APPENDIX

We define here two measures. Let $O \triangleq \{\tilde{o}_j\}_{j=1}^n$ be a set of objects, and let P and $\dot{P}$ be two soft clusterings. Let $\alpha_P(\tilde{o}_{j_1}, \tilde{o}_{j_2})$ be the number of clusters shared by objects $\tilde{o}_{j_1}$ and $\tilde{o}_{j_2}$ in clustering P, and let $\alpha_{\dot{P}}(\tilde{o}_{j_1}, \tilde{o}_{j_2})$ be similarly defined. The extended bcubed (EBC) measure [33] is based on extended bcubed precision and extended bcubed recall:

$$\mathrm{EBCP}(P,\dot{P}) \triangleq \frac{1}{n}\sum_{j_1=1}^{n} \frac{\sum_{j_2=1}^{n} \min\{\alpha_P(\tilde{o}_{j_1}, \tilde{o}_{j_2}), \alpha_{\dot{P}}(\tilde{o}_{j_1}, \tilde{o}_{j_2})\}}{\sum_{j_2=1}^{n} \alpha_P(\tilde{o}_{j_1}, \tilde{o}_{j_2})} \tag{14a}$$

$$\mathrm{EBCR}(P,\dot{P}) \triangleq \frac{1}{n}\sum_{j_1=1}^{n} \frac{\sum_{j_2=1}^{n} \min\{\alpha_P(\tilde{o}_{j_1}, \tilde{o}_{j_2}), \alpha_{\dot{P}}(\tilde{o}_{j_1}, \tilde{o}_{j_2})\}}{\sum_{j_2=1}^{n} \alpha_{\dot{P}}(\tilde{o}_{j_1}, \tilde{o}_{j_2})}. \tag{14b}$$

EBC is defined by default as

$$\mathrm{EBC}(P,\dot{P}) \triangleq 2 \cdot \frac{\mathrm{EBCP}(P,\dot{P}) \cdot \mathrm{EBCR}(P,\dot{P})}{\mathrm{EBCP}(P,\dot{P}) + \mathrm{EBCR}(P,\dot{P})}. \tag{15}$$

The other measure is defined as follows. Let $\beta_P(\tilde{o}_j)$ be the number of clusters to which object $\tilde{o}_j$ belongs in P minus 1, and let $\beta_{\dot{P}}(\tilde{o}_j)$ be similarly defined. The agreements and disagreements associated with a pair (j_1, j_2) are

$$a_G^{P,\dot{P}}(\tilde{o}_{j_1}, \tilde{o}_{j_2}) = \min\left\{\alpha_P(\tilde{o}_{j_1}, \tilde{o}_{j_2}), \alpha_{\dot{P}}(\tilde{o}_{j_1}, \tilde{o}_{j_2})\right\} + \sum_{i=1}^{2} \min\left\{\beta_P(\tilde{o}_{j_i}), \beta_{\dot{P}}(\tilde{o}_{j_i})\right\} \tag{16}$$

and

$$d_G^{P,\dot{P}}(\tilde{o}_{j_1}, \tilde{o}_{j_2}) = |\alpha_P(\tilde{o}_{j_1}, \tilde{o}_{j_2}) - \alpha_{\dot{P}}(\tilde{o}_{j_1}, \tilde{o}_{j_2})| + \sum_{i=1}^{2} |\beta_P(\tilde{o}_{j_i}) - \beta_{\dot{P}}(\tilde{o}_{j_i})|, \tag{17}$$

respectively. Summing up $a_G^{P,\dot{P}}(\tilde{o}_{j_1}, \tilde{o}_{j_2})$ and $d_G^{P,\dot{P}}(\tilde{o}_{j_1}, \tilde{o}_{j_2})$ over all ordered pairs $(\tilde{o}_{j_1}, \tilde{o}_{j_2})$ $(j_1 < j_2)$ results in the following overall measures of agreements ($a_G^{P,\dot{P}}$) and disagreements ($d_G^{P,\dot{P}}$) between P and $\dot{P}$:

$$a_G^{P,\dot{P}} = \sum_{j_1=1}^{n-1} \sum_{j_2=j_1+1}^{n} a_G^{P,\dot{P}}(\tilde{o}_{j_1}, \tilde{o}_{j_2}) \quad \text{and} \tag{18}$$

$$d_G^{P,\dot{P}} = \sum_{j_1=1}^{n-1} \sum_{j_2=j_1+1}^{n} d_G^{P,\dot{P}}(\tilde{o}_{j_1}, \tilde{o}_{j_2}) \tag{19}$$

The Campello soft index (CSI) [36] is defined by

$$\mathrm{CSI}(P,\dot{P}) \triangleq \frac{a_G^{P,\dot{P}}}{a_G^{P,\dot{P}} + d_G^{P,\dot{P}}}. \tag{20}$$

We hereafter assume that P, $\dot{P}$, and $\hat{P}$ are the soft clustering representations of B, $\dot{B}$, and $\hat{B}$, respectively.

Proposition 1. *There are non-equivalent biclusterings B and $\dot{B}$ such that $P \equiv \dot{P}$.*

Proof: Let B be a biclustering such that some of the matrix entries are not biclustered, and let $\dot{B}$ be the biclustering B with a new bicluster added that has only one entry from the matrix entries not biclustered in B. This new bicluster is transformed into a singleton for $\dot{P}$ by Eq. 12, whereas Eq. 13 creates an equivalent singleton for P. In other words, the bicluster added in B to produce $\dot{B}$ is superfluous from the point of view of our transformation. □

Proposition 2. *If B and $\dot{B}$ are non-equivalent non-degenerate biclusterings, then $P \not\equiv \dot{P}$.*

Proof: Let k and q be the number of biclusters in B and $\dot{B}$, respectively. Suppose that $P \triangleq \{P_i\}_{i=1}^{\overline{k}} \equiv \dot{P} \triangleq \{\dot{P}_i\}_{i=1}^{\overline{q}}$. Thus, $\overline{k} = \overline{q}$ and there is a bijection $\{(t_i, y_i)\}_{i=1}^{\overline{k}}$ such that $P_{t_i} \equiv \dot{P}_{y_i}$ for all i. Without loss of generality, suppose that $P_{t_1}, P_{t_2}, \ldots, P_{t_{\underline{k}}}$ (respectively, $\dot{P}_{y_1}, \dot{P}_{y_2}, \ldots, \dot{P}_{y_{\underline{q}}}$) are the non singletons. Clearly, $\underline{k} = \underline{q} = k = q$. The bijection $\{(t_i, y_i)\}_{i=1}^{\underline{k}}$ implies that there is a corresponding bijection between B and $\dot{B}$ making $B \equiv \dot{B}$, which contradicts the assumption that $B \not\equiv \dot{B}$. □

We will adopt the following notation in several proofs. Let C_1 and C_2 be two sets of objects from $O = \{\tilde{o}_j\}_{j=1}^n$, and let $f(\cdot,\cdot)$ be a function on $O \times O$. When proving some property of the $\mathbb{S}_{\mathrm{csi}}$ measure, $f(C_1, C_2) = x$ means $f(\tilde{o}_{j_1}, \tilde{o}_{j_2}) = x$ for all $\tilde{o}_{j_1} \in C_1$ and $\tilde{o}_{j_2} \in C_2$ such that $j_1 \neq j_2$. When proving some property of the $\mathbb{S}_{\mathrm{ebc}}$ measure, $f(C_1, C_2) = x$ means $f(\tilde{o}_{j_1}, \tilde{o}_{j_2}) = x$ for all $\tilde{o}_{j_1} \in C_1$ and $\tilde{o}_{j_2} \in C_2$. In both cases, for a function $f(\cdot)$, $f(C_1) = x$ means $f(\tilde{o}_j) = x$ for all $\tilde{o}_j \in C_1$.

Proposition 3. *The $\mathbb{S}_{\mathrm{csi}}$ measure is sensitive to the size of spurious biclusters (Def. 1).*

Proof: Let B, $\hat{B}$, and $\dot{B}$ be biclusterings, as in Def. 1, and remember that O is the set of objects. Let $\{P_{t_i}\}_{i=1}^x$ be the set of soft clusters corresponding to the spurious biclusters $\{B_{t_i}\}_{i=1}^x$ in B, and similarly define $\{\hat{P}_{t_i}\}_{i=1}^x$ for $\hat{B}$. Define $C_s^1 \triangleq \bigcup_{i=1}^x P_{t_i}$ and $C_s^2 \triangleq \bigcup_{i=1}^x \hat{P}_{t_i}$. Note that $C_s^1 \subset C_s^2$. We know that

$$\begin{aligned}
\alpha_{\dot{P}}(C_s^2, O) = \beta_{\dot{P}}(C_s^2) = 0,\\
\alpha_P(O, O - C_s^2) = \alpha_{\hat{P}}(O, O - C_s^2),\\
\beta_P(O - C_s^2) = \beta_{\hat{P}}(O - C_s^2),\\
\alpha_P(C_s^2, C_s^2) \leq \alpha_{\hat{P}}(C_s^2, C_s^2), \text{ and}\\
\beta_P(C_s^2) \leq \beta_{\hat{P}}(C_s^2).
\end{aligned}$$

