Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition 41 (2008) 2630–2644
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/pr

Beyond FCM: Graph-theoretic post-processing algorithms for learning and representing the data structure

Nikolaos A. Laskaris*, Stefanos P. Zafeiriou

Artificial Intelligence and Information Analysis Laboratory, Department of Informatics, Aristotle University, Biology Building, BOX 451, GR-54124 Thessaloniki, Greece

Received 19 September 2007; received in revised form 25 January 2008; accepted 6 February 2008

Abstract

We show that when fuzzy C-means (FCM) algorithm is used in an over-partitioning mode, the resulting membership values can be further utilized for building a *connectivity graph* that represents the relative distribution of the computed centroids. Standard graph-theoretic procedures and recent algorithms from manifold learning theory are subsequently applied to this graph. This facilitates the accomplishment of a great variety of data-analysis tasks. The definition of optimal cluster number C_o, the detection of intrinsic geometrical constraints within the data, and the faithful low-dimensional representation of the original structure are all performed efficiently, by working with just a down-sampled version (comprised of the centroids) of the data. Our approach is extensively demonstrated using synthetic data and actual brain signals.

Keywords: Fuzzy clustering; Manifold learning; Prototyping; Spectral-graph theory; Visual data mining

1. Introduction

The detection of meaningful structure in the data, the proper description of this structure and its subsequent economical representation are unavoidable steps for discovering knowledge from large datasets and using it later for modeling and prediction. Clustering, very often, plays an instrumental role in such explorations. Using the algorithm of choice, the ensemble of data is partitioned into homogenous data chunks with a single prototype representing each one of the produced groups. The derived, down-sampled, version of the dataset, if/when organized properly can be used for skimming (i.e. it can serve as an intelligible summary of the whole spectrum of information contained within the original dataset). Such organization, which is inherent in some neural network algorithms like the Kohonen's map, can be achieved with a suitable post-processing (e.g. [1]). In the simplest case, the individual dataset items are distributed among different clear-cut categories, which are readily recognized by the clustering routine.

Dimensionality reduction offers an alternative way to look for putative structure in the dataset by representing the items in a reduced space and therefore facilitating visual inspection (e.g. PCA and MDS). Recently, there has been a renewed interest in dimensionality reduction techniques that evolved into a distinct branch of data analysis, namely the manifold learning theory [2–5]. The emerging techniques have a common theme: they attempt to recover, in a fully unsupervised manner, the intrinsic dimensionality of the data by starting from high-dimensional observations and detecting geometrical constraints in their local distribution. They offer a natural re-parameterization of the items, and this is equivalent with an efficient structural description for a given dataset. Depending on the particular method, the *learning* of *manifolds* (i.e. the tracing of constrained surfaces) can be accomplished by means of spectral-graph theory, semidefinite programming, etc.

The conjunction of clustering and dimensionality reduction has a long history. First, distance-preserving maps were suggested as visual tools that can support the selection of true cluster number. Then, some self-organizing nets had inbuilt an interaction between the clustering process and the mapping of the formed groups on a low-dimensional grid. Finally, spectral

* Corresponding author. Tel.: +30 2310 998706; fax: +30 2310 998453.
E-mail address: laskaris@aiia.csd.auth.gr (N.A. Laskaris).

0031-3203/$30.00
doi:10.1016/j.patcog.2008.02.005

graph-theoretic algorithms for cluster analysis (e.g. [6,7]) were built over the eigenanalysis (and therefore the low-dimensional description) of pairwise relational data. More recently, manifold learning algorithms were shown to result in low-dimensional embeddings that either enhance the clustering tendencies of the original data or could be easily handled via traditional cluster algorithms like C-means [8]. Specifically, an elongated cohesive structure, which would otherwise split into multiple groups, can be transformed after embedding into a single, compact group easily detectable by standard clustering [9]. We can try to conceptualize the essence of these algorithms as follows. An overall graph is built that encapsulates all the dataset items. This graph conveys all important pairwise relationships and in such a way that coherencies in the partial distribution of items are emphasized. A distinct graph component corresponds to a meaningful data structure, which might be characterized by intricate idiosyncratic geometry. A graph simplification/flattening step produces the final low-dimensional embedding in which the different graph components can be detected and traced easily.

Despite the commonsense practice to apply clustering after dimensionality reduction, in this work we wanted to explore the possibility of reversing this order as a means to achieve efficient and economical description of the structure in our data. The original motivation was the practical problem that large datasets are inaccessible to many of the current manifold learning algorithms. Especially in a visual data mining setting, where the dynamic interaction of the user with the embedding is desired, the execution time is prohibitive even for medium-sized datasets (of $\sim$ 2000 items). Apparently the effective size of a large dataset can be reduced by via *random sampling* or *prototyping* and the derived down-sampled version could be directly fed to any manifold learning algorithm. However, the former strategy cannot resolve this situation, since delicate structure is revealed only when sufficient number of items is taken into account. Only prototyping can indeed alleviate the problem, but only up to some extent since the geometrical relationships among prototypes cannot always reveal the true distribution over the manifold (see for instance Fig. 1).

The principal goal was to advance fuzzy C-means (FCM) algorithm, a widely available and indisputably efficient clustering technique, in such a way that the computed prototypes would provide a faithful skeleton of the underlying dataset structure. The core idea was that using the membership values, available after the execution of FCM, refined geometrical relationships among prototypes could be estimated with respect to the underlying data distribution. In this paper we show that, with simple algorithmic steps, a $(C \times C)$ connectivity graph is built containing all the essential information for the structural description and amenable (after some simple transformations) to recent manifold learning algorithms and standard graph-theoretic procedures for organizing datasets.

The paper is organized as follows. In Section 2 we review the standard FCM algorithm. In Section 3 we describe some ideas borrowed from topology-representing neural networks and show how these can be implemented within the framework of fuzzy clustering. In Section 4 we outline the known algorithms as adapted here for the post-processing of FCM outcomes, while in Section 5 we provide selected worked examples. Section 6 is devoted to the correlation of the presented work with recent literature and some concluding remarks.

2. The FCM algorithm

Fuzzy clustering can be traced back to Ref. [10] and the seminal work of Bezdek [11–14]. It incorporates various algorithms in which an object can simultaneously belong to several classes but with different degrees of certainty, as these measured by a membership function. In general, any such algorithm attempts to estimate a partition matrix $\mathbf{U}(\mathbf{X})$ of a given dataset $\mathbf{X} = \{X_1, X_2, \ldots, X_N\}$, that defines its segmentation into a number $C \ll N$ of homogeneous and contrasted subsets $(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_c)$, the clusters. The partition matrix U is of size $C \times N$ and has the form

$$\mathbf{U}(\mathbf{X}) = [u_{ij}]_{C\times N} = \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_C \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1N} \\ u_{21} & u_{22} & \cdots & u_{21} \\ \vdots & \vdots & \vdots & \vdots \\ u_{C1} & u_{C2} & & u_{CN} \end{bmatrix}.$$

In the above notation $u_{ij} \in [0, 1]$ denotes the membership grade of jth object to ith cluster, satisfying the usual conditions in fuzzy clustering

$$\text{(a) } 0 < \sum_{j=1}^{N} u_{ij} < N, \quad \forall i, \qquad \text{(b) } \sum_{i=1}^{C} u_{ij} = 1, \quad \forall j,$$

$$\text{(c) } \sum_{i=1}^{C}\sum_{j=1}^{N} u_{ij} = N.$$

In the particular case of FCM algorithm, the objects to be clustered are represented by vectors in R^p and therefore the dataset can take the form of a large matrix $\mathbf{X} = [x_{ij}]_{N\times p} = [X_1|X_2|\cdots X_N]$ with row-vectors of the form $X_i = [x_{i1}, x_{i2}, \ldots, x_{ip}]$. The algorithm derives C cluster centroids O_i in R^p, by minimizing the following objective function that includes the memberships as additional unknowns in the estimation of an aggregate, weighted and within-cluster dispersion

$$J_m(\mathbf{U}, \mathbf{O}) = \sum_{i=1}^{C}\sum_{j=1}^{N} u_{ij}^m \|X_j - O_i\|^2. \tag{1}$$

The matrix $\mathbf{O}$ denotes the set of centroids tabulated in the form $\mathbf{O}=[o_{ij}]_{C\times p}=[O_1|O_2|\cdots O_C]$, $m \in (1, \infty)$ is a weighting parameter [14,15] that controls the fuzziness in the classification process and the distance function $\|\cdot\|$ is the standard Euclidean norm. After a proper initialization of the partition matrix, an iterative optimization scheme is followed that alternates between the re-estimation of cluster centroids and the evaluation of new memberships for the objects according to the following

two formulae:

$$O_i = \frac{\sum_{j=1}^{N} u_{ij}^m X_j}{\sum_{j=1}^{N} u_{ij}^m}, \quad 1 \leqslant i \leqslant C \tag{2}$$

$$u_{ij} = \left[\sum_{k=1}^{C} \left(\frac{\|X_j - O_i\|^2}{\|X_j - O_k\|^2} \right)^{1/(m-1)} \right]^{-1}, \quad 1 \leqslant i \leqslant C, 1 \leqslant j \leqslant N. \tag{3}$$

Every time this dual step is executed, the emerged partition matrix is compared with its former version and the whole procedure terminates when the changes have become negligible. Alternatively, the termination can be based on the comparison of successive J_m values.

