Computer Methods and Programs in Biomedicine (2004) **76**, 115—129

ELSEVIER

Computer Methods and Programs in Biomedicine

www.intl.elsevierhealth.com/journals/cmpb

Protein synthesis profiling in the developing brain: a graph theoretic clustering approach☆

George Potamias[a,b,*], **Catherine R. Dermon**[c]

[a] *Institute of Computer Science, Foundation for Research and Technology—Hellas (FORTH), Vassilika Vouton, 711 10 Heraklion, Crete, Greece*
[b] *Department of Computer Science, University of Crete, 714 09 Heraklion, Crete, Greece*
[c] *Department of Biology, University of Crete, 71409 Heraklion, Crete, Greece*

Received 9 February 2004; received in revised form 5 May 2004; accepted 6 May 2004

KEYWORDS
Clustering;
Minimum spanning tree;
Neuroscience;
Brain development;
Protein synthesis

Summary Mapping regional brain development in terms of protein synthesis (PS) activity yields insight on specific spatio-temporal ontogenetic patterns. The biosynthetic activity of an individual brain nucleus is represented as a time-series object, and clustering of time-series contributes to the problem of inducing indicative patterns of brain developmental events and forming respective PS chronological maps. Clustering analysis of PS chronological maps, in comparison with epigenetic influences of α_2 adrenoceptors treatment, reveals relationships between distantly located brain structures. Clustering is performed with a novel graph theoretic clustering approach (GTC). The approach is based on the weighted graph arrangement of the input objects and the iterative partitioning of the corresponding minimum spanning tree. The final result is a hierarchical clustering-tree organization of the input objects. Application of GTC on the PS patterns in developing brain revealed five main clusters that correspond to respective brain development indicative profiles. The induced profiles confirm experimental findings, and provide evidence for further experimental studies.

1. Introduction

The complex architecture of the brain creates unique problems for its proper assembly. In vertebrate brains, neurons with similar long-distance connections are aggregated into cerebral nuclei (anatomic modules) and the process of their differentiation involves specification of cell fate, migration of cells, outgrowth of axons, as well as formation and modification of neural circuitry. These developmental events are based on intrinsic genetic programs and epigenetic factors controlling the spatio-temporal pattern of gene expression, to produce proteins—the fundamental components of structural and functional elements in brain tissue. Protein synthesis (PS) indeed underlies long-term events in the developing brain that involve changes in morphology and remodeling [1] and reflects these major morphogenetic processes during ontogeny. Ageing is accompanied by reduced protein synthesis rates [2,3] attributed to age related decline in the rate of elongation of the polypeptide chain. Moreover, protein synthesis is extremely valuable cell function marker, e.g., its stimulation parallels learning paradigm in chick [4], while de-

☆ Supplementary data associated with this article can be found at doi: 10.1016/j.cmpb.2004.05.002.

*Corresponding author. Tel.: +30 2810 391693; fax: +30 2810 391601.
E-mail addresses: potamias@ics.forth.gr (G. Potamias), dermon@biology.uoc.gr (C.R. Dermon).

0169-2607/$ — see front matter © 2004 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.cmpb.2004.05.002

generation and apoptotic events [5] are associated with its suppression.

Avian embryo being an excellent model for developmental studies [6] was used to address the issue of brain developmental plasticity and the epigenetic influence of transmission of signals between differentiating cells. Specifically, the effect of the manipulation of a neurotransmitter system was studied, by application of an alpha 2 (α_2) adrenergic agonist, clonidine. The noradrenergic system, known for its role in learning and memory [7] and its early appearance in developing central nervous system (CNS) [8,9], suggests its significant role in neural plasticity mechanisms.

During the course of development the establishment and refinement of proper connections among neurons concerns a problem of pattern formation. Therefore, brain developmental processes could be approached by a series of sequential events as captured by the chronological PS patterns of the brain nuclei and represented as a collection of time-series.

Expanding interest in data mining and knowledge discovery has contributed to an increase of research awareness in mining time-series data. Mining tasks referring to linear precedence phenomena, i.e., ordering of elements (events) in a sequence as a relation over the time axis, include prediction, characterization, and clustering [10]. In this paper we introduce a novel graph theoretic clustering (GTC) approach and apply it on experimental brain developmental time-series data. The approach is based on the arrangement of the objects in a weighted graph, the construction of the graph's minimum spanning tree (MST), and an algorithm that iteratively partitions the tree. The final result is a hierarchical clustering-tree organization of the input objects.

A special feature of GTC is the combination of different information sources in order to compute the distance between the input objects. Domain background knowledge could be utilized in order to compute the distance between objects and arrange them in the weighted graph. Then, iterative partitioning of the respective MST is done with reference to the original feature-based description of data. This hybrid characteristic makes the whole data analysis process more 'knowledgeable' in the sense that established domain knowledge guides the clustering process.

In the present study, we focus on the discovery of indicative and descriptive patterns in order to 'uncover' hidden relations and yield insight on the order of chronological and topographical maps of avian brain, providing profiling rules that possibly guide its development. The results and their biological interpretation contribute to the identification of hierarchical rules underlying the origin of brain structures and provide possible homologies with the mammalian brain. The induced profiles confirm experimental findings, parallel them with established neurobiological knowledge, and provide evidence for further experimental studies.

2. Background

2.1. Brain development data

The experimental data for the present study concern 30 experimental animals—chick embryos and hatchlings, at different developmental stages (ages). These stages were selected based on data from our previous studies [8]. A total of 57 brain-nuclei were identified and targeted, in agreement to their histological and neurochemical characteristics previously defined [11]. Local cerebral PS activity was measured (refer also to Section 4.1) over six age/time-stamps: E11, E13, E15, E17, E19 (5 embryonic days), and P1 (1 post-hatching day). In Table 1, classification and background knowledge concerning the nomenclature, major anatomy, topography, and density of α_2 adrenergic receptors as well as known function of the 57 targeted brain-nuclei is summarized.

For each brain nucleus the PS respective age's means over all chicks were recorded. The final outcome is a set of 57 time-series in a time-span of the six age/time-points. We refer to this dataset as CTRL. The respective data acquired by clonidine treatment are referred as CLON (refer to Section 4.1 for details on the followed biomedical methods, treatments and protocols).

2.2. Indicative brain developmental profiles: a graph theoretic clustering approach

Mapping of regional cerebral protein synthesis activity in developing animals forms the basis to determine developmentally evolved changes in adult brain. Towards this goal, and representing each brain-nucleus PS chronological pattern as a time-series object, we introduce and apply a novel graph theoretic clustering approach on the collection of these time-series objects. The careful biological interpretation of the resulted clusters would provide the basis for the discovery of indicative cerebral developmental maps of PS activity, the extraction of rules and relationships that may govern normal ontogenetic processes, and corre-

