Similarity Measures for Comparing Biclusterings

Danilo Horta and Ricardo J.G.B. Campello

Abstract—The comparison of ordinary partitions of a set of objects is well established in the clustering literature, which comprehends several studies on the analysis of the properties of similarity measures for comparing partitions. However, similarity measures for clusterings are not readily applicable to biclusterings, since each bicluster is a tuple of two sets (of rows and columns), whereas a cluster is only a single set (of rows). Some biclustering similarity measures have been defined as minor contributions in papers which primarily report on proposals and evaluation of biclustering algorithms or comparative analyses of biclustering algorithms. The consequence is that some desirable properties of such measures have been overlooked in the literature. We review 14 biclustering similarity measures. We define eight desirable properties of a biclustering measure, discuss their importance, and prove which properties each of the reviewed measures has. We show examples drawn and inspired from important studies in which several biclustering measures convey misleading evaluations due to the absence of one or more of the discussed properties. We also advocate the use of a more general comparison approach that is based on the idea of transforming the original problem of comparing biclusterings into an equivalent problem of comparing clustering partitions with overlapping clusters.

Index Terms—Biclustering similarity measure, gene expression, external evaluation, validity index

1 Introduction

Gene expression data are the product of microarray experiments, in which the expression levels of typically thousands of genes are recorded under varying conditions (e.g., organ tissues, blood samples, and phases of cell cycle) [1]. These data are usually represented by a data matrix $A \in \mathbb{R}^{n \cdot p}$, where rows represent genes and columns represent conditions. Hartigan [2] first presented algorithms capable of simultaneously clustering both rows and columns of a data matrix. Later, Mirkin [3] defined this new type of data clustering as biclustering (also called co-clustering or two-mode/two-way clustering). However, this type of clustering algorithm started to draw the scientific community's attention only after the work of Cheng and Church [4]. The biclustering paradigm has become popular in the gene expression field, as a set of genes will rarely be similar to each other under all investigated conditions, and vice versa [4], [5], [6], [7], [8], [9], [10]. To a lesser degree, the biclustering approach has also gained attention in the text, web-log, and market-basket analysis fields [11]. In text analysis, for example, one may wish to find similar documents and their interplay with word clusters.

Since 2000, researchers have developed dozens of biclustering algorithms (surveys in [6], [9], [12], [13]) for gene expression. The proposition of a new algorithm is usually accompanied by a comparative study that includes other biclustering algorithms. Four approaches have been used to evaluate the efficacy of the proposed algorithms in these experiments. The first depends on biological analyses and interpretations by human experts [8], who rely on visualization methods (e.g., parallel coordinate plots and heat maps of the data matrix) and previous knowledge about genes and conditions [14], [15], [16], [17], [18]. This method is frequently accompanied by other approaches due to its subjective nature and is impractical when several algorithms are compared [8].

A more objective and popular approach consists in comparing solutions by their biological significance [13], [19], [20]. For example, one can apply the algorithms to real data sets whose genes are annotated in the Gene Ontology database [21] and then perform an enrichment analysis, which will provide p-values indicating the degree of randomness of the biclusters found. Such an analysis is appealing, but it does not consider the clustered columns (conditions), and it cannot be used in solutions derived from synthetic data sets.

The third method of comparison consists in using indices of internal evaluation [22], [23], [24], which are capable of assessing solutions using only information inherent to the data set. Cheng and Church [4] proposed the mean squared residue that measures the goodness of the pattern found in the gene expression matrix.[1] Internal indices consider both the gene and condition dimensions, therefore the performance of a biclustering algorithm can be fully assessed. However, internal indices make stringent assumptions about the patterns that a bicluster should have, but the gene expression patterns a biological process may exhibit is still an open question.

An external evaluation can be performed when a reference solution is known (a ground truth, e.g., in experiments with synthetic data sets). A similarity measure can then be used to directly compare the found solutions with the reference one [13], [19], [20], and no assumption about gene expression patterns has to be made. It has been mentioned in [8] that an external evaluation is preferable to assess an algorithm in a given data set, whereas an internal evaluation can be performed to investigate

• *The authors are with the Instituto de Ciências Matemáticas e de Computação – ICMC. Universidade de São Paulo – Campus de São Carlos. Caixa Postal 668, 13560-970, São Carlos-SP, Brazil. E-mail: {horta, campello}@icmc.usp.br.*

Manuscript received 27 Feb. 2013; revised 22 Nov. 2013; accepted 7 May 2014. Date of publication 15 May 2014; date of current version 2 Oct. 2014. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below. Digital Object Identifier no. 10.1109/TCBB.2014.2325016

1. Aguilar-Ruiz [25] carried out an in-depth analysis of the mean square residue, identifying some of its drawbacks.

why a particular method does not perform well. Similarity measures are thus an important tool for comparative studies [8], [13], [26], and an analysis of their properties would be valuable.

The remainder of this paper is organized as follows. Section 2 reviews the related work and Section 3 establishes a common background our paper relies on. Section 4 reviews 14 measures for biclusterings comparison. Section 5 advocates the use of data matrix entries as objects to be clustered, transforming biclusterings into overlapping (soft) clusterings. Section 6 proposes eight properties that a measure for comparing biclusterings should have, discusses why they are important, and proves (by referring to the Appendix, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TCBB.2014.2325016) which properties each measure has. Section 7 provides some examples from comparative studies in which several measures convey misleading evaluations due to the absence of one or more of the discussed properties. The computational performance and memory footprint of the two biclustering measures we end up recommending are assessed in Section 7. Section 8 addresses the conclusions. The Appendix, available in the online supplemental material, defines two soft clustering measures adopted in our analysis, presents propositions and proofs, points to web pages having the implementations of the used biclustering algorithms, and describes the configuration of the biclustering algorithms.

2 Related Work

Meila [27] proposed 12 properties of measures for hard clusterings, discussed their importance, introduced the variation of information (VI) measure, and analyzed it along with some other popular measures. Meila [28], [29] also showed that some clustering measures can be completely characterized by a set of instructive axioms and used lattice graph as the mathematical tool in which the space of hard clusterings can be represented and studied.

Prelić et al. [8] briefly discussed the existing methods for comparing biclustering algorithms and introduced the bicluster relevance and bicluster recovery scores.

Patrikainen and Meila [30] proposed the first framework for comparing subspace clusterings. Most of the article was dedicated to the special case of axis-aligned subspace clusterings, in which each cluster is associated with a subset of attributes. Although biclustering and axis-aligned subspace clustering algorithms usually search for distinct types of structures in data, both produce the same type of clustering solution, which means that the techniques discussed in [30] can also be applied when comparing biclusterings. The same article proposed a set of desirable properties for comparing non-overlapping subspace clustering and analyzed some measures in terms of these properties.

Santamaría et al. [31] reviewed some internal, external, and relative validation indices for biclustering. Rosenberg and Hirschberg [32] highlighted the importance of two usually conflicting clustering aspects (homogeneity and completeness).

Amigó et al. [33] proposed four properties of measures for hard clusterings, including the homogeneity and completeness aspects drawn from [32]. The intuition behind these properties were validated in an experiment involving human assessments and compared with other properties in the literature. Lee et al. [24], [34] reviewed several measures for biclusterings and proposed two new ones.

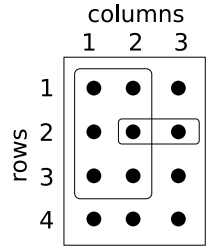


Fig. 1. Data matrix $\mathrm{A} \in \mathbb{R}^{4\cdot 3}$ biclustered into two biclusters.

3 Clustering Background

Let $O \triangleq \{\tilde{o}_1, \tilde{o}_2, \ldots, \tilde{o}_n\}$ be a set of objects. A *hard clustering* of O can be represented by a collection $P \triangleq \{P_1, P_2, \ldots, P_k\}$ of k subsets P_i (clusters), such that their union gives O, there is no empty set, and they are pairwise disjoint (i.e., $P_i \cap P_l = \emptyset$ for $i \neq l$). We call P a *soft clustering* if the last constraint (being pairwise disjoint) is removed [35], [36]. In traditional clustering analysis, objects in O usually represent the rows of a data matrix $\mathrm{A} \in \mathbb{R}^{n\cdot p}$, such that $\tilde{o}_j$ corresponds to the object represented by the jth row of A.

Let $R \triangleq \{1, 2, \ldots, n\}$ and $C \triangleq \{1, 2, \ldots, p\}$ be the sets of indices denoting rows (e.g., genes or documents) and columns (e.g., conditions or words), respectively, of data matrix A. In biclustering analysis, *bicluster* $B_i \triangleq (B_i^{\mathrm{r}}, B_i^{\mathrm{c}})$ is a tuple of two nonempty sets $B_i^{\mathrm{r}} \subset R$ and $B_i^{\mathrm{c}} \subset C$. A collection $B \triangleq \{B_1, B_2, \ldots, B_k\}$ of biclusters forms a *biclustering* of the data represented by A. Consider the biclustering represented in Fig. 1. Using the established notation, we have two biclusters $B_1 = (\{1, 2, 3\}, \{1, 2\})$ and $B_2 = (\{2\}, \{2, 3\})$. The set $B = \{B_1, B_2\}$ represents the corresponding biclustering. In this context, objects in O denote data matrix entries (i.e., row-column pairs).

Some biclustering definitions impose other conditions, such as no overlap between rows and between columns ($B_i^{\mathrm{r}} \cap B_l^{\mathrm{r}} = B_i^{\mathrm{c}} \cap B_l^{\mathrm{c}} = \emptyset$ for every $i \neq l$) and the requirement that $\cup_{i=1}^{k} B_i^{\mathrm{r}} = R$ and $\cup_{i=1}^{k} B_i^{\mathrm{c}} = C$ [6]. These assumptions are too restrictive and are not made in the present paper. In fact, let us consider, e.g., gene expression data, which typically contain thousands of genes or possibly the entire genome of an organism. Some genes will likely not participate in any biological process under the monitored conditions, which violates $\cup_{i=1}^{k} B_i^{\mathrm{r}} = R$. Moreover, in a transcriptomic data set multiple genetic pathways may be active under one condition, and a gene may participate in different genetic pathways under different conditions, which violates $B_i^{\mathrm{r}} \cap B_l^{\mathrm{r}} = \emptyset \, \forall i \neq l$. To detect these gene interactions, the biclusters must overlap [37]. The same applies to text documents described by bags of words, where a document may belong to different categories depending on the words considered. For these reasons, we have adopted here the most general form of biclustering definition.

We say that biclusters B_i and B_l are *equivalent*, $B_i \equiv B_l$, iff[2] B_i and B_l are constituted by the same rows and columns. Biclusters B_i and B_l in a solution are *equal* iff $i = l$. We say that $B \triangleq \{B_i\}_{i=1}^{k}$ and $\dot{B} \triangleq \{\dot{B}_i\}_{i=1}^{q}$ are *equivalent biclusterings*, $B \equiv \dot{B}$, iff $k = q$ and there is a bijection[3] $\{(t_i, y_i)\}_{i=1}^{k}$ for which $B_{t_i} \equiv \dot{B}_{y_i}$ for all i. Note that a solution may have biclusters consisting of one row and one column. We classify such a biclustering as *degenerate* mainly for two reasons: (i) this type of solution is hardly found in real tasks and (ii) some measures have certain properties only in the presence of non-degenerate solutions. Finally, we say that two biclusters B_i and B_l *overlap* iff $B_i^{\mathrm{r}} \times B_i^{\mathrm{c}} \cap B_l^{\mathrm{r}} \times B_l^{\mathrm{c}} \neq \emptyset$ (i.e., their corresponding submatrices in A overlap), and a solution having such biclusters is called *overlapping biclustering*.

4 Current Similarity Measures for Biclusterings

We assume that $B \triangleq \{B_i\}_{i=1}^{k}$ and $\dot{B} \triangleq \{\dot{B}_i\}_{i=1}^{q}$ are, respectively, the found and reference biclusterings. Dissimilarity measures were transformed into similarity measures for comparison purposes.