In the standard FCM-algorithm, both C and m are user-defined parameters. While for the later the usual choice is $m=2$, the definition of optimal number of classes C_o is generally considered and important issue that deserves a careful treatment since this number has to be tailored to the dataset at hand. Many validity indices, devoted specifically to fuzzy clustering, have been proposed (see Ref. [16] for an extensive comparison) and their relative success seems to be case-dependent. Among these indices, we briefly mention here the Bezdeck's fuzzy partition coefficient F_{PC} due to its affinity with the ideas presented later

$$F_{PC}(\mathbf{U}) = \frac{\mathrm{tr}(\mathbf{U}.\mathbf{U}^T)}{N}. \tag{4}$$

This index expresses the overall unshared membership of the data after being segregated in the C fuzzy clusters [15]. It is an aggregate 'anti-overlap' measure for which in general holds $1/C \leqslant F_{PC} \leqslant 1$. The higher the F_{PC} value, the clearer the induced partition is. When used for the definition of optimal cluster number, FCM is applied with variable cluster number C_i and the computed partition matrices $U(C_i)$ are used to identify

$$C_o = \underset{2 \leqslant C_i \leqslant C_{max}}{\arg\max} \; F_{PC}(U(C_i)).$$

The FCM algorithm has gained great popularity over the last years due to the concrete mathematical formulation, the straightforward implementation and the wide availability in software packages like Matlab. It has found a tremendous number of applications (e.g. [17–19]), and undergone a lot of modifications or extensions (e.g. [20–23]). Finally, post-processing procedures have been recently introduced with the aim of visualizing the computed memberships and therefore portraying inter-cluster overlap (see Ref. [24] for a detailed discussion of the topic).

3. Topology representing graphs

Martinez and Schulten, in a highly influential article [25], introduced the notion of *induced Delaunay triangulation* (IDT) and showed how IDT can be computed using a stochastic algorithm, that simulates Hebbian competitive rule, within the framework of neural-gas vector quantizer. Their implementation builds over the particular self-organizing neural network. However the core idea of a connectivity graph G_{ij}, $(i, j = 1, \ldots, C)$ that refers to the actual topological relations between prototypes and serves as a model-skeleton for even intricately structured distributions, applies to other prototyping procedures as well. In the sequel, we first describe the originally proposed procedure for defining $\mathbf{G}$, then mention briefly its basic properties and finally show how it can be estimated from the outcomes of FCM algorithm.

(step-1): The elements of a $C \times C$ matrix G are initialized to zeroes, with the entry G_{ij} corresponding to the strength of connection between prototypes O_i and O_j

(step-2): Each vector X_k, $k=1, \ldots, N$ is compared with all the prototypes according to Euclidean distance, the two nearest prototypes O_r and O_q are detected and the corresponding entries G_{rq} and G_{qr} are increased by 1.

The derived $\mathbf{G}$ is a symmetric matrix. Its elements $G_{ij} \geqslant 0$ are directly analogous to the extent of overlap between the receptive fields of O_i and O_j. More formally, nonzero values denote that the corresponding Voronoi regions are sharing a common boundary. With simple elementwise thresholding $\theta(C_{ij}) = A_{ij}$, where $\theta(\cdot)$ is the Heavyside step function, the adjacency matrix $\mathbf{A}$ of IDT is computed. This matrix corresponds to a meaningful subgraph of the Delaunay triangulation of the vectors O_i, $i = 1, \ldots C$, that perfectly preserves the topology of the data. It has the important property that a link connects two nearby prototypes only when they are natural neighbors over the manifold (for instance, see Fig. 1). Provided that the distribution of O_i is dense on M (i.e. C is sufficiently high), this implies that each edge $\overline{O_r O_q}$ characterized by $G_{rq} > 0$ (or equivalently $A_{rq} = A_{qr} = 1$) belongs to the data manifold. According to the authors of Ref. [25], this property opens the possibility of a discrete path preserving representation of the data manifold.

The connectivity graph $\mathbf{G}$ can be estimated easily after the execution of FCM algorithm, since the partition matrix $\mathbf{U}$ contains all the required information regarding the relative closeness of each vector X_j with all the prototypes O_i, $i = 1, 2, \ldots, C$. Using separately each column $\mathbf{U}_{(:,j)} = [u_{1j}, u_{2j}, \ldots, u_{cj}]^T$, we can estimate the ordering of distances $\{\|O_i - X_j\|\}_{i=1:C}$ directly through the reverse ordering of the corresponding membership values. Therefore, the identification of the two highest memberships u_{rj} and u_{qj} in the jth column casts a single vote, associated to the specific vector X_j, regarding the connection between O_r and O_q (i.e. the step-2 in the previous algorithm). By summing up the contributions from all the N vectors, the $\mathbf{G}$ is built.

Let $\mathbf{U}^+$ the matrix produced after replacing the two highest memberships, in each column of $\mathbf{U}$, with ones and zeroing all the remaining elements. Then the computation of $\mathbf{G}$ takes the following compact form:

$$\mathbf{G} = \mathbf{G}(\mathbf{U}) = \mathbf{U}^+ \cdot \mathbf{U}^{+^T}. \tag{5}$$

Actually this formula provides C extra values tabulated in the diagonal of $\mathbf{G}$. These values report the relative distribution of vectors in the different clusters and sum up to the value $2N$. They do not carry any direct information about the topological

relations between the prototypes and therefore can be ignored. The overall computation of $\mathbf{G}$ will be later simplified even further, by replacing the $\mathbf{U}^{+}$ matrix with the one produced in a simpler way.

Before introducing this simplification, let us first compare the above equation with Eq. (4). Apart from a scaling factor and the trace operator, there is an obvious correspondence between these two equations. The matrix multiplication $\mathbf{U}.\mathbf{U}^{\mathrm{T}}$ results in a $C \times C$ matrix with the off-diagonal elements having the form

$$[\mathbf{U}.\mathbf{U}^{\mathrm{T}}]_{rq} = \sum_{j=1}^{N} u_{rj}u_{qj},$$

which is a known measure of the relative overlap of the fuzzy class intersection between the rth and qth fuzzy clusters [15]. It is a point-by-point correlation of the membership values corresponding to the two clusters. As this summation extends to all the vectors X_j, $j = 1, \ldots, N$, even those that have very small memberships to either of the two clusters, it has the tendency to produce nonzero values for all the cluster pairs. On the other hand, the point-by-point correlations implied by Eq. (5) produce useful topological measures since they follow a 'hardening'-step applied to each $\mathbf{U}_{(:,j)}$ vector. Inspired by this parallelism between topology-representation and fuzzy-classification metrics, the following algorithmic procedure is suggested for a computationally economical approximation of the connectivity graph $\mathbf{G}$:

The fuzzy connectivity graph (FCG):

step-1: Apply FCM algorithm, $(\mathbf{O}, \mathbf{U}) = \mathrm{FCM}(\mathbf{X}, C)$
step-2: Build $\mathbf{U}' = [u'_{ij}]_{C\times N}$ such that $u'_{ij} = u_{ij}.\theta(u_{ij} - \tau)$
step-3: $\mathbf{G}' = \mathbf{G}'(\mathbf{U}, \tau) = [g_{ij}']_{C\times C} = \mathbf{U}'.\mathbf{U}'^{\mathrm{T}}$

In this pseudocode, the first step denotes the application of standard FCM-algorithm using a high number C, i.e. in an over-clustering fashion. The second step denotes a simple thresholding scheme in which all the membership values below a user-defined threshold $0 < \tau < 1$ are zeroed. Hence, they do not contribute in the subsequent estimation of correlations via the execution of the third step. We have experimentally verified that (with a proper selection of the τ) the fuzzy connectivity graph (FCG) bears almost equivalent information, regarding the topological description of the data, with the $\mathbf{G}$. Apart from its appealing simplicity, the introduced approximation is of practical utility in the case of high N. In what follows (except Sections 5.1 and 5.2.) when we are talking about connectivity graph we mean its fuzzy approximation G' and, hence, the apostrophe is dropped hereafter.

4. Graph-based post-processing of the FCG

The FCG contains information about the intrinsic data topology, since each entry G_{ij} quantifies the adjacency between prototypes in the sense that its value is directly analogous to the overlap between the corresponding receptive fields. Provided a dense data sampling (i.e. high N), this estimation of adjacency is very reliable and more informative than using the naı̈ve pairwise Euclidean distance which assumes embedding of data in a flat space. In order to exploit the rich structural information contained in $\mathbf{G}$ for the benefit of data organization, we adopted the following graph-theoretic techniques.

4.1. Minimal spanning tree (MST) and MST-ordering

The minimal-spanning tree (MST) is a popular graph-theoretic tool that can provide, for a given connected weighted graph, both the '*nearest neighbor*' information about each node and the '*shortest linkage*' information about subsets of vertices. When used to handle data points, its Euclidean version (eMST) is applied. In this implementation there is a node for each data point on a fully connected graph, while the Euclidean pairwise distances serve as the link weights. The eMST was used in a recent work for organizing the prototypes O_i, $i = 1, 2, \ldots, C$ derived via neural-type vector-quantization [1]. There, a node was associated with the endpoint of each vector O_i and the eMST was the connected graph formed by connecting all these nodes with the unique set of $(C - 1)$ line segments having the minimal total length. MST-graphs, in general, are characterized by the existence of a unique path connecting every pair of nodes that can be followed in order to navigate, systematically, in the graphical structure. Associated with this property of MST, there is a graph-theoretic procedure for transforming this graph to an ordered list, in which similar ranks are assigned to nearby nodes. The so-called *MST-ordering* is based on a standard technique for traversing a tree-graph and accomplished by selecting one node of the MST as the '*root*' and following a *breadth first search* for the remaining ones [27]. The rank of each node is the order in which it is visited in this search (see Fig. 3).