Table 1 Brain nuclei nomenclature and classification

	Brain nuclei	Abbreviation	Class-Type			
			MAD	A2ST	BS	CAT
1.	Archistriatum dorsale	Ad	pallium	low	somatosensory	lateral
2.	Area corticoidea lateralis	CDL	pallium	medium	multimodal	dorsal
3.	Area parahippocampalis	APH	pallium	low	limbic	dorsal
4.	Cerebellar White	WH	cerebellum	low	white-matter	dorsal
5.	Cerebellum Granule cell Layer	GR	cerebellum	low	motor	dorsal
6.	Commissura anterior	CA	NA*	low	white-matter	ventral
7.	Commissura posterior	CP	NA	low	white-matter	dorsal
8.	Cortex piriformis	Cpi	pallium	medium	visual	lateral
9.	Ectostriatum	E	pallium	low	visual	lateral
10.	External Granule layer	EXTGR	cerebellum	medium	motor	dorsal
11.	Fasciculus prosencephali lateralis, a	FPLa	NA	low	white-matter	ventral
12.	Fasciculus prosencephali lateralis, p	FPLp	NA	low	white-matter	ventral
13.	Hippocampus	Hip	pallium	low	limbic	medial
14.	Hyperstriatum ventrale	HV	pallium	high	multimodal	dorsal
15.	Lobus paroflactorius	LPO	subpallium	medium	motor	medial
16.	Locus ceruleus	Loc	pons	high	limbic	dorsal
17.	N. accumbens	Ac	subpallium	high	limbic	medial
18.	N. basalis	Bas	pallium	high	somatosensory	ventral
19.	N. D. Branchium Conjactivum	nDBC	pons	medium	motor	medial
20.	N. dorsolateralis	DL	thalamus	low	somatosensory	lateral
21.	N. geniculatus lateralis, ventralis	GLv	thalamus	medium	visual	ventral
22.	N. intercollicularis	ICo	tectum	high	auditory/vocal	dorsal
23.	N. isthmi pars magnocellularis	Imc	tegmentum	low	auditory/vocal	lateral
24.	N. isthmo-opticus	IO	pons	low	visual	dorsal
25.	N. lemnisci lateralis. intermedia	LLi	pons	low	somatosensory	lateral
26.	N. linearis caudalis	LC	pons	medium	somatosensory	medial
27.	N. mammilaris medialis	MM	hypothalamus	high	limbic	medial
28.	N. mesencephalic lateralis, dorsalis	Mld	tectum	low	auditory/vocal	dorsal
29.	N. nervi oculomotorii	OcM	pons	low	motor	dorsal
30.	N. ovoidalis	Ov	thalamus	low	auditory/vocal	medial
31.	N. pontis lateralis	PL	pons	medium	motor	ventral
32.	N. pontis medialis	PM	pons	medium	motor	ventral
33.	N. preopticus medialis	POM	hypothalamus	high	limbic	ventral
34.	N. reticularis pontis	RPO	pons	low	somatosensory	ventral
35.	N. tegm.pend-ponti., compacta	TPc	tegmentum	medium	motor	ventral
36.	N. Vestibularis, medialis	VeM	pons	medium	motor	dorsal
37.	N.anterior medialis hypothalami	AM	hypothalamus	high	limbic	medial
38.	N.dorsomedialis	DM	thalamus	high	limbic	medial
39.	Neostriatum	N	pallium	low	multimodal	lateral
40.	Neostriatum intermedium	NI	pallium	low	multimodal	medial
41.	Nucleus opticus basalis	nBOR	tegmentum	medium	visual	ventral
42.	Nucleus pretectalis	PT	thalamus	high	visual	dorsal
43.	Nucleus rotundus	Rt	thalamus	low	visual	lateral
44.	Nucleus semilunaris	Slu	tegmentum	low	somatosensory	lateral
45.	Nucleus septalis lateralis	SL	subpallium	high	limbic	medial
46.	Nucleus septalis medialis	SM	subpallium	high	limbic	medial
47.	Nucleus spiriformis lateralis	SPI	tegmentum	medium	visual	lateral
48.	Nucleus subpretectalis	SP	tegmentum	medium	somatosensory	ventral
49.	Nucleus taeniae	Tn	pallium	low	limbic	ventral
50.	Paleostiatum augmentatum	PA	subpallium	low	motor	lateral
51.	Paleostriatum primitivum	PP	subpallium	high	motor	ventral
52.	Str. griseum et fibrosum super, sup	SGFSs	tectum	high	visual	lateral
53.	Str.griseum fibrosum. super deep	SGFSd	tectum	medium	visual	lateral
54.	Stratum album centrale	SAC	tectum	low	white-matter	lateral
55.	Stratum griseum centrale	SGC	tectum	medium	visual	lateral
56.	Substantia grisea centralis	GCt	tectum	high	auditory/vocal	medial
57.	Tractus Opticus	Tov	NA	low	white-matter	ventral

The targeted class-types (utilized in the present study) and their values are shown; **MAD**—'major anatomic divisions'; **A2ST**—'A2 synaptic transmission'; **BS**—'brain system'; and **CAT**—'cerebral axes topography'.

late critical periods with specific structures during development.

The GTC clustering approach is based on the arrangement of the objects in a weighted graph, the construction of the graph's minimum spanning tree, and an algorithm that iteratively partitions the tree. The final result is a hierarchical clustering-tree organization of the input objects. With GTC there is no need to specify the number of clusters in advance (a prerequisite of other clustering approaches such as *k*-means [12]). In contrast, a 'termination' condition, implemented with an information-theoretic formula, is applied on each of the nodes of the growing clustering-tree and decides to stop or, to further expand the tree at that node.

2.3. Related work

MST-based clustering is not a new idea. It was first introduced by Zahn [13] and Page [14]. Recently, a similar approach that follows a different partitioning strategy was also introduced and applied on gene-expression profiling tasks [15]; the method is implemented in the core of the EXCAVATOR gene-expression analysis system (http://www.apocom.com/geneexpression.html).

These approaches follow a 'one-shot' MST partition strategy with the identification of 'weak' (or 'long') MST edges, which are then cut. Because of their one-shot partitioning strategy these methods could not identify special relations in the data as for example the potential of a hierarchical organization. In addition, all approaches demand the presetting of the number of desired clusters. In most cases such a demand is problematic, especially in exploratory data analysis where, the analyst possesses no hints about the potential number of clusters. For the approach in [15] an estimate for the optimal number of clusters is computed in advance, a pre-processing step of high computational cost.

Moreover, GTC exploits a 'hybrid' characteristic. Assuming that the assignment of objects to classes is known in advance then, the VDM metric (refer to Section 4.2.2) is used to utilize information that comes from external (to the feature-based description of the objects) modality. The clustering is to be performed on a (potentially) different distance-based arrangement of the object, and the final hierarchical clustering outcome reflects both: (a) the feature-based description of the objects (in our case the brain developmental PS profiles), and (b) their class assignments. So, conjectures made from one source of information may be used to confirm (or, reject) conjectures from the other, and vice versa. In this setting, pre-established domain-knowledge is utilized in order to discover regularities and confirm/reject hypotheses. In that sense, GTC presents a 'knowledgeable' exploratory data analysis approach. This is in contrast to other MST-based clustering approaches where, the computation of distances between objects relies solely on the feature-based description of the objects and the corresponding 'geometric' arrangement of them. In this mode, clustering is not coupled with background domain knowledge, a crucial source of information in order to decide where to cut the MST (especially for 'borderline' cases).

3. Design considerations

Our aim is to discover indicative and characteristic patterns in the developing brain. In this context, we rely on the history of in vivo PS activity of specific brain areas. Profiling the respective developmental patterns may yield insight on their maturation patterns, and reveal relationships between distantly located structures.

Towards this goal we follow a clustering methodology in order to induce groups of brain-nuclei that exhibit similar PS chronological profiles. With the careful inspection of the induced clusters, and their relationship, researchers in the field (i.e., neuroscientists) may uncover and reveal reliable brain developmental models that not only confirm established neurobiological knowledge but also, provide hints for further experimental studies.

Clustering aims to group together objects with similar properties. This can also be viewed as the reduction of the dimensionality of the system or, the discovery of 'structure in the data'. Here, we present a novel graph theoretic clustering approach, as adjusted for time-series data analysis that follows two basic steps.

Step 1: Assume a feature-value description of the input objects. In the case of the brain development domain the features are the protein-synthesis stamped time-points. Following a dynamic discretization procedure the continuous value of each time-point is assigned to a discrete nominal value. The distances between all the discretised time-series are computed. The distances may be computed taking in consideration various modalities. For the brain development data the distance may reflect different class-type assigned to the brain nuclei, just like the ones shown in Table 1 (refer also to Section 2.1 for details). Currently, GTC employs class assign-

ment information in the computation of the time-series distances and realizes this information by the utilization of a special distance measure, the VDM metric (refer to Section 4.2.2).

Step 2: A fully connected weighted graph is devised with the objects as nodes, and edge-weights the computed distances. The minimum spanning tree of the graph is computed and formed. The MST reserves the minimum distance between the objects in a way that low-distant objects are arranged in neighboring areas of the tree. Then, the MST is cut to sub-trees following an iterative partitioning algorithm that concludes to the hierarchical clustering of the input objects.

