

Metacluster-based Projective Clustering Ensembles

Francesco Gullo · Carlotta Domeniconi ·
Andrea Tagarelli

Received: 25 May 2012 / Accepted: 11 June 2013 / Published online: 29 June 2013
© The Author(s) 2013

Abstract The Projective Clustering Ensemble (PCE) problem is a recent clustering advance aimed at combining the two powerful tools of *clustering ensembles* and *projective clustering*. PCE has been formalized as either a two-objective or a single-objective optimization problem. Two-objective PCE has been recognized as more accurate than its single-objective counterpart, although it is unable to jointly handle the object-based and feature-based cluster representations.

In this paper, we push forward the current PCE research, aiming to overcome the limitations of all existing PCE formulations. We propose a novel single-objective PCE formulation so that (i) the object-based and feature-based cluster representations are jointly considered, and (ii) the resulting optimization strategy follows a metacluster-based methodology borrowed from traditional clustering ensembles. As a result, the proposed formulation features best suitability to the PCE problem, thus guaranteeing improved effectiveness. Experiments on benchmark datasets have shown how the proposed approach achieves better average accuracy than all existing PCE methods, as well as efficiency superior to the most accurate existing metacluster-based PCE method on larger datasets.

Keywords Clustering · Clustering ensembles · Projective clustering · Subspace clustering · Dimensionality reduction · Optimization

Editors: Emmanuel Müller, Ira Assent, Stephan Günnemann, Thomas Seidl, Jennifer Dy.

F. Gullo (✉)
Yahoo! Research, 08018 Barcelona, Spain
e-mail: gullo@yahoo-inc.com

C. Domeniconi
Dept. of Computer Science, George Mason University, 22030 Fairfax, VA, USA
e-mail: carlotta@cs.gmu.edu

A. Tagarelli
DIMES Dept., University of Calabria, 87036 Rende (CS), Italy
e-mail: tagarelli@deis.unical.it

1 Introduction

After more than four decades, a large number of algorithms has been developed for data clustering, focusing on different aspects such as data types, algorithmic features, and application targets (Gan et al. 2007). In the last few years, there has been an increasing interest in developing advanced tools for data clustering. In this respect, *projective clustering* and *clustering ensembles* represent two of the most important directions of research.

The goal of projective clustering (or projected clustering) (Ng et al. 2005; Yiu and Mamoulis 2005; Achtert et al. 2006; Domeniconi et al. 2007; Moise et al. 2008) is to discover *projective clusters*, i.e., subsets of the input data having different (possibly overlapping) subsets of features (subspaces) associated with them. Projective clustering is closely related to the *subspace clustering* problem (Agrawal et al. 1998; Parsons et al. 2004; Kriegel et al. 2009; Moise et al. 2009), as both detect clusters that exist in different subspaces; however, the goal of subspace clustering is to search for *all* clusters in *all* meaningful subspaces, whereas projective clustering methods output a single partition of the input dataset. Projective clustering aims to solve issues that typically arise in high-dimensional data, such as sparsity and concentration of distances (Beyer et al. 1999; Hinneburg et al. 2000; Tomasev et al. 2011). Existing projective clustering methods (Kriegel et al. 2009; Moise et al. 2009) can be classified into four main approaches: *bottom-up* (it finds subspaces recognized as “interesting” and assigns each data object to the most similar subspace, Moise et al. 2008; Sequeira and Zaki 2004), *top-down* (it finds the subspaces starting from the full feature space, Ng et al. 2005; Liu et al. 2000; Yip et al. 2004, 2005; Achtert et al. 2006; Aggarwal et al. 1999; Woo et al. 2004; Böhm et al. 2004), *soft* (it produces soft data clusterings, Moise et al. 2008; Chen et al. 2008, and/or clusterings having differently weighted feature-to-cluster assignments, Domeniconi et al. 2007; Chen et al. 2008), and *hybrid* (it combines elements of both projective and subspace clustering approaches, Procopiuc et al. 2002; Achtert et al. 2007; Kriegel et al. 2005).

The problem of clustering ensembles (Strehl and Ghosh 2002; Topchy et al. 2005; Domeniconi and Al-Razgan 2009; Ghosh and Acharya 2011), also known as consensus clustering (Nguyen and Caruana 2007) or aggregation clustering (Gionis et al. 2007), is stated as follows: given a set of clustering solutions, or *ensemble*, to derive a *consensus clustering* that properly summarizes the solutions in the ensemble. The input ensemble is typically generated by varying one or more aspects of the clustering process, such as the clustering algorithm, the parameter setting, and the number of features, objects or clusters. The majority of the existing clustering ensemble methods follow the *instance-based* approach or the *metacluster-based* approach, or a combination of both (i.e., *hybrid*). In particular, the metacluster-based approach lies in the principle “to cluster clusters” (Bradley and Fayyad 1998; Strehl and Ghosh 2002; Boulis and Ostendorf 2004), whereby a new dataset is inferred whose objects are the clusters that belong to the clustering solutions in the ensemble. This dataset is then partitioned in order to produce a set of *metaclusters* (i.e., sets of clusters), and the consensus clustering is finally computed by assigning each data object to the metacluster that optimizes a specific criterion (e.g., majority voting).

Projective clustering and clustering ensembles have been recently treated in a unified framework (Gullo et al. 2009, 2013). The underlying motivation of that study is that many real-world application problems are high dimensional and lack a-priori knowledge. Examples are: clustering of multi-view data, privacy preserving clustering, news or document retrieval based on pre-defined categorizations, and distributed clustering of high-dimensional data. To address both issues simultaneously, the problem of *projective clustering ensembles* (PCE) is hence formalized, whose goal is to compute a *projective consensus clustering* from

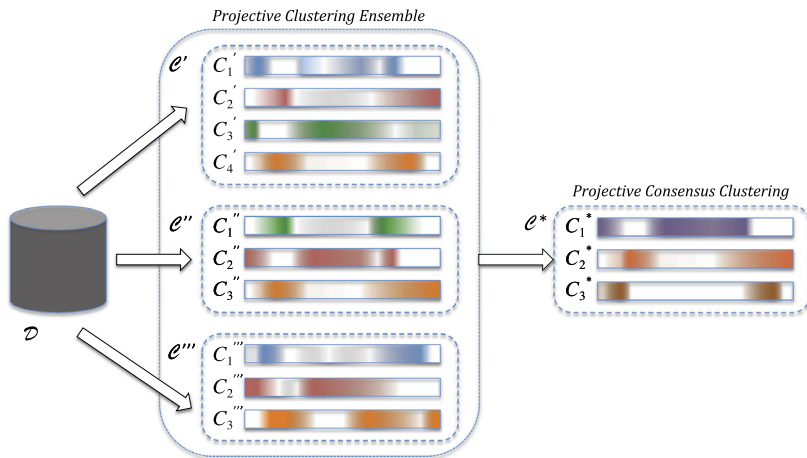


Fig. 1 Illustration of a projective clustering ensemble and derived consensus clustering. Each gradient refers to the cluster memberships over all objects. Colors denote different feature subspaces associated with the projective clusters. (The color version of this figure is available only in the electronic edition)

an ensemble of projective clustering solutions. Intuitively, each projective cluster is characterized by a distribution of memberships of the objects as well as a distribution over the features that belong to the subspace of that cluster. Figure 1 illustrates a projective clustering ensemble with three projective clustering solutions, which are obtained according to different views over the same dataset. A projective cluster is graphically represented as a rectangle filled with a color gradient, where higher intensities correspond to larger membership values of objects to the cluster. Clusters of the same clustering may overlap with their gradient (i.e., objects can have multiple assignments with different degrees of membership), and colors change to denote that different groupings of objects are associated with different feature subspaces. In the figure, a projective consensus clustering is derived by suitably “aggregating” the ensemble members. In particular, the first projective consensus cluster is derived by summarizing C_1' , C_2'' , and C_3''' , the second is derived from C_2' , C_3'' , and C_3''' , and the third is derived from C_4' , C_1'' , and C_1''' . Note that the resulting color in each projective consensus cluster resembles a merge of colors in the original projective clusters, which means that a projective consensus cluster is associated with a subset of features shared by the objects in the original clusters.

Two formulations of PCE have been proposed in Gullo et al. (2013), namely *two-objective* PCE and *single-objective* PCE. The former consists in the optimization of two objective functions, which separately consider the data object clustering and the feature-to-cluster assignment. The latter embeds in one objective function the object-based and feature-based representations of the various clusters. A heuristic developed for two-objective PCE, called *MOEA-PCE*, has shown to be particularly accurate, although it has drawbacks concerning efficiency, parameter setting, and interpretability of results. In contrast, the heuristic developed for single-objective PCE, called *EM-PCE*, has shown better efficiency while being outperformed by two-objective PCE in terms of effectiveness. An attempt to improve the accuracy of single-objective PCE has been proposed in Gullo et al. (2010); the approach is based on a corrective term designed to achieve a better balance between the object-to-cluster assignment and the feature-to-cluster assignment when measuring the error of a candidate projective consensus clustering. Despite the achieved improvement in accuracy, the heuristics in Gullo et al. (2010) are still outperformed by two-objective PCE, thus suggesting that

the path indicated by the two-objective formulation is the one to be followed. Nevertheless, the two-objective PCE suffers from a major weakness: it does not take into consideration the interrelation between the object-based and the feature-based cluster representations. In a nutshell, according to the early two-objective PCE formulation, consensus clusterings whose clusters have both object-based and feature-based representations that well comply with the input ensemble but that are not correctly “coupled” with one another might mistakenly be recognized as ideal. This fact can lead to projective consensus clustering solutions that contain conceptual flaws in their cluster composition. By preventing this scenario, one can improve the two-objective PCE formulation and corresponding methods.

Contributions In this paper, we pursue a new approach to the study of PCE, which is motivated by our insights on the theoretical foundations of two-objective and single-objective PCE formulations. Our aim is to provide a stronger tie between the PCE and the traditional clustering ensemble problem. By investigating the opportunity of adapting existing approaches for clustering ensembles to the PCE problem, we propose a new single-objective formulation of PCE which resembles a *metacluster-based* clustering ensemble approach and extends our first attempt in this regard (Gullo et al. 2011). The key idea underlying our proposal is to define a function that measures the distance of a projective clustering solution from a given ensemble, in such a way that the object-based and the feature-based cluster representations are considered as a whole. We show that the new PCE formulation is theoretically sound and has advantages over the ones proposed previously (Gullo et al. 2011, 2013). It enables the development of heuristic algorithms that can exploit the results obtained by the majority of existing clustering ensemble methods. Specifically, we define a heuristic that follows a metacluster-based approach, called *Enhanced MetaCluster-Based Projective Clustering Ensembles* (E-CB-PCE), which computes the consensus clustering starting from a partition (i.e., metaclustering) inferred from the set of clusters belonging to the ensemble components. Compared to the previously developed metacluster-based heuristics in Gullo et al. (2011), E-CB-PCE involves more constraints aimed to ensure a total coverage of the ensemble components in terms of clusters selected for deriving the projective consensus clustering solution. Experimental results have revealed that the proposed E-CB-PCE is on average more accurate than all previous PCE methods, according to each of the selected assessment criteria. Moreover, E-CB-PCE has shown to improve the efficiency of its most direct competitor CB-PCE (Gullo et al. 2011) on larger datasets up to two orders of magnitude.

Organization of the paper The remainder of this paper is organized as follows. Section 2 provides the background on the PCE problem and its early single-objective and two-objective formulations. Section 3 focuses on the proposed metacluster-based PCE approach. Section 4 presents the two developed heuristics for the metacluster-based PCE along with an analysis of their computational complexity. Section 5 describes the experimental evaluation and presents the results, and Sect. 6 concludes the paper. Finally, we give the proofs of all the theoretical results of the paper in the [Appendix](#).

2 Early Projective Clustering Ensembles (PCE)

Let \mathcal{D} be a set of data objects, where each $\mathbf{o} \in \mathcal{D}$ is an $|\mathcal{F}|$ -dimensional point defined over a feature space \mathcal{F} .¹ A *projective cluster* C defined over \mathcal{D} is a pair $\langle \Gamma_C, \Delta_C \rangle$, where

¹Vectorial notation here denotes row vectors.

- Γ_C denotes the *object-based* representation of C . It is a $|\mathcal{D}|$ -dimensional real-valued vector whose components $\Gamma_{C,o} \in [0, 1], \forall o \in \mathcal{D}$, represent the *object-to-cluster* assignment of o to C , i.e., the probability $\Pr(o|C)$ that the object o belongs to C ;
- Δ_C denotes the *feature-based* representation of C . It is an $|\mathcal{F}|$ -dimensional real-valued vector whose components $\Delta_{C,f} \in [0, 1], \forall f \in \mathcal{F}$, represent the *feature-to-cluster* assignments of the f -th feature to C , i.e., the probability $\Pr(f|C)$ that the feature f is *informative* for cluster C (f belongs to the subspace associated with C).

The object-based (Γ_C) and the feature-based (Δ_C) representations of a projective cluster C implicitly define the *projective cluster representation matrix* (for short, *projective matrix*) \mathbf{X}_C of C . \mathbf{X}_C is a $|\mathcal{D}| \times |\mathcal{F}|$ matrix that stores, $\forall o \in \mathcal{D}, f \in \mathcal{F}$, the probability of the intersection of the events “object o belongs to C ” and “feature f belongs to the subspace associated with C ”. Under the assumption of independence between the two events, such a probability is equal to $\Pr(C|o) = \Gamma_{C,o}$ joint with $\Pr(f|C) = \Delta_{C,f}$. Hence, given $\mathcal{D} = \{o_1, \dots, o_{|\mathcal{D}|}\}$ and $\mathcal{F} = \{1, \dots, |\mathcal{F}|\}$, the matrix \mathbf{X}_C can formally be defined as:

$$\mathbf{X}_C = \Gamma_C^T \Delta_C = \begin{pmatrix} \Gamma_{C,o_1} \times \Delta_{C,1} & \dots & \Gamma_{C,o_1} \times \Delta_{C,|\mathcal{F}|} \\ \vdots & & \vdots \\ \Gamma_{C,o_{|\mathcal{D}|}} \times \Delta_{C,1} & \dots & \Gamma_{C,o_{|\mathcal{D}|}} \times \Delta_{C,|\mathcal{F}|} \end{pmatrix}. \tag{1}$$

A couple of interesting results put in relation the projective matrix and the object/feature-based representation of a projective cluster. Such results are formalized in the following two propositions. We will exploit them in the remainder of the paper.

Proposition 1 *For any two projective clusters C, C' it holds that $\mathbf{X}_C = \mathbf{X}_{C'}$ if and only if $\Gamma_C = \Gamma_{C'}$ and $\Delta_C = \Delta_{C'}$.*

Proposition 2 *For any projective cluster C , its object-based representation $\Gamma_C = (\Gamma_{C,o_1}, \dots, \Gamma_{C,o_{|\mathcal{D}|}})$ and feature-based representation $\Delta_C = (\Delta_{C,1}, \dots, \Delta_{C,|\mathcal{F}|})$ can uniquely be derived from its projective matrix \mathbf{X}_C as follows:*

$$\Gamma_{C,o_i} = \sum_{j=1}^{|\mathcal{F}|} X_C(i, j), \quad \Delta_{C,j} = \frac{X_C(1, j)}{\Gamma_{C,o_1}} = \dots = \frac{X_C(|\mathcal{D}|, j)}{\Gamma_{C,o_{\mathcal{D}}}},$$

where $X_C(i, j)$ denotes the element (i, j) of the matrix \mathbf{X}_C .

A *projective clustering solution*, denoted by \mathcal{C} , is defined as a set of projective clusters that satisfy the following conditions:

$$\sum_{C \in \mathcal{C}} \Gamma_{C,o} = 1, \quad \forall o \in \mathcal{D}, \quad \text{and} \quad \sum_{f \in \mathcal{F}} \Delta_{C,f} = 1, \quad \forall C \in \mathcal{C}. \tag{2}$$

The semantics of any projective clustering \mathcal{C} is that for each projective cluster $C \in \mathcal{C}$, the objects belonging to C are close to each other if (and only if) they are projected onto the subspace associated with C .

A *projective ensemble* \mathcal{E} is defined as a set of projective clustering solutions. No information about the ensemble generation strategy (algorithms and/or setups), nor original feature values of the objects within \mathcal{D} are provided along with \mathcal{E} . Moreover, each projective clustering solution in \mathcal{E} may contain in general a different number of clusters.

The goal of PCE is to derive a *projective consensus clustering* that properly summarizes the projective clustering solutions within the input projective ensemble.

2.1 Single-objective PCE

A first PCE formulation proposed in Gullo et al. (2013) is based on a single-objective function:

$$C^* = \arg \min_c \sum_{C \in \mathcal{C}} \sum_{\mathbf{o} \in \mathcal{D}} \Gamma_{C,\mathbf{o}}^\alpha X_{C,\mathbf{o}}, \tag{3}$$

where

$$X_{C,\mathbf{o}} = \sum_{f \in \mathcal{F}} (\Delta_{C,f} - \Lambda_{\mathbf{o},f})^2, \tag{4}$$

$$\Lambda_{\mathbf{o},f} = \frac{1}{|\mathcal{E}|} \sum_{\hat{c} \in \mathcal{E}} \sum_{\hat{c} \in \hat{\mathcal{C}}} \Gamma_{\hat{c},\mathbf{o}} \Delta_{\hat{c},f}, \tag{5}$$

and $\alpha > 1$ is a positive integer that ensures non-linearity of the objective function w.r.t. $\Gamma_{C,\mathbf{o}}$. To solve the optimization problem based on the above function, the *EM-based Projective Clustering Ensembles (EM-PCE)* heuristic is defined. EM-PCE iteratively looks for the optimal values of $\Gamma_{C,\mathbf{o}}$ (resp. $\Delta_{C,f}$) while keeping $\Delta_{C,f}$ (resp. $\Gamma_{C,\mathbf{o}}$) fixed, until convergence.