Before reviewing the measures that will be analyzed in this paper, it is worth mentioning that Turner et al. [38] adapted the F-measure to biclustering, but they used a concept from a specific model of biclustering (plaid model) to establish the correspondence between the found and reference biclusters that severely narrows its applicability. For this reason, we will not include this measure in our study.

4.1 Measures $\mathbb{S}_{\mathrm{prel}}$ and $\mathbb{S}_{\mathrm{prec}}$

Prelić et al. [8] defined two measures that consider only the gene dimension, categorizing them as measures for clusterings comparison. The overall match scores, which consider both gene and condition dimensions, were proposed in their supplementary material, available online. Let

$$S_{\mathrm{r}}(B, \dot{B}) \triangleq \frac{1}{k}\sum_{i=1}^{k} \max_{l \in \mathbb{N}_{1,q}} \left\{ \frac{|B_i^{\mathrm{r}} \cap \dot{B}_l^{\mathrm{r}}|}{|B_i^{\mathrm{r}} \cup \dot{B}_l^{\mathrm{r}}|} \right\} \text{ and} \tag{1}$$

$$S_{\mathrm{c}}(B, \dot{B}) \triangleq \frac{1}{k}\sum_{i=1}^{k} \max_{l \in \mathbb{N}_{1,q}} \left\{ \frac{|B_i^{\mathrm{c}} \cap \dot{B}_l^{\mathrm{c}}|}{|B_i^{\mathrm{c}} \cup \dot{B}_l^{\mathrm{c}}|} \right\} \tag{2}$$

be match scores for rows and columns, respectively. The overall relevance and recovery match scores are

$$\mathbb{S}_{\mathrm{prel}}(B, \dot{B}) \triangleq \sqrt{S_{\mathrm{r}}(B, \dot{B}) \cdot S_{\mathrm{c}}(B, \dot{B})} \quad \text{and}$$
$$\mathbb{S}_{\mathrm{prec}}(B, \dot{B}) \triangleq \mathbb{S}_{\mathrm{prel}}(\dot{B}, B).$$

4.2 Measures $\mathbb{S}_{\mathrm{rnia}}$ and $\mathbb{S}_{\mathrm{ce}}$

Patrikainen and Meila [30] introduced four measures for comparing subspace clusterings. In the following, we define two of them that are theoretically superior according to their analysis and were the only ones used in their experimental study. Let N_{j_1,j_2} and $\dot{\mathrm{N}}_{j_1,j_2}$ be the number of biclusters the matrix entry at the j_1th row and j_2th column belongs to in biclusterings B and $\dot{B}$, respectively. The sizes of union and intersection sets that consider overlapping are

$$|U| \triangleq \sum_{j_1,j_2} \max\{\mathrm{N}_{j_1,j_2}, \dot{\mathrm{N}}_{j_1,j_2}\} \quad \text{and} \tag{3}$$

$$|I| \triangleq \sum_{j_1,j_2} \min\{\mathrm{N}_{j_1,j_2}, \dot{\mathrm{N}}_{j_1,j_2}\}. \tag{4}$$

Let

$$U_B \triangleq \bigcup_{i=1}^{k} B_i^{\mathrm{r}} \times B_i^{\mathrm{c}} \tag{5}$$

be the usual union set of a biclustering B. We have $|U| = |U_B \cup U_{\dot{B}}|$ and $|I| = |U_B \cap U_{\dot{B}}|$ for non-overlapping biclusterings B and $\dot{B}$. The relative non-intersecting area measure [30] is

$$\mathbb{S}_{\mathrm{rnia}}(B, \dot{B}) \triangleq 1 - \frac{|U| - |I|}{|U|} = \frac{|I|}{|U|}.$$

Let $\{(t_i, y_i)\}_{i=1}^{\min\{k,q\}}$ be a unique relation[4] that maximizes

$$d_{\max} \triangleq \sum_{i=1}^{\min\{k,q\}} |B_{t_i}^{\mathrm{r}} \times B_{t_i}^{\mathrm{c}} \cap \dot{B}_{y_i}^{\mathrm{r}} \times \dot{B}_{y_i}^{\mathrm{c}}|. \tag{6}$$

The clustering error [30] is given by

$$\mathbb{S}_{\mathrm{ce}}(B, \dot{B}) \triangleq 1 - \frac{|U| - d_{\max}}{|U|} = \frac{d_{\max}}{|U|}.$$

4.3 Measure $\mathbb{S}_{\mathrm{l\&w}}$

Liu and Wang [39] defined the popular [40], [41], [42] measure

$$\mathbb{S}_{\mathrm{l\&w}}(B, \dot{B}) \triangleq \frac{1}{k}\sum_{i=1}^{k} \max_{l \in \mathbb{N}_{1,q}} \left\{ \frac{|B_i^{\mathrm{r}} \cap \dot{B}_l^{\mathrm{r}}| + |B_i^{\mathrm{c}} \cap \dot{B}_l^{\mathrm{c}}|}{|B_i^{\mathrm{r}} \cup \dot{B}_l^{\mathrm{r}}| + |B_i^{\mathrm{c}} \cup \dot{B}_l^{\mathrm{c}}|} \right\}.$$

4.4 Measure $\mathbb{S}_{\mathrm{stm}}$

Let

$$\mathbb{D}(B_i, \dot{B}_l) \triangleq \frac{2 \cdot |B_i^{\mathrm{r}} \times B_i^{\mathrm{c}} \cap \dot{B}_l^{\mathrm{r}} \times \dot{B}_l^{\mathrm{c}}|}{|B_i^{\mathrm{r}} \times B_i^{\mathrm{c}}| + |\dot{B}_l^{\mathrm{r}} \times \dot{B}_l^{\mathrm{c}}|}$$

be the Dice index [43] applied to B_i and $\dot{B}_l$. Santamaría et al. [31] proposed the measure

$$\mathbb{S}_{\mathrm{stm}}(B, \dot{B}) \triangleq \frac{1}{k}\sum_{i=1}^{k} \max_{l \in \mathbb{N}_{1,q}} \{\mathbb{D}(B_i, \dot{B}_l)\}.$$

4.5 Measures $\mathbb{S}_{\mathrm{wjac}}$ and $\mathbb{S}_{\mathrm{wdic}}$

Let

$$\mathbb{J}(B_i, \dot{B}_l) \triangleq \frac{|B_i^{\mathrm{r}} \times B_i^{\mathrm{c}} \cap \dot{B}_l^{\mathrm{r}} \times \dot{B}_l^{\mathrm{c}}|}{|B_i^{\mathrm{r}} \times B_i^{\mathrm{c}} \cup \dot{B}_l^{\mathrm{r}} \times \dot{B}_l^{\mathrm{c}}|}$$

2. We use "iff" as a shorthand for "if and only if".

3. This bijection is between $\mathbb{N}_{1,k}$ and $\mathbb{N}_{1,q}$. We will henceforth omit this detail in similar contexts.

4. By unique relation we mean left-unique, right-unique relation. For example, $\{(1,3),(4,2)\}$ is a unique relation between $\mathbb{N}_{1,4}$ and $\mathbb{N}_{1,4}$, but $\{(1,3),(4,3)\}$ is not.

be the Jaccard index [44] applied to B_i and $\dot{B}_l$. The measures proposed by Lee et al. [34] account for the size of the biclusters:

$$\mathbb{S}_{\text{wjac}}(B,\dot{B}) \triangleq \frac{\sum_{i=1}^{k} |B_i^{\text{r}} \times B_i^{\text{c}}| \cdot \max_{l\in\mathbb{N}_{1,q}}\{\mathbb{J}(B_i,\dot{B}_l)\}}{\sum_{i=1}^{k} |B_i^{\text{r}} \times B_i^{\text{c}}|} \text{ and } \quad (7)$$

$$\mathbb{S}_{\text{wdic}}(B,\dot{B}) \triangleq \frac{\sum_{i=1}^{k} |B_i^{\text{r}} \times B_i^{\text{c}}| \cdot \max_{l\in\mathbb{N}_{1,q}}\{\mathbb{D}(B_i,\dot{B}_l)\}}{\sum_{i=1}^{k} |B_i^{\text{r}} \times B_i^{\text{c}}|}. \quad (8)$$

The $\mathbb{S}_{\text{wdic}}$ measure differs from $\mathbb{S}_{\text{stm}}$ as the former assigns more weight to the evaluation of larger biclusters.

4.6 Measure $\mathbb{S}_{\text{fabi}}$

Hochreiter et al. [37] stated that previous measures designed specifically for biclusterings neither account for overlapping biclusters nor consider the number of biclusters in the found and reference solutions. As for $\mathbb{S}_{\text{ce}}$, let $\{(t_i,y_i)\}_{i=1}^{\min\{k,q\}}$ be a unique relation that maximizes $\sum_{i=1}^{\min\{k,q\}} \mathbb{J}(B_{t_i},\dot{B}_{y_i})$. The fabia measure is

$$\mathbb{S}_{\text{fabi}}(B,\dot{B}) \triangleq \frac{\sum_{i=1}^{\min\{k,q\}} \mathbb{J}(B_{t_i},\dot{B}_{y_i})}{\max\{k,q\}}. \quad (9)$$

4.7 Measures $\mathbb{S}_{\text{u}}$ and $\mathbb{S}_{\text{e}}$

Bozdag et al. [26] defined the following two measures:

$$\mathbb{S}_{\text{u}}(B,\dot{B}) \triangleq 1 - \frac{|U_{\dot{B}}| - |U_B \cap U_{\dot{B}}|}{|U_{\dot{B}}|} = \frac{|U_B \cap U_{\dot{B}}|}{|U_{\dot{B}}|} \text{ and } \quad (10)$$
$$\mathbb{S}_{\text{e}}(B,\dot{B}) \triangleq \mathbb{S}_{\text{u}}(\dot{B},B).$$

The first is concerned with the uncovered portion of the reference biclustering and the second is concerned with the extra portion of the found biclustering.

Ayadi et al. [45] used the measures

$$\mathbb{S}_{\text{sh}}(B,\dot{B}) \triangleq \frac{|U_B \cap U_{\dot{B}}|}{|U_{\dot{B}}|} \text{ and}$$
$$\mathbb{S}_{\text{nsh}}(B,\dot{B}) \triangleq 1 - \frac{|U_B - (U_B \cap U_{\dot{B}})|}{|U_{\dot{B}}|}$$

based on the work of Cano et al. [22]. Note that $\mathbb{S}_{\text{sh}}(B,\dot{B}) = \mathbb{S}_{\text{u}}(B,\dot{B})$ and $\mathbb{S}_{\text{nsh}}$ can assign negative evaluations, whereas $\mathbb{S}_{\text{sh}}$ assume values in $[0,1]$. On the other hand, both $\mathbb{S}_{\text{u}}$ and $\mathbb{S}_{\text{e}}$ assume values in $[0,1]$ and are symmetric in relation to the parameter order. We thus analyze only the $\mathbb{S}_{\text{u}}$ and $\mathbb{S}_{\text{e}}$ measures.

4.8 Measure $\mathbb{S}_{\text{ay}}$

Ayadi et al. [46] proposed the measure

$$\mathbb{S}_{\text{ay}}(B,\dot{B}) \triangleq \frac{1}{k}\sum_{i=1}^{k} \max_{l\in\mathbb{N}_{1,q}} \frac{|B_i^{\text{r}} \cap \dot{B}_l^{\text{r}}||B_i^{\text{c}} \cap \dot{B}_l^{\text{c}}|}{|B_i^{\text{r}} \cup \dot{B}_l^{\text{r}}||B_i^{\text{c}} \cup \dot{B}_l^{\text{c}}|}.$$

Note that $|B_i^{\text{r}} \cap \dot{B}_l^{\text{r}}||B_i^{\text{c}} \cap \dot{B}_l^{\text{c}}| = |B_i^{\text{r}} \times B_i^{\text{c}} \cap \dot{B}_l^{\text{r}} \times \dot{B}_l^{\text{c}}|$, but $|B_i^{\text{r}} \cup \dot{B}_l^{\text{r}}||B_i^{\text{c}} \cup \dot{B}_l^{\text{c}}| \neq |B_i^{\text{r}} \times B_i^{\text{c}} \cup \dot{B}_l^{\text{r}} \times \dot{B}_l^{\text{c}}|$ in general (the former is always greater than or equal to the latter), which makes $\mathbb{S}_{\text{ay}}$ and $\mathbb{S}_{\text{stm}}$ different.

