CrossMark

METHODOLOGIES AND APPLICATION

Distributed proximity-based granular clustering: towards a development of global structural relationships in data

Witold Pedrycz · Rami Al-Hmouz · Ali Morfeq · Abdullah Saeed Balamash

Abstract The study is focused on a development of a global structure in a family of distributed data realized on a basis of locally discovered structures. The local structures are revealed by running fuzzy clustering (Fuzzy C-Means), whereas building a global view is realized by forming global proximity matrices on a basis of the local proximity matrices implied by the partition matrices formed for the individual data sets. To capture the diversity of local structures, a global perspective at the structure of the data is captured in terms of a granular proximity matrix, which is built by invoking a principle of justifiable granularity with regard to the aggregation of individual proximity matrices. The three main scenarios are investigated: (a) designing a global structure among the data through building a granular proximity matrix, (b) refining a local structure (expressed in the form of a partition matrix) by engaging structural knowledge conveyed at the higher level of the hierarchy and provided in the form of the granular proximity matrix, (c) forming a consensus-building scheme and updating all local structures with the aid of the proximity dependences available at the upper layer of the hierarchy. While the first scenario delivers a passive approach to the development of the global structure, the two others are of an active nature by facilitating a structural feedback between the local and global level of the hierarchy of the developed structures. The study is illustrated through a series of experiments carried out for synthetic and publicly available data sets.

1 Introduction

Clustering is about revealing a structure in a single data set. Distributed clustering is concerned with the same problem present in situations where there is a family of data sets for which clustering is carried out separately. The term distributed clustering is quite often encountered in the literature. Furthermore, this type of clustering quite often comes with a remarkable variety of terminology, methods, evaluation measures, extensions (Corsini et al. 2005; Pedrycz and Rai 2008; Pedrycz 2005) and applications (Coppi et al. 2010; Graves et al. 2012; Peters 2011). The results of clustering are conveniently interpreted as information granules in the sequel benefiting from a wealth of conceptual developments of Granular Computing (Apolloni et al. 2006; Pedrycz 2013, 2007). Likewise the distributed nature of the data may imply cases when objects (patterns) are described in different feature spaces or all data sets are in the same feature space (Mali and Mitra 2003; Pedrycz and Rai 2008). There are also combinations of these two alternatives (Pedrycz and Rai 2008).

There are two essential aspects that have to be raised with regard to distributed clustering:

W. Pedrycz (✉)
Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6R 2V4, Canada
e-mail: wpedrycz@ualberta.ca

W. Pedrycz · R. Al-Hmouz · A. Morfeq · A.S. Balamash
Department of Electrical and Computer Engineering, Faculty of Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia

W. Pedrycz
Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

(i) *a way of communicating local findings* (viz. the structure of the local data). The locally formed structures of the data (expressed in terms of partition matrices and prototypes) need to be compared. If we are concerned with different patterns, however, all of them are expressed in the same feature space, the structural signatures of the data that can be compared across the data come in the form of the prototypes of the clusters. For the same data (patterns), which are formed in different feature spaces, the local partition matrices are viable constructs using which we can build more abstract constructs (proximity matrices), which in the sequel can initiate some sharing findings across the sets of data. We stress that a direct comparison of partition matrices is not feasible as there are different numbers of clusters and there is no correspondence among the clusters constructed for the individual data sets.

(ii) *a way of building a global view at the data.* There are two variants. Depending whether locally available results are refined or left intact we distinguish between active and passive approaches.

In this study, we are concerned with a category of objective function-based clustering in which the results of clustering come in the form of partition matrices and prototypes (or other numeric representatives). K-Means and Fuzzy C-Means (FCM) (Bezdek 1981) are highly visible alternatives of the clustering methods falling under this category. To facilitate a sound comparison of local structures, one has to look at them from a more general and abstract point of view than the one being conveyed by partition matrices. Proximity matrices (Bezdek 1981; Pedrycz 2005) come here as a viable alternative as they are abstracted from the number of clusters. Their dimensionality is $N \times N$ meaning that all locally formed proximity matrices can be compared (matched) as their dimensionality does not depend explicitly upon the number of clusters. The usage of proximity matrices in clustering problems has been reported in so-called proximity-based clustering, see (Graves et al. 2012; Pedrycz et al. 2004; Pedrycz 2004).

The main objective of this study is to develop a general concept of distributed clustering based on the principles of Granular Computing (Pedrycz et al. 1998) and their constructs (Apolloni et al. 2006; Pedrycz et al. 2004). Here our focus is on the data described in different feature spaces, which implies that a communication vehicle is established in terms of proximity matrices. Based on this form of interaction, we discuss three main conceptual settings. The one is of a passive nature, which concentrates on a granular characterization of proximity-based structure with the invocation of granular proximity matrices of a global character. The two other active-like alternatives invoke some structural feedback to refine local structures on a basis of the global result (viz. a granular proximity matrix).

Our investigations come with several well-articulated aspects of originality. The formulation of the problem is original: although some facets of collaborative clustering have been investigated in the literature, those approaches focus on the passive mode meaning that the results of clustering are aggregated, however, an active facet is not considered at all meaning that no mechanisms adjusting local clustering findings were developed given some global findings. Let us recall that a passive mode implies that the locally available clustering results are provided and some aggregation mechanism is invoked, which gives rise to a general (global) view of the results. In this process, irrespectively of the result obtained at the global level, the local structures (clusters) are not modified (affected). In contrast, when talking about an active mode, a feedback loop is being formed so that in an iterative fashion the local results give rise to some global results. In the sequel, those are contrasted with the results available at the lower level and as a result the local results are modified following a certain adjustment strategy so that in the next iteration there is some improvement observed at the level of the global results.

It is worth noting that there have been some interesting earlier studies on interval-valued clustering, cf. (Souza and Carvalho 2004; Gacek and Pedrycz 2013; Hathaway et al. 1996; Pedrycz et al. 1998; Hwang and Rhee 2007; Mali and Mitra 2003; Wong and Hu 2013; Zhang et al. 2014). There is, however, an important difference between the undertaken research and the previous line of investigation. Here we are concerned with numeric data whereas information granules come as a result of reconciliation of results and are reflective of the diversity of the local findings. The previous studies were focused on granular clustering, more specifically, interval-valued data

The paper is structured as follows. In Sect. 2, we highlight an essence of the problem and identify a role of information granularity being played in this setting. In the sequel, we briefly outline the essence of the main classes of problems (Sect. 3). In Sect. 4, all associated optimization problems are formulated and solved. More specifically, we discuss a way of forming granular proximity matrices through the use of the principle of justifiable granularity (Pedrycz 2013), and look at the techniques of refining local partition matrices based on the gradient-based optimization and particle swarm optimization (PSO) as well as a hybrid of these two techniques. Two ways of characterization of granular proximity matrices are discussed. Numeric studies are covered in Sect. 5.

In the study, we adhere to the standard notation encountered in pattern recognition, clustering and system modeling. To emphasize the origin of the locally available data and the resulting constructs, we use indexes placed in square brackets, say $c[ii]$, $U[ii]$, $u_{ik}[ii]$, etc.

2 The essence of the problem and underlying facet of information granularity

Let us consider a collection of p data sets $\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_p$ originating from a certain problem. For instance, those sets could be data describing a certain system for which formed are individual, local views associated with their local data. The data originated from different collections may be described in different feature spaces $\mathbf{F}_1, \mathbf{F}_2, \ldots, \mathbf{F}_p$. In general, we also assume that some data points are shared among the data sets meaning that an intersection of them is non-empty, namely $\mathbf{D} = \mathbf{D}_1 \cap \mathbf{D}_2 \cap \cdots \cap \mathbf{D}_p$ card $(\mathbf{D}) = N$. Formally speaking, a selected object o_k belonging to the intersection of $\mathbf{D}_1, \mathbf{D}_2, \ldots$, and $\mathbf{D}_p$ comes with its own vectors of features $\mathbf{x}_k[1], \mathbf{x}_k[2], \ldots, \mathbf{x}_k[p]$ defined in the corresponding feature spaces. The data $\{\mathbf{x}_k[ii]\}, k = 1, 2, \ldots N$ forming the data set $\mathbf{D}_{ii}$ are clustered in the corresponding feature space resulting in the corresponding partition matrix $U[ii]$. Clustering is completed for other data sets subsequently giving rise to partition matrices $U[1], U[2], \ldots$, and $U[p]$, respectively. In general, the number of clusters associated with these partition matrices, namely $c[1], c[2], \ldots$, and $c[p]$, could vary from one data set to another. The partition matrices formed in this way exhibit a local character, viz. they are concerned with the findings being confined to the given feature space and produced for the particular locally available data set. Our key objective of this study is to discover (or reconcile) a global structure in the data based on the reconciliation of the local views conveyed through the already constructed partition matrices.

There are two essential and general observations to be made here in the context of the problem under study:

(a) It is apparent that any aggregation of the partition matrices is not feasible because of the fact that the number of clusters could vary from one partition matrix to another. To proceed with any comparison of partition matrices, this process cannot be realized directly but through comparing proximity matrices induced by the corresponding partition matrices.

(b) as the resulting proximity matrices exhibit an evident diversity, we may contemplate to use an aggregation

mechanism that fully reflects and quantifies this diversity. This, in turn, brings a concept of granular proximity matrices as the constructs capturing this facet of the existing variety among the local proximity matrices.

In the study, when developing a structure of a global nature, we rely on the fundamental concept of proximity matrices implied by partition matrices. Recall that for any partition matrix $U[ii] = [u_{ik}[ii]]$, $i = 1, 2, \ldots, c[ii]$, $k = 1, 2, \ldots, N$, the corresponding proximity matrix $P[ii] = [p_{kl}[ii]]$ comes in the following form

$$p_{kl}[ii] = \sum_{i=1}^{c[ii]} \min(u_{ik}[ii], u_{il}[ii]) \tag{1}$$

$k, l = 1, 2, \ldots, N$.

It is instructive to provide a brief example to highlight the essence of the approach, motivate its origin, and shed light at the multistep processing dwelling on a formation of proximity matrices. The five two-dimensional data positioned in several feature subspaces are shown in Fig. 1.

