantees for large data such as data from clinical, molecular and social web domains.
Noise robustness	Handling of uncertainty relations observed in social networks [50] and stock markets [68]; artefacts in multivariate physiological data (such as electroencephalograms [41]), experimental errors in molecular arrays [56].
Handling of missing values	Adequate mining of incomplete and/or sparse matrices derived from biological networks, web social contexts, and healthcare data.
Flexible coherency	Constant models for non-differential (yet coherent) functional associations; additive and multiplicative factors to model the distinct responsiveness and experimental bias of biological molecules and physiological signals; symmetries to simultaneously capture activation and repression mechanisms and opposed (yet correlated) regularities associated with trading, tweeting, browsing and (e-)commerce activity; plaid models for overlapping regulatory influence in biological contexts and cumulative effects in social/biological networks [60,62,61].
Parameterizable level of coherency	Dynamic definition of the desirable coherency strength for an adequate multi-level analysis of matrices derived from expression data (optimum number of expression levels [86]), scored networks, collaborative filtering data (grading scale), and physiological signals (adequate resolution [39]).
Flexible structures	Overlapping groups of molecular units, physiological features, patients, web users and transactions with varying size and configurations.
Annotated significance	Testing the statistical significance of biclustering solutions (guaranteeing that their coherence does not occur by chance) to further validate their use to support critical decisions, such as medical and financial decisions.
Constraint-driven searches	Discovery of non-trivial biclusters and ability to focus the search on specific biclusters of interest (e.g. specific regulatory behavior, high-order SNPs from genome-wide data, web users with a specific behavior, health records related with particular medical conditions, domain-guidance from background knowledge [42,66]).
Biclustering-based classification	Support for classification tasks from matrices with a large number of uninformative elements (benefiting from local views), including computer-aided diagnosis, phenotype discrimination and user recommendations [27,43].

is named antecedent and the right-hand side consequent. Given an itemset database D , the *support* of a rule, $sup_{P \rightarrow P'}$, is given by $sup(P \cup P')$, and the *confidence* of a rule, $conf_{P \rightarrow P'}$, is given by $\frac{sup(P \cup P')}{sup(P)}$. Confidence reveals the strength of the rule (the conditional probability that a transaction that contains the items in the antecedent also contains the items in consequent).

Definition 1.1. Given an itemset database D and a minimum support and confidence thresholds, θ and δ :

- frequent itemset mining (FIM) problem consists of computing the set $\{P | P \subseteq \mathcal{L}, sup_P \geq \theta\}$;
- association rule mining aims to compute $\{(P, P') | P \subseteq \mathcal{L}, P' \subseteq \mathcal{L}, sup_{P \rightarrow P'} \geq \theta, conf_{P \rightarrow P'} \geq \delta\}$.

A *frequent itemset* or a pattern is an itemset with $sup_P \geq \theta$. To illustrate these concepts, consider the following itemset database, $D_{ex} = \{(t_1, \{B, E, G\}), (t_2, \{A, B, C, E, H, J\}), (t_3, \{A, B, D, H, J\}), (t_4, \{D, H, J\}), (t_5, \{A, H, J\}), (t_6, \{A, G\})\}$, with $|\mathcal{L}|=12$. We have $\Phi_{\{B,J\}} = \{t_2, t_3\}$ and $sup_{\{B,J\}} = |\{t_2, t_3\}|/6 = 0.3$. An illustrative rule in D_{ex} is $R_1 : \{H, J\} \rightarrow \{A\}$ with $sup_{R_1} = 0.5$ and $conf_{R_1} = 0.75$. For $\theta=4$, the FIM tasks returns $\{\{A\}, \{H\}, \{J\}, \{H, J\}\}$.

Consider two itemsets P and P' , where $P' \subseteq P$, and a predicate M . M is *monotonic* when $M(P) \Rightarrow M(P')$ and *anti-monotonic* when $\neg M(P') \Rightarrow \neg M(P)$. FIM approaches rely on these properties: the support of P is bounded by the support of P' and, if P' is not frequent, then P is also not frequent. Table 3 shows three major search variants that rely on these properties.

Since FIM proposal [2], multiple extensions have been proposed, including principles to enhance the scalability of pattern miners, and condensed and approximate pattern representations [24,54].

Pattern mining has been additionally applied over structured datasets, leading to contributions in different fields, including sequential pattern mining [79], graph mining [129] and cube computation [55].

Biclustering: Biclustering allows the discovery of subspaces, each defining a subset of rows that show a coherent pattern that is observed for a subset of the overall columns.

Definition 1.2. Given a matrix, $A=(X,Y)$, with a set of rows $X=\{x_1, \dots, x_n\}$, a set of columns $Y=\{y_1, \dots, y_m\}$, and elements $a_{ij} \in \mathbb{R}$ relating row i and column j :

- A *bicluster* $B=(I,J)$ is a $r \times s$ submatrix of A , where $I=(i_1, \dots, i_r) \subset X$ is a subset of rows and $J=(j_1, \dots, j_s) \subset Y$ is a subset of columns;
- The *biclustering task* is to identify a structure of biclusters $B=\{B_1, \dots, B_p\}$ such that each bicluster $B_k=(I_k, J_k)$ satisfies specific criteria of *homogeneity* and *significance*.

The *homogeneity* criteria is commonly guaranteed through the use of a merit function to guide the search [98]. An illustrative merit function is the variance of values in the rows or columns in the bicluster. Merit functions can either define the homogeneity of each bicluster (intra-bicluster homogeneity) or the homogeneity of a set of biclusters (inter-bicluster homogeneity), allowing some biclusters to deviate from the expected homogeneity as long as the overall criterion is preserved. The merit function is the simplest way to affect the coherency, quality and structure. The *coherency* of a bicluster is defined by the observed correlation of values (Definition 1.3). Biclusters can follow dense, constant, additive, multiplicative, plaid or order-preserving coherencies, either across rows or columns [82]. The *quality* of a bicluster is defined by the type and amount of accommodated noise. The *structure* is defined by the number,² size and positioning of biclusters. Flexible structures are characterized by an arbitrary-high set of (possibly overlapping) biclusters. The statistical *significance* of a bicluster determines how its probability of occurrence deviates from expectations. Following the taxonomy proposed by Madeira and Oliveira [82], Table 4 synthesizes the main biclustering approaches according to their search paradigm.

Definition 1.3. Let the elements in a bicluster $a_{ij} \in (I,J)$ have *coherency across rows* given by $a_{ij} = k_j + \gamma_i + \eta_{ij}$, where k_j is the expected value for column j , γ_i is the adjustment for row i , and η_{ij} is the noise factor. Given a dataset A and a specific coherency strength $\delta \in [0, \max_A - \min_A]$, $a_{ij} = k_j + \gamma_i + \eta_{ij}$ where $\eta_{ij} \in [-\delta/2, +\delta/2]$. The γ factors define the coherency assumption: *constant* when $\gamma=0$, *multiplicative* if a_{ij} is better described by $k_j \gamma_i + \eta_{ij}$, and *additive* otherwise. A *plaid* assumption considers the cumulative contributions from multiple biclusters on areas where their rows and columns overlap.

² The number of outputted biclusters can either be fixed (restrictive setting), parameterized by the user [29,67], dynamically parameterized based on the size and stochastic properties of the input matrix [63], or variable [119].

Table 3
Three major search strategies to perform frequent itemset mining.

Strategy	Principles	Optimizations [54,22,76]	Criticism
Apriori-based [2]	Monotonicity principle (an itemset is candidate if all its subsets are frequent): $(k-1)$ -itemsets are combined to create new candidate k -itemsets in k scans until no new candidate group can be generated.	Incremental mining; Hashing; Use of bit-sets; Reduced scans; Partitioning and sampling; Dynamic itemset counting;	Inefficient for dense data (density above $\sim 20\%$).
Pattern growth [1]	Divide-and-conquer without candidate generation and multiple scans. A frequent-pattern tree is built (from an ordered list of frequent items) and mined (based on prefix paths co-occurring with growing suffix patterns). By using the least frequent items as a suffix, a good selectivity is achieved.	Depth-first tree generation; Alternative trees; Combined bottom-up and top-down traversals; Array-based structures.	Not able to deliver the supporting transactions of a pattern (required for biclustering). Adequate for dense matrices and low supports.
Vertical projection [135]	Eclat, a representative vertical method, builds the transaction-set for each item and grows the itemsets under a depth-first strategy (similar to FP-growth) by intersecting transaction-sets to avoid multiple scans.	Specialized structures; Bit-set operations;	Optimized for flattened matrices ($n > m$).

Table 4
Classes of biclustering approaches according to merit-guided searches and optima guarantees.

Paradigm	Optimality guarantees
<i>Divide-and-conquer</i> approaches to exploit the matrix recursively with the branching following a global merit function [57,128,137]. Although efficient, the structure of biclusters is restrictive and the initial assumptions can easily lead to the missing of relevant biclusters. <i>Greedy</i> iterative approaches with the selection, addition and removal of rows and columns being performed until a local merit function is maximized [35,89,131,94,15].	<i>Local optima</i> (local searches dependent on initial assumptions and convergence behavior)
<i>Two-way clustering</i> approaches under merit functions to produce the clusters on both dimensions of the data matrix and to derive biclusters from their combinations [49,120,47]; <i>Stochastic</i> approaches that model data with a multivariate distribution [105,112,17,113] and learn a parametric model that maximizes a merit function. This model is used to derive biclusters.	<i>Distance-based guarantees</i> as learners rely on approximative views (clustering abstractions or generative models)
<i>Ensemble methods</i> [56] that use a merit function to aggregate a large set of biclustering solutions from the iterative application of multiple biclustering approaches.	Dependent on selected approaches
<i>Exhaustive</i> approaches under constraints (e.g. fix number of biclusters, differential expression) [119,126,110], which rely on heuristics based on merit functions to guide the space exploration.	<i>Global optima</i>

PM-based biclustering: While traditional biclustering approaches rely on flexible merit functions to guide the space exploration, PM-based approaches require these functions to be defined in terms of support and, eventually, confidence or other interestingness metrics. This restriction enables a scalable exhaustive space search that produces an arbitrarily high number of biclusters within a flexible structure.

Definition 1.4. Let A be a matrix whose values in \mathbb{R} are assigned to a set of items \mathcal{L} . A bicluster under a *constant model* can either follow: an overall orientation where $a_{ij} \in \mathcal{L}$; a column-based orientation where $a_{ij} = k_j$ and $k_j \in \mathcal{L}$; or a row-based orientation where $a_{ij} = k_i$ and $k_i \in \mathcal{L}$. A bicluster following an *additive* (or *multiplicative*) *model* has $a_{ij} = k_j + \gamma_i$ (or $a_{ij} = k_i \times \gamma_j$), where $k_i \in \mathbb{R}$ and $\gamma_j \in \mathbb{R}$ define the column and row contributions. A bicluster under a *symmetric model* either considers symmetries on rows $c_i \times a_{ij}$ or columns $c_j \times a_{ij}$, where $c_i \in \{-1, 1\}$.

Definition 1.5. Given a matrix A whose elements are the concatenation of the observed values $a_{ij} \in \mathcal{L}$ with their column (or row) indexes. Let \mathcal{P}_P of an itemset P in A be its set of indexes. set of *biclusters* $\cup_k(I_k, J_k)$ can be derived from a set of frequent itemsets $\cup_k P_k$ by mapping $(I_k, J_k) = B_k$, where $B_k = (\Phi_{P_k}, \Psi_{P_k})$, to compose biclusters with coherency across rows, or $(I_k, J_k) = (\Psi_{P_k}, \Phi_{P_k})$ for column-coherency.

Two classes of PM-based biclustering approaches can be considered: (1) a first class targeting discrete matrices by using as-is pattern miners, and (2) a second class targeting numeric matrices by

extending methods based on the introduced monotonic (or Apriori) property [2]. The first class of methods rely on an itemization step followed by the application of FIM under a low support threshold. The itemization step maps a real-value or discrete matrix into an itemset database. For real-value matrices, normalization and discretization procedures are applied. Then, the discrete value of each element is concatenated with its column index. Each transaction of the target itemset database corresponds to a row with these new values. FIM is then applied over this database to mine frequent patterns for composing biclusters with coherency across rows. The second class of methods relies on variants of the FIM task to learn frequent patterns directly from the real-valued matrix. In both classes, the coherency strength δ is implicitly defined by the number of items or the maximum allowed distance. Biclusters with coherency across columns can be mined using the transpose matrix. Finally, biclusters with coherent values overall can be discovered by mining one item (or range of values) at a time. Fig. 1 illustrates how to deliver these different types of biclusters using frequent patterns when considering the constant model.

