The final form of this manuscript was published in the Applied Soft Computing journal by Elsevier: https://doi.org/10.1016/j.asoc.2019.105688.

Experimental correlation analysis of bicluster coherence measures and gene ontology information

Victor Alexandre Padilha*, André Carlos Ponce de Leon Ferreira de Carvalho

Institute of Mathematics and Computer Sciences, University of São Paulo, Av. Trabalhador são-carlense, 400, São Carlos – SP, 13566-590, Brazil

Abstract

Biclustering algorithms have become popular tools for gene expression data analysis. They can identify local patterns defined by subsets of genes and subsets of samples, which cannot be detected by traditional clustering algorithms. In spite of being useful, biclustering is an NP-hard problem. Therefore, the majority of biclustering algorithms look for biclusters optimizing a pre-established coherence measure. Many heuristics and validation measures have been proposed for biclustering over the last 20 years. However, there is a lack of an extensive comparison of bicluster coherence measures on practical scenarios. To deal with this lack, this paper experimentally analyzes 17 bicluster coherence measures and external measures calculated from information obtained in the gene ontologies. In this analysis, results were produced by 10 algorithms from the literature in 19 gene expression datasets. According to the experimental results, a few pairs of strongly correlated coherence measures could be identified, which suggests redundancy. Moreover, the pairs of strongly correlated measures might change when dealing with normalized or non-normalized data and biclusters enriched by different ontologies. Finally, there was no clear relation between coherence measures and assessment using information from gene ontology.

Keywords: Biclustering, Coherence measures, Gene ontology, Gene expression data

*Corresponding author

Email addresses: victorpadilha@usp.br (Victor Alexandre Padilha), andre@icmc.usp.br (André Carlos Ponce de Leon Ferreira de Carvalho)

1. Introduction

High-throughput technologies, such as microarrays [1], allow researchers to monitor the behavior of thousands of genes under specific biological samples. Normally, the samples correspond to different points in a time series, different types of tissues, different environmental conditions, different organs and/or different individuals [2].

Gene expression data analysis studies investigate the behavior of thousands of genes from an organism under multiple biological samples. Their results and conclusions can support a better understanding of gene functions, biological processes, effects of treatments, among others [3, 4]. For such, these studies use a data matrix representation, which is obtained by concatenating the results from multiple high-throughput experiments. In such a matrix, each row corresponds to a gene, each column corresponds to a sample and each element quantifies the expression level of a gene in a specific sample [2, 5].

Traditional clustering algorithms are often used to analyze gene expression data. They allow researchers to improve their understanding of the functions of the genes from an organism. However, some studies argue that a biological process may be active only under subsets of genes and subsets of samples [6, 2, 7], which characterizes clusters in subspaces of the original dataset. Besides, some genes or samples may not take part in any cluster at all. Thus, a traditional clustering method may not be able to answer some important research questions.

Biclustering overcomes the previously discussed clustering limitations. It looks for local patterns, called biclusters, comprising subsets of genes and subsets of samples, which are usually obfuscated by the high dimensionality of a dataset. Additionaly, biclustering may allow the presence of overlapped biclusters and unclustered genes or samples.

Although biclustering has proved its importance, the size of its search space is in the order of 2^{N+M} for N genes and M samples, which characterizes an NP-hard problem [6, 4, 2, 5]. Therefore, many algorithms are based on the op-

timization of a bicluster coherence measure through (meta-)heuristics, in order to produce approximate results in an acceptable amount of time.

The selection or proposal of an appropriate coherence measure is crucial in the development of a biclustering algorithm [7]. Each measure can detect a particular set of patterns and it is the main component that guides the algorithm search for a good solution.

Since 2000, many biclustering algorithms and coherence measures have been proposed in the literature. At the same time, extensive surveys [2, 8, 9, 10] and studies for the comparison of algorithms were carried out [11, 12, 13, 14, 15]. However, few studies have investigated the behaviors of different coherence measures and to what extent they agree with the external biological information available.

In a preliminary study [16], we investigated the correlations of 15 biclustering coherence measures for results generated by 9 biclustering algorithms in 19 gene expression datasets. For such, we considered two experimental scenarios on normalized data to analyze relations between coherence criteria and biological significance of biclusters. The present study extends this work, presenting the following contributions:

- Analysis of correlations between 17 coherence measures for results obtained by 10 biclustering algorithms in the 19 gene expression datasets, to present evidence able to reduce the use of redundant measures during evaluation;

- Evaluation of results for 16 different experimental scenarios, which encompass normalized and non-normalized data, separate and aggregated analyses with the available ontologies. Then, we have more evidence to assess if the performances of coherence measures agree with those achieved by evaluation using external knowledge; and

- Computational complexity analyses of the measures, which are usually not provided in their original studies, and were only provided in the supple-

mentary material of [16]. These analyses are important for applications where a large number of biclusters needs to be assessed.

This paper is organized as follows. Section 2 presents the main related works found by the authors. Section 3 introduces the coherence and external measures, biclustering algorithms and the gene expression datasets selected for this study. Section 4 presents the experiments carried out and discusses their results. Finally, Section 5 presents the main conclusions from this study.

2. Related work

There are several studies which propose new biclustering algorithms and/or coherence measures. However, there is a lack of extensive comparisons between distinct measures on the results of different algorithms. For instance, few related studies discuss how biclustering coherence measures relate to each other.

The first study of biclustering in the context of gene expression data can be found in [6]. The well-known Mean Squared Residue (MSR) coherence measure and an algorithm for its optimization were proposed. According to the experimental results obtained, the biclustering paradigm used affects the gene expression data analysis. This study became the main benchmark adopted when developing new bicluster coherence measures and algorithms.

In [17], the authors formally analyze the MSR measure, showing its limitation for identifying scaling patterns in biclusters, due to its high dependence on the variances of scaling factors. Other studies, such as [18, 19, 20, 21, 22, 23, 24], introduced bicluster coherence measures able to overcome the limitations of MSR on certain types of patterns. The improvements obtained were shown in the performance on synthetic and/or real data.

In [25], the authors proposed a coherence measure and an internal biclustering evaluation index. They also discussed the main advantages and disadvantages when using relative, internal and external measures. The authors tested their proposals on synthetic datasets for the task of hyperparameter selection for two biclustering algorithms.

A new measure, the Minimal Mean Squared Error (MMSE), to detect linear patterns in biclusters, was proposed in [26]. The authors compared MMSE with five measures from the literature. They also adapted the algorithm proposed in [6] to optimize the new measure and performed experiments on synthetic and real datasets. This modified algorithm was compared with 6 biclustering and 2 clustering algorithms from the literature. According to the authors, the new measure and algorithm detected patterns not usually found by other measures and algorithms.

A large number of coherence measures, 14 altogether, were discussed in [7]. From these 14 measures, 13 were tested on synthetic datasets and on 4 real datasets. In the experiments with synthetic datasets, to assess these coherence measures when the biclusters do not follow perfect patterns, they were tested on 3 types of bicluster patterns subject to different noise levels. In the experiments with real datasets, the measures were applied to biclusters found by an evolutionary algorithm previously proposed. However, as this algorithm includes one of the investigated measures in its fitness function, there is a bias in the experimental results. Afterwards, the coherence measures were compared to values obtained from external biological information. According to the authors, the correlation between their results and the biological measures according to a normalized Mutual Information (MI) score showed a relation between many coherence measures with the biological information.

This paper goes one step forward in the previous analysis by comparing a larger number of 17 biclustering measures in a larger number of datasets (19). Besides, in order to reduce algorithmic bias, each measure was evaluated using 10 biclustering algorithms. In order to reduce dependence on estimation algorithms or on data binning to calculate MI for continuous variables, the Pearson and Spearman correlations were used. Additionally, the Wilcoxon signed-rank test was applied to the results to assess any evidence of differences between the results from the two correlation measures.

3. Methods

This section has a description of the main methods used. For such, it is organized as follows. In Section 3.1, we discuss the types of numeric bicluster patterns. In Section 3.2, we present the coherence measures investigated. In Section 3.3, we describe the 10 algorithms used in the experiments. In Section 3.4, we present the 19 datasets selected. In Section 3.5, we discuss the external evaluation using GO ontologies and the quantities calculated from them. In Section 3.6, we detail our experimental methodology. Finally, in Section 3.7, we describe the hyperparameter settings used for the algorithms.