It is apparent that the structures vary from one feature space to another. Assume that the partition matrices $U[1]$, $U[2]$, and $U[3]$ have the following entries which reflect a distribution of the data. Likewise, by visualizing the structures of the data, the number of clusters varies from 2 to 3.

$$U[1] = \begin{bmatrix} 0.9 & 0.8 & 0.3 & 0.1 & 0 \\ 0.1 & 0.2 & 0.7 & 0.9 & 1 \end{bmatrix}$$

$$U[2] = \begin{bmatrix} 1.0 & 0.1 & 0.9 & 0.0 & 0.0 \\ 0.0 & 0.8 & 0.1 & 0.0 & 0.0 \\ 0.0 & 0.1 & 0.0 & 1.0 & 1.0 \end{bmatrix}$$

$$U[3] = \begin{bmatrix} 1.0 & 0.1 & 0.2 & 0.3 & 1 \\ 0.0 & 0.9 & 0.8 & 0.7 & 0 \end{bmatrix}$$

In light of the varying dimensionality of these matrices, they cannot be compared directly. Instead, we have to consider some constructs built at the higher level of generality whose representation does not explicitly involve the number of clusters. Here, the corresponding proximity matrices are formed.
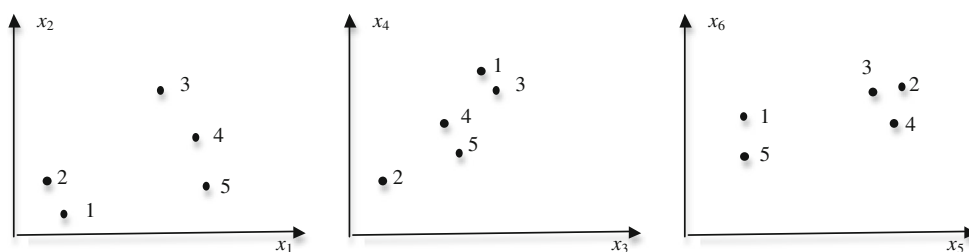

Fig. 1 Example data positioned in three subspaces

These proximity matrices are determined following (1) and come in the form

$$P[1] = \begin{bmatrix} 1.0 & 0.3 & 0.4 & 0.2 & 0.1 \\ & 1.0 & 0.5 & 0.3 & 0.2 \\ & & 1.0 & 0.8 & 0.7 \\ & & & 1.0 & 0.9 \\ & & & & 1.0 \end{bmatrix}$$

$$P[2] = \begin{bmatrix} 1.0 & 0.1 & 0.0 & 0.0 & 0.0 \\ & 1.0 & 0.2 & 0.1 & 0.1 \\ & & 1.0 & 0.0 & 0.0 \\ & & & 1.0 & 0.0 \\ & & & & 1.0 \end{bmatrix}$$

$$P[3] = \begin{bmatrix} 1.0 & 0.1 & 0.2 & 0.3 & 0.0 \\ & 1.0 & 0.9 & 0.8 & 0.1 \\ & & 1.0 & 0.9 & 0.2 \\ & & & 1.0 & 0.3 \\ & & & & 1.0 \end{bmatrix}$$

The proximity matrices exhibit different entries and as a result of their summarization produces the construct at the next level of generality, namely, a granular proximity matrix.

The granular proximity matrix visualizes an emergence of groups of data that are kept close to each other. One can observe a jump in the values of the entries pointing at the occurrence of the well-formed clusters.

3 Main classes of problems

We can distinguish three general categories of problems where the fundamental ideas outlined so far can be fully exploited. The essence of these problems is visualized through a series of figures that help contrast different tasks being studied here.

3.1 Formation of a general description of data

The underlying objective of this problem is to describe the data at the global level by aggregating individual findings conveyed by the locally available proximity matrices, see Fig. 2.

The granular character of the proximity matrix formed at the upper level of hierarchy quantifies the diversity of the local structures. The entries of the granular proximity matrix $G(P) = [p_{kl}^{-}, p_{kl}^{+}], k, l = 1, 2, \ldots, N$, which are intervals with the lower and upper bounds p_{kl}^{-} and p_{kl}^{+}, quantify a strength of linkage occurring between a pair of data, say k and l.
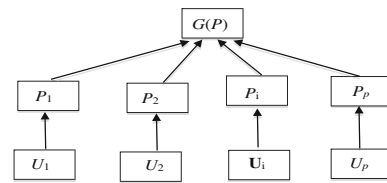
Fig. 2 Building a general description of structure—from local proximity matrices $P_1, P_2, \ldots, P_p$ to granular proximity matrix of global character $G(P)$
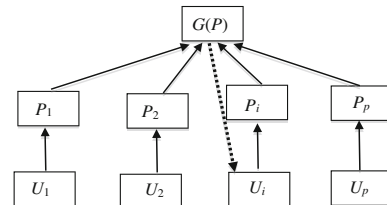
Fig. 3 Updating partition matrix U_i by invoking a feedback loop involving a global proximity matrix at the upper level of the hierarchy
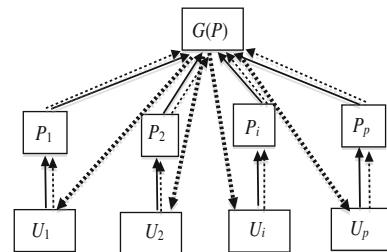
Fig. 4 Building consensus on a basis of a collection of local proximity matrices: observe a feedback loop involving all partition matrices and the global granular proximity matrix

3.2 Refinement of the locally discovered structure of data

In this situation, we are focused on a certain locally available data and attempt to reconcile the findings by augmenting (affecting) the already constructed partition matrix by the additional source of structural knowledge formed at the upper level of the hierarchy, see Fig. 3. In light of the structure of the auxiliary knowledge, this development falls under an umbrella of so-called knowledge-based clustering,

The optimization mechanism is the one discussed in depth in Sect. 4.2.

3.3 Building consensus

In this scenario, we are concerned with a reconciliation of the structural findings obtained locally. The consensus building is an iterative process as illustrated in Fig. 4.

The result formed at the global level is used to adjust the individual partition matrices $U_1, U_2, \ldots, U_p$. The optimization is realized following the scheme outlined in the previous

section. The updated partition matrices are used to build proximity matrices and those give rise to the granular proximity matrix. In turn, the new proximity matrix offers a navigation of optimization of the local partition matrices and the iterations are continued. The convergence of the process becomes critical and for this a suitable index needs to be established.

4 Associated optimization problems, their solutions and characterizations

The three categories of problems outlined in the previous section call for a certain way of formulating the ensuing optimization problem, building its solution and finally characterizing the quality of the obtained solutions. We come up with some design procedures and show their arrangement when solving the three categories of problems formulated above.

4.1 Development of granular proximity matrix

A formulation of a granular proximity matrix is a common task encountered in the three classes of problems discussed above. Given a collection of proximity matrices $P[1], P[2], \ldots, P[p]$ we realize a granular proximity matrix $G(P)$ so that the inherent granularity of this construct is captured and quantified. Proximity matrices exhibit an inherent diversity. Being cognizant of this fact, we can assume that the aggregation result is a granular proximity matrix $G(P)$ whose entries are intervals located in the [0,1] interval. In other words, the granular proximity matrix comes with interval-values entries, $G(P) = [g_{kl}^{-}, g_{kl}^{+}], k, l = 1, 2, \ldots, N$ The interval-valued character of the construct is reflective of the variability in the local findings.

The construction of the granular proximity matrix $G(P)$ realized for individual entries of the matrix is realized by invoking a principle of justifiable granularity (Pedrycz 2013). In a nutshell, this principle states that when aggregating some numeric experimental evidence, in the face of the diversity of the existing pieces of evidence, the result is a certain information granule (instead of another numeric outcome) such that it is supported enough by the experimental data while simultaneously demonstrating sufficient specificity thus coming with a well-conveyed semantics. The principle is applied to the individual entries of the proximity matrices. Let us consider the (k, l)-th entry of the proximity matrices $P[1], P[2], \ldots, P[p]$, namely consider a set $\mathbf{P} = \{p_{kl}[1], p_{kl}[2], \ldots, p_{kl}[p]\}$. We also assume that some initial numeric representative (say mean or median) of $\mathbf{P}$ is provided. For the (k, l)-th entry we denote it by m_{kl}.

In the simplest scenario, the principle of justifiable granularity gives rise to an interval representation $[g_{kl}^{-}, g_{kl}^{+}]$ of the data through the maximization of the coverage of the data (the requirement of sufficient experimental evidence) and the

specificity of the information granule (semantics constraint) where these two fundamental requirements are expressed as follows (below we are concerned with the upper bound of the interval; the determination of the lower bound is realized in the same manner).

experimental evidence

(coverage of data) $\quad f_1(p_{kl}^{+}) = \text{card}\,\{p_{kl}|p_{kl} > m_{kl}\}$ (2)

specificity requirement $\quad f_2(p_{kl}^{+}) = \exp(-\alpha\,|m_{kl} - p_{kl}|), \alpha \geq 0$

It is apparent that these two requirements are in conflict; any improvement of one of them deteriorates the performance of another one. A compromise is set up when the product of the terms $V = f_1 * f_2$ attains its maximal value. In other words, the optimal upper bound of the interval, say $g_{kl}^{+} = \arg \text{Max}_g V(g)$ is achieved.

