

## Computational Neuroscience

## On the discovery of group-consistent graph substructure patterns from brain networks

Nantia D. Iakovidou<sup>a,\*</sup>, Stavros I. Dimitriadis<sup>b,c</sup>, Nikolaos A. Laskaris<sup>c</sup>, Kostas Tsihlias<sup>a</sup>, Yannis Manolopoulos<sup>a</sup>

<sup>a</sup> Data Engineering Laboratory, Department of Informatics, Aristotle University Thessaloniki, 54124, Greece

<sup>b</sup> Electronics Laboratory, Department of Physics, University of Patras, Greece

<sup>c</sup> Artificial Intelligence Information Analysis Lab, Department of Informatics, Aristotle University of Thessaloniki, 54124, Greece

## HIGHLIGHTS

- ▶ We introduce a potential alternative to current techniques for generating motif repertoire in brain connectivity research.
- ▶ We indicate, using actual functional connectivity graphs, various ways to exploit the new technique for gaining insights to assembled connectivity graph datasets.
- ▶ Using the proposed contrastive learning scheme for motif extraction we finally score and rank the detected motifs according to their importance.

## ARTICLE INFO

## Article history:

Received 10 July 2012

Received in revised form

19 December 2012

Accepted 20 December 2012

## Keywords:

Functional networks

Connectivity key-subgraphs

Graph mining

## ABSTRACT

Complex networks constitute a recurring issue in the analysis of neuroimaging data. Recently, network motifs have been identified as patterns of interconnections since they appear in a significantly higher number than in randomized networks, in a given ensemble of anatomical or functional connectivity graphs. The current approach for detecting and enumerating motifs in brain networks requires a predetermined motif repertoire and can operate only with motifs of small size (consisting of few nodes).

There is a growing interest in methodologies for frequent graph-based pattern mining in large graph datasets that can facilitate adaptive design of motifs. The results presented in this paper are based on the graph-based Substructure pattern mining (gSpan) algorithm and introduce a manifold of ways to exploit it for data-driven motif extraction in connectomics research.

Functional connectivity graphs from electroencephalographic (EEG) recordings during resting state and mental calculations are used to demonstrate our approach. Relying on either time-invariant or time-evolving graphs, characteristic motifs associated with various frequency bands were derived and compared. With a suitable manipulation, the gSpan discovers motifs which are specific to performing mental arithmetics. Finally, the subject-dependent temporal signatures of motifs' appearance revealed the transient nature of the evolving functional connectivity (math-related motifs "come and go").

© 2012 Elsevier B.V. All rights reserved.

## 1. Introduction

Graphs are generic data structures which are employed in various scientific domains such as Computer Science (Faloutsos et al., 1999), Biology (De Silva and Stumpf, 2005) and Neuroscience (Rubinov and Sporns, 2010). In general, a graph models a network of relationships between objects. There is a growing interest in the theoretical aspects of network analysis in an attempt to model (Simpson et al., 2012; Vértés et al., 2012), describe (Rubinov

and Sporns, 2010) and propose new measures (Joyce et al., 2010) for a better understanding of complex systems. Complex network analysis was emerged with the scope of characterizing and understanding the underlying mechanisms that govern complex systems. The key idea is to recognize and assess important properties of complex systems by quantifying the topologies of their network representations (Bullmore and Sporns, 2009; Rubinov and Sporns, 2010). Having originated in the mathematical study of networks, this approach shares many methodologies and algorithms with the well-established branch of graph theory.

A hypothesis regarding network complexity is the following: as the complexity of a network increases, a restricted repertoire of initially existing simpler networks is preserved, extended, and combined, while it is less likely that more complex structures

\* Corresponding author at: Data Engineering Laboratory, Department of Informatics, Aristotle University of Thessaloniki, 54124, Greece. Tel.: +30 2310991924.