1.2. Related work

To our knowledge, BicPAM [62], BiModule [96], DeBi [111], Bellay's et al. [14], GenMiner [86] and BiP [60] are the state-of-the-art methods for the first class of PM-based biclustering. BiModule [96,97] allows a parameterized multi-value itemization of the input matrix to discover constant biclusters derived from (closed) frequent

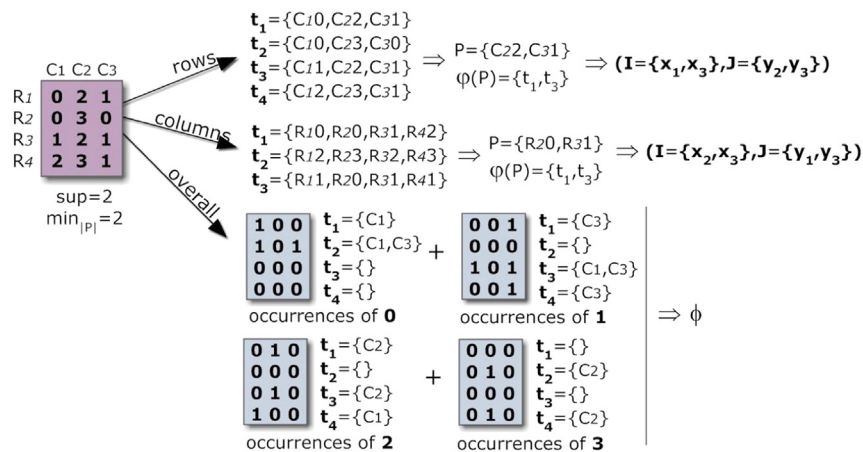


Fig. 1. Mining biclusters with constant assumptions over itemset matrices. To discover biclusters with constant values on the rows, the input matrix needs to be itemized. Column identifiers are combined with the observed values, and FIM applied under a parameterizable support threshold ($\theta = 2 \wedge |P| \geq 2$). Constant values on columns can be mined using the transpose matrix. To find biclusters with constant values overall, each item needs to be separately mined. In each iteration, only the elements containing the selected item are included in the transactions.

patterns using the LCM miner [125]. DeBi [111] derives biclusters from (maximal) frequent patterns mined over binarized matrices using the MAFA miner [22], and places key post-processing principles to adjust them in order to guarantee their statistical significance. The recently proposed BicPAM [62], parameterized with the F2G miner [65] by default, extends the constant assumption of previous approaches to find biclusters with symmetric, additive and multiplicative factors by performing iterative corrections on the input matrix. BicPAM also surpasses discretization problems by introducing the possibility to assign multiple discrete values to a single element, and offers new strategies to robustly handle noise and missing values. Bellay's et al. method [14] uses the Apriori miner [2] with additional principles to evaluate the functional coherency of the discovered biclusters against the background noise. This is one of diverse PM-based attempts to exhaustively discover dense biclusters in either unweighted networks [13,90,133,80] or, more interestingly, in scored networks [32,30]. GenMiner [86] includes external knowledge within the input matrix to derive biclusters from association rules that relate annotations (external grouping of rows or columns) with clusters derived from (closed) frequent patterns using CLOSE [102]. BiP [60] is prepared to discover plaid models by relying on noise-tolerant association rules for the recovery of apparent noisy areas due to the presence of cumulative effects on the overlapping areas between biclusters.

The itemization step is optional for the second class of methods [8]. To our knowledge, RAP [101], RCB discovery [8] and ET-bicluster [52] are state-of-the-art methods here. RAP [101] plugs an adapted range-based metric to mine constant biclusters on rows (or columns), while RCB discovery targets biclusters with constant values overall [8]. ET-bicluster extends the previous approaches to discover noisy biclusters, although an exhaustive enumeration of biclusters is not guaranteed [52]. Alternative support metrics with dedicated Apriori-based searches have been additionally proposed [69,115,53].

2. PM-based biclustering

We propose a structured view of PM-based biclustering according to a set of dimensions of decision. We rely on state-of-the-art literature to characterize each dimension. These dimensions gather principles on different steps with impact on the biclusters' type, structure and quality, as illustrated in Figs. 2 and 3. Throughout this paper we define a set of principles for each step.

Different options for PM-based biclustering can be grouped according to its three major steps: *mapping* (preprocessing), *mining* (pattern discovery), and *closing* (postprocessing). The core step is the *mining* step, corresponding to the application of the target pattern miners. This step is driven by the chosen paradigm, target patterns and search properties. The *mapping* step (optional for methods able to deal with non-discrete data) is responsible for the itemization of a (real-value) matrix and for other preprocessing options to handle outlier, noisy and missing elements. Finally, the *closing* step includes the postprocessing of the mined patterns to affect the structure and quality of the target biclustering solutions.

These options impact the homogeneity of the biclustering solutions. The homogeneity criteria can be intentionally controlled to search for biclusters with a specific coherency (underlying pattern correlation), *structure* (number, size and positioning of biclusters) and *quality* (amount and type noise within a particular bicluster or set of biclusters).

Section 2.1 covers the core PM-based biclustering paradigms. Sections 2.2 – 2.3 detail the remaining mapping and closing dimensions and discuss their implications in the behavior of PM-based approaches.

2.1. Mining options: discovery of biclusters using pattern mining

Flexible scenarios where the number and position of biclusters is not constrained require efficient algorithms [111,81]. The adequate use of PM approaches is critical to guarantee the flexibility and scalability of the biclustering algorithm, and depends essentially on four variables discussed below: (1) the chosen PM-based approach to biclustering, (2) the application schema, (3) the target pattern representations, and (4) the search strategies.

2.1.1. Mining approaches to compose biclusters

In what follows, we overview the state-of-the-art options using: (1) frequent pattern mining, (2) association rule mining, (3) structured pattern mining, and (4) hybrid approaches to compose biclusters.

2.1.1.1. Frequent pattern mining. Two main strategies can be considered: (1) relying on frequent itemset mining (FIM) support metric as-is; and (2) defining new (anti-)monotonic support metrics for a dedicated yet efficient search.

Fig. 1 illustrates how PM can be applied to find biclusters with constant items overall, on rows and on columns. When ignoring the closing step, the discovered biclusters are the frequent itemsets. The support threshold defines the minimum number of rows in a

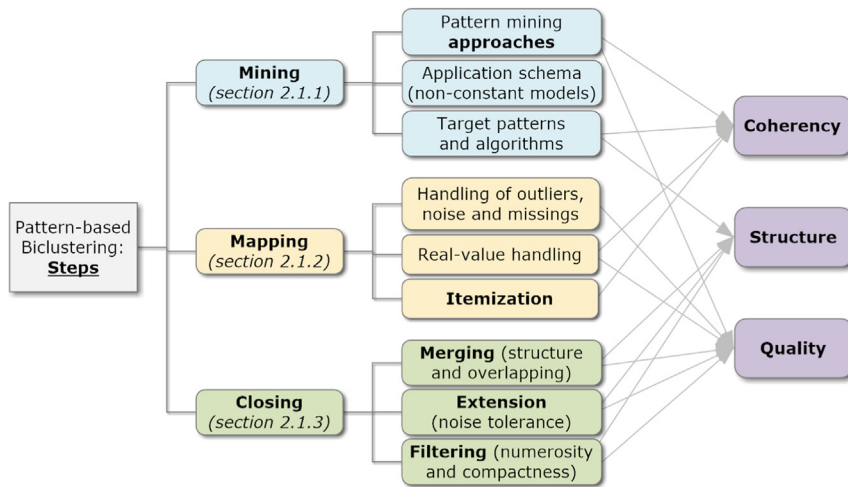


Fig. 2. Process-view dimensions and their impact on biclustering solutions.

GOAL \ STEPS	Target coherency	Target structures	Target quality
Mining options (section 2.1)	<i>PM approaches based on itemsets, rules and sequences to mine constant, plaid and order-preserving models. Extensions for additive/multiplicative/symmetric models. Search alternatives for row- and column-based biclusters.</i>	<i>Pattern types (as condensed representations) and mining approaches (as rule-based mining) with effect on the number and shape of biclusters.</i>	<i>Metrics (support-confidence-correlation), mining approaches and pattern types impact the level of noise allowance.</i>
Mapping options (section 2.2)	<i>Number of items (or value's ranges) to determine coherency strength.</i>	<i>Number and positioning of biclusters mainly affected by the itemization (noise relaxations and missings handling) and discretization.</i>	<i>Multi-item assignments for discretization. Handlers of missings and outliers. Noise-relaxations.</i>
Closing options (section 2.3)	<i>Affect coherency strength with extension and filtering. Strategies to recover areas explained by plaid effects.</i>	<i>Merging and summarization options enable to customize structures and control their volume.</i>	<i>Type and amount of noise defined by extension, filtering and merging criteria.</i>

Fig. 3. Structured view of PM-based biclustering: illustrative options across the major dimensions. It groups critical decision dimensions (corresponding to either a row, a column or a cell of the framework) to support the design of PM-based biclustering approaches. A set of principles for each dimension is illustrated and detailed throughout this work for each biclustering step (mining, mapping and closing) and biclustering goal (defined according to a specific type, structure and quality of biclustering solutions).

bicluster. By decreasing this threshold we are degrading the efficiency of the task, but searching for a broader set of biclusters with smaller sizes. In the context of gene expression, this is critical since small groups of genes can be functionally related. Additionally, the search can allow the pruning of itemsets below a minimum number of columns and above a maximum number of rows and columns.

From the point of view of an itemized database, the FIM-based biclusters are perfect biclusters, that is, they do not allow value-variations in any of its elements. Contrasting, from the point of view of the input real-value matrix, these biclusters can handle noise as different values may be assigned with the same item. The number of items can be flexibly parameterized to control the level of noise-tolerance, which contrasts with traditional biclustering approaches over discrete matrices³ [94,119]. Although BiModule [97,96] allows a parameterizable number of items and support threshold, the structural data noise and the applied itemization procedure often leads to the partitioning of large biclusters into smaller ones (with many of them filtered out as no longer satisfy the support criterion). Contrasting, although DeBi [111] and Bellay's et al. method [14] alleviate this problem by providing postprocessing strategies to improve the functional coherence of the discovered biclusters, they require the input data to be binarized.

FIM-based approaches suffer from the risk of assigning elements with similar real-values to different items. We refer to this drawback as the items-boundary problem. In order to address this problem, the notion of support of an itemset can be redefined. As

long as the new support metric is (anti-)monotonic, its inclusion within Apriori-based frameworks [101] can be easily handled with efficiency. Patterns are thus generated using breadth-first level-wise pattern tree.

Han et al. proposed Min-Apriori [53], an algorithm to deal with ordinal items. Steinbach et al. [115] introduced a framework to generalize the notion of support to extend association analysis to continuous-based patterns. An alternative support function [69] has been proposed to mine hyperclique patterns (groups of columns or rows strongly related) over numeric matrices. Calders et al. [25] proposed the use of rank-based measures to score the similarity of sets of numeric attributes within new support metrics by extending τ , Spearman's ρ , and Spearman's Footrule F correlation metrics [71]. Here, efficient algorithms are designed to deal with the ranks of attribute values, but not with the original numeric values. However, these approaches do not capture key properties of real-valued matrices, such as the need to ensure that the values of items in a transaction are within a range to guarantee coherence and distinguish positive from negative values.

More recent approaches propose range-based support metrics to either discover coherency on rows, such as RAP [101]. RAP is defined under a sign-coherence constraint, enforcing that a transaction can only contribute to the support of a pattern if the values of all the items in it have the same sign.⁴ An alternative, RCB

³ Illustrating, xMotif [94] relies on greedy search and uses a size merit function and a noise threshold to guarantee the discovery of large and interesting biclusters, and SAMBA-based approaches [119] map binarized matrices into a weighted bipartite graph to find subgraphs that maximize a weight merit function.

⁴ For a matrix $A=(X,Y)$ and $I \subseteq X, J \subseteq Y$, the support metric is defined as $sup(I) = \sum_{i \in X} S(i,J)$, with:
 $S(i,J) = \begin{cases} \min_{j \in J} |a_{ij}| & \text{if } (\max_{j \in J} a_{ij} - \min_{j \in J} a_{ij}) \leq \sigma \min_{j \in J} |a_{ij}| \wedge (\forall_j a_{ij} > 0 \vee \forall_j a_{ij} < 0) \\ 0 & \text{otherwise.} \end{cases}$

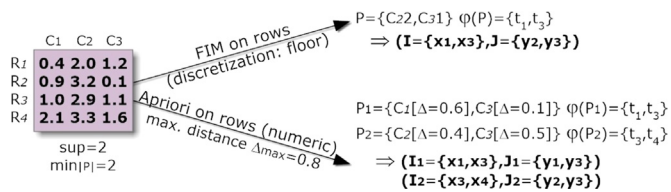


Fig. 4. Biclustering using FIM over discrete matrices versus range-based searches over numeric matrices. Biclusters discovered with range-based support metrics are less prone to the items-boundary problem.

discovery method [8], verifies range constraints on both dimensions (rows and columns) using a monotonic range measure. The Apriori-based method is slightly modify in order to grow homogeneous-squares that are then used to compose rectangles (biclusters). Finally, ET-bicluster model [52] revises the previous support metrics for the discovery of noisy biclusters by guaranteeing that each supporting transaction of a pattern does not exceed a specific error-threshold. Although this support metric is not anti-monotonic and thus does not guarantee the exhaustive search of all possible patterns, optimality distances can be given.

Despite the relevance of this type of hyperclique-based approaches to avoid the items-boundary problem, they require the definition and parameterization of (anti-)monotonic metrics. Additionally, PM principles to enhance scalability and to discover condensed representations for these range-based patterns cannot be directly applied. Fig. 4 provides an illustrative application of this type of enhanced FIM-based approaches against traditional FIM-based approaches.

In labeled datasets, FIM-based approaches have been extended for the discovery of class-discriminative biclusters (biclusters with significantly higher support for a particular class) [43,116,95].

2.1.1.2. Association rule mining. Association rule mining can alternatively be used to compose biclusters [60]. Its core task is the support-guided discovery and confidence-guided combination of frequent itemsets [54]. Given a matrix A, a simple association rule relates columns ($J \rightarrow J'$) or, when transposed, relates rows ($I \rightarrow I'$). An illustrative rule from a transposed gene expression matrix is $\{gene_A \downarrow\} \rightarrow \{gene_B \uparrow, gene_C \uparrow\}$, meaning that when $gene_A$ is under-expressed, it is very likely that genes B and C are over-expressed. An arbitrary high number of states/items can be considered. When using association rules to compose biclusters, the items on the antecedent and consequent of a rule, as well as the supporting transactions from both sides, are considered to derive each bicluster. Thus, association rules can be used to capture accommodate noise when confidence levels are below 100%, as illustrated in Fig. 5. Consider the illustrative rule $R_1 : \{g_2, g_3\} \rightarrow \{g_4, g_5\}$ with confidence below 90%. Instead of using the conditions that support $\{g_2, g_3, g_4, g_5\}$ to build the bicluster, one can extend it by considering the conditions that support uniquely

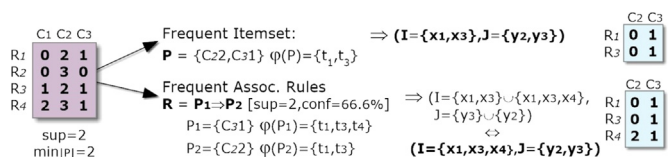


Fig. 5. Discovering biclusters from association rules: comparing noise-intolerant biclusters from frequent itemsets vs. noise-tolerant biclusters from association rules.

R_1 antecedent, $\{g_2, g_3\}$. Confidence is thus seen as a homogeneity indicator.