Weaknesses of single-objective PCE The objective function at the basis of the problem in (3) does not allow for a perfect balance between object- and feature-to-cluster assignments when measuring the error of a candidate projective consensus clustering solution. This weakness is formally shown in Gullo et al. (2010) and avoided by adjusting (3) with a corrective term. The final form of the problem based on the corrected objective function is the following:

$$C^* = \arg \min_c \sum_{C \in \mathcal{C}} \sum_{\mathbf{o} \in \mathcal{D}} \Gamma_{C,\mathbf{o}}^\alpha \left(\frac{1}{2|\mathcal{E}|} X_{C,\mathbf{o}} + \frac{1}{|\mathcal{D}| - 1} X'_{C,\mathbf{o}} \right), \tag{6}$$

where

$$X'_{C,\mathbf{o}} = \sum_{\mathbf{o}' \neq \mathbf{o}} \left(1 - \frac{\Gamma_{C,\mathbf{o}'}}{|\mathcal{E}|} \sum_{\hat{c} \in \mathcal{E}} \sum_{\hat{c} \in \hat{\mathcal{C}}} \Gamma_{\hat{c},\mathbf{o}} \Gamma_{\hat{c},\mathbf{o}'} \right).$$

The above optimization problem is tackled in Gullo et al. (2010) by proposing two different heuristics. The first one, called *E-EM-PCE*, follows the same scheme as the EM-PCE algorithm for the early single-objective PCE formulation. The second heuristic, called *E-2S-PCE*, consists of two sequential steps that handle the object-to-cluster and the feature-to-cluster assignments separately.

2.2 Two-objective PCE

PCE is also formulated in Gullo et al. (2013) as a two-objective optimization problem, whose functions take into account the object-based (function Ψ_o) and the feature-based (function Ψ_f) cluster representations of a given projective ensemble \mathcal{E} , respectively:

$$C^* = \arg \min_c \{ \Psi_o(C, \mathcal{E}), \Psi_f(C, \mathcal{E}) \}, \tag{7}$$

where the arg min function is over all possible projective clustering solutions \mathcal{C} that satisfy the conditions reported in (2) (this makes the searching space in principle infinite), and

$$\Psi_o(\mathcal{C}, \mathcal{E}) = \sum_{\hat{\mathcal{C}} \in \mathcal{E}} \bar{\psi}_o(\mathcal{C}, \hat{\mathcal{C}}), \quad \Psi_f(\mathcal{C}, \mathcal{E}) = \sum_{\hat{\mathcal{C}} \in \mathcal{E}} \bar{\psi}_f(\mathcal{C}, \hat{\mathcal{C}}). \tag{8}$$

Functions $\bar{\psi}_o$ and $\bar{\psi}_f$ are defined as $\bar{\psi}_o(\mathcal{C}', \mathcal{C}'') = \frac{1}{2}(\psi_o(\mathcal{C}', \mathcal{C}'') + \psi_o(\mathcal{C}'', \mathcal{C}'))$ and $\bar{\psi}_f(\mathcal{C}', \mathcal{C}'') = \frac{1}{2}(\psi_f(\mathcal{C}', \mathcal{C}'') + \psi_f(\mathcal{C}'', \mathcal{C}'))$, respectively, where

$$\psi_o(\mathcal{C}', \mathcal{C}'') = \frac{1}{|\mathcal{C}'|} \sum_{\mathcal{C}' \in \mathcal{C}'} \left(1 - \max_{\mathcal{C}'' \in \mathcal{C}''} J(\mathbf{\Gamma}_{\mathcal{C}'}, \mathbf{\Gamma}_{\mathcal{C}''}) \right),$$

$$\psi_f(\mathcal{C}', \mathcal{C}'') = \frac{1}{|\mathcal{C}'|} \sum_{\mathcal{C}' \in \mathcal{C}'} \left(1 - \max_{\mathcal{C}'' \in \mathcal{C}''} J(\mathbf{\Delta}_{\mathcal{C}'}, \mathbf{\Delta}_{\mathcal{C}''}) \right),$$

and $J(\mathbf{u}, \mathbf{v}) = (\mathbf{u}\mathbf{v}^T) / (\|\mathbf{u}\|_2^2 + \|\mathbf{v}\|_2^2 - \mathbf{u}\mathbf{v}^T) \in [0, 1]$ denotes the extended Jaccard similarity coefficient (also known as Tanimoto coefficient) between any two real-valued vectors \mathbf{u} and \mathbf{v} (Strehl et al. 2000).

The problem defined in (7) is solved by a heuristic, called *MOEA-PCE*, in which a *Pareto-based Multi-Objective Evolutionary Algorithm* is exploited to avoid combining the two objective functions into a single one.

Weaknesses of two-objective PCE Experimental evidence in Gullo et al. (2013) has shown that the two-objective PCE formulation is much more accurate than the single-objective counterpart. Nevertheless, the original two-objective PCE also suffers from an important conceptual issue which has been firstly identified in Gullo et al. (2011), and we further investigate in this paper. The existence of this issue proves that the accuracy of two-objective PCE can still be improved, which is a major goal of this work. We unveil this issue in the following example.

Example 1 Let \mathcal{E} be a projective ensemble defined over a set \mathcal{D} of data objects and a set \mathcal{F} of features. Suppose that \mathcal{E} contains only one projective clustering solution \mathcal{C} and that \mathcal{C} in turn contains two projective clusters \mathcal{C}' and \mathcal{C}'' , whose object- and feature-based representations are different from one another, i.e., $\exists \mathbf{o} \in \mathcal{D}$ s.t. $\Gamma_{\mathcal{C}', \mathbf{o}} \neq \Gamma_{\mathcal{C}'', \mathbf{o}}$, and $\exists f \in \mathcal{F}$ s.t. $\Delta_{\mathcal{C}', f} \neq \Delta_{\mathcal{C}'', f}$.

Let us consider two candidate projective consensus clusterings $\mathcal{C}_1 = \{\mathcal{C}'_1, \mathcal{C}''_1\}$ and $\mathcal{C}_2 = \{\mathcal{C}'_2, \mathcal{C}''_2\}$. We assume that $\mathcal{C}_1 = \mathcal{C}$, whereas \mathcal{C}_2 is defined as follows. Cluster \mathcal{C}'_2 has object- and feature-based representations given by $\mathbf{\Gamma}_{\mathcal{C}'}$ (i.e., the object-based representation of the first cluster \mathcal{C}' within \mathcal{C}) and $\mathbf{\Delta}_{\mathcal{C}''}$ (i.e., the feature-based representation of the second cluster \mathcal{C}'' within \mathcal{C}), respectively; cluster \mathcal{C}''_2 has object- and feature-based representations given by $\mathbf{\Gamma}_{\mathcal{C}''}$ (i.e., the object-based representation of the second cluster \mathcal{C}'' within \mathcal{C}) and $\mathbf{\Delta}_{\mathcal{C}'}$ (i.e., the feature-based representation of the first cluster \mathcal{C}' within \mathcal{C}), respectively. According to (8), it is easy to see that:

$$\Psi_o(\mathcal{C}_1, \mathcal{E}) = \Psi_o(\mathcal{C}_2, \mathcal{E}) = 0, \quad \text{and} \quad \Psi_f(\mathcal{C}_1, \mathcal{E}) = \Psi_f(\mathcal{C}_2, \mathcal{E}) = 0.$$

Thus, both candidates \mathcal{C}_1 and \mathcal{C}_2 minimize the objectives of the early two-objective PCE formulation reported in (7), and hence, they are both recognized as optimal solutions. This conclusion is conceptually wrong, because only \mathcal{C}_1 should be recognized as an optimal solution, since only \mathcal{C}_1 exactly corresponds to the unique solution of the ensemble. Conversely,

\mathcal{C}_2 is not well-representative of the ensemble \mathcal{E} , as the object- and the feature-based representations of its clusters are inversely associated to each other w.r.t. the associations in \mathcal{C} . Indeed, in \mathcal{C}_2 , $C'_1 = \langle \Gamma_{C'}, \Delta_{C''} \rangle$ and $C''_1 = \langle \Gamma_{C''}, \Delta_{C'} \rangle$, whereas, the solution $\mathcal{C} \in \mathcal{E}$ is such that $C' = \langle \Gamma_{C'}, \Delta_{C'} \rangle$ and $C'' = \langle \Gamma_{C''}, \Delta_{C''} \rangle$.

The issue described in the above example arises because the two-objective PCE formulation ignores that the object-based and the feature-based representations of a projective cluster are strictly coupled to one another and, therefore, they need to be considered as a whole. In other words, in order to effectively evaluate the quality of a candidate projective consensus clustering, detecting the correct object-based and feature-based representations of the various clusters in a standalone fashion is not enough; instead, a mapping between the two (i.e., which object-based representation a feature-based representation should be coupled with and vice versa) should be properly discovered as well. For this purpose, the objective functions Ψ_o and Ψ_f should not be kept separated, but they should somehow be put in relation to one another. We show next how this can be overcome by employing a *metacluster-based* PCE formulation.

3 Metacluster-based PCE

3.1 Early metacluster-based PCE formulation

In our previous work (Gullo et al. 2011), we attempted to solve the main drawback of two-objective PCE shown in Example 1 by proposing the following alternative formulation based on a single-objective function:

$$C^* = \arg \min_{\mathcal{C}} \Psi_{of}(C, \mathcal{E}), \tag{9}$$

where Ψ_{of} is a function designed to measure the “distance” of any well-defined projective clustering solution \mathcal{C} from \mathcal{E} in terms of both data clustering and feature-to-cluster assignment. To define Ψ_{of} , we resorted to the early two-objective PCE formulation and adapted the (asymmetric) measure therein involved to the new setting:

$$\Psi_{of}(C, \mathcal{E}) = \sum_{\hat{C} \in \mathcal{E}} \overline{\psi}_{of}(C, \hat{C}), \tag{10}$$

where

$$\overline{\psi}_{of}(C', C'') = \frac{1}{2} (\psi_{of}(C', C'') + \psi_{of}(C'', C')), \tag{11}$$

and

$$\psi_{of}(C', C'') = \frac{1}{|C'|} \sum_{C' \in C'} \left(1 - \max_{C'' \in C''} \hat{J}(\mathbf{X}_{C'}, \mathbf{X}_{C''}) \right). \tag{12}$$

In order to measure the similarity between any pair C', C'' of projective clusters jointly in terms of object-based representation and feature-based representation, the corresponding projective matrices $\mathbf{X}_{C'}$ and $\mathbf{X}_{C''}$ are compared to each other. To accomplish this, the two matrices are first linearized as mono-dimensional vectors, and then compared by means of some distance measure between real-valued vectors. We resorted to the Tanimoto similarity coefficient (also known as extended Jaccard coefficient), as it represents a trade-off solution

between Euclidean and Cosine measures in terms of scale/translation invariance (Strehl et al. 2000). Moreover, it has a fixed-range codomain ($[0, 1]$), which is a desirable property in the design of the proposed objective function. More precisely, the generalized definition of the Tanimoto coefficient operating on real-valued matrices is as follows:

$$\hat{J}(\mathbf{X}, \hat{\mathbf{X}}) = \frac{\sum_{i=1}^{|\text{rows}(\mathbf{X})|} \mathbf{X}_i \hat{\mathbf{X}}_i^T}{\|\mathbf{X}\|_2^2 + \|\hat{\mathbf{X}}\|_2^2 - \sum_{i=1}^{|\text{rows}(\mathbf{X})|} \mathbf{X}_i \hat{\mathbf{X}}_i^T}, \tag{13}$$

where $\mathbf{X}_i \hat{\mathbf{X}}_i^T$ denotes the scalar product between the i -th rows of matrices \mathbf{X} and $\hat{\mathbf{X}}$.

It is easy to note that the PCE formulation reported in (9) is well-suited to measure the quality of a candidate consensus clustering in terms of both object-to-cluster and feature-to-cluster assignments as a whole. As shown in Gullo et al. (2011), this allows for overcoming the conceptual disadvantages of both early single-objective and two-objective PCE. Particularly, the issue described in Example 1 does not arise anymore with the PCE formulation in (9). Indeed, considering again the two candidate projective consensus clusterings C_1 and C_2 of Example 1, it is straightforward to see that:

$$\Psi_{of}(C_1, \mathcal{E}) = 0, \quad \text{and} \quad \Psi_{of}(C_2, \mathcal{E}) > 0.$$

and hence, C_1 would be correctly recognized as an optimal solution, whereas C_2 would not.

Weaknesses of early metacluster-based PCE formulation Although the early metacluster-based PCE formulation in (9) mitigates the issues of two-objective PCE, we will show how to further improve that formulation. For this purpose, next we provide a detailed analysis of the theoretical properties of the formulation in (9), with the ultimate goal of defining an *enhanced* metacluster-based PCE formulation that discards the controversial aspects of the earlier formulation.

Let us denote the expression $1 - \hat{J}(\mathbf{X}_{C'}, \mathbf{X}_{C''})$ with $T(\mathbf{X}_{C'}, \mathbf{X}_{C''})$ or, more simply, $T(C', C'')$, for any two clusters C', C'' . Since $\hat{J} \in [0, 1]$, $T(\cdot, \cdot)$ is regarded as Tanimoto distance (Gullo et al. 2011). Combining (10), (11), and (12), we have that:

$$\begin{aligned} \Psi_{of}(C, \mathcal{E}) &= \sum_{\hat{C} \in \mathcal{E}} \bar{\Psi}_{of}(C, \hat{C}) \\ &= \sum_{\hat{C} \in \mathcal{E}} \frac{1}{2} \left(\frac{1}{|C|} \sum_{C \in C} \left(1 - \max_{\hat{C} \in \hat{C}} \hat{J}(\mathbf{X}_C, \mathbf{X}_{\hat{C}}) \right) + \frac{1}{|\hat{C}|} \sum_{\hat{C} \in \hat{C}} \left(1 - \max_{C \in C} \hat{J}(\mathbf{X}_{\hat{C}}, \mathbf{X}_C) \right) \right). \end{aligned}$$

By discarding the constant terms and observing that

$$1 - \max_{C \in C} \hat{J}(\mathbf{X}_C, \mathbf{X}_{\hat{C}}) = \min_{C \in C} (1 - \hat{J}(\mathbf{X}_C, \mathbf{X}_{\hat{C}})) = \min_{C \in C} T(C, \hat{C}),$$

we can rewrite Ψ_{of} as:

$$\Psi_{of}(C, \mathcal{E}) = \underbrace{\sum_{\hat{C} \in \mathcal{E}} \sum_{C \in C} \min_{C \in C} T(C, \hat{C})}_{\Psi'_{of}(C, \mathcal{E})} + \underbrace{\sum_{C \in C} \sum_{\hat{C} \in \mathcal{E}} \min_{\hat{C} \in \hat{C}} T(C, \hat{C})}_{\Psi''_{of}(C, \mathcal{E})}. \tag{14}$$

Thus, the objective function Ψ_{of} corresponds to a sum of two objective functions Ψ'_{of} and Ψ''_{of} , which are the focus of the following discussion. Let us introduce the variable $x(\hat{C}, C) \in$

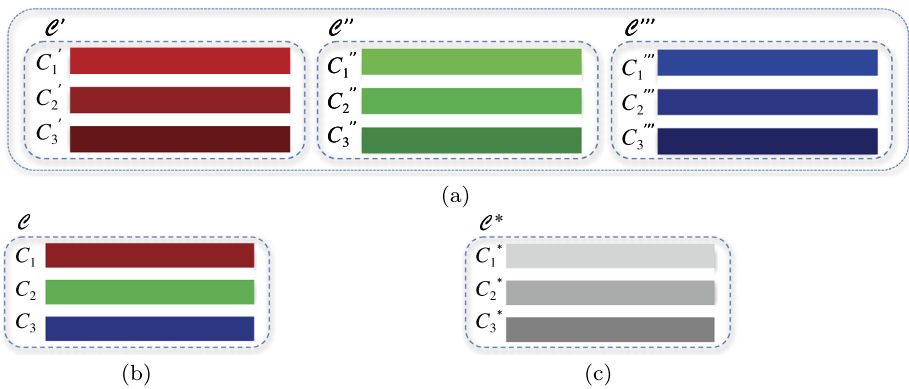


Fig. 2 (a) A projective ensemble where, for each projective cluster, different projective matrix representations are depicted with different color shades, (b) optimal projective consensus clustering according to the Ψ'_{of} objective function, (c) expected projective consensus clustering. (The color version of this figure is available only in the electronic edition)

$\{0, 1\}$, which is 1 if cluster \hat{C} , belonging to any projective clustering solution within the ensemble \mathcal{E} , is “mapped” to the cluster C of the candidate projective consensus clustering \mathcal{C} , and 0 otherwise. The meaning of the “mapping” between clusters is clarified next.

Explaining the Ψ'_{of} function The Ψ'_{of} function models a modified version of the K -Means problem (Jain and Dubes 1988), where (i) the input dataset for the clustering task corresponds to the set of projective clusters belonging to all solutions in the ensemble \mathcal{E} , i.e., the set $\{\hat{C} \mid \hat{C} \in \hat{\mathcal{C}}, \hat{C} \in \mathcal{E}\}$, (ii) each “object” in such a dataset is hence a projective cluster represented by its projective matrix, (iii) the centers have the form of projective matrices satisfying the constraints defined in (2), and (iv) the distance between objects and centers is computed according to the Tanimoto distance $T(\cdot, \cdot)$, rather than the classic squared Euclidean distance. Formally, the Ψ'_{of} function models the following optimization problem:

$$C^* = \arg \min_{\mathcal{C}} \sum_{C \in \mathcal{C}} \sum_{\hat{C} \in \mathcal{E}} \sum_{\hat{C} \in \hat{\mathcal{C}}} x(\hat{C}, C) T(\hat{C}, C) \tag{15}$$

s.t. \mathcal{C} satisfies (2),

$$x(\hat{C}, C) \in \{0, 1\}, \quad \forall \hat{C} \in \mathcal{E}, \forall \hat{C} \in \hat{\mathcal{C}}, \forall C \in \mathcal{C},$$

$$\sum_{C \in \mathcal{C}} x(\hat{C}, C) = 1, \quad \forall \hat{C} \in \mathcal{E}, \forall \hat{C} \in \hat{\mathcal{C}}. \tag{16}$$

Due to the analogies with K -Means, it is easy to see that the solution for the above problem is a set of K “center” projective matrices (representing the various clusters in the optimal C^*) that minimize the sum of the Tanimoto distances between each projective cluster in the ensemble and its closest center. In the following example we show that such a problem might not be particularly appropriate in the context of PCE.