E-mail address: [niakovid@csd.auth.gr](mailto:niakovid@csd.auth.gr) (N.D. Iakovidou).

are generated entirely by new sets of subnetworks (Sporns et al., 2000; Sporns and Kötter, 2004). The above hypothesis is supported by the observation that complex networks arise from the addition of connections in positions that maximize the efficiency in overall information processing (Sporns and Kötter, 2004). An important approach to exploring the rules that govern the structure of complex networks is to study the distribution of characteristic building blocks which are called “motifs”. Network motifs have been examined in the setting of gene regulatory, metabolic, and other biological and artificial networks (Milo et al., 2002, 2004). Motif fingerprints have been discovered in various neurobiological data (Macaque and cat cortex, *Caenorhabditis elegans*) where the relationship between structural and functional motifs has been examined (Sporns and Kötter, 2004).

A considerable drawback of the encountered motifs in previous studies is the a priori selection of the size of motifs to be detected (usually motifs of 2, 3 or 4 nodes are considered). There is a practical reason for fixing the size of motifs to a small number  $N$ , since all the possible motifs are first constructed and “tabulated” and then matched within the given graph(s). The number of all  $N$ -node motifs increases exponentially (3-node  $\rightarrow$  13 possible classes, 4-node  $\rightarrow$  199 possible classes etc.) (Milo et al., 2002; Sporns and Kötter, 2004; Rubinov and Sporns, 2010) and soon this number becomes prohibitive. The original network motif detection approach attempts to find significant frequent subnetworks in one network with unlabeled vertices that are indistinguishable from each other and also with a fixed number of motif size (Milo et al., 2002). A relevant data mining problem with many scientific and commercial applications is the detection of frequent graph-based substructure patterns that correspond to the same labeled nodes without a priori selection of the motif size, among a given set of graphs (Yan and Han, 2002). By resorting to these techniques we can develop a bottom-up approach (i.e. data-driven) to construct a motif-repertoire. An important approach of this kind, based on Depth-First Search (DFS) lexicographic order, is the gSpan (graph-based Substructure pattern mining) algorithm (Yan and Han, 2002). Detecting common subgraphs among a number of graphs is an important issue that could improve both descriptive analysis and modeling of network data (Kolaczyk, 2009). Subsequently, these overrepresented sub-networks may be linked with a potential functional contribution to the global functionality of the entire network.

These new complex network methodologies show great promise in studying brain connectivity datasets. Brain connectomics is an emerging field that encounters networks of brain regions connected by anatomical tracts or by functional associations (Rubinov and Sporns, 2010). The brain is a complex system, and its characterization using the complex network approach is not only meaningful but also necessary. Such methods for characterizing the structural or functional networks are becoming extremely popular (Achard et al., 2006; Bassett and Bullmore, 2006, 2009; Stam and Reijneveld, 2007; Bullmore and Sporns, 2009). Connectivity can be recovered from brain activity recorded with various modalities such as EEG, magnetoencephalography (MEG) and functional magnetic resonance imaging (fMRI). The functional coupling among distinct brain regions can be quantified based on suitable estimators depending on the nature of recorded data. In particular, for fast-recording modalities (mainly EEG/MEG) the issue of non-stationarity is of great importance and functional connectivity studies fall in two distinct categories. Those adopting a static network approach (Micheloyannis et al., 2009; Rubinov et al., 2009; Dimitriadis et al., 2009) and those considering time-varying connectivity graphs (De Vico Fallani et al., 2007a,b; Valencia et al., 2008; Dimitriadis et al., 2010a,b). Network motifs have been employed as descriptors for both static (Sporns and Kötter, 2004) and time-varying functional connectivity graphs (De Vico Fallani et al., 2007). They constitute an important measure of local

network topology that describe local patterns of interconnections and indicate important circuits for information processing (Milo et al., 2002, 2004).

The scope of this work is twofold: (1) to introduce in brain connectivity research a potential alternative to current techniques for generating a repertoire of motifs and (2) to indicate, using actual functional connectivity graphs, various ways to exploit the new technique for gaining insights to assembled connectivity graph datasets. For demonstration purposes we considered brain networks constructed from scalp EEG recordings. These recordings correspond to normal subjects who were either performing mental arithmetic calculations or were at rest. At this point, we need to clarify that in our framework the detected motifs refer to subgraph patterns built over a given set of labeled nodes (the recording sites).