3.1. Bicluster patterns

Let $X = (R, C)$ be a gene expression matrix, where R is a set of N rows (genes) and C is a set of M columns (samples). A bicluster corresponds to a submatrix $B = (I, J)$, $I \subseteq R$, $J \subseteq C$, which presents some patterns between its values. Several numeric patterns have been described in the literature. The most general among them are [10]:

- *Shifting pattern*, where each bicluster element b_{ij} can be defined by a constant/typical value π_i for the i^{th} row added to an adjustment factor β_j for the j^{th} column. Thus, $b_{ij} = \pi_i + \beta_j$.
- *Scaling pattern*, where each bicluster element b_{ij} is described by the constant/typical value π_i for the i^{th} row multiplied by an adjustment factor α_j for the j^{th} column. Thus, $b_{ij} = \pi_i \alpha_j$.
- *Shifting-scaling pattern*, where the bicluster presents both patterns simultaneously. Each bicluster element b_{ij} is obtained by multiplying π_i by α_j and adding the result to β_j. Thus, $b_{ij} = \pi_i \alpha_j + \beta_j$. Note that shifting and scaling biclusters are special cases of shifting-scaling patterns when $\alpha_j = 1$ and $\beta_j = 0$ $\forall j \in J$, respectively.

From the aforementioned patterns, some specific patterns that are also widely referenced in the literature can be extracted, such as [2]:

- *Constant pattern*, where all of the bicluster elements are equal to the same constant value μ. Thus, $\pi_i = \mu \ \forall i \in I$, $\alpha_j = 1$ and $\beta_j = 0 \ \forall j \in J$.

- *Constant row pattern*, where the elements of each row of the bicluster are equal to the same constant value, which can be different from one row to another. Thus, $\alpha_j = 1$ and $\beta_j = 0 \ \forall j \in J$.

- *Constant column pattern*, where the elements of each column of the bicluster are equal to the same constant value, which can be different from one column to another. Thus, $\pi_i = 1 \ \forall i \in I$.

It must be mentioned that, in real gene expression data, the expression values may be obfuscated by the presence of noise. Therefore, one cannot expect the biclusters to always present the perfect patterns previously described. Thus, for each element x_{ij} of the original data matrix X, there is generally an unknown η_{ij} value associated to it, which represents its amount of noise [2]. This motivates the use of coherence measures, which quantify the extent of agreement between a noisy bicluster and a desired ideal pattern.

3.2. Coherence measures

The use of coherence measures is an important step to evaluate a set of biclusters that were produced by one or more biclustering algorithms. These measures require only the data available and inspect the quality of the biclusters' elements regarding a set of predefined patterns. By using different measures, the results can be assessed from different perspectives and, as a consequence, cover different aspects of the data based on distinct approaches, such as: the variability of the bicluster's values (Variance-based), correlations among genes or biological samples (Correlation-based), and correspondence of the bicluster's elements with a general tendency pattern that models their behavior (Standardization-based).

In this section, we introduce the coherence measures investigated in this paper. They are the same measures investigated in [7] and four additional measures which, to the best of our knowledge, have not been previously investigated in related studies: three were the main contributions of the bicluster evaluation

work in [25] that assesses constant patterns, constant row patterns and constant column patterns; the fourth, proposed in [26], can capture shifting-scaling biclusters that, although is the most general bicluster model discussed in the literature, is the hardest one to deal with and only few measures are able to properly evaluate it.

Next, we present the measures, organized in the following categories, according to the similarities of their approaches: Variance-based (Section 3.2.1), Correlation-based (Section 3.2.2) and Standardization-based (Section 3.2.3). We also provide the time complexity analyses for the measures, which are usually not provided in their original publications. In Table 1, we present a summary of the measures: range of values, objectives (i.e., if a measure must be maximized or minimized) and time complexity.

3.2.1. Variance-based measures

The measures from this category evaluate the coherence of the values of a bicluster regarding their expected values predicted using quantities, such as the bicluster mean or the bicluster row and column means. In this paper, b_{iJ}, b_{Ij} and b_{IJ} stand for the mean of the i^{th} row, the mean of the j^{th} column and the mean of all elements of a bicluster B, respectively. These measures are presented next.

1. *Variance (VAR)* [27] is used to detect constant patterns:

$$\mathrm{VAR}(B) = \sum_{i=1}^{|I|}\sum_{j=1}^{|J|}(b_{ij} - b_{IJ})^2. \tag{1}$$

   Clearly, the smaller the value, the closer a bicluster is to a constant pattern.

   Time complexity analysis. The calculation of b_{IJ} costs $O(|I||J|)$. The sum of the squared terms also costs $O(|I||J|)$. Overall, the time complexity of VAR is $O(|I||J|)$.

2. *Mean Squared Residue (MSR)* [6] is based on the shifting bicluster model, and produces smaller values for biclusters that agree more with this model.

MSR is defined as:

$$\mathrm{MSR}(B) = \frac{1}{|I||J|} \sum_{i=1}^{|I|} \sum_{j=1}^{|J|} (b_{ij} - b_{iJ} - b_{Ij} + b_{IJ})^2. \tag{2}$$

Time complexity analysis. The calculations of $b_{iJ}\, \forall i \in I$, $b_{Ij}\, \forall j \in J$ and b_{IJ} require $O(|I||J|)$ steps. The sum of all squared terms also requires $O(|I||J|)$ steps. Overall, the time complexity of MSR is $O(|I||J|)$.

3. *Mean Absolute Residue (MAR)* [28] is also based on the shifting bicluster model. The only difference between MAR and MSR is that MAR takes the absolute difference between the bicluster elements and their expected values predicted by the row, column and bicluster means. It is defined as:

$$\mathrm{MAR}(B) = \frac{1}{|I||J|} \sum_{i=1}^{|I|} \sum_{j=1}^{|J|} |b_{ij} - b_{iJ} - b_{Ij} + b_{IJ}|. \tag{3}$$

Time complexity analysis. MAR has the same time of complexity of MSR, which is $O(|I||J|)$.

4. *Relevance Index (RI)* [29] identifies the constant columns pattern based on the local and global variances of the columns in the bicluster. It is formulated as:

$$\mathrm{RI}(B) = \sum_{j=1}^{|J|} R_j, \tag{4}$$

where

$$R_j = 1 - \frac{\sigma_{Ij}^2}{\sigma_j^2}, \tag{5}$$

σ_{Ij}^2 is the variance of the j^{th} column of B and σ_j^2 is the variance of the j^{th} column of the full dataset.

Time complexity analysis. The calculation of each σ_{Ij}^2 costs $O(|I|)$. The calculation of each σ_j^2 costs $O(N)$. Therefore, any R_j requires $O(|I|) + O(N) = O(N)$ steps. Since B has $|J|$ columns, the complexity of RI is $O(N\,|J|)$.

5. *Constancy by rows (C_r)* [25] quantifies the agreement of a bicluster with the constant row pattern:

$$C_r(B) = \frac{1}{|I|} \sum_{i=1}^{|I|-1} \sum_{k=i+1}^{|I|} \sqrt{\sum_{j=1}^{|J|} (b_{ij} - b_{kj})^2}. \tag{6}$$

Time complexity analysis. The sum of the squared terms costs $O(|J|)$. In a bicluster, there is a total of $(|I|\,(|I|-1))\,/\,2 = O(|I|^2)$ pairs of rows. Overall, C_r runs in $O(|I|^2|J|)$.

6. *Constancy by columns (C_c)* [25] expresses the extent to which the values of a bicluster present a constant column pattern. It is the transposed version of C_r.

Time complexity analysis. Since C_c is the transposed version of C_r, its time complexity is $O(|I||J|^2)$.

7. *Overall Constancy (OC)* [25] minimizes its value when evaluating constant biclusters. For such, it integrates the constancy by rows and the constancy by columns formulae:

$$\mathrm{OC}(B) = \frac{|I|C_r(B) + |J|C_c(B)}{|I| + |J|}. \tag{7}$$

Time complexity analysis. Since it requires the calculation of C_r and C_c, OC runs in $O(\max(|I|^2\,|J|, |I|\,|J|^2))$.