To mine specific rules of interest, other interestingness metrics have been used to augment the support-confidence framework, including lift, conviction, chi-square, cosine and all-confidence [117]. BiP explores the thresholds of these metrics can be explored

to discover biclusters with varying quality [60]. In matrices with numerous correlations, the support should be set low, the confidence set high, and constraints incorporated to deal with the explosion of rules from frequent itemsets. For instance, the rule-based GenMiner approach [86] imposes rules to be non-redundant with minimal antecedent and maximal consequent (minimal non-redundant rule for short) in order to avoid the explosion of rules. Alternatively, association rules can be pruned based on their statistical/biological significance [5] according to hypotheses verified by correlation coefficients (such as Pearson's Product Moment Correlation, Spearman's Rank-order Correlation Coefficient and Kendall's Tau).

Carmona-Saez et al. [26] and GenMiner [86] extended simple rules by integrating annotations from semantic sources and (biomedical) knowledge bases. Illustrative rules include: $annotation_1 \Rightarrow \{c_1 \downarrow, c_2 \uparrow\}$, meaning that a group of genes (with the same annotation) is likely to be under-expressed in condition c_1 and over-expressed in condition c_2 , or $\{c_1 \downarrow, c_2 \uparrow\} \Rightarrow annotation_1$, meaning that a group of genes with the expression profile given by c_1 and c_2 is likely to have specific annotations.

Finally, and similarly to support customization, confidence and other interestingness metrics can be customized and plugged within an Apriori-based framework. However, to our knowledge, there are not yet implementations of this type of rule-based approaches.

2.1.1.3. Structured pattern mining. Approaches that target different types of patterns provide alternative search paradigms for biclustering and hold the potential to discover biclusters with specific properties. This set of approaches includes:

- **Constraint-based pattern mining** or actionable pattern discovery approaches. Biclusters are declaratively defined through the use of flexible pattern constraints that specify the target homogeneity criteria. In this context, a bicluster is a specific formal concept called bi-set. A bi-set satisfies, at least, a local constraint: the column set (or intent) is the maximal set of columns that are true for the supporting set of rows (or extent) [19,4];
- **Sequential pattern mining (SPM) approaches**: SPM can be used to mine order-preserving biclusters [61,63,78]. A bicluster is order-preserving if there is a permutation of its columns under which the sequence of values in every row is (either monotonically or strictly) increasing. For this aim, the indexes of the elements in the matrix are reordered per row; the ordered set of indexes are mapped into a sequential database; SPM is applied; and the biclusters are mapped from the frequent sequences and their supporting transactions (Fig. 6). OP-Clustering [78] was the first attempt to SPM-based biclustering. More recently, BicSPAM [63] was proposed to address the efficiency bottlenecks and noise-intolerance of previous algorithms, and allow a parameterizable variation of the degree of co-occurrences versus precedences to affect the order-preserving coherency;
- **Two-way PM-based clustering approaches**: This is a promising direction since the patterns underlying the clusters on each dimension can be used to affect the structure, quality and type of biclusters [49];
- **Graph mining approaches**: Real-value matrices can be mapped into weighted bipartite graphs, and thus biclustering can be mapped into the task of finding maximal cliques [84] or other substructures from graphs derived from binarized matrices [119]. Despite its computational complexity, structured pattern mining over weighted bipartite graphs is a direction with growing attention [28];
- **Cube computation approaches**: Cube computation shares similarities with frequent pattern analysis, being well-suited to deal with matrices in \mathbb{R}^n when $n > 2$ [55,132]. The additional dimensions can be used to capture additional informative views (such as time points or replicates) [3], to model

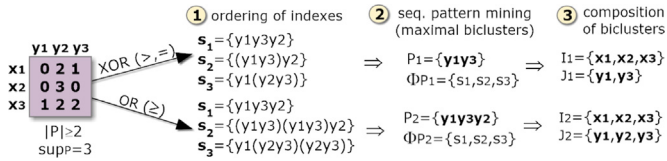


Fig. 6. Mining order-preserving biclusters in real-valued matrices with sequential pattern mining.

contributions from overlapping areas of biclusters under a plaid model assumption [60], or to find biclusters' consensus over cubes with different pre-processing and closing criteria.

2.1.1.4. Hybrid approaches. Biclustering can rely on multiple types of patterns discovered by different PM approaches. Valid options include the definition of ensemble methods combining plain and structured patterns or the output of multiple PM methods (parameterized with different support-confidence thresholds). Frequent itemsets can be also used to produce an initial solution, while rules can be posteriorly mined to shape the discovered biclusters by accommodating noise. An alternative ensemble model can rely on the multiple results from the iterative parameterization of a PM method with different PM-based constraints of interest. To our knowledge, these hybrid possibilities have not been systemically studied in literature.

2.1.2. Application schema

The previous pattern mining approaches can be iteratively applied with a decreasing support threshold until a stopping criteria is achieved [62]. BicPAM makes available distinct stopping criteria, including a minimum coverage of the elements in the input matrix by the discovered biclusters or, alternatively, an approximate number of biclusters (after or prior to postprocessing) [62]. Such criteria can either be driven from user expectations or dynamically derived from the properties of the input matrix [63].

Furthermore, iterative corrections can be applied on the matrix to enable the discovery of more flexible coherencies. BicPAM makes use of the observed differences and of the least common divisor between the observed values for a given column (or row) in the matrix in order to perform iterative corrections across rows (or columns) and thus identify shifting and scaling factors. The removal of these factors in the matrix allows the discovery of additive models and multiplicative models [62]. Similarly, BicPAM can also rely on combinatorial sign-adjustments across rows (or columns) to model symmetries, and integrate them with shifting and scaling factors [62]. Pruning strategies are considered to avoid redundant calculus and reduce the computational complexity of these iterative corrections.

BiP relies on the converging application of PM for learning plaid models [60], based on the observation that, by incrementally removing overlapping contributions, the residual values become closer to the underlying unstructured noise. For this aim, BiP performs checks between iterative applications of PM searches in order to recover areas explained by cumulative effects (contributions on overlapping areas between biclusters) and to remove noisy areas that are not described by a plaid assumption. Without degradation of efficiency levels, it also provides relaxations to model overlapping contributions characterized by noisy and non-linear cumulative effects [60].

2.1.3. Pattern representation

Depending on the chosen PM approach, different patterns, such as frequent itemsets, association rules, sequential patterns or structured patterns, can be considered. Each of these patterns can have different representations, being the most common: simple, maximal, closed, pseudo-closed, approximated, rare, top-K, multilevel and erasable [24]. In particular, when targeting association rules,

additional representations can be considered as indirect, minimal, non-redundant, approximative, quantitative and sporadic rules [117]. Although an analysis of the impact of using each representation on the biclustering solutions is possible, we consider simple, maximal and closed representations for simplicity sake.

Definition 2.1. Given an itemset matrix, a support threshold θ , and the coverage function $\phi : 2^L \rightarrow 2^D$ that maps an itemset P to its set of supporting transactions. A closed frequent itemset is a frequent itemset that has no superset with the same support ($\forall P' \supset P |P'| < |P|$). A maximal frequent itemset is a frequent itemset with all supersets being infrequent, $\forall P' \supset P |\Phi(P')| < \theta$.

Given an itemset database $D_{ex} = \{\{A, B, H, J\}, \{D, H, J\}, \{C, D, H, J\}\}$, and thresholds $\theta = 2$ ($|\Phi P| \geq 2$) and $|P| \geq 2$, there is one maximal frequent itemset ($\{D, H, J\}$) and there are two closed frequent itemsets ($\{D, H, J\}$ and $\{H, J\}$). The selection of the pattern representation essentially depends on the type and structure of the target biclusters, and on the post-processing needs.

Maximal itemsets for biclustering, such as those used in DeBi [111], are associated with biclusters with the columns' size maximized. Such flattened biclusters are only of interest when there is an extension step to be performed to include new rows. However, since both vertical and smaller biclusters are lost, this representation leads to incomplete solutions. The opposite alternative is the use of all frequent itemsets for biclustering. This solution leads to a high number of potentially redundant biclusters (if contained by another bicluster), which can degrade the performance of the mining and closing steps. Finally, the search for closed itemsets, such as FIM-based BiModule [96] and rule-based GenMiner [86], allows the discovery of overlapping biclusters if a reduction on the number of columns results in a higher number of rows. Closed pattern solutions are thus enabling the return of all maximal biclusters (set of biclusters that are not included in other biclusters). The properties of these three alternative representations are illustrated in Fig. 7.

2.1.4. Search strategies

The choice of the search strategy depends essentially on the target biclustering task and on the properties of the considered implementation. Generally, PM searches are centered on computing the set of frequent patterns, which is the core task of all pattern miners.

The choice of whether to use a vertical or an horizontal data format depends essentially on the type of biclusters we are targeting. To find constant items on the rows or on both dimensions, we usually benefit from using searches over horizontal data. This is particularly true for matrices where the total number of rows largely exceeds the total number of columns. To find constant items on the columns (when $n > m$), a vertical data format should be the choice, as the performance of searches using the horizontal format degrades exponentially with the increase in the number of items.

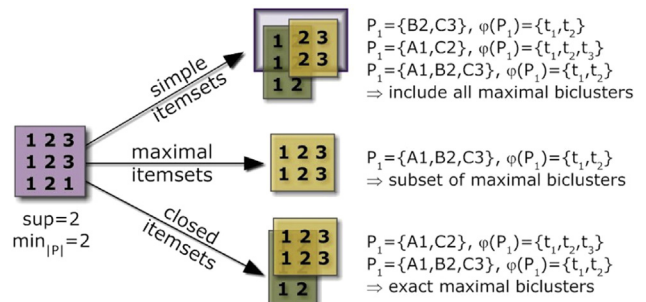


Fig. 7. Comparison of biclustering solutions using simple, maximal and closed patterns.

The choice of whether to use an Apriori-based, pattern-growth or combined approach, depends on three variables: (1) the type of PM-based approaches (range-based approaches cannot rely on pattern-growth methods), (2) the density of the resulting itemset matrix, and (3) the ability to retrieve the supporting transaction set for each frequent itemset without degrading the overall efficiency. This analysis is detailed in supplementary material. When biclusters with constant values overall are targeted, the resulting matrices are sparser (Fig. 1) and, therefore, an Apriori strategy is preferred. For denser matrices, pattern-growth strategies are preferable.

In particular, the discovery of patterns together with their supporting transactions has been tackled using extensions over Apriori and vertical-based algorithms by relying on bitset vectors to capture the supporting transactions per pattern [86,111,96]. However, bitset vectors offer efficiency problems in terms of memory and time for large and dense datasets. Henriques et al. [65] study efficient alternatives and propose a pattern-growth algorithm to discover full-patterns with heightened time and memory efficiency.

An additional key aspect is the chosen implementation. The use of bit-set operations and either reduced number of scans or efficient tree-traversals are usually key for a top performance. Efficient implementations include algorithms to mine closed itemsets under an Apriori search (LCM [125], Charm [136]), vertical search (TD-Close [77]) or pattern-growth search (FPClose [51]); and to mine maximal itemsets under an Apriori search (MaxMiner [11]), vertical search (Mafia [22]) or pattern-growth search (AFOPT [76]). Similarly, multiple implementation variants can be found to compose association rules [138,87] and to mine structured patterns. For instance, sequence miners can either use Apriori, pattern-growth and vertical searches, and find closed and maximal sequential patterns [79]. In DeBi [111], BiModule [97] and GenMiner [86] use Mafia [22], LCM [125] and CLOSE [102] implementations, respectively. Range-based variants use Apriori [2]. Additional principles proposed in literature [138,99,100] can be seized to guarantee the scalability of the search when mining large biclusters from dense or large data settings.

2.2. Mapping options: preprocessing input data

Previous section covered essential mining options with impact on the coherency, structure and quality of PM-based biclustering solutions. However, their optimum application requires the input matrices to be correctly normalized⁵ and (depending on the PM-based approach) discretized. The problem of defining an adequate coherency strength is identical for range-based approaches (distance thresholds as a function of data domain values) and discrete PM-based approaches (number of items). Although discretization may imply loss of information, it alleviates the noise dilemma [26,31].

Since discretization is a key step for the class of PM-based methods that relies on itemset databases, having key implications on the target solution, we study two variables: (1) the number of items (also referred to as symbols or expression levels) and (2) the method used to map the normalized real-value matrix into a itemset database. A sensitivity analysis on the impact of the number of items on the quality and size of biclusters was, first, performed in Bidens [83] and BiModule [96]. Fig. 8 illustrates how simple discretization options can lead to different solutions. The itemization (concatenation of the item with the column-index) implies that the resulting number of items is at most $m \times l$, being l the number of items specified by the user. The use of fixed ranges (potentially equal sized intervals between the observed

maximum and minimum) is the simplest discretization option, but it usually leads to an accentuated weak distribution of items and it is prone to the items-boundary problem. The first problem can be corrected using a percentage-based method for the depth partitioning of items that leads to intervals containing approximately the same number of elements. Alternatively, distributions combine the properties of the previous solutions. In the example, a Gaussian distribution is able to minimize the loss of potentially relevant biclusters. By finding multiple suitable curves (for each row or column) or one suitable overall curve to approximate the matrix, one can either use threshold methods [26,31] or compute the statistical cutoff points to create equally-distributed areas. Nordi [86] is a Gaussian-based method used in GenMiner [86] that statistically detects outliers (using the Grubbs method), applies normality tests (using QQ-plot and Lilliefors) to transform the initial row distributions into a "more" Normal distribution, and computes cutoff thresholds using the z-score methodology. In the presence of matrices with multimodal distributions, more expedite methods based on a mixture of distributions must be considered.

A unique advantage of PM-based approaches is the fact that they can easily address the items-boundary problem of discretization procedures by assigning two or more items to an element in the original matrix with a real value that is near a discretization boundary (or cut-off point). This is possible since PM is able to learn from transactions (mapped from the rows of an itemized matrix) with an arbitrary number of items. Despite the critical relevance of this strategy, its impact was not yet systemically assessed.