4. Biomedical and computational methods

4.1. Biomedical methods and data acquisition

In most studies the measurement of rates of protein synthesis is based on the use of radiolabeled precursors [16,17]. The late embryonic development between days 11 (E11) and 19 (E19) as well as the post hatching day 1 (P1) were studied. For the study of epigenetic effect of manipulation of noradrenergic system, clonidine (alpha 2 agonist; catapressan) was administrated (100 ng/kg) in the air sac of the egg every second day at embryonic days E4, E6, E8, E10, E12, E14, E16, E18. Control animals were injected with saline. Protein synthesis activity was studied 36 h after the last clonidine application. Our study determines the relative incorporation of radioactivity, since the specific activity of the precursor amino acid pool [17] is not known for avian species.

Each subject received an intraperitoneal injection of l-[1-^{14}C] leucine (56 μCi/mmol; Amersham) at a dose of 100 μCi/kg in sterile saline. At the end of 60 min experimental time the brain was immediately dissected out of the skull, frozen on dry ice, and stored at −75 °C until sectioned for autoradiographic experiments. All brains were cut in the coronal plane with a Leica cryostat at −20 °C in 20 μm thick sections. Sections were fixed overnight in 30% formalin, washed under running water for 2 h, dried, exposed to Amersham ^{14}C-sensitive autoradiographic Hyperfilm, along with a set of ^{14}C-methylmethacrylate standards (Amersham), as described in [18]. A series of adjacent sections was counterstained with cresyl violet for cytoarchitectonic identification, using nomenclature based on atlas [19] and the ''Avian Brain Nomenclature Exchange'' web site (http://jarvis.neuro.duke.edu/nomen/index.html). After 3 weeks of exposure, the films were developed and 57 discrete brain structures were analyzed with an image analysis system (NIH Image; http://rsb.info.nih.gov/nih-image/), as described in [19]. Each structure was outlined and measured in four to six consecutive sections depending on its antero-posterior extent. In part (a) of Fig. 4, the basic components of the followed biomedical methodology are illustrated.

4.2. Computational methods

4.2.1. Time-series discretization

Discovering sequential relationships in a time sequence is important to many application domains. In data mining applications, and especially in prediction and clustering tasks, it is often necessary to search within a series collection for time-series that matches a pre-specified query series. During the last years a great-deal of work is devoted on such research aspects [20,21].

Measuring the distance between objects is a crucial issue in many data retrieval and data mining applications. The typical task is to define a distance function dist(a,b) (the dual problem is to define a similarity function), between two sequences a and b, which represents how 'distant' ('similar') they are to each other. A simple starting point would be to measure the distance of time-series by using a normal distance metric (e.g., Euclidean). But, for time-series this way of measuring distance is not appropriate, since the sequences can have outliers, different scaling factors, and baselines. As it is noted in [20], reliable time-series matching and clustering operations should take in consideration the following functions: (i) ignore small or not-significant parts of the series; (ii) translate the offset of the series in order to align them vertically; and (iii) scale the amplitude of the series so that each of the respective segments lies within an envelope of fixed width.

The above problems could be tackled by discretising the series. That is, each value of a time-series is transformed into a representative nominal one. In the present work, we follow and adjust the qualitative dynamic discretization (QDD) method presented in [22]. The basic idea underlying the QDD discretization method is the use of statistical information about the preceding values observed from the series in order to select the discrete value that

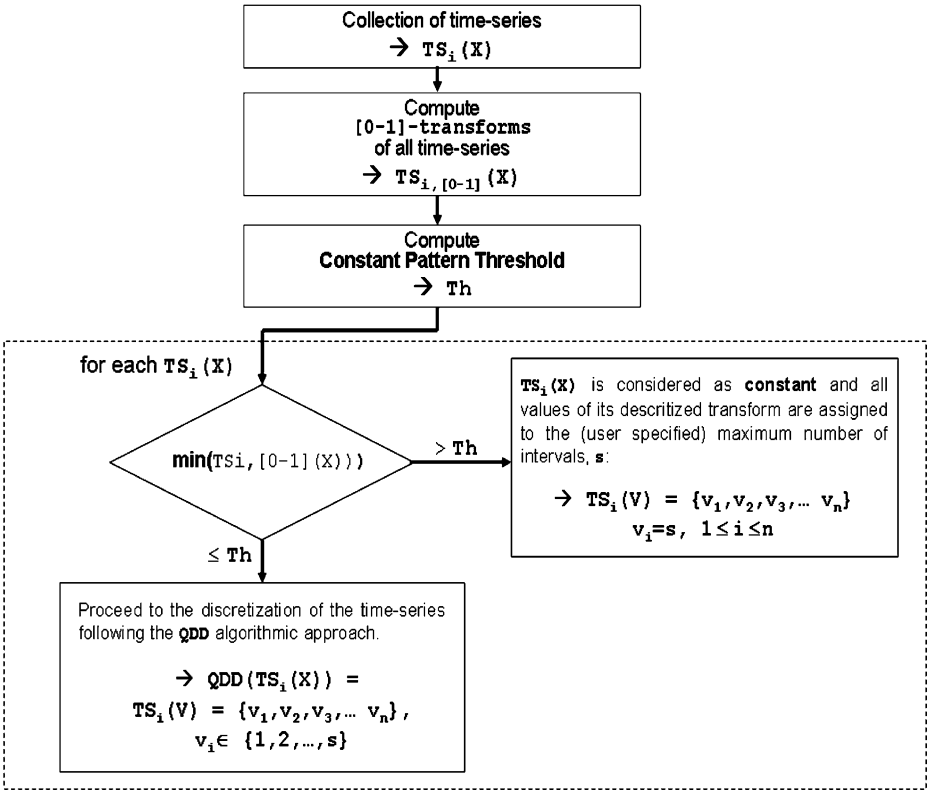

Fig. 1 Illustration of the overall time-series discretization process and the formation of the respective time-series discretized transforms.

corresponds to a new continuous value from the series. A new continuous value will be assigned to the same discrete value as its preceding values if the continuous value belongs to the same population (to be decided with a Student's *t*-statistic). Otherwise a static discrete transformation measure will assign a new discrete value to the continuous one. The overall time-series discretization process is illustrated in Fig. 1.

4.2.1.1. Constant patterns: coping with 'insignificant' changes. With the QDD method it is very-difficult to model 'constant' time-series, i.e., series with values fluctuating in 'small' or, insignificant ranges (see Fig. 2). We refined and enhanced the QDD method by computing a threshold value in order to decide if the series should be considered as constant or not (internal rectangle of Fig. 2). First, for each series $TS(X) = \{X_1, X_2, \ldots, X_m\}$ its [0,1]-transform is computed, $TS_{[0-1]}(X) = \{X_{1,[0-1]}, X_{2,[0-1]}, \ldots, X_{m,[0-1]}\}$. This is done by dividing all the values of the series by the series' maximum value so that the values of the series range in the [0,1] interval. Then, we use the formula below to compute the threshold value.

$$Th = \max(\min(TS_{[0-1],i})) - \text{S.D.}(\min(TS_{[0-1],i})) \quad (1)$$

In the computation of Th all the input time-series are considered. So, $\max(\min(TS_{[0-1],i})$ is the maximum of the list of all time-series' minimum values, and $\text{S.D.}(\min(TS_{[0-1],i}))$ is the standard deviation of this list. For each time-series a test is applied that identifies a time-series as 'constant' or not. If the minimum value of the time-series [0—1]-transform is greater than the computed threshold then the series is considered as constant, and the discrete value *s* (i.e., the user specified number of discrete values) is assigned to all of its values ('if' condition in Figs. 1 and 2). Otherwise the discretization process is triggered (the QDD algorithm) where, the continuous values of the series are assigned to respective nominal values.