The rest of the paper is organized as follows. In Section 2, after a short description of the available graph databases, we introduce the method and three alternative implementation modes. Section 3 includes indicative results while placing them in neuroscientific context and finally a discussion follows in Section 4.

## 2. Materials and methods

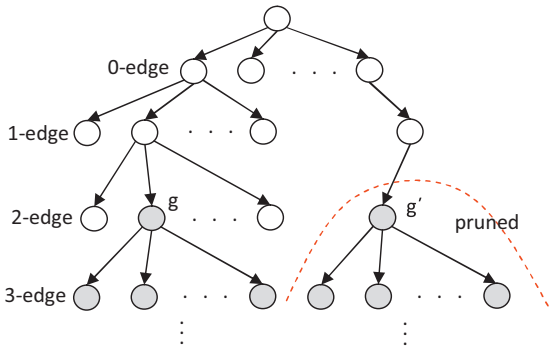
### 2.1. Functional connectivity data

The functional connectivity graph dataset was created as an intermediate result in two previous studies (Dimitriadis et al., 2010a,b, 2012) and hence a detailed description can be found therein. It consists of two types of graphs (time invariant and time-varying graphs) that correspond to two different brain states (resting state and mental calculations) and reflect neural synchrony in each of the five standard frequency bands ( $\theta$ ,  $\alpha_1$ ,  $\alpha_2$ ,  $\beta$  and  $\gamma$ ). The original EEG signals had been recorded from 18 healthy adults while they were performing multiplications (active condition) or doing nothing (baseline condition). Starting from the signals recorded (at sampling frequency of 500 Hz with 30 sensors covering the head according to International 10–20 system) and filtered within a particular frequency band, we derived functional connectivity graphs (FCGs) by means of a phase synchrony estimator (in particular phase locking value, PLV, Lachaux et al., 1999) that quantifies functional dependence in pairwise fashion. Each FCG has a tabular format, i.e. a  $[30 \times 30]$  weighted-adjacency matrix  $W$  description, with elements  $w_{ij} \in [0,1]$  denoting the strength of functional association between the  $i$ th and the  $j$ th sensor (and implicitly between the cortical areas underneath). In the case of static FCGs, each graph represents connectivity estimates corresponding to subject’s brain activity for long recording periods (8 s). In the case of time-varying FCGs (time-indexed networks), a time-series of connectivity estimates is represented, with each connectivity snapshot corresponding to subject’s brain activity lasting for few milliseconds.

### 2.2. The gSpan algorithm

We consider the problem of identifying frequently occurring patterns (motifs) in a set of unweighted graphs. In particular, provided a graph set  $G = \{G_1, G_2, \dots, G_N\}$  of  $N$  undirected graphs and a real number  $\rho \in (0,1]$  we wish to find all subgraphs with minimum support  $\rho$ . The support  $\rho$  means that the subgraphs occur in at least  $\lfloor \rho \cdot N \rfloor$  (with “ $\lfloor \cdot \rfloor$ ” denoting the floor function) graphs in the set  $G$ . An algorithm that can provide a solution to this problem is **gSpan** (Yan and Han, 2002). gSpan has been experimentally verified to be one of the most efficient algorithms for graph mining. In the following, we provide a bird’s-eye view of the algorithm.

The algorithm discovers frequent subgraphs by extending them by one node at a time. This means that large motifs are generated after all smaller motifs that constitute them have also been



**Fig. 1.** The hierarchical decomposition of the search space. The discovery starts from trivial motifs (0-edge motifs that consist only of frequent single nodes) and then proceeds to larger motifs. The tree is traversed in a DFS manner. Assuming that  $g = g'$ , there is no need to traverse the subtree of  $g'$  since the generated motifs have been already reported by the traversal made in the subtree of  $g$  and thus the whole subtree of  $g'$  is pruned during the search process.

discovered. This is based on the *frequency antimonotone property* (Vanetik et al., 2006). This property states that if a graph  $g$  is frequent then any subgraph of  $g$  is also frequent and if  $g$  is not frequent then any graph that contains  $g$  is also not frequent.