Thus,

$$\begin{aligned}
a_G^{P,\dot{P}}(C_s^2, O - C_s^2) &= \min\{\beta_P(O - C_s^2), \beta_{\dot{P}}(O - C_s^2)\}\\
&= \min\{\beta_{\hat{P}}(O - C_s^2), \beta_{\dot{P}}(O - C_s^2)\}\\
&= a_G^{\hat{P},\dot{P}}(C_s^2, O - C_s^2),
\end{aligned}$$

$$a_G^{P,\dot{P}}(C_s^2, C_s^2) = 0 = a_G^{\hat{P},\dot{P}}(C_s^2, C_s^2), \text{ and}$$
$$a_G^{P,\dot{P}}(O - C_s^2, O - C_s^2) = a_G^{\hat{P},\dot{P}}(O - C_s^2, O - C_s^2).$$

Thus, $a_G^{P,\dot{P}}(O,O) = a_G^{\hat{P},\dot{P}}(O,O)$. Observe that

$$\begin{aligned}
d_G^{P,\dot{P}}(C_s^2, O - C_s^2) &= \alpha_P(C_s^2, O - C_s^2) + \beta_P(C_s^2) \\
&\quad + |\beta_P(O - C_s^2) - \beta_{\dot{P}}(O - C_s^2)| \\
&= \alpha_{\hat{P}}(C_s^2, O - C_s^2) + \beta_P(C_s^2) \\
&\quad + |\beta_{\hat{P}}(O - C_s^2) - \beta_{\dot{P}}(O - C_s^2)| \\
&\leq \alpha_{\hat{P}}(C_s^2, O - C_s^2) + \beta_{\hat{P}}(C_s^2) \\
&\quad + |\beta_{\hat{P}}(O - C_s^2) - \beta_{\dot{P}}(O - C_s^2)| \\
&= d_G^{\hat{P},\dot{P}}(C_s^2, O - C_s^2), \\
d_G^{P,\dot{P}}(C_s^2, C_s^2) &= \alpha_P(C_s^2, C_s^2) + \beta_P(C_s^2) + \beta_P(C_s^2) \\
&\leq \alpha_{\hat{P}}(C_s^2, C_s^2) + \beta_{\hat{P}}(C_s^2) + \beta_{\hat{P}}(C_s^2) \qquad (21) \\
&= d_G^{\hat{P},\dot{P}}(C_s^2, C_s^2), \text{ and} \\
d_G^{P,\dot{P}}(O - C_s^2, O - C_s^2) &= d_G^{\hat{P},\dot{P}}(O - C_s^2, O - C_s^2).
\end{aligned}$$

Thus, $d_G^{P,\dot{P}}(O,O) \leq d_G^{\hat{P},\dot{P}}(O,O)$.

Let P_t be the soft cluster corresponding to a spurious bicluster B_t that was increased, giving rise to $\hat{B}_t$ and $\hat{P}_t$. Let $\tilde{o}_{j_1}$ and $\tilde{o}_{j_2}$ be two objects from $\hat{P}_t$ such that $\tilde{o}_{j_1} \in P_t$ and $\tilde{o}_{j_2} \notin P_t$. Thus, $\alpha_{\hat{P}}(\tilde{o}_{j_1}, \tilde{o}_{j_2}) > \alpha_P(\tilde{o}_{j_1}, \tilde{o}_{j_2})$, $d_G^{P,\dot{P}}(\tilde{o}_{j_1}, \tilde{o}_{j_2}) < d_G^{\hat{P},\dot{P}}(\tilde{o}_{j_1}, \tilde{o}_{j_2})$ (Eq. (21)), and $\mathrm{CSI}(P,\dot{P}) > \mathrm{CSI}(\hat{P},\dot{P})$. □

Proposition 4. *The $\mathbb{S}_{\mathrm{ebc}}$ measure is sensitive to the size of spurious biclusters (Def. 1).*

Proof: Let B, $\hat{B}$, and $\dot{B}$ be biclusterings, as in Def. 1, and remember that O is the set of objects. Let C_s^1 and C_s^2 be the sets defined in the proof of Proposition 3. Note that

$$\begin{aligned}
\alpha_P(C_s^2, C_s^2) &\geq \alpha_{\dot{P}}(C_s^2, C_s^2), \\
\alpha_{\hat{P}}(C_s^2, C_s^2) &\geq \alpha_{\dot{P}}(C_s^2, C_s^2), \text{ and} \\
\alpha_P(O, O - C_s^2) &= \alpha_{\hat{P}}(O, O - C_s^2).
\end{aligned}$$

The nominators of Eqs. (14) are equal if one compares P with $\dot{P}$ or $\hat{P}$ with $\dot{P}$. Thus, $\mathrm{EBCR}(P,\dot{P}) = \mathrm{EBCR}(\hat{P},\dot{P})$. $\mathrm{EBCP}(P,\dot{P}) \geq \mathrm{EBCP}(\hat{P},\dot{P})$ because $\alpha_{\hat{P}}(O,O) \geq \alpha_P(O,O)$.

Let P_t be the soft cluster corresponding to a spurious bicluster B_t that was increased, giving rise to $\hat{B}_t$ and $\hat{P}_t$. Let $\tilde{o}_{j_1}$ and $\tilde{o}_{j_2}$ be two objects from $\hat{P}_t$ such that $\tilde{o}_{j_1} \in P_t$ and $\tilde{o}_{j_2} \notin P_t$. Thus, $\alpha_{\hat{P}}(\tilde{o}_{j_1}, \tilde{o}_{j_2}) > \alpha_P(\tilde{o}_{j_1}, \tilde{o}_{j_2})$, $\mathrm{EBCP}(P,\dot{P}) > \mathrm{EBCP}(\hat{P},\dot{P})$, and $\mathrm{EBC}(P,\dot{P}) > \mathrm{EBC}(\hat{P},\dot{P})$. □

Proposition 5. *The $\mathbb{S}_{\mathrm{rnia}}$ and $\mathbb{S}_{\mathrm{ce}}$ measures penalize solutions that do not cover all reference biclusters (Def. 2).*

Proof: Let B and $\dot{B}$ be as given in Def. 2. We have $\dot{\mathrm{N}}_{j_1,j_2} \geq \mathrm{N}_{j_1,j_2}$ for all j_1 and j_2, and there are j_1 and j_2 such that $\dot{\mathrm{N}}_{j_1,j_2} > \mathrm{N}_{j_1,j_2}$ (Eqs. (3) and (4)). Thus, $|U| > |I|$, and $\mathbb{S}_{\mathrm{rnia}}$ follows the property given by Def. 2. We also have $\mathbb{S}_{\mathrm{rnia}}(B,\dot{B}) \geq \mathbb{S}_{\mathrm{ce}}(B,\dot{B})$ by Proposition 1 in [30]. Thus, $\mathbb{S}_{\mathrm{ce}}$ also has the property. □

Proposition 6. *The $\mathbb{S}_{\mathrm{prec}}$, $\mathbb{S}_{\mathrm{u}}$, and $\mathbb{S}_{\mathrm{erec}}$ measures do not always penalize solutions that do not cover all reference biclusters (Def. 2).*

Proof: Let B and $\dot{B}$ be as given in Def. 2, where $B \triangleq \{B_1\}$ and $\dot{B} \triangleq \{\dot{B}_1, \dot{B}_2\}$ such that $B_1 \equiv \dot{B}_1 \equiv \dot{B}_2$. We would have $\mathbb{S}_{\mathrm{prec}}(B,\dot{B}) = \mathbb{S}_{\mathrm{u}}(B,\dot{B}) = \mathbb{S}_{\mathrm{erec}}(B,\dot{B}) = 1$, violating the condition given by Def. 2. □

Proposition 7. *The $\mathbb{S}_{\mathrm{csi}}$ measure penalizes solutions that do not cover all reference biclusters (Def. 2).*

Proof: Let B and $\dot{B}$ be as given in Def. 2. We have $\alpha_P(j_1, j_2) \leq \alpha_{\dot{P}}(j_1, j_2)$ for all $j_1 \neq j_2$, and the inequality is attained for at least one pair (j_1, j_2). Thus, $d_G^{P,\dot{P}} > 0$ and $\mathrm{CSI}(P,\dot{P}) < 1$. □

Proposition 8. *The $\mathbb{S}_{\mathrm{ebc}}$ measure penalizes solutions that do not cover all reference biclusters (Def. 2).*

Proof: Let B and $\dot{B}$ be as given in Def. 2. We have $\alpha_P(j_1, j_2) \leq \alpha_{\dot{P}}(j_1, j_2)$ for all j_1 and j_2. The inequality is attained for at least one pair (j_1, j_2), implying that $\mathrm{EBCR}(P,\dot{P}) < 1$ and $\mathrm{EBC}(P,\dot{P}) < 1$. □

Proposition 9. *The $\mathbb{S}_{\mathrm{rnia}}$ measure penalizes solutions for non-intersecting area (Def. 3).*