Example 2 Figure 2(a) shows an ensemble composed by three projective clustering solutions. Projective clusters in each solution are depicted as colored rectangles whose color shade corresponds to a certain representation provided by its corresponding projective ma-

trix. Similar shades denote similar projective matrices, and, therefore, similar clusters.² In this example, the clusters within the same projective clustering solution are highly similar to each other, and are highly dissimilar from the clusters of the other solutions. Assuming $K = 3$ clusters in the output projective consensus clustering, the optimal partition of the ensemble of Fig. 2(a) according to the problem defined in (15)–(16) would correspond to the one identified by the solutions in the ensemble themselves. This leads to the optimal consensus clustering reported in Fig. 2(b), whose clusters $C_1, C_2,$ and C_3 take the colors *red, green,* and *blue* as a result of the summarization of the sets $\{C'_1, C'_2, C'_3\}, \{C''_1, C''_2, C''_3\}, \{C'''_1, C'''_2, C'''_3\}$, respectively. This solution is actually not very intuitive for PCE, since a good projective consensus clustering is supposed to summarize the information available from the ensemble by putting in relation clusters that belong to different solutions of the ensemble. Therefore, it would be much more meaningful if a consensus clustering would have the form reported in Fig. 2(c), whose *grey*-shaded clusters $C^*_1, C^*_2,$ and C^*_3 derive from mixing the *red, green,* and *blue* colors of the sets $\{C'_1, C''_1, C'''_1\}, \{C'_2, C''_2, C'''_2\}, \{C'_3, C''_3, C'''_3\}$, respectively.

Explaining the Ψ''_{of} function The optimization problem defined by the Ψ''_{of} function in (14) can be rewritten as follows:

$$C^* = \arg \min_C \sum_{\hat{C} \in \mathcal{C}} \sum_{C \in \mathcal{E}} \sum_{\hat{C} \in \hat{\mathcal{C}}} x(\hat{C}, C) T(\hat{C}, C) \tag{17}$$

s.t. \mathcal{C} satisfies (2),

$$x(\hat{C}, C) \in \{0, 1\}, \quad \forall \hat{C} \in \mathcal{E}, \forall \hat{C} \in \hat{\mathcal{C}}, \forall C \in \mathcal{C},$$

$$\sum_{\hat{C} \in \hat{\mathcal{C}}} x(\hat{C}, C) = 1, \quad \forall \hat{C} \in \mathcal{E}, \forall C \in \mathcal{C}. \tag{18}$$

Let us informally explain what is given above. Consider first the case where the number K of clusters in the output consensus clustering is equal to 1 and consider all subsets \mathbf{C} of $\{\hat{C} \mid \hat{C} \in \mathcal{C}, C \in \mathcal{E}\}$ that satisfy the following condition: each solution within the ensemble \mathcal{E} has exactly one cluster in \mathbf{C} ; also, for each \mathbf{C} , let $\mathbf{X}_{\mathbf{C}}$ denote the projective matrix that (i) satisfies (2), and (ii) minimizes the sum $d_{\mathbf{C}}$ of the Tanimoto distances between itself and (the projective matrices of) all members of \mathbf{C} . The problem defined by Ψ''_{of} (for $K = 1$) aims to find a matrix $\mathbf{X}_{\mathbf{C}^*}$ such that $d_{\mathbf{C}}$ is minimum; the final output would be a projective clustering C^* composed by only one cluster whose projective matrix corresponds to the matrix $\mathbf{X}_{\mathbf{C}^*}$. This interpretation can easily be generalized to the case $K > 1$; indeed, in this case, all K optimal solution matrices (i.e., projective clusters in C^*) would necessarily have the same form, as formally stated in the next proposition.

Proposition 3 *Given a projective ensemble \mathcal{E} , let Υ denote the set of all clusterings obtained by taking exactly one cluster from each ensemble member, i.e., $\Upsilon = \{\mathbf{C} \mid \mathbf{C} \subseteq \{\hat{C} \mid \hat{C} \in \mathcal{C} \wedge \hat{C} \in \mathcal{E}\} \wedge |\hat{C} \cap \hat{C}| = 1, \forall \hat{C} \in \mathcal{E}\}$. Moreover, let $\mathbf{X}^* = \arg \min_{\hat{\mathbf{X}}} \min_{\mathbf{C} \in \Upsilon} \sum_{C \in \mathbf{C}} T(\hat{\mathbf{X}}, \mathbf{X}_{\mathbf{C}})$ s.t. $\hat{\mathbf{X}}$ satisfies (2). Given an integer $K \geq 1$, it holds that the optimal projective consensus clustering for the optimization problem defined in (17)–(18) is $C^* = \{C^*_1, \dots, C^*_K\}$ s.t. $\mathbf{X}_{C^*_1} = \dots = \mathbf{X}_{C^*_K} = \mathbf{X}^*$.*

²The meaning of the colors here is different from Fig. 1, where different colors and gradients refer to the distinct parts of the two-fold projective cluster representation, i.e., the feature-to-cluster assignment and the object-to-cluster assignment, respectively. Here, instead, we focus on the unified view of the representation provided by the projective matrix.

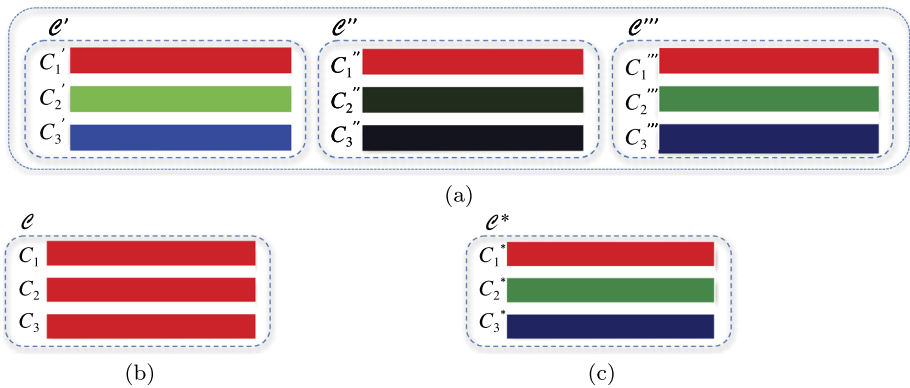


Fig. 3 (a) A projective ensemble where, for each projective cluster, different projective matrix representations are depicted with different color shades, (b) optimal projective consensus clustering according to the Ψ''_{of} objective function, (c) expected projective consensus clustering. (The color version of this figure is available only in the electronic edition)

The above result provides a clearer explanation of the optimization problem based on the objective function Ψ''_{of} . Some issues that may arise with Ψ''_{of} are discussed in the following example.

Example 3 The ensemble illustrated in Fig. 3(a) contains three projective clustering solutions whose clusters are very different from each other (different color shades), but are similar to some clusters from other solutions in the ensemble. In particular, C'_1, C''_1, C'''_1 are exactly the same (red shade), and the other groups of similar clusters are C'_2, C''_2, C'''_2 (green shade), and C'_3, C''_3, C'''_3 (blue shade). According to Proposition 3, the optimal solution for the optimization problem based on Ψ''_{of} with $K = 3$ is the one depicted in Fig. 3(b), as the clusters in the red-shaded group are more similar to each other than the green-shaded and blue-shaded groups. This solution is conceptually far away from the ideal one reported in Fig. 3(c), which is expected to be composed by three different clusters, each one summarizing a group of clusters with different shades.

3.2 Enhanced metacluster-based PCE formulation

We have discussed above the limitations due to the treatment of the functions Ψ'_{of} and Ψ''_{of} as separate (Gullo et al. 2011). Combining Ψ'_{of} and Ψ''_{of} into a single function somehow mitigates their respective undesired effects. Nevertheless, it is hard to understand to which extent each one of the two objective functions really contributes to hide the weaknesses of the other. Hence, it is more appropriate to have a problem formulation based on a single objective function which does not suffer from either of the drawbacks illustrated in Examples 2 and 3. We achieve this goal by proposing a modified metacluster-based PCE formulation, whose details are discussed next.

Let us first consider again the issue due to the function Ψ'_{of} described in Example 2. We recall that a major problem in this case is that the summarization provided by the clusters in the optimal consensus clustering might put in relation clusters from the same solutions in the ensemble, rather than coupling clusters from different ensemble components. It can be noted that this arises because the mappings of the various clusters in the ensemble to the center matrices are unconstrained (cf. the formal definition of the problem in (15)–(16)).

The issue can be overcome by defining such mappings so to guarantee that each center is associated with (at least) one cluster from each different solution in the ensemble. These constraints actually correspond to those defined in (18), which is indeed the optimization problem based on function Ψ''_{of} .

Focusing now on the function Ψ''_{of} , a possible solution to the issue described in Example 3 is to constrain the clusters forming the output projective consensus clustering to be different from each other. Alternatively, one can observe that the issue of Example 3 is mainly due to the fact that not all clusters in the ensemble are required to be mapped to a cluster of the projective consensus clustering. Based on this consideration, we can fix such an issue by resorting to some constraints of the problem based on the other function Ψ'_{of} , particularly the constraints listed in (16). Thus, in conclusion, the constraints reported in (16) and (18) represent a solution to the issues pertaining the functions Ψ'_{of} and Ψ''_{of} , respectively. As such, we define our enhanced metacluster-based PCE formulation by involving both these constraints. We hereinafter refer to this problem as CB-PCE ENHANCED.

Problem 1 (CB-PCE enhanced) Given a projective ensemble \mathcal{E} defined over a set \mathcal{D} of objects and a set \mathcal{F} of features, and an integer $K > 0$, find a projective clustering solution \mathcal{C}^* such that $|\mathcal{C}^*| = K$ and:

$$\mathcal{C}^* = \arg \min_{\mathcal{C}} \sum_{C \in \mathcal{C}} \sum_{\hat{C} \in \mathcal{E}} \sum_{\hat{C} \in \hat{\mathcal{C}}} x(\hat{C}, C) T(\hat{C}, C) \tag{19}$$

s.t. \mathcal{C} satisfies (2),

$$\begin{aligned} x(\hat{C}, C) &\in \{0, 1\}, \quad \forall \hat{C} \in \mathcal{E}, \forall \hat{C} \in \hat{\mathcal{C}}, \forall C \in \mathcal{C}, \\ \sum_{C \in \mathcal{C}} x(\hat{C}, C) &\geq 1, \quad \forall \hat{C} \in \mathcal{E}, \forall \hat{C} \in \hat{\mathcal{C}}, \end{aligned} \tag{20}$$

$$\sum_{\hat{C} \in \hat{\mathcal{C}}} x(\hat{C}, C) \geq 1, \quad \forall C \in \mathcal{C}. \tag{21}$$

Note that the constraints (20) and (21) have been slightly modified w.r.t. the original ones in (16) and (18), in order to handle the cases where the number of clusters of some ensemble solutions is smaller or larger than the number of clusters in the output consensus clustering. In summary, the constraints in (20) force each cluster \hat{C} of the ensemble to be mapped to at least one cluster in the candidate projective consensus clustering \mathcal{C} , while the constraints in (21) ensure that each cluster $C \in \mathcal{C}$ is coupled with at least one cluster from each ensemble solution $\hat{\mathcal{C}}$.

4 Heuristics for CB-PCE enhanced

The CB-PCE ENHANCED problem can formally be shown to be NP-hard. The detailed proof, which is reported in the Appendix, is based on a reduction from the JACCARD MEDIAN problem (Chierichetti et al. 2010). This reduction shows that CB-PCE ENHANCED remains hard even on a restricted version of the problem where the input dataset is a singleton and the number of clusters in both the output projective consensus clustering and each ensemble solution is 1.

Lemma 1 Let CB-PCE RESTRICTED be a special version of the CB-PCE ENHANCED problem where (i) $|\mathcal{D}| = 1$, (ii) $K = 1$ (K denotes the number of clusters in the output

Algorithm 1 CB-PCE

Input: a projective ensemble \mathcal{E} ; the number K of clusters in the output projective consensus clustering.
Output: the projective consensus clustering C^* .

- 1: $\mathbf{P} \leftarrow \text{pairwiseClusterDistances}(\mathcal{E})$
 - 2: $\mathbf{M} \leftarrow \text{metaclusters}(\mathcal{E}, \mathbf{P}, K)$
 - 3: $C^* \leftarrow \emptyset$
 - 4: **for all** $\mathcal{M} \in \mathbf{M}$ **do**
 - 5: compute $\Gamma_{\mathcal{M}}^*$ according to Theorem 2
 - 6: compute $\Delta_{\mathcal{M}}^*$ according to Theorem 3
 - 7: $C^* \leftarrow C^* \cup \{(\Gamma_{\mathcal{M}}, \Delta_{\mathcal{M}}^*)\}$
 - 8: **end for**
-

projective consensus clustering), (iii) $|\hat{C}| = 1, \forall \hat{C} \in \mathcal{E}$, (iv) $n(\Delta_{\hat{C}}) = n(\Delta_{\hat{C}'}) = n_{\mathcal{E}}, \forall \hat{C}, \hat{C}'$, where $n(\Delta_{\hat{C}}) = \sum_{f \in \mathcal{F}} \mathbb{I}[\Delta_{\hat{C}, f} > 0]$, and (v) $\Delta_{\hat{C}, f} = \frac{1}{n_{\mathcal{E}}}, \forall f \in \mathcal{F}, \forall \hat{C} \in \hat{\mathcal{C}}, \forall \hat{C}' \in \mathcal{E}$. CB-PCE RESTRICTED is **NP-hard**.

Theorem 1 CB-PCE ENHANCED is **NP-hard**.

The above results prompted us to develop heuristics for approximating CB-PCE ENHANCED. In this respect, consider the formulation of CB-PCE ENHANCED reported in (19)–(21) and suppose that, for any input ensemble \mathcal{E} , the optimal mappings between the clusters \hat{C} within \mathcal{E} and the clusters C of the output projective consensus partition are available, i.e., suppose we know in advance the optimal $x(\hat{C}, C)$ values, $\forall \hat{C} \in \mathcal{E}, \forall \hat{C}' \in \hat{\mathcal{C}}, \forall C \in \mathcal{C}$. Within this view, a metacluster might be derived for each $C \in \mathcal{C}$ that contains all clusters \hat{C} in the ensemble that are mapped to C (i.e., such that $x(\hat{C}, C) = 1$). In this way, the optimum for CB-PCE ENHANCED would be found by computing, for each metacluster, the projective matrix that minimizes the sum of the distances from all members of that metacluster under the constraints in (2). Unfortunately, neither the optimal $x(\hat{C}, C)$ mappings can be known in advance as they are part of the optimization process, nor the computation of the optimal projective matrices given an optimal mapping is feasible, as it is easy to observe from Lemma 1 that even this subproblem is **NP-hard**.

Nevertheless, the above reasoning interestingly reveals that CB-PCE ENHANCED can conceptually be split into two different sub-problems: the first one concerning the mapping of the clusters within the ensemble to the clusters in the output projective consensus clustering, and the second one consisting in finding optimal projective matrices given that mapping. To tackle these sub-problems, we improve upon the CB-PCE heuristic originally introduced in Gullo et al. (2011), as it still complies with the two-step nature of CB-PCE ENHANCED.

4.1 The CB-PCE heuristic

Algorithm 1 sketches the CB-PCE heuristic defined in Gullo et al. (2011). For clarity of presentation, the following symbols are used in addition to the notation provided in Sect. 2: \mathbf{M} denotes a set of metaclusters (i.e., a set of sets of clusters), $\mathcal{M} \in \mathbf{M}$ denotes a metacluster (i.e., a set of clusters), and $M \in \mathcal{M}$ denotes a cluster (i.e., a set of data objects). The mapping of clusters in the ensemble to clusters in the output projective consensus clustering is approximated in CB-PCE by exploiting a basic idea in traditional metacluster-based clustering ensembles, namely grouping all clusters in the ensemble in order to form a set of metaclusters (cf. Sect. 2). This is achieved by invoking the function *metaclusters* (Line 2),

which aims to cluster the set of clusters from all solutions within the input ensemble \mathcal{E} . This function exploits the matrix \mathbf{P} of pairwise distances between all clusters in \mathcal{E} (Line 1), which are computed using the Tanimoto distance defined in (13).

The second conceptual step of the CB-PCE ENHANCED problem consists in finding a suitable projective matrix representation for each cluster in the output consensus clustering, given the various metaclusters (Lines 4–8). As shown above, the ideal solution concerning the direct optimization of the sum of the Tanimoto distances is NP-hard. Therefore the idea underlying CB-PCE is to derive projective matrices by focusing on the optimization of a different criterion which is easier to solve and well-suited to find reasonable and effective approximations. Specifically, the solution provided by CB-PCE is to adapt the widely used *majority voting* (Strehl and Ghosh 2002) to the context at hand.

Deriving projective matrices from metaclusters For each metacluster we derive a projective matrix that optimizes the majority voting criterion. This is sufficient to obtain an approximation of the solution of the problem, since deriving the corresponding object- and feature-based cluster representations given a projective matrix can be performed easily by using Proposition 2. However, due to the use of the majority voting criterion, it is possible to derive object- and feature-based representation vectors directly, without requiring the computation of the projective matrix.

The object- and feature-based representations of each projective cluster to be included into the output consensus clustering \mathcal{C}^* are denoted as $\Gamma_{\mathcal{M}}^*$ and $\Delta_{\mathcal{M}}^*$, $\forall \mathcal{M} \in \mathbf{M}$, respectively. More precisely, $\Gamma_{\mathcal{M}}^*$ (resp. $\Delta_{\mathcal{M}}^*$) is the object-based (resp. feature-based) representation of the projective cluster within \mathcal{C}^* corresponding to the metacluster \mathcal{M} . Let us derive the values of $\Gamma_{\mathcal{M}}^*$ first. Since the ensemble can in principle contain projective clusterings that are soft at the clustering level, the majority voting criterion leads to the definition of the following optimization problem:

$$\{\Gamma_{\mathcal{M}}^* | \mathcal{M} \in \mathbf{M}\} = \arg \min_{\{\Gamma_{\mathcal{M}} | \mathcal{M} \in \mathbf{M}\}} Q \tag{22}$$

$$\text{s.t. } \sum_{\mathcal{M} \in \mathbf{M}} \Gamma_{\mathcal{M}, \mathbf{o}} = 1, \quad \forall \mathbf{o} \in \mathcal{D}, \tag{23}$$

$$\Gamma_{\mathcal{M}, \mathbf{o}} \geq 0, \quad \forall \mathcal{M} \in \mathbf{M}, \forall \mathbf{o} \in \mathcal{D}, \tag{24}$$

where

$$Q = \sum_{\mathcal{M} \in \mathbf{M}} \sum_{\mathbf{o} \in \mathcal{D}} \Gamma_{\mathcal{M}, \mathbf{o}}^\alpha A_{\mathcal{M}, \mathbf{o}}, \quad A_{\mathcal{M}, \mathbf{o}} = \frac{1}{|\mathcal{M}|} \sum_{M \in \mathcal{M}} (1 - \Gamma_{M, \mathbf{o}}),$$

and $\alpha > 1$ is an integer that guarantees the non-linearity of the objective function Q w.r.t. $\Gamma_{\mathcal{M}, \mathbf{o}}$, needed to ensure $\Gamma_{\mathcal{M}, \mathbf{o}}^* \in [0, 1]$ (rather than $\{0, 1\}$).³ The final solution for such a problem is stated in the next theorem. The details of its derivation are in the Appendix.