4.9 Measures $\mathbb{S}_{\text{erel}}$ and $\mathbb{S}_{\text{erec}}$

Eren et al. [13] introduced the measures

$$\mathbb{S}_{\text{erel}}(B,\dot{B}) \triangleq \frac{1}{k}\sum_{i=1}^{k} \max_{l\in\mathbb{N}_{1,q}}\{\mathbb{J}(B_i,\dot{B}_l)\} \text{ and}$$
$$\mathbb{S}_{\text{erec}}(B,\dot{B}) \triangleq \mathbb{S}_{\text{erel}}(\dot{B},B).$$

The $\mathbb{S}_{\text{erel}}(B,\dot{B})$ measure computes the relevance of the found biclustering, whereas $\mathbb{S}_{\text{erec}}(B,\dot{B})$ measures the recovery of the reference biclustering.

5 New Approach to Evaluate Biclustering Solutions

Patrikainen and Meila [30] proposed the use of the set of data matrix entries as the base element set, instead of the data set objects, to compare axis-aligned non-overlapping subspace clusterings. They redefined the concept of intersection and union sizes used by some similarity measures for handling the overlapping case, giving rise to the $\mathbb{S}_{\text{rnia}}$ and $\mathbb{S}_{\text{ce}}$ measures defined in Section 4.2 and considered in our theoretical and empirical analysis. We propose a similar approach for representing biclusterings (for both overlapping and non-overlapping cases) in that the set of data matrix entries are used as the base element set, but without relying on the redefinition of intersection and union sizes. This approach consists in representing a biclustering by a soft clustering, which allows taking advantage of measures designed for comparing this type of clusterings.

Consider a data matrix $\mathrm{A} \in \mathbb{R}^{n\cdot p}$. Each entry in A is now an object of a set $O \triangleq \{\tilde{o}_1,\tilde{o}_2,\ldots,\tilde{o}_{n\cdot p}\}$ of row-column pairs, such that $\mathrm{A}_{1,1}$ is represented by $\tilde{o}_1$, $\mathrm{A}_{2,1}$ by $\tilde{o}_2,\ldots,\mathrm{A}_{1,2}$ by $\tilde{o}_{n+1}$, and so on. Precisely, the mapping is defined as

$$\pi(j,s) \triangleq j + n(s-1) \quad \forall j \in \mathbb{N}_{1,n}, \forall s \in \mathbb{N}_{1,p}, \quad (11)$$

where $\tilde{o}_{\pi(j,s)}$ represents the matrix element of the jth row in the sth column (i.e., element $\mathrm{A}_{j,s}$). Any bicluster $B_i \triangleq (B_i^{\text{r}}, B_i^{\text{c}}) \in B$, where $B \triangleq \{B_1,B_2,\ldots,B_k\}$, can be converted into a ordinary cluster P_i. To do that, we define P_i as

$$P_i \triangleq \bigcup_{j\in B_i^{\text{r}}, s\in B_i^{\text{c}}} \{\tilde{o}_{\pi(j,s)}\}. \quad (12)$$

That is, the entries of A biclustered in B_i are clustered in P_i. Performing this transformation for every $i \in \mathbb{N}_{1,k}$ and defining the set $\{P_1,P_2,\ldots,P_k\}$ produces a soft clustering of the row-column pairs (i.e., entries in A). Note that, in principle, some row-column pairs may not be clustered. For example, many of the elements in a gene expression data matrix will not exhibit a pattern [6] and, hopefully, will not be clustered. We can assign each of these noisy elements to singletons (i.e., sets having only one element), as these elements should not be clustered with any other element. Specifically, we define P as an augmented set given by

$$P \triangleq \{P_1,P_2,\ldots,P_k,P_{k+1},\ldots,P_{k+h}\}, \quad (13)$$

where P_i for $i \in \mathbb{N}_{1,k}$ is given by Eq. (12) and the remaining clusters are the singletons corresponding to the non-biclustered entries of A.

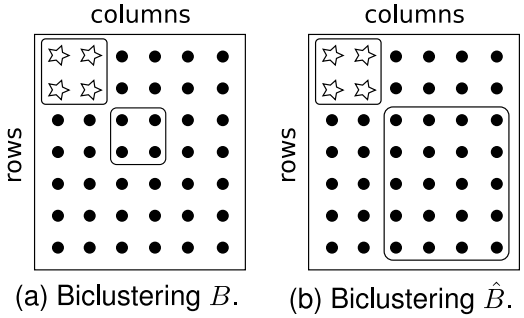


Fig. 2. Two found biclusterings differing only in the size of the noisy biclusters.

TABLE 1
Evaluation of the Biclusterings in Fig. 2

Meas.	Fig. 2a	Fig. 2b	Meas.	Fig. 2a	Fig. 2b
$\mathbb{S}_{\text{prel}}$	0.500	0.500	$\mathbb{S}_{\text{fabi}}$	0.500	0.500
$\mathbb{S}_{\text{prec}}$	1.000	1.000	$\mathbb{S}_{\text{u}}$	1.000	1.000
$\mathbb{S}_{\text{rnia}}$	**0.500**	**0.167**	$\mathbb{S}_{\text{e}}$	**0.500**	**0.167**
$\mathbb{S}_{\text{ce}}$	**0.500**	**0.167**	$\mathbb{S}_{\text{ay}}$	0.500	0.500
$\mathbb{S}_{\text{lw}}$	0.500	0.500	$\mathbb{S}_{\text{erel}}$	0.500	0.500
$\mathbb{S}_{\text{stm}}$	0.500	0.500	$\mathbb{S}_{\text{erec}}$	1.000	1.000
$\mathbb{S}_{\text{wjac}}$	**0.500**	**0.167**	$\mathbb{S}_{\text{csi}}$	**0.500**	**0.031**
$\mathbb{S}_{\text{wdic}}$	**0.500**	**0.167**	$\mathbb{S}_{\text{ebc}}$	**0.963**	**0.708**

As an example, let us consider the biclustering represented in Fig. 1. As discussed above, we have the biclustering $B = \{B_1, B_2\}$, where $B_1 = (\{1,2,3\},\{1,2\})$ and $B_2 = (\{2\},\{2,3\})$. Using the representation given in Eq. (11), we now have the set $O = \{\tilde{o}_1, \tilde{o}_2, \ldots, \tilde{o}_{12}\}$ of row-column pairs. Applying Eqs. (12) and (13) provides $P_1 = \{\tilde{o}_1, \tilde{o}_2, \tilde{o}_3, \tilde{o}_5, \tilde{o}_6, \tilde{o}_7\}$, $P_2 = \{\tilde{o}_6, \tilde{o}_{10}\}$, $P_3 = \{\tilde{o}_4\}$, $P_4 = \{\tilde{o}_8\}$, $P_5 = \{\tilde{o}_9\}$, $P_6 = \{\tilde{o}_{11}\}$, $P_7 = \{\tilde{o}_{12}\}$, and $P = \{P_1, P_2, \ldots, P_7\}$.

After transforming the found and reference biclusterings into soft clusterings P and $\dot{P}$, the final step of the proposed evaluation approach validates P using $\dot{P}$. We selected two measures for soft clusterings that we believe are the most promising ones, according to our experience:

$$\mathbb{S}_{\text{csi}}(B, \dot{B}) \triangleq \text{CSI}(P, \dot{P}) \quad \text{and}$$
$$\mathbb{S}_{\text{ebc}}(B, \dot{B}) \triangleq \text{EBC}(P, \dot{P}),$$

where CSI and EBC are defined by Eqs. (20) and (15) in the Appendix, available in the online supplemental material.

Similarly to biclusters, we say that P_i and P_l are *equivalent*, $P_i \equiv P_l$, iff P_i and P_l have the same objects. Clusters P_i and P_l in a solution are *equal* iff $i = l$. We say that $P \triangleq \{P_i\}_{i=1}^{k}$ and $\dot{P} \triangleq \{\dot{P}_i\}_{i=1}^{q}$ are *equivalent* clusterings, $P \equiv \dot{P}$, iff $k = q$ and there is a bijection $\{(t_i, y_i)\}_{i=1}^{k}$ such that $P_{t_i} \equiv \dot{P}_{y_i}$ for all i. Note that two non-equivalent biclusterings can be transformed into the same soft clustering (Proposition 1 in the Appendix, available in the online supplemental material). However, this is possible only for degenerate solutions (Proposition 2).

6 Theoretical Comparison of Similarity Measures

This section compares the measures discussed so far in terms of eight properties that we consider relevant for evaluating biclusterings. Let

$$\text{I}(B_i, B_l) \triangleq \left(B_i^{\text{r}} \times B_i^{\text{c}}\right) \cap \left(B_l^{\text{r}} \times B_l^{\text{c}}\right)$$

denote intersection between two biclusters.

Definition 1 (Size of Spurious Biclusters). *Let $\{B_{t_i}\}_{i=1}^{x}$ be the set of biclusters in B such that $\text{I}(B_{t_i}, \dot{B}_l) = \emptyset$ for all $l \in \mathbb{N}_{1,q}$ and $i \in \mathbb{N}_{1,x}$. $\{B_{t_i}\}_{i=1}^{x}$ is called the set of spurious biclusters in B. Let $\hat{B}$ be a biclustering equivalent to B, except that one or more spurious biclusters were increased in size and are still spurious. We say that $\mathbb{S}$ is sensitive to the size of spurious biclusters iff $\mathbb{S}(B, \dot{B}) > \mathbb{S}(\hat{B}, \dot{B})$.*

Since noisy entries should not be grouped with other entries, a bicluster containing more noisy entries should be evaluated as lower in quality. Fig. 2 illustrates this case. The stars denote the only bicluster of the reference biclustering $\dot{B}$, and the filled circles represent noisy entries. Figs. 2a and 2b illustrate two biclusterings B and $\hat{B}$ that have two biclusters each. Each solution contains a bicluster composed of noisy entries, but such spurious biclusters have different sizes. Table 1 shows that most of the measures ignore the change in the size of the spurious bicluster.

The union size defined by Eq. (3) increases whenever a spurious bicluster is increased, which explains why the $\mathbb{S}_{\text{rnia}}$ and $\mathbb{S}_{\text{ce}}$ measures are sensitive to the size of spurious biclusters. The $\mathbb{S}_{\text{wjac}}$ and $\mathbb{S}_{\text{wdic}}$ measures are also sensitive because the denominators of Eqs. (7) and (8) increase and the nominators do not change when a spurious bicluster increases. The $\mathbb{S}_{\text{e}}$ measure is clearly sensitive if the domain of biclusterings is restricted to the non-overlapping ones, but not for the general domain as spurious biclusters can be increased without necessarily increasing $|U_B|$ (Eq. (5)). The $\mathbb{S}_{\text{csi}}$ and $\mathbb{S}_{\text{ebc}}$ measures are sensitive, according to Propositions 3 and 4.

Definition 2 (Coverage). *Assume that B has less biclusters than $\dot{B}$ (i.e., $k < q$) and that each bicluster in B is equivalent to a bicluster in $\dot{B}$. Specifically, B is given by a proper subset of the biclusters in $\dot{B}$. Thus, $\mathbb{S}$ is a measure that penalizes solutions for not covering all reference biclusters iff $\mathbb{S}(B, \dot{B}) < 1$.*

Fig. 3 illustrates a case in which the found biclustering does not cover all biclusters of the reference solution (represented by blank shapes), which is clearly undesirable. However, Table 2 shows that half of the measures evaluate B in Fig. 3 as a perfect solution.

Measure $\mathbb{S}_{\text{fabi}}$ does not attain 1 for solutions having different numbers of biclusters, implying that it has the coverage property. Proposition 5 shows that $\mathbb{S}_{\text{rnia}}$ and $\mathbb{S}_{\text{ce}}$ also have the property. Despite their results in Table 2,

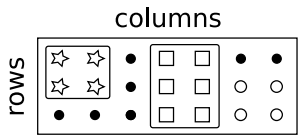


Fig. 3. Example of a biclustering B not covering the entire reference solution $\dot{B}$.