Proof: Let B, $\dot{B}$, and $\hat{B}$ be three biclusterings, as in Def. 3, and let S be the set of matrix elements, as in Def. 3. The matrix elements in S are those corresponding to j_1 and j_2 such that $\dot{\mathrm{N}}_{j_1,j_2} = 0$. We have $\mathrm{N}_{j_1,j_2} = \hat{\mathrm{N}}_{j_1,j_2}$ for all j_1 and j_2 such that $\dot{\mathrm{N}}_{j_1,j_2} > 0$. Thus, $\min\{\mathrm{N}_{j_1,j_2}, \dot{\mathrm{N}}_{j_1,j_2}\} = \min\{\hat{\mathrm{N}}_{j_1,j_2}, \dot{\mathrm{N}}_{j_1,j_2}\}$ for all j_1 and j_2. Since $\mathrm{N}_{j_1,j_2} \leq \hat{\mathrm{N}}_{j_1,j_2}$ for all j_1 and j_2, we have $\max\{\mathrm{N}_{j_1,j_2}, \dot{\mathrm{N}}_{j_1,j_2}\} \leq \max\{\hat{\mathrm{N}}_{j_1,j_2}, \dot{\mathrm{N}}_{j_1,j_2}\}$ for all j_1 and j_2. Therefore, $\mathbb{S}_{\mathrm{rnia}}(B,\dot{B}) \geq \mathbb{S}_{\mathrm{rnia}}(\hat{B},\dot{B})$. Since $\mathrm{N}_{j_1,j_2} < \hat{\mathrm{N}}_{j_1,j_2}$ for at least a pair (j_1, j_2) such that $\dot{\mathrm{N}}_{j_1,j_2} = 0$, we have $\max\{\mathrm{N}_{j_1,j_2}, \dot{\mathrm{N}}_{j_1,j_2}\} < \max\{\hat{\mathrm{N}}_{j_1,j_2}, \dot{\mathrm{N}}_{j_1,j_2}\}$ for such a pair and $\mathbb{S}_{\mathrm{rnia}}(B,\dot{B}) > \mathbb{S}_{\mathrm{rnia}}(\hat{B},\dot{B})$. □

Proposition 10. *The $\mathbb{S}_{\mathrm{ce}}$ measure penalizes solutions for non-intersecting area (Def. 3).*

Proof: Let B, $\dot{B}$, and $\hat{B}$ be three biclusterings, as in Def. 3. We know from the proof of Proposition 9 that $|U|$ increases from comparing B with $\dot{B}$ to comparing $\hat{B}$ with $\dot{B}$. On the other hand, $d_{\max}$ (Eq. 6) does not change from comparing B with $\dot{B}$ to comparing $\hat{B}$ with $\dot{B}$. Thus, $\mathbb{S}_{\mathrm{ce}}(B,\dot{B}) > \mathbb{S}_{\mathrm{ce}}(\hat{B},\dot{B})$. □

Proposition 11. *The $\mathbb{S}_{\mathrm{wjac}}$ and $\mathbb{S}_{\mathrm{wdic}}$ measures do not always penalize solutions for non-intersecting area (Def. 3).*

Proof: Consider a data matrix $\mathrm{A} \in \mathbb{R}^{4 \cdot 4}$. Let $B \triangleq \{B_1, B_2\}$, $B_1 \triangleq (\{2,3,4\},\{1,2\})$, $B_2 \triangleq (\{2,3,4\},\{3\})$, $\dot{B} \triangleq \{\dot{B}_1\}$, $\dot{B}_1 \triangleq (\{2,3,4\},\{3,4\})$, $\hat{B} \triangleq \{\hat{B}_1, \hat{B}_2\}$, $\hat{B}_1 \triangleq (\{2,3,4\},\{1,2\})$, and $\hat{B}_2 \triangleq (\{1,2,3,4\},\{3\})$. Note that B, $\dot{B}$, and $\hat{B}$ follow the biclustering definitions given in Def.

3. However, $\mathbb{S}_{\text{wjac}}(B,\dot{B}) = 0.167 < 0.171 = \mathbb{S}_{\text{wjac}}(\hat{B},\dot{B})$ and $\mathbb{S}_{\text{wdic}}(B,\dot{B}) = 0.22 < 0.24 = \mathbb{S}_{\text{wdic}}(\hat{B},\dot{B})$. □

Proposition 12. *Let B, $\dot{B}$, and $\hat{B}$ be three biclusterings, as in Def. 3. We have $\mathbb{S}_{\text{csi}}(B,\dot{B}) \geq \mathbb{S}_{\text{csi}}(\hat{B},\dot{B})$.*

Proof: Let S be the set of matrix elements, as in Def. 3. Define C as the set of elements from O corresponding to the matrix elements in S, and let $\overline{C} \triangleq O - C$. We have

$$\begin{aligned}
\alpha_P(\overline{C},\overline{C}) &= \alpha_{\hat{P}}(\overline{C},\overline{C}),\\
\alpha_P(C,\overline{C}) &\leq \alpha_{\hat{P}}(C,\overline{C}),\\
\alpha_P(C,C) &\leq \alpha_{\hat{P}}(C,C),\\
\alpha_{\dot{P}}(C,\overline{C}) &= \alpha_{\dot{P}}(C,C) = 0,\\
\beta_P(\overline{C}) &= \beta_{\hat{P}}(\overline{C}),\\
\beta_P(C) &\leq \beta_{\hat{P}}(C), \text{ and}\\
\beta_{\dot{P}}(C) &= 0.
\end{aligned}$$

Thus, $a_G^{P,\dot{P}}(\overline{C},\overline{C}) = a_G^{\hat{P},\dot{P}}(\overline{C},\overline{C})$ and $d_G^{P,\dot{P}}(\overline{C},\overline{C}) = d_G^{\hat{P},\dot{P}}(\overline{C},\overline{C})$. We have

$$\begin{aligned}
a_G^{P,\dot{P}}(C,\overline{C}) &= \min\{\beta_P(\overline{C}),\beta_{\dot{P}}(\overline{C})\}\\
&= \min\{\beta_{\hat{P}}(\overline{C}),\beta_{\dot{P}}(\overline{C})\}\\
&= a_G^{\hat{P},\dot{P}}(C,\overline{C})
\end{aligned}$$

and

$$\begin{aligned}
d_G^{P,\dot{P}}(C,\overline{C}) &= \alpha_P(C,\overline{C}) + \beta_P(C) + |\beta_P(\overline{C}) - \beta_{\dot{P}}(\overline{C})|\\
&= \alpha_P(C,\overline{C}) + \beta_P(C) + |\beta_{\hat{P}}(\overline{C}) - \beta_{\dot{P}}(\overline{C})|\\
&\leq \alpha_{\hat{P}}(C,\overline{C}) + \beta_{\hat{P}}(C) + |\beta_{\hat{P}}(\overline{C}) - \beta_{\dot{P}}(\overline{C})| \qquad (22)\\
&= d_G^{\hat{P},\dot{P}}(C,\overline{C}).
\end{aligned}$$

Note that

$$a_G^{P,\dot{P}}(C,C) = 0 = a_G^{\hat{P},\dot{P}}(C,C)$$

and

$$\begin{aligned}
d_G^{P,\dot{P}}(C,C) &= \alpha_P(C,C) + 2\beta_P(C)\\
&\leq \alpha_{\hat{P}}(C,C) + 2\beta_{\hat{P}}(C) \qquad (23)\\
&= d_G^{\hat{P},\dot{P}}(C,C).
\end{aligned}$$

Thus, $a_G^{P,\dot{P}} = a_G^{\hat{P},\dot{P}}$, $d_G^{P,\dot{P}} \leq d_G^{\hat{P},\dot{P}}$, $\text{CSI}(P,\dot{P}) \geq \text{CSI}(\hat{P},\dot{P})$, and $\mathbb{S}_{\text{csi}}(B,\dot{B}) \geq \mathbb{S}_{\text{csi}}(\hat{B},\dot{B})$. □

Proposition 13. *The $\mathbb{S}_{\text{csi}}$ measure penalizes solutions for non-intersecting area (Def. 3) in the domain of non-degenerate biclusterings.*

Proof: Let B, $\dot{B}$, and $\hat{B}$ be three biclusterings, as in Def. 3, with the additional restriction of being non-degenerates. Let S be the set of matrix elements, as in Def. 3. Define C as the set of elements from O corresponding to the matrix elements in S. If there are biclusters in $\hat{B}$ originated from the expansion of biclusters in B, there is an object $\tilde{o}_j \in C$ that belongs to more clusters in $\hat{P}$ than in P, meaning that $\beta_{\hat{P}}(\tilde{o}_j) > \beta_P(\tilde{o}_j)$. If there are new biclusters in $\hat{B}$, let $\hat{B}_i$ be one of these. There is a pair $\tilde{o}_{j_1},\tilde{o}_{j_2} \in C$ ($j_1 \neq j_2$) that belongs to more clusters in $\hat{P}$ than in P, meaning that $\alpha_{\hat{P}}(\tilde{o}_{j_1},\tilde{o}_{j_2}) > \alpha_P(\tilde{o}_{j_1},\tilde{o}_{j_2})$. In the first case, $d_G^{\hat{P},\dot{P}} > d_G^{P,\dot{P}}$ because of Ineq. (22). In the second case, $d_G^{\hat{P},\dot{P}} > d_G^{P,\dot{P}}$ because of Ineq. (23). Therefore, $\text{CSI}(P,\dot{P}) > \text{CSI}(\hat{P},\dot{P})$ and $\mathbb{S}_{\text{csi}}(B,\dot{B}) > \mathbb{S}_{\text{csi}}(\hat{B},\dot{B})$. □