Alternative discretization options that aim to deal with this problem include: (1) adaptive discretization based on dynamic threshold selection policy [107]; (2) statistical methods to detect differential activity of elements as the basis to create partitions [31] (commonly adopted as a binarization method); (3) distance-based subspace clustering models [75] to flexibly partition the values while preserving meaningful and significant clusters; (4) fuzzification approaches where a continuous domain is partition into fuzzy sets, provided to be more robust to noise when compared with other simple binning techniques [47]; and (5) supervised discretization methods [45] (when descriptive labels per row or column are present or computed using clustering methods), where a row or column is partitioned into a number of disjoint intervals in such a way that the entropy of the partition is minimal.

An additional preprocessing concern appears for matrices with arbitrary-high number of missing elements. Although multiple imputation methods have been proposed [122,38,58] to alleviate this problem, they can introduce additional noise and undesirably affect the homogeneity of the output biclusters. BicPAM [62] and BicSPAM [63] consider varying relaxations to surpass this problem, including a relaxed setting where the missing element is replaced by all the available items (leading to transactions with varying size), and a medium-constrained setting to consider a parameterizable number of items around its value-estimation.

2.3. Closing options: postprocessing biclustering solutions

PM-based biclustering approaches produce exhaustive solutions with flexible structures (arbitrary number and positioning of biclusters). These non-exhaustive, non-exclusive structures, where overlapping is allowed, are the most suitable option to tackle the applications listed in Table 1.

Two key challenges of exhaustive solutions are: handling noise and dealing with the potential explosion of valid biclusters. Part of these questions can be answered in the mapping step by selecting the number of items and discretization setting able to handle the items-boundary problem. However, postprocessing may be required to avoid the following two challenges of the noise dilemma. The first results from a too restrictive noise tolerance, commonly associated with a high number of items, which leads to many small sized biclusters. The second is related to heightened

⁵ Normalization options are often applied before biclustering to enhance differences across rows and/or columns and, consequently, to improve the ability to discover biclusters. de Souto et al. [34] compare three normalization procedures (z-score, scaling and rank-based procedures) over gene expression datasets using alternative clustering algorithms. Additional methods for preprocessing the input matrix have been reported [118,83,25].

levels of noise allowance, commonly occurring in binarized partitions or through the use of rule-based approaches under a relaxed level of confidence. To handle these challenges we propose the use of a set of criteria structured according to three major postprocessing steps (merging, filtering and extension) described below.

Merging options: Merging biclusters may serve two goals: noise allowance (to avoid solutions composed uniquely of small biclusters) and overall biclustering structure manipulation. The first goal is driven by the observation that when two biclusters share a significant area it is probable that their merging composes a larger bicluster still respecting some homogeneity criteria. Commonly, such decomposition is related to the items-boundary problem or with a missing value. The simplest criterion to allow the merging is either to rely on the overlapping area (as a percentage of the smaller bicluster), to compute the overall noisy percentage after the merging, or both. Additional homogeneity criteria relying on the real-values provided by the input matrix can be formulated. Henriques et al. [61] performed a comparison between three distinct efficient merging techniques. Bellay et al. [14] proposed a Markov Clustering (MCL) algorithm to both summarize biclustering solutions and allow for the creation of larger biclusters.

Filtering options: Filtering is needed at two levels: (1) at the row/column level and (2) at the bicluster level. The first type of filtering is needed to exclude rows or columns from a particular bicluster in order to improve its homogeneity. This is usually the case when a low number of items is considered, leading to highly noise-tolerant biclusters. For this purpose, we can rely on statistical tests on each row and column of a particular bicluster to identify removals [111]; use existing greedy-iterative approaches to maximize a merit function until a parameterizable reduction in size is

verified [29]; or discover patterns under more restrictive conditions (as higher support and confidence thresholds) and use them guide the removal of rows and columns [62,63].

The second type of filtering is required to guarantee the dissimilarity of biclusters, removing biclusters partially contained in larger biclusters. BiModule [96] filters small biclusters by sorting biclusters following the score $a_{ij} \times \log_2 |I| \times \log_2 |J|$ and biclusters whose cells overlap by more than 25% with a higher scored bicluster. The work by Bellay et al. [14] separates biclusters that represent biological phenomena from false discoveries (emerging from the background data distributions) using randomized data scores.

Extension options: Three optional and non-exclusive strategies can be used to extend the discovered biclusters so that the resulting solution still satisfies some pre-defined homogeneity significance criteria. First strategy consists on the use of statistical tests to include rows or columns from each bicluster. DeBi [111] uses statistical tests to extend biclusters obtained over binary matrices by evaluating the association strength between key columns of a bicluster and a new row using Fisher's exact test for independence on a contingency table. This guarantees that each row in the bicluster shows a statistical difference between the columns in the bicluster and the columns not in the bicluster, leading to more functionally coherent biclusters. Second strategy is to rely on traditional merit functions for further (greedy) extensions over PM-based biclusters. Third strategy is to discover patterns under more relaxed criteria (such as lower support-confidence thresholds) and use them to guide the extension step [62]. When considering lower supports, new columns and rows can be added to the original frequent patterns. Similarly, more relaxed association rules, with less restrictive ways to group the antecedent-consequent, can be used to guide extensions.

Alternatives to merging, filtering and extension options: Alternatives to previously introduced closing options to deal with large sets of small biclusters include: (1) summarization techniques based on simple and hierarchical clustering methods or on the definition of similarity measures to compare biclusters [18]; (2) user-driven formal constraints and querying expressions [19,20]; (3) co-clustering for exclusively partition both dimensions to select representative biclusters [36]; (4) pre- and post-pruning techniques (including item-based constraints and discrimination metrics) [88]; (5) patterns based on half-spaces (as quantitative rules) in which external sources of information are used as a filtering basis [48]; and (6) verification techniques based on metrics computed using external data sources as,

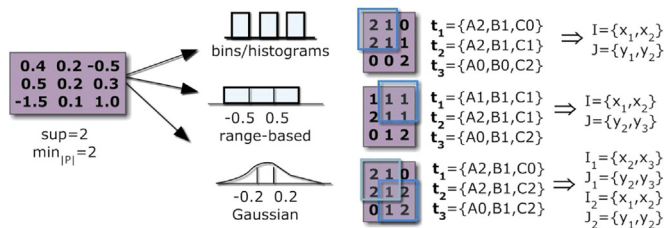


Fig. 8. Comparison of alternative discretization options by addressing their impact on the itemization and biclustering solutions with constant values on columns.

Table 5 Systemic comparison of the two major classes of PM-based biclustering approaches.

Approach	Major benefits	Challenges	Proposed principles to tackle challenges
PM-based biclustering	<ul style="list-style-type: none"> - Exhaustive searches; - Handle missings and noise; - Biclusters with multi-levels of coherency strength; - Extensions to discover flexible coherencies; - Flexible structures; - Flexible searches; - Constraint-based guidance; 	<ol style="list-style-type: none"> 1. Deterioration of efficiency levels for large data (in the absence of PM scalability principles); 2. Not natively prepared to capture additive, multiplicative, symmetric and plaid coherencies (their discovery can be computationally expensive); 3. High number of mined biclusters (memory usage); 4. Need to fix thresholds for the standard (customized) support metric; 	<ol style="list-style-type: none"> 1. Data partitioning methods; PM in distributed settings; approximated patterns (discovered under specific performance guarantees) [54,134]; 2. Iterative data mappings on rows/ columns (with pruning heuristics) to mine non-constant biclusters [62]; merging procedures sensitive to overlapping plaid effects [60]; 3. Adequate data structures; filtering options pushed into mining step; 4. Use of multi-thresholds (iterative method); data-driven estimation;
Range-based support biclustering	<ul style="list-style-type: none"> - Range-based support addresses the items-boundary problem; - Easy extension of Apriori methods to seize efficiency gains when dealing with multiple distances (support thresholds); 	<ol style="list-style-type: none"> 1. Separation of positive and negative values to guarantee monotonicity, resulting in biclusters without simultaneous under- and over-expressed values; 2. Dedicated Apriori-based methods do not allow the direct use of PM scalability principles; 	<ol style="list-style-type: none"> 1. Merging of biclusters with shared columns (or rows) but different signs to avoid the violation of the (anti-)monotonic property; 2. Dedicated extensions to mine patterns with tree structures (required for dense datasets), and to make use of (scalable) data partitions;

for instance, term enrichment (in gene expression data) to affect the addition-removal of columns-rows per bicluster.

2.4. A systematic comparison of PM-based biclustering approaches

In what follows, we provide a synthesis of the benefits and challenges of using PM-based biclustering approaches together with principles on how to tackle existing challenges. Table 5 focuses on PM-based biclustering classes in general, while Table 6 focuses on each surveyed approach in particular.

Understandably, different applications may be better tackled by different PM-based biclustering approaches. BicPAM, BiModule and RAP are default options for settings where meaningful biclusters can only be found using multiple coherency levels, which is often the case with scored biological/social networks, expression data and physiological data [62,96,101]. DeBi and BicPAM are critical for the analysis of large Boolean datasets, such as the ones derived from (web) text data or genomic structural variations [111,62]. GenMiner's ability to incorporate external knowledge is relevant for biological and clinical contexts [86]. The constant overall assumption of RCB is critical to efficiently mine biclusters with a specific behavior or rating in web social data and collaborative filtering data [8]. The noise-tolerance of ET-biclusters and BicPAM is relevant to deal with experimental errors and instance-based variations of physiological, molecular and clinical data [52,62]. Finally, BiP and BicPAM are the choice for the analysis of non-trivial (yet coherent) behavior across biomedical and social domains as they allow the discovery of flexible (yet meaningful and significant) coherencies [60,62].

3. Performance evaluation of PM-based biclustering approaches

This section evaluates the performance of PM-based biclustering approaches. We first describe the quality evaluation methodology and then present preliminary results on synthetic and real data.

3.1. Methodology

Effective evaluation of PM-based biclustering solutions is challenged by three major issues. First, a large variety of metrics and synthetic datasets have been proposed (with many being biased to the specificities of a particular approach) [98]. This is the case either when a variant of the optimized merit function is used to evaluate the approach, or when a developed approach is optimized towards specific data settings. Second, there is no ground truth to evaluate biclusters observed in real data. Finally, existing efforts to develop a standard evaluation [92,108] only cover a subset of all aspects, often leading to wrong assumptions regarding the performance of the assessed approaches.

Evaluating biclustering solutions on both synthetic and real data is essential. In synthetic data, a set of biclusters $\mathcal{H} = \{H_1, \dots, H_g\}$ (referred as hidden or true biclusters) is typically planted. Objective metrics can be formulated since an approximate solution is known a priori, including the relative non-intersecting area (RNAI) [21] and its extension (CE subspace⁶) [103], match scores [108,67], and clustering metrics⁷ (such as entropy, recall and precision) [6,7].

⁶ RNAI cannot distinguish if several or a single found biclusters cover a hidden bicluster, thus CE maps each found bicluster to at most one hidden bicluster and each hidden bicluster to at most one found bicluster.

⁷ Clustering metrics are applied to one dimension at a time (rows or columns). Typical objective functions aim high *intra-cluster similarity* (overall pattern for rows within a bicluster is similar across all columns) and low *inter-cluster similarity* (patterns differ for rows from different biclusters). Entropy combines these views by measuring the homogeneity of the found clusters \mathcal{B} against the hidden clusters \mathcal{H} . Alternatively, F-measure (and its precision and recall components) evaluate how

In particular, we rely on Jaccard-based match scores (MS) to assess the similarity of \mathcal{B} and \mathcal{H} [108]. $MS(\mathcal{B}, \mathcal{H})$ defines the extent to which found biclusters cover the hidden biclusters (completeness), while $MS(\mathcal{H}, \mathcal{B})$ reflects how well hidden biclusters are recovered (precision).

$$MS(\mathcal{B}, \mathcal{H}) = \frac{1}{|\mathcal{B}|} \sum_{(I_1, J_1) \in \mathcal{B}} \max_{(I_2, J_2) \in \mathcal{H}} \frac{|I_1 \cap I_2|}{|I_1 \cup I_2|}$$

Since MS scores are not sensitive to the number of biclusters in both sets, Hochreiter et al. [67] introduced a consensus (FC) by computing similarities between the pairs of closest biclusters between \mathcal{B} and \mathcal{H} . Let S_1 and S_2 be, respectively, the larger and smaller set of biclusters from $\{\mathcal{B}, \mathcal{H}\}$, and MP be the assigned pairs using the Munkres method based on overlapping areas [93].

$$FC(\mathcal{B}, \mathcal{H}) = \frac{1}{|S_1|} \sum_{(I_1, J_1) \in S_1, (I_2, J_2) \in S_2 \in MP} \frac{|I_1 \cap I_2| \times |J_1 \cap J_2|}{|I_1| \times |J_1| + |I_2| \times |J_2| - |I_1 \cap I_2| \times |J_1 \cap J_2|}$$

In the absence of hidden biclusters, only subjective metrics can be formulated. Merit functions can be applied as long as they are not biased towards the merit functions used within the approaches under comparison. A detailed comparison of merit functions is provided by Orzechowski [98]. Complementarily, domain-driven scores can be computed using the groups of rows and columns in biclustering solutions retrieved from real datasets against annotations extracted from (biomedical) knowledge bases, such as Gene Ontology and Yeastract [85,121], semantic sources or bibliographic databases. Biclusters can be ranked using a p -value p from testing the hypergeometric hypothesis against these annotations [16,130]. In biological domains, this score can be used to investigate whether the discovered biclusters show significant enrichment with respect to terms in gene ontologies, transcription factors, protein-interaction networks and metabolic pathways using varying levels of significance and correction procedures [96,111].

We propose a methodology for evaluating PM-based biclustering approaches according to three major decision axes. The first axis concerns the target set of synthetic and real datasets. Synthetic datasets must have varying sizes and configurations and be able to exploit different biclustering solutions with respect to their coherency, size, noise and overlapping degree. The second axis includes the set of biclustering approaches and parameterizations to establish comparisons. Finally, the third axis defines the set of metrics to be used. It should assess: (1) time and memory efficiency; (2) accuracy from synthetic data using match scores, CE subspace or FC consensus; and (3) domain relevance scores from real data.

3.2. Results

Below we collect initial empirical evidence that shows the relevance of PM-based biclustering approaches. The following experiments were computed using an Intel Core i3 1.80 GHz with 6GB of RAM.