8. *Scaling Mean Squared Residue (SMSR)* [21] is a modification of the MSR measure that is able to detect scaling biclusters:

$$\mathrm{SMSR}(B) = \frac{1}{|I||J|} \sum_{i=1}^{|I|} \sum_{j=1}^{|J|} \frac{(b_{iJ}\, b_{Ij} - b_{ij}\, b_{IJ})^2}{b_{iJ}^2\, b_{Ij}^2}. \tag{8}$$

As in MSR, smaller values indicate biclusters that better suit the desired model.

Time complexity analysis. SMSR requires the same quantities as MSR and MAR ($b_{iJ}\,\forall i \in I$, $b_{Ij}\,\forall j \in J$ and b_{IJ}) to determine the differences among the values of the bicluster elements and their expected values. Therefore, the complexity of SMSR is $O(|I||J|)$.

9. *Minimal Mean Squared Error (MMSE)* [26] is based on the shifting, scaling and shifting-scaling models. Its authors argue that it is better suited than previous measures, such as MSR and SMSR, to identify negative correlated linear patterns. This measure is formally expressed as:

$$\text{MMSE}(B) = \frac{1}{|I||J|}\left[\sum_{i=1}^{|I|}\sum_{j=1}^{|J|} d_{ij}^2 - \lambda_{\max}(DD^T)\right], \tag{9}$$

where $d_{ij} = b_{ij} - b_{iJ}$, D is the matrix containing all d_{ij} elements, and $\lambda_{\max}(DD^T)$ is the eigenvalue of DD^T with maximum absolute value.
The time complexity of MMSE is $O(\min(|I|, |J|)\ |I||J|)$. The complete analysis is provided in the original paper.

3.2.2. Correlation-based measures

These measures assess the similarity between gene/sample behaviors, instead of the magnitudes or deviations among their values [7]. For such, they use either the Pearson or the Spearman correlation to measure gene/sample similarities. In this paper, the former is denoted as $r(\cdot,\cdot)$ while the latter is represented as $\rho(\cdot,\cdot)$. In addition, the i^{th} row and the j^{th} column of B are denoted as b_{i*} and b_{*j}, respectively. These measures are detailed next.

1. *Average Correlation (AC)* [23] was proposed to detect shifting, scaling and shifting-scaling biclusters by calculating the average Pearson correlation between its rows:

$$\text{AC}(B) = \frac{2}{|I|(|I|-1)}\sum_{i=1}^{|I|-1}\sum_{k=i+1}^{|I|} r(b_{i*}, b_{k*}). \tag{10}$$

Time complexity analysis. The calculation of each $r(b_{i*}, b_{k*})$ costs $O(|J|)$. There are $|I|\,(|I|-1)\,/\,2 = O(|I|^2)$ pairs of rows in B. Overall, AC requires $O(|I|^2\,|J|)$ steps.

2. *Sub-matrix Correlation Score (SCS)* [30] was proposed to detect shifting or scaling patterns. It takes into account correlations between rows and between columns. The ideal bicluster would present strong correlations

on both dimensions. SCS is formally defined as:

$$\mathrm{SCS}(B) = \min\{S_{\mathrm{row}}(B), S_{\mathrm{col}}(B)\}, \tag{11}$$

where

$$S_{\mathrm{row}}(B) = \min_{i=1,\cdots,|I|} \left\{ 1 - \frac{1}{|I|-1} \sum_{\substack{k=1 \\ k\neq i}}^{|I|} |r(b_{i*}, b_{k*})| \right\}, \tag{12}$$

$$S_{\mathrm{col}}(B) = \min_{j=1,\cdots,|J|} \left\{ 1 - \frac{1}{|J|-1} \sum_{\substack{l=1 \\ l\neq j}}^{|J|} |r(b_{*j}, b_{*l})| \right\}. \tag{13}$$

Time complexity analysis. The calculation of each $r(b_{i*}, b_{k*})$ and each $r(b_{*j}, b_{*l})$ costs $O(|J|)$ and $O(|I|)$, respectively. The calculations of all S_{row} values and all S_{col} values require $O(|I|^2\,|J|)$ and $O(|I|\,|J|^2)$ steps, respectively. Overall, the time complexity of SCS is $O(\max(|I|^2\,|J|, |I|\,|J|^2))$.

3. *Average Correlation Value (ACV)* [20] was designed to identify shifting or scaling models. For such, it gives higher values for biclusters containing rows or columns presenting a strong average Pearson correlation value:

$$\begin{aligned} \mathrm{ACV}(B) = \max \Bigg\{ & \frac{2}{|I|(|I|-1)} \sum_{i=1}^{|I|-1} \sum_{k=i+1}^{|I|} |r(b_{i*}, b_{k*})|, \\ & \frac{2}{|J|(|J|-1)} \sum_{j=1}^{|J|-1} \sum_{l=j+1}^{|J|} |r(b_{*j}, b_{*l})| \Bigg\}. \end{aligned} \tag{14}$$

Time complexity analysis. The average absolute correlation among the rows of B requires $O(|I|^2\,|J|)$ steps. The average absolute correlation between the columns of B costs $O(|I|\,|J|^2)$. Therefore, ACV runs in $O(\max(|I|^2\,|J|, |I|\,|J|^2))$.

4. *Average Spearman's Rho (ASR)* [31] was proposed to overcome any sensitivity of the ACV measure due to using the Pearson correlation. It is formulated as:

$$\begin{aligned} \mathrm{ASR}(B) = \max \Bigg\{ & \frac{2}{|I|(|I|-1)} \sum_{i=1}^{|I|-1} \sum_{k=i+1}^{|I|} \rho(b_{i*}, b_{k*}), \\ & \frac{2}{|J|(|J|-1)} \sum_{j=1}^{|J|-1} \sum_{l=j+1}^{|J|} \rho(b_{*j}, b_{*l}) \Bigg\}. \end{aligned} \tag{15}$$

Time complexity analysis. The Spearman coefficient measures the correlation between the ranks of the elements of two vectors. For such, it requires a sorting step, which can be performed in $O(n \log n)$ for n elements. Thus, each $\rho(b_{i*}, b_{k*})$ and each $\rho(b_{*j}, b_{*l})$ cost $O(|J| \log |J|)$ and $O(|I| \log |I|)$, respectively. The first argument of max runs in $O(|I|^2 |J| \log |J|)$. The latter argument of max requires $O(|J|^2 |I| \log |I|)$ steps. Overall, ASR runs in $O(\max(|I|^2 |J| \log |J|, |J|^2 |I| \log |I|))$.

5. *Spearman's Biclustering Measure (SBM)* [24] was introduced to detect shifting or scaling patterns by calculating the average Spearman correlation coefficient between the rows and columns of a bicluster and weighting their influences in the final result. Formally, this measure is defined as:

$$\text{SBM}(B) = \psi(B)\, \omega(B)\, \bar{\rho}_I(B)\, \bar{\rho}_J(B), \tag{16}$$

where

$$\bar{\rho}_I(B) = \frac{2}{|I|(|I|-1)} \sum_{i=1}^{|I|-1} \sum_{k=i+1}^{|I|} |\rho(b_{i*}, b_{k*})|, \tag{17}$$

$$\bar{\rho}_J(B) = \frac{2}{|J|(|J|-1)} \sum_{j=1}^{|J|-1} \sum_{l=j+1}^{|J|} |\rho(b_{*j}, b_{*l})|, \tag{18}$$

$\psi(B)$ and $\omega(B)$ are hyperparameters that refer to the importance of the rows and the columns of a bicluster. Their values are set by the user. In this paper, we used $\omega(B) = 1$ and

$$\psi(B) = \begin{cases} 1, & \text{if } |J| > 9, \\ \dfrac{|J|}{M}, & \text{otherwise,} \end{cases} \tag{19}$$

which are the default values used by the original authors.

Time complexity analysis. SBM is calculated in constant time after $\bar{\rho}_I(B)$ and $\bar{\rho}_J(B)$ are obtained. Therefore, SBM has the same time complexity of ASR: $O(\max(|I|^2 |J| \log |J|, |J|^2 |I| \log |I|))$.