Proposition 14. *Let B, $\dot{B}$, and $\hat{B}$ be three biclusterings, as in Def. 3. We have $\mathbb{S}_{\text{ebc}}(B,\dot{B}) \geq \mathbb{S}_{\text{ebc}}(\hat{B},\dot{B})$.*

Proof: Let S be the set of matrix elements, as in Def. 3. Define C as the set of elements from O corresponding to the matrix elements in S, and let $\overline{C} \triangleq O - C$. We have $\alpha_P(\overline{C},\overline{C}) = \alpha_{\hat{P}}(\overline{C},\overline{C})$, $\alpha_{\dot{P}}(C,\overline{C}) = 0$, $\alpha_{\dot{P}}(\tilde{o}_{j_1},\tilde{o}_{j_2}) = 0$ for all $\tilde{o}_{j_1},\tilde{o}_{j_2} \in C$ s.t. $j_1 \neq j_2$, $\alpha_P(\tilde{o}_j,\tilde{o}_j) \geq \alpha_{\dot{P}}(\tilde{o}_j,\tilde{o}_j)$ for all $\tilde{o}_j \in C$, and $\alpha_{\hat{P}}(\tilde{o}_j,\tilde{o}_j) \geq \alpha_{\dot{P}}(\tilde{o}_j,\tilde{o}_j)$ for all $\tilde{o}_j \in C$. Thus, $\min\{\alpha_P(O,O),\alpha_{\dot{P}}(O,O)\} = \min\{\alpha_{\hat{P}}(O,O),\alpha_{\dot{P}}(O,O)\}$. Since

$$\alpha_{\hat{P}}(C,\overline{C}) \geq \alpha_P(C,\overline{C}) \text{ and} \qquad (24)$$

$$\alpha_{\hat{P}}(C,C) \geq \alpha_P(C,C), \qquad (25)$$

we have $\text{EBC}(P,\dot{P}) \geq \text{EBC}(\hat{P},\dot{P})$ and $\mathbb{S}_{\text{ebc}}(B,\dot{B}) \geq \mathbb{S}_{\text{ebc}}(\hat{B},\dot{B})$. □

Proposition 15. *The $\mathbb{S}_{\text{ebc}}$ measure penalizes solutions for non-intersecting area (Def. 3) in the domain of non-degenerate biclusterings.*

Proof: Let B, $\dot{B}$, and $\hat{B}$ be three biclusterings, as in Def. 3, with the additional restriction of being non-degenerate. Let S be the set of matrix elements, as in Def. 3. Define C as the set of elements from O corresponding to the matrix elements in S, and let $\overline{C} \triangleq O - C$. If there are biclusters in $\hat{B}$ originated from the expansion of biclusters in B, there is a pair $\tilde{o}_{j_1},\tilde{o}_{j_2} \in C$ s.t. $j_1 \neq j_2$ or a pair $\tilde{o}_{j_1} \in C, \tilde{o}_{j_2} \in \overline{C}$ that belongs to more clusters in $\hat{P}$ than in P, meaning that $\alpha_{\hat{P}}(\tilde{o}_{j_1},\tilde{o}_{j_2}) > \alpha_P(\tilde{o}_{j_1},\tilde{o}_{j_2})$. If there are new biclusters in $\hat{B}$, let $\hat{B}_i$ be one of these. There is a pair $\tilde{o}_{j_1},\tilde{o}_{j_2} \in C$ s.t. $j_1 \neq j_2$ that belongs to more clusters in $\hat{P}$ than in P, meaning that $\alpha_{\hat{P}}(\tilde{o}_{j_1},\tilde{o}_{j_2}) > \alpha_P(\tilde{o}_{j_1},\tilde{o}_{j_2})$. In both cases, we can conclude from Ineqs. (24) and (25) that $\text{EBC}(P,\dot{P}) > \text{EBC}(\hat{P},\dot{P})$ and $\mathbb{S}_{\text{ebc}}(B,\dot{B}) > \mathbb{S}_{\text{ebc}}(\hat{B},\dot{B})$. □

Proposition 16. *The $\mathbb{S}_{\text{csi}}$ measure penalizes solutions for multiple biclusters coverage (Def. 4).*

Proof: Let B and $\dot{B}$ be two biclusterings, as in Def. 4. We have $\alpha_P(O,O) \geq \alpha_{\dot{P}}(O,O)$ and $\beta_P(O) = \beta_{\dot{P}}(O) = 0$. There are $\tilde{o}_{j_1}$ and $\tilde{o}_{j_2}$ s.t. $j_1 \neq j_2$ such that $\alpha_P(\tilde{o}_{j_1},\tilde{o}_{j_2}) > \alpha_{\dot{P}}(\tilde{o}_{j_1},\tilde{o}_{j_2})$, implying that $d_G^{P,\dot{P}} > 0$, $\text{CSI}(P,\dot{P}) < 1$, and $\mathbb{S}_{\text{csi}}(B,\dot{B}) < 1$. □

Proposition 17. *The $\mathbb{S}_{\text{ebc}}$ measure penalizes solutions for multiple biclusters coverage (Def. 4).*

Proof: Let B and $\dot{B}$ be two biclusterings, as in Def. 4. Note that $\alpha_P(O,O) \geq \alpha_{\dot{P}}(O,O)$. There are $\tilde{o}_{j_1}$ and $\tilde{o}_{j_2}$ s.t. $j_1 \neq j_2$ such that $\alpha_P(\tilde{o}_{j_1},\tilde{o}_{j_2}) > \alpha_{\dot{P}}(\tilde{o}_{j_1},\tilde{o}_{j_2})$, implying that $\text{EBCP}(P,\dot{P}) < 1$, $\text{EBC}(P,\dot{P}) < 1$, and $\mathbb{S}_{\text{ebc}}(B,\dot{B}) <$

1. □

Proposition 18. *The $\mathbb{S}_{\text{prel}}$, $\mathbb{S}_{\text{prec}}$, and $\mathbb{S}_{\text{l\&w}}$ measures penalize solutions for multiple biclusters coverage (Def. 4).*

Proof: Let B and $\dot{B}$ be two biclusterings, as in Def. 4. Since the biclusters in $\dot{B}$ do not overlap, we have the proper subset relationships $\dot{B}_i^{\text{r}} \subset B_1^{\text{r}}$ for all i or $\dot{B}_i^{\text{c}} \subset B_1^{\text{c}}$ for all i. Thus, $|B_1^{\text{r}} \cup \dot{B}_i^{\text{r}}| > |B_1^{\text{r}} \cap \dot{B}_i^{\text{r}}|$ for all i or $|B_1^{\text{c}} \cup \dot{B}_i^{\text{c}}| > |B_1^{\text{c}} \cap \dot{B}_i^{\text{c}}|$ for all i. We have $S_{\text{r}}(B, \dot{B}) < 1$ or $S_{\text{c}}(B, \dot{B}) < 1$ (Eqs. (1) and (2)), implying that $\mathbb{S}_{\text{prel}}(B, \dot{B}) < 1$. Similarly, $\mathbb{S}_{\text{prec}}(B, \dot{B}) < 1$ and $\mathbb{S}_{\text{l\&w}}(B, \dot{B}) < 1$. □

Proposition 19. *The $\mathbb{S}_{\text{stm}}$, $\mathbb{S}_{\text{wjac}}$, $\mathbb{S}_{\text{wdic}}$, $\mathbb{S}_{\text{ay}}$, $\mathbb{S}_{\text{erel}}$, and $\mathbb{S}_{\text{erec}}$ measures penalize solutions for multiple biclusters coverage (Def. 4).*

Proof: Let B and $\dot{B}$ be two biclusterings, as in Def. 4. Since the biclusters in $\dot{B}$ do not overlap, we have the proper subset relationship $\dot{B}_i^{\text{r}} \times \dot{B}_i^{\text{c}} \subset B_1^{\text{r}} \times B_1^{\text{c}}$ for all i. Thus, $\mathbb{D}(B_1, \dot{B}_i) < 1$ and $\mathbb{J}(B_1, \dot{B}_i) < 1$ for all i, implying that $\mathbb{S}_{\text{stm}}(B, \dot{B}) < 1$, $\mathbb{S}_{\text{wjac}}(B, \dot{B}) < 1$, $\mathbb{S}_{\text{wdic}}(B, \dot{B}) < 1$, $\mathbb{S}_{\text{erel}}(B, \dot{B}) < 1$, and $\mathbb{S}_{\text{erec}}(B, \dot{B}) < 1$. From $\mathbb{S}_{\text{ay}}(B, \dot{B}) < \mathbb{S}_{\text{erel}}(B, \dot{B})$, we also have $\mathbb{S}_{\text{ay}}(B, \dot{B}) < 1$. □

Proposition 20. *The $\mathbb{S}_{\text{csi}}$ measure penalizes solutions with repetitive biclusters (Def. 5).*

Proof: Let B, $\dot{B}$, and $\hat{B}$ be biclusterings, as in Def. 5. Let C_{r} be the set of objects from O corresponding to the matrix entries of the biclusters in B that were replicated. We have

$$\begin{aligned} \alpha_P(O, O - C_{\text{r}}) &= \alpha_{\hat{P}}(O, O - C_{\text{r}}), && (26)\\ \alpha_{\hat{P}}(C_{\text{r}}, C_{\text{r}}) &> \alpha_P(C_{\text{r}}, C_{\text{r}}) \geq \alpha_{\dot{P}}(C_{\text{r}}, C_{\text{r}}), && (27)\\ \beta_P(O - C_{\text{r}}) &= \beta_{\hat{P}}(O - C_{\text{r}}), \text{ and} \\ \beta_{\hat{P}}(C_{\text{r}}) &> \beta_P(C_{\text{r}}) \geq \beta_{\dot{P}}(C_{\text{r}}). \end{aligned}$$

We conclude that $a_G^{P,\dot{P}}(O, O) = a_G^{\hat{P},\dot{P}}(O, O)$ and $a_G^{P,\dot{P}} = a_G^{\hat{P},\dot{P}}$.