4.2.2. Time-series distances

The distance between two time-series, $TS_a(X)$ and $TS_b(X)$, of m time-points both, is computed by the distance between their corresponding discretized transforms, $TS_a(V)$ and $TS_b(V)$.

$$\text{distance}(TS_a(X), TS_b(X)) = \text{distance}(TS_a(V), TS_b(V)) = \frac{\sum_{j=1}^{m} \text{dist}(v_{a,j}, v_{b,j})}{m} \quad (2)$$

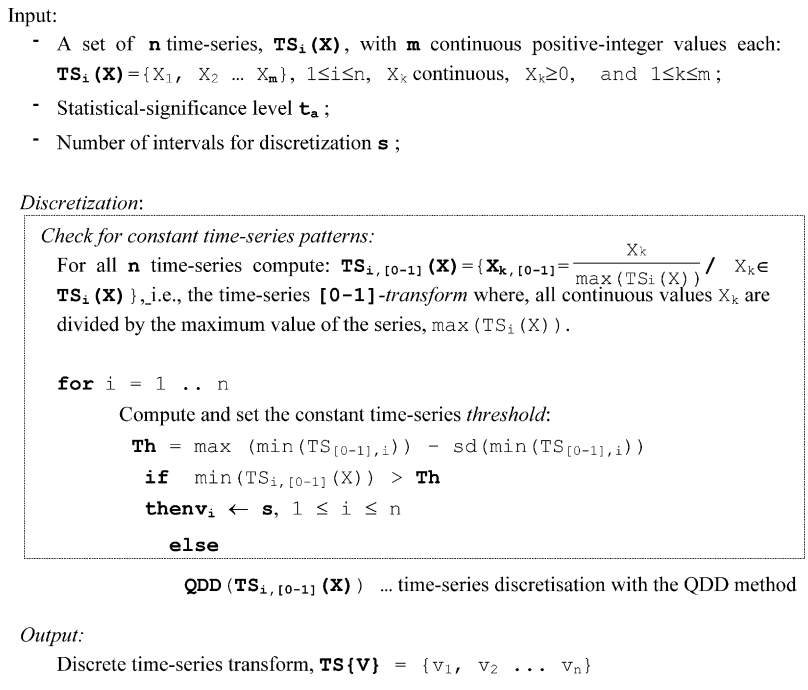

Fig. 2 Details of the time-series discretization process focusing on the identification and formation of constant time-series patterns (internal rectangle).

The 'dist' function, in formula 2 above, between two nominal values, $v_{a,j}$, $v_{b,j}$, may be computed utilizing different metrics. Below the 'NOM_dist' metric is given.

$$\text{NOM\_dist}(v_{a,j}, v_{b,j}) = \begin{bmatrix} 1 & \text{if } v_{a,j} \neq v_{b,j} \\ 0 & \text{otherwise} \end{bmatrix}$$

The current GTC implementation incorporates a variety of other time-series distance computation approaches (refer also to Section 5). Among others the value difference metric (VDM) is utilized and implemented.

4.2.2.1. The VDM distance: a knowledgeable metric. VDM combines information about the input objects that originates from different modalities. For example, the a priori assignment of brain nuclei to specific class-type values (see Table 1) could be utilized. The VDM distance metric, given by formula 3 below, takes into account this information [23].

$$\text{VDM}_a(V_a = x,\ Va = y) = \sum_{c=1}^{C} \left| \frac{N_{a,x,c}}{N_{a,x}} - \frac{N_{a,y,c}}{N_{a,y}} \right|^2 \qquad (3)$$

where $V_a = x$: x is the value of feature a; $N_{a,x}$ the number of objects with value x for feature a; $N_{a,x,c}$ the number of class c objects with value x for feature a; and C is the total number of classes.

Using VDM we may conclude into a distance arrangement of the objects that differs from the one that results when the used distance-metric does not utilize objects' class information. So, the final hierarchical clustering outcome will confront not only to the distance between the feature-based description of the objects but to their class resemblance as well. As the assignment of classes to objects reflect to some form of established domain knowledge the whole clustering operation becomes more 'knowledgeable'.

4.2.3. Graph theoretic clustering

Having on our disposal two different sources of information (a) the feature-based description of the objects, and (b) the knowledge depended distances between them; the question is how we utilize both sources of information in order to form a reliable clustering of the objects. Towards this target, we elaborate on an innovative graph theoretic clustering approach realized by the following procedures.

(a) *Minimum spanning tree construction*: Given a set E of n objects, the minimum spanning tree of the fully-connected weighted graph of the objects is constructed; the formed MST contains exactly $n - 1$ edges. In the current GTC implementation we use Prims' method for the construction of the MST [24]. A basic characteristic of the MST is that it reserves the shortest distances between the objects. This guarantees that objects lying in 'close areas'

of the tree exhibit low distances. So, finding the 'right' cuts of the tree could result in a reliable grouping of the objects.

(b) *Iterative MST partition*: It is implemented within the following three steps.

Step 1: *Binary splitting*. At each node (i.e., sub-cluster) in the so-far formed hierarchical tree, each of the edges in the corresponding node's sub-MST is cut. With each cut a binary split of the objects is formed. If the current node includes n objects then $n-1$ such splits are formed. The two sub-clusters, formed by the binary split, plus the clusters formed so far (excluding the current node) compose a potential partition.

Step 2: *Best split*. The *category utility* (CU) (by formula 4, exemplified into the sequel) of all formed $n-1$ potential partitions are computed. The one that exhibits the highest CU is selected as the best partition of the objects in the current node.

Step 3: *Iteration and termination criterion*. Following a depth-first tree-growing process, steps 1 and 2 are iteratively performed. The category utility of the 'current' best partition, $CU_{current}$, is tested against the 'so-far' formed clusters, CU_{so_far}. If $CU_{current} > CU_{so\_far}$ then, the node is split (Step 1), otherwise we stop further expansion of the current clustering-tree node.

The final outcome is a hierarchical clustering tree where (by default) the termination nodes are the final clusters. After visual inspection of the hierarchical tree the user may decide to use higher levels of the tree as the final clusters. Note that there is no need to determine the number of clusters in advance—a task left to the node growing/termination criterion (Step 3). As an example of the GTC output, in Fig. 3 we show the MST and the clusters that were induced for the brain development time-series data (see Section 6). The tree was plotted with the aid of the GraphViz/dot graph-visualization software (http://www.graphviz.org/).

For the computation and estimation of the utility that each set of clusters exhibits we rely on the established and well-known category utility formula [25]. The CU metric resembles an information theoretic one, and it is based on the distribution of the objects' feature-values in a set of object groups $\{G_1, G_2, \ldots, G_g\}$.

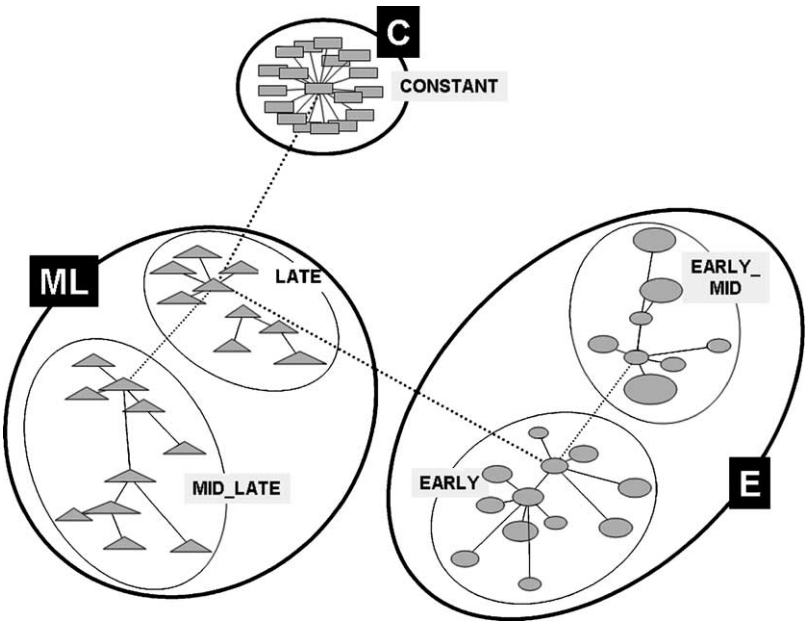

Fig. 3 The minimum spanning tree (MST) and the resulted clusters for the neurophysiologic CTRL experiment (see Section 5). The final set of clusters was induced applying the following GTC parameterization: three discretization intervals; 99% statistical significance threshold; and at least 14% of the total number of objects in each cluster. Dotted lines indicated clusters' separation; bolded ovals indicate high-level clusters of the clustering-tree (i.e., clusters 'E', 'ML', and 'C'); and ovals indicate the respective sub-clusters (i.e., the final low-levels of the clustering-tree).