In [7], the Pearson correlation was also included in this category. However, this measure can only be calculated between pairs of rows or pairs of columns and

not for a whole bicluster. The authors did not mention how they summarized all the Pearson correlation values for gene or sample pairs of a bicluster. The most simple approach would be to return the average value. However, this is exactly what the AC measure does. For this reason, the Pearson correlation was not considered as a bicluster coherence measure by itself in this study.

3.2.3. Standardization-based measures

These coherence measures are based on standardization evaluation of the bicluster's rows/columns tendencies by scaling their values to make them comparable [7]. Thus, these measures are calculated on the standardized bicluster B', whose elements are defined as:

$$b'_{ij} = \frac{b_{ij} - \mu_i}{\sigma_i}, \tag{20}$$

where μ_i and σ_i are the mean and the standard deviation of the i^{th} row (gene) of B, respectively. These measures are detailed next.

1. *Maximal Standard Area (MSA)* [19] defines a band for the set of columns of a bicluster, which corresponds to the maximum and minimum values of each column. The value of MSA is the total area of this band. This measure, which has been applied to detect shifting or scaling bicluster patterns, is defined as:

$$\text{MSA}(B) = \sum_{j=1}^{|J|-1} \left| \frac{\max_j^{B'} - \min_j^{B'} + \max_{j+1}^{B'} - \min_{j+1}^{B'}}{2} \right|, \tag{21}$$

where $\max_j^{B'}$ and $\min_j^{B'}$ correspond to the maximum and minimum values of the jth column of B', respectively.

Time complexity analysis. $\max_j^{B'}$ and $\min_j^{B'}$ require $O(|I|)$ steps. Since we have $|J|$ columns in the bicluster, MSA runs in $O(|I|\,|J|)$.

2. *Virtual Error (VE)* [18] calculates the difference between the bicluster elements and a virtual row (gene) pattern that captures the general trend of the bicluster values [7]. It is minimized when evaluating biclusters with

shifting or scaling patterns. This measure is defined as:

$$\mathrm{VE}(B) = \frac{1}{|I||J|} \sum_{i=1}^{|I|} \sum_{j=1}^{|J|} |b'_{ij} - p'_j|, \tag{22}$$

where p is the mean row vector of B, and p' is its standardized version.

Time complexity analysis. p requires $O(|I||J|)$ steps to be calculated. The standardization of B takes $O(|I||J|)$ steps. The standardization of p costs $O(|J|)$. The absolute differences between the elements of B' and the elements of p' require $O(|I||J|)$. Overall, VE runs in $O(|I||J|)$.

3. *Transposed Virtual Error (VEt)* [22] is the VE measure applied in B^T. It is able to detect all the patterns identified by VE and also the shifting-scaling pattern.

Time complexity analysis. VEt requires the same number of steps as VE: $O(|I||J|)$.

Table 1: Summary of the investigated measures.

Category	Measure	Reference	Range	Objective	Time complexity
Variance-based	VAR	[27]	$[0, \infty)$	Min	$O(\lvert I\rvert \lvert J\rvert)$
	MSR	[6]	$[0, \infty)$	Min	$O(\lvert I\rvert \lvert J\rvert)$
	MAR	[28]	$[0, \infty)$	Min	$O(\lvert I\rvert \lvert J\rvert)$
	RI	[29]	$(-\infty, \lvert J\rvert]$	Max	$O(N \lvert J\rvert)$
	C_r	[25]	$[0, \infty)$	Min	$O(\lvert I\rvert^2 \lvert J\rvert)$
	C_c	[25]	$[0, \infty)$	Min	$O(\lvert I\rvert \lvert J\rvert^2)$
	OC	[25]	$[0, \infty)$	Min	$O(\max(\lvert I\rvert^2 \lvert J\rvert, \lvert I\rvert \lvert J\rvert^2))$
	MMSE	[26]	$[0, \infty)$	Min	$O(\min(\lvert I\rvert, \lvert J\rvert) \lvert I\rvert \lvert J\rvert)$
	SMSR	[21]	$[0, \infty)$	Min	$O(\lvert I\rvert \lvert J\rvert)$
Correlation-based	AC	[23]	$[-1, 1]$	Max	$O(\lvert I\rvert^2 \lvert J\rvert)$
	SCS	[30]	$[0, 1]$	Min	$O(\max(\lvert I\rvert^2 \lvert J\rvert, \lvert I\rvert \lvert J\rvert^2))$
	ACV	[20]	$[0, 1]$	Max	$O(\max(\lvert I\rvert^2 \lvert J\rvert, \lvert I\rvert \lvert J\rvert^2))$
	ASR	[31]	$[-1, 1]$	Max	$O(\max(\lvert I\rvert^2 \lvert J\rvert \log \lvert J\rvert, \lvert J\rvert^2 \lvert I\rvert \log \lvert I\rvert))$
	SBM	[24]	$[0, \infty)$	Max	$O(\max(\lvert I\rvert^2 \lvert J\rvert \log \lvert J\rvert, \lvert J\rvert^2 \lvert I\rvert \log \lvert I\rvert))$
Standardization-based	MSA	[19]	$[0, \infty)$	Min	$O(\lvert I\rvert \lvert J\rvert)$
	VE	[18]	$[0, \infty)$	Min	$O(\lvert I\rvert \lvert J\rvert)$
	VEt	[22]	$[0, \infty)$	Min	$O(\lvert I\rvert \lvert J\rvert)$

3.3. Algorithms

To investigate the behavior of the coherence measures, we selected 10 biclustering algorithms often used in the literature, which have already been extensively studied and have free implementations which are publicly available. These algorithms are based on different formulations and use diverse types of heuristics (e.g., greedy, divide-and-conquer, exhaustive enumeration, etc.) to deal with biclustering tasks. Hence, they are able to identify different types of bicluster patterns and bicluster structures (e.g., exclusive row or column biclusters, non-overlapping biclusters in checkerboard structures, arbitrarily positioned biclusters, etc.). Thus, they model different particularities of a dataset and reduce the bias towards a specific coherence measure when evaluating the identified biclusters. These algorithms are:

- *Cheng and Church's Algorithm (CCA)* [6], which starts with the full data matrix as a bicluster. Next, it iteratively prunes rows and columns out of the bicluster, minimizing the MSR measure, until it satisfies a desired threshold. As a last step, some rows or columns are added back to the bicluster as long as they do not violate the MSR threshold.

- *Statistical-Algorithmic Method for Bicluster Analysis (SAMBA)* [4], which constructs a bipartite graph for the dataset, where one set of vertices represents the genes and the other set corresponds to the samples. Next, based on a likelihood model, it enumerates the most significant complete bipartite subgraphs (bicliques). Each biclique corresponds to a bicluster in the final solution.

- *Order Preserving Sub-Matrix (OPSM)* [32], which mines biclusters containing columns that induce a permutation where the values of each row strictly increases. The search procedure is performed by a greedy heuristic guided by a probabilistic score.

- *Spectral* [33] which searches for constant biclusters organized in a checkerboard structure. For such, it applies the singular value decomposition to

the input matrix. Then, it clusters rows and columns independently by projecting them on their best partitioning eigenvectors and applying the k-means algorithm.

- *Plaid* [34], which represents a set of biclusters as a sum of linear layers plus an additional layer that models noise and background effects in the data. The optimization problem consists of a sum of squared errors minimization between the plaid model and the data, which is solved by a binary least squares algorithm.

- *Binary Inclusion Maximal Biclustering Algorithm (Bimax)* [11], which discretizes the input dataset into a binary matrix based on the threshold $(\min(A) + \max(A)) / 2$, where $\min(A)$ and $\max(A)$ indicate the maximum and minimum values of the matrix. Next, it searches for upregulated biclusters whose values are all equal to one, using an enumerative divide-and-conquer approach.

- *Bayesian Biclustering (BBC)* [35], which assumes the plaid model for the input dataset, but restricts the overlap between biclusters to occur only in genes or only in samples. For the plaid model fitting, it uses a Gibbs sampling procedure.

- *Large Average Submatrices (LAS)* [36], which assumes a Gaussian random matrix as a null model for the data and searches for biclusters with average values that significantly deviate from such a model. For such, it uses a greedy procedure to optimize a Bonferroni-based significance score that takes into account the size of a bicluster and its average value.