Having in mind to emphasize the intrinsic geometrical relations, the above ideas are employed here by first applying standard Prim's algorithm [26] to a dissimilarity-matrix $^{-}\mathbf{G}$ derived from the FCG as

$$^{-}\mathbf{G} = [^{-}g_{ij}]_{C\times C} : {}^{-}g_{ij} = \max_{i,j}(g_{ij}) - g_{ij}. \tag{6}$$

The derived MST is a set of $C-1$ links that can be thought of as the skeleton of data structure. Hence, it constitutes an appealing tool for detecting degeneracies in the approximation of data structure via dimensionality reduction techniques. Any crossing between different links in the new low-dimensional space indicates defects in the representation of local topological relationships between the centroids. In addition the derived MST can be utilized for the rough re-parameterization of centroids. The subsequent MST-ordering provides a seriation $O_{[i]}$ $i = 1, 2, \ldots, C$ that, depending on the selection of 'root', may reveal the most important degrees of variation in our data. This selection can be simplified through dimensionality reduction, as discussed later.

4.2. Laplacian eigenmap (LE) and locality preserving projection (LPP)

LEs belong to the early scientific contributions that founded the field of manifold leaning. Based on ideas from

spectral-graph theory, and in particular the graph-Laplacian, they provide a low-dimensional geometric representation of a data manifold [28]. First, a neighborhood graph is built over the data points. Weights representing the similarity between nodes are then assigned to the formed edges. After that, the Laplacian operator $\boldsymbol{L}$ is applied to the produced graph and finally a geometric embedding is produced, via eigenanalysis, in which associated nodes are mapped to nearby positions. The LE algorithm has a locality preserving character that makes it robust to outlier and noise. Moreover, is known to produce a mapping in which possible clustering tendencies in the data are enhanced. The core idea implemented by this algorithm is such that it can be directly adopted for visualizing the structure of any given weighted graph (as long as positive symmetric weights are denoting the association between nodes) in a coordinate space [29]. Hence, it can be applied to the FCG graph, giving rise to a point diagram denoted hereafter as $\mathrm{FCG_{LE}}$.

FCG-related Laplacian eigenmap (FCG_{LE}):

step-1: Solve the generalized eigenvector problem $\mathbf{L}\ F_i = \lambda_i \mathbf{D} F_i$ with $\mathbf{D} = [d_{ij}]_{C\times C}$: $d_{ii} = \sum_{j=1}^{C} g_{ij} \vee d_{ij} = 0$, $j \neq i$ and $\mathbf{L} = \mathbf{D} - \mathbf{G}$

step-2: Define the order $F_{[i]}$, $i = 1, \ldots, C$ of eigenvectors according to the order of their eigenvalues: $0 = \lambda_{[1]} < \lambda_{[2]} < \cdots < \lambda_{[C]}$

step-3: Output $\mathbf{P}=[p_{ij}]_{C\times r}$: $O_i \rightarrow P_i = (F_{[2]}(i), F_{[3]}(i), \ldots, F_{[r+1]}(i))$

The jth-node of FCG is mapped to a point. Hence, a nonlinear mapping is implicitly performed $O_i \rightarrow P_i$, $i = 1, 2 \ldots C$, from R^p to R^r, in which $r < p$. Specifically the choice $r = 2$ (or even $r = 3$) facilitates the direct visualization of connectivity between the fuzzy groups.

LPP, on the other hand, came with the second wave of advances in manifold learning theory that extended the potential of early techniques by including additional characteristics, like the possibility of extrapolating the mapping to data points outside the initial training set. It was introduced [30] as a linear approximation of LE. In this work we resort to LPP since it engages, apart from the FCG, also the centroids O_i in the computation of low-dimensional map. Due to linearity, the derived map is smoother than LE (a useful property when data structure is characterized more by gradual variations than large discontinuities) and very helpful for selecting the 'root' needed for a successful MST ordering of the FCG. The steps of LPP technique have been adopted as follows:

FCG-related locality preserving projection map (FCG_{LPP}):

step-1: Solve the generalized eigenvector problem $\mathbf{O}^{\mathrm{T}}\mathbf{L}\mathbf{O}F_i = \lambda_i \mathbf{O}^{\mathrm{T}}\mathbf{D}\mathbf{O}F_i$ with $\mathbf{O} = [o_{ij}]_{C\times p} = [O_1|O_2|\ldots O_C]$, $\mathbf{L} = \mathbf{D} - \mathbf{G}$ and $\mathbf{D} = [d_{ij}]_{C\times C}$: $d_{ii} = \sum_{j=1}^{C} g_{ij} \vee d_{ij} = 0$, $j \neq i$

step-2: Define the order $F_{[i]}$, $i = 1, \ldots, C$ of eigenvectors according to the order of their eigenvectors: $\lambda_{[1]} < \lambda_{[2]} < \cdots < \lambda_{[C]}$

step-3: Select the first r eigenvectors and tabulate them as different columns in $\mathbf{A} = [a_{ij}]_{p\times r} = [F_{[1]}, F_{[2]}, \ldots, F_{[r]})$

step-4: Output $\mathbf{P} = [p_{ij}]_{C\times r} = \mathbf{OA}$

The mapping rule from R^p to R^r, defined by matrix $\mathbf{A}$, can be generalized to any vector including the original ones, i.e. $X_i \rightarrow Q_i = X_i\ \mathbf{A}$, $i = 1, 2 \ldots, N$. By doing so, the whole data sample can appear in a low-dimensional space as a way to provide a more vivid picture of the overlap between different clusters (in the spirit of Ref. [24]).

4.3. ISOMAP

The pioneering paper of Tenenbaum et al. [2] introduced ISOMAP, one of the milestones in manifold learning [5]. Graph-theoretic ideas were engaged with a linear dimensionality reduction technique, namely the multi-dimensional scaling (MDS), to discover intrinsically low-dimensional structures embedded in high-dimensional datasets. Starting from a neighborhood graph built over the N dataset items, shortest-path distances are computed and tabulated in a $(N \times N)$ dissimilarity matrix that is then fed to classic MDS. A point diagram with N elements is then computed, via eigenanalysis, in a space of low-dimensionality r (usually 2 or 3) as an interface helpful in recovering the essential degrees of freedom in the data. In the standard implementation (available at http://isomap.stanford.edu/), the input data are expected in a relational format. This format implies a fully connected graph with pairwise distances as weights. Then the user can select among two strategies for trimming redundant links from the original graph. In the simplest case, the links with weights above a threshold $\boldsymbol{\varepsilon_o}$ are removed and remains a sparse neighborhood graph, over which graph distances are computed. There is a strong affinity between the notion of a path-preserving representation of data manifolds (see Section 3) and the function of ISOMAP which, in essence, is the unfolding of shortest-path distances in a low-dimensional space. This suggested the engagement of this technique in the visualization of FCG.

FCG-related ISOMAP (FCG_{ISOMAP}):

step-1: Define a small $\boldsymbol{\varepsilon_o}$ such that after trimming a single connected neighborhood-graph is constructed

step-2: Run ISOMAP with the default option for a 10-d coordinate space $[\mathbf{Q}, R] = \mathrm{ISOMAP}\ (^{-}\mathbf{G}, \text{'}epsilon\text{'}, \varepsilon_o)$, $\mathbf{Q} = [q_{ij}]_{C\times 10}$

step-3: Output the first r-rows from $\mathbf{Q}$, $\mathbf{P} = [p_{ij}]_{C\times r}$, $p_{kl} = [q_{kl}]$, $k = 1, \ldots, C$, $l = 1, 2, \ldots, r$

With the first step, we secure that all nodes will be seen in the final r-dimensional point diagram (the selection of threshold $\boldsymbol{\varepsilon_o}$ is not difficult, since there is a corresponding diagnostic in the ISOMAP-routine and moreover its value ranges in (0,1)). Using the dissimilarity matrix $^{-}\mathbf{G}$ of Eq. (6), a 10D point diagram is computed. The successive values of residual variance R indicate the relative improvement in the approximation of $^{-}\mathbf{G}$ (by the point diagram) as dimensionality is increased. These values can therefore facilitate the selection of parameter r used in the final step. As in the case of Laplacian eigenmap, a nonlinear mapping $O_i \rightarrow P_i$, $i = 1, 2, \ldots, C$, such that $r < p$ is performed also with ISOMAP. The latter is expected to perform better when relative smooth manifolds are to be recovered, since a global

view of the data is pursued (via the computation of shortest-path distances).

5. Experiments

In this section some worked examples are included with the aim of clarifying the previously presented ideas and demonstrating the different ways in which can be applied to the data at hand. Having in mind an exploratory-data-analysis scenario, where the user wants to gain some insights to his data and provide a meaningful summarization through graphs, the overall scheme takes the following form (for instance see Fig. 4). We begin by running FCM in an over-partitioning mode. We next construct the FCG and derive the corresponding MST. We then apply LPP in order to select the root for MST-ordering and also sketch the FCG in two dimensions. Finally we perform either LE or ISOMAP and overplot the MST in the reduced space. A color map, that follows the MST-based ranks, is adopted so as to enhance further the visualization of data structure. User–machine interaction can readily 'optimize' the few control parameters, namely the partition size C, the threshold τ and the type/size of dimensionality reduction.