Synthetic datasets were generated⁸ by varying the size of the matrices, the number and shape of the planted biclusters and the number of items ($|\mathcal{L}| \in \{5, 10, 20\}$). The properties are described in Table 7. The number of rows and columns for each bicluster followed a Uniform distribution over the ranges presented in Table 7. We allow for overlapping biclusters and a random noise factor (up to $\pm 15\%$ of the range of values), which can difficult the recovery of planted biclusters. For each of these settings we instantiated 20 matrices: 10 matrices with background values

(footnote continued)

well the hidden clusters are represented [7,91]. The underlying principle is that biclusters should cover many rows of a particular hidden cluster but few rows from other hidden clusters.

⁸ Available in <http://web.ist.utl.pt/rmch/software/bicpam>.

Table 6

Benefits, challenges and possible improvements of state-of-the-art PM-based biclustering approaches. PM-based biclustering benefits and challenges in Table 5 apply to DeBi, BiModule, GenMiner and BicPAM/BiP, while both PM-based and range-based biclustering benefits and challenges in Table 5 apply to RAP, RCB and ET-Biclusters.

Approach	Major benefits	Challenges	Principles to tackle challenges
DeBi	Complete and statistical rigorous options for post-processing biclustering solutions; discovery adapted to the target significance level; (see <i>PM-based benefits</i>)	Efficiency deterioration from post-processing extension procedures; discovery of maximal patterns (loss of a large number of potentially significant biclusters); binarization of data; (see <i>PM-based challenges</i>)	Discovery of closed patterns (removes the need for an exhaustive extension of biclusters); multi-level discretization (standardly as remaining PM-based approaches); (see <i>PM-based principles</i>)
BiModule	Multi-level discretization with removal of outliers; (see <i>PM-based benefits</i>)	No merging-extension options for handling noise and growing biclusters; (see <i>PM-based challenges</i>)	Inclusion of the surveyed closing options; (see <i>PM-based principles</i>)
GenMiner	More complete frame to derive noisy biclusters from rules (non-perfect confidence levels); allows extracting relations between genes and real-world annotations; (see <i>PM-based benefits</i>)	Require annotations from knowledge bases; non-parameterized levels of expression (only 3); (see <i>PM-based challenges</i>)	Retrieval of annotations from the dataset under analysis when knowledge bases are not available; delivery of rules without the need annotations for annotation on the antecedent or consequent; inclusion of the surveyed mapping options; (see <i>PM-based principles</i>)
BicPAM/ BiP	Discovery of additive/multiplicative/symmetric/plaid models; robustness to discretization, noise and missings; dedicated PM searches to explore further efficiency gains; (see <i>PM-based benefits</i>)	Efficiency levels of the search for non-constant models rapidly deteriorates for very large matrices; (see <i>PM-based challenges</i>)	New heuristics, scalability principles, approximative searches (replacing the exhaustive criteria), or constraint-based guidance to learn non-constant models; (see <i>PM-based principles</i>)
RAP	(see <i>PM & range-based benefits</i>)	Not able to deal with noisy biclusters; (see <i>PM- and range-based principles</i>)	Inclusion of closing framework (merging and extension strategies); (see <i>PM- and range-based challenges</i>)
RCB Discovery	(see <i>PM and range-based benefits</i>)	Constant coherency overall excludes biclusters with meaningful differences across columns (rows); joining squares (discovered patterns) to compose rectangles (biclusters) is a combinatorial problem that impacts efficiency 8; (see <i>PM- and range-based challenges</i>)	Combined results with other approach biclustering solutions (e.g. RAP); alternative computational methods; (see <i>PM- and range-based principles</i>)
ET-Bicluster	Parameterizable discovery of biclusters based on the allowed amount of noise; (see <i>PM and range-based benefits</i>)	Inclusion of error-based thresholds on the Apriori-method violates the (anti-)monotonic property, thus not guaranteeing exhaustive solutions; (see <i>PM- and range-based challenges</i>)	Adoption of more relaxed thresholds to avoid loosing biclusters of interest with a post-filtering of biclusters non-satisfying criteria; inference of bounds on the performance guarantees; (see <i>PM- and range-based principles</i>)

Table 7

Properties of the generated set of synthetic datasets.

Matrix size (#rows × # cols)	100 × 30	500 × 60	1000 × 100	2000 × 200	4000 × 400
Nr. of hidden biclusters	3	5	10	15	20
Nr. columns in biclusters	[5,7]	[6,8]	[6,10]	[6,14]	[6,20]
Nr. rows in biclusters	[10,20]	[15,30]	[20,40]	[40,70]	[60,100]

following a Uniform distribution, $U(1, |\mathcal{L}|)$, and 10 matrices according to a Gaussian distribution, $N(\frac{\mathcal{L}}{2}, \frac{\mathcal{L}}{6})$.

Comparison: We selected 15 state-of-the-art approaches⁹: FABIA method with sparse prior option [67], ISA [70], OPSM [15], CC [29], Samba [119], xMotifs [94], OP-Clustering [78], BicSPAM [63], Bexpa [106], BCPlaid [123] and the PM-based BiModule [96], DeBi [111], RAP [101], BicPAM [62] and BiP [60] biclustering approaches. We used the following software: R packages fabia¹⁰ and biclust¹¹, BicAT [10], Expander¹², (Evo-)Bexpa [106], RAP¹³ and BicPAMS¹⁴. In particular, we adjust BicPAM behavior according to the

⁹ The specified number of biclusters for FABIA, Bexpa, CC, xMotifs and ISA (number of starting points) was the number of hidden biclusters plus 10%: $|\mathcal{H}| \times 1.1$. Note that this specification guides the search, optimistically biasing Fabia Consensus (FC) levels. The default number of iterations for the OPSM method was varied from 10 to 200 iterations. Remaining parameterizations were set by default.

¹⁰ <http://www.bioinf.jku.at/software/fabia/fabia.html>.

¹¹ <http://cran.r-project.org/web/packages/biclust>

¹² <http://acgt.cs.tau.ac.il/expander>.

¹³ <http://www.mybiosoftware.com/rap-association-analysis-approach-biclust.html>.

¹⁴ <https://web.ist.utl.pt/rmch/software/bicpams>.

proposed principles in this work by considering closed patterns, multiple levels of coherency strength ($|\Sigma| \in \{3, 5, 7\}$), an assignment of two items for elements with values near item-boundaries, and merging ($> 70\%$ overlap) and filtering options. The support threshold was incrementally decreased 10% until the area of the discovered biclusters covered at least 5% of the input matrix. Fig. 9 compares the ability of these state-of-the-art approaches to discover planted biclusters with constant coherency on rows. Results confirm the superior performance of PM-based biclustering approaches both in terms of the MS (\mathcal{B}, \mathcal{H}) (correctness) and MS (\mathcal{H}, \mathcal{B}) (completeness) as they provide exhaustive and flexible searches. Superiority is also verified for non-constant models. Fig. 10 compares the performance of biclustering methods prepared to discover shifting-scaling factors when the planted biclusters follow additive and multiplicative models. A closer look to the performance of PM-based biclustering, when multiple levels of coherency strength are considered, is provided in Fig. 11.

Efficiency: Fig. 12 shows the boundaries on efficiency of PM-based biclustering approaches when considering 20.000 rows (magnitude of the human genome). We varied the number of columns,

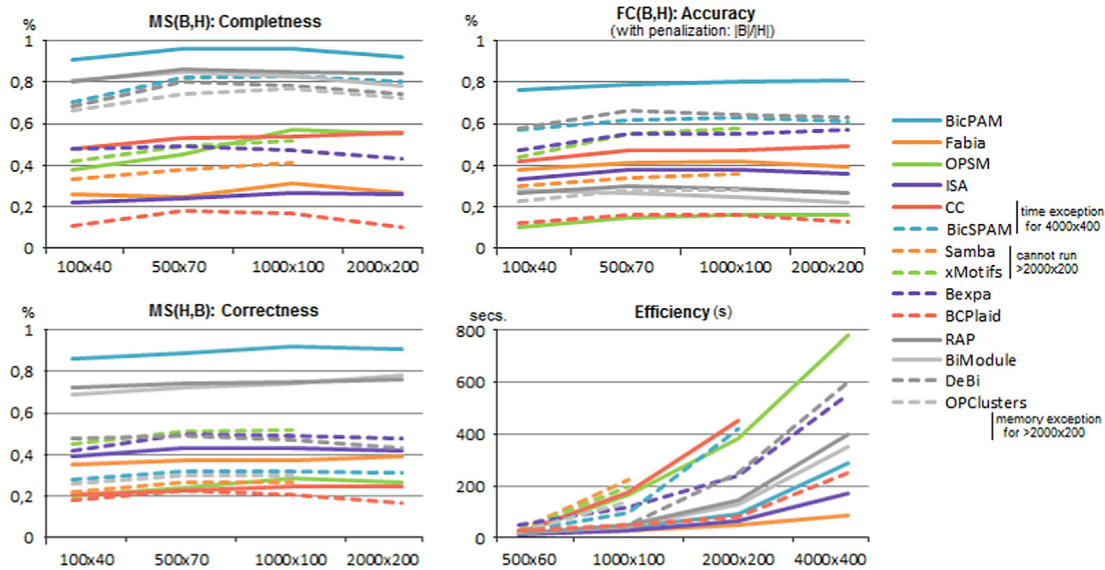


Fig. 9. Comparison of the performance of state-of-the-art biclustering approaches on data settings with varying properties and constant coherencies.

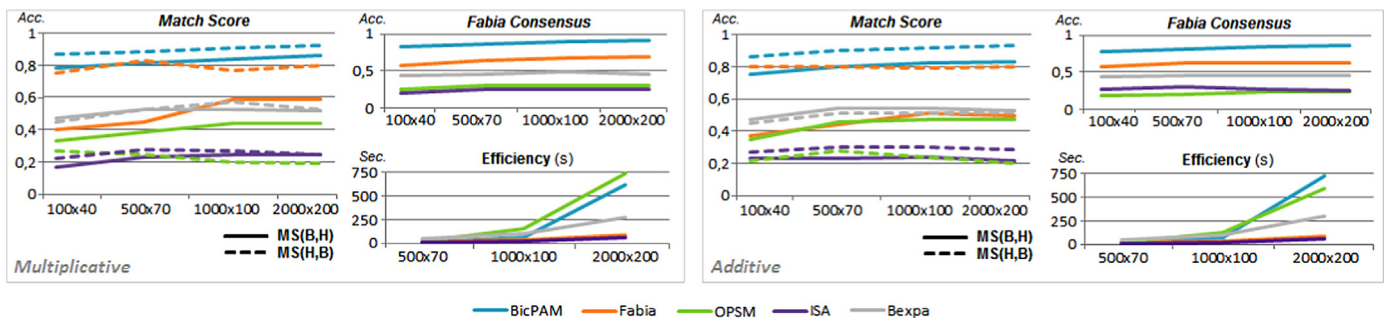


Fig. 10. Comparison of biclustering approaches to recover biclusters with non-constant coherency.

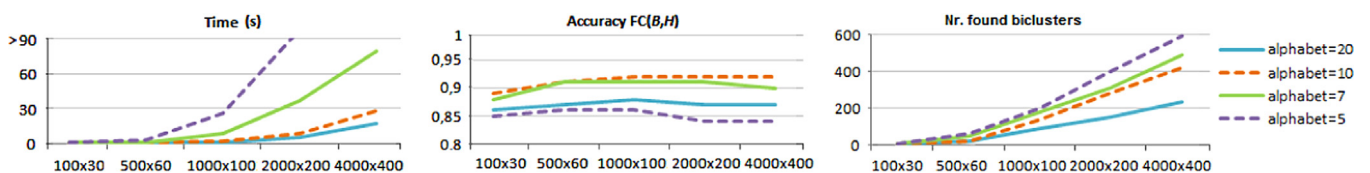


Fig. 11. Performance of PM-based biclustering for data settings with varying coherency strength.

items ($|\mathcal{L}| \in \{5, 7\}$) and considered a simple merging option ($> 70\%$ overlap). We planted 15 biclusters to occupy 2% of the area. Charm [136], an efficient pattern miner to deliver closed patterns (maximal biclusters), was used. Generally, we observe that PM-based biclustering approaches are scalable for these dense and large matrices. Understandably, the number of items has strong impact in efficiency as it defines the density of the itemset database. The scalability of pattern mining methods can be guaranteed for even harder settings by adopting some of the largely researched parallelization, distribution, streaming and error-bounding PM principles [54]. Additionally, hyperclique patterns [52], which require item-pairwise support-similarity, can be also considered to promote the efficiency of the mining procedure.

Impact of mining options: Fig. 13 illustrates the impact of the chosen search and pattern representations (simple, closed, maximal) in the efficiency and MS levels of PM-based biclustering approaches when using a discretization step with 10 items and the

1000 × 100 data setting. The FIM methods were tested using SPMF¹⁵ and F2G [65]. FPGrowth [65] and Eclat [135] are the most competitive choices for small support thresholds, while Apriori [2] is the best option for medium-to-large support levels. Additionally, the use of simple patterns (using FPGrowth [1]) degrades $MS(\mathcal{B}, \mathcal{H})$, while the use of maximal patterns (using CharmMFI [136]) penalizes $MS(\mathcal{H}, \mathcal{B})$ as it discards biclusters with a non-large number of columns (even if they have larger number of rows).

Impact of closing options: We planted additional levels of noise, by varying the amount of noisy elements from 0 to 10%, for the 1000 × 100 setting. Fig. 14 describes the impact of alternative strategies to extend, merge and filter biclusters using Charm. When increasing the planted noise, extension options are critical to maintain attractive levels of accuracy (20pp higher than the baseline option). Fig. 13(b) illustrates

¹⁵ <http://www.philippe-fournier-viger.com/spmf>.

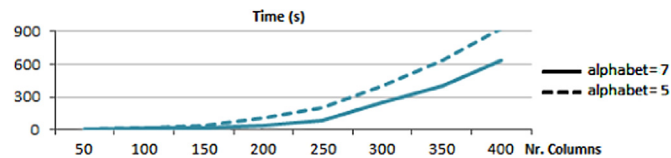


Fig. 12. Efficiency bounds of PM-based biclustering in the absence of scalability principles for datasets with 20,000 rows.