- *Qualitative Biclustering (QUBIC)* [37], which represents the data as a graph, with genes as vertices, edge weights equal to the number of samples for which two genes are similar. The algorithm consists of a greedy procedure that extracts biclusters that correspond to heavy subgraphs where the genes present similar expression patterns in the same subset of samples.

- *Factor Analysis for Bicluster Acquisition (FABIA)* [38], which assumes a sum of multiplicative layers for a dataset, where each layer represents a different bicluster, plus a noise layer. To fit this model, FABIA uses an expectation-maximization approach for likelihood maximization.

In Table 2, we summarize the software packages used to implement the algorithms used in the experiments of this paper. The algorithms are available in R, Java, C and Python packages. In the experimental phase, we used biclustlib [15], which is a Python library that provides wrappers for these implementations.

Table 2: Algorithms' software packages.

Algorithm	Language	Availability
CCA	R	https://cran.r-project.org/web/packages/biclust/index.html
SAMBA	Java	http://acgt.cs.tau.ac.il/expander/
OPSM	Java	https://sop.tik.ee.ethz.ch/bicat/
Spectral	Python	https://scikit-learn.org/stable/
Plaid	R	https://cran.r-project.org/web/packages/biclust/index.html
Bimax	R	https://cran.r-project.org/web/packages/biclust/index.html
BBC	C	http://www.people.fas.harvard.edu/~junliu/BBC/
LAS	Python	https://github.com/padilha/biclustlib
QUBIC	C	https://github.com/maqin2001/qubic
FABIA	Python	https://github.com/bioinf-jku/pyfabia

3.4. Data Collection

The experiments were performed using 19 datasets associated with the *Saccharomyces cerevisiae* organism, one of the organisms most comprehensively studied in biology and, as a consequence, with extensive and high-quality Gene Ontology information available [11, 39]. This collection consists of the main biclustering benchmarks of this organism available in the literature. They are represented by dense real-valued data matrices obtained from time series mi-

croarray experiments, including the datasets used in [6][1] and [11][2], included in most biclustering studies, and the benchmark of 17 datasets introduced by [40][3], whose data were systematically collected from previous gene expression data analyses studies [41, 42, 43] and were already used in clustering [44] and biclustering [15] analyses. The main aspects of these datasets are summarized in Table 3.

Table 3: Gene expression datasets.

Name	# of genes	# of samples	Reference
Alpha factor	1099	18	[40]
Cdc 15	1086	24	[40]
Cdc 28	1044	17	[40]
Elutriation	935	14	[40]
1mM menadione	1050	9	[40]
1M sorbitol	1030	7	[40]
1.5mM diamide	1038	8	[40]
2.5mM DTT	991	8	[40]
Constant 32nM H2O2	976	10	[40]
Diauxic shift	1016	7	[40]
Complete DTT	962	7	[40]
Heat shock 1	988	8	[40]
Heat shock 2	999	7	[40]
Nitrogen depletion	1011	10	[40]
YPD 1	1011	12	[40]
YPD 2	1022	10	[40]
Yeast sporulation	1171	7	[40]
S. cerevisiae	2993	173	[11]
Tavazoie	2884	17	[6]

[1] http://arep.med.harvard.edu/biclustering/
[2] https://sop.tik.ee.ethz.ch/bimax/
[3] http://lapad-web.icmc.usp.br/repositories/ieee-tcbb-2013/index.html

3.5. External Bicluster Evaluation Measures

For the external evaluation, we performed the gene enrichment analysis of the biclusters found using the Gene Ontology (GO)[4] [45] knowledge base, which provides three ontologies: Biological Process, Molecular Function and Cellular Component. Each ontology contains a structured general vocabulary comprising "is-a" and "part-of" relationships between its terms to describe the role of the genes in an organism [46].

In this study, we performed four different analyses using the GO database: (i) using all the three ontologies; (ii) using only the Biological Process ontology; (iii) using only the Cellular Component ontology; and (iv) using only the Molecular Function ontology. For each analysis, after identifying the GO terms in each bicluster, the Fisher test was applied to assess the over-representation of each term [4, 11, 37, 15]. In this study, a GO term was considered significant in a bicluster if its p-value, after performing the Benjamini and Hochberg multiple test correction [47], was lower than 0.05 [13, 15]. Three different measures were extracted for each bicluster containing at least one significant term [7]: the mean p-value, the best p-value and the number of significant terms. The experiments investigated correlations of these quantities with the coherence measures discussed in Section 3.2.

3.6. Experimental methodology

Briefly, the experimental methodology has 6 steps, which are illustrated in Figure 1. We will now explain each step.

Given a dataset X_i, two different scenarios were considered in step (1) before applying any algorithm and coherence measure. In the first scenario, the features (samples) of each dataset were standardized to zero mean and unit variance. In the latter scenario, the algorithms and coherence measures were applied to the original (non-normalized) data.

[4] http://www.geneontology.org/

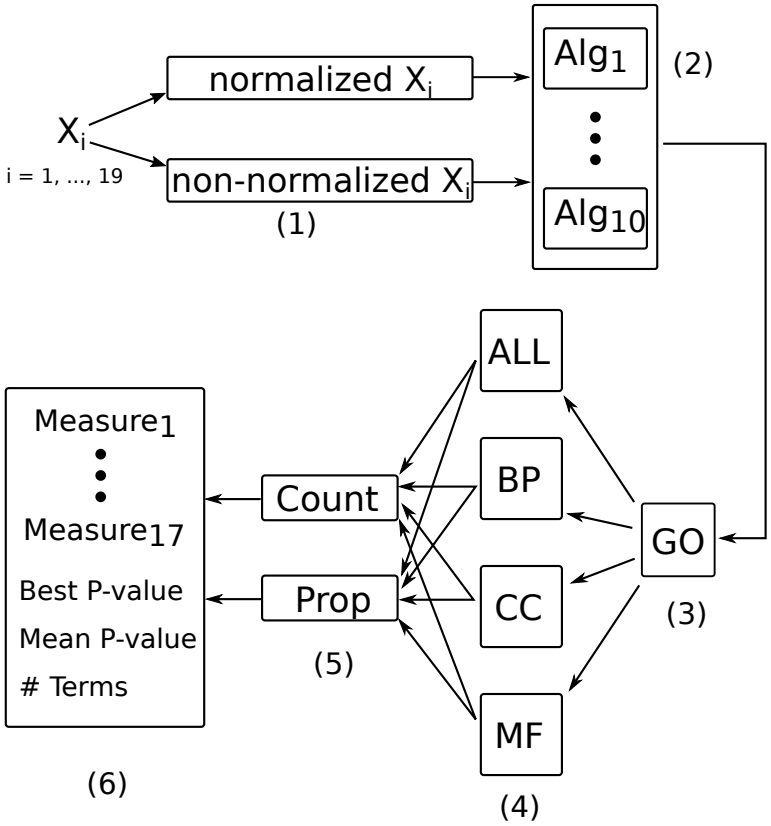

Figure 1: Experimental methodology followed to obtain the results of the coherence and GO measures.

In step (2), the selected algorithms were run on both versions of X_i. Deterministic and non deterministic algorithms were selected for this study. For each dataset, the deterministic algorithms (SAMBA, OPSM, Bimax and QUBIC) were run once, while the non deterministic algorithms (CCA, Spectral, Plaid, BBC, LAS and FABIA) were run 30 times.

In step (3), the biclusterings found by each algorithm were compared with the GO external evaluation. Four different scenarios were considered for the GO evaluation, which are illustrated in step (4): (i) using all GO ontologies (ALL); (ii) using only the Biological Process ontology (BP); (iii) using only the Cellular Component ontology (CC); and (iv) using only the Molecular Function

ontology (MF).

Given that 6 out of the 10 investigated algorithms are non deterministic, a pre-established procedure was adopted to select which of their biclusterings would be analyzed for each dataset. In step (5), two different approaches were followed. The first, called "count", selects, for each dataset, the biclustering solution that contains the median number of significant biclusters. The second, called "prop", selects, for each dataset, the biclustering solution that contains the median proportion of significant biclusters for the total number of biclusters in the solution[5]. Both approaches do not discard empty biclustering solutions to calculate the median.

Finally, in step (6), the 17 coherence measures from Section 3.2 and the 3 GO measures from Section 3.5 are calculated for each bicluster containing at least one significant GO term, according to the GO scenario being considered.