Theorem 2 *The optimum of the problem defined in (22)–(24) is ($\forall \mathcal{M}, \forall \mathbf{o}$):*

$$\Gamma_{\mathcal{M}, \mathbf{o}}^* = \left[\sum_{\mathcal{M}' \in \mathbf{M}} \left(\frac{A_{\mathcal{M}, \mathbf{o}}}{A_{\mathcal{M}', \mathbf{o}}} \right)^{\frac{1}{\alpha-1}} \right]^{-1}.$$

³An alternative way of obtaining $\Gamma_{\mathcal{M}, \mathbf{o}}^* \in [0, 1]$ is to employ regularization terms (Li and Mukaidono 1999).

A similar argument applies to $\Delta_{\mathcal{M}}^*$. In this case, the problem to be solved is as follows:

$$\{\Delta_{\mathcal{M}}^* | \mathcal{M} \in \mathbf{M}\} = \arg \min_{\{\Delta_{\mathcal{M}} | \mathcal{M} \in \mathbf{M}\}} \sum_{\mathcal{M} \in \mathbf{M}} \sum_{f \in \mathcal{F}} \Delta_{\mathcal{M},f}^\beta B_{\mathcal{M},f} \tag{25}$$

$$\text{s.t. } \sum_{f \in \mathcal{F}} \Delta_{\mathcal{M},f} = 1, \quad \forall \mathcal{M} \in \mathbf{M}, \tag{26}$$

$$\Delta_{\mathcal{M},f} \geq 0, \quad \forall \mathcal{M} \in \mathbf{M}, \forall f \in \mathcal{F}, \tag{27}$$

where $B_{\mathcal{M},f} = |\mathcal{M}|^{-1} \sum_{M \in \mathcal{M}} 1 - \Delta_{M,f}$ and β plays the same role as α in the function Q . The solution of such a problem is similar to that derived for $\Gamma_{\mathcal{M},\sigma}^*$.

Theorem 3 *The optimum of the problem defined in (25)–(27) is $(\forall \mathcal{M}, \forall f)$:*

$$\Delta_{\mathcal{M},f}^* = \left[\sum_{f' \in \mathcal{F}} \left(\frac{B_{\mathcal{M},f}}{B_{\mathcal{M},f'}} \right)^{\frac{1}{\beta-1}} \right]^{-1}.$$

4.2 The E-CB-PCE heuristic

Although CB-PCE provides an approximation to CB-PCE ENHANCED that reasonably exploits its intrinsic two-step nature, it does not take into account the new findings of this work. In particular, a major issue of CB-PCE is due to the approximation of the optimal mappings between clusters in the ensembles and clusters in the projective consensus clustering: it satisfies only the constraints listed in (20), and it ignores the constraints in (21). Indeed, the clustering of all the clusters in the ensemble in CB-PCE is carried out by using a standard clustering algorithm, and this is clearly not sufficient to enforce that each output metacluster must contain at least one cluster from each different projective clustering solution of the ensemble, which is what the constraints in (21) require. For this purpose, we describe next a modified version of CB-PCE, called *Enhanced CB-PCE (E-CB-PCE)*. E-CB-PCE follows the overall scheme of CB-PCE reported in Algorithm 1. The only difference is that it incorporates a well-suited (local-search) procedure to be used as the *metaclusters* subroutine in Line 2, whose main goal is to produce metaclusters satisfying the constraints in (21) along with those in (20).

The outline of the proposed method is given in Function 2. The set of metaclusters is initialized in such a way that the constraints in (20)–(21) are satisfied (Line 1). A score V for the set of metaclusters is computed, corresponding to the sum of all pairwise Tanimoto distances between the clusters in the same metacluster (Line 2). The procedure performs an iterative step aimed at improving the score V (Lines 3–20). In particular, we employ a local search that moves a projective cluster M from its source metacluster \mathcal{M}_s to a target metacluster $\mathcal{M}_t \neq \mathcal{M}_s$ (Lines 7–15). The move that causes the largest decrease in the score V is performed (Lines 16–19). The method terminates when no valid move that improves the current score V is available, i.e., a local minimum of V has been reached.

The evaluation of the move of a cluster M from its source metacluster \mathcal{M}_s to a target metacluster $\mathcal{M}_t \neq \mathcal{M}_s$ is performed by the method *evaluateMove* presented in Function 3. This method returns the relative score \hat{V} of the move along with a “swap” projective cluster \hat{M} within \mathcal{M}_t that aims to replace M in \mathcal{M}_s . If no swapping is required, no swap cluster is returned (i.e., $\hat{M} = \mathbf{nil}$). The move is evaluated by distinguishing between two cases. Let us denote with \mathcal{C}_M the projective clustering solution in the input ensemble which contains M . If the source metacluster \mathcal{M}_s contains one or more clusters of \mathcal{C}_M besides M , removing

Function 2 metaclusters($\mathcal{E}, \mathbf{P}, K$)

Input: a projective ensemble \mathcal{E} ; a matrix \mathbf{P} storing the pairwise distances between clusters in $\bigcup_{\hat{C} \in \mathcal{E}} \hat{C}$; the number K of output metaclusters.

Output: a set \mathbf{M} of K metaclusters.

```

1: initialize  $\mathbf{M} = \{\mathcal{M}_1, \dots, \mathcal{M}_K\}$  so to satisfy the constraints in (20)–(21)
2: compute initial  $V$  from  $\mathbf{P}$  as  $V \leftarrow \sum_{\mathcal{M} \in \mathbf{M}} \sum_{M_1, M_2 \in \mathcal{M}} T(M_1, M_2)$ 
3:  $V' \leftarrow V$ 
4: repeat
5:    $V \leftarrow V'$ 
6:    $M_{move} \leftarrow \mathbf{nil}, M_{swap} \leftarrow \mathbf{nil}, \mathcal{M}_{old} \leftarrow \mathbf{nil}, \mathcal{M}_{new} \leftarrow \mathbf{nil}$ 
7:   for all  $\mathcal{M}_s \in \mathbf{M}, M \in \mathcal{M}_s$  do
8:     for all  $\mathcal{M}_t \in \mathbf{M} \setminus \{\mathcal{M}_s\}$  do
9:        $(\hat{V}, \hat{M}) \leftarrow \text{evaluateMove}(M, \mathcal{M}_s, \mathcal{M}_t, \mathbf{P})$ 
10:      if  $V + \hat{V} < V'$  then
11:         $V' \leftarrow V + \hat{V}$ 
12:         $M_{move} \leftarrow M, M_{swap} \leftarrow \hat{M}, \mathcal{M}_{old} \leftarrow \mathcal{M}_s, \mathcal{M}_{new} \leftarrow \mathcal{M}_t$ 
13:      end if
14:    end for
15:  end for
16:  if  $V' < V$  then
17:    move  $M_{move}$  from  $\mathcal{M}_{old}$  to  $\mathcal{M}_{new}$ 
18:    if  $M_{swap} \neq \mathbf{nil}$ , move  $M_{swap}$  from  $\mathcal{M}_{new}$  to  $\mathcal{M}_{old}$ 
19:  end if
20: until no  $M$  changes metacluster
21: return  $\mathbf{M}$ 

```

Function 3 evaluateMove($M, \mathcal{M}_s, \mathcal{M}_t, \mathbf{P}$)

Input: a projective cluster M ; a source metacluster \mathcal{M}_s ; a target metacluster \mathcal{M}_t ; a matrix \mathbf{P} storing the pairwise distances between clusters in $\bigcup_{\hat{C} \in \mathcal{E}} \hat{C}$.

Output: a pair (\hat{V}, \hat{M}) , where \hat{V} is a real value and \hat{M} is a projective cluster.

```

1: for any projective cluster  $M$  and metacluster  $\mathcal{M}$ , let  $g(M, \mathcal{M})$  denote  $\sum_{M' \in \mathcal{M}} T(M, M')$ 
2:  $\hat{M} \leftarrow \mathbf{nil}$ 
3: let  $\mathcal{C}_M$  the projective clustering in  $\mathcal{E}$  s.t.  $M \in \mathcal{C}_M$ 
4: if  $\mathcal{M}_s \cap \mathcal{C}_M \setminus \{M\} \neq \emptyset$  then
5:    $\hat{V} \leftarrow g(M, \mathcal{M}_t) - g(M, \mathcal{M}_s)$ 
6: else
7:    $\hat{M} \leftarrow \arg \min_{M' \in \mathcal{M}_t \cap \mathcal{C}_M} [(g(M, \mathcal{M}_t \setminus \{M'\}) - g(M, \mathcal{M}_s)) + (g(M', \mathcal{M}_s \setminus \{M\}) - g(M', \mathcal{M}_t))]$ 
8:    $\hat{V} \leftarrow [(g(M, \mathcal{M}_t \setminus \{\hat{M}\}) - g(M, \mathcal{M}_s)) + (g(\hat{M}, \mathcal{M}_s \setminus \{M\}) - g(\hat{M}, \mathcal{M}_t))]$ 
9: end if
10: return  $(\hat{V}, \hat{M})$ 

```

M from \mathcal{M}_s does not violate the constraints in (21). Therefore, in this case, the score \hat{V} is computed as the sum of the Tanimoto distances of M from all the clusters in the new metacluster \mathcal{M}_t minus the distances between M and all the members of the old metacluster \mathcal{M}_s (Lines 4–5). Otherwise, if \mathcal{M}_s does not contain any other cluster of \mathcal{C}_M , the move evaluation also takes into account that a “swap” is needed to ensure that the constraints in (21) are satisfied (Lines 7–8). More precisely, M needs to be replaced with another cluster from \mathcal{C}_M that currently belongs to the target metacluster \mathcal{M}_t . Such a cluster from \mathcal{M}_t is chosen in such a way that the resulting score \hat{V} is minimized.

4.3 Computational analysis

The bottleneck of the early CB-PCE heuristic is the computation of the pairwise Tanimoto distances between the clusters in the ensemble (Line 1 in Algorithm 1). For any single pair of clusters, the Tanimoto distance in principle takes $\mathcal{O}(|\mathcal{D}||\mathcal{F}|)$ time, as it requires to consider all the elements of the $|\mathcal{D}| \times |\mathcal{F}|$ projective matrices of those clusters. Nevertheless, we show in Proposition 4 that the Tanimoto distance formula can be rewritten in such a way that its computation takes instead $\mathcal{O}(|\mathcal{D}| + |\mathcal{F}|)$ time.

Proposition 4 *For any two projective clusters C, C' it holds that:*

$$T(\mathbf{X}_C, \mathbf{X}_{C'}) = 1 - \frac{p(C, C')}{q(C) + q(C') - p(C, C')},$$

where

$$p(C, C') = (\mathbf{\Gamma}_C \mathbf{\Gamma}_{C'}^T) \times (\mathbf{\Delta}_C \mathbf{\Delta}_{C'}^T), \quad q(C) = (\|\mathbf{\Gamma}_C\|_2^2) \times (\|\mathbf{\Delta}_C\|_2^2).$$

Thanks to the above result, the proposed E-CB-PCE improves upon the efficiency of the early CB-PCE. In particular, E-CB-PCE can now achieve a better time complexity for the Tanimoto distance of a single pair of clusters without introducing any approximation, which was the solution exploited in our previous work to obtain a speed-up (cf. FCB-PCE heuristic in Gullo et al. 2011).

Time complexity of E-CB-PCE Next we discuss in detail the computational complexity of the proposed E-CB-PCE. We are given: a set \mathcal{D} of data objects, each one defined over a feature space \mathcal{F} , a projective ensemble \mathcal{E} defined over \mathcal{D} and \mathcal{F} , and a positive integer K representing the number of clusters in the output projective consensus clustering. It is also reasonable to assume that the size $|\mathcal{C}|$ of each solution \mathcal{C} in \mathcal{E} is $\mathcal{O}(K)$. The complexity of E-CB-PCE can be broken down into three stages:

1. *Pre-processing*: it concerns the computation of the pairwise distances between clusters, by applying the Tanimoto distance $T(\cdot, \cdot)$ to projective matrices. According to Proposition 4, any single distance computation can be performed in $\mathcal{O}(|\mathcal{D}| + |\mathcal{F}|)$ time. Thus, computing the overall pairwise matrix takes $\mathcal{O}(K^2|\mathcal{E}|^2(|\mathcal{D}| + |\mathcal{F}|))$ time;
2. *Meta-clustering*: it concerns the clustering of the $\mathcal{O}(K|\mathcal{E}|)$ clusters of all the solutions in the ensemble according to the procedure described in Function 2. This procedure is based on a local-search optimization strategy of an objective function that is quadratic in the number of all clusters within the input ensemble. Thus, its time complexity is $\mathcal{O}(IK^2|\mathcal{E}|^2)$, where I is the number of iterations to convergence.
3. *Post-processing*: it concerns the assignment of objects and features to the metaclusters based on Theorems 2 and 3. According to those theorems, both the object and the feature assignments need to look up all the clusters in each metacluster only once; thus, for each object and for each feature, it takes $\mathcal{O}(K|\mathcal{E}|)$ time. Performing this step for all objects and features leads to a total cost of $\mathcal{O}(K|\mathcal{E}|(|\mathcal{D}| + |\mathcal{F}|))$ for the entire post-processing step.

It can be noted that the first step is an *offline* phase, i.e., a phase to be performed only once in case of a multi-run execution, whereas the second and third are *online* steps. Thus, as summarized in Table 1 (where we also report the complexities of the earlier MOEA-PCE,

Table 1 Computational complexities of PCE methods

Method	Total	Online	Offline
<i>MOEA-PCE</i> (Gullo et al. 2009, 2013)	$\mathcal{O}(ItK^2 \mathcal{E} (\mathcal{D} + \mathcal{F}))$	$\mathcal{O}(ItK^2 \mathcal{E} (\mathcal{D} + \mathcal{F}))$	–
<i>EM-PCE</i> (Gullo et al. 2009, 2013)	$\mathcal{O}(K \mathcal{E} \mathcal{D} \mathcal{F})$	$\mathcal{O}(IK \mathcal{D} \mathcal{F})$	$\mathcal{O}(K \mathcal{E} \mathcal{D} \mathcal{F})$
<i>E-EM-PCE</i> (Gullo et al. 2010)	$\mathcal{O}(K \mathcal{E} \mathcal{D} (IK + \mathcal{F}))$	$\mathcal{O}(IK \mathcal{D} (K \mathcal{E} + \mathcal{F}))$	$\mathcal{O}(K \mathcal{E} \mathcal{D} \mathcal{F})$
<i>E-2S-PCE</i> (Gullo et al. 2010)	$\mathcal{O}(K \mathcal{E} \mathcal{D} (\mathcal{D} + \mathcal{F}))$	$\mathcal{O}(K \mathcal{D} (I \mathcal{D} + \mathcal{E} \mathcal{F}))$	$\mathcal{O}(K \mathcal{E} \mathcal{D} (\mathcal{D} + \mathcal{F}))$
<i>CB-PCE</i> (Gullo et al. 2011)	$\mathcal{O}(K^2 \mathcal{E} ^2 \mathcal{D} \mathcal{F})$	$\mathcal{O}(K \mathcal{E} (K \mathcal{E} + \mathcal{D} + \mathcal{F}))$	$\mathcal{O}(K^2 \mathcal{E} ^2 \mathcal{D} \mathcal{F})$
<i>FCB-PCE</i> (Gullo et al. 2011)	$\mathcal{O}(K^2 \mathcal{E} ^2(\mathcal{D} + \mathcal{F}))$	$\mathcal{O}(K \mathcal{E} (K \mathcal{E} + \mathcal{D} + \mathcal{F}))$	$\mathcal{O}(K^2 \mathcal{E} ^2(\mathcal{D} + \mathcal{F}))$
<i>E-CB-PCE</i>	$\mathcal{O}(K^2 \mathcal{E} ^2(I + \mathcal{D} + \mathcal{F}))$	$\mathcal{O}(K \mathcal{E} (IK \mathcal{E} + \mathcal{D} + \mathcal{F}))$	$\mathcal{O}(K^2 \mathcal{E} ^2(\mathcal{D} + \mathcal{F}))$

EM-PCE (Gullo et al. 2013), E-EM-PCE, E-2S-PCE (Gullo et al. 2010), and CB-PCE, FCB-PCE (Gullo et al. 2011) methods,⁴ we can conclude that the *offline*, *online*, and *total* (i.e., offline + online) complexities of E-CB-PCE are $\mathcal{O}(K^2|\mathcal{E}|^2(|\mathcal{D}| + |\mathcal{F}|))$, $\mathcal{O}(K|\mathcal{E}|(IK|\mathcal{E}| + |\mathcal{D}| + |\mathcal{F}|))$, and $\mathcal{O}(K^2|\mathcal{E}|^2(I + |\mathcal{D}| + |\mathcal{F}|))$, respectively.

5 Experimental evaluation

We evaluated accuracy and efficiency of the proposed E-CB-PCE algorithm and compared it with the early PCE methods, i.e., MOEA-PCE, EM-PCE (Gullo et al. 2009, 2013), E-EM-PCE, E-2S-PCE (Gullo et al. 2010), and CB-PCE (Gullo et al. 2011). In the following, we introduce our evaluation methodology which includes the selected datasets, the strategy used for generating the projective ensembles, the setup of the proposed algorithms, and the assessment criteria for the projective consensus clusterings. Finally, we discuss obtained experimental results.