We have

$$\begin{aligned} d_G^{P,\dot{P}}(O, O - C_{\text{r}}) &= |\alpha_P(O, O - C_{\text{r}}) - \alpha_{\dot{P}}(O, O - C_{\text{r}})| \\ &\quad + |\beta_P(O) - \beta_{\dot{P}}(O)| \\ &\quad + |\beta_P(O - C_{\text{r}}) - \beta_{\dot{P}}(O - C_{\text{r}})| \\ &= |\alpha_{\hat{P}}(O, O - C_{\text{r}}) - \alpha_{\dot{P}}(O, O - C_{\text{r}})| \\ &\quad + |\beta_P(O) - \beta_{\dot{P}}(O)| \\ &\quad + |\beta_{\hat{P}}(O - C_{\text{r}}) - \beta_{\dot{P}}(O - C_{\text{r}})| \\ &\leq d_G^{\hat{P},\dot{P}}(O, O - C_{\text{r}}) \end{aligned}$$

and

$$\begin{aligned} d_G^{P,\dot{P}}(C_{\text{r}}, C_{\text{r}}) &= |\alpha_P(C_{\text{r}}, C_{\text{r}}) - \alpha_{\dot{P}}(C_{\text{r}}, C_{\text{r}})| \\ &\quad + 2|\beta_P(C_{\text{r}}) - \beta_{\dot{P}}(C_{\text{r}})| \\ &< |\alpha_{\hat{P}}(C_{\text{r}}, C_{\text{r}}) - \alpha_{\dot{P}}(C_{\text{r}}, C_{\text{r}})| \\ &\quad + 2|\beta_{\hat{P}}(C_{\text{r}}) - \beta_{\dot{P}}(C_{\text{r}})| \\ &= d_G^{\hat{P},\dot{P}}(C_{\text{r}}, C_{\text{r}}) \end{aligned}$$

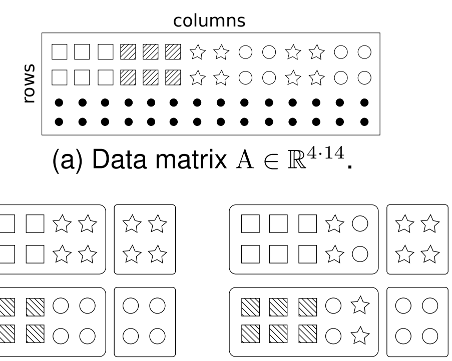
Thus, $d_G^{\hat{P},\dot{P}} > d_G^{P,\dot{P}}$, $\text{CSI}(P, \dot{P}) > \text{CSI}(\hat{P}, \dot{P})$, and $\mathbb{S}_{\text{csi}}(B, \dot{B}) > \mathbb{S}_{\text{csi}}(\hat{B}, \dot{B})$. □

Proposition 21. *The $\mathbb{S}_{\text{ebc}}$ measure penalizes solutions with repetitive biclusters (Def. 5).*

Proof: Let B, $\dot{B}$, and $\hat{B}$ be biclusterings, as in Def. 5. Let C_{r} be the set of objects from O corresponding to the matrix entries of the biclusters in B that were replicated. We know from Eqs. (26) and (27) that the nominators of Eqs. (14) do not change from comparing P with $\dot{P}$ to comparing $\hat{P}$ with $\dot{P}$, and that $\alpha_{\hat{P}}(O, O) \geq \alpha_P(O, O)$. We have $\tilde{o}_{j_1}$ and $\tilde{o}_{j_2}$ (for which $j_1 = j_2$ is allowed) such that $\alpha_{\hat{P}}(\tilde{o}_{j_1}, \tilde{o}_{j_2}) > \alpha_P(\tilde{o}_{j_1}, \tilde{o}_{j_2})$. Therefore, $\text{EBCP}(P, \dot{P}) > \text{EBCP}(\hat{P}, \dot{P})$, $\text{EBC}(P, \dot{P}) > \text{EBC}(\hat{P}, \dot{P})$, and $\mathbb{S}_{\text{ebc}}(B, \dot{B}) > \mathbb{S}_{\text{ebc}}(\hat{B}, \dot{B})$. □

Proposition 22. *The $\mathbb{S}_{\text{prec}}$ and $\mathbb{S}_{\text{erec}}$ measures do not have the homogeneity property (Def. 7).*

Proof: Consider the non-overlapping biclusterings B and $\hat{B}$ represented by Figs. 13b and 13c and the reference solution $\dot{B}$ represented by Fig. 13a. We have $\mathbb{S}_{\text{prec}}(B, \dot{B}) = \mathbb{S}_{\text{prec}}(\hat{B}, \dot{B}) = 0.74$ and $\mathbb{S}_{\text{erec}}(B, \dot{B}) = \mathbb{S}_{\text{erec}}(\hat{B}, \dot{B}) = 0.55$.



(a) Data matrix $\mathrm{A} \in \mathbb{R}^{4 \cdot 14}$.

(b) Biclustering B. (c) Biclustering $\hat{B}$.

Fig. 13: Difference in homogeneity.

□

Proposition 23. *The $\mathbb{S}_{\text{csi}}$ measure has the homogeneity property (Def. 7).*

Proof: Let B, $\hat{B}$, and $\dot{B}$ be three biclusterings, as in Def. 7. Let C_{s}^1 be the x objects swapped from P_{i_1}, C_{n}^1 be the objects from the minor category in P_{i_1} that were not swapped, and C_{r}^1 be the objects from the main category in P_{i_1}. Analogously, define C_{s}^2, C_{n}^2, and C_{r}^2. Since B, $\hat{B}$, and $\dot{B}$ are non-overlapping solutions, we have $\beta_P(O) = \beta_{\hat{P}}(O) = \beta_{\dot{P}}(O) = 0$. Thus, the eventual difference between $a_G^{P,\dot{P}}(\tilde{o}_{j_1}, \tilde{o}_{j_2})$ (respectively, $d_G^{P,\dot{P}}(\tilde{o}_{j_1}, \tilde{o}_{j_2})$) and $a_G^{\hat{P},\dot{P}}(\tilde{o}_{j_1}, \tilde{o}_{j_2})$ ($d_G^{\hat{P},\dot{P}}(\tilde{o}_{j_1}, \tilde{o}_{j_2})$) can only be due to the eventual difference between $\alpha_P(\tilde{o}_{j_1}, \tilde{o}_{j_2})$ and $\alpha_{\hat{P}}(\tilde{o}_{j_1}, \tilde{o}_{j_2})$ (see Eqs. (16) and (17)).

We conclude from Table 12 that $a_G^{P,\dot{P}} = a_G^{\hat{P},\dot{P}} + |C_{\text{s}}^1||C_{\text{n}}^1| + |C_{\text{s}}^2||C_{\text{n}}^2| = a_G^{\hat{P},\dot{P}} + x(|C_{\text{n}}^1| + |C_{\text{n}}^2|)$, where $x = |C_{\text{s}}^1| = |C_{\text{s}}^2|$, as in Def. 7. Table 13 shows the differences in

TABLE 12: Differences in $a_G^{\cdot,\cdot}(\cdot,\cdot)$ by changing P to $\hat{P}$.

$a_G^{P,\dot{P}}(C_\mathrm{s}^1, C_\mathrm{n}^1) = \min\{1,1\} > \min\{0,1\} = a_G^{\hat{P},\dot{P}}(C_\mathrm{s}^1, C_\mathrm{n}^1)$
$a_G^{P,\dot{P}}(C_\mathrm{s}^2, C_\mathrm{n}^2) = \min\{1,1\} > \min\{0,1\} = a_G^{\hat{P},\dot{P}}(C_\mathrm{s}^2, C_\mathrm{n}^2)$
$a_G^{P,\dot{P}}(C_\mathrm{s}^1, C_\mathrm{r}^1) = \min\{1,0\} = \min\{0,0\} = a_G^{\hat{P},\dot{P}}(C_\mathrm{s}^1, C_\mathrm{r}^1)$
$a_G^{P,\dot{P}}(C_\mathrm{s}^2, C_\mathrm{r}^2) = \min\{1,0\} = \min\{0,0\} = a_G^{\hat{P},\dot{P}}(C_\mathrm{s}^2, C_\mathrm{r}^2)$
$a_G^{P,\dot{P}}(C_\mathrm{s}^1, C_\mathrm{n}^2) = \min\{0,0\} = \min\{1,0\} = a_G^{\hat{P},\dot{P}}(C_\mathrm{s}^1, C_\mathrm{n}^2)$
$a_G^{P,\dot{P}}(C_\mathrm{n}^1, C_\mathrm{s}^2) = \min\{0,0\} = \min\{1,0\} = a_G^{\hat{P},\dot{P}}(C_\mathrm{n}^1, C_\mathrm{s}^2)$
$a_G^{P,\dot{P}}(C_\mathrm{s}^1, C_\mathrm{r}^2) = \min\{0,0\} = \min\{1,0\} = a_G^{\hat{P},\dot{P}}(C_\mathrm{s}^1, C_\mathrm{r}^2)$
$a_G^{P,\dot{P}}(C_\mathrm{r}^1, C_\mathrm{s}^2) = \min\{0,0\} = \min\{1,0\} = a_G^{\hat{P},\dot{P}}(C_\mathrm{r}^1, C_\mathrm{s}^2)$