5.1. The topology representing graph G

We commence by providing a few examples of connectivity graph G built with the original procedure. Using the 2D example of Fig. 1, we demonstrate the rich structural information that can be conveyed by such a graph and discuss the role of C. Fig. 1a$_1$ shows the data distribution as thin dots and

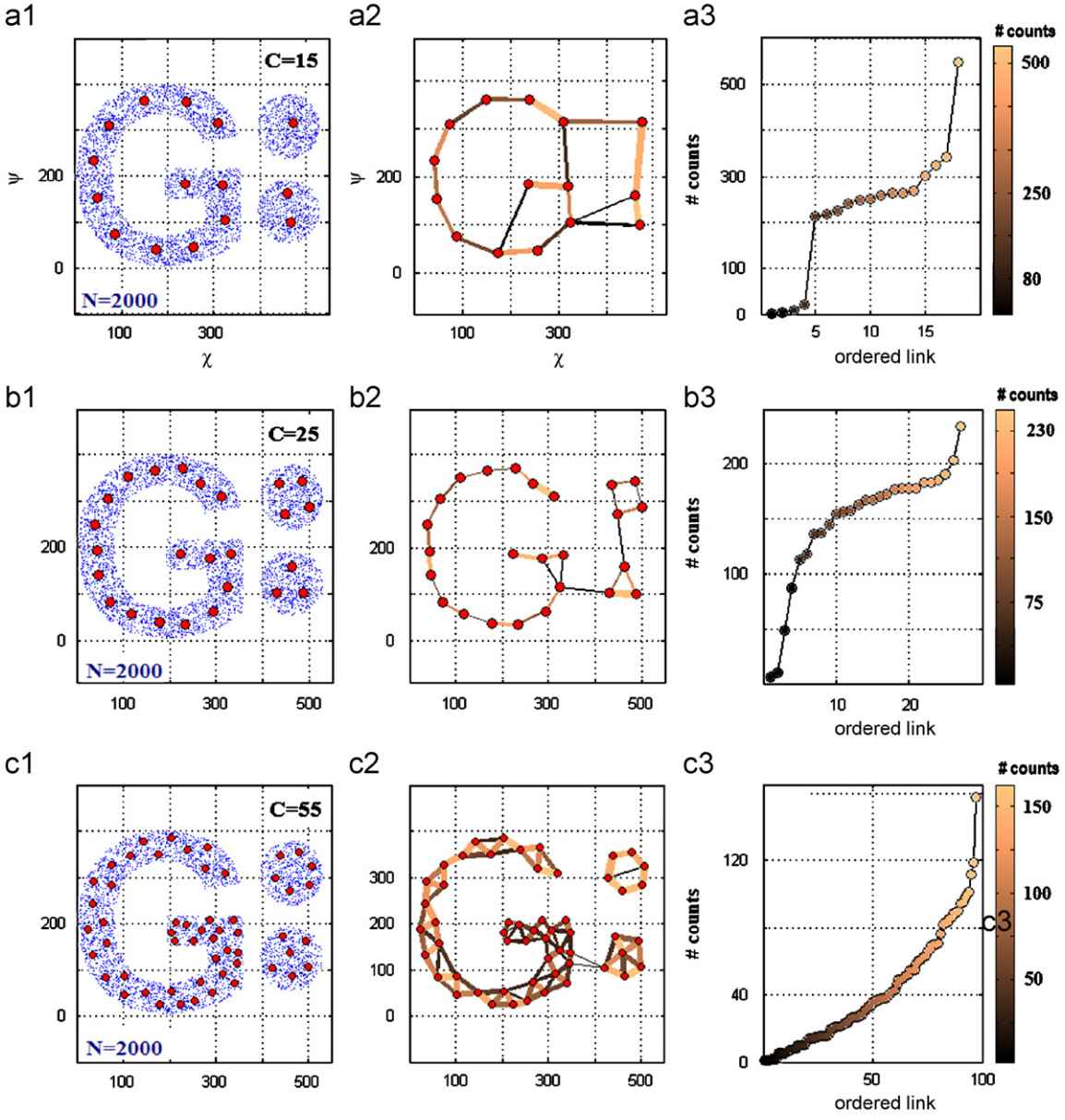

Fig. 1. Examples of connectivity graph G built with varying the number C of fuzzy clusters.

the $C = 15$ computed centroids as tiny spheres. Fig. 1a$_2$ provides the connectivity graph G drawn over the previous centroids. The relative strength of each connection (link) is denoted using a twin code: by the thickness of the corresponding edge and its luminance. Hence, thin dark links correspond to weaker connections. Fig. 1a$_3$ shows the whole distribution of strength for all the detected connections. It has been produced by ordering all entries G_{ij}, $j > i$ and plotting the connectivity strength (expressed in counts) versus the order of the connection. The attached color bar, shows the correspondence between luminance and connection strength and is in common for Figs. 1a$_2$ and 1a$_3$. The middle and lower panels, comprised of Figs. 1b$_{1-3}$ and 1c$_{1-3}$ correspondingly, contain similar information, but for different number of centroids. It is clearly evident (middle column) that as the number C of fuzzy clusters increases, the connectivity graph is cleared from erroneous connections. A sufficiently high number C is therefore needed. On the other hand, a very large C (i.e. in the order of data size N) is not a reasonable choice as well, since both the computational cost and the error in estimating connection strength will increase. The decision on a C that constitutes a sensible trade-off can be assisted by graphs in the form of right-most column. When C is very small (Fig. 1a$_3$) the values of G tend to cover different isolated ranges. As C increases (Fig. 1b$_3$) there is no such isolation. With a sufficiently high C (Fig. 1c$_3$), the strengths tend to distribute more uniformly and within a smaller range. Therefore, as a rule of thumb, we can suggest a value of C such that the connection strengths will tend to distribute uniformly, but without collapsing to a very small range of values. Obviously, whether or not a specific C value is appropriate, can easily be decided through a graph in the form of those seen in the middle column. This is the reason why we suggest, in any case (not only whenever the data are 2D as in this particular example), a similar presentation of interconnections in the reduced 2D-space produced via LPP (see Fig. 4d).

5.2. "The FCG approximates G"

This subsection is devoted to the comparison of original connectivity graph G with its fuzzy counterpart G' regarding the implied structural information. The FCG is computed with the introduced algorithm (see Section 3), in which a threshold parameter τ needs to be set. For both the examples included in Fig. 2 the value $\tau = 0.1$ was used.

Fig. 2 contains two different panels in which the numbering of included graphs has been duplicated since they are in full correspondence. In the upper part, the structure of previous 2D point distribution is studied. For the lower part a uniformly distributed 2D dataset has been used. Fig. 2b portrays the strength of connectivity between the centroids seen in Fig. 2a, while Fig. 2c shows the distribution of the strength values. Both graphs correspond to $\mathbf{G}$ matrix and have been produced as detailed in the previous section. On the other hand, Fig. 2d and Fig. 2e correspond to the matrix $\mathbf{G}$'. The two diagrams have been produced in a similar way with the previous ones. The only technicality needs to be mentioned is that the strength values have been brought to the range (0, 1], with the maximum entry of $\mathbf{G}$' corresponding to 1. The contrast of Figs. 2b and d provides clear evidence about the resemblance between the topological information contained in the two graphs $\mathbf{G}$ and $\mathbf{G}$'.

The comparison of G and G' was also carried out in a more quantitative way. Using the *Hubert's Γ-statistic*, we correlate the structure between the two Graphs derived for several datasets of varying dimensionality and structural complexity. A highly significant similarity was revealed in every case. As it was expected, this similarity was found to depend on the employed threshold τ. However, the optimal choice of τ was depending on the dataset nature and the employed C (to a much lesser extend) as well. Based on extensive experimentation, we can suggest a value in the range (0.085–0.15). The visual inspection of connectivity graphs produced with different τ can provide the near-optimal value resulting in the most interesting structural description. In the sequel, the value $\tau = 0.1$ is everywhere implied.

5.3. Analyzing the FCG from synthetic data

The classical 'two-moons' dataset is used to demonstrate the potential of FCG$_{LE}$ for enhancing clustering tendencies in the data, even when these are obscured by the nonEuclidean nature of the intrinsic topology. The 50 centroids are shown in Fig. 3a along with the original 2D dataset. The visualized connectivity of Fig. 3b clearly indicates the existence of two elongated substructures approaching at a specific site. After choosing the 'root' on a different site, the MST-ordering assigned ranks as seen in Fig. 3c. To enhance visualization even further, the color map seen beside has been used to color the nodes according to MST-based ranks. The result vividly represents that we can navigate in the dataset using the MST-based parameterization (using as a single parameter the rank of each centroid). An even clearer re-parameterization of the dataset can be achieved with the application of FCG$_{LE}$. Fig. 3d shows the derived point diagram in a 2D coordinate space. Please notice that this mapping did not actually realize dimensionality reduction, but rather a nonlinear transformation of the original structure (reduction would have been performed if only $\boldsymbol{r}_1$ dimension had been kept). In the new space the original MST has been appended and thenodes have been assigned the previously defined ranks. It is apparent that the FCG$_{LE}$ is helpful in tearing the 'two moons' apart. Even via simple thresholding along the r_1 dimension the two elongated clusters can be separated, whereas traditional FCM with $C=2$ will be confused around the nearby sites of the two moons since it is dominated by the assumption of spherical clusters.

A second example is discussed next, in which dimensionality reduction is actually mandatory for mediating the structural representation. The dataset is 3D and synthesized from four distinct multivariate Gaussian distributions, with two of them characterized by elongated structure (and containing 1000 vectors) and the others being spherical (and containing 500 vectors).