5.1 Evaluation methodology

5.1.1 Datasets

We selected 22 publicly available datasets with different characteristics in terms of number of objects, features, and classes. In the summary provided in Table 2, the first fifteen datasets are from the UCI Machine Learning Repository (Asuncion and Newman 2010), the next four datasets are from the UCR Time Series Classification/Clustering Page (Keogh et al. 2003), whereas the last three datasets are synthetically generated and selected from Müller et al.

⁴In Table 1, t denotes the population size for the genetic algorithm at the basis of MOEA-PCE.

Table 2 Datasets used in the experiments

Dataset	Objects	Features	Classes
Iris	150	4	3
Wine	178	13	3
Glass	214	10	6
Ecoli	327	7	5
Yeast	1,484	8	10
Multiple-Features	2,000	585	10
Segmentation	2,310	19	7
Abalone	4,124	7	17
Waveform	5,000	40	3
Letter	7,648	16	10
Isolet	7,797	617	26
Gisette	13,500	5,000	2
p53-Mutants	300	5,409	2
Amazon	120	10,000	4
Arcene	200	10,000	2
Shapes	160	1,614	9
Tracedata	200	275	4
ControlChart	600	60	6
Twopat	800	128	4
N30	1,356	20	8
D75	1,365	75	7
S2500	2,262	20	8

(2009). Note that the synthetic datasets originally had overlapping clusters. We selected for each dataset the maximal subset of data objects forming a partition and the corresponding natural subspace for each cluster in the partition.

5.1.2 Projective ensemble generation

We adopted a basic strategy for projective ensemble generation, which consists in selecting a (projective) clustering algorithm and varying the parameter(s) of that algorithm in order to guarantee the diversity of the solutions within the projective ensemble. We were not interested in comparing projective clustering algorithms and assessing the impact of their performance on projective ensemble generation, since generating projective ensembles with the highest quality is not a goal of this work; nevertheless, we resorted to a state-of-the-art algorithm, LAC, whose effectiveness in the context of projective clustering has been already proven (Domeniconi et al. 2007). The diversity of the projective clustering solutions was ensured by randomly choosing the initial centroids and varying the LAC's parameter h .

Note that LAC yields projective clusterings that have hard object-to-cluster assignments and have weighted feature-to-cluster assignments. Therefore, in order to test the ability of the proposed algorithm to also deal with soft clustering solutions and with solutions having unweighted feature-to-cluster assignments, we generated each projective ensemble \mathcal{E} as a composition of four equally-sized subsets, denoted as \mathcal{E}_1 , \mathcal{E}_2 , \mathcal{E}_3 , and \mathcal{E}_4 and defined as follows:

- \mathcal{E}_1 contains solutions that have hard object-to-cluster assignments and weighted feature-to-cluster assignments, i.e., solutions as provided by standard LAC;
- \mathcal{E}_2 contains solutions that have hard object-to-cluster assignments and unweighted feature-to-cluster assignments. Starting from a LAC solution \mathcal{C} defined over a set \mathcal{D} of data objects and a set \mathcal{F} of features, a projective clustering \mathcal{C}' having unweighted feature-to-cluster assignments is derived such that $\Delta_{\mathcal{C}',f} = \mathbf{I}[\Delta_{\mathcal{C},f} \geq |\mathcal{F}|^{-1} \sum_{f' \in \mathcal{F}} \Delta_{\mathcal{C},f'}]$, $\forall \mathcal{C}' \in \mathcal{C}'$, $\forall f \in \mathcal{F}$, where $\mathbf{I}[A]$ is the *indicator function*, which is equal to 1 when the event A is true, and 0 otherwise;
- \mathcal{E}_3 contains solutions that have soft object-to-cluster assignments and weighted feature-to-cluster assignments. Starting from a LAC solution \mathcal{C} , a soft projective clustering \mathcal{C}'' is derived by computing the $\Gamma_{\mathcal{C}'',\mathbf{o}}$ values ($\forall \mathcal{C}'' \in \mathcal{C}'', \forall \mathbf{o} \in \mathcal{D}$), proportionally to the distance of \mathbf{o} from the centroids $\bar{\mathcal{C}}''$ of the clusters \mathcal{C}'' :

$$\Gamma_{\mathcal{C}'',\mathbf{o}} = \frac{\sum_{f \in \mathcal{F}} (o_f - \bar{\mathcal{C}}''_f)^2}{\sum_{C \in \mathcal{C}''} \sum_{f \in \mathcal{F}} (o_f - \bar{\mathcal{C}}_f)^2},$$

where the f -th feature $\bar{\mathcal{C}}_f$ of the centroid of any cluster C is defined as $\bar{\mathcal{C}}_f = |C|^{-1} \sum_{\mathbf{o} \in C} o_f$.

- \mathcal{E}_4 contains solutions that have soft object-to-cluster assignments and unweighted feature-to-cluster assignments. The solutions are derived from standard LAC solutions according to the methods employed for generating \mathcal{E}_2 and \mathcal{E}_3 , respectively.

For each dataset, we generated 10 different projective ensembles; all results we present in the following correspond to averages over these projective ensembles.

5.1.3 Assessment criteria

We assessed the quality of a projective consensus clustering \mathcal{C} using both external and internal cluster validity criteria: the former is based on the similarity of \mathcal{C} w.r.t. a reference classification, whereas the latter is based on the average similarity w.r.t. the solutions in the input projective ensemble \mathcal{E} .

Similarity w.r.t. the reference classification (external evaluation) This evaluation stage exploits the availability of a reference classification, hereinafter denoted as $\tilde{\mathcal{C}}$, for any given dataset \mathcal{D} . Note that all the selected datasets are coupled with a reference classification that provides information about the ideal object-to-cluster assignments $\Gamma_{\tilde{\mathcal{C}},\mathbf{o}}$ ($\forall \tilde{\mathcal{C}} \in \tilde{\mathcal{C}}, \forall \mathbf{o} \in \mathcal{D}$), which are hard assignments. The $\Delta_{\tilde{\mathcal{C}},f}$ feature-to-cluster assignments are instead defined according to the following approaches:

- For the synthetic datasets N30, D75, and S2500, which already provide information about the ideal subspaces assigned to each group of objects identified by the reference classification, these subspaces are directly used to define unweighted $\Delta_{\tilde{\mathcal{C}},f}$ feature-to-cluster assignments in $\tilde{\mathcal{C}}$.
- For the remaining datasets, the $\Delta_{\tilde{\mathcal{C}},f}$ values are derived by applying the procedure suggested in Domeniconi et al. (2007) to the reference classification $\tilde{\mathcal{C}}$: given the $\Gamma_{\tilde{\mathcal{C}},\mathbf{o}}$ values ($\forall \tilde{\mathcal{C}} \in \tilde{\mathcal{C}}, \forall \mathbf{o} \in \mathcal{D}$) originally provided along with $\tilde{\mathcal{C}}$, the $\Delta_{\tilde{\mathcal{C}},f}$ values are computed as:

$$\Delta_{\tilde{\mathcal{C}},f} = \frac{\exp(-U(\tilde{\mathcal{C}}, f)/h)}{\sum_{f' \in \mathcal{F}} \exp(-U(\tilde{\mathcal{C}}, f')/h)},$$

where the LAC parameter h is set to 0.2 and:

$$U(\tilde{C}, f) = \left(\sum_{o \in \mathcal{D}} \Gamma_{\tilde{C}, o} \right)^{-1} \sum_{o \in \mathcal{D}} \Gamma_{\tilde{C}, o} (\bar{C}_f - o_f)^2, \quad \bar{C}_f = \left(\sum_{o \in \mathcal{D}} \Gamma_{\tilde{C}, o} \right)^{-1} \sum_{o \in \mathcal{D}} \Gamma_{\tilde{C}, o} \times o_f.$$

In order to compute the similarity between a projective consensus clustering \mathcal{C} and a reference classification \tilde{C} , we resort to the popular *F1-measure* (van Rijsbergen 1979). Particularly, here we provide a definition of *F1-measure* that enables a comparison between projective clustering having *soft* object/feature-to-cluster assignments. Given a projective cluster $C \in \mathcal{C}$, the precision $P(C)$ and the recall $R(C)$ are defined as:

$$P(C) = \frac{\max_{\tilde{C} \in \tilde{\mathcal{C}}} \text{overlap}(\tilde{C}, C)}{\text{size}(C)}, \quad R(C) = \frac{\max_{\tilde{C} \in \tilde{\mathcal{C}}} \text{overlap}(\tilde{C}, C)}{\text{size}(\arg \max_{\tilde{C} \in \tilde{\mathcal{C}}} \text{overlap}(\tilde{C}, C))},$$

and the *F1-measure* is defined as:

$$F1(\tilde{C}, \mathcal{C}) = \frac{1}{|\mathcal{C}|} \sum_{C \in \mathcal{C}} \frac{2P(C)R(C)}{P(C) + R(C)}.$$

The values of the *F1-measure* belong to the interval $[0, 1]$, where larger values indicate more accurate projective consensus clusterings. The *overlap*(\cdot, \cdot) and *size*(\cdot) functions quantify the degree of overlap between two projective clusters and the size of a projective cluster, respectively. Such functions are defined based on three ways of comparing the various projective clusters, namely *object-based* (o), *feature-based* (f), and *object & feature-based* (of), which respectively account for the object-based representations only of the projective clusters to be compared, the feature-based representation only, or both. We hence define variants of the *overlap*(\cdot, \cdot) and the *size*(\cdot) functions to handle each of the three cases:

- Object-based (measure $F1_o$): $\text{overlap}(C', C'') = \sum_{o \in \mathcal{D}} \Gamma_{C', o} \Gamma_{C'', o}$, $\text{size}(C) = \sum_{o \in \mathcal{D}} \Gamma_{C, o}$;
- feature-based (measure $F1_f$): $\text{overlap}(C', C'') = \sum_{f \in \mathcal{F}} \Delta_{C', f} \Delta_{C'', f}$, $\text{size}(C) = \sum_{f \in \mathcal{F}} \Delta_{C, f}$;
- object & feature-based (measure $F1_{of}$): $\text{overlap}(C', C'') = (\sum_{o \in \mathcal{D}} \Gamma_{C', o} \Gamma_{C'', o}) \times (\sum_{f \in \mathcal{F}} \Delta_{C', f} \Delta_{C'', f})$, $\text{size}(C) = (\sum_{o \in \mathcal{D}} \Gamma_{C, o}) (\sum_{f \in \mathcal{F}} \Delta_{C, f})$.

Similarity w.r.t. the projective ensemble solutions (internal evaluation) Any valid projective consensus clustering \mathcal{C} should comply with the information available from the input projective ensemble \mathcal{E} . In this respect, we carried out an evaluation stage to measure the average similarity between a projective consensus clustering and the solutions within \mathcal{E} . We define the object & feature-based measure $\overline{F1}_{of}$ (object-based $\overline{F1}_o$ and feature-based $\overline{F1}_f$ are defined similarly) as follows:

$$\overline{F1}_{of}(C) = \frac{1}{|\mathcal{E}|} \sum_{\hat{C} \in \mathcal{E}} \max\{F1(C, \hat{C}), F1(\hat{C}, C)\}.$$

All these measures range from $[0, 1]$; moreover, the larger the values $\overline{F1}_{of}$, $\overline{F1}_o$, or $\overline{F1}_f$ are, the larger the similarity between the projective consensus clustering \mathcal{C} and the solutions within the projective ensemble is, and hence the better the quality of \mathcal{C} .

Table 3 Evaluation w.r.t. the reference classification ($F1_{of}$)

Dataset	MOEA-PCE	EM-PCE	E-EM-PCE	E-2S-PCE	CB-PCE	E-CB-PCE
Iris	0.649(0.025)	0.588(0.002)	0.532(0.046)	0.543(0.074)	0.640(0.071)	0.673(0.017)
Wine	0.345(0.025)	0.300(0.003)	0.303(0.001)	0.266(0.047)	0.327(0.028)	0.356(0.001)
Glass	0.279(0.009)	0.298(0.007)	0.307(0.009)	0.253(0.035)	0.311(0.042)	0.420(0.032)
Ecoli	0.518(0.020)	0.564(0.013)	0.573(0.007)	0.540(0.104)	0.599(0.063)	0.550(0.035)
Yeast	0.288(0.009)	0.237(0.004)	0.232(0.005)	0.202(0.025)	0.317(0.025)	0.308(0.008)
Mult.-Feat.	0.270(0.019)	0.300(0.015)	0.290(0.048)	0.405(0.041)	0.174(0.021)	0.180(0.001)
Segmentation	0.334(0.018)	0.400(0.008)	0.401(0.001)	0.382(0.026)	0.364(0.054)	0.452(0.029)
Abalone	0.116(0.003)	0.112(0.003)	0.111(0.004)	0.080(0.006)	0.116(0.004)	0.122(0.002)
Waveform	0.339(0.056)	0.338(0.006)	0.267(0.058)	0.296(0.012)	0.263(0.026)	0.253(0.012)
Letter	0.181(0.025)	0.155(0.007)	0.155(0.006)	0.141(0.021)	0.120(0.023)	0.185(0.012)
Isolet	0.141(0.004)	0.138(0.001)	0.138(0.001)	0.106(0.025)	0.140(0.017)	0.148(0.001)
Gisette	0.595(0.015)	0.532(0.006)	0.515(0.022)	0.541(0.002)	0.499(0.004)	0.545(0.001)
p53-Mutants	0.464(0.021)	0.411(0.020)	0.411(0.002)	0.397(0.001)	0.412(0.011)	0.418(0.001)
Amazon	0.441(0.019)	0.388(0.006)	0.388(0.019)	0.446(0.067)	0.437(0.011)	0.442(0.022)
Arcene	0.367(0.012)	0.142(0.002)	0.142(0.001)	0.153(0.007)	0.148(0.016)	0.165(0.002)
Shapes	0.243(0.009)	0.294(0.007)	0.297(0.005)	0.208(0.031)	0.281(0.025)	0.314(0.006)
Tracedata	0.438(0.010)	0.432(0.012)	0.432(0.001)	0.437(0.019)	0.491(0.031)	0.470(0.014)
ControlChart	0.092(0.013)	0.203(0.020)	0.195(0.008)	0.250(0.021)	0.063(0.008)	0.122(0.007)
Twopat	0.144(0.025)	0.070(0.002)	0.070(0.001)	0.071(0.005)	0.070(0.007)	0.073(0.002)
N30	0.098(0.005)	0.108(0.003)	0.107(0.005)	0.091(0.013)	0.110(0.015)	0.124(0.003)
D75	0.033(0.002)	0.038(0.001)	0.038(0.001)	0.032(0.003)	0.038(0.005)	0.039(0.002)
S2500	0.116(0.004)	0.122(0.005)	0.121(0.007)	0.098(0.010)	0.131(0.009)	0.127(0.004)
min	0.033	0.038	0.038	0.032	0.038	0.039
max	0.649	0.588	0.573	0.543	0.640	0.673
avg	0.295	0.280	0.274	0.270	0.275	0.295

5.1.4 Parameter setting

To set the parameters α and β of the proposed E-CB-PCE, we performed a *leave-one-dataset-out* approach: for each dataset the performance of E-CB-PCE on the other datasets was assessed for different values of the parameter(s), and the value(s) that achieved the maximum $F1_{of}$ was then used to obtain a projective clustering solution for the left-out dataset. In general, we observed that the settings were scarcely influenced by any specific dataset, which indicates that a relatively easy setup can be performed on new datasets for which a reference classification or other a-priori knowledge is not available. Particularly, the best-performance scores were mostly reached by setting both α and β to 2. For the competing methods, we set the parameters as suggested in their respective papers.

5.2 Results

Accuracy Tables 3, 4, 5 give the results of the external evaluation w.r.t. the reference classification (assessment criteria $F1_{of}$, $F1_o$, and $F1_f$, respectively), and Tables 6, 7, 8 give

Table 4 Evaluation w.r.t. the reference classification ($F1_o$)

Dataset	MOEA-PCE	EM-PCE	E-EM-PCE	E-2S-PCE	CB-PCE	E-CB-PCE
Iris	0.967(0.026)	0.880(0.002)	0.784(0.076)	0.733(0.158)	0.948(0.137)	1(0.027)
Wine	0.835(0.037)	0.731(0.009)	0.735(0.001)	0.639(0.152)	0.855(0.082)	0.957(0.003)
Glass	0.474(0.053)	0.509(0.012)	0.537(0.021)	0.433(0.052)	0.507(0.068)	0.571(0.038)
Ecoli	0.760(0.014)	0.667(0.021)	0.665(0.021)	0.641(0.127)	0.789(0.063)	0.761(0.007)
Yeast	0.417(0.007)	0.333(0.007)	0.328(0.014)	0.298(0.030)	0.424(0.034)	0.395(0.014)
Mult.-Feat.	0.319(0.069)	0.369(0.020)	0.363(0.053)	0.493(0.063)	0.326(0.035)	0.269(0.007)
Segmentation	0.443(0.062)	0.568(0.014)	0.567(0.007)	0.554(0.029)	0.523(0.082)	0.659(0.047)
Abalone	0.208(0.006)	0.169(0.005)	0.164(0.006)	0.152(0.014)	0.198(0.014)	0.210(0.004)
Waveform	0.515(0.070)	0.415(0.001)	0.387(0.056)	0.382(0.018)	0.437(0.027)	0.437(0.009)
Letter	0.331(0.030)	0.306(0.005)	0.308(0.002)	0.299(0.037)	0.307(0.031)	0.305(0.017)
Isolet	0.959(0.031)	0.978(0.001)	0.978(0.001)	0.722(0.208)	0.977(0.080)	1(0.002)
Gisette	0.728(0.015)	0.674(0.007)	0.658(0.030)	0.581(0.140)	0.667(0.003)	0.669(0.001)
p53-Mutants	0.728(0.038)	0.619(0.028)	0.642(0.031)	0.596(0.002)	0.650(0.054)	0.733(0.001)
Amazon	0.555(0.036)	0.488(0.009)	0.480(0.047)	0.553(0.071)	0.525(0.011)	0.518(0.029)
Arcene	0.705(0.009)	0.626(0.001)	0.614(0.001)	0.618(0.071)	0.647(0.021)	0.644(0.019)
Shapes	0.681(0.017)	0.693(0.015)	0.704(0.013)	0.531(0.052)	0.716(0.041)	0.786(0.017)
Tracedata	0.614(0.002)	0.628(0.059)	0.628(0.001)	0.619(0.032)	0.629(0.029)	0.617(0.010)
ControlChart	0.319(0.020)	0.332(0.003)	0.342(0.026)	0.345(0.040)	0.311(0.014)	0.321(0.019)
Twopat	0.355(0.011)	0.296(0.002)	0.296(0.002)	0.302(0.021)	0.313(0.007)	0.316(0.001)
N30	0.807(0.219)	0.884(0.013)	0.881(0.017)	0.752(0.117)	0.915(0.096)	0.997(0.028)
D75	0.857(0.146)	0.952(0.018)	0.955(0.022)	0.717(0.118)	0.931(0.103)	0.981(0.046)
S2500	0.880(0.156)	0.895(0.031)	0.889(0.056)	0.747(0.093)	0.983(0.070)	0.941(0.014)
min	0.208	0.169	0.164	0.152	0.198	0.210
max	0.967	0.978	0.978	0.752	0.983	1
avg	0.612	0.591	0.587	0.532	0.617	0.640

the results of the internal evaluation w.r.t. the projective ensemble solutions (assessment criteria $\overline{F1}_{of}$, $\overline{F1}_o$, and $\overline{F1}_f$, respectively). All algorithms involved in the comparison are nondeterministic, thus all tables contain average results over 50 different runs along with the corresponding standard deviations (in parentheses). Moreover, to improve the readability of the results, for each competitor and assessment criterion, we summarize the average gain of E-CB-PCE w.r.t. the competing method in Table 9.