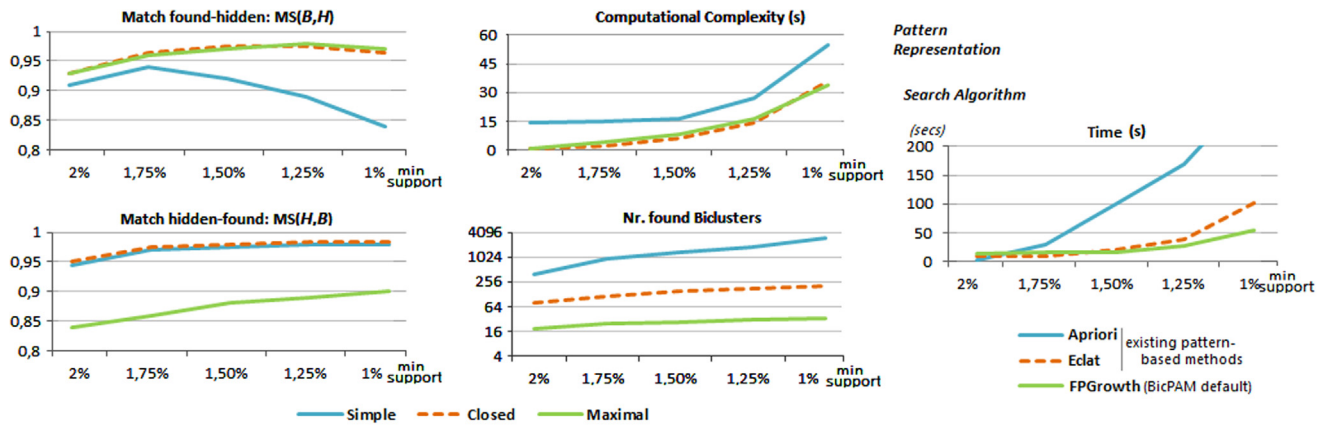


Fig. 13. Comparison of mining searches and pattern representations for the 1000×100 setting.

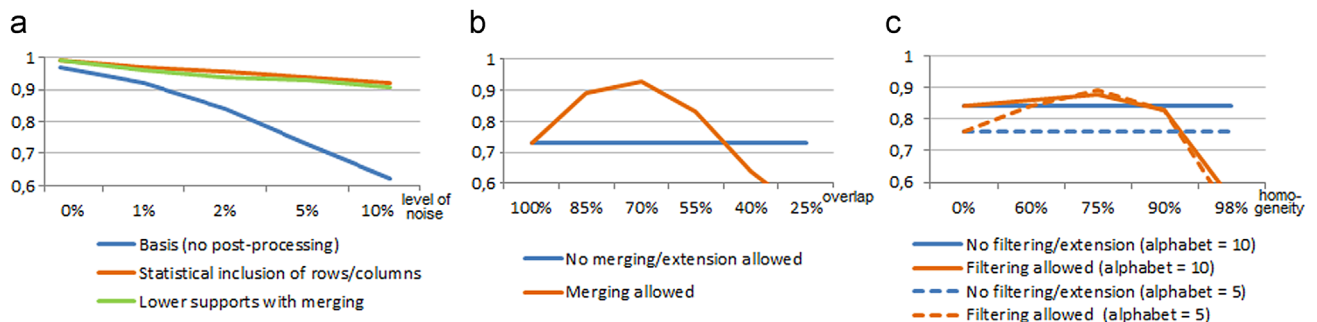


Fig. 14. Impact of extending, merging and filtering options. (a) Extending biclusters for varying levels of noise. (b) Merging for varying overlapping degrees (5% of planted noise). (c) Filtering for varying homogeneity degrees (2% of planted noise).

Table 8

Illustrative set of PM-based biclusters with unique properties and heightened biological relevance.

ID	Dataset	Pattern	Items	Closing options	# Genes	# Conds	# <i>p</i> -values < 0.01	# <i>p</i> -values [0.01,0.05]	Best <i>p</i> -value
B1	dlblc	FAABFFF	A-F	Merging with tight overlapping	83	7	41	21	1.97E-10
B2	dlblc	AAABCA	A-C	Extensions allowed (with tight merging)	153	8	9	1	2.27E-12
B3	hughes	EEEEEE	A-E	Merging allowed	581	6	12	7	1.31E-25
B4	hughes	CCDCBCCC	A-E	Merging with relaxed overlapping	654	10	16	4	1.31E-17

the impact of *merging* biclusters with large overlapping areas assuming a level of planted noise of 5%. When decreasing the overlapping threshold, *MS* levels increase up to a certain threshold (near 70% for this experimental setting). A correct identification of this threshold can lead to significant gains (near 15pp in this setting). Finally, the use of *filtering* strategies to remove rows and columns can also enhance the recovery of the planted biclusters, as it is illustrated in Fig. 14(c). Similarly to the merging option, *MS* increases up to a 75% homogeneity (given by $1 - MSR$ [29]) and decreases above this threshold since the homogeneity criteria becomes too restrictive.

Domain relevance: To assess the relevance of PM-based biclustering in biological settings we used two gene expression datasets: (1) *dlblc* dataset (660 genes, 180 conditions, human genome) [109], and (2) *hughes* dataset (6300 genes, 300 conditions, yeast genome) [74].

For each dataset standard PM-based biclustering (closed FIM) was applied using multiple levels of expression $|\mathcal{L}| \in \{4..7\}$ and different closing options: (1) merging (70% overlap), (2) relaxed merging (55% overlap) with filtering of rows, and (3) tight merging (90% overlap) with extensions on rows that appear in another bicluster sharing a minimum 50% of conditions. The biological relevance of each bicluster was obtained using the Gene Ontology (GO) annotations using the GoToolBox [85]. Table 8 shows an illustrative set of PM-based biclusters with significantly enriched GO terms (after Bonferroni correction). These biclusters could hardly be discovered by peer biclustering methods, since many of them include conditions with multiple degrees of expression (such as B1, B2 and B4). All of them have heightened biological significance as observed by the number of highly enriched terms. Interestingly, we also observe that different

closing options lead to distinct biclusters. Complementary analyzes supporting the biological relevance of PM-based biclustering are provided in [62,60,111].

4. Conclusions

This work provides a structure view on pattern mining-based approaches to biclustering as they are increasingly positioned as the means to perform exhaustive searches under relaxed conditions (flexible structures of biclusters with parameterizable coherency and quality) with heightened efficiency. In this context, this work surveys and integrates the contributions of existing PM-based biclustering approaches, evaluates their performance, and discusses their relevance for pattern recognition applications.

A set of principles were synthesized, covering alternative design options to guide the definition of PM-based biclustering approaches: (1) mining paradigms (including frequent itemset mining, association rule mining, sequential PM, constraint-based PM and structured PM), principles to define support-confidence-correlation metrics, pattern representations (as simple, condensed and approximate), searches, and extensions to consider flexible coherencies; (2) pre-processing options, including strategies to deal with the items-boundary problem when discretization procedures are considered and with noisy and missing elements; and (3) strategies to compose adequate structures of biclusters through extension-merging-filtering steps without the need to adapt the core task. As such, this work introduces a highly-parameterizable environment to design PM-based biclustering approaches, where the behavior can be dynamically defined according to the input dataset and the target biclustering type, structure and quality. In particular, the quality of a target solution can be easily affected through the mining options, such as the confidence of association rules to define the level of tolerated noise; mapping options, such as the number of items (coherency strength) and multi-item assignments; and merging, filtering and extension options based, respectively, on the allowed noise (overlapping degree), dissimilarity and homogeneity of biclusters.

A qualitative comparison of the state-of-the-art PM-based biclustering approaches was provided, as well as initial empirical evidence supporting the accuracy, efficiency and biological relevance of this class of algorithms.

Following this comprehensive work, new research can embrace several promising directions, including: (1) development of new integrative PM-based biclustering approaches; (2) proposal of statistical tests to effectively assess the significance of biclusters with varying coherency and quality; (3) integration of principles from domain-driven PM to incorporate constraints in PM-based biclustering when background knowledge is available; and (4) design of robust classifiers based on discriminative PM-based biclusters.

Conflict of interest

None declared.

Acknowledgments

This work was supported by Fundação para a Ciência e a Tecnologia under the projects UID/CEC/50021/2013 and the PhD grant SFRH/BD/75924/2011 to RH.

References

- [1] Ramesh C. Agarwal, Charu C. Aggarwal, V. Prasad, A tree projection algorithm for generation of frequent item sets, *J. Parallel Distrib. Comput.* 61 (March 3) (2001) 350–371.
- [2] Tomasz Imieliński, Rakesh Agrawal, Arun Swami, Mining association rules between sets of items in large databases, *SIGMOD Rec.* 22 (June (2)) (1993) 207–216.
- [3] H.A. Ahmed, P. Mahanta, D.K. Bhattacharyya, J.K. Kalita, A. Ghosh, Intersected coexpressed subcube miner: An effective triclustering algorithm, in: *WICT*, December 2011, pp. 846–851.
- [4] Faris Alqadah, Joel S. Bader, Rajul Anand, Chandan K. Reddy, Query-based biclustering using formal concept analysis, in: *SDM*, SIAM/Omnipress, Anaheim, California, USA, 2012, pp. 648–659.
- [5] Ronnie Alves, Domingo S. Rodriguez-Baena, Domingo S. Rodriguez-Baena Ronnie Alves, Jesús S. Aguilar-Ruiz, Gene association analysis: a survey of frequent pattern mining from gene expression data, *Brief. Bioinform.* 11 (2) (2010) 210–224.
- [6] I. Assent, R. Krieger, E. Muller, T. Seidl, DUSC: Dimensionality unbiased subspace clustering, in: *ICDM*, 2007.
- [7] Assent Ira, Müller Emmanuel, Krieger Ralph, Jansen Timm, Seidl Thomas, Machine learning and knowledge discovery in databases, in: Daelemans Walter, Goethals Bart, Morik Katharina (Eds.), *Pleiades: Subspace Clustering and Evaluation*, Lecture Notes in Computer Science, 5212, Springer, Berlin Heidelberg, 2008, pp. 666–671, ISBN: 978-3-540-87480-5, http://dx.doi.org/10.1007/978-3-540-87481-2_44.
- [8] Gowtham Atluri, Jeremy Bellay, Gaurav Pandey, Chad Myers, Vipin Kumar, Discovering coherent value bicliques in genetic interaction data, in: *BLOKDD*, 2000.
- [9] R. Rathipriya, K. Thangavel, J. Bagyamani, Binary particle swarm optimization based biclustering of web usage data, *CoRR abs/11080748* (2011).
- [10] Simon Barkow, Stefan Bleuler, Amela. Prelić, Philip Zimmermann, Eckart Zitzler, Bicat: a biclustering analysis toolbox, *Bioinformatics* 22 (May (10)) (2006) 1282–1283.
- [11] Roberto J. Bayardo Jr., Efficiently mining long patterns from databases, *SIGMOD Rec.* 27 (June 2) (1998) 85–93.
- [12] Gürkan Bebek, Jong Yang, Pathfinder: mining signal transduction pathway segments from protein–protein interaction networks, *BMC Bioinform.* 8 (2007).
- [13] Jeremy Bellay, Gowtham Atluri, Tina L. Sing, Kiana Toufighi, Michael Costanzo, Philippe Souza Moraes Ribeiro, Gaurav Pandey, Joshua Baller, Benjamin VanderSluis, Magali Michaut, Sangjo Han, Philip Kim, Grant W. Brown, Brenda J. Andrews, Charles Boone, Vipin Kumar, Chad L. Myers, Putting genetic interactions in context through a global modular decomposition, *Genome Res.* 21 (8) (2011) 1375–1387.
- [14] Jeremy Bellay, et al., Putting genetic interactions in context through a global modular decomposition, *Genome Res.* 21 (8) (2011) 1375–1387.
- [15] Amir Ben-Dor, Benny Chor, Richard Karp, Zohar Yakhini, Discovering local structure in gene expression data: the order-preserving submatrix problem, *RECOMB*, ACM, New York, NY, USA (2002) 49–57.
- [16] G.F. Berriz, O.D. King, B. Bryant, C. Sander, F.P. Roth, Characterizing gene sets with FuncAssociate, *Bioinformatics* 19 (2003) 2502–2504.
- [17] Manuele Bicego, Pietro Lovato, Alberto Ferrarini, Massimo Delledonne, Biclustering of expression microarray data with topic models, in: *IC on Pattern Recognition*, IEEE, 2010, pp. 2728–2731.
- [18] Sylvain Blachon, Ruggero Pensa, Jérémy Besson, Céline Robardet, Jean-Francois Boulicaut, Olivier Gandrillon, Clustering formal concepts to discover biologically relevant knowledge from gene expression data, *In Silico Biol.* 7 (July) (0033) (2007).
- [19] Jean-François Boulicaut, Jérémy Besson, Actionability and formal concepts: a data mining perspective, in: *IC on Formal Concept Analysis*, Springer-Verlag, Berlin, Heidelberg, 2008, pp. 14–31.
- [20] Jean-François Boulicaut, Inductive databases and multiple uses of frequent itemsets: The cInQ approach, in: Rosa Meo, PierLuca Lanzi, and Mika Klemettinen (Eds.), *Database Sup. for Data Mining App.*, LNCS, vol. 2682, Springer, Berlin, Heidelberg, 2004, pp. 1–23.
- [21] Doruk Bozdağ, Ashwin S. Kumar, V. Catalyurek, Comparative analysis of biclustering algorithms, *Bioinformatics and Computational Biology*, ACM, New York, NY, USA (2010) 265–274.
- [22] Douglas Burdick, Manuel Calimlim, Johannes Gehrke, Mafia: a maximal frequent itemset algorithm for transactional databases, in: *ICDE*, IEEE Computer Society, Washington, DC, USA, 2001, pp. 443–452.
- [23] Stanislav Busygin, Nikita Boyko, Panos M. Pardalos, Michael Bewernitz, Georges Ghacibeh, Biclustering EEG data from epileptic patients treated with vagus nerve stimulation, *Data Mining, Systems Analysis and Optimization in Biomedicine*, 953, AIP Publishing, Gainesville, Florida, USA (2007) 220–231.
- [24] Toon Calders, Bart Goethals, Mining all non-derivable frequent itemsets, in: *PKDD*, Springer-Verlag, London, UK, 2002, pp. 74–85.
- [25] Toon Calders, Bart Goethals, Szymon Jaroszewicz, Mining rank-correlated sets of numerical attributes, in: *ACM SIGKDD*, ACM, New York, NY, USA, 2006, pp. 96–105.
- [26] Pedro Carmona-Saez, Monica Chagoyen, Andres Rodriguez, Oswaldo Trelles, JoseM Carazo, Alberto Pascual-Montano, Integrated analysis of gene expression by association rules discovery, *BMC Bioinform.* 7 (2006) 1–16.
- [27] André Valério Carreiro, Artur J. Ferreira, Mrioá AT. Figueiredo, Sara Cordeiro Madeira, Towards a classification approach using meta-biclustering: impact of discretization in the analysis of expression time series, *J. Integr. Bioinf.* 9 (3) (2012) 207.
- [28] Malika Charrad, Mohamed Ben Ahmed, Simultaneous clustering: a survey, *Pattern Recognition and Machine Intelligence*, in: Kuznetsov Sergei O.,