TABLE 2
Evaluation of the Biclusterings in Fig. 3

Meas.	Fig. 3	Meas.	Fig. 3
$\mathbb{S}_{\text{prel}}$	1.000	$\mathbb{S}_{\text{fabi}}$	**0.667**
$\mathbb{S}_{\text{prec}}$	**0.770**	$\mathbb{S}_{\text{u}}$	**0.714**
$\mathbb{S}_{\text{rnia}}$	**0.714**	$\mathbb{S}_{\text{e}}$	1.000
$\mathbb{S}_{\text{ce}}$	**0.714**	$\mathbb{S}_{\text{ay}}$	1.000
$\mathbb{S}_{\text{lw}}$	1.000	$\mathbb{S}_{\text{erel}}$	1.000
$\mathbb{S}_{\text{stm}}$	1.000	$\mathbb{S}_{\text{erec}}$	**0.667**
$\mathbb{S}_{\text{wjac}}$	1.000	$\mathbb{S}_{\text{csi}}$	**0.778**
$\mathbb{S}_{\text{wdic}}$	1.000	$\mathbb{S}_{\text{ebc}}$	**0.923**

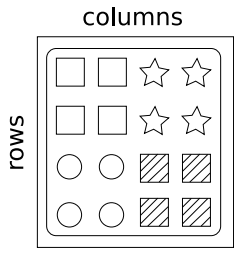


Fig. 4. Found biclustering with one bicluster and reference biclustering with four biclusters.

TABLE 3
Evaluation of the Biclusterings in Fig. 4

Meas.	Fig. 4	Meas.	Fig. 4
$\mathbb{S}_{\text{prel}}$	**0.500**	$\mathbb{S}_{\text{fabi}}$	**0.062**
$\mathbb{S}_{\text{prec}}$	**0.500**	$\mathbb{S}_{\text{u}}$	1.000
$\mathbb{S}_{\text{rnia}}$	1.000	$\mathbb{S}_{\text{e}}$	1.000
$\mathbb{S}_{\text{ce}}$	**0.250**	$\mathbb{S}_{\text{ay}}$	**0.250**
$\mathbb{S}_{\text{lw}}$	**0.500**	$\mathbb{S}_{\text{erel}}$	**0.250**
$\mathbb{S}_{\text{stm}}$	**0.400**	$\mathbb{S}_{\text{erec}}$	**0.250**
$\mathbb{S}_{\text{wjac}}$	**0.250**	$\mathbb{S}_{\text{csi}}$	**0.200**
$\mathbb{S}_{\text{wdic}}$	**0.400**	$\mathbb{S}_{\text{ebc}}$	**0.400**

Proposition 6 shows that $\mathbb{S}_{\text{prec}}$, $\mathbb{S}_{\text{u}}$, and $\mathbb{S}_{\text{erec}}$ do not have the property. Propositions 7 and 8 prove that $\mathbb{S}_{\text{csi}}$ and $\mathbb{S}_{\text{ebc}}$ have the property.

Definition 3 (Non-intersecting Area). *Let B and $\dot{B}$ be two biclusterings, and let S be the matrix elements that are not biclustered by $\dot{B}$. Let $\hat{B}$ be a biclustering that differs from B only by adding elements from S into biclusters of B and/or by creating other biclusters with elements only from S. Measure $\mathbb{S}$ penalizes solutions for non-intersecting area iff $\mathbb{S}(B,\dot{B}) > \mathbb{S}(\hat{B},\dot{B})$.*

The above property is more general than the property with the same name in [30] because S in Definition 3 can have elements biclustered in B. The non-intersecting area property subsumes the intuitive idea that increasing solution B without adding matrix elements from $\dot{B}$ should make the resulting solution less similar to $\dot{B}$.

If a measure penalizes solutions for non-intersecting area, it also has the property given in Definition 1. The only measures that might satisfy Definition 3 are thus $\mathbb{S}_{\text{rnia}}$, $\mathbb{S}_{\text{ce}}$, $\mathbb{S}_{\text{wjac}}$, $\mathbb{S}_{\text{wdic}}$, $\mathbb{S}_{\text{csi}}$, and $\mathbb{S}_{\text{ebc}}$.

Measure $\mathbb{S}_{\text{e}}$ follows the property only in the domain of non-overlapping biclusterings, which is clear from Eq. (10). Propositions 9 and 10 prove that $\mathbb{S}_{\text{rnia}}$ and $\mathbb{S}_{\text{ce}}$ have the discussed property. Measures $\mathbb{S}_{\text{csi}}$ and $\mathbb{S}_{\text{ebc}}$ also have the property, but in the domain of non-degenerate solutions, according to Propositions 13 and 15. Proposition 11 shows a case in which $\mathbb{S}_{\text{wjac}}$ and $\mathbb{S}_{\text{wdic}}$ fail to comply with the property.

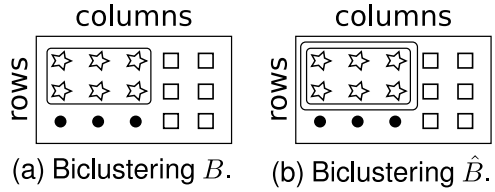


Fig. 5. Example of a repetitive bicluster.

TABLE 4
Evaluation of the Biclusterings in Fig. 5

Meas.	Fig. 5a	Fig. 5b	Meas.	Fig. 5a	Fig. 5b
$\mathbb{S}_{\text{prel}}$	1.000	1.000	$\mathbb{S}_{\text{fabi}}$	0.500	0.500
$\mathbb{S}_{\text{prec}}$	0.577	0.577	$\mathbb{S}_{\text{u}}$	0.600	0.600
$\mathbb{S}_{\text{rnia}}$	**0.600**	**0.375**	$\mathbb{S}_{\text{e}}$	1.000	1.000
$\mathbb{S}_{\text{ce}}$	**0.600**	**0.375**	$\mathbb{S}_{\text{ay}}$	1.000	1.000
$\mathbb{S}_{\text{lw}}$	1.000	1.000	$\mathbb{S}_{\text{erel}}$	1.000	1.000
$\mathbb{S}_{\text{stm}}$	1.000	1.000	$\mathbb{S}_{\text{erec}}$	0.500	0.500
$\mathbb{S}_{\text{wjac}}$	1.000	1.000	$\mathbb{S}_{\text{csi}}$	**0.714**	**0.125**
$\mathbb{S}_{\text{wdic}}$	1.000	1.000	$\mathbb{S}_{\text{ebc}}$	**0.889**	**0.800**

Definition 4 (Multiple Coverage). *Let $B \triangleq \{B_1\}$ and $\dot{B} \triangleq \{\dot{B}_i\}_{i=1}^{q}$ be two biclusterings, such that $q > 1$, $B_1^{\text{r}} \times B_1^{\text{c}} = \bigcup_{i=1}^{q} \dot{B}_i^{\text{r}} \times \dot{B}_i^{\text{c}}$, and $\dot{B}$ has no overlapping biclusters. Thus, $\mathbb{S}$ is a measure that penalizes solutions for multiple biclusters coverage iff $\mathbb{S}(B,\dot{B}) < 1$.*

We should strive to generate biclusterings that cover the entire reference solution, but without mixing matrix entries from different biclusters. The above property formalizes this idea and was proposed in [30].

Fig. 4 shows an example in which a measure has to recognize the difference between the solutions. Table 3 shows that only $\mathbb{S}_{\text{rnia}}$, $\mathbb{S}_{\text{u}}$, and $\mathbb{S}_{\text{e}}$ could not identify the difference. Measure $\mathbb{S}_{\text{fabi}}$ does not attain 1 for solutions having different numbers of biclusters, implying that it has the property given by Definition 4. Patrikainen and Meila [30] showed that $\mathbb{S}_{\text{ce}}$ has the property. Propositions 16, 17, 18, and 19 prove that the remaining measures also have the property.

Definition 5 (Repetitive Biclusters). *Let $\dot{B}$ be a non-overlapping reference biclustering. Let B be a biclustering that has one or more biclusters that perfectly match a bicluster from $\dot{B}$. These are called ideal biclusters. Let $\hat{B}$ be a biclustering equivalent to B, except that there is one or more ideal biclusters in B that were replicated. Thus, $\mathbb{S}$ is a measure that penalizes solutions with repetitive biclusters iff $\mathbb{S}(B,\dot{B}) > \mathbb{S}(\hat{B},\dot{B})$.*

Fig. 5a illustrates a biclustering B that has one ideal bicluster. This bicluster was replicated, giving rise to $\hat{B}$ in Fig. 5b. $\dot{B}$ is defined by blank shapes. Clearly, B is more similar to $\dot{B}$ than $\hat{B}$ is. However, Table 4 shows that most of the measures could not identify this difference. An inspection of Eq. (3) leads to the conclusion that $\mathbb{S}_{\text{rnia}}$ and $\mathbb{S}_{\text{ce}}$ follow the property given by Definition 5. The $\mathbb{S}_{\text{csi}}$ and $\mathbb{S}_{\text{ebc}}$ measures also have the property, according to Propositions 20 and 21.

Though Definition 5 applies to the specific case of identical biclusters, Section 7 shows examples generated by biclustering algorithms in which the measures that do not

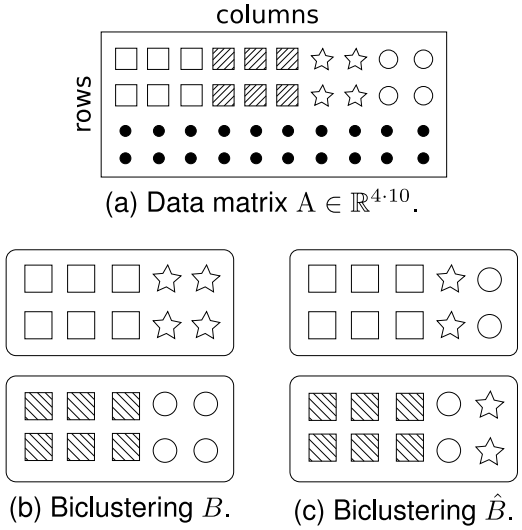


Fig. 6. Difference in homogeneity.

have the property have difficulty in penalizing solutions with several similar biclusters.

Definition 6 (Symmetry). *Measure $\mathbb{S}$ is symmetric iff $\mathbb{S}(B, \dot{B}) = \mathbb{S}(\dot{B}, B)$ for any B and $\dot{B}$.*

The symmetry property is important because it makes the measure more understandable [27]. We refer the reader to Table 6, which indicates the presence or absence of the symmetry property for each measure. The proofs are straightforward and will be omitted.

Definition 7 (Homogeneity). *Let B, $\hat{B}$, and $\dot{B}$ be non-overlapping biclusterings. Let $B_{i_1} \in B$ be a bicluster containing only matrix elements from biclusters $\dot{B}_{\mathrm{ma}(i_1)} \in \dot{B}$ and $\dot{B}_{\mathrm{mi}(i_1)} \in \dot{B}$ such that $|\mathrm{I}(B_{i_1}, \dot{B}_{\mathrm{ma}(i_1)})| > |\mathrm{I}(B_{i_1}, \dot{B}_{\mathrm{mi}(i_1)})|$. In words, $\dot{B}_{\mathrm{ma}(i_1)}$ is the main category in B_{i_1} and the remaining matrix elements come from $\dot{B}_{\mathrm{mi}(i_1)}$. Let B_{i_2}, $\dot{B}_{\mathrm{ma}(i_2)}$, and $\dot{B}_{\mathrm{mi}(i_2)}$ be analogously defined, such that $\mathrm{mi}(i_1) \neq \mathrm{mi}(i_2)$, $\mathrm{mi}(i_1) \neq \mathrm{ma}(i_2)$, and $\mathrm{mi}(i_2) \neq \mathrm{ma}(i_1)$. Let $\hat{B}$ be a biclustering equivalent to B, except that $x > 0$ matrix entries from the minor category in B_{i_1} were swapped for x matrix entries from the minor category in B_{i_2}. Thus, $\mathbb{S}$ is a measure that penalizes less homogeneous solutions iff $\mathbb{S}(B, \dot{B}) \geq \mathbb{S}(\hat{B}, \dot{B})$, such that $\mathbb{S}(B, \dot{B}) = \mathbb{S}(\hat{B}, \dot{B})$ iff $x = |\mathrm{I}(B_{i_1}, \dot{B}_{\mathrm{mi}(i_1)})| = |\mathrm{I}(B_{i_2}, \dot{B}_{\mathrm{mi}(i_2)})|$.*

The homogeneity has already been proposed [29], [32], [33] (discussed in [29] as the "problem of matching") as a desirable feature of measures for comparing clusterings, but without such a formalization. A measure should not be sensitive only to the main category in a found bicluster, but it should also consider how the rest of the found bicluster is organized.