In the sequel, the optimal bound, b_{opt}, is obtained as a result of the following optimization problem

$$b_{\text{opt}} = \arg \text{Max } V \quad (3)$$

The range of possible values of α requires some clarification. The smallest value of α is equal to 0; in this case, the corresponding optimal bounds of a and b are the lowest and highest values of z_k, $a_{opt} = \arg \min\{z_1, z_2, \ldots, z_N\}$ and $b_{\text{opt}} = \arg \max\{z_1, z_2, \ldots, z_N\}$. The higher the value of α, the more specific the resulting interval. In other words, the maximal value of α is the one, which results in the shortest interval. To come up with the detailed computing, let us consider a subset of the original data $\{z_1, z_2, \ldots, z_N\}$ whose elements are greater than the numeric representative. Furthermore arrange them in an increasing order which yields a set $\{p_1, p_2, \ldots p_M\}$ where $r_1 < r_2 < \ldots$ and $M < N$. Likewise, the associated sequence of the weights is given as $w_1, w_2, \ldots w_M$. The maximal value of α, $\alpha_{\max}$ is then the one, which satisfies all inequalities listed below.

$$\exp(-\alpha|m - r_1|) > (w_1 + w_2) * \exp(-\alpha|m - r_2|)$$
$$w_1 * \exp(-\alpha|m - r_1|) > (w_1 + w_2 + w_3) * \exp(-\alpha|m - r_3|)$$
$$\ldots \quad (4)$$
$$w_1 * \exp(-\alpha|m - r_1|) > (w_1 + w_2 + w_3 + \ldots + w_M)$$
$$* \exp(-\alpha|m - r_M|)$$

The same process is realized for the value of α associated with the lower bound; the result is denoted by $\alpha_{\max}$'. Now we can realize normalization by admitting a unified [0,1] range of values of α which helps us form a series of intervals being formed by a single value of a for their lower and upper bounds. In other words for any value of α it is transformed to its internal value by scaling it to $\alpha * \alpha_{\max}$ and $\alpha * \alpha_{\max}$', respectively. It is worth noting that these intervals indexed by successive values of α are α-cuts of a certain fuzzy set. In other words, here the result of the principle of justifiable granularity becomes a fuzzy set.

This procedure is directly applicable to the construction of granular partition matrix $G(P)$.

The two main steps are envisioned here:

(i) formation of clusters and partition matrices for $\mathbf{D}_1$, $\mathbf{D}_2$, ...$\mathbf{D}_p$, viz. $U[1]$, $U[2]$,..., $U[p]$.
(ii) building proximity matrices $P(U[1])$, $P(U[2])$, ... $P(U[p])$. Note that they are produced for all pairs of the data belonging to $\mathbf{D}$.
(iii) use of the principle of justifiable granularity to form the granular construct $G(P)$. The interval-valued proximity matrix is built for a certain predetermined value of α.

4.2 Refinement of local partition matrix

The crux of this scenario has been captured in Fig. 4. From the optimization perspective, we first form a granular proximity matrix $G(P)$ and afterwards use it in the refinement of some locally constructed partition matrix.

For some given partition matrix $U[ii]$, we proceed with its modifications (adjustments) in such a way that $P(U[ii])$ is "contained" in $G(P)$ to the highest extent. The adjustments are made possible by engaging an idea of optimal allocation of information granularity. The underlying idea is to adjust the entries of $U[ii]$, $ii = 1, 2, \ldots, p$ in such away that the modified partition matrix produces a proximity matrix whose values are included in the interval-valued entries of $G(P)$.

In what follows we elaborate on a detailed algorithm. As the method is the same for any local partition matrix, we omit the index (ii) and use a simplified notation $U = [u_{ik}]$, $i = 1, 2 \ldots c$, $k = 1, 2, ..N$ (as noted earlier we are concerned with the data belonging only to the intersection of the local data, namely $\mathbf{D}$). The entries of $G(P)$ are intervals denoted as $[g_{kl}^-, g_{kl}^+]$, $k, l = 1, 2, \ldots, N$. To express the request of inclusion we introduce the following criterion

$$V(U) = ||P(U) \in G(P)|| \qquad (5)$$

Here $||a \in A||$ stands for a degree of inclusion of numeric value "a" in the interval A. Obviously one can study here Boolean (0-1) predicates however its multivalued counterpart of the inclusion predicate is more suitable for optimization purposes as it offers some desirable aspects of continuity assuming truth values ranging from 0 to 1.

Let us rewrite (5) using a multivalued inclusion predicate (ϕ)

$$V(U) = \sum_{k,l=1}^{N} (p_{kl}\varphi g_{kl}^+)(g_{kl}^-\varphi p_{kl}) \qquad (6)$$

where p_{kl} is the entry of the proximity matrix for the pair of (k, l) data. In more detail the inclusion predicate shown above is defined in the following form

$$a\phi b = \begin{cases} 1 & \text{if } a \le b \\ b/a, & \text{if } a > b \end{cases} \qquad (7)$$

$a, b \in [0,1]$.

The adjustment of the entries of the partition matrix is done in an iterative fashion by following the gradient-based optimization scheme

$$u_{st}(\text{iter} + 1) = u_{st}(\text{iter}) - \xi \frac{\partial}{\partial u_{st}} V(U) \qquad (8)$$

$\xi > 0$ is a certain learning rate controlling intensity of the learning process .

Proceeding with the detailed formulas of the gradient we complete the following calculations

$$\frac{\partial V}{\partial u_{st}} = \sum_{k,l} \left[\frac{\partial(p_{kl}\varphi g_{kl}^+)}{\partial u_{st}}(g_{kl}^-\varphi p_{kl}) + \frac{\partial(g_{kl}^-\varphi p_{kl})}{\partial u_{st}}(p_{kl}\varphi g_{kl}^+) \right] \qquad (9)$$

$s = 1, 2, \ldots, c$ and (k, l) pertains top the pair of data in $\mathbf{D}$. Proceeding with the detailed calculations, we note that $\frac{\partial(p_{kl}\varphi g_{kl}^+)}{\partial u_{st}} = \frac{\partial(p_{kl}\varphi g_{kl}^+)}{\partial p_{kl}} \frac{\partial p_{kl}}{\partial u_{st}}$. This yields

$$\frac{\partial(p_{kl}\varphi g_{kl}^+)}{\partial p_{kl}} = \frac{\partial}{\partial p_{kl}} \begin{cases} 1, & \text{if } p_{kl} \le g_{kl}^+ \\ \frac{g_{kl}^+}{p_{kl}}, & \text{if } p_{kl} > g_{kl}^+ \end{cases} = \begin{cases} 0, & \text{if } p_{kl} \le g_{kl}^+ \\ \frac{-g_{kl}^+}{p_{kl}^2}, & \text{if } p_{kl} > g_{kl}^+ \end{cases} \qquad (10)$$

$$\frac{\partial p_{kl}}{\partial u_{st}} = \frac{\partial}{\partial u_{st}} \sum_{w=1}^{c} \min(u_{wk}, u_{wl}) = \sum_{w=1}^{c} \frac{\partial}{\partial u_{st}} \min(u_{wk}, u_{wl}) \qquad (11)$$

and in the sequel we have

$$\frac{\partial}{\partial u_{st}} \min(u_{wk}, u_{wl})$$
$$= \begin{cases} 1 & \text{if } (u_{wk} \le u_{wl}, w=s, k=t) \text{ or } (u_{wl} \le u_{wk}, w=s, k=t) \\ 0, & \text{otherwise} \end{cases} \qquad (12)$$

Proceeding with the second part of the (9), the detailed formula reads as follows

$$\frac{\partial(g_{kl}^-\varphi p_{kl})}{\partial p_{kl}} = \frac{\partial}{\partial p_{kl}} \begin{cases} 1 & \text{if } g_{kl}^- \le p_{kl} \\ \frac{p_{kl}}{g_{kl}^-}, & \text{if } g_{kl}^- > p_{kl} \end{cases} = \begin{cases} 0, & \text{if } g_{kl}^- \le p_{kl} \\ \frac{1}{g_{kl}^-}, & \text{if } g_{kl}^- > p_{kl} \end{cases} \qquad (13)$$

Note that to retain the values of the partition matrix in the unit interval, a clipping operation, if required, is invoked that is at any iteration the values of $u_{st}(\text{iter}+1)$ are kept within the [0,1] interval. Furthermore, we complete an additional normalization operation to keep the sum of the values of u_{st} (being summed over "s" for any "t") equal to 1. The initial point of the iteration scheme is the original partition matrix locally available for each data.

The gradient-based mechanism can be considered as a stand-a-lone optimization scheme or could be considered in