$$CU(G_1, G_2, \ldots, G_g) = \frac{\sum_{k=1}^{g} p(G_k)\left[\left(\sum_i \sum_j p(A_i = V_{ij}/G_k)^2\right) - \left(\sum_i \sum_j p(A_i = V_{ij})^2\right)\right]}{g} \quad (4)$$

where

$$p(G_k) = \frac{\text{number\_of\_objects\_in\_}G_k}{\text{number\_of\_total\_objects}}$$

$$p(A_i = V_{ij}/G_k) = \frac{\text{number\_of\_objects\_in\_}G_k\text{\_with\_value\_}V_{ij}\text{\_for\_feature\_}A_i}{\text{number\_of\_objects\_in\_}G_k}$$

and

$$p(A_i = V_{ij}) = \frac{\text{number\_of\_objects\_with\_value\_}V_{ij}\text{\_for\_feature\_}A_i}{\text{number\_of\_total\_objects}}$$

4.2.4. Assessing the utility of background knowledge

As it was already mentioned, a unique GTC feature is the ability to utilize domain background knowledge and guide the clustering process in a 'knowledgeable' way. With this feature it is possible to test the 'fitness' of the utilized background knowledge to the discovered clusters. In other words, we may access the degree to which the experimental data parallels, i.e., confirms or, rejects specific domain theories.

We assume that the utilized background knowledge comes in the form of class pre-assignment to the input objects—just like the class-types to which the targeted brain-nuclei belong (refer to Table 1). To do this, we introduce the *Impurity Index* metric, given by formula 5, below:

$$II_C = \frac{\sum_{c \in C} II_{C,c}}{|C|} \quad (5)$$

The impurity index, II_C, is based on the well known 'diversity index' formula [26] and it measures the descriptive power of a cluster with respect to the classes assigned to the input objects. In other words, it helps to give answers to questions like: ''*how probable is to find mostly objects of a specific class-type in a specific cluster*''.

The impurity index of a class-type C, II_C, is computed as the average over all the impurity-indices of the class-type's values c, $II_{C,c}$ ($|C|$ is the cardinality of class-type C). The impurity-index of a class-type value is given by the following entropic formula:

$$II_{C,c} = -\sum_k p(C = c|k)\log(p(C = c|k)) \quad (6)$$

where k ranges over all the induced clusters, and $p(C = c|k)$ the conditional probability of class-type value c given cluster k, i.e., the distribution of class-type value c in cluster k.

5. System description

The current GTC version is implemented in SWI Prolog (http://www.swi-prolog.org). The implementation provides a batch of distance-computation metrics such as: the normal and square-rooted 'Euclidean' metrics; the 'Pearson' linear- and rank-correlation metrics; the 'Edit' distance metric, the 'NOM' distance (refer to Section 4.2.2), and the VDM metric.

Moreover, some extra features are implemented that let the user: (i) to select and focus the clustering analysis on 'parts' (i.e., periods) of the input time-series; (ii) to specify the statistical significance level being used in the time-series discretization procedure; (iii) to import a distances-file and not to compute distances from the feature-based description of the input objects; and (iv) to specify the percentage minimum number of objects (relative to the total number of input objects) that each sub-cluster is required to include—a quite useful feature that controls the generalization level of the clustering-tree.

In part (b) of Fig. 4 we illustrate the basic steps and components for the followed computational methodology.

5.1. Time complexity

The core of GTC (i.e., the MST iterative partitioning) time-complexity depends: (i) on the complexity of computing the category utility indices, and (ii) on the depth of the resulted clustering tree. Denote with F, the number of features; V, the mean number of values per feature and n, the total number of input objects. The category utility computation needs a time linear to the total number of the features' values, $\sim O(F \times V)$.

In the worst case the maximum depth of the tree is $n - 1$. That is, at the zero level (i.e., all ob-

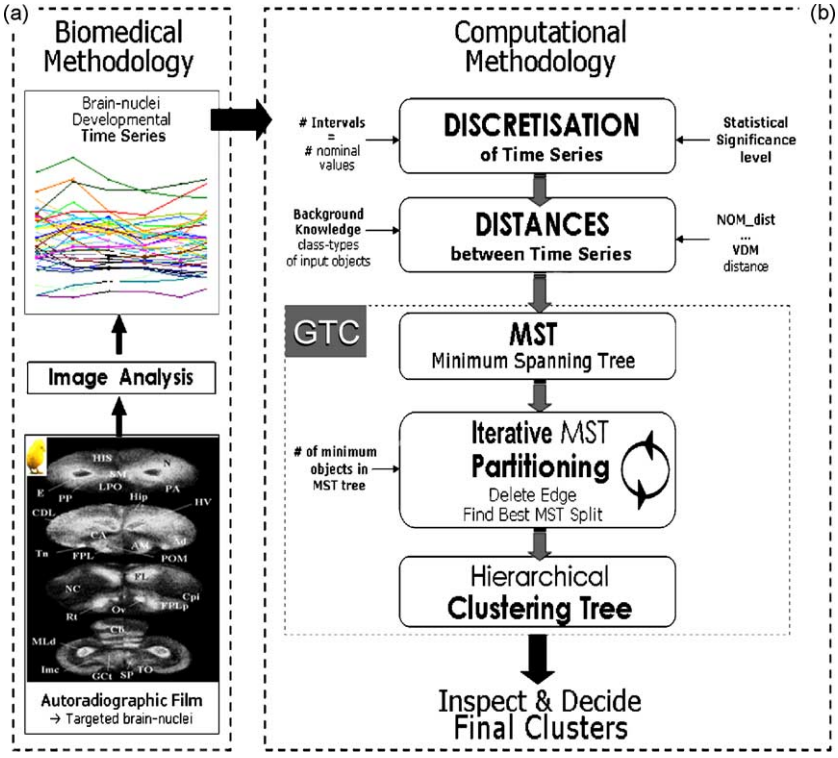

Fig. 4 (a) Biomedical methodology—brain developmental data acquisition and time-series data presentation, (b) computational methodology: discretization of time-series (the user enters the number of time-series intervals and respective values as well as the statistical-significance level for transforming a numeric time-series value to its corresponding nominal value); computation of distances between time-series (based on the availability of background knowledge the user specifies the respective input file); the core of the GTC algorithm where, the computed MST is computed and iteratively partitioned to construct the final clustering tree. The resulted clustering tree may be inspected by the user and the final clusters are selected.

jects in one group) the resulted sub-clusters have 1 and $n-1$ objects, respectively. The sub-clusters are formed after performing a total of $n-1$ CU computations (i.e., edge-cuts or, splits of the corresponding MST tree). At the second level the cluster with the $n-1$ objects is partitioned into two sub-clusters with 1 and $n-2$ objects, respectively, after performing a total of $n-2$ CU computations. At the last level, $n-1$, there are $n-(n-2)$ objects, and a total of $n-(n-2)-1=1$ CU computations are to be performed. So, the total number of CU computations is equal to $1+2+\cdots+(n-1) = n(n-1)/2$. As a result, and for the worst case, the GTC algorithm exhibits a quadratic to the total number of input objects, and linear to the number of features and the mean number of feature values, time-complexity, i.e., $\sim O(n^2 \times F \times V)$.

The quadratic complexity figure is in accordance to hierarchical clustering approaches that use dynamic closest pairing techniques [27], and with *k*-means approaches when the preset number of clusters is equal to the total number of input objects. In all the conducted experiments, and for datasets with ~1000—5000 objects and ~10—20 features, the real execution time of the Prolog-based GTC implementation ranges from ~2 to ~30 min (on a 1.7 MHz, 0.5 G RAM PC).

6. Status report and lessons learned

The application of GTC on the CTRL dataset, using the NOM_dist, concludes to a set of five clusters. The same number of clusters was retained for the CLON dataset. In Fig. 3, we illustrate the MST organization of these clusters, both in high-levels—the three super-clusters 'E': for the Early development pattern; 'ML': for the Mid-Late development pattern; and 'C': for the Constant pattern, and in low-levels of the clustering-tree—the five sub-clusters; 'EARLY' and 'EARLY_MID' developmental sub-patterns of the 'E' indicative cluster pattern; 'MID_LATE' and 'LATE' developmental sub-patterns of the 'ML' indicative cluster pattern; and 'CONSTANT' developmental pattern being identical to the 'C' indicative cluster pattern.