The original dataset is shown in Fig. 4a together with the 50 derived prototypes. The corresponding FCG-based MST is first portrayed in the original space along with the related ranks

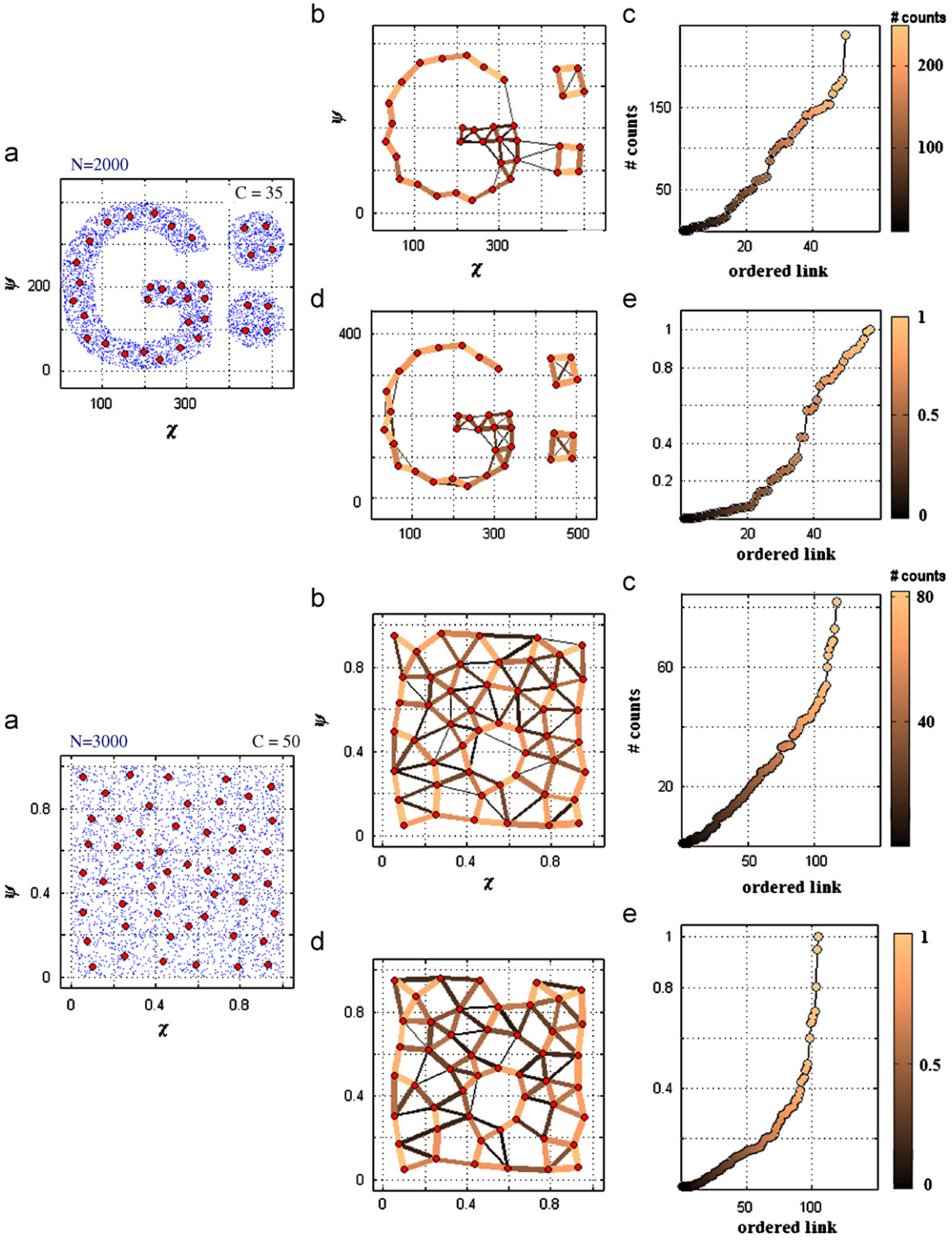

Fig. 2. Comparing connectivity graph G with its fuzzy counterpart FCG.

Fig. 4b. The FCG_{LPP} point diagram is shown in Fig. 4c, with the original MST overplotted. A few crossings between the links of MST can be seen and this indicates that the original structure cannot be perfectly embedded in two dimensions. The same point diagram is used in Fig. 4d in order to visualize the FCG-related connectivity. Augmented by the connectivity links, it makes evident the existence of four distinct components in the original data and their differences in shape. The corresponding FCG_{LE} point diagram is shown in Fig. 4e with the ranks denoted as previously. The later point diagram intensifies clustering even more and facilitates the easy delineation of the four distinct graph components as seen in Fig. 4f. This delineation

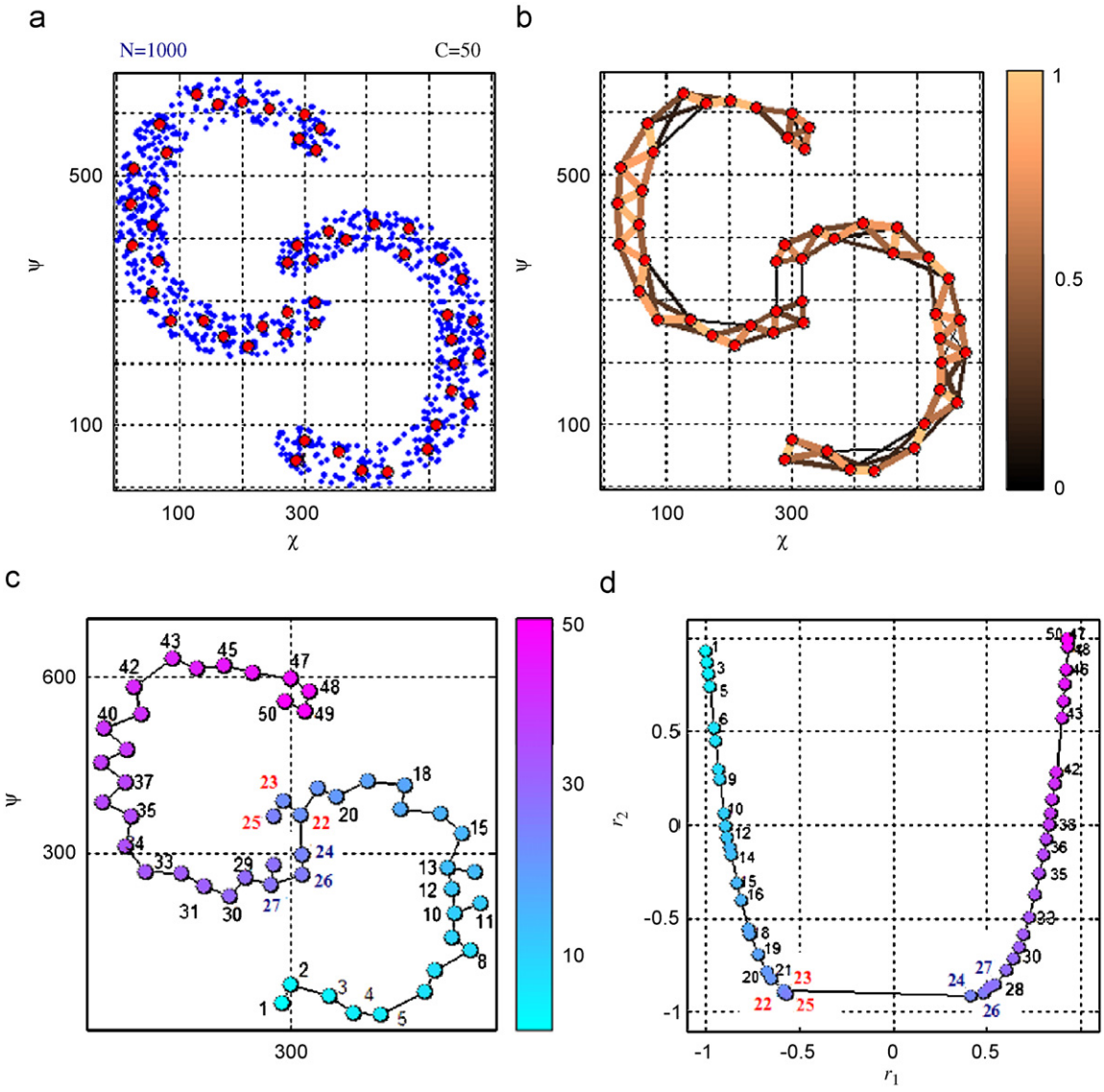

Fig. 3. FCG related analysis applied to the 'two moons' dataset (see Section 5.3).

can be done manually using interactive graphics, through standard clustering applied in the reduced space or alternatively by removing the three weaker edges from the MST graph.Next, with the inverse mapping $P_i \rightarrow O_i$, $i = 1, 2, \ldots, C$, from R^r to R^p, each one of the four swarms of centroids are identified in the original 3D space and used to segment the data as seen in Fig. 4g. The color of the four graph components has been transferred to the initial vectors X_i according to the simple nearest-neighbor classification rule (i.e. *maximum defuzzification* in the fuzzy logic terminology [15]). This result can be directly compared with the output from standard FCM with $C = 4$, which is presented in Fig. 4h in the same format. It is clear that the standard FCM fails to completely separate the elongated from the spherical substructures.