Table 1 Statistical characteristics of synthetic two- and three-dimensional data

Set	c[p]				
$\mathbf{D}_1$	2	$m=[-3\ 9]\ \Sigma=\begin{bmatrix}2.2 & 0\\0 & 0.6\end{bmatrix}$	$m=[-2\ 4]\ \Sigma=\begin{bmatrix}2.4 & 0\\0 & 0.8\end{bmatrix}$		
$\mathbf{D}_2$	2	$\mathbf{m}=[2\ -9\ 1]\ \Sigma=\begin{bmatrix}2.1 & 0 & 0\\0 & 1.7 & 0\\0 & 0 & 0.4\end{bmatrix}$	$m=[8\ 6\ 4]\ \Sigma=\begin{bmatrix}2.2 & 0 & 0\\0 & 1.7 & 0\\0 & 0 & 0.7\end{bmatrix}$		
$\mathbf{D}_3$	3	$m=[9\ 1]\ \Sigma=\begin{bmatrix}1.3 & 0\\0 & 1.9\end{bmatrix}$	$m=[-10\ -6]\ \Sigma=\begin{bmatrix}1.0 & 0\\0 & 0.1\end{bmatrix}$	$m=[-1\ -6]\ \Sigma=\begin{bmatrix}2.8 & 0\\0 & 1\end{bmatrix}$	
$\mathbf{D}_4$	3	$m=[2\ 4]\ \Sigma=\begin{bmatrix}2.2 & 0\\0 & 0.1\end{bmatrix}$	$m=[-2\ -3]\ \Sigma=\begin{bmatrix}1.8 & 0\\0 & 0.9\end{bmatrix}$	$m=[10\ 2]\ \Sigma=\begin{bmatrix}1.6 & 0\\0 & 0.5\end{bmatrix}$	
$\mathbf{D}_5$	4	$m=[8\ 9]\ \Sigma=\begin{bmatrix}1.0 & 0\\0 & 0.5\end{bmatrix}$	$m=[6\ -3]\ \Sigma=\begin{bmatrix}2.4 & 0\\0 & 0.6\end{bmatrix}$	$m=[2\ 2]\ \Sigma=\begin{bmatrix}0.9 & 0\\0 & 0.1\end{bmatrix}$	$m=[-3\ -1]\ \Sigma=\begin{bmatrix}1.9 & 0\\0 & 1.9\end{bmatrix}$
$\mathbf{D}_6$	2	$m=[3\ 0\ -10]\ \Sigma=\begin{bmatrix}1.9 & 0 & 0\\0 & 0.4 & 0\\0 & 0 & 0.1\end{bmatrix}$	$m=[-10\ 15]\ \Sigma=\begin{bmatrix}2.6 & 0 & 0\\0 & 1.8 & 0\\0 & 0 & 0.2\end{bmatrix}$		
$\mathbf{D}_7$	4	$m=[10\ -4]\ \Sigma=\begin{bmatrix}1.0 & 0\\0 & 1.2\end{bmatrix}$	$m=[-9\ -9]\ \Sigma=\begin{bmatrix}2.5 & 0\\0 & 1.8\end{bmatrix}$	$m=[-9\ 7]\ \Sigma=\begin{bmatrix}2.4 & 0\\0 & 1.6\end{bmatrix}$	$m=[-2\ -4]\ \Sigma=\begin{bmatrix}0.1 & 0\\0 & 0.4\end{bmatrix}$
$\mathbf{D}_8$	2	$m=[6\ 10\ -5]\ \Sigma=\begin{bmatrix}0.5 & 0 & 0\\0 & 0.3 & 0\\0 & 0 & 0.9\end{bmatrix}$	$m=[1\ 8\ -5]\ \Sigma=\begin{bmatrix}2.9 & 0 & 0\\0 & 0.3 & 0\\0 & 0 & 0.6\end{bmatrix}$		
$\mathbf{D}_9$	2	$m=[-8\ 10]\ \Sigma=\begin{bmatrix}2.1 & 0 & 0\\0 & 0.1 & 0\\0 & 0 & 0.8\end{bmatrix}$	$m=[8\ -7\ 0]\ \Sigma=\begin{bmatrix}1.7 & 0 & 0\\0 & 1.7 & 0\\0 & 0 & 0.1\end{bmatrix}$		
$\mathbf{D}_{10}$	2	$m=[1\ -4]\ \Sigma=\begin{bmatrix}2.2 & 0\\0 & 0.1\end{bmatrix}$	$m=[-7\ 8]\ \Sigma=\begin{bmatrix}1.7 & 0\\0 & 0.6\end{bmatrix}$		
$\mathbf{D}_{11}$	2	$m=[-8\ 7]\ \Sigma=\begin{bmatrix}1.8 & 0\\0 & 1.7\end{bmatrix}$	$m=[-6\ -1]\ \Sigma=\begin{bmatrix}2.1 & 0\\0 & 0.4\end{bmatrix}$		
$\mathbf{D}_{12}$	4	$m=[-5\ 5]\ \Sigma=\begin{bmatrix}2.9 & 0\\0 & 0.6\end{bmatrix}$	$m=[-5\ 6]\ \Sigma=\begin{bmatrix}0.6 & 0\\0 & 0.4\end{bmatrix}$	$m=[8\ -5]\ \Sigma=\begin{bmatrix}0.5 & 0\\0 & 1.7\end{bmatrix}$	$m=[-9\ 10]\ \Sigma=\begin{bmatrix}1.2 & 0\\0 & 1.6\end{bmatrix}$
$\mathbf{D}_{13}$	2	$m=[6\ 1\ 8]\ \Sigma=\begin{bmatrix}0.9 & 0 & 0\\0 & 1.8 & 0\\0 & 0 & 0.5\end{bmatrix}$	$m=[-3\ 8\ 8]\ \Sigma=\begin{bmatrix}2.9 & 0 & 0\\0 & 1.3 & 0\\0 & 0 & 0.8\end{bmatrix}$		
$\mathbf{D}_{14}$	4	$m=[6\ 10\ -5]\ \Sigma=\begin{bmatrix}0.5 & 0 & 0\\0 & 0.3 & 0\\0 & 0 & 0.9\end{bmatrix}$	$m=[1\ 8\ -5]\ \Sigma=\begin{bmatrix}2.9 & 0 & 0\\0 & 0.3 & 0\\0 & 0 & 0.6\end{bmatrix}$	$m=[-9\ -6\ 5]\ \Sigma=\begin{bmatrix}2.0 & 0 & 0\\0 & 0.1 & 0\\0 & 0 & 0.2\end{bmatrix}$	$m=[-9\ 1\ 8]\ \Sigma=\begin{bmatrix}1.7 & 0 & 0\\0 & 1.9 & 0\\0 & 0 & 0.9\end{bmatrix}$
$\mathbf{D}_{15}$	2	$m=[2\ 6\ -6]\ \Sigma=\begin{bmatrix}1.5 & 0 & 0\\0 & 0.2 & 0\\0 & 0 & 0.7\end{bmatrix}$	$m=[-3\ -7\ 9]\ \Sigma=\begin{bmatrix}2.9 & 0 & 0\\0 & 2.0 & 0\\0 & 0 & 0.3\end{bmatrix}$		

**Fig. 5** Plots of synthetic data sets
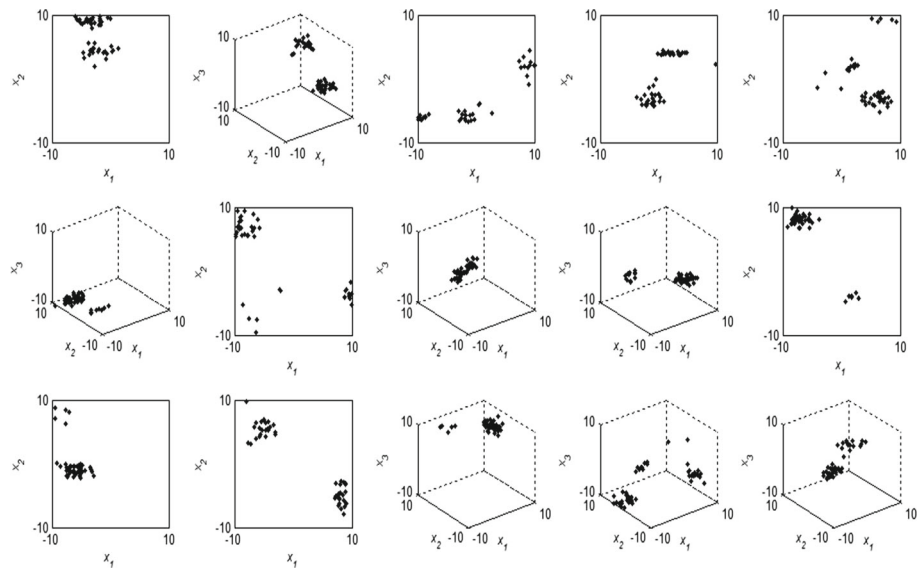
conjunction with more advanced population-based optimization such as Particle Swarm Optimization (PSO) and establishes a hybrid optimization scheme in which both of these optimization mechanisms are arranged in a certain sequence with an ultimate intent of avoiding local minima.

With regard to the optimization, several hybridizations of the generic optimization mechanisms are worth investigating, namely, tandems of PSO-gradient method and gradient optimization- PSO where we capitalize on the key properties of these techniques. PSO as the population-based technique is beneficial in realizing a global-oriented search, whereas the gradient-oriented method comes with a very detailed search capabilities, however, it is also prone to being stuck in possible local minima. A hybrid scheme of the form of PSO followed by the gradient-based technique comes as a sound alternative emphasizing the advantages of the contributing methods.

4.3 A general scheme of consensus building

In contrast to the two previously outlined processes in which granular proximity matrices are involved, consensus building is an iterative process and its dynamics comes into play. We proceed in an iterative fashion by forming a granular proximity matrix $G(P)$ on the basis of locally formed partition matrices (proximity matrices) and then update each $U[1], U[2], \ldots, U[p]$ as discussed in the second scheme. Then these updated partition matrices lead to the proximity matrices and subsequently the new granular proximity matrix is produced. This complete iterative loop is repeated. The process is monitored with respect to its convergence. Some parameters of the method, especially the values of α can impact the convergence process and their impact can be assessed in an experimental fashion.
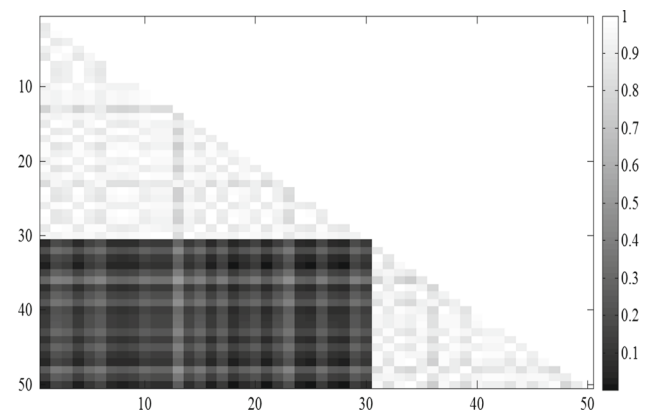
Fig. 6 Granular proximity matrix for $\mathbf{D}_1$. A sudden jump in the levels of brightness (resemblance values) occurring along the 30th data point is a result of the occurence of two well-formed clusters, see Fig. 8. The first one involves the first 30 data points while the rest of the data (from 31 to 50) form the second cluster

### 4.4 Characterization of granular proximity matrices

There are several indicators that can be used as sound descriptors of the produced granular proximity matrix supporting also the quality of the convergence process encountered in consensus building.