- Mandal Deba P., Kundu Malay K., Pal Sankar K (Eds.), *Simultaneous Clustering: A Survey*, 6744, Springer, Berlin Heidelberg, ISBN 978-3-642-21785-22011, pp. 370–375. http://dx.doi.org/10.1007/978-3-642-21786-9_60.
- [29] Yizong Cheng, George M. Church, *Biclustering of expression data*, *Intelligent Systems for Molecular Biology*, AAAI Press, La Jolla, California, USA (2000) 93–103.
- [30] Recep Colak, Flavia Moser, Jeffrey Shih-Chieh Chu, Alexander Schönhuth, Nansheng Chen, Martin Ester, *Module discovery by exhaustive search for densely connected, co-expressed regions in biomolecular interaction networks*, *PLoS One* 5 (10) (2010) e13348.
- [31] Chad Creighton, Samir Hanash, *Mining gene expression databases for association rules*, *Bioinformatics* 19 (1) (2003) 79–86.
- [32] Phuong Dao, Recep Colak, Raheleh Salari, Flavia Moser, Elai Davicioni, Alexander Schönhuth, Martin Ester, *Inferring cancer subnetwork markers using density-constrained biclustering*, *Bioinformatics* 26 (18) (2010) 625–631.
- [33] P.A.D. de Castro, F.O. de Franga, H.M. Ferreira, F.J. von Zuben, *Applying biclustering to perform collaborative filtering*, *Intell. Syst. Des. Appl.* (October) (2007) 421–426.
- [34] M.C.P. de Souto, D.S.A. de Araujo, I.G. Costa, R. Soares, T.B. Ludermir, A. Schliep, *Comparative study on normalization procedures for cluster analysis of gene expression datasets*, in: *IJCNN*, June, 2008, pp. 2792–2798.
- [35] Zhaohong Deng, Kup-Sze Choi, Fu-Lai Chung, Shitong Wang, *Enhanced soft subspace clustering integrating within-cluster and between-cluster information*, *Pattern Recognit.* 43 (3) (2010) 767–781.
- [36] Inderjit S. Dhillon, Subramanyam Mallela, Dharmendra S. Modha, *Information-theoretic co-clustering*, in: *KDD*, ACM, New York, NY, USA, 2003, pp. 89–98.
- [37] Chris Ding, Ya Zhang, Tao Li, Stephen R. Holbrook, *Biclustering protein complex interactions with a biclique finding algorithm*, in: *ICDM*, IEEE Computer Society, Washington, DC, USA, 2006, pp. 178–187.
- [38] A.R. Donders, G.J. van der Heijden, T. Stijnen, K.G. Moons, *Review: a gentle introduction to imputation of missing values*, *J. Clin. Epidemiol.* 59 (10) (2006) 1087–1091.
- [39] E. Elhamifar, R. Vidal, *Sparse subspace clustering*, in: *Computer Vision and Pattern Recognition*, June 2009, pp. 2790–2797.
- [40] Kemal Eren, Mehmet Deveci, Onur Küçükünç, Ümit V. Çatalyürek, M. Deveci, *A comparative analysis of biclustering algorithms for gene expression data*, *Brief. Bioinf.* 14 (3) (2013) 279–292.
- [41] Nikita Boyko, Neng Fan, Panos M. Pardalos, in: *Wanpracha Chaovalitwongse, Panos M. Pardalos, Petros Xanthopoulos (Eds.), Recent advances of data biclustering with application in computational neuroscience*, *Computational Neuroscience*, 38, Springer Optimization and Its Applications Springer, New York, ISBN 978-0-387-88629-92010, pp. 85–112. http://dx.doi.org/10.1007/978-0-387-88630-5_6.
- [42] Gang Fang, Majda Haznadar, Wen Wang, Haoyu Yu, Michael Steinbach, Timothy R. Church, William S. Oetting, Brian Van Ness, Vipin Kumar, *High-order SNP combinations associated with complex diseases: efficient discovery, statistical power and functional interactions*, *Plos One* 7 (2012).
- [43] Gang Fang, Rui Kuang, Gaurav Pandey, Michael Steinbach, Chad L. Myers, Vipin Kumar, *Subspace differential coexpression analysis: problem definition and a general approach*, in: *Pacific Symposium on Biocomputing*, World Scientific Publishing, 2010, pp. 145–156.
- [44] Paolo Favaro, René Vidal, Paolo Favaro, Avinash Ravichandran, *A closed form solution to robust subspace estimation and clustering*, in: *Computer Vision and Pattern Recognition*, IEEE, Colorado Springs, USA, 2011, pp. 1801–1807.
- [45] Usama M. Fayyad, Keki B. Irani, *Multi-interval discretization of continuous-valued attributes for classification learning*, in: *IJCAI*, 1993, pp. 1022–1029.
- [46] Adelaide Freitas, Wassim Ayadi, Mourad Elloumi, José Luis, Jin-Kao Hao Oliveira, *Survey on biclustering of gene expression data*, *Biological Knowledge Discovery Handbook* (2012) 591–608.
- [47] Guojun Gan, Jianhong Wu, *A convergence theorem for the fuzzy subspace clustering (fsc) algorithm*, *Pattern Recognit.* 41 (6) (2008) 1939–1947.
- [48] Elisabeth Georgii, Lothar Richter, Ulrich Rückert, Stefan Kramer, *Analyzing microarray data using quantitative association rules*, *Bioinformatics* 21 (January 2) (2005) 123–129.
- [49] Gad Getz, Erel Levine, and Eytan Domany, *Coupled two-way clustering analysis of gene microarray data*, *Proc. Natl. Acad. Sci.* 97 (22) (2000) 12079–12084.
- [50] Dmitry Gnatyshak, Dmitry Ignatov, Alexander Semenov, Jonas Poelmans, *Gaining insight in social networks with biclustering and triclustering of LNBIP*, in: *Perspectives in Business Informatics Research*, vol. 128, Springer, Berlin Heidelberg, 2012, pp. 162–171.
- [51] Gösta Grahne, Jianfei Zhu, *Efficiently using prefix-trees in mining frequent itemsets*, in: *FIMI*, vol. 90, 2003.
- [52] Rohit Gupta, Navneet Rao, Vipin Kumar, *Discovery of error-tolerant biclusters from noisy gene expression data*, *BMC Bioinf.* 12 (12) (2011) 1–17.
- [53] E.H. Han, G. Karypis, V. Kumar, *Min-apriori: an algorithm for finding association rules in data with continuous attributes*, Department of Computer Science, University of Minnesota, Minneapolis (1997).
- [54] Jiawei Han, Hong Cheng, Dong Xin, Xifeng Yan, *Frequent pattern mining: current status and future directions*, *Data Min. Knowl. Discov.* 15 (August (1)) (2007) 55–86.
- [55] Jiawei Han, Jian Pei, Guozhu Dong, Ke Wang, *Efficient computation of iceberg cubes with complex measures*, *SIGMOD Rec.* 30 (May (2)) (2001) 1–12.
- [56] Blaise Hanczar, Mohamed Nadif, *Ensemble methods for biclustering tasks*, *Pattern Recognit.* 45 (11) (2012) 3938–3949.
- [57] J.A. Hartigan, *Direct clustering of a data matrix*, *J. Am. Stat. Assoc.* 67 (337) (1972) 123–129.
- [58] Trond Hellem, Bjarte Dysvik, Inge Jonassen, *LImpute: accurate estimation of missing values in microarray data with least squares methods*, *Nucleic Acids Res.* e32+ (February (3)) (2004) 34.
- [59] R. Henriques, C. Antunes, *Learning predictive models from integrated healthcare data: extending pattern-based and generative models to capture temporal and crossattribute dependencies*, in: *System Sciences (HICSS)*, January 2014, pp. 2562–2569.
- [60] R. Henriques, S. Madeira, *Biclustering with flexible plaid models to unravel interactions between biological processes*, in: *IEEE/ACM Trans. Comput. Biol. Bioinf.* 2015 (volume pp), (99), p. 1, <http://dx.doi.org/10.1109/TCBB.2014.2388206>.
- [61] Rui Henriques, Cláudia Antunes, Sara C. Madeira, *Methods for the efficient discovery of large item-indexable sequential patterns*, in: *Lecture Notes in Computer Science*, Springer Int. Pub., 2014, pp. 100–116 http://dx.doi.org/10.1007/978-3-319-08407-7_7.
- [62] Rui Henriques, Sara Madeira, *Bicpam: pattern-based biclustering for biomedical data analysis*, *Algorithms Mol. Biol.* 9 (1) (2014) 27.
- [63] Rui Henriques, Sara Madeira, *Bicspam: flexible biclustering using sequential patterns*, *BMC Bioinf.* 15 (2014) 130.
- [64] Rui Henriques, Sara C. Madeira, Cláudia Antunes, *F2g: efficient discovery of full-patterns*, in: *ECML/PKDD IW on New Frontiers in Mining Complex Patterns*, Prague, 2013.
- [65] Rui Henriques, Sílvia Moura Pina, Cláudia Antunes, *Temporal mining of integrated healthcare data: methods, revealings and implications*, in: *SDM IW on Data Mining for Medicine and Healthcare*, SIAM, Austin, US, 2013, pp. 56–64.
- [66] Sepp Hochreiter, Ulrich Bodenhofer, Martin Heusel, Andreas Mayr, Andreas Mitterecker, Adetayo Kasim, Tatsiana Khamiakova, Suzy Van Sanden, Dan Lin, Willem Talloen, Luc Bijnens, Hinrich W.H. Göhlmann, Ziv Shkedy, Djork-Arné Clevert, *FABIA: factor analysis for bicluster acquisition*, *Bioinformatics* 26 (June (12)) (2010) 1520–1527.
- [67] Qinghua Huang, *A biclustering technique for mining trading rules in stock markets*, in: *Dehuai Zeng (Ed.), Applied Informatics and Communication, of Communications in Computer and Information Science*, vol. 224, Springer, Berlin, Heidelberg, 2011, pp. 16–24.
- [68] Yaochun Huang, Hui Xiong, Weili Wu, Sam Y. Sung, *Mining quantitative maximal hyperclique patterns: a summary of results*, in: *PAKDD*, Springer-Verlag, Berlin, Heidelberg, 2006, pp. 552–556.
- [69] Jan Ihmels, Sven Bergmann, Naama Barkai, *Defining transcription modules using large-scale gene expression data*, *Bioinformatics* 20 (September (13)) (2004) 1993–2003.
- [70] Maurice G. Kendall, *Rank Correlation Methods*, Griffin, London, 1948.
- [71] Adam Kirsch, Michael Mitzenmacher, Andrea Pietracaprina, Geppino Pucci, Eli Upfal, and Fabio Vandin, *An efficient rigorous approach for identifying statistically significant frequent itemsets*, in: *ACM SIGMOD Symposium on Principles of Database Systems*, PODS '09, ACM, New York, NY, USA, 2009, pp. 117–126.
- [72] L. Lazzeroni, A. Owen, *Plaid models for gene expression data*, *Stat. Sin.* 12 (2002) 61–86.
- [73] William Lee, Desiree Tillo, Nicolas Bray, HRandall Morse, Ronald W. Davis, Timothy R. Hughes, Corey Nislow, *A high-resolution atlas of nucleosome occupancy in yeast*, *Nat. Genet.* 39 (September (10)) (2007) 1235–1244.
- [74] Guimei Liu, Jinyan Li, Kelvin Sim, and Limsoon Wong, *Distance based subspace clustering with flexible dimension partitioning*, in: *ICDE*, IEEE, 2007, pp. 1250–1254.
- [75] Guimei Liu, Hongjun Lu, Wenwu Lou, Jeffrey Xu Yu, *On computing, storing and querying frequent patterns*, in: *ACM SIGKDD*, ACM, New York, NY, USA, 2003, pp. 607–612.
- [76] Hongyan Liu, Jiawei Han, Dong Xin, Zheng Shao, *Top-down mining of interesting patterns from very high dimensional data*, in: *ICDE*, IEEE Computer Society, Washington, DC, USA, 2006, p. 114.
- [77] Jinze Liu, Wei Wang, *Op-cluster: clustering by tendency in high dimensional space*, in: *ICDM*, IEEE Computer Society, Washington, DC, USA, Melbourne, Florida, USA, 2003, p. 187.
- [78] Nizar R. Mabroukeh, C.I. Ezeife, *A taxonomy of sequential pattern mining algorithms*, *ACM Comput. Surv.* 43 (December (1)) (2010) 31–341.
- [79] Jamie I. MacPherson, Jonathan N. Dickerson, John W. Pinney, David L. Robertson, *Patterns of HIV-1 protein interaction identify perturbed host-cellular subsystems*, *PLoS Comput. Biol.* 6 (7) (2010) e1000863.
- [80] Sara Madeira, Miguel Nobre Parreira Cacho Teixeira, Isabel Sá-Correia, and Arlindo Oliveira, *Identification of regulatory modules in time series gene expression data using a linear time biclustering algorithm*, *IEEE/ACM Trans. Comput. Biol. Bioinf.* 1 (January) (2010) 153–165.
- [81] Sara C. Madeira, Arlindo L. Oliveira, *Biclustering algorithms for biological data analysis: A survey*, *IEEE/ACM Trans. Comput. Biol. Bioinf.* 1 (January (1)) (2004) 24–45.
- [82] M.A. Mahfouz, M.A. Ismail, *Bidens: iterative density based biclustering algorithm with application to gene expression analysis*, in: *PWASET*, vol. 37 2009, pp. 342–348.
- [83] Kazuhisa Makino, Takeaki Uno, *New algorithms for enumerating all maximal cliques of LNCS*, in: *SWAT*, vol. 3111, Springer, 2004, pp. 260–272.