As observed in the summary reported in Table 9, E-CB-PCE achieved better average accuracy performance w.r.t. both MOEA-PCE and CB-PCE, thus showing that the most recent findings of this work incorporated into the proposed E-CB-PCE actually give the expected outcome. E-CB-PCE was more accurate than both MOEA-PCE and CB-PCE on 16 out of 22 datasets on average, while achieving average gains up to 0.147 ($\overline{F1}_{of}$ assessment criterion) and 0.078 ($F1_f$ assessment criterion) w.r.t. MOEA-PCE and CB-PCE, respectively.

Larger improvements were produced by E-CB-PCE w.r.t. the early single-objective PCE methods, i.e., EM-PCE, E-EM-PCE, and E-2S-PCE. The average gains achieved by E-CB-PCE w.r.t. EM-PCE, E-EM-PCE, and E-2S-PCE reported in Table 9 were in general larger than those observed w.r.t. the remaining competing methods MOEA-PCE and CB-PCE.

Table 5 Evaluation w.r.t. the reference classification ($F1_f$)

Dataset	MOEA-PCE	EM-PCE	E-EM-PCE	E-2S-PCE	CB-PCE	E-CB-PCE
Iris	0.974(0.019)	0.667(0.001)	0.667(0.001)	0.667(0.001)	0.829(0.044)	0.807(0.001)
Wine	0.643(0.051)	0.426(0.006)	0.432(0.022)	0.437(0.017)	0.472(0.058)	0.476(0.091)
Glass	0.804(0.023)	0.662(0.029)	0.671(0.037)	0.662(0.028)	0.748(0.086)	0.734(0.032)
Ecoli	0.906(0.021)	0.970(0.013)	0.983(0.027)	1(0.001)	0.880(0.035)	1(0.047)
Yeast	0.846(0.011)	0.774(0.010)	0.772(0.018)	0.726(0.025)	0.827(0.026)	0.956(0.018)
Multi.-Feat.	0.768(0.013)	0.795(0.001)	0.802(0.027)	0.887(0.003)	0.502(0.034)	0.567(0.026)
Segmentation	0.861(0.026)	0.747(0.029)	0.737(0.047)	0.706(0.001)	0.745(0.115)	0.971(0.056)
Abalone	0.822(0.023)	0.716(0.006)	0.713(0.007)	0.670(0.001)	0.750(0.008)	0.754(0.006)
Waveform	0.660(0.067)	0.792(0.001)	0.792(0.001)	0.792(0.001)	0.722(0.043)	0.700(0.036)
Letter	0.643(0.037)	0.595(0.008)	0.598(0.009)	0.580(0.001)	0.593(0.031)	0.686(0.027)
Isolet	0.171(0.004)	0.143(0.001)	0.144(0.005)	0.142(0.005)	0.149(0.005)	0.150(0.005)
Gisette	0.876(0.016)	0.797(0.001)	0.794(0.005)	0.797(0.001)	0.739(0.041)	0.818(0.001)
p53-Mutants	0.721(0.009)	0.671(0.001)	0.671(0.001)	0.673(0.001)	0.673(0.015)	0.692(0.001)
Amazon	0.890(0.005)	0.828(0.001)	0.828(0.001)	0.829(0.001)	0.840(0.013)	0.849(0.003)
Arcene	0.536(0.014)	0.264(0.003)	0.262(0.001)	0.260(0.001)	0.268(0.013)	0.483(0.010)
Shapes	0.428(0.011)	0.448(0.002)	0.449(0.001)	0.405(0.016)	0.451(0.018)	0.444(0.005)
Tracedata	0.787(0.010)	0.800(0.006)	0.811(0.015)	0.796(0.029)	0.795(0.027)	0.894(0.034)
ControlChart	0.322(0.044)	0.673(0.002)	0.673(0.003)	0.673(0.003)	0.263(0.027)	0.473(0.035)
Twopat	0.451(0.079)	0.233(0.001)	0.233(0.001)	0.233(0.001)	0.281(0.039)	0.339(0.021)
N30	0.131(0.002)	0.119(0.001)	0.120(0.002)	0.113(0.004)	0.119(0.003)	0.338(0.005)
D75	0.041(0.001)	0.039(0.001)	0.039(0.001)	0.036(0.001)	0.040(0.003)	0.045(0.001)
S2500	0.141(0.002)	0.124(0.002)	0.127(0.003)	0.125(0.005)	0.130(0.003)	0.343(0.001)
min	0.041	0.039	0.039	0.036	0.040	0.045
max	0.974	0.970	0.983	1	0.880	1
avg	0.610	0.558	0.560	0.555	0.537	0.615

Looking at the standard deviations reported in Tables 3–8, it can be observed that the proposed E-CB-PCE was quite insensitive to its random component. The standard deviations were in the order of 10^{-3} in most cases, while being in the order of 10^{-2} in the remaining cases.

Efficiency Table 10 shows the runtimes (in milliseconds) of the various algorithms involved in the comparison. As expected, the proposed E-CB-PCE was slower than EM-PCE on most datasets, while clearly outperforming MOEA-PCE. The runtimes of E-CB-PCE were one or two orders of magnitude smaller than those of MOEA-PCE on average, up to four orders on Isolet. Only on one dataset, MOEA-PCE was more efficient than E-CB-PCE (Glass), even though the runtimes of the two methods remained of the same order of magnitude.

Compared to CB-PCE, the proposed E-CB-PCE was faster on 11 datasets, resulting in one order faster on 3 of them (i.e., Multiple-Features, Waveform, Amazon), and two orders faster on 4 datasets (i.e., Isolet, Gisette, p53-Mutants, Arcene). In general, we observed that the performance of E-CB-PCE mainly depended on the number of iterations needed for

Table 6 Evaluation w.r.t. the projective ensemble solutions ($\overline{F1}_{of}$)

Dataset	<i>MOEA-PCE</i>	<i>EM-PCE</i>	<i>E-EM-PCE</i>	<i>E-2S-PCE</i>	<i>CB-PCE</i>	E-CB-PCE
Iris	0.858(0.010)	0.830(0.002)	0.768(0.049)	0.638(0.153)	0.886(0.107)	0.961(0.004)
Wine	0.603(0.008)	0.569(0.002)	0.571(0.001)	0.498(0.085)	0.716(0.035)	0.765(0.001)
Glass	0.430(0.007)	0.437(0.009)	0.413(0.018)	0.438(0.035)	0.557(0.044)	0.613(0.020)
Ecoli	0.630(0.009)	0.625(0.004)	0.623(0.019)	0.496(0.052)	0.819(0.033)	0.846(0.009)
Yeast	0.610(0.008)	0.586(0.017)	0.589(0.024)	0.525(0.046)	0.768(0.017)	0.789(0.007)
Mult.-Feat.	0.221(0.008)	0.233(0.013)	0.221(0.033)	0.180(0.016)	0.237(0.014)	0.200(0.005)
Segmentation	0.269(0.009)	0.308(0.003)	0.308(0.003)	0.309(0.020)	0.320(0.026)	0.326(0.008)
Abalone	0.675(0.011)	0.541(0.011)	0.551(0.008)	0.571(0.034)	0.741(0.022)	0.772(0.013)
Waveform	0.146(0.004)	0.110(0.001)	0.105(0.005)	0.119(0.005)	0.462(0.003)	0.470(0.003)
Letter	0.196(0.005)	0.146(0.002)	0.146(0.002)	0.108(0.009)	0.337(0.011)	0.356(0.006)
Isolet	0.751(0.009)	0.790(0.001)	0.790(0.001)	0.593(0.155)	0.876(0.075)	0.914(0.001)
Gisette	0.597(0.003)	0.505(0.002)	0.483(0.022)	0.481(0.018)	0.622(0.003)	0.669(0.004)
p53-Mutants	0.628(0.005)	0.632(0.001)	0.588(0.060)	0.634(0.001)	0.671(0.014)	0.627(0.036)
Amazon	0.468(0.008)	0.444(0.015)	0.435(0.039)	0.520(0.023)	0.539(0.024)	0.518(0.022)
Arcene	0.589(0.009)	0.421(0.004)	0.419(0.001)	0.397(0.043)	0.800(0.027)	0.836(0.001)
Shapes	0.444(0.009)	0.468(0.009)	0.475(0.009)	0.307(0.045)	0.713(0.026)	0.764(0.002)
Tracedata	0.723(0.008)	0.716(0.001)	0.716(0.001)	0.603(0.079)	0.830(0.074)	0.909(0.008)
ControlChart	0.186(0.005)	0.037(0.002)	0.037(0.001)	0.049(0.001)	0.334(0.015)	0.270(0.006)
Twopat	0.370(0.012)	0.144(0.001)	0.144(0.001)	0.154(0.007)	0.526(0.045)	0.501(0.001)
N30	0.323(0.013)	0.301(0.003)	0.302(0.004)	0.276(0.025)	0.462(0.029)	0.510(0.014)
D75	0.423(0.014)	0.429(0.004)	0.429(0.004)	0.362(0.034)	0.573(0.039)	0.596(0.014)
S2500	0.442(0.013)	0.449(0.010)	0.439(0.020)	0.345(0.036)	0.604(0.015)	0.605(0.004)
min	0.146	0.037	0.037	0.049	0.237	0.200
max	0.858	0.830	0.790	0.638	0.886	0.961
avg	0.481	0.442	0.434	0.391	0.609	0.628

Function 2 to converge. In particular, due to the nature of the local-search moves performed at each iteration, the runtime of Function 2 could negatively compensate the performance gain achieved by E-CB-PCE w.r.t. CB-PCE in the other steps of the heuristic. However, as observed in the measurements reported, this mostly affected the smaller datasets.

6 Conclusion

In our previous work (Gullo et al. 2009) we introduced a framework in which projective clustering and clustering ensembles are addressed simultaneously. This resulted in the formulation of a new problem called projective clustering ensembles (PCE). Since the original formulation, research efforts have been made to design a single-objective function that keeps the object-based and the feature-based cluster representations joined together, and at the same time facilitates the adaptation of a conventional clustering ensemble approach to the PCE problem. In this paper, we presented the latest advance of PCE by proposing

Table 7 Evaluation w.r.t. the projective ensemble solutions ($\overline{F\bar{I}_o}$)

Dataset	<i>MOEA-PCE</i>	<i>EM-PCE</i>	<i>E-EM-PCE</i>	<i>E-2S-PCE</i>	<i>CB-PCE</i>	E-CB-PCE
Iris	0.915(0.007)	0.842(0.002)	0.785(0.044)	0.684(0.142)	0.886(0.093)	0.954(0.004)
Wine	0.723(0.010)	0.662(0.005)	0.667(0.003)	0.575(0.105)	0.716(0.067)	0.766(0.004)
Glass	0.513(0.008)	0.511(0.010)	0.482(0.021)	0.505(0.046)	0.557(0.055)	0.600(0.029)
Ecoli	0.810(0.013)	0.804(0.005)	0.816(0.034)	0.697(0.093)	0.819(0.029)	0.852(0.006)
Yeast	0.754(0.011)	0.681(0.017)	0.682(0.026)	0.609(0.059)	0.768(0.023)	0.790(0.008)
Mult.-Feat.	0.216(0.009)	0.262(0.015)	0.249(0.040)	0.186(0.020)	0.237(0.015)	0.176(0.008)
Segmentation	0.313(0.008)	0.343(0.005)	0.345(0.002)	0.346(0.019)	0.320(0.039)	0.356(0.015)
Abalone	0.749(0.015)	0.592(0.011)	0.601(0.009)	0.628(0.036)	0.741(0.022)	0.771(0.012)
Waveform	0.489(0.008)	0.473(0.001)	0.444(0.028)	0.433(0.041)	0.462(0.008)	0.477(0.004)
Letter	0.324(0.005)	0.309(0.004)	0.308(0.002)	0.274(0.025)	0.337(0.011)	0.364(0.005)
Isolet	0.876(0.010)	0.875(0.001)	0.875(0.001)	0.669(0.176)	0.876(0.064)	0.908(0.001)
Gisette	0.693(0.004)	0.615(0.003)	0.595(0.020)	0.516(0.133)	0.622(0.018)	0.603(0.031)
p53-Mutants	0.720(0.006)	0.679(0.001)	0.611(0.106)	0.677(0.001)	0.671(0.033)	0.649(0.045)
Amazon	0.519(0.008)	0.473(0.017)	0.459(0.049)	0.551(0.026)	0.539(0.033)	0.514(0.030)
Arcene	0.840(0.007)	0.800(0.001)	0.799(0.002)	0.754(0.142)	0.800(0.008)	0.805(0.002)
Shapes	0.692(0.010)	0.701(0.012)	0.708(0.009)	0.547(0.056)	0.713(0.028)	0.762(0.002)
Tracedata	0.829(0.008)	0.830(0.001)	0.831(0.001)	0.732(0.093)	0.830(0.063)	0.890(0.008)
ControlChart	0.347(0.004)	0.307(0.004)	0.311(0.011)	0.342(0.007)	0.334(0.012)	0.367(0.006)
Twopat	0.528(0.006)	0.463(0.002)	0.463(0.001)	0.488(0.030)	0.526(0.047)	0.573(0.001)
N30	0.435(0.013)	0.469(0.004)	0.470(0.006)	0.430(0.039)	0.462(0.042)	0.501(0.009)
D75	0.551(0.015)	0.580(0.005)	0.579(0.006)	0.507(0.043)	0.573(0.039)	0.595(0.016)
S2500	0.562(0.015)	0.575(0.011)	0.567(0.019)	0.504(0.043)	0.604(0.029)	0.594(0.006)
min	0.216	0.262	0.249	0.186	0.237	0.176
max	0.915	0.875	0.875	0.754	0.886	0.954
avg	0.609	0.584	0.575	0.530	0.609	0.630

a metacluster-based formulation and related heuristics, which are theoretically and experimentally proven to best fit the PCE problem.

We expect that alternative approaches to the PCE problem will be developed in the next few years. One general direction we envision is moving the burden of the computation from the clustering side to a proper representation model. In this respect, we argue that an approach worth exploring is tensorial models and related tensor decomposition methods (Kolda and Bader 2009). In a nutshell, a tensor model (e.g., a third-order tensor) can provide an integrated representation of the relevant dimensions in the input ensemble (i.e., the objects, the features, and the clusters); in addition, during the tensor decomposition, a consensus clustering, or even multiple consensus clusterings, can be induced in a straightforward manner. Interestingly, the ability of some existing tensor decomposition methods to generate an unfolding of the tensor in which the correlations among all dimensions (i.e., aspects in the ensemble) are preserved, might play an important role in establishing natural mappings between the clusters in the input ensemble and the clusters of the consensus

Table 8 Evaluation w.r.t. the projective ensemble solutions ($\overline{F1}_f$)

Dataset	<i>MOEA-PCE</i>	<i>EM-PCE</i>	<i>E-EM-PCE</i>	<i>E-2S-PCE</i>	<i>CB-PCE</i>	E-CB-PCE
Iris	0.964(0.008)	0.989(0.001)	0.989(0.001)	0.986(0.002)	0.990(0.004)	0.992(0.001)
Wine	0.861(0.008)	0.856(0.002)	0.857(0.003)	0.852(0.006)	0.910(0.023)	0.905(0.034)
Glass	0.833(0.007)	0.843(0.003)	0.844(0.006)	0.849(0.013)	0.871(0.030)	0.846(0.016)
Ecoli	0.848(0.007)	0.818(0.006)	0.814(0.004)	0.804(0.001)	0.884(0.010)	0.893(0.021)
Yeast	0.870(0.005)	0.890(0.005)	0.888(0.006)	0.907(0.011)	0.909(0.018)	0.912(0.010)
Mult.-Feat.	0.656(0.008)	0.662(0.001)	0.665(0.014)	0.697(0.004)	0.715(0.046)	0.695(0.010)
Segmentation	0.854(0.014)	0.815(0.005)	0.812(0.013)	0.824(0.001)	0.942(0.071)	0.935(0.023)
Abalone	0.944(0.004)	0.976(0.002)	0.979(0.004)	0.994(0.001)	0.977(0.010)	0.984(0.007)
Waveform	0.266(0.013)	0.279(0.001)	0.279(0.001)	0.279(0.001)	0.323(0.004)	0.341(0.008)
Letter	0.604(0.011)	0.453(0.009)	0.448(0.009)	0.401(0.001)	0.686(0.061)	0.644(0.019)
Isolet	0.838(0.004)	0.891(0.001)	0.891(0.001)	0.861(0.019)	0.897(0.012)	0.903(0.001)
Gisette	0.750(0.005)	0.771(0.001)	0.770(0.002)	0.771(0.001)	0.751(0.010)	0.771(0.010)
p53-Mutants	0.858(0.002)	0.920(0.001)	0.920(0.001)	0.918(0.001)	0.908(0.007)	0.916(0.001)
Amazon	0.874(0.002)	0.945(0.001)	0.945(0.001)	0.943(0.001)	0.932(0.005)	0.939(0.001)
Arcene	0.664(0.010)	0.546(0.006)	0.544(0.003)	0.534(0.006)	0.558(0.029)	0.592(0.001)
Shapes	0.641(0.005)	0.629(0.004)	0.631(0.002)	0.585(0.009)	0.679(0.011)	0.693(0.002)
Tracedata	0.853(0.003)	0.848(0.003)	0.849(0.001)	0.820(0.011)	0.894(0.028)	0.930(0.007)
ControlChart	0.510(0.022)	0.154(0.001)	0.154(0.001)	0.154(0.001)	0.556(0.038)	0.396(0.022)
Twopat	0.685(0.024)	0.342(0.001)	0.343(0.005)	0.343(0.001)	0.599(0.105)	0.472(0.073)
N30	0.810(0.005)	0.670(0.010)	0.668(0.019)	0.676(0.015)	0.842(0.055)	0.808(0.052)
D75	0.799(0.008)	0.741(0.006)	0.743(0.004)	0.717(0.017)	0.832(0.040)	0.832(0.009)
S2500	0.803(0.007)	0.772(0.009)	0.770(0.028)	0.673(0.028)	0.818(0.049)	0.864(0.012)
min	0.266	0.154	0.154	0.154	0.323	0.341
max	0.964	0.989	0.989	0.994	0.990	0.992
avg	0.763	0.719	0.718	0.709	0.794	0.785

Table 9 Average gains of E-CB-PCE w.r.t. the competing methods

Criterion	<i>MOEA-PCE</i>	<i>EM-PCE</i>	<i>E-EM-PCE</i>	<i>E-2S-PCE</i>	<i>CB-PCE</i>
$F1_{of}$	0	0.014	0.021	0.025	0.020
$F1_o$	0.029	0.049	0.054	0.108	0.023
$F1_f$	0.005	0.056	0.055	0.060	0.078
$\overline{F1}_{of}$	0.147	0.186	0.194	0.237	0.019
$\overline{F1}_o$	0.021	0.046	0.055	0.100	0.021
$\overline{F1}_f$	0.022	0.066	0.067	0.076	-0.010
avg	0.037	0.070	0.074	0.101	0.025

clustering. Moreover, the common problem due to the unavailability of feature relevance values could be relaxed in a tensor modeling, thus enabling a new generation of clustering ensemble methods.