$d_G^{\cdot,\cdot}(\cdot,\cdot)$, implying that $d_G^{P,\dot{P}} = d_G^{\hat{P},\dot{P}} - |C_\mathrm{s}^1||C_\mathrm{n}^1| - |C_\mathrm{s}^2||C_\mathrm{n}^2| - |C_\mathrm{s}^1||C_\mathrm{n}^2| - |C_\mathrm{n}^1||C_\mathrm{s}^2| - |C_\mathrm{s}^1||C_\mathrm{r}^2| - |C_\mathrm{r}^1||C_\mathrm{s}^2| + |C_\mathrm{s}^1||C_\mathrm{r}^1| + |C_\mathrm{s}^2||C_\mathrm{r}^2| = d_G^{\hat{P},\dot{P}} - 2x(|C_\mathrm{n}^1| + |C_\mathrm{n}^2|)$. Thus,

$$\mathrm{CSI}(\hat{P}, \dot{P}) = \frac{a_G^{P,\dot{P}} - x(|C_\mathrm{n}^1| + |C_\mathrm{n}^2|)}{a_G^{P,\dot{P}} + d_G^{P,\dot{P}} + x(|C_\mathrm{n}^1| + |C_\mathrm{n}^2|)}$$

and $\mathbb{S}_\mathrm{csi}(B, \dot{B}) \geq \mathbb{S}_\mathrm{csi}(\hat{B}, \dot{B})$.

TABLE 13: Differences in $d_G^{\cdot,\cdot}(\cdot,\cdot)$ by changing P to $\hat{P}$.

$d_G^{P,\dot{P}}(C_\mathrm{s}^1, C_\mathrm{n}^1) = \lvert 1-1\rvert < \lvert 0-1\rvert = d_G^{\hat{P},\dot{P}}(C_\mathrm{s}^1, C_\mathrm{n}^1)$
$d_G^{P,\dot{P}}(C_\mathrm{s}^2, C_\mathrm{n}^2) = \lvert 1-1\rvert < \lvert 0-1\rvert = d_G^{\hat{P},\dot{P}}(C_\mathrm{s}^2, C_\mathrm{n}^2)$
$d_G^{P,\dot{P}}(C_\mathrm{s}^1, C_\mathrm{r}^1) = \lvert 1-0\rvert > \lvert 0-0\rvert = d_G^{\hat{P},\dot{P}}(C_\mathrm{s}^1, C_\mathrm{r}^1)$
$d_G^{P,\dot{P}}(C_\mathrm{s}^2, C_\mathrm{r}^2) = \lvert 1-0\rvert > \lvert 0-0\rvert = d_G^{\hat{P},\dot{P}}(C_\mathrm{s}^2, C_\mathrm{r}^2)$
$d_G^{P,\dot{P}}(C_\mathrm{s}^1, C_\mathrm{n}^2) = \lvert 0-0\rvert < \lvert 1-0\rvert = d_G^{\hat{P},\dot{P}}(C_\mathrm{s}^1, C_\mathrm{n}^2)$
$d_G^{P,\dot{P}}(C_\mathrm{n}^1, C_\mathrm{s}^2) = \lvert 0-0\rvert < \lvert 1-0\rvert = d_G^{\hat{P},\dot{P}}(C_\mathrm{n}^1, C_\mathrm{s}^2)$
$d_G^{P,\dot{P}}(C_\mathrm{s}^1, C_\mathrm{r}^2) = \lvert 0-0\rvert < \lvert 1-0\rvert = d_G^{\hat{P},\dot{P}}(C_\mathrm{s}^1, C_\mathrm{r}^2)$
$d_G^{P,\dot{P}}(C_\mathrm{r}^1, C_\mathrm{s}^2) = \lvert 0-0\rvert < \lvert 1-0\rvert = d_G^{\hat{P},\dot{P}}(C_\mathrm{r}^1, C_\mathrm{s}^2)$

If $x = |\mathrm{I}(B_{i_1}, \dot{B}_{\mathrm{mi}(i_1)})| = |\mathrm{I}(B_{i_2}, \dot{B}_{\mathrm{mi}(i_2)})|$, then $|C_\mathrm{n}^1| = |C_\mathrm{n}^2| = 0$ and $\mathbb{S}_\mathrm{csi}(B, \dot{B}) = \mathbb{S}_\mathrm{csi}(\hat{B}, \dot{B})$. If $\mathbb{S}_\mathrm{csi}(B, \dot{B}) = \mathbb{S}_\mathrm{csi}(\hat{B}, \dot{B})$, then $|C_\mathrm{n}^1| + |C_\mathrm{n}^2| = 0$ (because $x > 0$, Def. 7) and $x = |\mathrm{I}(B_{i_1}, \dot{B}_{\mathrm{mi}(i_1)})| = |\mathrm{I}(B_{i_2}, \dot{B}_{\mathrm{mi}(i_2)})|$.

□

Proposition 24. *The $\mathbb{S}_\mathrm{ebc}$ measure has the homogeneity property (Def. 7).*

Proof: Let B, $\hat{B}$, and $\dot{B}$ be three biclusterings, as in Def. 7. Let C_s^1 be the x objects swapped from P_{i_1}, C_n^1 be the objects from the minor category in P_{i_1} that were not swapped, and C_r^1 be the objects from the main category in P_{i_1}. Analogously, define C_s^2, C_n^2, and C_r^2. Let

$$\delta_\mathrm{p}^{P,\dot{P}}(\tilde{o}_{j_1}) \triangleq \frac{\sum_{j_2=1}^n \min\{\alpha_P(\tilde{o}_{j_1}, \tilde{o}_{j_2}), \alpha_{\dot{P}}(\tilde{o}_{j_1}, \tilde{o}_{j_2})\}}{\sum_{j_2=1}^n \alpha_P(\tilde{o}_{j_1}, \tilde{o}_{j_2})}$$

such that $\mathrm{EBCP}(P, \dot{P}) = (1/n) \sum_{j_1=1}^n \delta_\mathrm{p}^{P,\dot{P}}(\tilde{o}_{j_1})$.

Note that

$$\sum_{\tilde{o}_{j_1} \notin P_{i_1} \cup P_{i_2}} \delta_\mathrm{p}^{P,\dot{P}}(\tilde{o}_{j_1}) = \sum_{\tilde{o}_{j_1} \notin P_{i_1} \cup P_{i_2}} \delta_\mathrm{p}^{\hat{P},\dot{P}}(\tilde{o}_{j_1}).$$

TABLE 14: Differences in $\delta_\mathrm{p}^{\cdot,\cdot}(\cdot)$ by changing P to $\hat{P}$.

$\delta_\mathrm{p}^{P,\dot{P}}(C_\mathrm{s}^1) = \frac{\lvert C_\mathrm{s}^1 \cup C_\mathrm{n}^1\rvert}{\lvert P_{i_1}\rvert}$	$\delta_\mathrm{p}^{\hat{P},\dot{P}}(C_\mathrm{s}^1) = \frac{\lvert C_\mathrm{s}^1\rvert}{\lvert P_{i_2}\rvert}$
$\delta_\mathrm{p}^{P,\dot{P}}(C_\mathrm{s}^2) = \frac{\lvert C_\mathrm{s}^2 \cup C_\mathrm{n}^2\rvert}{\lvert P_{i_2}\rvert}$	$\delta_\mathrm{p}^{\hat{P},\dot{P}}(C_\mathrm{s}^2) = \frac{\lvert C_\mathrm{s}^2\rvert}{\lvert P_{i_1}\rvert}$
$\delta_\mathrm{p}^{P,\dot{P}}(C_\mathrm{n}^1) = \frac{\lvert C_\mathrm{s}^1 \cup C_\mathrm{n}^1\rvert}{\lvert P_{i_1}\rvert}$	$\delta_\mathrm{p}^{\hat{P},\dot{P}}(C_\mathrm{n}^1) = \frac{\lvert C_\mathrm{n}^1\rvert}{\lvert P_{i_1}\rvert}$
$\delta_\mathrm{p}^{P,\dot{P}}(C_\mathrm{n}^2) = \frac{\lvert C_\mathrm{s}^2 \cup C_\mathrm{n}^2\rvert}{\lvert P_{i_2}\rvert}$	$\delta_\mathrm{p}^{\hat{P},\dot{P}}(C_\mathrm{n}^2) = \frac{\lvert C_\mathrm{n}^2\rvert}{\lvert P_{i_2}\rvert}$
$\delta_\mathrm{p}^{P,\dot{P}}(C_\mathrm{r}^1) = \frac{\lvert C_\mathrm{r}^1\rvert}{\lvert P_{i_1}\rvert}$	$\delta_\mathrm{p}^{\hat{P},\dot{P}}(C_\mathrm{r}^1) = \frac{\lvert C_\mathrm{r}^1\rvert}{\lvert P_{i_1}\rvert}$
$\delta_\mathrm{p}^{P,\dot{P}}(C_\mathrm{r}^2) = \frac{\lvert C_\mathrm{r}^2\rvert}{\lvert P_{i_2}\rvert}$	$\delta_\mathrm{p}^{\hat{P},\dot{P}}(C_\mathrm{r}^2) = \frac{\lvert C_\mathrm{r}^2\rvert}{\lvert P_{i_2}\rvert}$