Linkage analysis

As the matrix $G(P)$ is of granular character, its quantification of content is realized as interval-valued strength of linkage or a fuzzy set of linkage. Moving at the more synthesized level of description by summing the elements of the k-th row of the matrix, namely
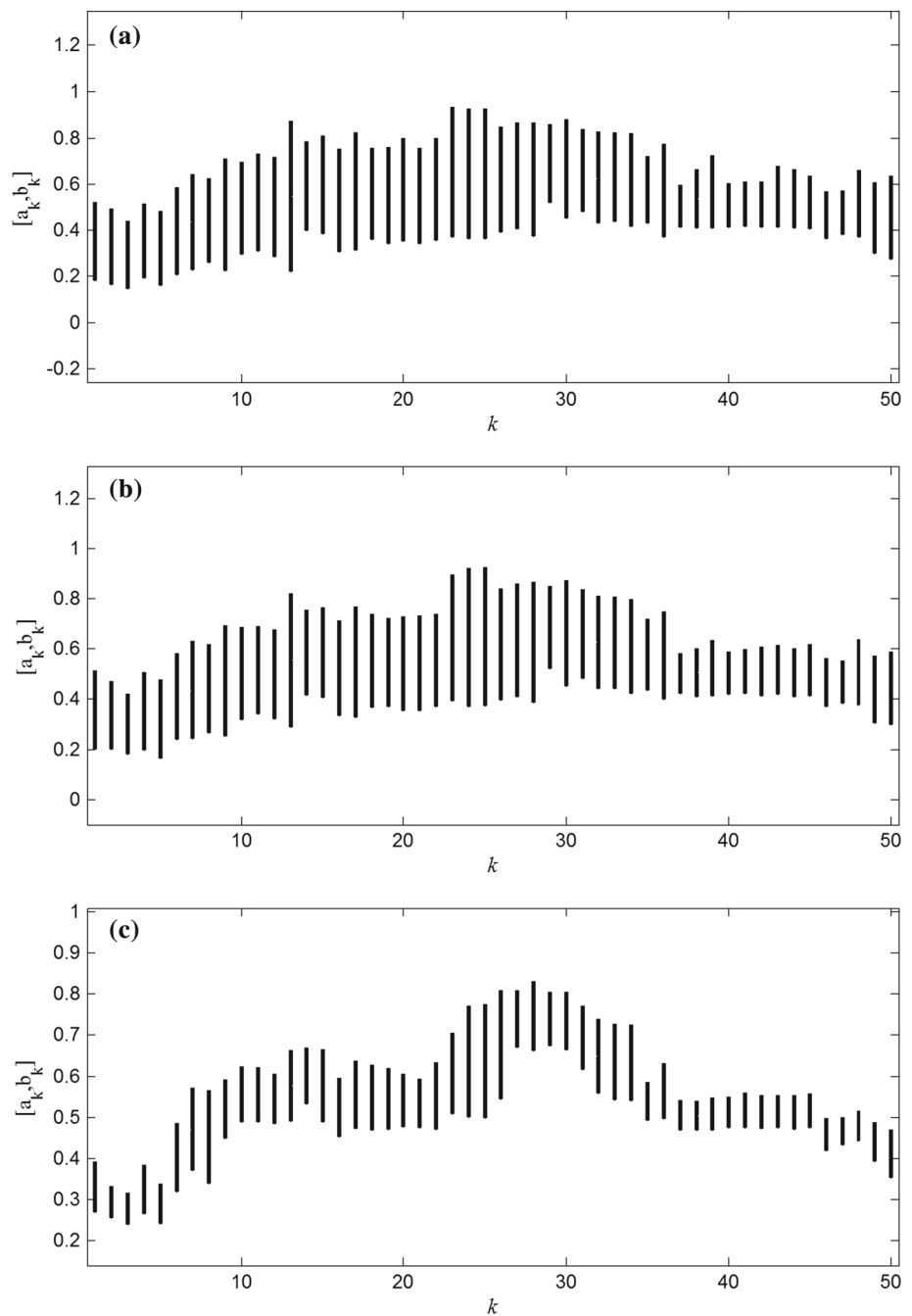
$$[a_k, b_k] = \frac{1}{N-1} \sum_{\substack{l=1 \\ l \neq k}}^{N} [g_{kl}^-, g_{kl}^+]$$

$$= \left[\frac{1}{N-1} \sum_{\substack{l=1 \\ l \neq k}}^{N} g_{kl}^-, \frac{1}{N-1} \sum_{\substack{l=1 \\ l \neq k}}^{N} g_{kl}^+ \right] \qquad (14)$$

we can identify data that are potential outliers—those are those data points for which the interval $[a_k, b_k]$ becomes located close to zero. Furthermore, the length of the inter-

val, namely $|b_k - a_k|$, becomes reflective of the diversity of the evaluations of proximity values of the k-th data delivered locally. This allows us to rank the data in terms of their associations with other data and tag outliers.

Overall granularity of granular proximity matrix

This index is useful when building consensus in an iterative process. The updated partition matrices are used to build proximity matrices and those give rise to the granular proximity matrix. In turn, the new proximity matrix offers a navi-

Fig. 8 Performance index V in the course of optimization (reported in successive generations of PSO or iterations of the gradient-based optimization) ($\alpha = 0.3$) using **a** gradient-based method, **b** PSO, **c** gradient- PSO and **d** PSO-gradient
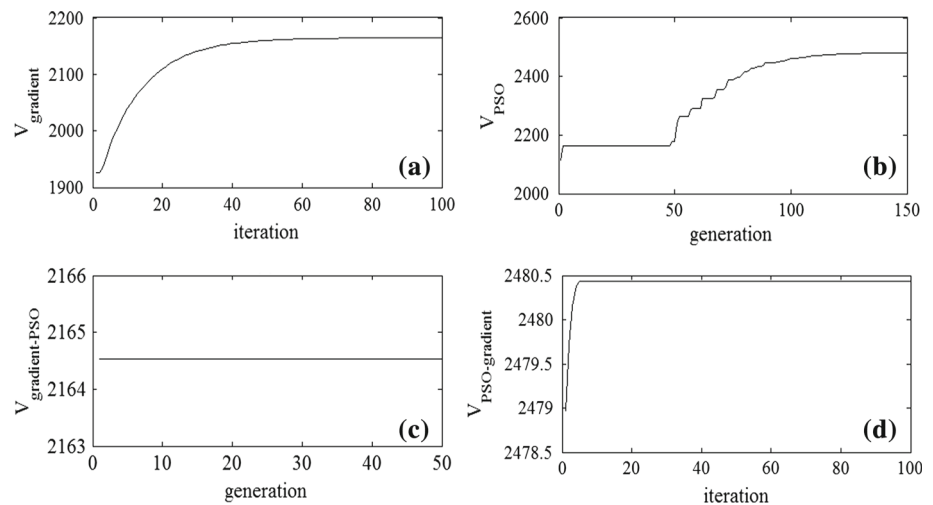
gation of optimization of the local partition matrices and the iterations are continued. The convergence of the process can be monitored by an overall level of granularity of the granular partition matrix, denoted by Gran produced in the successive iteration steps,

$$\text{Gran} = \frac{1}{N^2 - N} \sum_{\substack{k,l \\ k \neq l}} (g_{kl}^+ - g_{kl}^-) \tag{15}$$

The decreasing values of this index are indicative of the reduced diversity of the locally formed structures and in this way they point at the increasing agreement among different locally produced views.

5 Numeric studies

In this section, a series of experiments involving both synthetic and real-world data are presented to illustrate how different schemes discussed above operate and a form of the results formed. In all experiments, we consider Fuzzy C-Means (FCM) algorithm (Bezdek 1981) run with the fuzzification coefficient set to 2, $m = 2$.

5.1 Synthetic data

Here, we consider 15 synthetic data sets, both two- and three-dimensional ones, coming as mixtures of data governed by Gaussian distributions with some mean vectors m and covariance matrices Σ, $N(m, \Sigma)$. The statistical characteristics of the data are summarized in Table 1 while Fig. 5 displays their distribution in the corresponding feature spaces. Each set consists of 50 data points. As we assume the knowledge of the structure of the data, the number of clusters was set as $c_p[1], c_p[2]....c_p[ii], ii = 1, 2..c[p]$

In the sequel, we elaborate on the three models of building global structures of data or reconciling structural characteristics of the local nature.

Formation of granular proximity matrix The objective is to build a granular proximity matrix of a general nature positioned at the upper level of the hierarchy. As discussed, we first form a collection of proximity matrices and afterwards form their granular generalization. For illustration, Fig. 6 visualizes a proximity matrix obtained for $\mathbf{D}_1$.

Figure 6 Granular proximity matrix for $\mathbf{D}_1$. A sudden jump in the levels of brightness (resemblance values) occurring along the 30th data point is a result of the occurence of two well-formed clusters, see Fig. 8. The first one involves the first 30 data points while the rest of the data (from 31 to 50) form the second cluster.

The granular proximity matrix visualizes an emergence of groups of data that are kept close to each other. One can observe a jump in the values of the entries pointing at the occurrence of the well-formed clusters.

The formation of granular proximity matrix was considered for several values of α. Proceeding with the link analysis, we obtain the results shown in Fig. 7. The link analysis helps us visualize the points that weakly linked with the rest of the data as well as highlight those for which the length of the interval is excessively large.

These results are reported for several values of α (namely, 0.3, 0.6, and 1). There is a visible tendency of a stronger and visible revealing of the outliers when the levels of binding locally available structures are made stronger. For instance, as shown in Fig. 7c, it is apparent that there are some collections of data (those indexed as 1–5 and 35–50) which are different with regard to the locally present structures.

Involvement of global granular results in the enhancements of the local partition matrix. Here, as discussed earlier, the optimization strategy involves plain gradient-based