Table 10 Execution times (milliseconds)

Dataset	MOEA-PCE	EM-PCE	E-EM-PCE	E-2S-PCE	CB-PCE	E-CB-PCE
Iris	2,056	37	109	253	74	1,492
Wine	2,558	29	88	163	144	1,223
Glass	7,712	56	615	248	500	9,177
Ecoli	14,401	59	685	625	1,147	9,337
Yeast	227,878	757	20,438	21,560	30,384	83,008
Mult.-Feat.	490,602	116,334	151,582	88,713	1,562,139	114,368
Segmentation	233,951	1,854	17,385	40,014	25,692	31,780
Abalone	3,411,116	4,105	156,959	430,974	275,053	857,164
Waveform	125,247	4,912	15,905	91,844	10,487	2,179
Letter	2,248,695	9,591	126,222	772,883	70,458	135,832
Isolet	20,676,754	10,100	10,000	8,488	66,136	1,447
Gisette	966,108	34,260	38,216	29,700	93,148	1,450
p53-Mutants	58,695	22,209	22,168	19,347	65,501	1,490
Amazon	395,988	21,556	21,619	20,914	135,446	12,737
Arcene	120,537	21,557	21,499	17,405	81,433	2,413
Shapes	211,654	17,752	19,152	12,473	282,857	106,180
Tracedata	12,777	1,062	1,108	960	9,000	4,716
ControlChart	50,798	1,522	5,397	2,801	13,900	20,708
Twopat	31,850	2,946	5,706	4,606	9,788	5,344
N30	164,969	1,340	13,781	16,243	22,904	44,517
D75	135,297	4,558	13,414	15,477	32,938	30,631
S2500	290,408	2,717	29,223	44,008	39,039	55,607

Appendix: Proofs

A.1 Proofs of Sect. 2

Proposition 1 For any two projective clusters C, C' it holds that $\mathbf{X}_C = \mathbf{X}_{C'}$ if and only if $\Gamma_C = \Gamma_{C'}$ and $\Delta_C = \Delta_{C'}$.

Proof Let us prove both the directions of the implication. Proving $\mathbf{X}_C = \mathbf{X}_{C'} \Leftrightarrow \Gamma_C = \Gamma_{C'} \wedge \Delta_C = \Delta_{C'}$ is immediate: by definition, $\mathbf{X}_C = \Gamma_C^T \Delta_C$ and $\mathbf{X}_{C'} = \Gamma_{C'}^T \Delta_{C'}$. Regarding $\mathbf{X}_C = \mathbf{X}_{C'} \Rightarrow \Gamma_C = \Gamma_{C'} \wedge \Delta_C = \Delta_{C'}$, we first note that:

$$\begin{aligned} \mathbf{X}_C = \mathbf{X}_{C'} &\Rightarrow X_C(i, j) = X_{C'}(i, j), \quad \forall i, \forall j \\ &\Rightarrow \Gamma_{C, o_i} \times \Delta_{C, j} = \Gamma_{C', o_i} \times \Delta_{C', j}, \quad \forall i, \forall j. \end{aligned} \tag{28}$$

By summing the statements in (28) over j , we obtain:

$$\begin{aligned} \Gamma_{C, o_i} \times \Delta_{C, j} &= \Gamma_{C', o_i} \times \Delta_{C', j}, \quad \forall i, \forall j \\ \Rightarrow \sum_{j=1}^{|\mathcal{F}|} \Gamma_{C, o_i} \times \Delta_{C, j} &= \sum_{j=1}^{|\mathcal{F}|} \Gamma_{C', o_i} \times \Delta_{C', j}, \quad \forall i \end{aligned}$$

$$\Rightarrow \Gamma_{C, \mathbf{o}_i} \sum_{j=1}^{|\mathcal{F}|} \Delta_{C,j} = \Gamma_{C', \mathbf{o}_i} \sum_{j=1}^{|\mathcal{F}|} \Delta_{C',j}, \quad \forall i,$$

which implies $\Gamma_{C, \mathbf{o}_i} = \Gamma_{C', \mathbf{o}_i}, \forall i$, since $\sum_{j=1}^{|\mathcal{F}|} \Delta_{C,j} = \sum_{j=1}^{|\mathcal{F}|} \Delta_{C',j} = 1$ by definition. Similarly, by summing the statements in (28) over i leads to:

$$\begin{aligned} \Gamma_{C, \mathbf{o}_i} \times \Delta_{C,j} &= \Gamma_{C', \mathbf{o}_i} \times \Delta_{C',j}, \quad \forall i, \forall j \\ \Rightarrow \sum_{i=1}^{|\mathcal{D}|} \Gamma_{C, \mathbf{o}_i} \times \Delta_{C,j} &= \sum_{i=1}^{|\mathcal{D}|} \Gamma_{C', \mathbf{o}_i} \times \Delta_{C',j}, \quad \forall j \\ \Rightarrow \Delta_{C,j} \sum_{i=1}^{|\mathcal{D}|} \Gamma_{C, \mathbf{o}_i} &= \Delta_{C',j} \sum_{i=1}^{|\mathcal{D}|} \Gamma_{C', \mathbf{o}_i}, \quad \forall j \\ \Rightarrow \Delta_{C,j} &= \Delta_{C',j}, \quad \forall j, \end{aligned}$$

where the last derivation holds since the result $\Gamma_{C, \mathbf{o}_i} = \Gamma_{C', \mathbf{o}_i}, \forall i$, shown above clearly implies $\sum_{i=1}^{|\mathcal{D}|} \Gamma_{C, \mathbf{o}_i} = \sum_{i=1}^{|\mathcal{D}|} \Gamma_{C', \mathbf{o}_i}$. □

Proposition 2 For any projective cluster C , its object-based representation $\Gamma_C = (\Gamma_{C, \mathbf{o}_1}, \dots, \Gamma_{C, \mathbf{o}_{|\mathcal{D}|}})$ and feature-based representation $\Delta_C = (\Delta_{C,1}, \dots, \Delta_{C,|\mathcal{F}|})$ can uniquely be derived from its projective matrix \mathbf{X}_C as follows:

$$\Gamma_{C, \mathbf{o}_i} = \sum_{j=1}^{|\mathcal{F}|} X_C(i, j), \quad \Delta_{C,j} = \frac{X_C(1, j)}{\sum_{j=1}^{|\mathcal{F}|} X_C(1, j)} = \dots = \frac{X_C(|\mathcal{D}|, j)}{\sum_{j=1}^{|\mathcal{F}|} X_C(|\mathcal{D}|, j)},$$

where $X_C(i, j)$ denotes the element (i, j) of the matrix \mathbf{X}_C .

Proof The first statement holds since:

$$\sum_{j=1}^{|\mathcal{F}|} X_C(i, j) = \sum_{j=1}^{|\mathcal{F}|} \Gamma_{C, \mathbf{o}_i} \times \Delta_{C,j} = \Gamma_{C, \mathbf{o}_i} \sum_{j=1}^{|\mathcal{F}|} \Delta_{C,j} = \Gamma_{C, \mathbf{o}_i}.$$

Once derived the expression for Γ_{C, \mathbf{o}_i} , the second statement is immediate, because each element $X_C(i, j)$ is by definition equal to $\Gamma_{C, \mathbf{o}_i} \times \Delta_{C,j}$. □

A.2 Proofs of Sect. 3

Proposition 3 Given a projective ensemble \mathcal{E} , let Υ denote the set of all clusterings obtained by taking exactly one cluster from each ensemble member, i.e., $\Upsilon = \{C \mid C \subseteq \hat{C} \mid \hat{C} \in \hat{\mathcal{C}} \wedge \hat{C} \in \mathcal{E}\} \wedge |\hat{C} \cap C| = 1, \forall \hat{C} \in \mathcal{E}$. Moreover, let $\mathbf{X}^* = \arg \min_{\hat{X}} \min_{C \in \Upsilon} \sum_{C \in \mathcal{C}} T(\hat{X}, \mathbf{X}_C)$ s.t. \hat{X} satisfies (2). Given an integer $K \geq 1$, it holds that the optimal projective consensus clustering for the optimization problem defined in (17)–(18) is $C^* = \{C_1^*, \dots, C_K^*\}$ s.t. $\mathbf{X}_{C_1^*} = \dots = \mathbf{X}_{C_K^*} = \mathbf{X}^*$.

Proof The constraints in (18) imply that each cluster C of the optimal projective consensus clustering should be associated with exactly one cluster \hat{C} of each ensemble solution, but no limitations are imposed about associations of \hat{C} to multiple C , which are therefore allowed.

Within this view, the set \mathcal{Y} identifies a sound and complete searching space for the mappings $x(\hat{C}, C)$. This implies that the matrix \mathbf{X}^* , which is the matrix that minimizes the sum of the Tanimoto distances between itself and all clusters within a set of clusters \mathbf{C} , over all matrices satisfying (2) and all $\mathbf{C} \in \mathcal{Y}$, would represent an optimal solution for the case $K = 1$. The generalization to the case $K > 1$ is straightforward, as $\mathcal{C}^* = \{C_1^*, \dots, C_K^*\}$ s.t. $\mathbf{X}_{C_1^*} = \dots = \mathbf{X}_{C_K^*} = \mathbf{X}^*$ represents a feasible solution to the problem and any other solution where $\mathbf{X}_{C_i^*} \neq \mathbf{X}^*$ for some $i \in [1..K]$ would have a non-smaller objective function value due to the optimality of \mathbf{X}^* . \square

A.3 Proofs of Sect. 4

Lemma 2 Given an m -dimensional, binary vector \mathbf{u} , let \mathbf{u}' denote the vector $\frac{1}{n(\mathbf{u})}\mathbf{u}$, where $n(\mathbf{u}) = \sum_{j=1}^m u_j$. For any two m -dimensional binary vectors \mathbf{u} and \mathbf{v} , it holds that $J(\mathbf{u}', \mathbf{v}') = J(\mathbf{u}, \mathbf{v})$.

Proof

$$\begin{aligned} J(\mathbf{u}', \mathbf{v}') &= \frac{\sum_{j=1}^m u'_j v'_j}{\sum_{j=1}^m (u'_j)^2 + \sum_{j=1}^m (v'_j)^2 - \sum_{j=1}^m u'_j v'_j} \\ &= \frac{\sum_{j=1}^m \frac{u_j}{n(\mathbf{u})} \frac{v_j}{n(\mathbf{v})}}{\sum_{j=1}^m \left(\frac{u_j}{n(\mathbf{u})}\right)^2 + \sum_{j=1}^m \left(\frac{v_j}{n(\mathbf{v})}\right)^2 - \sum_{j=1}^m \frac{u_j}{n(\mathbf{u})} \frac{v_j}{n(\mathbf{v})}} \\ &= \frac{\frac{1}{n(\mathbf{u})n(\mathbf{v})} \sum_{j=1}^m u_j v_j}{\frac{1}{n(\mathbf{u})^2} \sum_{j=1}^m u_j^2 + \frac{1}{n(\mathbf{v})^2} \sum_{j=1}^m v_j^2 - \frac{1}{n(\mathbf{u})n(\mathbf{v})} \sum_{j=1}^m u_j v_j}. \end{aligned}$$

Since \mathbf{u} and \mathbf{v} are binary vectors, it holds that $\sum_{j=1}^m u_j^2 = \sum_{j=1}^m u_j = n(\mathbf{u})$ and $\sum_{j=1}^m v_j^2 = \sum_{j=1}^m v_j = n(\mathbf{v})$; thus, it results that:

$$\begin{aligned} J(\mathbf{u}', \mathbf{v}') &= \frac{\frac{1}{n(\mathbf{u})n(\mathbf{v})} \sum_{j=1}^m u_j v_j}{\frac{1}{n(\mathbf{u})^2} \sum_{j=1}^m u_j^2 + \frac{1}{n(\mathbf{v})^2} \sum_{j=1}^m v_j^2 - \frac{1}{n(\mathbf{u})n(\mathbf{v})} \sum_{j=1}^m u_j v_j} \\ &= \frac{\frac{1}{n(\mathbf{u})n(\mathbf{v})} \sum_{j=1}^m u_j v_j}{\frac{1}{n(\mathbf{u})} + \frac{1}{n(\mathbf{v})} - \frac{1}{n(\mathbf{u})n(\mathbf{v})} \sum_{j=1}^m u_j v_j} \\ &= \frac{\frac{1}{n(\mathbf{u})n(\mathbf{v})} \sum_{j=1}^m u_j v_j}{\frac{1}{n(\mathbf{u})n(\mathbf{v})} (n(\mathbf{u}) + n(\mathbf{v}) - \sum_{j=1}^m u_j v_j)} \\ &= \frac{\sum_{j=1}^m u_j v_j}{\sum_{j=1}^m u_j^2 + \sum_{j=1}^m v_j^2 - \sum_{j=1}^m u_j v_j} = J(\mathbf{u}, \mathbf{v}). \end{aligned}$$

\square

Lemma 1 Let CB-PCE RESTRICTED be a special version of the CB-PCE ENHANCED problem where (i) $|\mathcal{D}| = 1$, (ii) $K = 1$ (K denotes the number of clusters in the output projective consensus clustering), (iii) $|\hat{\mathcal{C}}| = 1$, $\forall \hat{C} \in \mathcal{E}$, (iv) $n(\Delta_{\hat{C}}) = n(\Delta_{\hat{C}'}) = n_{\mathcal{E}}$, $\forall \hat{C}, \hat{C}'$, where $n(\Delta_{\hat{C}}) = \sum_{f \in \mathcal{F}} \mathbb{I}[\Delta_{\hat{C}, f} > 0]$, and (v) $\Delta_{\hat{C}, f} = \frac{1}{n_{\mathcal{E}}}$, $\forall f \in \mathcal{F}, \forall \hat{C} \in \hat{\mathcal{C}}, \forall \hat{C}' \in \mathcal{E}$. CB-PCE RESTRICTED is NP-hard.

Proof First, it is easy to note that, due to the restrictions w.r.t. the general CB-PCE ENHANCED, the projective matrix of each cluster \hat{C} within a valid input ensemble for CB-PCE RESTRICTED corresponds to its (mono-dimensional) feature-based representation $\mathbf{A}_{\hat{C}}$. To show that CB-PCE RESTRICTED is NP-hard, we define a reduction from the JACCARD MEDIAN problem, which has been proven to be NP-hard in Chierichetti et al. (2010). JACCARD MEDIAN is defined as follows: given a set U of m -dimensional, binary vectors, find a (possibly new) binary vector \mathbf{x} such that $\sum_{\mathbf{u} \in U} (1 - J(\mathbf{u}, \mathbf{x}))$ is minimized. The corresponding decision problem is: given a set U of m -dimensional binary vectors and a real value r , check whether a binary vector \mathbf{x} exists such that $\sum_{\mathbf{u} \in U} (1 - J(\mathbf{u}, \mathbf{x})) \leq r$. Given a set U of vectors in input to JACCARD MEDIAN, we define a valid instance of our problem by building a new set $U' = \{\mathbf{u}' \mid \mathbf{u}' = \frac{1}{n(\mathbf{u})}\mathbf{u}, \mathbf{u} \in U\}$. It is easy to see that U' is a valid input for CB-PCE RESTRICTED, as all vectors in U' represent valid projective matrices as their values sum to one, and this is the only condition to satisfy for an ensemble \mathcal{E} having $|\hat{C}| = 1, \forall \hat{C} \in \mathcal{E}$ and $|\mathcal{D}| = 1$. The key point is to show that, for each r , U has a solution for the decision version of JACCARD MEDIAN if and only if U admits a solution for the decision version of CB-PCE RESTRICTED. Let us prove the first side of the implication, i.e., if U is a *yes* instance for JACCARD MEDIAN then U' is a *yes* instance for CB-PCE RESTRICTED, for all r . To this end, let \mathbf{x} denote the binary vector that makes U a *yes* instance for JACCARD MEDIAN, i.e., the binary vector such that $\sum_{\mathbf{u} \in U} (1 - J(\mathbf{u}, \mathbf{x})) \leq r$. According to Lemma 2, the Jaccard similarities, and therefore the Jaccard distances defined as $1 - J$, are preserved if each input vector \mathbf{u} is scaled up by a factor $n(\mathbf{u})$. This implies that $\sum_{\mathbf{u} \in U} (1 - J(\mathbf{u}, \mathbf{x})) = \sum_{\mathbf{u}' \in U'} (1 - J(\mathbf{u}', \mathbf{x}')) \leq r$. Thus, $\mathbf{x}' = \frac{1}{n(\mathbf{x})}\mathbf{x}$ makes U' a *yes* instance for CB-PCE RESTRICTED (note that, like each vector $\mathbf{u}' \in U'$, \mathbf{x}' satisfies the constraints of CB-PCE RESTRICTED on the output format because it represents a valid projective matrix). The second side of the implication, i.e., if U is a *no* instance for JACCARD MEDIAN then U' is a *no* instance for CB-PCE RESTRICTED, for all r , can be proved similarly. Indeed, according to Lemma 2 again, if no binary vector is within a distance r from the vectors in U , then no vector can exist with such a property for U' either. \square

Theorem 1 CB-PCE ENHANCED is NP-hard.