With the third example, we meant to demonstrate the insensitivity to noise. We started by constructing the 2D synthetic dataset of Fig. 5a, that follows *Archimedean spiral* and therefore is characterized by intrinsic dimensionality 1. The original dataset $\mathbf{X}^1$ was of $[4000 \times 2]$ size. With a simple linear transformation, using a matrix $\mathbf{W}_{[2\times 10]}$ containing randomly distributed vectors, we created the following dataset $\mathbf{X}_{[4000\times 10]} = \mathbf{X}^1.\mathbf{W}$. This was equivalent with embedding the original structure in a 10D space. On the top of these data we added Gaussian noise distributed uniformly and independently in all 10 dimensions. Two different perspectives of the resulting dataset are provided in Fig. 5b and Fig. 5c. The FCG-based MST-ordering is indicated in Fig. 5d (original space) and Fig. 5e (reduced space). The connectivity is visualized in Fig. 5f, while the point diagram from LE is provided in Fig. 5g. Specifically the comparison of the last two figurines indicates the basic difference between the $\mathrm{FCG_{LE}}$ and $\mathrm{FCG_{LPP}}$ point diagrams. LPP is a linear technique (although it emphasizes the nonlinear characteristics of a given data structure), therefore it is not expected to recover perfectly the

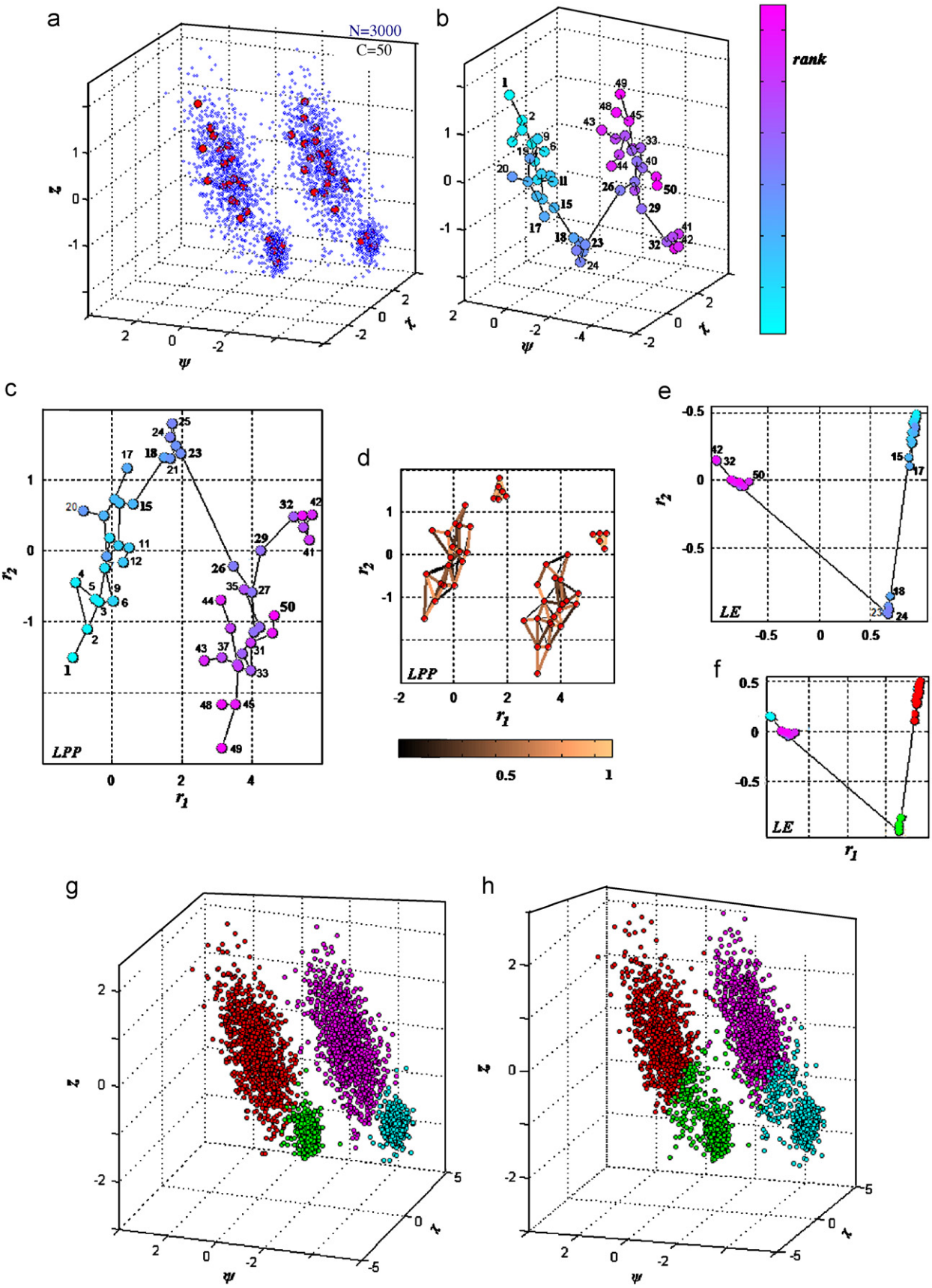

Fig. 4. (a–g) FCG based clustering of 3D data. (h) Standard FCM clustering (see Section 5.3).

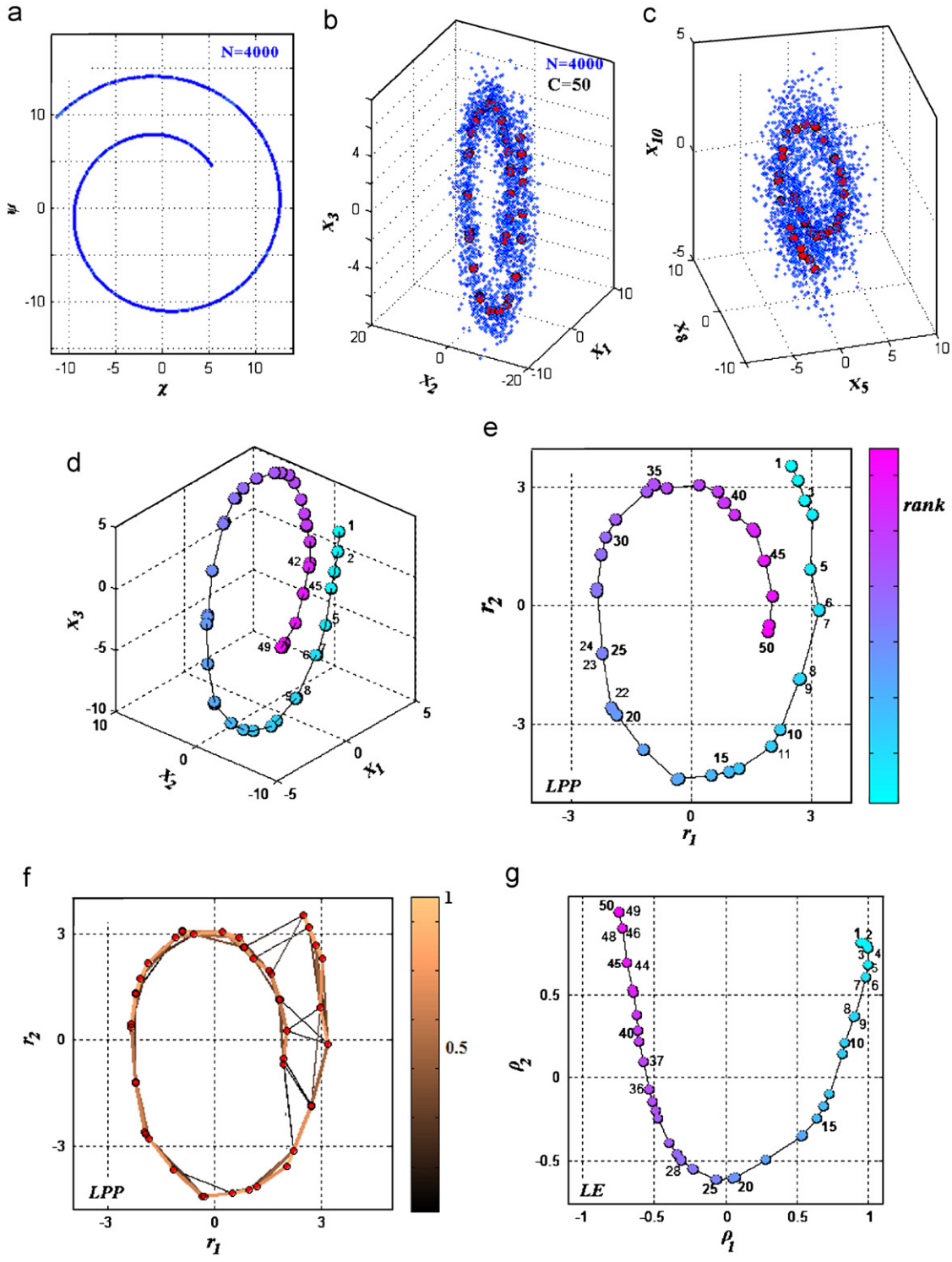

Fig. 5. Parameterization of an intrinsically 1D structure recovered from 10D data (see Section 5.3).

highly nonlinear structure of Archimedean spiral. LE, being of nonlinear nature, goes one step further and unfolds the detected structure in the reduced space. Regardless of this difference, it should be noticed that both techniques effectively portray the main topological characteristics even from data corrupted by noise.