Fig. 6a depicts a reference biclustering $\dot{B}$. Figs. 6b and 6c show an example of biclusterings B and $\hat{B}$ given in Definition 7. B (Fig. 6b) is clearly a less disrupted solution than $\hat{B}$ (Fig. 6c) and, therefore, should be preferred. However, Table 5 shows that most of the measures did not detect the difference, as they evaluated both solutions as equal in quality. Proposition 22 shows that neither $\mathbb{S}_{\mathrm{prec}}$ nor $\mathbb{S}_{\mathrm{erec}}$ satisfy the condition for homogeneity compliance, despite their results in Table 5. Propositions 23 and 24 show that $\mathbb{S}_{\mathrm{csi}}$ and $\mathbb{S}_{\mathrm{ebc}}$ have the homogeneity property.

TABLE 5
Evaluation of the Biclusterings in Fig. 6

Meas.	Fig. 6b	Fig. 6c	Meas.	Fig. 6b	Fig. 6c
$\mathbb{S}_{\mathrm{prel}}$	0.775	0.775	$\mathbb{S}_{\mathrm{fabi}}$	0.300	0.300
$\mathbb{S}_{\mathrm{prec}}$	**0.707**	**0.619**	$\mathbb{S}_{\mathrm{u}}$	1.000	1.000
$\mathbb{S}_{\mathrm{rnia}}$	1.000	1.000	$\mathbb{S}_{\mathrm{e}}$	1.000	1.000
$\mathbb{S}_{\mathrm{ce}}$	0.600	0.600	$\mathbb{S}_{\mathrm{ay}}$	0.600	0.600
$\mathbb{S}_{\mathrm{lw}}$	0.714	0.714	$\mathbb{S}_{\mathrm{erel}}$	0.600	0.600
$\mathbb{S}_{\mathrm{stm}}$	0.750	0.750	$\mathbb{S}_{\mathrm{erec}}$	**0.500**	**0.383**
$\mathbb{S}_{\mathrm{wjac}}$	0.600	0.600	$\mathbb{S}_{\mathrm{csi}}$	**0.467**	**0.347**
$\mathbb{S}_{\mathrm{wdic}}$	0.750	0.750	$\mathbb{S}_{\mathrm{ebc}}$	**0.864**	**0.800**

Definition 8 (Conditions for Maximum). *We say that $\mathbb{S}$ follows the necessary and sufficient conditions for the maximum if $\mathbb{S}$ is such that: $\mathbb{S}(B, \dot{B}) = 1$ iff B and $\dot{B}$ are equivalent biclusterings.*

The above property is important because it guarantees that no better solution exists if the measure attains the maximum. Proposition 25 shows that $\mathbb{S}_{\mathrm{ce}}(B, \dot{B}) = 1$ iff $B \equiv \dot{B}$, which was not shown in [30] for the case in which overlapping is allowed. Proposition 26 shows that $\mathbb{S}_{\mathrm{fabi}}$ also follows the conditions for the maximum. The $\mathbb{S}_{\mathrm{csi}}$ and $\mathbb{S}_{\mathrm{ebc}}$ measures do not obey such conditions (Proposition 27) even for non-degenerate solutions (Proposition 28). Tables 1, 4, and 5 show that the remaining measures do not follow the conditions for the maximum.

In summary, Table 6 discriminates each measure according to the discussed properties.

6.1 Remarks

Because $\mathbb{S}_{\mathrm{prel}}$, $\mathbb{S}_{\mathrm{prec}}$, $\mathbb{S}_{\mathrm{ce}}$, $\mathbb{S}_{\mathrm{l\&w}}$, $\mathbb{S}_{\mathrm{stm}}$, $\mathbb{S}_{\mathrm{wjac}}$, $\mathbb{S}_{\mathrm{wdic}}$, $\mathbb{S}_{\mathrm{fabi}}$, $\mathbb{S}_{\mathrm{ay}}$, $\mathbb{S}_{\mathrm{erel}}$, and $\mathbb{S}_{\mathrm{erec}}$ rely on bicluster-to-bicluster assignments, they ignore the relationship between the matrix entries that do not belong to the main category of the respective bicluster. This type of evaluation is analogous to the set-matching measures for clustering comparison (e.g., Meila and Heckerman's criterion [47], Larsen and Aone's criterion [48], van Dongen's metric [49]), which present the analogous problem of ignoring what occurs in the unmatched part of each cluster [27], [32], [33]. Not coincidentally, none of these biclustering measures have the homogeneity property given by Definition 7 (see Table 6). We also know that matching clusters (or biclusters) between found and reference solutions is somewhat arbitrary [50], [51]; it can be manipulated to either generate more or less favorable evaluations, as clustering algorithms (as well as biclustering algorithms) do not provide such an assignment. The $\mathbb{S}_{\mathrm{rnia}}$, $\mathbb{S}_{\mathrm{u}}$, and $\mathbb{S}_{\mathrm{e}}$ measures are even more extreme because they entirely ignore the relationship between the matrix entries by relying their evaluation only on whether a given matrix entry has been biclustered. Conversely, both $\mathbb{S}_{\mathrm{csi}}$ and $\mathbb{S}_{\mathrm{ebc}}$ analyze the relationship between each pair of matrix entries, similarly to well-known pair-based measures for clusterings, such as Rand index [52], Jaccard index [44], and adjusted Rand index [53]. This is the reason why $\mathbb{S}_{\mathrm{csi}}$ and $\mathbb{S}_{\mathrm{ebc}}$ have the homogeneity property.

Measures $\mathbb{S}_{\mathrm{ce}}$, $\mathbb{S}_{\mathrm{csi}}$, and $\mathbb{S}_{\mathrm{ebc}}$ stand out as the top ones in our theoretical analysis. They differ in what regards the homogeneity and maximum properties only.

TABLE 6
Measure Discrimination According to the Discussed Properties

Properties	$\mathbb{S}_{\text{prel}}$	$\mathbb{S}_{\text{prec}}$	$\mathbb{S}_{\text{rnia}}$	$\mathbb{S}_{\text{ce}}$	$\mathbb{S}_{\text{l\&w}}$	$\mathbb{S}_{\text{stm}}$	$\mathbb{S}_{\text{wjac}}$	$\mathbb{S}_{\text{wdic}}$	$\mathbb{S}_{\text{fabi}}$	$\mathbb{S}_{\text{u}}$	$\mathbb{S}_{\text{e}}$	$\mathbb{S}_{\text{ay}}$	$\mathbb{S}_{\text{erel}}$	$\mathbb{S}_{\text{erec}}$	$\mathbb{S}_{\text{csi}}$	$\mathbb{S}_{\text{ebc}}$
Spur. Bic.			✓	✓			✓	✓			✓[a]				✓	✓
Coverage			✓	✓					✓						✓	✓
Non-int. Area			✓	✓							✓[a]				✓[b]	✓[b]
Mult. Cover.	✓	✓		✓	✓	✓	✓	✓	✓			✓	✓	✓	✓	✓
Rep. Bic.			✓	✓											✓	✓
Symmetry			✓	✓					✓						✓	✓
Homogeneity															✓	✓
Maximum				✓					✓							

[a] *Property valid in the domain of non-overlapping biclusterings.*
[b] *Property valid in the domain of non-degenerate biclusterings.*

7 Experiments

This section describes several experiments whose bicluster models and biclustering algorithms used were drawn from important studies in the biclustering literature of gene expression. These experiments show that the discussed measure properties help understand the results of real comparative experiments. We also present an experiment showing that the evaluation performed by biclustering measures can be used to assess the robustness of a given biclustering algorithm. We then compare two measures in terms of computational performance and memory footprint.

Table 15 in the Appendix, available in the online supplemental material, shows all the biclustering algorithms used in the experiments along with their references and sources from which we obtained the implementations. Table 16 in the Appendix, available in the online supplemental material, shows the parameter values used by the biclustering algorithms in the experiments.

7.1 Empirical Comparative Evaluation

7.1.1 Experiment 1

This experiment follows the constant up-regulated model of bicluster adopted in an empirical comparative analysis of biclustering algorithms recently published [13]. We generated a data matrix with 50 rows and 10 columns[5] and inserted a bicluster having an expression level of 5. The background values were i.i.d. drawn from the standard normal $N(0,1)$. The bicluster inserted in $\dot{B}$ is given by $\dot{B}_1 \triangleq (\{1,2,\ldots,30\},\{1,2,\ldots,8\})$. Fig. 7a shows the data set.

We applied the biclustering algorithms listed in Table 15 to the data matrix. Table 7 shows the results of two interesting cases (from pcluster and bibit algorithms) and of an artificial biclustering constituted of only the first bicluster from the pcluster solution. The pcluster algorithm [54] generated three highly similar biclusters:

$$B_1 = (\{1,2,\ldots,30\},\{1,2,\ldots,8,10\}),$$
$$B_2 = (\{1,2,\ldots,30,46,48\},\{1,2,\ldots,8\}), \quad \text{and}$$
$$B_3 = (\{1,2,\ldots,30,46,47,50\},\{1,2,\ldots,8\}).$$

Note that the $\mathbb{S}_{\text{rnia}}$, $\mathbb{S}_{\text{ce}}$, $\mathbb{S}_{\text{csi}}$, and $\mathbb{S}_{\text{ebc}}$ measures, which have the property of detecting replicated biclusters (Definition 5), considerably penalized the pcluster solution. Although not having the above property, the $\mathbb{S}_{\text{fabi}}$ measure severely penalized the pcluster solution, which can be explained by the difference in the number of biclusters between the found and reference solutions. The third column shows the results regarding a biclustering having only B_1 from the pcluster solution. Measures $\mathbb{S}_{\text{l\&w}}$ and $\mathbb{S}_{\text{u}}$ showed small to no difference between the full pcluster solution, having three nearly identical biclusters, and the almost perfect biclustering $\{B_1\}$. Measures $\mathbb{S}_{\text{prel}}$, $\mathbb{S}_{\text{prec}}$, $\mathbb{S}_{\text{stm}}$, $\mathbb{S}_{\text{wjac}}$, $\mathbb{S}_{\text{wdic}}$, $\mathbb{S}_{\text{ay}}$, $\mathbb{S}_{\text{erel}}$, and $\mathbb{S}_{\text{erec}}$ evaluated $\{B_1\}$ as a worse solution than the full pcluster one, which is counterintuitive.

The bibit algorithm [55] found 31 biclusters. All of them encompass a large portion of the data matrix, which led $\mathbb{S}_{\text{prec}}$, $\mathbb{S}_{\text{u}}$, and $\mathbb{S}_{\text{erec}}$ to attain 1. This is not dramatic per se because these three measures should be taken together with their pairs $\mathbb{S}_{\text{prel}}$, $\mathbb{S}_{\text{e}}$, and $\mathbb{S}_{\text{erel}}$ in an analysis. However, we believe that a good measure should evaluate the found solution as very poor because of the big difference in the number of biclusters. Table 7 shows that the measures that have the property of detecting replicated biclusters along with $\mathbb{S}_{\text{fabi}}$ attained evaluations close to zero for the bibit solution.