Fig. 9 Partition matrix for **D**₁
a obtained locally, **b** refinement
realized when $\alpha = 0.3$ and
c refinement realized when
$\alpha = 0.6$
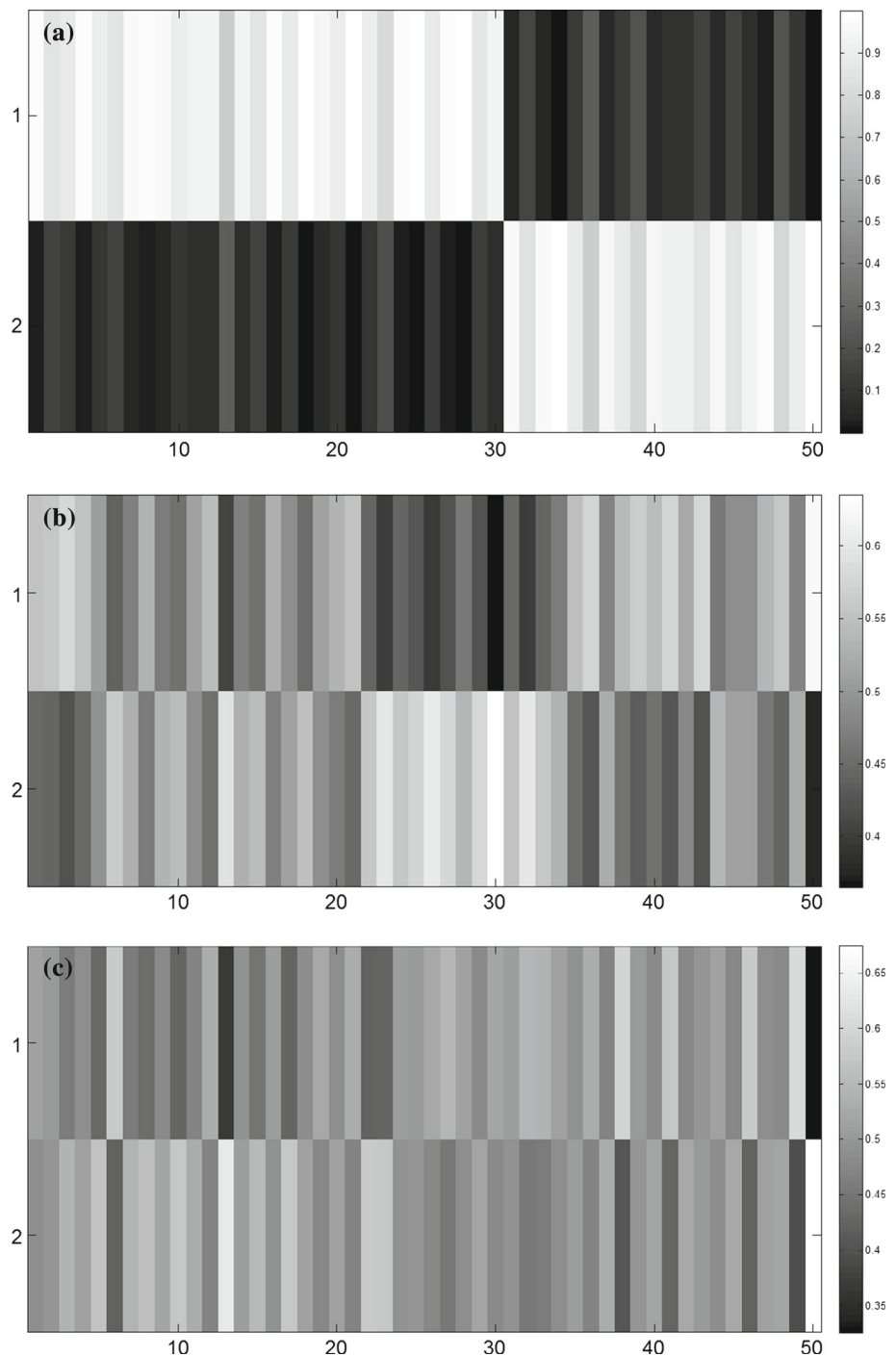
method and PSO as well as several of their hybrid approaches combining these generic methods.

The results are reported in terms of the performance index V whereas the initial learning rate was set to $\xi = 0.01$.

The hybrid optimization method PSO-gradient, Fig. 8d, outperforms other optimization schemes. The partition matrix U generated with the use of the PSO method, Fig. 8b, serves as a sound initial condition for the gradient-based method, which is helpful in carrying out fine-tuning of the

entries of U. For the other hybrid method (gradient-PSO) Fig. 8c, it is clearly shown that PSO does not produce further improvement for V as it is eventually stuck in some local maxima.

It is noted that the value of the learning rate ξ was made quite low with intent of making the process stable when it comes to the fine-tuning phase of the entries of $G(P)$. Furthermore, we have adopted some dynamic changes of the values of the learning rate. If the value of V decreases in a

certain iteration, this iteration is ignored and the value of ξ is decreased, say being made $0.75*\xi$. Next we continue the learning with this new value of ξ until V decreases again, etc.
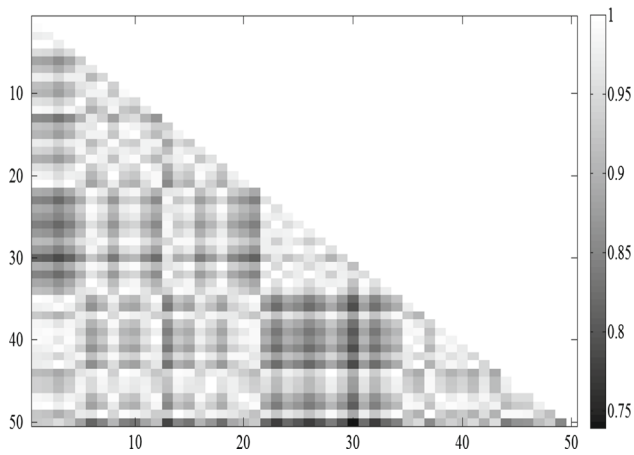
Fig. 10 Proximity matrix of after refinement partition matrix ($\alpha = 0.3$) obtained for $\mathbf{D}_1$

The original partition matrix (formed locally) and its refinement produced by the involvement of the granular proximity matrix are shown in Fig. 9.

While the local structure is quite apparent, the refinements guided by the globally produced structure lead to some changes of the structure. This is not surprising as some global structure is considered and its impacts become clear.

The results obtained after the detailed refinements of partition matrix are visualized in Fig. 10.

Figure 9 reveals some interesting relationships. When only local data are considered, there is a well-delineated structure, which points at two clusters. For the increasing values of α, $\alpha = 0.3$ and 0.6, we witness an increasingly influential impact of the global structure so the clusters are not as distinct as in the first case. Obviously, this is not surprising, as now we have started accommodating a global view (structure), which might not be in full agreement with the local topology of the data. Furthermore, the partition matrices displayed in this figure identify data points, which are mostly impacted by the global structure. This is a useful insight into the nature of the individual data, which helps pinpoint the elements, which are
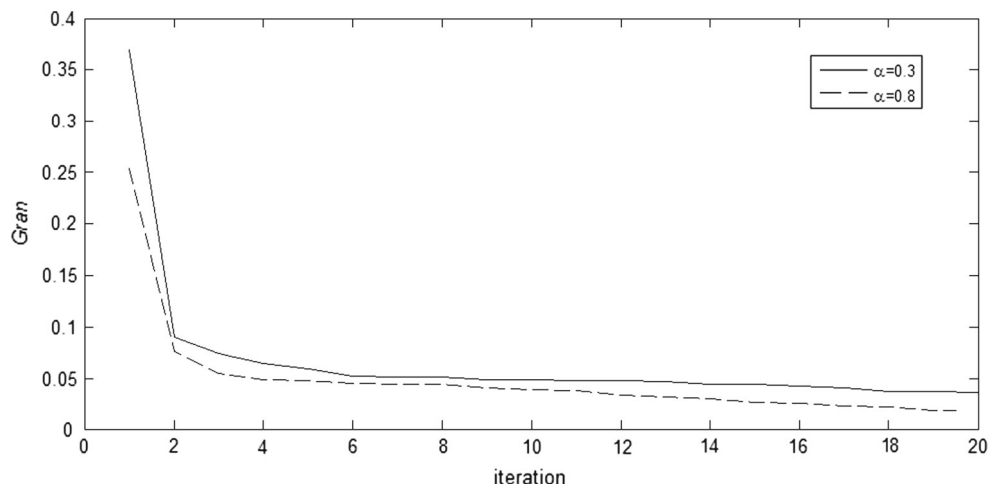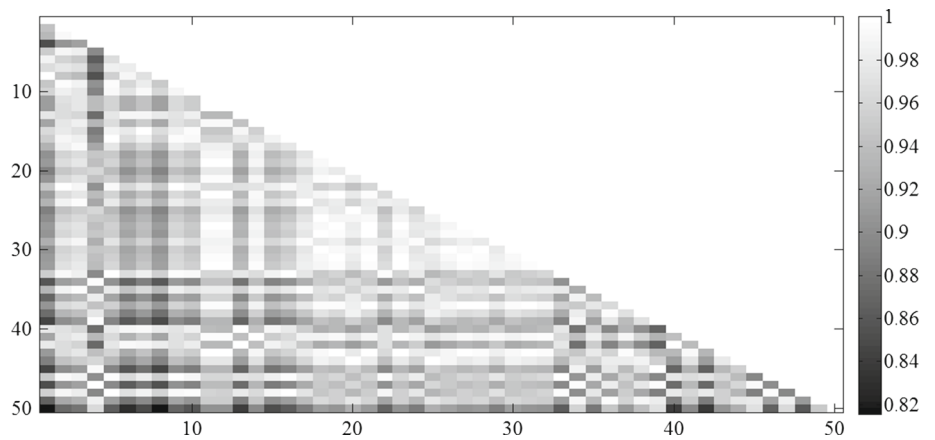