The last included example, using synthetic data, deals with the learning of "*Swiss-roll*" dataset, a standard benchmark in manifold learning theory. The original dataset is shown in Fig. 6a along with $C = 300$ derived centroids. In full accordance with the previously discussed figures, the MST-ordering is depicted in Fig. 6b. The ISOMAP-based representation of

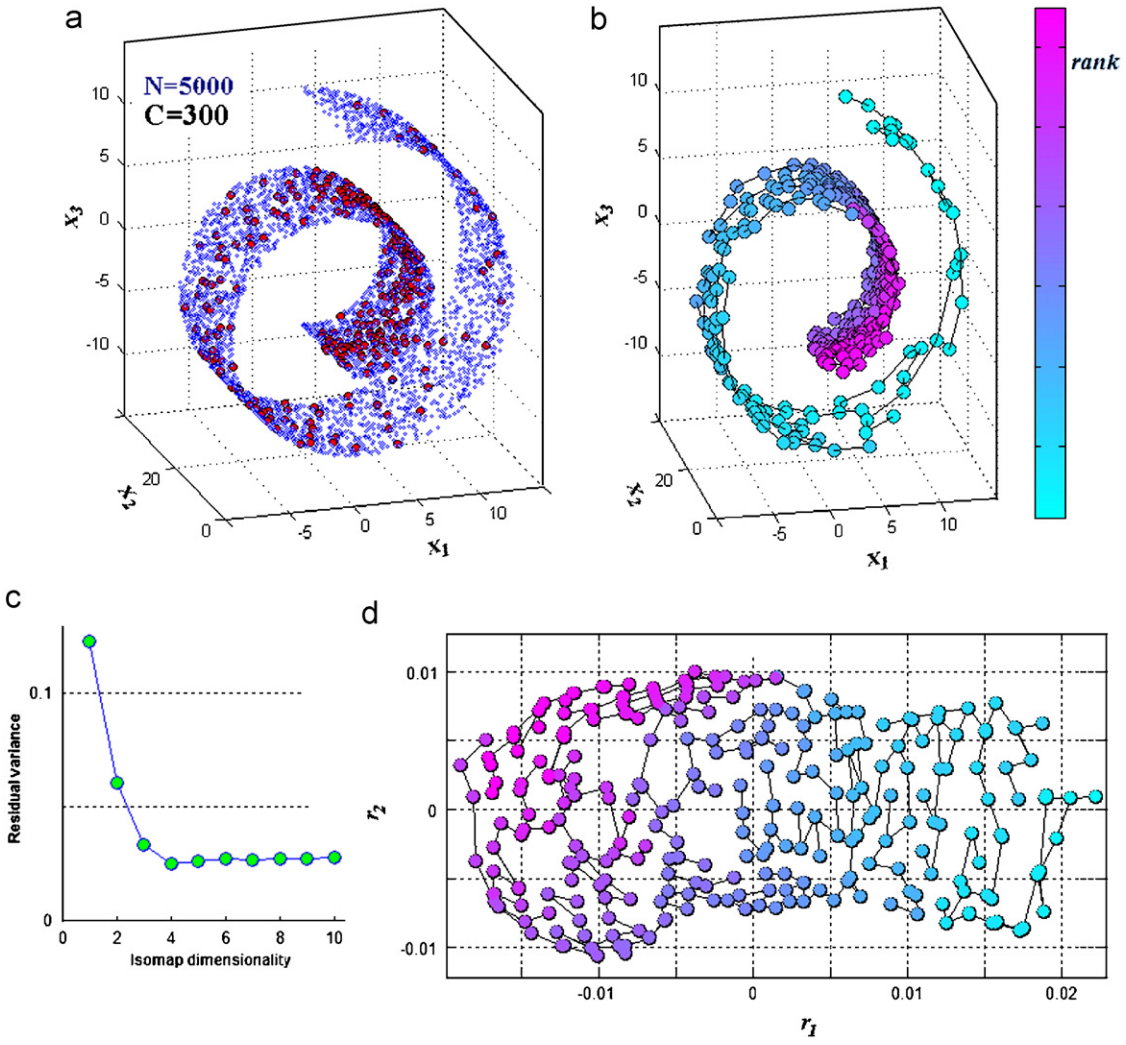

Fig. 6. Unfolding the 'Swiss-roll' data with the FCG-related ISOMAP (see Section 5.3).

the FCG is shown in Fig. 6d, where it is clear that the induced graph flattening faithfully represents the hidden structure. The successive values of residual variance, seen in Fig. 6c, serves as a further indication that two dimensions were sufficient for the representation of graph structure. Two points need to be mentioned here. First, that the corresponding FCG_{LE} mapping (not shown here) was relatively less successful. This can be explained by the smoothness of specific nonlinear manifold. The second point is that ISOMAP could not be applied directly due to dataset size. Using FCG-based ISOMAP, we achieve a meaningful subsampling along with structure representation.

5.4. Application of FCG-analysis to neuromagnetic responses

Finally, we chose to include a realistic example in which the proposed methodology is applied to neuroscientific data with the purpose of knowledge discovery from a large dataset of encephalographic responses. This particular choice was motivated by our own experience with similar techniques applied previously to the same data [1] and facilitated the discussion of relative improvements.

The data correspond to single-trial (ST) responses from a simple visual experiment targeting at the early neuromagnetic response known as *N70m*. A detailed description of the experimental data can be found in Ref. [31]. In short, the MEG signal was recorded continuously with a sampling rate of 625 Hz after low pass filtering at 200 Hz using the 151-channel whole-head Omega biomagnetometer (CTF Systems Inc). The stimulus was a circular checkerboard patch of 4.5° radius and the whole experiment consisted of 240 trials of pattern reversal at the rate of 1.43 Hz. After standard preprocessing, trials were extracted in the range from −100 to 200 ms relative to the onset of each pattern reversal. A spatial operator was applied to the multichannel data in order to extract a single temporal pattern conveying the neural activity from the occipital region, during the ith ST:

$$x_i(t), i = 1, \ldots, \mathrm{N}, \quad t = -100{:}\, T_s{:}\, 200\,\mathrm{ms},$$
$$T_s = 1/f_s = 1/625\,\mathrm{s}$$

A simple data-driven procedure was adopted for extracting features from each temporal pattern which were then used in the detailed study of N70m responses. Based on the pattern of

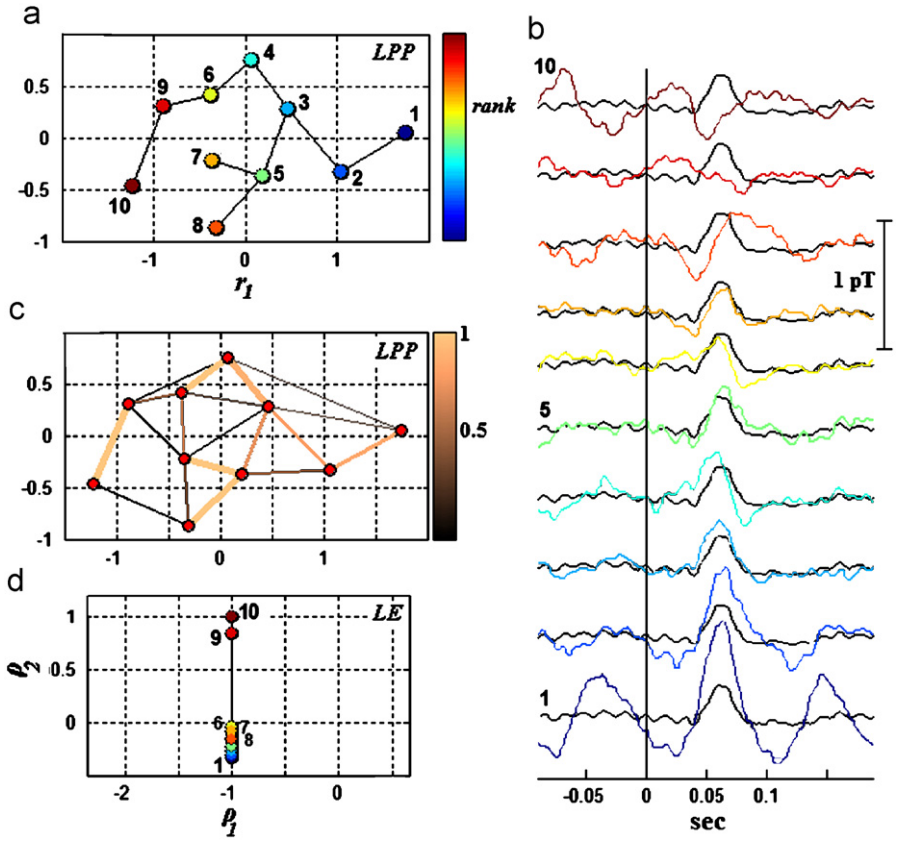

Fig. 7. The visual summary produced via FCG-based prototyping from a dataset of MEG responses (see Section 5.4).

the ensemble average, a set of p latencies around the latency $t_{\max}$ of the N70m peak was defined from the zero crossings around it (see Ref. [1] or [31]). The chain of signal values at these latencies (i.e. the specific segment) constituted the set of extracted features. In this way, the feature vector extracted from the ith ST-pattern $x_i(t)$ was a p-dimensional vector of the form

$$X_i = \left[x_i\left(t_{\max} - \frac{p-1}{2}T_s\right), \ldots, x_i\left(t_{\max} + \frac{p-1}{2}T_s\right)\right],$$
$$t_{\max} = 59.2\,\text{ms}, \quad p = 21.$$

The ST-segments were tabulated, in a corresponding $(N \times p)$ data-matrix $\mathbf{X}$, that can be thought of as containing the different snapshots of regional brain activity during the specific time interval. We need to notice here that dealing with such a dataset is not a trivial task, especially when the study of response variability is among the objectives. At the ST level, the 'true' brain response can be hidden by the brain waves related to ubiquitous ongoing brain activity. Thus, we need to rely exclusively on efficient data-mining techniques in order to organize and represent the related information.