7.1.2 Experiment 2

The seminal paper by Madeira and Oliveira [6] proposed four major classes of biclusters. One of these classes consists of biclusters with constant values in rows or columns. We then created a data set with 30 rows and 10 columns having a constant column-wise bicluster and a constant row-wise bicluster, depicted in Fig. 7b. Table 8 displays the results of four algorithms. Pcluster generated the extreme amount of 497 biclusters, but several of the measures assigned relatively high scores to it. As in the previous experiment, only the measures that have the property of detecting replicated biclusters together with $\mathbb{S}_{\text{fabi}}$ attained close to zero evaluations.

Xmotifs [56] found only the constant row-wise bicluster. The $\mathbb{S}_{\text{rnia}}$, $\mathbb{S}_{\text{ce}}$, $\mathbb{S}_{\text{fabi}}$, $\mathbb{S}_{\text{csi}}$, and $\mathbb{S}_{\text{ebc}}$ measures, which

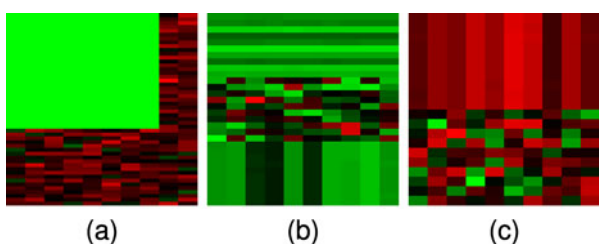


Fig. 7. Data sets for identifying pathological cases.

5. We generated a data set smaller than the ones in [13] for didactic purposes.

TABLE 7
Evaluations Regarding the Data Set in Fig. 7a

Meas.	Pcluster	Pcluster $\{B_1\}$	Bibit
$\mathbb{S}_{\text{prel}}$	0.956	0.943	0.718
$\mathbb{S}_{\text{prec}}$	1.000	0.943	1.000
$\mathbb{S}_{\text{rnia}}$	0.304	0.889	0.024
$\mathbb{S}_{\text{ce}}$	0.304	0.889	0.024
$\mathbb{S}_{\text{lw}}$	0.950	0.974	0.679
$\mathbb{S}_{\text{stm}}$	0.954	0.941	0.688
$\mathbb{S}_{\text{wjac}}$	0.911	0.889	0.531
$\mathbb{S}_{\text{wdic}}$	0.954	0.941	0.677
$\mathbb{S}_{\text{fabi}}$	0.312	0.889	0.032
$\mathbb{S}_{\text{u}}$	1.000	1.000	1.000
$\mathbb{S}_{\text{e}}$	0.795	0.889	0.480
$\mathbb{S}_{\text{ay}}$	0.912	0.889	0.532
$\mathbb{S}_{\text{erel}}$	0.912	0.889	0.549
$\mathbb{S}_{\text{erec}}$	0.938	0.889	1.000
$\mathbb{S}_{\text{csi}}$	0.083	0.790	0.004
$\mathbb{S}_{\text{ebc}}$	0.687	0.940	0.025

TABLE 8
Evaluations Regarding the Data Set in Fig. 7b

Meas.	Pcluster	Xmotifs	Bcca	Msbe
$\mathbb{S}_{\text{prel}}$	0.625	1.000	1.000	0.408
$\mathbb{S}_{\text{prec}}$	1.000	0.707	0.707	1.000
$\mathbb{S}_{\text{rnia}}$	0.004	0.500	0.500	0.855
$\mathbb{S}_{\text{ce}}$	0.004	0.500	0.500	0.855
$\mathbb{S}_{\text{lw}}$	0.620	1.000	1.000	0.405
$\mathbb{S}_{\text{stm}}$	0.607	1.000	1.000	0.333
$\mathbb{S}_{\text{wjac}}$	0.439	1.000	1.000	0.855
$\mathbb{S}_{\text{wdic}}$	0.603	1.000	1.000	0.855
$\mathbb{S}_{\text{fabi}}$	0.004	0.500	0.500	0.333
$\mathbb{S}_{\text{u}}$	1.000	0.500	0.500	1.000
$\mathbb{S}_{\text{e}}$	0.678	1.000	1.000	0.862
$\mathbb{S}_{\text{ay}}$	0.392	1.000	1.000	0.333
$\mathbb{S}_{\text{erel}}$	0.443	1.000	1.000	0.333
$\mathbb{S}_{\text{erec}}$	0.950	0.500	0.500	1.000
$\mathbb{S}_{\text{csi}}$	0.001	0.500	0.500	0.932
$\mathbb{S}_{\text{ebc}}$	0.038	0.802	0.802	0.950

follow the coverage property given by Definition 2, detected the difference between the found and reference solutions. The $\mathbb{S}_{\text{prec}}$, $\mathbb{S}_{\text{u}}$, and $\mathbb{S}_{\text{erec}}$ measures, which do not follow the coverage property, also detected the difference. The reason is that these three measures violate the coverage property only in certain convoluted situations, as in Proposition 6. Similarly, the bcca algorithm [57] found only the constant column-wise bicluster, and the results were the same.

The msbe algorithm [39] found six biclusters, in which the first two perfectly match the reference ones and the others have no overlap with the reference biclusters:

$$\begin{aligned}
B_1 &= (\{1,2,\ldots,10\},\{1,2,\ldots,10\}),\\
B_2 &= (\{21,22,\ldots,30\},\{1,2,\ldots,10\}),\\
B_3 &= (\{12,15,17\},\{6,8,10\}),\\
B_4 &= (\{11,19,20\},\{1,4,9\}),\\
B_5 &= (\{12,15,16,19\},\{3,4\}),\text{and}\\
B_6 &= (\{11,13,17,20\},\{2,6\}).
\end{aligned}$$

This solution brings to light other measure characteristics not directly related to the studied properties. The $\mathbb{S}_{\text{stm}}$, $\mathbb{S}_{\text{fabi}}$, $\mathbb{S}_{\text{ay}}$, and $\mathbb{S}_{\text{erel}}$ measures matched biclusters B_1 and B_2 with the corresponding reference ones and summed their contribution to the evaluation. The summation result was then divided by the number of found biclusters, explaining why they attained value $2/6$. The $\mathbb{S}_{\text{wjac}}$ and $\mathbb{S}_{\text{wdic}}$ measures reached higher values because the sizes of the correct biclusters are greater than the sizes of the spurious ones. Measures $\mathbb{S}_{\text{csi}}$ and $\mathbb{S}_{\text{ebc}}$ attained high values due to their pair-wise based approach of evaluation. For example, the submatrix corresponding to bicluster B_1 has $\binom{100}{2} = 4,950$ pairs of matrix entries, which are individually evaluated by $\mathbb{S}_{\text{csi}}$ and $\mathbb{S}_{\text{ebc}}$, whereas the submatrix corresponding to the spurious bicluster B_3 has only $\binom{9}{2} = 36$ pairs. On the other hand, the other measures evaluate the solutions in terms of matrix entries, meaning that the found spurious biclusters have higher relevance in the evaluation. For example, both $\mathbb{S}_{\text{rnia}}$ and $\mathbb{S}_{\text{ce}}$ attained 0.855.

7.1.3 Experiment 3

We created another data set with 20 rows and 10 columns and only one constant column-wise bicluster, depicted in Fig. 7c. The las algorithm [58] found the biclusters

$$\begin{aligned}
B_1 &= (\{1,2,\ldots,10\},\{1,2,\ldots,10\}) \quad \text{and}\\
B_2 &= (\{11,13,17,18\},\{1,4,5,6,7,9,10\}),
\end{aligned}$$

and the cc algorithm [4] found

$$\begin{aligned}
B_1 &= (\{1,2,\ldots,10\},\{1,2,\ldots,10\}) \quad \text{and}\\
B_2 &= (\{11,15,16\},\{6,10\}).
\end{aligned}$$

The cc solution is better than las' because the spurious bicluster is smaller. Table 9 shows that the measures that are not sensitive to spurious biclusters (Definition 1) did not detect the difference between the solutions or evaluated las' solution as better than cc's.

7.1.4 Prelić's Experiments

The experiments of Scenario 1 in [8] consist in data sets with 10 biclusters with 10 rows and 5 columns each, placed in the diagonal of the data matrix. Fig. 8 illustrates one of the data sets,[6] to which we applied the same biclustering algorithms used in [8]. The bimax algorithm [8] found only one bicluster, which matches one of the reference ones. Similarly to the results from the xmotifs and bcca algorithms in Section 7.1.2, Table 10 shows that several measures assessed the bimax's solution as a perfect one.

Measure $\mathbb{S}_{\text{ebc}}$ attained an unexpected high evaluation for bimax's solution, which can be explained as follows. The precision given by Eq. (14a) is 1, as expected. The recall given by Eq. (14b) is an average over object evaluations, which means an average over matrix entry evaluations by following our approach. A matrix entry that is not

6. It can be downloaded from http://www.tik.ee.ethz.ch/~sop/bimax/SupplementaryMaterial/Datasets/InSilico/Scenario1/data/em_1,n_0.15.1.txt.h.

TABLE 9
Evaluations Regarding the Data Set in Fig. 7

Meas.	Las	Cc
$\mathbb{S}_{\text{prel}}$	0.652	0.548
$\mathbb{S}_{\text{prec}}$	1.000	1.000
$\mathbb{S}_{\text{rnia}}$	0.781	0.943
$\mathbb{S}_{\text{ce}}$	0.781	0.943
$\mathbb{S}_{\text{lw}}$	0.646	0.543
$\mathbb{S}_{\text{stm}}$	0.500	0.500
$\mathbb{S}_{\text{wjac}}$	0.781	0.943
$\mathbb{S}_{\text{wdic}}$	0.781	0.943
$\mathbb{S}_{\text{fabi}}$	0.500	0.500
$\mathbb{S}_{\text{u}}$	1.000	1.000
$\mathbb{S}_{\text{e}}$	0.781	0.943
$\mathbb{S}_{\text{ay}}$	0.500	0.500
$\mathbb{S}_{\text{erel}}$	0.500	0.500
$\mathbb{S}_{\text{erec}}$	1.000	1.000
$\mathbb{S}_{\text{csi}}$	0.929	0.997
$\mathbb{S}_{\text{ebc}}$	0.928	0.987

biclustered in found and reference solutions attains evaluation 1. Since by far most matrix entries corresponde to that case, the recall attained the high evaluation of 0.91, explaining the value given by $\mathbb{S}_{\text{ebc}}$. On the other hand, $\mathbb{S}_{\text{csi}}$ evaluates each pair of matrix entries and consolidates them in Eqs. (18) and (19). The pairs of noisy entries (the reddish elements in Fig. 8), which are by far the most abundant type of pairs in this problem and on which the found and reference solutions agree, do not contribute to the $\mathbb{S}_{\text{csi}}$ evaluation because both agreement and disagreement functions (Eqs. (16) and (17) assign 0 to them. The remaining pairs consist in pairs of greenish entries and pairs of greenish and reddish entries, on which the found and reference solutions generally disagree. Precisely, the agreement and disagreement terms (Eqs. (18) and (19) attained values 1,225 and 11,025, respectively.

7.1.5 Eren's Experiments

We generated a data matrix with 50 rows and 20 columns having three biclusters that follow the shift model used in [13]. Fig. 9 depicts the data matrix. The reference biclusters are

$$\dot{B}_1 \triangleq (\{1,2,\ldots,25\},\{1,2,\ldots,10\}),$$
$$\dot{B}_2 \triangleq (\{1,2,\ldots,25\},\{11,12,\ldots,20\}), \quad \text{and}$$
$$\dot{B}_3 \triangleq (\{26,27,\ldots,50\},\{1,2,\ldots,10\}).$$

TABLE 10
Evaluations Regarding the Data Set in Fig. 8

Meas.	Bimax	Meas.	Bimax
$\mathbb{S}_{\text{prel}}$	1.000	$\mathbb{S}_{\text{fabi}}$	0.100
$\mathbb{S}_{\text{prec}}$	0.100	$\mathbb{S}_{\text{u}}$	0.100
$\mathbb{S}_{\text{rnia}}$	0.100	$\mathbb{S}_{\text{e}}$	1.000
$\mathbb{S}_{\text{ce}}$	0.100	$\mathbb{S}_{\text{ay}}$	1.000
$\mathbb{S}_{\text{lw}}$	1.000	$\mathbb{S}_{\text{erel}}$	1.000
$\mathbb{S}_{\text{stm}}$	1.000	$\mathbb{S}_{\text{erec}}$	0.100
$\mathbb{S}_{\text{wjac}}$	1.000	$\mathbb{S}_{\text{csi}}$	0.100
$\mathbb{S}_{\text{wdic}}$	1.000	$\mathbb{S}_{\text{ebc}}$	0.954

We applied the same biclustering algorithms used in [13]. The bbc algorithm [16] found only $\dot{B}_2$ and $\dot{B}_3$, but Table 11 shows that several measures evaluated the bbc's solution as a perfect one, which again can be explained by their lack of the coverage property given by Definition 2.