We conclude from Table 14 and after some algebraic manipulation that

$$\sum_{\tilde{o}_{j_1} \in P_{i_1} \cup P_{i_2}} \delta_\mathrm{p}^{P,\dot{P}}(\tilde{o}_{j_1}) = \frac{(|C_\mathrm{s}^1| + |C_\mathrm{n}^1|)^2}{|P_{i_1}|} + \frac{|C_\mathrm{r}^1|^2}{|P_{i_1}|} + \frac{(|C_\mathrm{s}^2| + |C_\mathrm{n}^2|)^2}{|P_{i_2}|} + \frac{|C_\mathrm{r}^2|^2}{|P_{i_2}|}$$

and

$$\sum_{\tilde{o}_{j_1} \in P_{i_1} \cup P_{i_2}} \delta_\mathrm{p}^{\hat{P},\dot{P}}(\tilde{o}_{j_1}) = \frac{|C_\mathrm{s}^2|^2 + |C_\mathrm{n}^1|^2 + |C_\mathrm{r}^1|^2}{|P_{i_1}|} + \frac{|C_\mathrm{s}^1|^2 + |C_\mathrm{n}^2|^2 + |C_\mathrm{r}^2|^2}{|P_{i_2}|}.$$

From $x = |C_\mathrm{s}^1| = |C_\mathrm{s}^2|$, we have

$$\mathrm{EBCP}(\hat{P}, \dot{P}) - \mathrm{EBCP}(P, \dot{P}) = \frac{-2x}{n}\Big(\frac{|C_\mathrm{n}^1|}{|P_{i_1}|} + \frac{|C_\mathrm{n}^2|}{|P_{i_2}|}\Big). \quad (28)$$

Let

$$\delta_\mathrm{r}^{P,\dot{P}}(\tilde{o}_{j_1}) \triangleq \frac{\sum_{j_2=1}^n \min\{\alpha_P(\tilde{o}_{j_1}, \tilde{o}_{j_2}), \alpha_{\dot{P}}(\tilde{o}_{j_1}, \tilde{o}_{j_2})\}}{\sum_{j_2=1}^n \alpha_{\dot{P}}(\tilde{o}_{j_1}, \tilde{o}_{j_2})}$$

such that $\mathrm{EBCR}(P, \dot{P}) = (1/n) \sum_{j_1=1}^n \delta_\mathrm{r}^{P,\dot{P}}(\tilde{o}_{j_1})$. The only difference between $\delta_\mathrm{p}^{P,\dot{P}}(\tilde{o}_{j_1})$ and $\delta_\mathrm{r}^{P,\dot{P}}(\tilde{o}_{j_1})$ lies in the denominators. Thus,

$$\sum_{\tilde{o}_{j_1} \in P_{i_1} \cup P_{i_2}} \delta_\mathrm{r}^{P,\dot{P}}(\tilde{o}_{j_1}) = \frac{(|C_\mathrm{s}^1| + |C_\mathrm{n}^1|)^2}{|\dot{P}_{\mathrm{mi}(i_1)}|} + \frac{(|C_\mathrm{s}^2| + |C_\mathrm{n}^2|)^2}{|\dot{P}_{\mathrm{mi}(i_2)}|} + \frac{|C_\mathrm{r}^1|^2}{|\dot{P}_{\mathrm{ma}(i_1)}|} + \frac{|C_\mathrm{r}^2|^2}{|\dot{P}_{\mathrm{ma}(i_2)}|}$$

and

$$\sum_{\tilde{o}_{j_1} \in P_{i_1} \cup P_{i_2}} \delta_\mathrm{r}^{\hat{P},\dot{P}}(\tilde{o}_{j_1}) = \frac{|C_\mathrm{s}^1|^2 + |C_\mathrm{n}^1|^2}{|\dot{P}_{\mathrm{mi}(i_1)}|} + \frac{|C_\mathrm{s}^2|^2 + |C_\mathrm{n}^2|^2}{|\dot{P}_{\mathrm{mi}(i_2)}|} + \frac{|C_\mathrm{r}^1|^2}{|\dot{P}_{\mathrm{ma}(i_1)}|} + \frac{|C_\mathrm{r}^2|^2}{|\dot{P}_{\mathrm{ma}(i_2)}|}.$$

Therefore,

$$\mathrm{EBCR}(\hat{P}, \dot{P}) - \mathrm{EBCR}(P, \dot{P}) = \frac{-2x}{n}\Big(\frac{|C_{\mathrm{n}}^1|}{|\dot{P}_{\mathrm{mi}(i_1)}|} + \frac{|C_{\mathrm{n}}^2|}{|\dot{P}_{\mathrm{mi}(i_2)}|}\Big). \tag{29}$$

We conclude from (28) and (29) that $\mathrm{EBC}(P, \dot{P}) \geq \mathrm{EBC}(\hat{P}, \dot{P})$ and $\mathbb{S}_{\mathrm{ebc}}(B, \dot{B}) \geq \mathbb{S}_{\mathrm{ebc}}(\hat{B}, \dot{B})$. If $x = |\mathrm{I}(B_{i_1}, \dot{B}_{\mathrm{mi}(i_1)})| = |\mathrm{I}(B_{i_2}, \dot{B}_{\mathrm{mi}(i_2)})|$, then $|C_{\mathrm{n}}^1| = |C_{\mathrm{n}}^2| = 0$ and $\mathbb{S}_{\mathrm{ebc}}(B, \dot{B}) = \mathbb{S}_{\mathrm{ebc}}(\hat{B}, \dot{B})$. If $\mathbb{S}_{\mathrm{ebc}}(B, \dot{B}) = \mathbb{S}_{\mathrm{ebc}}(\hat{B}, \dot{B})$, then $|C_{\mathrm{n}}^1| = |C_{\mathrm{n}}^2| = 0$ (because $x > 0$, Def. 7) and $x = |\mathrm{I}(B_{i_1}, \dot{B}_{\mathrm{mi}(i_1)})| = |\mathrm{I}(B_{i_2}, \dot{B}_{\mathrm{mi}(i_2)})|$.

□

Proposition 25. *We have* $\mathbb{S}_{\mathrm{ce}}(B, \dot{B}) = 1$ *iff* $B \equiv \dot{B}$.

Proof: Clearly, $B \equiv \dot{B}$ implies that $\mathbb{S}_{\mathrm{ce}} = 1$. Let k be the number of biclusters in B and q the number of biclusters in $\dot{B}$. Suppose that $\mathbb{S}_{\mathrm{ce}}(B, \dot{B}) = 1$, and let $\{(t_i, y_i)\}_{i=1}^{\min\{k,q\}}$ be the unique relation that maximizes Eq. (6). Note that $|U| \geq \sum_{j_1,j_2} \mathrm{N}_{j_1,j_2} = \sum_{i=1}^{k} |B_i^{\mathrm{r}} \times B_i^{\mathrm{c}}| \geq d_{\max}$ and $|U| \geq \sum_{j_1,j_2} \dot{\mathrm{N}}_{j_1,j_2} = \sum_{i=1}^{q} |\dot{B}_i^{\mathrm{r}} \times \dot{B}_i^{\mathrm{c}}| \geq d_{\max}$. We have from $|U| = d_{\max}$ that $\sum_{i=1}^{k} |B_i^{\mathrm{r}} \times B_i^{\mathrm{c}}| = d_{\max} \leq \sum_{i=1}^{\min\{k,q\}} |B_{t_i}^{\mathrm{r}} \times B_{t_i}^{\mathrm{c}}| \leq \sum_{i=1}^{k} |B_i^{\mathrm{r}} \times B_i^{\mathrm{c}}|$, implying that $\sum_{i=1}^{k} |B_i^{\mathrm{r}} \times B_i^{\mathrm{c}}| = \sum_{i=1}^{\min\{k,q\}} |B_{t_i}^{\mathrm{r}} \times B_{t_i}^{\mathrm{c}}|$ and $k = \min\{k, q\}$. Similarly, $\sum_{i=1}^{q} |\dot{B}_i^{\mathrm{r}} \times \dot{B}_i^{\mathrm{c}}| = \sum_{i=1}^{\min\{k,q\}} |\dot{B}_{y_i}^{\mathrm{r}} \times \dot{B}_{y_i}^{\mathrm{c}}|$ and $q = \min\{k, q\}$. Thus, $k = q$. From $d_{\max} = \sum_{i=1}^{\min\{k,q\}} |B_{t_i}^{\mathrm{r}} \times B_{t_i}^{\mathrm{c}}|$ we have $B_{t_i}^{\mathrm{r}} \times B_{t_i}^{\mathrm{c}} \subseteq \dot{B}_{y_i}^{\mathrm{r}} \times \dot{B}_{y_i}^{\mathrm{c}}$ for all i. Similarly, $\dot{B}_{y_i}^{\mathrm{r}} \times \dot{B}_{y_i}^{\mathrm{c}} \subseteq B_{t_i}^{\mathrm{r}} \times B_{t_i}^{\mathrm{c}}$ for all i, and then $B_{t_i}^{\mathrm{r}} \times B_{t_i}^{\mathrm{c}} = \dot{B}_{y_i}^{\mathrm{r}} \times \dot{B}_{y_i}^{\mathrm{c}}$ for all i. Thus, $B \equiv \dot{B}$. □

Proposition 26. *We have* $\mathbb{S}_{\mathrm{fabi}}(B, \dot{B}) = 1$ *iff* $B \equiv \dot{B}$.