- [85] David Martin, Christine Brun, Elisabeth Remy, Pierre Mouren, Denis Thieffry, Bernard Jacq, Gotoolbox: functional analysis of gene datasets based on gene ontology, *Genome biology*, BioMed Central Ltd, 5(12), 2014, R101.
- [86] Ricardo Martinez, Claude Pasquier, Nicolas Pasquier, Genminer: Mining informative association rules from genomic data, *Bioinformatics and Biomedicine*, 2007, Nov, 15–22, <http://dx.doi.org/10.1109/BIBM.2007.49>.
- [87] Tara McIntosh, Sanjay Chawla, High confidence rule mining for microarray analysis, *IEEE/ACM Trans. Comput. Biol. Bioinf.* 4 (October (4)) (2007), 611–623.
- [88] Guy W. Mineau, Akshay Bissoon, Robert Godin, Simple pre- and post-pruning techniques for large conceptual clustering structures, *Electron. Trans. Artif. Intell.* 4 (C) (2000) 1–20.
- [89] Sushmita Mitra, Haider Banka, Multi-objective evolutionary biclustering of gene expression data, *Pattern Recognit.* 39 (December (12)) (2006) 2464–2477.
- [90] Anirban Mukhopadhyay, Ujjwal Maulik, Sanghamitra Bandyopadhyay, A novel biclustering approach to association rule mining for predicting HIV-1 human protein interactions, *PLoS One* 7 (4) (2012) e32289.
- [91] Emmanuel Müller, Ira Assent, Ralph Krieger, Stephan Günemann, Thomas Seidl, Densest: Density estimation for data mining in high dimensional spaces, in: *SDM*, SIAM, 2009, 173–184.
- [92] Emmanuel Müller, Stephan Günemann, Ira Assent, Thomas Seidl, Evaluating clustering in subspace projections of high dimensional data, *VLDB Endow.* 2 (August (1)) (2009) 1270–1281.
- [93] James Munkres, Algorithms for the assignment and transportation problems, *Soc. Ind. Appl. Math.* 5 (1) (1957) 32–38.
- [94] T.M. Murali, Simon Kasif, Extracting conserved gene expression motifs from gene expression data, in: *Pacific Symposium on Biocomputing*, 2003, pp. 77–88.
- [95] Omar Odibat, Chandan K. Reddy, Efficient mining of discriminative clusters from gene expression data, *Knowl. Inf. Syst.* (2013) 1–30.
- [96] Yoshifumi Okada, Wataru Fujibuchi, Paul Horton, A biclustering method for gene expression module discovery using closed itemset enumeration algorithm, *IJPS Trans. Bioinf.* 48 (SIG5) (2007) 39–48.
- [97] Yoshifumi Okada, Kosaku Okubo, Paul Horton, Wataru Fujibuchi, Exhaustive search method of gene expression modules and its application to human tissue data, *IAENG Int. J. Comput. Sci.* 34 (1) (2007) 119–126.
- [98] Patryk Orzechowski, Proximity measures and results validation in biclustering - a survey of LNCS, *Artificial Intelligence and Soft Computing*, vol. 7895, Springer, Berlin Heidelberg (2013) 206–217.
- [99] Feng Pan, Gao Cong, Anthony K.H. Tung, Jiong Yang, Mohammed Javeed Zaki, Carpenter: finding closed patterns in long biological datasets, in: *ACM SIGKDD*, 2003, pp. 637–642.
- [100] Feng Pan, A.K.H. Tung, Gao Cong, Xin Xu, Cobbler: combining column and row enumeration for closed pattern discovery, in: *Scientific and Statistical Database Management*, June 2004, pp. 21–30.
- [101] Gaurav Pandey, Gowtham Atluri, Michael Steinbach, Chad L. Myers, Vipin Kumar, An association analysis approach to biclustering, in: *ACM SIGKDD*, ACM, New York, NY, USA, 2009, pp. 677–686.
- [102] Nicolas Pasquier, Yves Bastide, Rafik Taouil, Lotfi Lakhal, Efficient mining of association rules using closed itemset lattices, *Inf. Syst.* 24 (March (1)) (1999) 25–46.
- [103] Anne Patrikainen, Marina Meila, Comparing subspace clusterings, *IEEE Trans. Knowl. Data Eng.* 18 (July (7)) (2006) 902–916.
- [104] René Peeters., The maximum edge biclique problem is np-complete, *Discrete Appl. Math.* 131 (September (3)) (2003) 651–654.
- [105] Liqing Peng, Junying Zhang, An entropy weighting mixture model for subspace clustering of high-dimensional data, *Pattern Recognit. Lett.* 32 (8) (2011) 1154–1161.
- [106] Beatriz Pontes, Raúl Giraldez, Jesús S Aguilar-Ruiz, Configurable pattern-based evolutionary biclustering of gene expression data, *Algorithms Mol. Biol.* 8(1) (2013) 4.
- [107] Ignacio Ponzoni, Francisco Azuaje, Juan Augusto, David Glass, Inferring adaptive regulation thresholds and association rules from gene expression data through combinatorial optimization learning, *IEEE/ACM Trans. Comput. Biol. Bioinf.* 4 (4) (2007) 624–634.
- [108] Amela Prelic, Stefan Bleuler, Philip Zimmermann, Anja Wille, Peter Bhlmannü, Wilhelm Gruissem, Lars Hennig, Lothar Thiele, Eckart Zitzler, A systematic comparison and evaluation of biclustering methods for gene expression data, *Bioinformatics* 22 (June (9)) (2006) 1122–1129.
- [109] Andreas Rosenwald, George Wright, Wing C. Chan, et al., The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma, *N. Engl. J. Med.* 346 (June 25) (2002) 1937–1947.
- [110] Swarup Roy, K Dhruva Bhattacharyya, KJugal Kalita, Cobi: pattern based co-regulated biclustering of gene expression data, *Pattern Recognit. Lett.* 34 (14) (2013) 1669–1678.
- [111] Akdes Serin, Martin Vingron, Debi: discovering differentially expressed biclusters using a frequent itemset approach, *Algorithms Mol. Biol.* 6 (2011) 1–12.
- [112] Fanhua Shang, L.C. Jiao, Fei Wang, Graph dual regularization non-negative matrix factorization for co-clustering, *Pattern Recognit.* 45 (6) (2012) 2237–2250 (Brain Decoding).
- [113] Qizheng Sheng, Yves Moreau, Bart De Moor, Biclustering microarray data by gibbs sampling, in: *ECCB*, 2003, pp. 196–205.
- [114] Kelvin Sim, Vivekanand Gopalkrishnan, Arthur Zimek, Gao Cong, A survey on enhanced subspace clustering, *Data Min. Knowl. Discov.* 26 (2) (2013) 332–397.
- [115] Michael Steinbach, Pang-Ning Tan, Hui Xiong, Vipin Kumar, Generalizing the notion of support, in: *ACM SIGKDD*, 2004, ACM, New York, NY, USA, pp. 689–694.
- [116] Michael Steinbach, Haoyu Yu, Gang Fang, Vipin Kumar, Using constraints to generate and explore higher order discriminative patterns of LNCS, in: *PAKDD*, vol. 6634, Springer, 2011, pp. 338–350.
- [117] Pang-Ning Tan, Vipin Kumar, Jaideep Srivastava, Selecting the right interestingness measure for association patterns, in: *ACM SIGKDD*, ACM, Edmonton, Alberta, Canada, 2002, pp. 32–41.
- [118] A. Tanay, R. Sharan, R. Shamir, Biclustering algorithms: a survey, in: *Handbook of Computational Molecular Biology*, 2004.
- [119] Amos Tanay, Roded Sharan, Ron Shamir, Discovering statistically significant biclusters in gene expression data, in: *ISMB*, 2002, pp. 136–144.
- [120] Chun Tang, Li Zhang, Murali Ramanathan, Aidong Zhang, Interrelated two-way clustering: an unsupervised approach for gene expression data analysis, in: *BIBE*, Washington, DC, USA, 2001, IEEE CS, p. 41.
- [121] Teixeira, Miguel Cacho and Monteiro, Pedro Tiago and Guerreiro, Joana Fernandes and Gonçalves, Joana Pinho and Mira, Nuno Pereira and dos Santos, Sandra Costa and Cabrito, Tânia Rodrigues and Palma, Margarida and Costa, Catarina and Francisco, Alexandre Paulo and others. The YEASTRACT database: an upgraded information system for the analysis of gene and genomic transcription regulation in *Saccharomyces cerevisiae*, *Nucleic Acids Res.* (database issue) (2014).
- [122] O. Troyanskaya, M. Cantor, G. Sherlock, P. Brown, T. Hastie, R. Tibshirani, D. Botstein, R.B. Altman, Missing value estimation methods for DNA microarrays, *Bioinformatics* 17 (6) (2001) 520–525. <http://dx.doi.org/10.1093/bioinformatics/17.6.520>.
- [123] Heather Turner, Trevor Bailey, Wojtek Krzanowski, Improved biclustering of microarray data demonstrated through systematic performance tests, *Comput. Stat. Data Anal.* 48 (2) (2005), 235–254.
- [124] Miranda van Uiter, Wouter Meuleman, Lodewyk Wessels, Biclustering sparse binary genomic data, *J. Comput. Biol.* 15 (10) (2008) 1329–1345.
- [125] Takeaki Uno, Masashi Kiyomi, Hiroki Arimura, Lcm ver.3: collaboration of array, bitmap and prefix tree for frequent itemset mining, in: *OSDM*, ACM, New York, NY, USA, 2005.
- [126] Haixun Wang, Wei Wang, Jiong Yang, Philip S. Yu, Clustering by pattern similarity in large data sets, in: *SIGMOD*, ACM, New York, NY, USA, 2002, pp. 394–405.
- [127] Shu Wang, Robin R Gutell, Daniel P Miranker, Biclustering as a method for rna local multiple sequence alignment, *Bioinformatics* 23 (24) (2007) 3289–3296.
- [128] Zhiguan Wang, Chi Wai Yu, Ray C.C. Cheung, Hong Yan, Hypergraph based geometric biclustering algorithm, *Pattern Recognit. Lett.* 33 (12) (2012) 1656–1665.
- [129] Takashi Washio, Hiroshi Motoda, State of the art of graph-based data mining, *SIGKDD Explor. Newslett.* 5 (July (1)) (2003) 59–68.
- [130] Peter H. Westfall, S. Stanley Young, Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment, John Wiley & Sons, 1993.
- [131] Hu Xia, Jian Zhuang, Dehong Yu, Novel soft subspace clustering with multi-objective evolutionary approach for high-dimensional data, *Pattern Recognit.* 46 (9) (2013) 2562–2575. <http://dx.doi.org/10.1016/j.patcog.2013.02.005>.
- [132] Dong Xin, Zheng Shao, Jiawei Han, Hongyan Liu, C-cubing: efficient computation of closed cubes by aggregation-based checking, in: *ICDE*, IEEE Computer Society, 2006, p. 4.
- [133] Hui Xiong, Xiao-Feng Heb, Chris Ding, Ya Zhang, Vipin Kumar, Stephen R Holbrook, Identification of functional modules in protein complexes via hyperclique pattern discovery, in: *Pacific Symposium on Biocomputing*, 2005.
- [134] Hui Xiong, Pang-Ning Tan, Vipin Kumar, Hyperclique pattern discovery, *Data Min. Knowl. Discov.* 13 (2) (2006) 219–242.
- [135] Mohammed J. Zaki, Karam Gouda, Fast vertical mining using diffsets, in: *ACM SIGKDD*, ACM, New York, NY, USA, 2003, pp. 326–335.
- [136] Mohammed J. Zaki, Ching J. Hsiao, CHARM: An Efficient Algorithm for Closed Itemset Mining.
- [137] Hongya Zhao, Kwok Leung Chan, Lee-Ming Cheng, L. Cheng, Hong Yan, A probabilistic relaxation labeling framework for reducing the noise effect in geometric biclustering of gene expression data, *Pattern Recognit.* 42 (11) (2009) 2578–2588. <http://dx.doi.org/10.1016/j.patcog.2009.03.016>.
- [138] Feida Zhu, Xifeng Yan, Jiawei Han, P.S. Yu, Hong Cheng, Mining colossal frequent patterns by core pattern fusion, in: *ICDE*, April 2007, pp. 706–715.

Rui Henriques received a M.Sc. degree in computer science and engineering from Instituto Superior Técnico (IST), Universidade de Lisboa. He is developing his Ph.D. studies in the field of learning from high-dimensional and structured data at IST and INESC-ID. He had received distinctions for his academic achievements by IST between 2006 and 2008, and a National Award for his merits by Caixa Geral de Depósitos, in 2009. He has also been a Business Analyst at McKinsey with wide exposure to real-life projects.

Cláudia Antunes received her Ph.D. from Instituto Superior Técnico (IST, University of Lisbon, Portugal) in the domain of data mining and machine learning, proposing new methods to deal with temporal data, in particular for mining event sequential patterns. She is currently a Professor at DEI department at IST and the scientific coordinator of two projects funded by FCT in the areas of domain-driven data mining and educational data mining. Cláudia has been working on methods for general pattern mining, from transactional to structured data. Her main interests are centered on mining complex knowledge from complex data, with emphasis on the incorporation of background knowledge in the pattern mining process.

Sara C. Madeira received a (5-year) B.Sc. degree in computer science from the University of Beira Interior, Covilhã, Portugal, in 2000, and the M.Sc. and Ph.D. degrees in computer science and engineering (CSE) at Instituto Superior Técnico (IST), Technical University of Lisbon, in 2002 and 2008. She is currently an Assistant Professor, at the CSE department at IST, and a Senior Researcher at INESC-ID, Lisbon. Her research interests include algorithms and data structures, data mining, machine learning, bioinformatics and medical informatics.