Overall, the only measures that did not show evident counterintuitive evaluations in the empirical analysis are the $\mathbb{S}_{\text{rnia}}$, $\mathbb{S}_{\text{ce}}$, and $\mathbb{S}_{\text{csi}}$ measures. $\mathbb{S}_{\text{fabi}}$ failed in Experiment 3 for not discriminating the clearly better solution from the other. The $\mathbb{S}_{\text{ebc}}$ measure attained a high evaluation for a very poor solution in Section 7.1.4, exposing a conceptual flaw in the application of $\mathbb{S}_{\text{ebc}}$ to soft clustering representation of biclustering. Most of the behavior exhibited by the measures can be explained with the help of the properties defined in Section 6.

7.2 Further Experiments with Selected Measures

In Section 7.2.1 we show an example of application where a biclustering measure can be used to assess the noise robustness of a biclustering algorithm. We employed $\mathbb{S}_{\text{ce}}$ and $\mathbb{S}_{\text{csi}}$, the only two measures that showed superior behavior in both the theoretical (Section 6) and empirical analyses (Section 7.1). Section 7.2.2 evaluates the computational performance and memory footprint of $\mathbb{S}_{\text{ce}}$ and $\mathbb{S}_{\text{csi}}$, pointing to fast implementations of both.

7.2.1 Noise Robustness Analysis

The fabia algorithm [37] assumes a multiplicative data set model described by

$$\mathrm{A} = \sum_{i=1}^{k} \boldsymbol{\lambda}_i \mathbf{z}_i^{\mathrm{T}} + \boldsymbol{\Upsilon},$$

where $\boldsymbol{\lambda}_i \in \mathbb{R}^n$ and $\mathbf{z}_i \in \mathbb{R}^p$ are sparse vectors defining the ith bicluster and $\boldsymbol{\Upsilon} \in \mathbb{R}^{n \cdot p}$ is the additive noise. The data sets

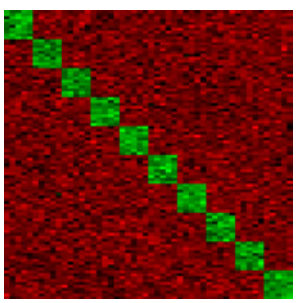


Fig. 8. A data set from Scenario 1 in [8].

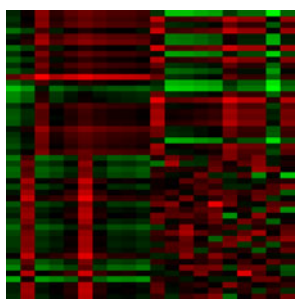


Fig. 9. A data set with shift biclusters [13].

TABLE 11
Evaluations Regarding the Data Set in Fig. 9

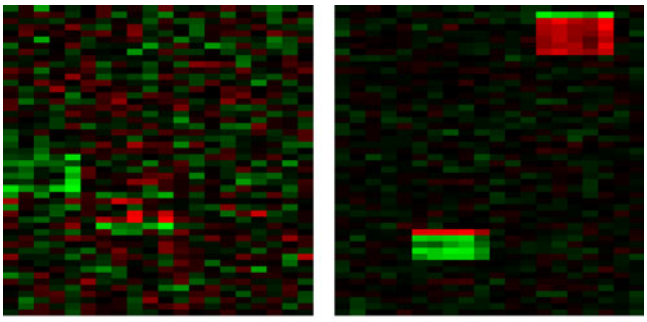
Meas.	Bbc	Meas.	Bbc
$\mathbb{S}_{\text{prel}}$	1.000	$\mathbb{S}_{\text{fabi}}$	0.667
$\mathbb{S}_{\text{prec}}$	1.000	$\mathbb{S}_{\text{u}}$	0.667
$\mathbb{S}_{\text{rnia}}$	0.667	$\mathbb{S}_{\text{e}}$	1.000
$\mathbb{S}_{\text{ce}}$	0.667	$\mathbb{S}_{\text{ay}}$	1.000
$\mathbb{S}_{\text{lw}}$	1.000	$\mathbb{S}_{\text{erel}}$	1.000
$\mathbb{S}_{\text{stm}}$	1.000	$\mathbb{S}_{\text{erec}}$	0.667
$\mathbb{S}_{\text{wjac}}$	1.000	$\mathbb{S}_{\text{csi}}$	0.667
$\mathbb{S}_{\text{wdic}}$	1.000	$\mathbb{S}_{\text{ebc}}$	0.858

for this section were generated similarly to the ones in [37] with $n = 50$ and $p = 20$ as follows. The $\boldsymbol{\lambda}_i$'s were generated by randomly choosing the number $n_i^{\lambda} \in \{5, 6, \ldots, 10\}$ of rows and $\mathbf{z}_i$'s by choosing the number $n_i^{z} \in \{5, 6, 7\}$ of columns. The n_i^{λ} components of $\boldsymbol{\lambda}_i$ and the n_i^{z} components of $\mathbf{z}_i$ (randomly chosen) were set to values drawn from $\mathcal{N}(\mu, 1)$. The remaining components from $\boldsymbol{\lambda}_i$ and $\mathbf{z}_i$ were drawn from $\mathcal{N}(0, 0.2^2)$. The $\boldsymbol{\Upsilon}$ components were drawn from $\mathcal{N}(0, 3^2)$. We generated 30 data sets for each $\mu \in \{0, 1, \ldots, 5\}$ using the above approach and applied the fabia algorithm (with the configuration given by Table 16) 30 times to each data set. The best evaluations according to $\mathbb{S}_{\text{ce}}$ and $\mathbb{S}_{\text{csi}}$ were retained for each data set. Two data sets are represented in Fig. 10 and the results are found in Fig. 11.

The evaluations show that the performance barely degradated from $\mu = 5$ to $\mu = 3$, showing that the fabia algorithm is robust for this range of signals and type of data set. The performance noticeably began to degradate for μ values smaller than 3. This type of analysis can be useful for assessing the robustness of competing biclustering algorithms.

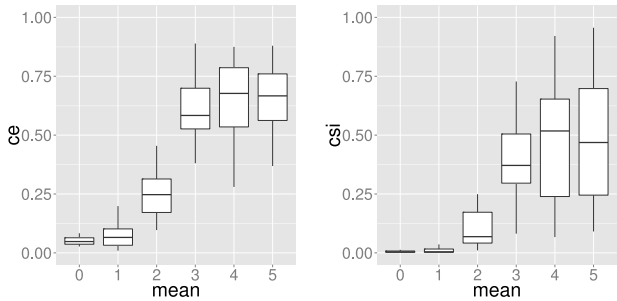
7.2.2 Performance

To assess the computational performance and memory footprint of the $\mathbb{S}_{\text{ce}}$ and $\mathbb{S}_{\text{csi}}$ implementations, we randomly generated 30 biclusterings having 10 biclusters each for data sets having varying numbers of rows $n \in \{50, 100, \ldots, 5,000\}$ and $p = 20$ columns. For each bicluster the numbers of rows $n_r \in \{5, 6, \ldots, n/10\}$ and columns $n_c \in \{5, 6, \ldots, p/2\}$ are randomly drawn, and the biclustered columns and rows are also randomly chosen. We evaluated a naïve and a fast $\mathbb{S}_{\text{csi}}$ implementation, both freely available at http://sn.im/26fzpck. The experiments were performed using the Matlab R2011a environment on a machine with i7 930 2.80 GHz CPU having four cores and 12 GBs of RAM.



(a) $\mu = 2.0$. (b) $\mu = 5.0$.

Fig. 10. Noisy data sets.



(a) The $\mathbb{S}_{\text{ce}}$ measure. (b) The $\mathbb{S}_{\text{csi}}$ measure.

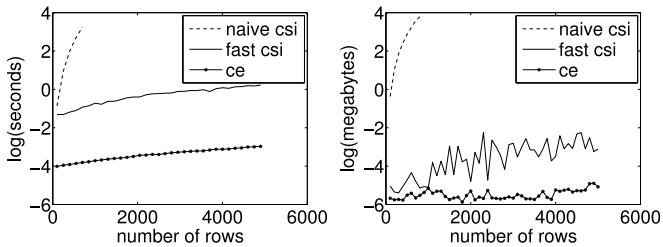
Fig. 11. Results for the experiments having noise.

Fig. 12 show that the naïve $\mathbb{S}_{\text{csi}}$ implementation is not only very slow but also could not handle the biclusterings from data sets having more than 800 rows because of the excessive use of memory. The $\mathbb{S}_{\text{ce}}$ implementation is faster and consume less memory than the fast $\mathbb{S}_{\text{csi}}$ implementation. However, the latter is still fast (took 1.3 seconds for $n = 5{,}000$) and modest in memory use (less than 1 MB for $n = 5{,}000$) for real applications.

8 Conclusions

This paper discussed the different types of evaluation approaches commonly encountered in the gene expression studies involving biclustering algorithms. One of these types of evaluation is the external one, which is of great importance for comparative studies, as explained in the introduction. We reviewed 14 measures used in external evaluations for comparing biclusterings and adapted an approach of subspace clustering comparison to measure the similarity between biclusterings. This approach allows the comparison between biclusterings by using measures for soft clusterings. We then reviewed and adopted two measures for soft clusterings that we believe are promising.

We formalized eight properties that a good biclustering measure should have, discussed why they are relevant, and proved which properties each measure has. The significance of the properties was assessed in experiments based on bicluster models and biclustering algorithms used in important studies. The $\mathbb{S}_{\text{ce}}$, $\mathbb{S}_{\text{csi}}$, and $\mathbb{S}_{\text{ebc}}$ measures stood out as the top ones in the theoretical comparison. However, we identified a problematic behavior of $\mathbb{S}_{\text{ebc}}$ in the empirical analysis, namely, the abundant noisy entries (which is not unusual in real gene expression data) dominated the $\mathbb{S}_{\text{ebc}}$ evaluation leading it to attain a high value to a clearly poor solution.



(a) Computational performance. (b) Memory footprint.

Fig. 12. Implementation performance.

Our study thus suggests that the $\mathbb{S}_{ce}$ and $\mathbb{S}_{csi}$ measures should be preferred for comparing biclusterings.

An interesting future work would be the an analysis of the measure behaviors for randomly generated biclusterings. It has been appreciated the importance of having measures for comparing clusterings that show a constant baseline evaluation for randomly generated solutions [59], [60], as a strategy to avoid biased evaluations.

The Matlab implementation of all the measures and the data sets we generated can be found at http://sn.im/26fzpck.

ACKNOWLEDGMENTS

The authors thank CNPq and FAPESP for their financial support. They thank Mehmet Deveci for sending the bbc files by personal communication.

REFERENCES

[1] H. L. Turner, T. C. Bailey, W. J. Krzanowski, and C. A. Hemingway, "Biclustering models for structured microarray data," *IEEE/ACM Trans. Comput. Biol. Bioinformat.*, vol. 2, no. 4, pp. 316–329, Oct. 2005.

[2] J. A. Hartigan, "Direct clustering of a data matrix," *J. Amer. Statist. Assoc.*, vol. 67, no. 337, pp. 123–129, 1972.

[3] B. Mirkin, *Mathematical Classification and Clustering*. New York, NY, USA: Springer, 1996.

[4] Y. Cheng and G. M. Church, "Biclustering of expression data," in *Proc. 8th Int. Conf. Intell. Syst. Mol. Biol.*, 2000, pp. 93–103.