3.7. Hyperparameter values used for the algorithms

The hyperparameter values used in this study were usually based on the default settings used or recommended by the original authors of each algorithm. However, to achieve results that best fit the investigated scenarios, they were modified for some of the biclustering techniques. These modifications are explained next.

CCA requires a maximum MSR threshold δ to produce biclusters. This quantity is usually different from one dataset to another. In this paper, $\delta = (\max(A) - \min(A))^2/12 \times 0.005$ [48], where $\max(A)$ and $\min(A)$ indicate the maximum and minimum values of a dataset, respectively. This setting provides an approximation for the δ values considered in the original work of Cheng and Church [6].

Before running its Gibbs sampling procedure, BBC normalizes the dataset. The interquartile range normalization (IQRN) on the features proposed by its

[5]This approach is different than "count" because it is not guaranteed that the heuristic adopted by each algorithm will always return the same number of biclusters.

original authors was not used here. Instead, we used the zero mean and unit variance normalization for the scenario of normalized data, to be in accordance with the other algorithms used in this study.

For the number of biclusters, 7 algorithms (CCA, Plaid, Bimax, BBC, LAS, QUBIC and FABIA) were executed to search for 30 biclusters in each dataset. Spectral was run to search for 15 gene clusters and 2 sample clusters. The other algorithms (SAMBA and OPSM) do not receive the number of biclusters as a hyperparameter. Thus, all biclusters returned by them were considered.

4. Results and discussion

Overall, we evaluated 16 different experimental scenarios, by combining: 2 versions of the datasets (normalized and non-normalized), 4 ontology analyses (ALL, BP, CC and MF), and 2 approaches to select biclusterings generated by non deterministic algorithms ("count" and "prop"). For each scenario, the biclusters found by all algorithms in all datasets were initially concatenated in an array. Next, the Pearson and Spearman correlations were calculated for the previously discussed coherence and external measures. The results are illustrated as heatmaps in Figures 2 and 3, where each element corresponds to the correlation value. To save space, only the correlations with the "count" approach, normalized data, and the three GO ontologies (ALL) are shown. The other 15 scenarios achieved similar results in most cases, allowing us to draw similar conclusions. Their respective figures are available in our supplementary material[6]. Minor differences are discussed in the text.

According to Figures 2 and 3, the measures from the external evaluation are not strongly correlated with any coherence criterion. These results were observed for all investigated scenarios. Therefore, biclusters with high biological significance from the GO point of view do not necessarily imply in good values for the coherence measures. Thus, it may be feasible to recommend using mul-

[6] http://padilha.github.io/asoc-2019-suppl

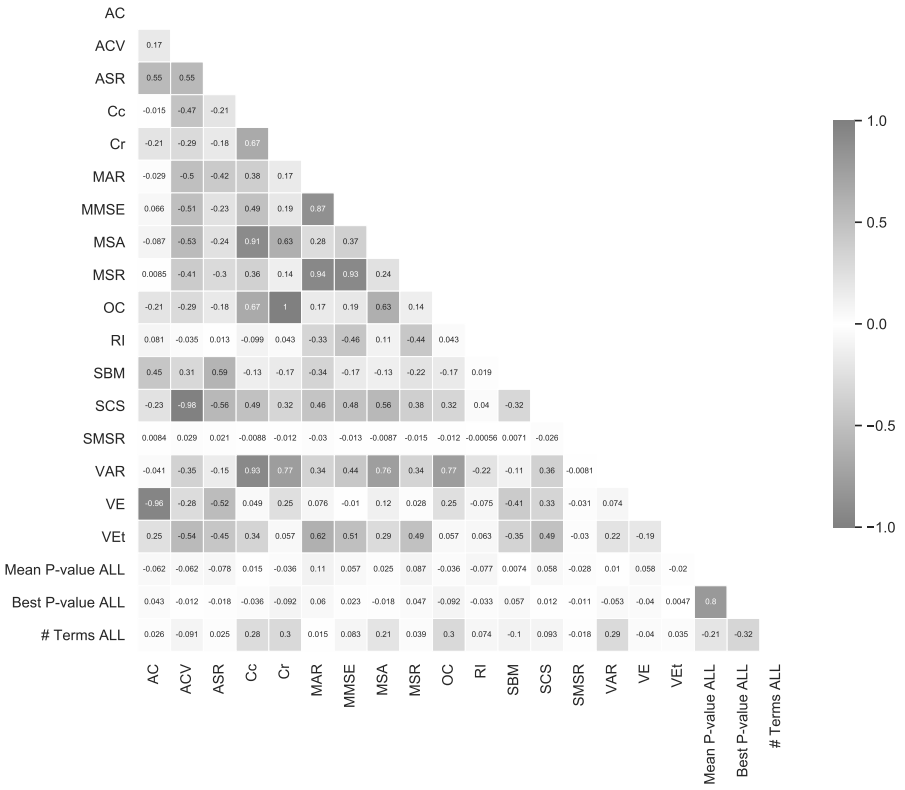

Figure 2: Results of the Pearson correlation using normalized data, all ontologies and the "count" approach.

tiple bicluster coherence criteria to complement the GO analysis. As a result, the biclusters will also be evaluated by a set of predefined patterns of interest and one can carefully inspect the quality of their trends.

It can be seen that some coherence criteria presented similar behavior according to the correlations. Measures that must be either maximized or minimized were selected. Thus, the interest is in strong correlations that can be either positive or negative. From the results, a few pairs of strongly correlated measures, with a correlation above 0.9 or below -0.9, can be extracted:

- (OC, C_r), (SCS, ACV) and (VE, AC) for both correlation coefficients in all experimental scenarios;

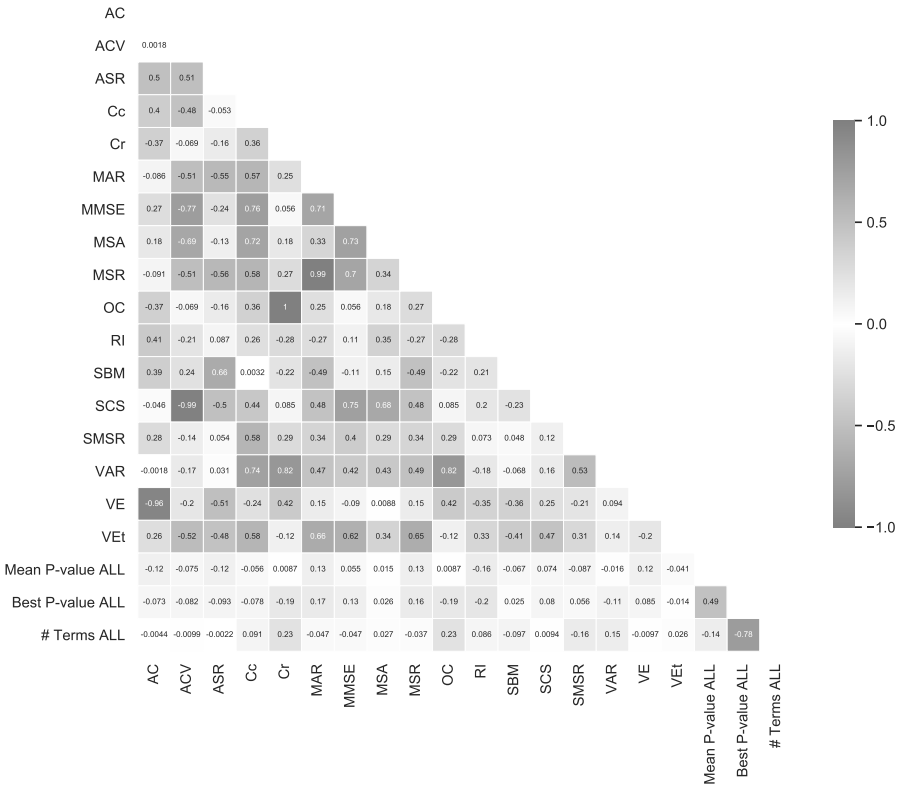

Figure 3: Results of the Spearman correlation using normalized data, all ontologies and the "count" approach.