Fig. 7 includes the results from a recording in which the dataset contained $N = 220$ brain responses and the proposed methodology was applied with $C = 10$. The FCG_{LE} point diagram is shown in Fig. 7a. with the MST-ranks being incorporated in the form of both colors and labels. The prototypical brain responses (i.e. temporal patterns) are presented in an orderly fashion in Fig. 7b based on their MST-ranks. They have been derived by first grouping the signals $x_i(t)$, $i = 1, \ldots, N$, according to the membership of the corresponding feature vectors X_i and then performing within-group averaging. Since the two diagrams share the same color code, the graphical representation of dataset variation in Fig. 7a has as natural counterpart the ordered list of signals in Fig. 7b. The FCG-related connectivity graph is visualized in Fig. 7c, while the FCG_{LE} point diagram clearly indicates the presence of two different types of brain responses. The first type (prototypes with ranks 1–8) shows a stimulus related reaction, while the second type (prototypes with ranks 9 and 10) indicates brain activity remained unaltered after the stimulus presentation. In our previous experimentations with different techniques on the same data ([1,31]) we have produced response-variability graphs of similar quality, but we have missed this indication for a bi-level kind of behavior. Such a kind of reaction (i.e. the identical stimulus is delivered but not always perceived by the subject in a repetitive stimulation paradigm) has recently been attributed to the existence of network attractor dynamics governing the spontaneous brain activity and switching between UP and DOWN states [32].

6. Discussion

We suggested the concept of fuzzy connectivity graph FCG and showed that it captures rich topological information

regarding the data structure. We further showed that FCG can be computed, effortlessly, after the execution of standard FCM routine and analyzed, readily, via graph-theoretic techniques ranging from graph-traversal to advanced graph-visualization schemes. Emphasis was put in suggesting easily implemented algorithmic procedures which were built over efficient, widely available routines. Therefore, simplicity and rapid execution, rather than sophistication and mathematical soundness, are the main characteristics of this work. Taking into account the time complexity of FCM, which is $O(NCp)$, it is easily deduced that the total complexity of the suggested methodologies is governed by the complexity of the included spectral procedures. This is $O(C^2)$ for both LPP and LE (using a sparse eigensolver) and $O(C^3)$ for ISOMAP. Hence, with a reasonable selection for $C < 0.1N$ the net computational load can be kept at the level of FCM complexity.

We should notice here, that throughout this presentation it is assumed that FCM has been executed properly. This implies a proper initialization, a check on convergence based on the evolution of J (Eq. (1)), and the avoidance of sparsely populated clusters (the latter situation may appear when C is of the same order with N or when outliers exist in the dataset). A second underlying assumption, inherent to all manifold learning approaches, is the existence of intrinsic data structure. The lack of detectable structure can be signaled by the sparseness of FCG and, mainly, by the similarity of this graph with the corresponding one from a surrogate dataset (produced from the original dataset via randomization [33]).

The core idea of this work is the duality between clustering and manifold learning, which is currently attracting the interest of data analysts (e.g. [34]). Our approach is governed by the philosophy of visual data mining: the user should be able to interact rabidly with graphs summarizing his data with the aim of extracting high-level semantics. Structure visualization schemes akin to the suggested ones have been recently introduced in the setting of Kohonen's map (see Ref. [35] and the papers cited there in). The advantage of our approach is that there are no topological restrictions imposed by a regular grid. Finally, structure visualization has been introduced, lately, in the setting of supervised learning [36]. There the overlap between predefined classes is portrayed in an appropriate reduced space.

This work can be extended in many ways. First, the connectivity mainframe from the FCG can be combined with data-dependent weights from the corresponding pairwise relationships. Second, since the fuzzy partition matrix is the main ingredient different algorithms form fuzzy clustering literature can be employed as well, with the potential to achieve extra characteristics like robustness to impulsive noise. Currently we are experimenting with different ways to test if C is sufficiently high regarding the topology of the given data.

Acknowledgments

N.L. highly recognizes Dr. Law's efforts to make available and keep up-dated the site cited below [5].

References

[1] N. Laskaris, S. Fotopoulos, A. Ioannides, Mining information from event-related recordings, IEEE Signal Process. Mag. 21 (3) (2004) 66–77.
[2] J.B. Tenenbaum, V. de Silva, J.C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (5500) (2000) 2319–2323.
[3] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500) (2000) 2323–2326.
[4] M.H. Law, A.K. Jain, Incremental nonlinear dimensionality reduction by manifold learning, IEEE Trans. Pattern Anal. Mach. Intell. 28 (3) (2006) 377–391.
[5] M. Law, Manifold Learning Resource Page, available at ⟨http://www.cse.msu.edu/~lawhiu/manifold/⟩.
[6] J. Shi, J. Malik, Normalized cuts and image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 22 (8) (2000) 888–905.
[7] A. Ng, M. Jordan, Y. Weiss, On spectral clustering: analysis and an algorithm, in: NIPS, 2001.
[8] L. Zelnik-Manor, P. Perona, Self-tuning spectral clustering, in: NIPS, 2004.
[9] B. Fischer, V. Roth, J.M. Buhmann, Clustering with the connectivity kernel, in: NIPS, 2003.
[10] J.C. Dunn, A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters, J. Cybernet. 3 (3) (1974) 32–57.
[11] J.C. Bezdek, Fuzzy mathematics in pattern classification, Ph.D. Dissertation, Cornell University, Ithaca, NY, 1973.
[12] J.C. Bezdek, Cluster validity with fuzzy sets, J. Cybernet. 3 (1974) 58–73.
[13] J.C. Bezdek, Numerical taxonomy with fuzzy sets, J. Math. Biol. 1 (1974) 57–71.
[14] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York, 1981.
[15] T. Ross, Fuzzy Logic with Engineering Applications, McGraw-Hill, New York, 1995.
[16] W. Wang, Y. Zhang, On fuzzy cluster validity indices, Fuzzy Sets Syst. 158 (2007) 2095–2117.
[17] J. Yang, S. Hao, P. Chung, Color image segmentation using fuzzy C-means and eigenspace projections, Signal Process. 82 (3) (2002) 461–472.
[18] S. Chen, D. Zhang, Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure, IEEE Trans. Syst. Man Cybern. A 34 (4) (2004) 1907–1916.
[19] A.A. Ioannides, G. Kostopoulos, N.A. Laskaris, L.C. Liu, T. Shibata, M. Schellens, V. Poghosyan, A. Khurshudyan, Timing and connectivity in the human somatosensory cortex from single trial mass electrical activity, Hum. Brain Map. 15 (4) (2002) 231–246.
[20] W. Pedrycz, Knowledge Based Clustering: From Data to Information Granules, Wiley, New Jersey, 2005.
[21] R. Krishnapuram, A. Joshi, Liyu Yi, A fuzzy relative of the k-medoids algorithm with application to web document and snippet clustering, in: Proceedings of the 1999 IEEE International Conference on Fuzzy Systems, vol. 3, 1999, pp. 1281–1286.
[22] D. Frossyniotis, C. Pateritsas, A. Stafylopatis, A multi-clustering fusion scheme for data partitioning, Int. J. Neural Syst. 15 (5) (2005) 391–401.
[23] D.-Q. Zhang, S.C. Chen, A novel kernalized fuzzy c-means algorithm with application in medical image segmentation, Artif. Intell. Med. 32 (1) (2004) 37–50.
[24] L. Rueda, Y. Zhang, Geometric visualization of clusters obtained from fuzzy clustering algorithms, Pattern Recognition 39 (8) (2006) 1415–1429.
[25] T. Martinez, K. Schulten, Topology representing networks, Neural Networks 7 (3) (1994) 507–522.
[26] R.C. Prim, Shortest connection networks and some generalizations, Bell Syst. Tech. J. 36 (1957) 1389–1401.
[27] J. Friedman, L. Rafsky, Graphics for the multivariate two-sample problem, J. Am. Stat. Assoc. 76 (1981) 277–287.
[28] M. Belkin, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, Neural Comput. 15 (6) (2003) 1373–1396.

[29] Y. Guo, J. Gao, Paul Kwan, Kernel Laplacian eigenmaps for visualization of non-vectorial data, in: Proceedings of the 19th ACS Australian Joint Conference on Artificial Intelligence (AI 2006), Hobart, TAS, Australia, 2006, pp. 1179–1183.
[30] X. He, P. Niyogi, Locality preserving projections, in: NIPS, 2003.
[31] N. Laskaris, L.C. Liu, A.A. Ioannides, Single-trial variability in early visual neuromagnetic responses: an explorative study based on the regional activation contributing to the N70m peak, Neuroimage 20 (2) (2003) 765–783.
[32] R. Cossart, D. Aronov, R. Yuste, Attractor dynamics of network UP states in the neocortex, Nature 423 (2003) 283–288.
[33] S. Dudoit, J. Fridlyand, A prediction-based resampling method for estimating the number of clusters in a dataset, Genome Biol. 3 (7) (2002) 0036.1–0036.21.
[34] D. Kushnir, M. Galun, A. Brandt, Fast multi scale clustering and manifold identification, Pattern Recognition 39 (10) (2006) 1876–1891.
[35] K. Tasdemir, E. Merényi, Data topology visualization for the self-organizing map, in: Proceedings of the European Symposium on Artificial Neural Networks (ESANN'2006), Bruges, Belgium, 2006, pp. 277–282.
[36] T. Iwata, K. Saito, N. Ueda, S. Stromsten, T. Griffiths, J. Tenenbaum, Parametric embedding for class visualization, in: NIPS, 2005.

About the author—NIKOLAOS LASKARIS received both the M.S. degree (1995) and Ph.D. degree (1998) from Patras University, Greece. He joined the Laboratory for Human Brain Dynamics in RIKEN-BSI, Japan (1999–2003). Currently, he is a Faculty Member of Informatics Department in AUTH, Greece and his research interests include computational intelligence, soft computing, data mining and their applications in biomedicine and neuroscience.

About the author—STEFANOS ZAFEIRIOU received both the B.Sc. degree in Informatics (2003) and the Ph.D.(2007) from the Aristotle University of Thessaloniki, Greece. He has co-authored more than 30 scientific publications. He is currently serving at Greek Army and his main research interests lie in the development and applications of computational intelligence methodologies.