The mean protein-syntheses over all the brain nuclei assigned into a cluster, i.e., the cluster's

centroid is used as the indicative (or representative) pattern of each cluster. Note that the means are computed over the [0—1]-transforms of the corresponding time-series. In Fig. 5, the plots of the clusters' representative patterns (for both CTRL and CLON experiments) are shown, accompanied with their respective colored-patterns. The colored patterns were derived with the aid of the GEPAS Web-services (http://gepas.bioinfo.cnio.es/cgi-bin/cluster), and after the corresponding 'newick trees' were generated with the aid of the treeview program (http://taxonomy.zoology.gla.ac.uk/rod/treeview.html).

6.1. Biological interpretation

The presented clustering analysis was used to identify the biochemical changes underlying the morphogenetic events in late embryonic development. Based on the regional differences in protein synthesis activity, the identified five histogenetic cluster-representative patterns led to the mature anatomically defined nuclei.

The 'CONSTANT' cluster included brain areas already formed (i.e., matured) brain regions (PS activity similar to post-hatching stages). Basically, these areas are not affected by the later ontogenetic events taking place at the studied stages. It is interesting that these regions are distributed though out longitudinal as well as the transverse brain axis.

The other clusters were characterized by altered protein synthesis rates (altered gene expression) that may represent altered function of the gray matter structure. Previous in vivo and in vitro studies showed that PS and nucleic acid synthesis decrease in mammalian brain during development [28]. This pattern is characteristic of 'EARLY' and 'EARLY_MID' clusters and represents about half of the brain regions studied, suggesting naturally occurring cell death, cell migration, and elimination of synapses.

In addition, 'MID_LATE' and 'LATE' patterns were identified, both showing an increase in PS during development, possibly reflecting addition of cells and specific qualitative changes in PS. For example, the production of proteins that appears to relate with late events of brain development, such as, synthesis of the myelin associated protein. The pattern of 'LATE' maturation characterized white matter structures, through out the antero-posterior axis that showed a delayed increase of protein synthesis in parallel with the late myelination and appearance of myelin proteins during development of the avian brain [29]. Moreover, patterning PS activity in immature cerebellum layers parallels known events in avian cerebellar development [8].

6.1.1. CTRL versus CLON

In order to assess the ways that clonidine affects the brain development process we identified the brain nuclei that change their cluster assignment between the corresponding GTC runs (i.e., CLON versus CTRL), and significantly change their developmental profile. Clonidine induced complex changes in nuclei located through out the brain axes: *CDL*, *GCT*, *GLv*, *LoC*, and *nDBC* change their profile from 'E' to 'C'; *CPi*, *Hip*, *POM*, *SGFSd*, and *TOv* nuclei form 'C' to 'E', and *FPLp* and *PL* nuclei from 'C' to 'ML'. Brain regions affected, contain significant levels of the α_2 adrenoceptor, in which clonidine acts as an agonist [8,30]. In most cases, stimulation of α_2 receptor subtype by clonidine application at different embryonic stages, delayed maturation of specific brain areas in several anatomical domains. This finding is in agreement with the pluri-segmental origin of the catecholaminergic system in the forebrain and midbrain [31], and the role of α_2 adrenoceptors in synapse formation and plasticity mechanisms [7].

A fundamental question, of great interest in brain development studies, is the potential of a 'hidden' program that guides the developmental process from which organized functional circuits emerge. A promising approach would be to examine the relation between the brain nuclei developmental profiles and various class-type values assigned to the nuclei. The targeted class-types and their values are (see Table 1): 'major anatomic divisions' (MAD): refers to major compartments of the brain—eight such divisions (values) are targeted that can be considered modular because they represent largely independent histogenetic units of neural tissue (pallium, subpallium, hypothalamus, thalamus, cerebellum, tectum, tegmentum, and pons) [32]; 'A_2 synaptic transmission' (A2ST): refers to the concentration-level of specific alpha 2 subtype of adrenoreceptors in the respective brain-nuclei [8]—three levels were identified (low, medium, and high); 'brain system' (BS): refers to the brain functional system organization including six such divisions (limbic, somatosensory, visual, motor, multimodal, and white-matter); and 'cerebral axes topography' (CAT): refers to the location of the area in respect to the dorso-ventral or medio-lateral axis of the brain—four such locations are identified (dorsal, ventral, lateral, and medial). The results are summarized in Table 2.

The analysis on the MAD class-type did not support the idea that forebrain divisions—telence-