- (MSR, MAR) for Spearman in all scenarios and for Pearson in all scenarios with normalized data and in the scenario with non-normalized data and MF analysis;
- (VAR, C_r) and (VAR, OC) for the Pearson coefficient in all scenarios using non-normalized data;
- (VAR, C_c) and (MSA, C_c) for the Pearson coefficient in all scenarios with normalized data; and
- (MSR, MMSE) for the Pearson correlation in all experimental scenarios.

It can be observed that the strong correlated pairs contain measures that

detect similar patterns. To avoid using paired criteria in the same application, since their results will be redundant, the one that is able to detect the most general numeric patterns is recommended.

Some evidence can be found that data normalization may be determinant in the behavior of some pairs of measures. This result was expected, since the algorithms do not return the same biclustering solutions when dealing with non-normalized or normalized data. Moreover, normalized data may alleviate the influence of different feature scales or outliers in the behaviors of the measures.

In addition, the correlations between measures might be different when considering different ontology scenarios for the enrichment, as was observed for (MSR, MAR) and the Pearson correlation.

Real applications may benefit from favouring measures with the lowest computational complexities. Table 1 summarizes the investigated measures and their computational complexities. Even if two coherence measures present lower correlations (e.g., around 0.7 or 0.8), those with lower complexity should be preferred, especially if a large number of biclusters is evaluated. From this table, the measures with the lowest complexities are: VAR, MSR, MAR, SMSR, MSA, VE and VEt.

The difference between the results from the Pearson and Spearman correlations, shown in Figure 4, were also analyzed. The difference observed was low for many pairs, which indicates that the two correlations were compatible in most cases. To statistically validate this finding, the Wilcoxon signed-rank test was applied to the difference matrix. Under a significance level of 0.05 no statistical evidence of difference was found, which supports the agreement of the correlation matrices. We repeated the Wilcoxon signed-rank test on each of the other 15 experimental scenarios. In all cases, we did not find statistical evidence to reject the null hypothesis.

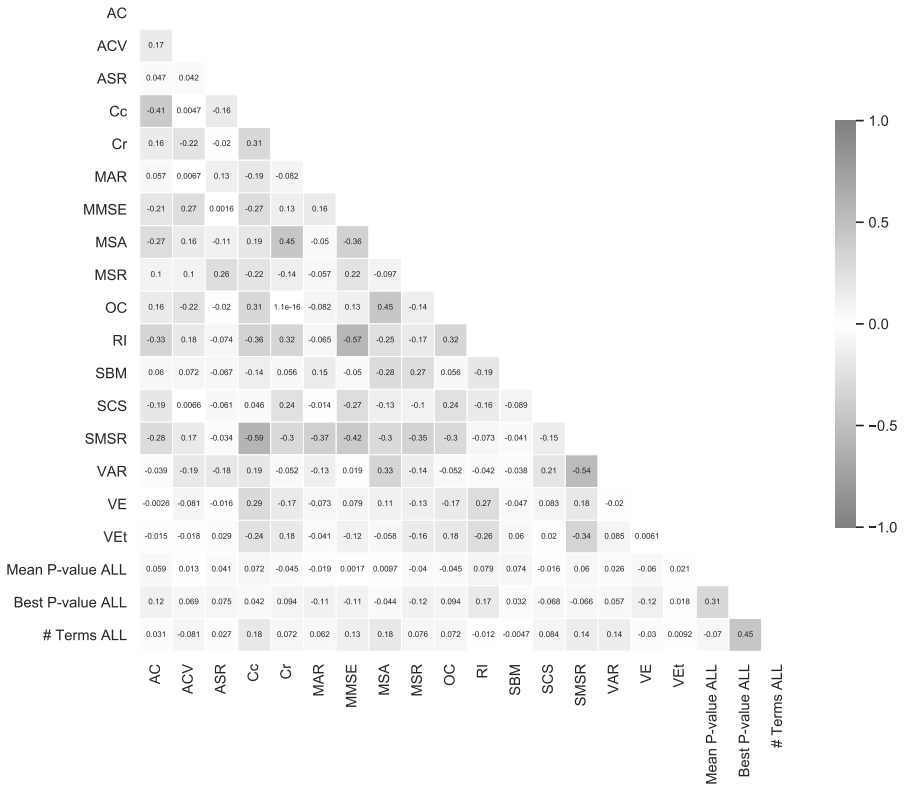

Figure 4: Difference between Pearson and Spearman correlations.

5. Conclusions

This paper extended the work of [7] by investigating the behavior of 17 bicluster coherence measures. We applied them to the results of 10 well-established biclustering algorithms. Our experiments were performed on a benchmark of 19 *Saccharomyces cerevisiae* time-course datasets.

The correlations among the coherence and the external GO criteria were analyzed using the Pearson and Spearman coefficients. According to the analysis, external GO evaluations did not agree with any coherence measure. These results suggest that a high GO significance does not automatically imply in good evaluations with coherence criteria. Besides, GO information may be incomplete [25]. Thus, the use of bicluster coherence measures together with the GO

analysis may be a better alternative to achieve more concrete conclusions.

These results conflict with those from [7], which claimed that the coherence measures present some dependence with the external biological measures. However, since this study employed 10 different algorithms, it reduced the bias regarding the evolutionary algorithm used in [7].

Overall, we analyzed 16 different experimental scenarios, which included: normalized and non-normalized data, evaluation using all GO ontologies, and 2 different approaches to select the results of non deterministic algorithms. We observed that normalization and the GO validation approach may be determinant, since some pairs of measures presented strong Pearson correlations in scenarios using either normalized or non-normalized data and specific ontologies for the enrichment.

In practical applications, the users of the measures must take into account the types of correlations among measures that they want to avoid. For such, we advise them to consider as similar only the pairs that presented a strong correlation in all scenarios for the desired coefficient (Pearson or Spearman) and data type (normalized or non-normalized).

This study also presented the time complexity analyses of the measures, usually not provided in their original studies. In many applications, the time complexities may be an important reason for choosing some measures rather than others. Mainly when a large number of biclusters need to be evaluated and/or the biclusters may be constituted by a large number of rows and columns, measures with the lowest complexities may be preferred.

Finally, the choice of the most appropriate bicluster coherence measure must also take into account the task to be solved. In a few practical scenarios, one may favor particular types of patterns compared to others and/or may prioritize measures with lower computational complexities. However, the use of heterogeneous measures allows the analysis of biclusters with different points of view. According to the experimental results reported in this paper, it is possible to avoid selecting a set of measures that present redundant behavior and may not bring new insights to the analysis.

Acknowledgements

The authors would like to thank the São Paulo Research Foundation (FAPESP), the Brazilian Federal Agency for Support and Evaluation of Graduate Education within the Ministry of Education of Brazil (CAPES), and Intel for the funding support.

Funding

This work was supported by FAPESP (grants #2017/02975-0, #2016/18615-0 and #2013/07375-0), the International Cooperation Program Probral CAPES/ DAAD (grant #88887.302257/2018-00), and Intel.

Competing interests

Declarations of interest: none.

References

References

[1] P. O. Brown, D. Botstein, Exploring the new world of the genome with dna microarrays, Nature genetics 21 (1s) (1999) 33.

[2] S. C. Madeira, A. L. Oliveira, Biclustering algorithms for biological data analysis: a survey, IEEE/ACM TCBB 1 (1) (2004) 24–45.

[3] A. Ben-Dor, R. Shamir, Z. Yakhini, Clustering gene expression patterns, Journal of computational biology 6 (3-4) (1999) 281–297.

[4] A. Tanay, R. Sharan, R. Shamir, Discovering statistically significant biclusters in gene expression data, Bioinformatics 18 (suppl_1) (2002) S136–S144.

[5] D. Jiang, C. Tang, A. Zhang, Cluster analysis for gene expression data: A survey, IEEE TKDE 16 (11) (2004) 1370–1386.

[6] Y. Cheng, G. M. Church, Biclustering of expression data., in: ISMB, Vol. 8, AAAI Press, 2000, pp. 93–103.

[7] B. Pontes, R. Girldez, J. S. Aguilar-Ruiz, Quality measures for gene expression biclusters, PloS one 10 (3) (2015) e0115497.

[8] A. Tanay, R. Sharan, R. Shamir, Biclustering algorithms: A survey, Handbook of computational molecular biology 9 (1-20) (2005) 122–124.