Fig. 11 Overall granularity obtained in successive iterations of the consensus-building process; results are shown for two selected values of $\alpha = 0.3$ and 0.8
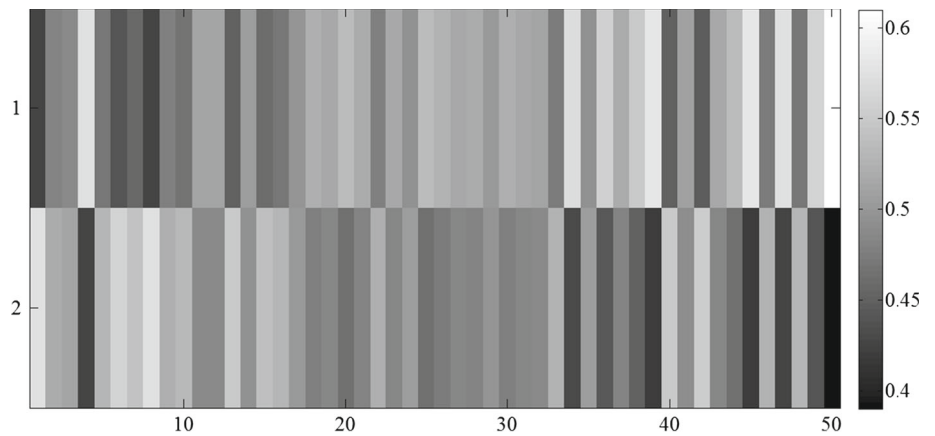
Fig. 12 Bounds of the linkage levels of the granular proximity matrix for produced as a result of consensus formation with $\alpha = 0.3$
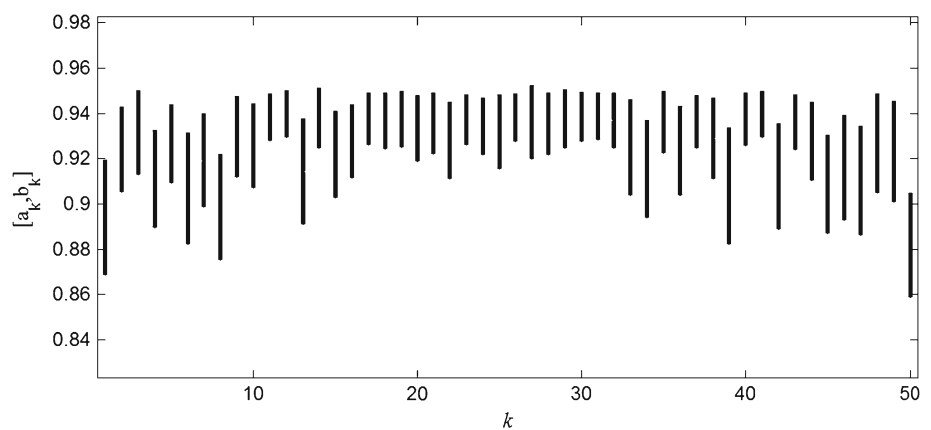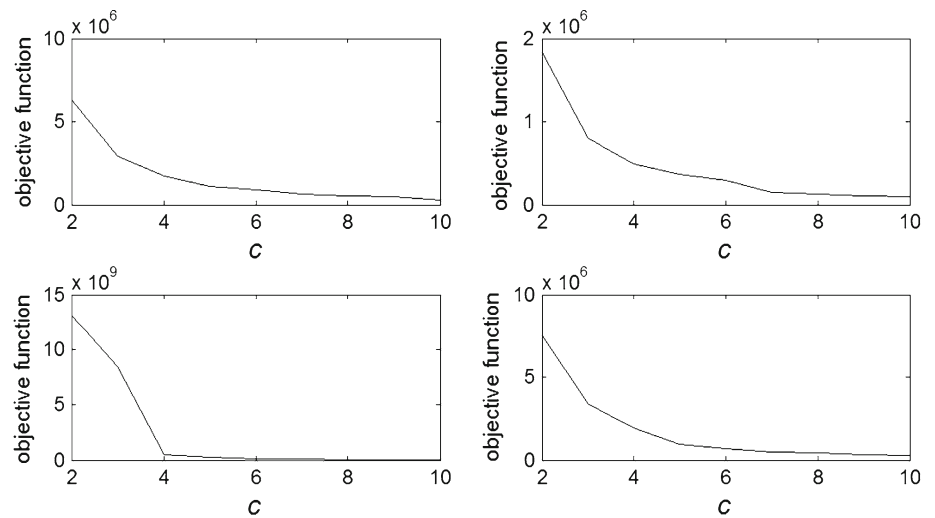
Fig. 13 proximity matrix for
$\mathbf{D}_1$ as a result of
consensus-building process
($\alpha = 0.3$)

Fig. 14 Partition matrix
obtained for $\mathbf{D}_1$ as a result of
consensus-building process
($\alpha = 0.3$)

Fig. 15 Objective functions
produced when clustering
individual locally available data

the least compatible with the global structure revealed at the higher level of the hierarchy.

Consensus building The most essential aspect of this process is concerned with the iterative process of forming consensus and its convergence. The results obtained when running the hybrid method (PSO-gradient option) are included in Fig. 11.

Table 2 Number of clusters based on the inspection of the objective function

	$\mathbf{D}_1$	$\mathbf{D}_2$	$\mathbf{D}_3$	$\mathbf{D}_4$
c	5	7	5	4

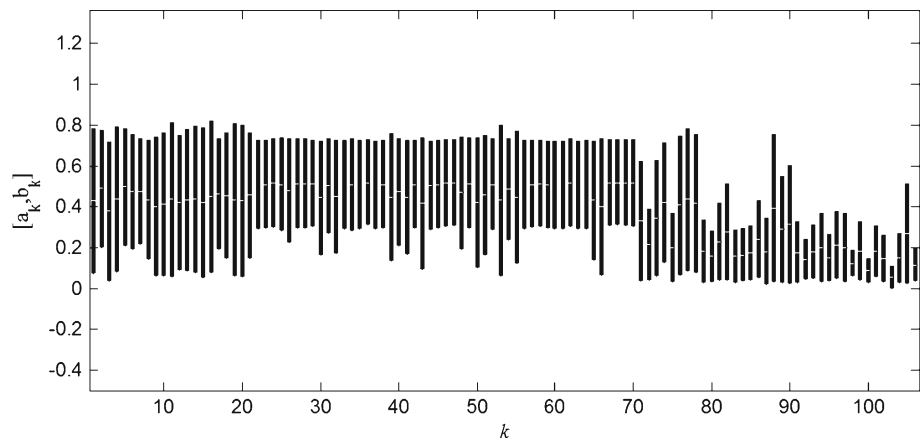The results of link analysis coming as a result of consensus formation are visualized in Fig. 12.
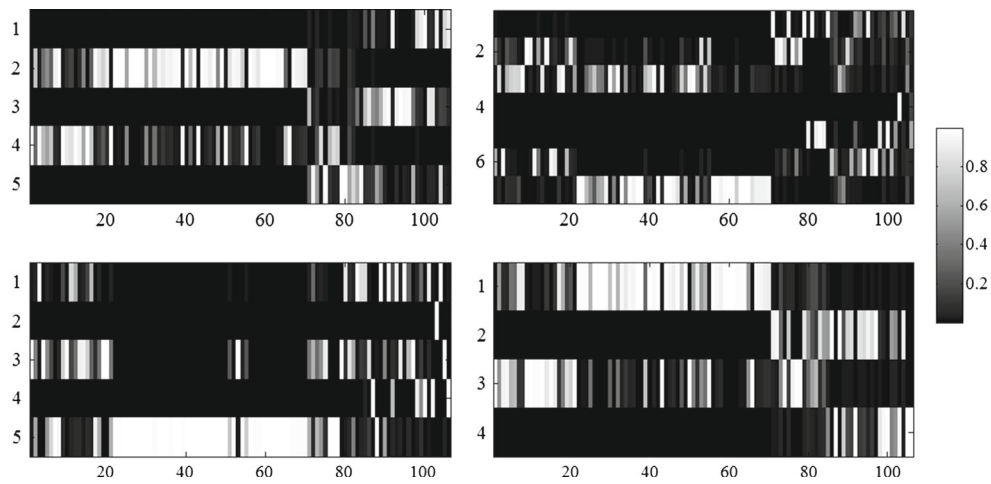
Fig. 16 Partition matrices obtained for the local data sets
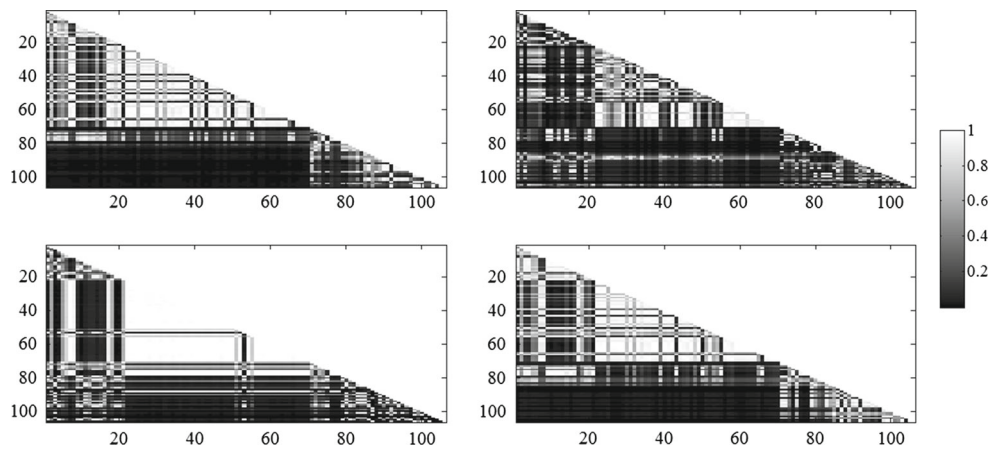
Fig. 17 Proximity matrices formed for the locally formed partition matrices

Fig. 18 Bounds of the linkage levels of the granular proximity matrix for $\alpha = 0.4$ for public set

It is also instructive to show how partition and proximity matrices changed once the consensus-building process has been completed; the results provided in the form of the gray scaled image are reported in Fig. 13.

The plots shown in Fig. 14 are useful in flagging the elements whose membership grades change significantly (in comparison with other data points). They could be candidates for further examination as their changed cluster membership
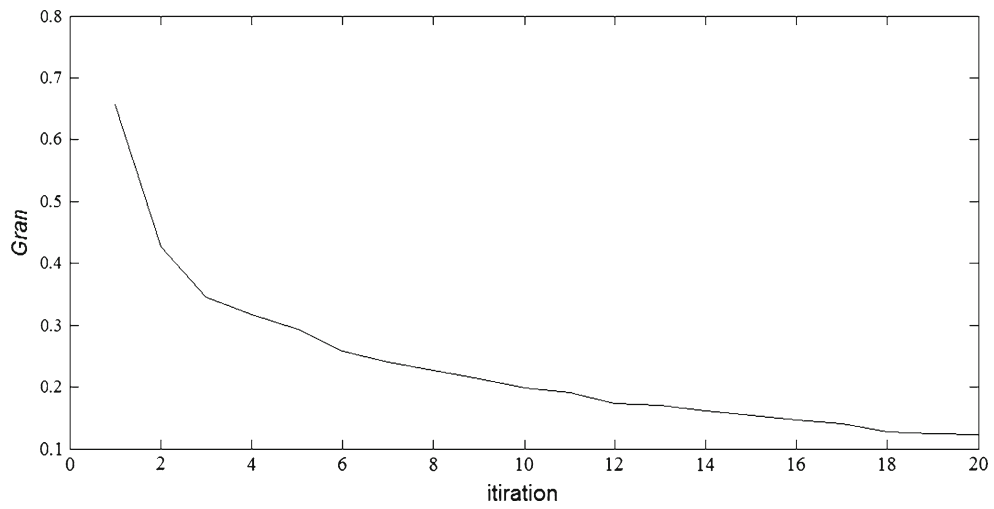
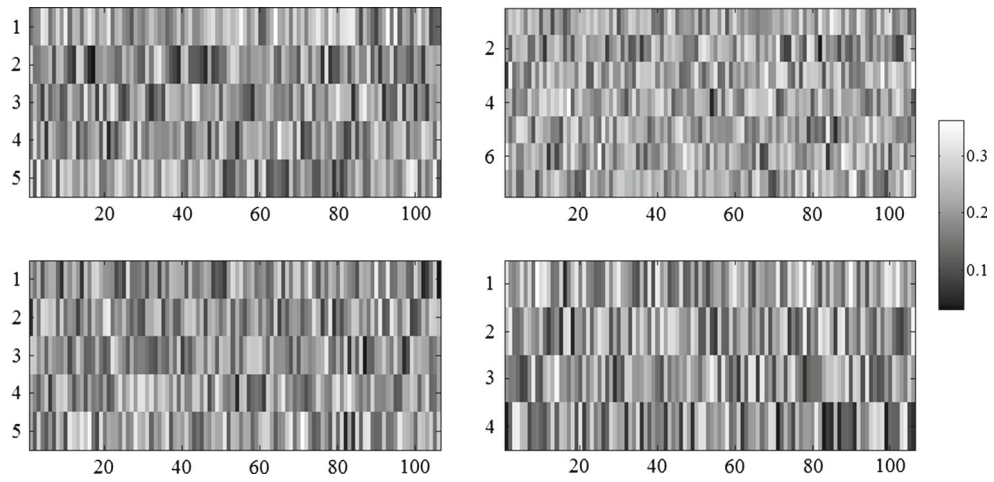**Fig. 19** Overall granularity reported in successive iterations; $\alpha = 0.4$