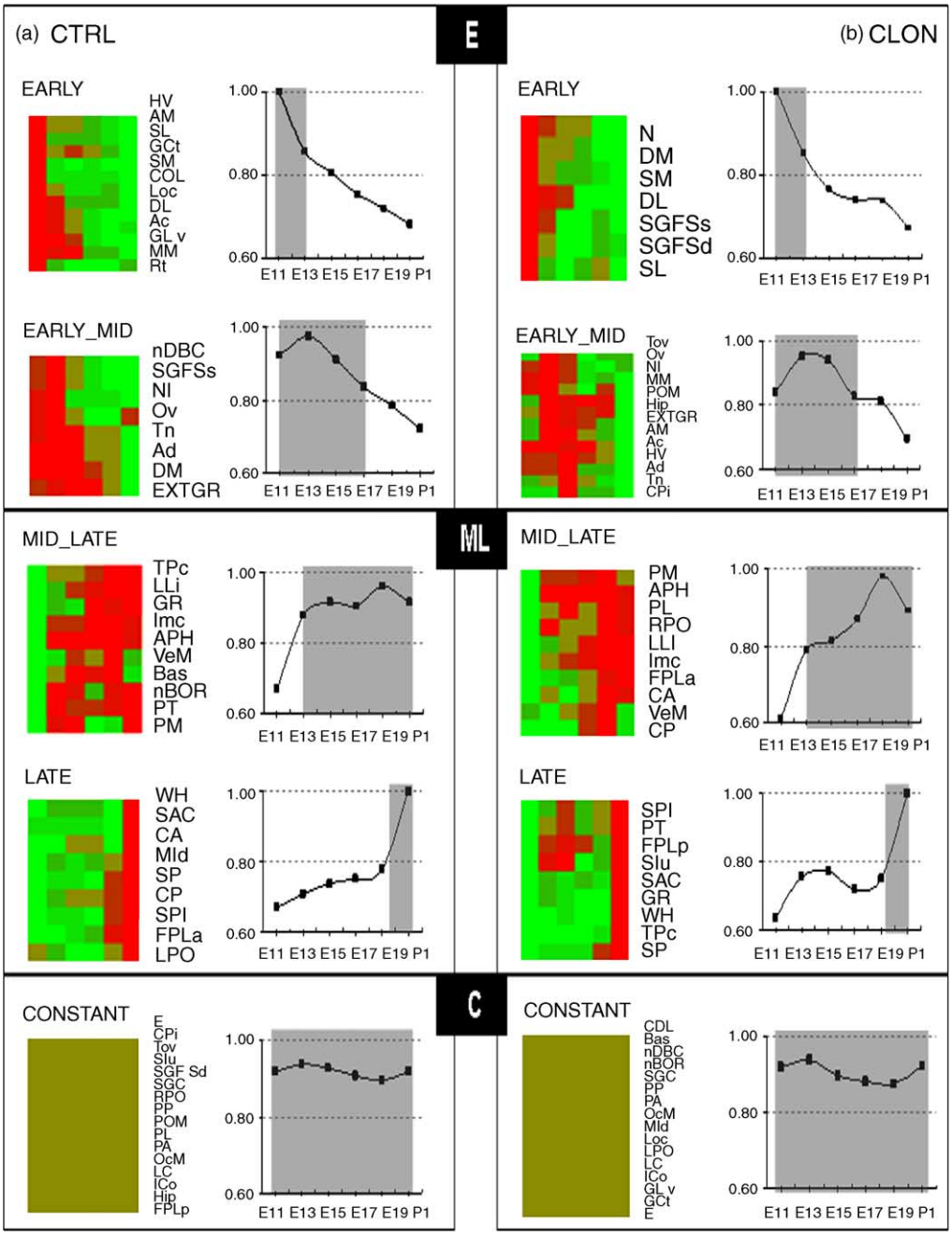

Fig. 5 The colored protein-synthesis profiles of the targeted brain nuclei, accompanied with the plots of the respective cluster's centroid; (a) for CTRL, and (b) for CLON experiments. Symbolic names were assigned based on the trend of the respective pattern—cluster E: with two sub-clusters identified, *EARLY*—high activity at *early* (E11, E13) stages (shaded area), followed by sharp decline during development and *EARLY_MID*—high activity at *early-to-mid* (E11—E17) stages gradually declining at later stages; cluster ML: with two sub-clusters identified, *MID_LATE*—low activity at the early, increasing at *mid-to-late* stages (from E13 to E19), and *LATE*—low activity in early-to-mid stages with highest activity at hatching (P1); and cluster C: a continuous *constant* developmental (high) activity over all the stamped time-points (from E11 to P1). The five clusters ('EARLY', 'EARLY_MID', 'MID_LATE', 'LATE', and 'CONSTANT') are retained between the two experiments. 'E' stands for the merged 'EARLY' and 'EARLY_MID' sub-clusters, and 'ML' for the merged 'MID_LATE' and 'LATE' sub-clusters.

Table 2 Distribution of brain regional class-type values in the induced clusters

	CTRL			CLON			Total
	C	E	ML	C	E	ML	
Major Anatomic Divisions (MAD)							
pallium	3	**6**	2	3	**7**	1	11
subpallium	2	3	1	3	3	0	6
thalamus	0	**5**	1	1	**4**	1	6
hypothalamus	1	2	0	0	**3**	0	3
tectum	3	2	2	**4**	2	1	7
tegmentum	1	0	**5**	1	0	5	6
pons	4	2	4	4	0	6	10
cerebellum	0	1	**2**	0	1	2	3
A2 Synaptic Transmission (A2ST)							
low	8	7	11	4	9	13	26
medium	5	4	7	7	3	6	16
high	3	**10**	2	5	**9**	1	15
Brain System (BS)							
auditory/vocal	1	2	2	**3**	1	1	5
visual	4	3	4	4	4	3	11
somatosensory	3	2	3	2	2	4	8
multimodal	0	**4**	0	1	**3**	0	4
limbic	2	**8**	1	1	**9**	1	11
motor	4	2	5	5	1	5	11
white-matter	2	0	**5**	0	1	6	7
Cerebral Axes Topography (CAT)							
dorsal	2	4	**8**	5	2	7	14
ventral	6	2	7	4	3	8	15
medial	2	**10**	1	4	**9**	0	13
lateral	6	5	4	3	7	5	15

Figures in bold indicate significant presence of areas (>50% of the total areas within the class-type) in 'C' (Constant), 'E' (Early) and 'ML' (Mid-Late) clusters.

phalon (pallium, subpallium) and diencephalon (thalamus, hypothalamus), develop later than the hindbrain divisions (pons, cerebellum). In fact an earlier histogenetic trend was determined in specific forebrain divisions (thalamus, pallium) compared to posterior located areas of hindbrain. However, our analysis is in accordance with the notion that, hindbrain nuclei in pons are pluri-segmental (originate from multiple embryonic divisions; for refs see [32], since they did not follow a specific ontogenetic pattern, and were distributed in all clusters.

Moreover, thalamic and pallial areas are grouped in the same discovered indicative developmental pattern of 'EARLY' histogenesis, supporting the significance of their reciprocal connections and the functional role of positional information in brain pattern formation. Subpallium derived structures follow diverse ontogenetic trends, providing an additional indication of their different modality, as suggested by their molecular/structural subdivision [33].

Furthermore, we questioned the possible correlation of the time course embryonic patterning with the hierarchical levels of brain nuclei (i.e., the BS class-type). Specific types of sensory information are analyzed in parallel by different functional systems (optic, acoustic, somatosensory); integration of many sensory modalities is performed by multi-modal association areas, while motor programs are generated by motor-related areas (e.g., subpallium; basal ganglia). Such brain systems represent functional neural units, processing specific information, composed by distantly located regions, derived from several embryonic divisions (plurisegmental) with molecular specificity. It has been suggested that embryonic modularity is transformed into functional modularity, in part by translating positional information [32]. We questioned the coincidence in the developmental PS pattern of differentially originated areas (different anatomic modality) that are functionally related (same functional modality), by grouping areas according to their involvement in specific neural circuits (BS, Table 2), but no specific pattern was followed among areas of the same functional modality. However, our analysis suggests a characteristic earlier maturation of multi-modal and limbic areas compared to those related to specific sensory or motor systems, as proposed by quantitative neuroanatomic studies in monkey cortex [34].

In addition to the division of brain along the longitudinal axis (i.e., MAD; pallium; anterior → cerebellum posterior), we grouped areas along the transverse (dorsal—ventral) and sagital (medial—lateral) axes (i.e., CAT class-types). This division clearly showed an advanced formation of medially located structures as expected due to their closer position to medial ventricular proliferation zones were cells are produced before migrating to their final position to form brain nuclei [11,35].

6.2. Exploiting neurobiological knowledge

Qualitative assessment of the clustering results, based on the impurity index formula (presented in Section 4.2.4), was performed in addition to the presented quantitative analysis. Specifically, we questioned whether: ''*A2ST level of brain-nuclei could predict its type of developmental profile*?''.

In Table 3, we present the impurity indices of the various class-types for different GTC runs: *NOM* run—the NOM_dist is employed, with no reference to brain nuclei classification; *VDM/MAD* run—the VDM distance (formula 3, Section 4.2.2) is utilized with brain nuclei assigned to brain 'ma-

Table 3 Impurity Indices. Bold figures indicate superiority with respect to the reference GTC run that utilizes the 'NOM' distance (i.e., no classification information is utilized; first column)

	NOM	VDM/ MAD	VDM/ A2ST	VDM/ BS	VDM/ CAT
MAD	0.13	*0.18*	**0.22**	**0.22**	*0.19*
A2ST	0.25	0.21	**0.37**	*0.29*	0.24
BS	0.19	0.18	**0.32**	*0.26*	*0.21*
CAT	0.22	0.20	**0.30**	*0.26*	*0.23*

Figures in bold indicate superiority over all runs.

jor anatomic divisions' class-types; *VDM/A2ST* run—the same with brain nuclei assigned to 'A2 synaptic transmission' class-types; *VDM/BS*—the same, with brain nuclei assigned to 'brain system' class-types; and *VDM/CAT*—the same, with brain-nuclei assigned to 'Cerebral Axes Topography' class-types.

The VDM/A2ST run produces superior impurity index figures with respect to all other runs (figures in bold in Table 3). The result of this analysis suggests a relation between protein-synthesis activity and the concentration-level of α_2 adrenoreceptors in the targeted brain-nuclei.

7. Conclusions and future plans

The primary objective of the presented study was the discovery of indicative and characteristic patterns in the developing brain. To this end, we introduced a novel graph-theoretic clustering (GTC) methodology adjusted for sequential events such as those occurring in the course of brain developmental.

The methodology relies on a careful discretization of time-series values. A weighted-graph structure is utilized in order to geometrically arrange the discretised time-series, and the respective minimum spanning-tree of the graph is formed. In the core of GTC is the iterative partitioning of the formed MST using a well-founded information-theoretic metric (category utility), which decides on the MST-edge to split, as well as, when to terminate partitioning. The final outcome is a hierarchical clustering organization of the input objects.

With GTC clustering we were able to uncover critical relations between targeted brain areas, and identify critical brain developmental events. Especially, the hierarchical clustering organization of the brain-nuclei revealed a more refined distinction between the discovered indicative brain developmental patterns.

In addition, hybridization between available neurobiological knowledge (i.e., the assignment of brain-nuclei to various brain divisions and relative class-types), and feature-based description of targeted brain-nuclei (i.e., the brain development PS profiles) unhide valuable information on histogenetic relationships between distantly located areas. Moreover, it reveals important evidence for 'area specific' modeling of the developing brain and confirms the fundamental role of α_2 adrenoceptors, providing the basis for further developmental studies.