[5] Y. Kluger, R. Basri, J. Chang, and M. Gerstein, "Spectral biclustering of microarray data: Coclustering genes and conditions," *Genome Res.*, vol. 13, no. 4, pp. 703–716, 2003.

[6] S. C. Madeira and A. L. Oliveira, "Biclustering algorithms for biological data analysis: A survey," *IEEE/ACM Trans. Comput. Biol. Bioinformat.*, vol. 1, no. 1, pp. 24–45, Jan. 2004.

[7] F. Divina and J. S. Aguilar-Ruiz, "Biclustering of expression data with evolutionary computation," *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 5, pp. 590–602, May 2006.

[8] A. Prelić, S. Bleuler, P. Zimmermann, A. Wille, P. Buhlmann, W. Gruissem, L. Hennig, L. Thiele, and E. Zitzler, "A systematic comparison and evaluation of biclustering methods for gene expression data," *Bioinformatics*, vol. 22, no. 9, pp. 1122–1129, 2006.

[9] S. Busygin, O. Prokopyev, and P. M. Pardalos, "Biclustering in data mining," *Comput. Oper. Res.*, vol. 35, no. 9, pp. 2964–2987, 2008.

[10] W.-H. Yang, D.-Q. Dai, and H. Yan, "Finding correlated biclusters from gene expression data," *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 4, pp. 568–584, Apr. 2011.

[11] I. S. Dhillon, S. Mallela, and D. S. Modha, "Information-theoretic co-clustering," in *Proc. 9th ACM SIGKDD Int. Conf. Knowl., Discov., Data Mining*, 2003, pp. 89–98.

[12] A. Tanay, R. Sharan, and R. Shamir, "Biclustering algorithms: A survey," in *Handbook of Computational Molecular Biology*, S. Aluru, Ed. Cleveland, OH, USA: CRC Press, 2005, pp. 26–1.

[13] K. Eren, M. Deveci, O. Küçüktunç, and U. V. Çatalyürek, "A comparative analysis of biclustering algorithms for gene expression data," *Briefings Bioinformat.*, vol. 14, pp. 279–292, 2012.

[14] J. A. Nepomuceno, A. T. Lora, J. S. Aguilar-Ruiz, and J. García-Gutiérrez, "Biclusters evaluation based on shifting and scaling patterns," in *Proc. 8th Int. Conf. Intell. Data Eng. Autom. Learn.*, 2007, pp. 840–849.

[15] R. Santamaría, R. Therón, and L. Quintales, "A visual analytics approach for understanding biclustering results from microarray data," *BMC Bioinformat.*, vol. 9, no. 1, p. 247, 2008.

[16] J. Gu and J. Liu, "Bayesian biclustering of gene expression data," *BMC Genomics*, vol. 9, no. Suppl 1, p. S4, 2008.

[17] K.-O. Cheng, N.-F. Law, W.-C. Siu, and A. Liew, "Identification of coherent patterns in gene expression data using an efficient biclustering algorithm and parallel coordinate visualization," *BMC Bioinformat.*, vol. 9, no. 1, p. 210, 2008.

[18] S. Dharan and A. Nair, "Biclustering of gene expression data using reactive greedy randomized adaptive search procedure," *BMC Bioinformat.*, vol. 10, no. Suppl 1, p. S27, 2009.

[19] B. Hanczar and M. Nadif, "Ensemble methods for biclustering tasks," *Pattern Recognit.*, vol. 45, no. 11, pp. 3938–3949, 2012.

[20] B. Gao, O. Griffith, M. Ester, H. Xiong, Q. Zhao, and S. Jones, "On the deep order-preserving submatrix problem: A best effort approach," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 2, pp. 309–325, Feb. 2012.

[21] G. O. Consortium. (2012). Gene ontology website. [Online]. Available: http://www.geneontology.org/

[22] C. Cano, L. Adarve, J. López, and A. Blanco, "Possibilistic approach for biclustering microarray data," *Comput. Biol. Med.*, vol. 37, no. 10, pp. 1426–1436, 2007.

[23] S. Gremalschi and G. Altun, "Mean squared residue based biclustering algorithms," in *Proc. 4th Int. Conf. Bioinformat. Res. Appl.*, 2008, pp. 232–243.

[24] Y. Lee, J. Lee, and C.-H. Jun, "Stability-based validation of bicluster solutions," *Pattern Recognit.*, vol. 44, no. 2, pp. 252–264, 2011.

[25] J. S. Aguilar-Ruiz, "Shifting and scaling patterns from gene expression data," *Bioinformatics*, vol. 21, no. 20, pp. 3840–3845, Oct. 2005.

[26] D. Bozdag, A. S. Kumar, and U. V. Çatalyürek, "Comparative analysis of biclustering algorithms," in *Proc. ACM Int. Conf. Bioinformat. Comput. Biol.*, 2010, pp. 265–274.

[27] M. Meila, "Comparing clusterings by the variation of information," in *Proc. 16th Annu. Conf. Learn. Theory Kernel Mach.*, 2003, vol. 2777, pp. 173–187.

[28] M. Meila, "Comparing clusterings: An axiomatic view," in *Proc. 22nd Int. Conf. Mach. Learn.*, 2005, pp. 577–584.

[29] M. Meila, "Comparing clusterings—An information based distance," *J. Multivar. Anal.*, vol. 98, pp. 873–895, May 2007.

[30] A. Patrikainen and M. Meila, "Comparing subspace clusterings," *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 7, pp. 902–916, Jul. 2006.

[31] R. Santamaría, L. Quintales, and R. Therón, "Methods to bicluster validation and comparison in microarray data," in *Proc. Intell. Data Eng. Autom. Learn.*, 2007, vol. 4881, pp. 780–789.

[32] A. Rosenberg and J. Hirschberg, "V-measure: A conditional entropy-based external cluster evaluation measure," in *Proc. Joint Conf. Empir. Methods Natural Lang. Process. Comput. Natural Lang. Learn.*, 2007, pp. 410–420.

[33] E. Amigó, J. Gonzalo, J. Artiles, and F. Verdejo, "A comparison of extrinsic clustering evaluation metrics based on formal constraints," *Inf. Retrieval*, vol. 12, no. 4, pp. 461–486, Aug. 2009.

[34] Y. Lee, J.-H. Lee, and C.-H. Jun, "Validation measures of bicluster solutions," *Ind. Eng. Manage. Syst.*, vol. 8, no. 2, pp. 101–108, 2009.

[35] B. S. Everitt, S. Landau, and M. Leese, *Cluster Analysis*. London, U. K.: Arnold Publishers, May 2001.

[36] R. J. G. B. Campello, "Generalized external indexes for comparing data partitions with overlapping categories," *Pattern Recognit. Lett.*, vol. 31, no. 9, pp. 966–975, 2010.

[37] S. Hochreiter, U. Bodenhofer, M. Heusel, A. Mayr, A. Mitterecker, A. Kasim, T. Khamiakova, S. Van Sanden, D. Lin, W. Talloen, L. Bijnens, H. W. H. Guhlmann, Z. Shkedy, and D.-A. Clevert, "FABIA: Factor analysis for bicluster acquisition," *Bioinformatics*, vol. 26, no. 12, pp. 1520–1527, 2010.

[38] H. Turner, T. Bailey, and W. Krzanowski, "Improved biclustering of microarray data demonstrated through systematic performance tests," *Comput. Statist. Data Anal.*, vol. 48, no. 2, pp. 235–254, 2005.

[39] X. Liu and L. Wang, "Computing the maximum similarity bi-clusters of gene expression data," *Bioinformatics*, vol. 23, no. 1, pp. 50–56, 2007.

[40] J. Xiao, L. Wang, X. Liu, and T. Jiang, "Finding additive biclusters with random background," in *Proc. 19th Annu. Symp. Combinatorial Pattern Matching*, 2008, vol. 5029, pp. 263–276.

[41] Q. Huang, M. Lu, and H. Yan, "An evolutionary algorithm for discovering biclusters in gene expression data of breast cancer," in *Proc. IEEE Congress Evol. Comput.*, 2008, pp. 829–834.

[42] A. Freitas, V. Afreixo, M. Pinheiro, J. Oliveira, G. Moura, and M. Santos, "Improving the performance of the iterative signature algorithm for the identification of relevant patterns," *Statist. Anal. Data Mining*, vol. 4, no. 1, pp. 71–83, 2011.

[43] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. pp. 297–302, 1945.

[44] P. Jaccard, "Nouvelles recherches sur la distribution florale," *Bull. Socit Vaudoise de Sci. Naturelles*, vol. 44, pp. 223–370, 1908.

[45] W. Ayadi, M. Elloumi, and J.-K. Hao, "Bicfinder: A biclustering algorithm for microarray data analysis," *Knowl. Inf. Syst.*, vol. 30, no. 2, pp. 341–358, Feb. 2012.
[46] W. Ayadi, O. Maatouk, and H. Bouziri, "Evolutionary biclustering algorithm of gene expression data," in *Proc. 23rd Int. Workshop Database Expert Syst. Appl.*, Sep. 2012, pp. 206–210.
[47] M. Meila and D. Heckerman, "An experimental comparison of model-based clustering methods," *Mach. Learn.*, vol. 42, pp. 9–29, 2001.
[48] B. Larsen and C. Aone, "Fast and effective text mining using linear-time document clustering," in *Proc. 5th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 1999, pp. 16–22.
[49] S. Dongen, "Performance criteria for graph clustering and markov cluster experiments," Nat. Res. Inst. Math. Comput. Sci., Amsterdam, The Netherlands, Tech. Rep. INS-R0012, 2000.
[50] D. Steinley, "Local optima in k-means clustering: What you don't know may hurt you." *Psychol. Methods*, vol. 8, no. 3, pp. 294–304, 2003.
[51] D. Steinley, "Properties of the hubert-arabie adjusted rand index." *Psychol. Methods*, vol. 9, no. 3, pp. 386–396, 2004.
[52] W. M. Rand, "Objective criteria for the evaluation of clustering methods," *J. Amer. Statist. Assoc.*, vol. 66, no. 336, pp. 846–850, 1971.
[53] L. Hubert and P. Arabie, "Comparing partitions," *J. Classification*, vol. 2, no. 1, pp. 193–218, Dec. 1985.
[54] H. Wang, W. Wang, J. Yang, and P. S. Yu, "Clustering by pattern similarity in large data sets," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2002, pp. 394–405.
[55] D. S. Rodriguez-Baena, A. J. Perez-Pulido, and J. S. Aguilar-Ruiz, "A biclustering algorithm for extracting bit-patterns from binary datasets," *Bioinformatics*, vol. 27, no. 19, pp. 2738–2745, 2011.
[56] T. M. Murali and S. Kasif, "Extracting conserved gene expression motifs from gene expression data," in *Proc. Pac. Symp. Biocomput.*, 2003, pp. 77–88.
[57] A. Bhattacharya and R. K. De, "Bi-correlation clustering algorithm for determining a set of co-regulated genes," *Bioinformatics*, vol. 25, no. 21, pp. 2795–2801, 2009.
[58] A. Shabalin, V. Weigman, C. Perou, and A. Nobel, "Finding large average submatrices in high dimensional data," *Ann. Appl. Statist.*, vol. 3, no. 3, pp. 985–1012, 2009.
[59] N. X. Vinh, J. Epps, and J. Bailey, "Information theoretic measures for clusterings comparison: Is a correction for chance necessary?" in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, 2009, pp. 1073–1080.
[60] N. X. Vinh, J. Epps, and J. Bailey, "Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance," *J. Mach. Learn. Res.*, vol. 11, pp. 2837–2854, 2010.

Danilo Horta received the BS degree in computer science from the Federal University of Itajubá, MG, Brazil, in 2007. He received the MS degree in the Department of Computer Sciences at the University of São Paulo (ICMC-USP 2007), and is currently working toward the PhD degree in the same department.

Ricardo J. G. B. Campello received the PhD degree in electrical engineering from the State University of Campinas, Brazil, in 2002. Since 2007, he is with the Department of Computer Sciences of the University of São Paulo at São Carlos, Brazil, currently as an associate professor. His current interests are mainly on data mining and machine learning.