Proof Straightforward, since CB-PCE ENHANCED is a more general version of CB-PCE RESTRICTED, which has been proven to be NP-hard in Lemma 1. \square

Theorem 2 The optimum of the problem P defined in (22)–(24) is $(\forall \mathcal{M}, \forall \mathbf{o})$:

$$\Gamma_{\mathcal{M}, \mathbf{o}}^* = \left[\sum_{\mathcal{M}' \in \mathbf{M}} \left(\frac{A_{\mathcal{M}, \mathbf{o}}}{A_{\mathcal{M}', \mathbf{o}}} \right)^{\frac{1}{\alpha-1}} \right]^{-1}.$$

Proof The optimal $\Gamma_{\mathcal{M}, \mathbf{o}}^*$ can be found by means of the conventional Lagrange multipliers method. We first consider the relaxed problem P' of P obtained by temporarily discarding the inequality constraints from the constraint set of P (i.e., the constraints defined in (24)).

We define the new unconstrained objective function Q' for P' as follows:

$$Q' = Q + \sum_{\mathbf{o} \in \mathcal{D}} \lambda_{\mathbf{o}} \left(\sum_{\mathcal{M}' \in \mathbf{M}} \Gamma_{\mathcal{M}', \mathbf{o}} - 1 \right). \tag{29}$$

The optimal $\Gamma_{\mathcal{M},\mathbf{o}}^*$ is computed by first retrieving the stationary points of Q' , i.e., the points for which

$$\nabla Q' = \left(\frac{\partial Q'}{\partial \Gamma_{\mathcal{M},\mathbf{o}}}, \frac{\partial Q'}{\partial \lambda_{\mathbf{o}}} \right) = 0.$$

Thus, we solve the following system of equations:

$$\frac{\partial Q'}{\partial \Gamma_{\mathcal{M},\mathbf{o}}} = \alpha A_{\mathcal{M},\mathbf{o}} (\Gamma_{\mathcal{M},\mathbf{o}})^{\alpha-1} + \lambda_{\mathbf{o}} = 0, \tag{30}$$

$$\frac{\partial Q'}{\partial \lambda_{\mathbf{o}}} = \sum_{\mathcal{M}' \in \mathbf{M}} \Gamma_{\mathcal{M}',\mathbf{o}} - 1 = 0. \tag{31}$$

By solving (30) w.r.t. $\Gamma_{\mathcal{M},\mathbf{o}}$ and substituting the solution in (31), we obtain:

$$\sum_{\mathcal{M}' \in \mathbf{M}} \left(\frac{-\lambda_{\mathbf{o}}}{\alpha A_{\mathcal{M}',\mathbf{o}}} \right)^{\frac{1}{\alpha-1}} = 1. \tag{32}$$

By solving (32) w.r.t. $\lambda_{\mathbf{o}}$ and substituting such a solution in (30), we obtain:

$$\alpha A_{\mathcal{M},\mathbf{o}} (\Gamma_{\mathcal{M},\mathbf{o}})^{\alpha-1} - \left[\sum_{\mathcal{M}' \in \mathbf{M}} \left(\frac{1}{\alpha A_{\mathcal{M}',\mathbf{o}}} \right)^{\frac{1}{\alpha-1}} \right]^{-(\alpha-1)} = 0. \tag{33}$$

Finally, by solving (33) w.r.t. $\Gamma_{\mathcal{M},\mathbf{o}}$, we obtain a stationary point whose expression is exactly equal to that reported in the claim of the theorem:

$$\Gamma_{\mathcal{M},\mathbf{o}}^* = \left[\sum_{\mathcal{M}' \in \mathbf{M}} \left(\frac{A_{\mathcal{M},\mathbf{o}}}{A_{\mathcal{M}',\mathbf{o}}} \right)^{\frac{1}{\alpha-1}} \right]^{-1}. \tag{34}$$

As it holds that (i) the stationary points of the Lagrangian function Q' are also stationary points of the original objective function Q , (ii) the feasible region of P , and hence the feasible region of P' , is a convex set, and (iii) Q is convex w.r.t. $\Gamma_{\mathcal{M},\mathbf{o}}$, it follows that such a stationary point represents a global minimum of Q , and, accordingly, the optimal solution of P' . Moreover, as $A_{\mathcal{M},\mathbf{o}} \geq 0, \forall \mathcal{M}, \forall \mathbf{o}$, it is trivial to observe that $\Gamma_{\mathcal{M},\mathbf{o}}^* \geq 0, \forall \mathcal{M}, \forall \mathbf{o}$. Therefore, the solution in (34) satisfies the inequality constraints that were temporarily discarded in order to define the relaxed problem P' (cf. (24)); thus, it represents the optimal solution of the original problem P , which proves the theorem. \square

Theorem 3 *The optimum of the problem defined in (25)–(27) is $(\forall \mathcal{M}, \forall f)$:*

$$\Delta_{\mathcal{M},f}^* = \left[\sum_{f' \in \mathcal{F}} \left(\frac{B_{\mathcal{M},f}}{B_{\mathcal{M},f'}} \right)^{\frac{1}{\beta-1}} \right]^{-1}.$$

Proof Analogous to Theorem 2. \square

Proposition 4 *For any two projective clusters C, C' it holds that:*

$$T(\mathbf{X}_C, \mathbf{X}_{C'}) = 1 - \frac{p(C, C')}{q(C) + q(C') - p(C, C')},$$

where

$$p(C, C') = (\mathbf{\Gamma}_C \mathbf{\Gamma}_{C'}^T) \times (\mathbf{\Delta}_C \mathbf{\Delta}_{C'}^T), \quad q(C) = (\|\mathbf{\Gamma}_C\|_2^2) \times (\|\mathbf{\Delta}_C\|_2^2).$$

Proof The Tanimoto distance $T(\mathbf{X}_C, \mathbf{X}_{C'})$ is equal to $1 - \hat{J}(\mathbf{X}, \hat{\mathbf{X}})$. Regarding \hat{J} , it holds that:

$$\begin{aligned} \hat{J}(\mathbf{X}, \hat{\mathbf{X}}) &= \frac{\sum_{i=1}^{|\text{rows}(\mathbf{X})|} \mathbf{X}_i^T \cdot \hat{\mathbf{X}}_i}{\|\mathbf{X}\|_2^2 + \|\hat{\mathbf{X}}\|_2^2 - \sum_{i=1}^{|\text{rows}(\mathbf{X})|} \mathbf{X}_i^T \cdot \hat{\mathbf{X}}_i} \\ &= \left(\sum_{\mathbf{o} \in \mathcal{D}} \sum_{f \in \mathcal{F}} \Gamma_{C,\mathbf{o}} \Delta_{C,f} \Gamma_{C',\mathbf{o}} \Delta_{C',f} \right) \times \left(\sum_{\mathbf{o} \in \mathcal{D}} \sum_{f \in \mathcal{F}} \Gamma_{C,\mathbf{o}} \Delta_{C,f} \right. \\ &\quad \left. + \sum_{\mathbf{o} \in \mathcal{D}} \sum_{f \in \mathcal{F}} \Gamma_{C',\mathbf{o}} \Delta_{C',f} - \sum_{\mathbf{o} \in \mathcal{D}} \sum_{f \in \mathcal{F}} \Gamma_{C,\mathbf{o}} \Delta_{C,f} \Gamma_{C',\mathbf{o}} \Delta_{C',f} \right)^{-1} \\ &= \left(\left(\sum_{\mathbf{o} \in \mathcal{D}} \Gamma_{C,\mathbf{o}} \Gamma_{C',\mathbf{o}} \right) \times \left(\sum_{f \in \mathcal{F}} \Delta_{C,f} \Delta_{C',f} \right) \right) \times \left(\left(\sum_{\mathbf{o} \in \mathcal{D}} \Gamma_{C,\mathbf{o}} \right) \times \left(\sum_{f \in \mathcal{F}} \Delta_{C,f} \right) \right. \\ &\quad \left. + \left(\sum_{\mathbf{o} \in \mathcal{D}} \Gamma_{C',\mathbf{o}} \right) \times \left(\sum_{f \in \mathcal{F}} \Delta_{C',f} \right) - \left(\sum_{\mathbf{o} \in \mathcal{D}} \Gamma_{C,\mathbf{o}} \Gamma_{C',\mathbf{o}} \right) \times \left(\sum_{f \in \mathcal{F}} \Delta_{C,f} \Delta_{C',f} \right) \right)^{-1} \\ &= \frac{(\mathbf{\Gamma}_C \mathbf{\Gamma}_{C'}^T) \times (\mathbf{\Delta}_C \mathbf{\Delta}_{C'}^T)}{(\|\mathbf{\Gamma}_C\|_2^2) \times (\|\mathbf{\Delta}_C\|_2^2) + (\|\mathbf{\Gamma}_{C'}\|_2^2) \times (\|\mathbf{\Delta}_{C'}\|_2^2) - (\mathbf{\Gamma}_C^T \cdot \mathbf{\Gamma}_{C'}) \times (\mathbf{\Gamma}_C \mathbf{\Gamma}_{C'}^T)} \\ &= \frac{p(C, C')}{q(C) + q(C') - p(C, C')}. \end{aligned}$$

□

References

Achtert, E., Böhm, C., Kriegel, H., Kröger, P., Müller-Gorman, I., & Zimek, A. (2006). Finding hierarchies of subspace clusters. In *Proc. European conf. on principles and practice of knowledge discovery in databases (PKDD)* (pp. 446–453).

Achtert, E., Böhm, C., Kriegel, H., Kröger, P., Müller-Gorman, I., & Zimek, A. (2007). Detection and visualization of subspace cluster hierarchies. In *Proc. int. conf. on database systems for advanced applications (DASFAA)* (pp. 152–163).

Aggarwal, C. C., Procopiuc, C. M., Wolf, J. L., Yu, P. S., & Park, J. S. (1999). Fast algorithms for projected clustering. In *Proc. ACM SIGMOD int. conf. on management of data* (pp. 61–72).

Agrawal, R., Gehrke, J., Gunopulos, D., & Raghavan, P. (1998). Automatic subspace clustering of high dimensional data for data mining applications. In *Proc. ACM SIGMOD int. conf. on management of data* (pp. 94–105).

Asuncion, A., & Newman, D. (2010). UCI Machine Learning Repository. <http://archive.ics.uci.edu/ml/>.

Beyer, K. S., Goldstein, J., Ramakrishnan, R., & Shaft, U. (1999). When is “Nearest neighbor” meaningful? In *Proc. int. conf. on database theory (ICDT)* (pp. 217–235).

Böhm, C., Kailing, K., Kriegel, H. P., & Kröger, P. (2004). Density connected clustering with local subspace preferences. In *Proc. IEEE int. conf. on data mining (ICDM)* (pp. 27–34).

Boulis, C., & Ostendorf, M. (2004). Combining multiple clustering systems. In *Proc. European conf. on principles and practice of knowledge discovery in databases (PKDD)* (pp. 63–74).

Bradley, P. S., & Fayyad, U. M. (1998). Refining initial points for K-means clustering. In *Proc. int. conf. on machine learning (ICML)* (pp. 91–99).

- Chen, L., Jiang, Q., & Wang, S. (2008). A probability model for projective clustering on high dimensional data. In *Proc. IEEE int. conf. on data mining (ICDM)* (pp. 755–760).
- Chierichetti, F., Kumar, R., Pandey, S., & Vassilvitskii, S. (2010). Finding the Jaccard median. In *Proc. ACM-SIAM symposium on discrete algorithms (SODA)* (pp. 293–311).
- Domeniconi, C., & Al-Razgan, M. (2009). Weighted cluster ensembles: methods and analysis. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 2(4), 17.
- Domeniconi, C., Gunopulos, D., Ma, S., Yan, B., Al-Razgan, M., & Papadopoulos, D. (2007). Locally adaptive metrics for clustering high dimensional data. *Data Mining and Knowledge Discovery*, 14(1), 63–97.
- Gan, G., Ma, C., & Wu, J. (2007) *Data clustering: theory, algorithms, and applications*. ASA-SIAM series on statistics and applied probability.
- Ghosh, J., Acharya, A.: (2011). Cluster ensembles. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(4), 305–315.
- Gionis, A., Mannila, H., & Tsaparas, P. (2007). Clustering aggregation. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(1), 4.
- Gullo, F., Domeniconi, C., & Tagarelli, A. (2009). Projective clustering ensembles. In *Proc. IEEE int. conf. on data mining (ICDM)* (pp. 794–799).
- Gullo, F., Domeniconi, C., & Tagarelli, A. (2010). Enhancing single-objective projective clustering ensembles. In *Proc. IEEE int. conf. on data mining (ICDM)* (pp. 833–838).
- Gullo, F., Domeniconi, C., & Tagarelli, A. (2011). Advancing data clustering via projective clustering ensembles. In *Proc. ACM SIGMOD int. conf. on management of data* (pp. 733–744).
- Gullo, F., Domeniconi, C., & Tagarelli, A. (2013). Projective clustering ensembles. *Data Mining and Knowledge Discovery (DAMI)*, 26(3), 452–511.
- Hinneburg, A., Aggarwal, C. C., & Keim, D. A. (2000). What is the nearest neighbor in high dimensional spaces? In *Proc. int. conf. on very large data bases (VLDB)* (pp. 506–515).
- Jain, A., & Dubes, R. (1988). *Algorithms for clustering data*. New York: Prentice Hall.
- Keogh, E., Xi, X., Wei, L., & Ratanamahatana, C. A. (2003). The UCR time series classification/clustering page. http://www.cs.ucr.edu/~eamonn/time_series_data/.
- Kolda, T. G., & Bader, B. W. (2009). Tensor decompositions and applications. *SIAM Review*, 51, 455–500.
- Kriegel, H., Kroger, P., Renz, M., & Wurst, S. (2005). A generic framework for efficient subspace clustering of high-dimensional data. In *Proc. IEEE int. conf. on data mining (ICDM)* (pp. 250–257).
- Kriegel, H., Kröger, P., & Zimek, A. (2009). Clustering high-dimensional data: a survey on subspace clustering, pattern-based clustering, and correlation clustering. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 3(1), 1–58.
- Li, R. P., & Mukaidono, M. (1999). Gaussian clustering method based on maximum-fuzzy-entropy interpretation. *Fuzzy Sets and Systems*, 102(2), 253–258.
- Liu, B., Xia, Y., & Yu, P. S. (2000). Clustering through decision tree construction. In *Proc. int. conf. on information and knowledge management (CIKM)* (pp. 20–29).
- Moise, G., Sander, J., & Ester, M. (2008). Robust projected clustering. *Knowledge and Information Systems*, 14(3), 273–298.
- Moise, G., Zimek, A., Kröger, P., Kriegel, H. P., & Sander, J. (2009). Subspace and projected clustering: experimental evaluation and analysis. *Knowledge and Information Systems (KAIS)*, 21(3), 299–326.
- Müller, E., Günnemann, S., Assent, I., & Seidl, T. (2009). Evaluating clustering in subspace projections of high dimensional data. *Proceedings of the VLDB Endowment (PVLDB)*, 2(1), 1270–1281. <http://dme.rwth-aachen.de/en/OpenSubspace/evaluation>.
- Ng, E. K. K., Fu, A. W. C., & Wong, R. C. W. (2005). Projective clustering by histograms. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 17(3), 369–383.
- Nguyen, N., & Caruana, R. (2007). Consensus clustering. In *Proc. IEEE int. conf. on data mining (ICDM)* (pp. 607–612).
- Parsons, L., Haque, E., & Liu, H. (2004). Subspace clustering for high dimensional data: a review. *SIGKDD Explorations*, 6(1), 90–105.
- Procopiu, C. M., Jones, M., Agarwal, P. K., & Murali, T. M. (2002). A Monte Carlo algorithm for fast projective clustering. In *Proc. ACM SIGMOD int. conf. on management of data* (pp. 418–427).
- van Rijsbergen, C. (1979). *Information retrieval*. Stoneham: Butterworths.
- Sequeira, K., & Zaki, M. (2004). SCHISM: a new approach for interesting subspace mining. In *Proc. IEEE int. conf. on data mining (ICDM)* (pp. 186–193).
- Strehl, A., & Ghosh, J. (2002). Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research*, 3, 583–617.
- Strehl, A., Ghosh, J., & Mooney, R. (2000). Impact of similarity measures on web-page clustering. In *Proc. of the AAAI workshop on artificial intelligence for web search* (pp. 58–64).
- Tomasev, N., Radovanovic, M., Mladenic, D., & Ivanovic, M. (2011). The role of hubness in clustering high-dimensional data. In *Proc. Pacific-Asia conf. on advances in knowledge discovery and data mining (PAKDD)* (pp. 183–195).

- Topchy, A. P., Jain, A. K., & Punch, W. F. (2005). Clustering ensembles: models of consensus and weak partitions. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 27(12), 1866–1881.
- Woo, K., Lee, J., Kim, M., & Lee, Y. (2004). FINDIT: a fast and intelligent subspace clustering algorithm using dimension voting. *Information and Software Technology*, 46(4), 255–271.
- Yip, K. Y., Cheung, D. W., & Ng, M. K. H. (2004). A practical projected clustering algorithm. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 16(11), 1387–1397.
- Yip, K. Y., Cheung, D. W., & Ng, M. K. (2005). On discovery of extremely low-dimensional clusters using semi-supervised projected clustering. In *Proc. IEEE int. conf. on data engineering (ICDE)* (pp. 329–340).
- Yiu, M. L., & Mamoulis, N. (2005). Iterative projected clustering by subspace mining. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 17(2), 176–189.