Proof: Clearly, $B \equiv \dot{B}$ implies that $\mathbb{S}_{\mathrm{fabi}} = 1$. Suppose that $\mathbb{S}_{\mathrm{fabi}}(B, \dot{B}) = 1$, and let $\{(t_i, y_i)\}_{i=1}^{\min\{k,q\}}$ be the optimal unique relation required by $\mathbb{S}_{\mathrm{fabi}}$. The denominator of Eq. (9) shows that $\mathbb{S}_{\mathrm{fabi}}$ attains 1 only when comparing solutions with the same number of biclusters, which makes $\{(t_i, y_i)\}_{i=1}^{\min\{k,q\}}$ a bijection. $\mathbb{S}_{\mathrm{fabi}}$ attains 1 only if $\mathbb{J}(B_{t_i}, \dot{B}_{y_i}) = 1$ for all i, and $\mathbb{J}(B_{t_i}, \dot{B}_{y_i})$ attains 1 only if $B_{t_i} \equiv \dot{B}_{y_i}$. Thus, $B \equiv \dot{B}$. □

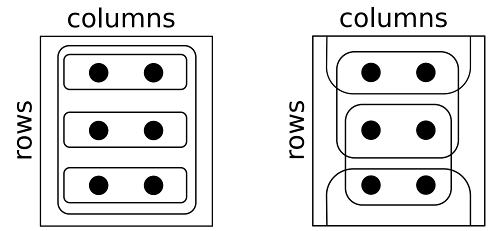
Proposition 27. *There are two non-equivalent biclusterings* B *and* $\dot{B}$ *such that* $\mathbb{S}_{\mathrm{csi}}(B, \dot{B}) = \mathbb{S}_{\mathrm{ebc}}(B, \dot{B}) = 1$.

Proof: Let B be a biclustering in which one or more matrix elements were not biclustered. Then add one bicluster in B that has only one of the matrix elements that were not biclustered, and let $\dot{B}$ be the resulting solution. Applying the transformation approach given in Section 5 to B and $\dot{B}$, we have two equivalent soft clusterings P and $\dot{P}$, implying that $\mathrm{CSI}(P, \dot{P}) = \mathrm{EBC}(P, \dot{P}) = 1$ and $\mathbb{S}_{\mathrm{csi}}(B, \dot{B}) = \mathbb{S}_{\mathrm{ebc}}(B, \dot{B}) = 1$. □

Proposition 28. *There are two non-equivalent non-degenerate biclusterings* B *and* $\dot{B}$ *for which* $\mathbb{S}_{\mathrm{csi}}(B, \dot{B}) = \mathbb{S}_{\mathrm{ebc}}(B, \dot{B}) = 1$.

Proof: The biclusterings in Fig. 14 are given by $B \triangleq \{B_i\}_{i=1}^4$, $\dot{B} \triangleq \{\dot{B}_i\}_{i=1}^3$, $B_1 \triangleq (\{1\}, \{1, 2\})$, $B_2 \triangleq (\{2\}, \{1, 2\})$, $B_3 \triangleq (\{3\}, \{1, 2\})$, $B_4 \triangleq (\{1, 2, 3\}, \{1, 2\})$, $\dot{B}_1 \triangleq (\{1, 2\}, \{1, 2\})$, $\dot{B}_2 \triangleq (\{2, 3\}, \{1, 2\})$, and $\dot{B}_3 \triangleq (\{1, 3\}, \{1, 2\})$. We have $P = \{P_i\}_{i=1}^4$ and $\dot{P} = \{\dot{P}_i\}_{i=1}^3$, where $P_1 = \{\tilde{o}_1, \tilde{o}_4\}$, $P_2 = \{\tilde{o}_2, \tilde{o}_5\}$, $P_3 = \{\tilde{o}_3, \tilde{o}_6\}$, $P_4 = \{\tilde{o}_j\}_{j=1}^6$, $\dot{P}_1 = \{\tilde{o}_1, \tilde{o}_2, \tilde{o}_4, \tilde{o}_5\}$, $\dot{P}_2 = \{\tilde{o}_2, \tilde{o}_3, \tilde{o}_5, \tilde{o}_6\}$, and $\dot{P}_3 = \{\tilde{o}_1, \tilde{o}_3, \tilde{o}_4, \tilde{o}_6\}$.

Note that $\alpha_P(\tilde{o}_{j_1}, \tilde{o}_{j_2}) = \alpha_{\dot{P}}(\tilde{o}_{j_1}, \tilde{o}_{j_2})$ for all j_1 and j_2, and we have $\beta_P(\tilde{o}_j) = \beta_{\dot{P}}(\tilde{o}_j)$ for all j. Thus, $d_G^{P,\dot{P}} = 0$, $\mathrm{CSI}(P, \dot{P}) = 1$, and $\mathbb{S}_{\mathrm{csi}}(B, \dot{B}) = 1$. Similarly, $\mathbb{S}_{\mathrm{ebc}}(B, \dot{B}) = 1$.



(a) Biclustering B. (b) Biclustering $\dot{B}$.

Fig. 14: Indistinguishable biclusterings for $\mathbb{S}_{\mathrm{csi}}$ and $\mathbb{S}_{\mathrm{ebc}}$.

□

Since $P \not\equiv \dot{P}$, the above proof contradicts Proposition 1 in [36], which states (in terms of matrices) that $\mathrm{CSI}(P, \dot{P}) = 1$ iff $P \equiv \dot{P}$.

TABLE 15: Biclustering algorithms used in the experiments.

Algorithm	Implementation	Source	Reference
bbc	C	http://www.people.fas.harvard.edu/~junliu/BBC/	[16]
bcca	Matlab	http://sn.im/26fzpck	[57]
bibit	Java	http://www.upo.es/eps/bigs/BiBit_algorithm.html	[55]
bimax	R biclust package	http://cran.r-project.org/web/packages/biclust/index.html	[8]
cc	R biclust package	http://cran.r-project.org/web/packages/biclust/index.html	[4]
las	Matlab	https://genome.unc.edu/las/	[58]
msbe	Java	http://www.cs.cityu.edu.hk/~lwang/software/msbe/help.html	[39]
pcluster	Windows binary	http://haixun.olidu.com/proj/delta.html	[54]
xmotifs	R biclust package	http://cran.r-project.org/web/packages/biclust/index.html	[56]
fabia	R fabia package	http://www.bioconductor.org/packages/2.12/bioc/html/fabia.html	[37]

TABLE 16: Algorithm configuration.

Algorithm	Configuration
bbc	• -k (number of biclusters to be found): k^* • -n (normalization method): zero-mean, unit-variance column normalization
bcca	• -minc (minimum number of columns): min(nc) • -maxk (maximum number of biclusters to be found): k^* • -theta (Pearson correlation threshold): 0.90
bibit	• -minr (minimum number of rows): min(nr) • -minc (minimum number of cols): min(nc)
bimax	• -minr (minimum number of rows): min(nr) • -minc (minimum number of cols): min(nc) • -k (number of biclusters to be found): k^*
cc	• -alpha (scaling factor): 1.2 (as in the original paper) • -delta (maximum acceptable score): $(\max(A)-\min(A))^2/12 * 0.005$ (first experiment in the original paper) • -k (top biclusters having the smallest errors to return): k^*
las	• -maxk (maximum number of biclusters to be found): k^*
msbe	• -alpha: 0.3 (original paper, page 54, central value) • -beta: 0.25 (original paper, page 54, central value) • -gamma: beta+0.8 (original paper, page 54, central value) • -bitype: additive • -nrr (number of reference rows): n • -nrc (number of reference columns): p
pcluster	• -delta (maximum number of biclusters to be found): 1 • -minr (minimum number of rows): min(nr) • -minc (minimum number of cols): min(nc)
xmotifs	• -ns (number of seeds): 10 (original paper) • -nr (number of repetitions): 1000 (original paper) • -ss (sample size): 5 (implementation's default value) • -alpha: 0.05 (implementation's default value)
fabia	• -n (normalization method): 0.75-0.25 quantile • -c (data centering): 2 • -alpha (sparseness loadings): 0.01 • -spl (sparseness prior loadings): 0 • -spz (sparseness factors): 0.5 • -k (number of biclusters to be found): k^*

Definitions. min(nr): minimum number of rows that the reference biclusters have (similarly for min(nc)); minr: refer to the rows of biclusters (similarly for minc); k^* is the ideal number of biclusters; and min(A): minimum value of the data matrix (similarly for max(A)).