[9] S. Busygin, O. Prokopyev, P. M. Pardalos, Biclustering in data mining, Computers & Operations Research 35 (9) (2008) 2964–2987.

[10] B. Pontes, R. Giráldez, J. S. Aguilar-Ruiz, Biclustering on expression data: A review, Journal of biomedical informatics 57 (2015) 163–180.

[11] A. Prelić, S. Bleuler, P. Zimmermann, A. Wille, P. Bühlmann, W. Gruissem, L. Hennig, L. Thiele, E. Zitzler, A systematic comparison and evaluation of biclustering methods for gene expression data, Bioinformatics 22 (9) (2006) 1122–1129.

[12] D. Bozdağ, A. S. Kumar, U. V. Catalyurek, Comparative analysis of biclustering algorithms, in: Proceedings of the First ACM International Conference on Bioinformatics and Computational Biology, ACM, 2010, pp. 265–274.

[13] K. Eren, M. Deveci, O. Küçüktunç, Ü. V. Çatalyürek, A comparative analysis of biclustering algorithms for gene expression data, Briefings in bioinformatics 14 (3) (2012) 279–292.

[14] A. Oghabian, S. Kilpinen, S. Hautaniemi, E. Czeizler, Biclustering methods: biological relevance and application in gene expression analysis, PloS one 9 (3) (2014) e90801.

[15] V. A. Padilha, R. J. G. B. Campello, A systematic comparative evaluation of biclustering techniques, BMC bioinformatics 18 (1) (2017) 55.

[16] V. A. Padilha, A. C. P. L. F. Carvalho, A study of biclustering coherence measures for gene expression data, in: 7th Brazilian Conference on Intelligent Systems, IEEE, 2018, pp. 546–551.

[17] J. S. Aguilar-Ruiz, Shifting and scaling patterns from gene expression data, Bioinformatics 21 (20) (2005) 3840–3845.

[18] F. Divina, B. Pontes, R. Giráldez, J. S. Aguilar-Ruiz, An effective measure for assessing the quality of biclusters, Computers in biology and medicine 42 (2) (2012) 245–256.

[19] R. Giraldez, F. Divina, B. Pontes, J. S. Aguilar-Ruiz, Evolutionary search of biclusters by minimal intrafluctuation, in: FUZZ-IEEE, IEEE, 2007, pp. 1–6.

[20] L. Teng, L. Chan, Discovering biclusters by iteratively sorting with weighted correlation coefficient in gene expression data, Journal of Signal Processing Systems 50 (3) (2008) 267–280.

[21] A. Mukhopadhyay, U. Maulik, S. Bandyopadhyay, A novel coherence measure for discovering scaling biclusters from gene expression data, Journal of bioinformatics and computational biology 7 (05) (2009) 853–868.

[22] B. Pontes, R. Giráldez, J. S. Aguilar-Ruiz, Measuring the quality of shifting and scaling patterns in biclusters, in: IAPR PRIB, Springer, 2010, pp. 242–252.

[23] J. A. Nepomuceno, A. Troncoso, J. S. Aguilar-Ruiz, Biclustering of gene expression data by correlation-based scatter search, BioData mining 4 (1) (2011) 3.

[24] J. L. Flores, I. Inza, P. Larrañaga, B. Calvo, A new measure for gene expression biclustering based on non-parametric correlation, Computer methods and programs in biomedicine 112 (3) (2013) 367–397.

[25] R. Santamaría, L. Quintales, R. Therón, Methods to bicluster validation and comparison in microarray data, in: International Conference on Intelligent Data Engineering and Automated Learning, Springer, 2007, pp. 780–789.

[26] S. Chen, J. Liu, T. Zeng, Measuring the quality of linear patterns in biclusters, Methods 83 (2015) 18–27.

[27] J. A. Hartigan, Direct clustering of a data matrix, Journal of the american statistical association 67 (337) (1972) 123–129.

[28] J. Yang, W. Wang, H. Wang, P. Yu, δ-clusters: capturing subspace correlation in a large data set, in: IEEE ICDE, IEEE, 2002, pp. 517–528.

[29] K. Y. Yip, D. W. Cheung, M. K. Ng, Harp: A practical projected clustering algorithm, IEEE TKDE 16 (11) (2004) 1387–1397.

[30] W.-H. Yang, D.-Q. Dai, H. Yan, Finding correlated biclusters from gene expression data, IEEE TKDE 23 (4) (2011) 568–584.

[31] W. Ayadi, M. Elloumi, J.-K. Hao, A biclustering algorithm based on a bicluster enumeration tree: application to dna microarray data, BioData mining 2 (1) (2009) 9.

[32] A. Ben-Dor, B. Chor, R. Karp, Z. Yakhini, Discovering local structure in gene expression data: the order-preserving submatrix problem, Journal of computational biology 10 (3-4) (2003) 373–384.

[33] Y. Kluger, R. Basri, J. T. Chang, M. Gerstein, Spectral biclustering of microarray data: coclustering genes and conditions, Genome research 13 (4) (2003) 703–716.

[34] H. Turner, T. Bailey, W. Krzanowski, Improved biclustering of microarray data demonstrated through systematic performance tests, Computational statistics & data analysis 48 (2) (2005) 235–254.

[35] J. Gu, J. S. Liu, Bayesian biclustering of gene expression data, BMC genomics 9 (1) (2008) S4.

[36] A. A. Shabalin, V. J. Weigman, C. M. Perou, A. B. Nobel, et al., Finding large average submatrices in high dimensional data, The Annals of Applied Statistics 3 (3) (2009) 985–1012.

[37] G. Li, Q. Ma, H. Tang, A. H. Paterson, Y. Xu, Qubic: a qualitative biclustering algorithm for analyses of gene expression data, Nucleic acids research 37 (15) (2009) e101–e101.

[38] S. Hochreiter, U. Bodenhofer, M. Heusel, A. Mayr, A. Mitterecker, A. Kasim, T. Khamiakova, S. Van Sanden, D. Lin, W. Talloen, et al., Fabia: factor analysis for bicluster acquisition, Bioinformatics 26 (12) (2010) 1520–1527.

[39] K. R. Christie, E. L. Hong, J. M. Cherry, Functional annotations for the saccharomyces cerevisiae genome: the knowns and the known unknowns, Trends in microbiology 17 (7) (2009) 286–294.

[40] P. A. Jaskowiak, R. J. G. B. Campello, I. G. Costa, Proximity measures for clustering gene expression microarray data: a validation methodology and a comparative analysis, IEEE/ACM TCBB 10 (4) (2013) 845–857.

[41] P. T. Spellman, G. Sherlock, M. Q. Zhang, V. R. Iyer, K. Anders, M. B. Eisen, P. O. Brown, D. Botstein, B. Futcher, Comprehensive identification of cell cycle–regulated genes of the yeast saccharomyces cerevisiae by microarray hybridization, Molecular biology of the cell 9 (12) (1998) 3273–3297.

[42] S. Chu, J. DeRisi, M. Eisen, J. Mulholland, D. Botstein, P. O. Brown, I. Herskowitz, The transcriptional program of sporulation in budding yeast, Science 282 (5389) (1998) 699–705.

[43] A. P. Gasch, P. T. Spellman, C. M. Kao, O. Carmel-Harel, M. B. Eisen, G. Storz, D. Botstein, P. O. Brown, Genomic expression programs in the

response of yeast cells to environmental changes, Molecular biology of the cell 11 (12) (2000) 4241–4257.

[44] P. A. Jaskowiak, R. J. Campello, I. G. Costa, On the selection of appropriate distances for gene expression data clustering, in: BMC bioinformatics, Vol. 15, BioMed Central, 2014, p. S2.

[45] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, et al., Gene ontology: tool for the unification of biology, Nature genetics 25 (1) (2000) 25.

[46] G. O. Consortium, The gene ontology (go) database and informatics resource, Nucleic acids research 32 (suppl_1) (2004) D258–D261.

[47] Y. Hochberg, Y. Benjamini, More powerful procedures for multiple significance testing, Statistics in medicine 9 (7) (1990) 811–818.

[48] D. Horta, R. J. G. B. Campello, Similarity measures for comparing biclusterings, IEEE/ACM TCBB 11 (5) (2014) 942–954.