This GTC methodological approach is currently being tested on other domains (e.g., economic time-series data), in order to explore the suitability of the proposed methodology in other domains. In addition, further experimentation with domains of huge volumes of data would provide important assessment of its scalability—in this context we have performed studies with real-world gene expression profiling datasets with encouraging preliminary results [36], Furthermore, we plan to improve and port GTC in Java, establish respective Web-based clustering services, and incorporate more sophisticated distance computation processes specifically suitable for time-series objects (e.g., dynamic time warping).

Acknowledgements

The authors thank A. Stamatakis for important contribution in the autoradiographic experiments. The work presented in this paper was partially supported by the EU project COLORED, QLRT-1999-31629.

References

[1] C. Kennedy, S. Suda, C.B. Smith, M. Miyaoka, M. Ito, L. Sokoloff, Changes in protein synthesis underlying functional plasticity in immature monkey visual system, Proc. Natl. Acad. Sci. U.S.A. 78 (1981) 3950—3953.
[2] M.C. Ingvar, P. Maeder, L. Sokoloff, C.B. Smith, Effects of ageing on local rates of cerebral protein synthesis in Sprague—Dawley rats, Brain 108 (1985) 155—170.
[3] S.I.S. Rattan, A. Derventzi, B.F.C. Clark, Protein synthesis, post-translational modifications and aging, Ann. N. Y. Acad. Sci. 663 (1992) 48—62.
[4] R. Schliebs, S.P. Rose, M.G. Stewart, Effects of passive avoidance training on in vitro protein synthesis in forebrain slices of day-old chicks, J. Neurochem. 44 (1985) 1014—1028.
[5] D.M. Hermann, E. Kilic, R. Hata, K.-A. Hossmann, G. Mies, Relationship between metabolic dysfunction, gene responses and delayed ischemic injury after mild focal cerebral ischemia in mice, Neuroscience 104 (2001) 947—950.

[6] E. Dupin, C. Ziller, N.M. Ledouarin, The avian embryo as a model in developmental studies: chimeras and in vitro clonal analysis, Curr. Top. Dev. Biol. 36 (1998) 1—35.
[7] A. Stamatakis, M.G. Stewart, C.R. Dermon, Passive avoidance learning involves alpha2 noradrenergic receptors in a day old chick, Neuroreport 9 (1998) 1679—1683.
[8] C.R. Dermon, E.D. Kouvelas, Binding properties, regional ontogeny and localization of adrenergic receptors in chick brain, Int. J. Dev. Neurosci. 6 (1988) 471—482.
[9] R. Revilla, R. Diez-Alarcia, R. Mostany, C.C. Perez, A. Fernandez-Lopez, Norepinephrine, epinephrine and MHPG levels in chick brain development, Neuropharmacology 41 (2001) 480—485.
[10] K. Morik, The representation race—preprocessing for handling time phenomena, LNAI 1810 (2000) 4—19.
[11] C.R. Dermon, B. Zikopoulos, L. Panagis, E. Harrison, C.L. Lancashire, R. Mileusnic, M.G. Stewart, Passive avoidance training enhances cell proliferation in 1-day-old chicks, Eur. J. Neurosci. 16 (2002) 1267—1274.
[12] J. Macqueen, Some methods for classification and analysis of multivariate observations, in: L.M. Le Cam, J. Neyman (Eds.), Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, University of California Press, Berkeley, CA, pp. 291—297.
[13] C.T. Zahn, Graph-theoretical methods for detecting and describing gestalt clusters, IEEE Trans. Comp. 20 (1971) 68—86.
[14] R.L. Page, Minimal spanning tree clustering method (ACM algorithm 479), CACM 17 (6) (1974) 321—323.
[15] D. Xu, V. Olman, L. Wang, Y. Xu, EXCAVATOR: a computer program for efficiently mining gene expression data, Nucleic Acids Res. 31 (19) (2003) 5582—5589.
[16] G. Mies, B. Djuricic, W. Paschen, K.-A. Hossmann, Quantitative measurement of cerebral protein synthesis in vivo: theory and methodological considerations, J. Neurosci. 76 (1997) 35—44.
[17] Y. Sun, G.E. Deibler, J. Jehle, J. Macedonia, I. Dumont, T. Dang, C.B. Smith, Rates of local cerebral protein synthesis in the rat during normal postnatal development, Am. J. Physiol. 268 (1995) 549—561.
[18] C.R. Dermon, A. Stamatakis, S. Giakoumaki, J. Balthazart, Differential effects of testosterone on protein synthesis activity in male and female quail brain, Neuroscience 123 (2003) 647—666.
[19] W.J. Kuenzel, M. Masson, A Stereotaxic Atlas of the Brain of the Chick (Gallus domesticus), The Johns Hopkins University Press, Baltimore, 1988.
[20] R. Agrawal, K. Lin, H.S. Sawhney, K. Shim, Fast similarity search in the presence of noise, in: Proceedings of the 21st VLDB Conference, Zurich, Switzerland, 1995, pp. 490—501.
[21] P. Laird, Identifying and using patterns in sequential data, in: K. Jantke, S. Kobayashi, E. Tomita, T. Yokomori (Eds.), Algorithmic Learning Theory, Springer-Verlag, Berlin, 1993, pp. 1—18.
[22] L.M. Lopez, I.F. Ruiz, R.M. Bueno, G.T. Ruiz, Dynamic discretisation of continuous values from time series, LNAI 1810 (2000) 290—291.
[23] D. Wilson, T. Martinez, Improved heterogeneous distance functions, J. Artif. Intell. Res. 6 (1997) 1—34.
[24] R. Prim, Shortest connection networks and some generalizations, Bell Syst. Tech. J. 36 (1957) 1389—1401.
[25] D. Fisher, Knowledge acquisition via incremental conceptual clustering, Machine Learning 2 (1987) 139—172.
[26] E.C. Pielou, The measurement of diversity in different types of biological collections, J. Theor. Biol. 13 (1966) 131—144.
[27] D. Epstein, Fast hierarchical clustering and other applications of dynamic closets pairs, J. ACM Exp. Algorithms 5 (2000) 1—23.
[28] D.E. Johnson, O.Z. Sellinger, Protein synthesis in neurons and glial cells of the developing rat brain: an in vivo study, J. Neurochem. 18 (1971) 1445—1460.
[29] W.B. Macklin, C.L. Weill, Appearance of myelin proteins during developmenta in the chick central nervous system, Dev. Neurosci. 7 (1985) 170—178.
[30] C.R. Dermon, E.D. Kouvelas, Quantitative analysis of localization of adrenergic receptors in chick brain, J. Neurosci. Res. 23 (1989) 297—303.
[31] A. Verney, N. Zecevic, L. Puelles, Structure of longitudinal brain zones that provide the origin for the sunstantia nigra ventral tegmental area in human embryos as revealed by cytoarchitecture tyrosine hydroxylase calretinin calbindin GABA immunoreactions, J. Comp. Neurol. 429 (2001) 22—44.
[32] C. Redies, L. Puelles, Modularity in vertebrate brain development and evolution, BioEssays 23 (2001) 1100—1111.
[33] M.C. Mione, C. Danevic, P. Boardman, B. Harris, J.G. Parnavelas, Lineage analysis reveals neurotransmitter (GABA or glutamate) but not calcium-binding protein homogeneity in clonally related cortical neurons, J. Neurosci. 14 (1994) 107—123.
[34] S.M. Dombrowski, C.C. Hilgetag, H. Barbas, Quantitative architecture distinguishes prefrontal cortical systems in the rhesus monkey, Cereb. Cortex 11 (2001) 975—988.
[35] A. Alvarez-Buylla, J.R. Kirn, Birth, migration, incorporation, and death of vocal control neurons in adult songbirds, J. Neurobiol. 33 (1997) 585—601.
[36] G. Potamias, The utility of sequence, function and transcriptional information in gene expression profiling (extended abstract), in: Proceedings EUNITE Workshop on Intelligent Technologies for Gene Expression Based Individualized Medicine, Jena, Germany, 2003, http://www.hki-jena.de/hki/hki_v031.htm.