Fig. 20 Partition matrices as a result of consensus formation
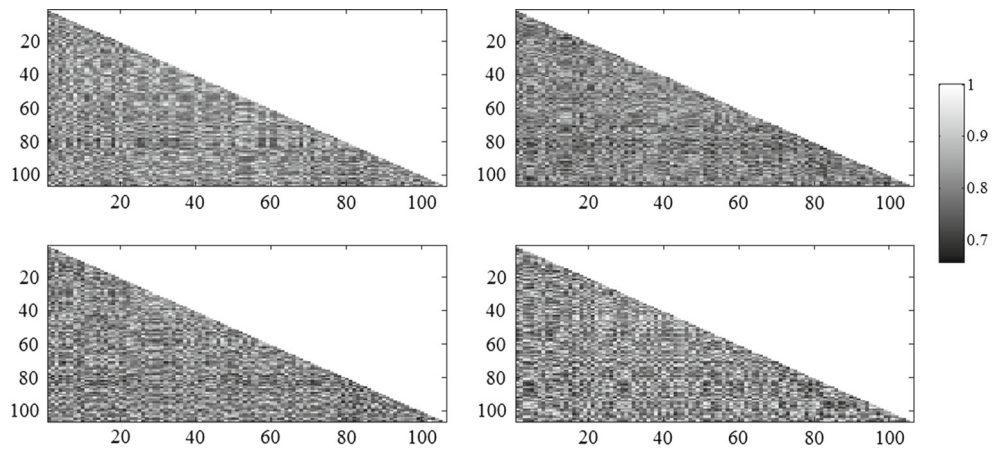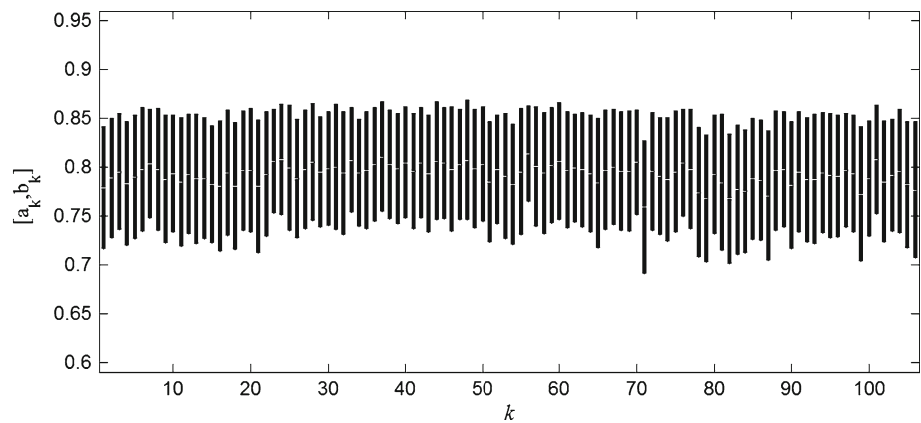

Fig. 21 Proximity matrices for partition matrices in Fig. 19, as result of consensus formation

Fig. 22 Bounds of the linkage levels of the granular proximity matrix for $\alpha = 0.4$ for public set, as a result of consensus formation

may stipulate characteristics that become revealed only in the presence of some external sources of knowledge.

Experiments with real-world data

We study here one of the publicly available real-world data (https://archive.ics.uci.edu/ml/datasets/Breast+Tissue), namely Breast tissues. This data set is concerned with electrical impedance measurements of freshly excised tissue samples from the breast. The features include the following:

1. I0 Impedivity (ohm) at zero frequency
2. PA500 phase angle at 500 KHz
3. HFS high-frequency slope of phase angle
4. DA impedance distance between spectral ends
5. AREA area under spectrum
6. A/DA area normalized by DA
7. MAX IP maximum of the spectrum
8. DR distance between I0 and real part of the maximum frequency point
9. P length of the spectral curve

For the purpose of this experiment, we split the data into four data sets by choosing several subsets of features in which the features are naturally related; each set is composed from 106 samples with the following subsets of features

D_1: 1, 7
D_2: 4, 8
D_3: 2, 3, 9
D_4: 5, 6

Proceeding with the fuzzy clustering realized with the use of the FCM algorithm, we determine local structures. The number of clusters is determined by inspecting the behavior of the minimized objective function being treated as a function of "c" and determining a "knee" point of the curve; refer to Fig. 15.

By inspecting the plots of the obtained objective function, we can choose a suitable value of the number of clusters, it

is the one at which a knee point of the relationship is visible. Adhering to this visual criterion, we select the numbers of clusters reported in Table 2.

The partition and proximity matrices are visualized in Figs. 16 and 17.

Considering $\alpha = 0.4$ when running the principle of justifiable granularity, we obtain the results shown in Fig. 18.

The result reported in terms of overall granularity, partition matrices, proximity matrices and linkage associated with the granular proximity matrix are shown in Figs. 19, 20, 21 and 22.

6 Conclusions

In this study, we have conceptualized, developed the algorithmic setting, and experimented with granular proximity matrices. It has been demonstrated that granularity of these matrices plays an important role in the realization of collaborative processes of forming views at the global structures not only facilitating this process, but also quantifying the diversity of locally available structures through the associated level of information granules of the granular proximity matrix. The guidance offered by global granular proximity matrices is an example of a realization of a structural feedback loop which augments the clustering processes by auxiliary sources of knowledge.

There are two open directions, which are worth further investigations:

Formation of structures exhibiting a higher type of granularity Higher level structures such as granular^2 proximity matrices (if a hierarchy having three levels is present) can be discussed.

Exploration of various formal ways of realizations of granular proximity matrices. While in this study, we are concerned with interval-valued proximity matrices (and this has been done for illustrative purposes), detailed considerations could involve other formalisms such as, e.g., fuzzy sets, rough sets and shadowed sets.

References

Apolloni B, Brega A, Malchiodi D, Palmas G, Zanaboni AM (2006) Learning rule representations from data. IEEE Trans Syst, Man Cybern, Part A: Syst Humans 36(5):1010–1028

Bezdek JC (1981) Pattern recognition with fuzzy objective function algorithms. Plenum Press, New York

Coppi R, D'Urso P, Giordani P (2010) A fuzzy clustering model for multivariate spatial time series. J Classif 27:54–88

Corsini P, Lazzerini B, Marcelloni F (2005) A new fuzzy relational clustering algorithm based on the fuzzy C-means algorithm. Soft Comput 9:439–447

de Souza RMCR, de Carvalho F (2004) Clustering of interval data based on city-block distances. Pattern Recognit Lett 25(3):353–365

Gacek A, Pedrycz W (2013) Description, analysis, and classification of biomedical signals: a computational intelligence approach. Soft Comput 17:1659–1671

Graves D, Noppen J, Pedrycz W (2012) Clustering with proximity knowledge and relational knowledge. Pattern Recognit 45(7):2633–2644

Hathaway R, Bezdek JC, Pedrycz W (1996) A parametric model for fusing heterogeneous fuzzy data. IEEE Trans Fuzzy Syst 4:270–281

Hwang C, Rhee FC-H (2007) Uncertain fuzzy clustering: interval type-2 fuzzy approach to C-Means. IEEE Trans Fuzzy Syst 15(1):107–120

Mali K, Mitra S (2003) Clustering and its validation in a symbolic framework. Pattern Recognit Lett 24(14):2367–2376

Pedrycz W, Bezdek JC, Hathaway RJ, Rogers GW (1998) A non-parametric model for fusing heterogeneous data. IEEE Trans Fuzzy Syst 6:411–425

Pedrycz W (2013) Granular computing: analysis and design of intelligent systems. CRC Press/Francis Taylor, Boca Raton

Pedrycz W, Rai P (2008) Collaborative clustering with the use of Fuzzy C-Means and its quantification. Fuzzy Sets Syst 15:2399–2427

Pedrycz W (2007) Granular computing—the emerging paradigm. J Uncertain Syst 1(1):38–61

Pedrycz W, Loia V, Senatore S (2004) P-FCM: a proximity-based fuzzy clustering. Fuzzy Sets Syst 148(1):21–41

Pedrycz W (2005) Knowledge-based fuzzy clustering. John Wiley, New York

Pedrycz W (2004) Fuzzy clustering with a knowledge-based guidance. Pattern Recognit Lett 25(4):469–480

Peters G (2011) Granular box regression. IEEE Trans Fuzzy Syst 19(6):1141–1152

Wong H, Hu BQ (2013) Application of interval clustering approach to water quality evaluation. J Hydrol 491:1–12

Zhang L, Pedrycz W, Lu W, Liu X, Zhang L (2014) An interval weighed fuzzy c-means clustering by genetically guided alternating optimization. Expert Syst Appl 41(13):5960–5971