To facilitate the discovery of motifs, the search space is hierarchically decomposed as shown in Fig. 1 giving rise to the *Depth-First Search (DFS) Code Tree*. The children of the root of this tree correspond to graphs which consist of a single node. Notice that in our case the set of nodes for all  $G_i$  is the same and as a result all single nodes are frequent. This is used only for initialization purposes, since single nodes are considered to be trivial motifs in our framework. These trivial motifs are then extended by other nodes into graphs that consist of two nodes and an edge. These new graphs are children of  $g$  in the DFS Code Tree and this procedure continues recursively until there is no node corresponding to a frequent graph that can be extended, in which case the algorithm backtracks.

To generate these graphs efficiently, a lexicographic label is assigned to each one of them. The smaller the label for a node the higher up in the DFS code tree resides. The enumeration of the subgraphs corresponding to nodes of the DFS code tree is done in lexicographic order, which is consistent with the DFS traversal of the search space. This is because the lexicographic label of  $g$  is a sequence of symbols, called the *DFS code*, generated by the DFS traversal of  $g$ . Since there are many DFS traversals which give rise to many DFS codes, we always choose the minimal one. For this coding scheme it holds that if a graph  $g$  is isomorphic to some other graph  $g'$  then their corresponding minimal codes are also identical. It is this minimal DFS code that allows for the efficient generation of candidate subgraphs that makes gSpan an efficient algorithm.

Another element of the gSpan algorithm is that in the exploration of the search space, represented conceptually by the DFS code tree, it may be the case that multiple nodes correspond to the same subgraph. In this case, when we visit a node  $v$  of the DFS code tree, which corresponds to a subgraph isomorphic to another node  $u$  that has already been visited, node  $v$  is discarded. In addition, because of the DFS traversal of the search space represented by the DFS code tree, all descendants of  $v$  are also discarded. Thus, the procedure does not need to explore the subtree of  $v$  since all frequent subgraphs in it have been reported in the subtree of  $u$ . Details of the algorithm can be found in (Yan and Han, 2002). Recently, extensions of gSpan (Jiang et al., 2010) have also been introduced. These versions extract weighted motifs in which the support takes into account the weights of the edges.

### 2.3. Exploiting gSpan for FCG mining

#### 2.3.1. Static graphs – single condition

For one particular brain condition we collect the set of  $N$  graphs that corresponds to the number of examined subjects. These graphs, originally, may be fully connected and weighted. A sparsification (binarization) step is recommended for reducing subsequent computational burden (to speed up the execution of gSpan to a small fraction of time) and as a “denoising” step as well, so as to retain the essence of underlying functional coupling manifold. Thresholding at a user-defined value of functional connectivity strength or based on statistical arguments (e.g. comparison with randomized networks, van Wijk et al., 2010) are the most common strategies for transforming connectivity  $W^{\text{subj.id}}$  matrices to adjacency matrices  $A^{\text{subj.id}}$  with ones only at entries corresponding to strongly coupled brain areas. We then adjust the gSpan algorithm to our framework, by setting appropriately parameter  $\rho$  and mine motifs that appear in the vast majority of these binary graphs (for example in a percentage  $\geq 90\%$  of graphs if  $\rho = 0.9$ ). In general, we prefer high  $\rho$  values (i.e. close to 1), since we want to reveal systematic trends in our data. Finally, we visualize the detected motifs in terms of the original brain topography.

#### 2.3.2. Static graphs – contrasting conditions

One of the most intriguing topics in the study of brain connectivity graphs is to compare two (or more) distinct set of FCGs, corresponding for instance to different clinical populations (Rubinov et al., 2009) or different recording conditions (Micheliyannis et al., 2005) and decide about graph-related features of high discriminability. We propose here a simple and effective strategy so as to exploit gSpan algorithm for discriminative learning. Given two sets of graphs  $G_1 = \{^1G_1, ^1G_2, \dots, ^1G_{N_1}\}$  and  $G_2 = \{^2G_1, ^2G_2, \dots, ^2G_{N_2}\}$ , our target is to find motifs that discriminate between these two sets, which means that we intend to detect motifs that exist in one set and do not appear in the other.

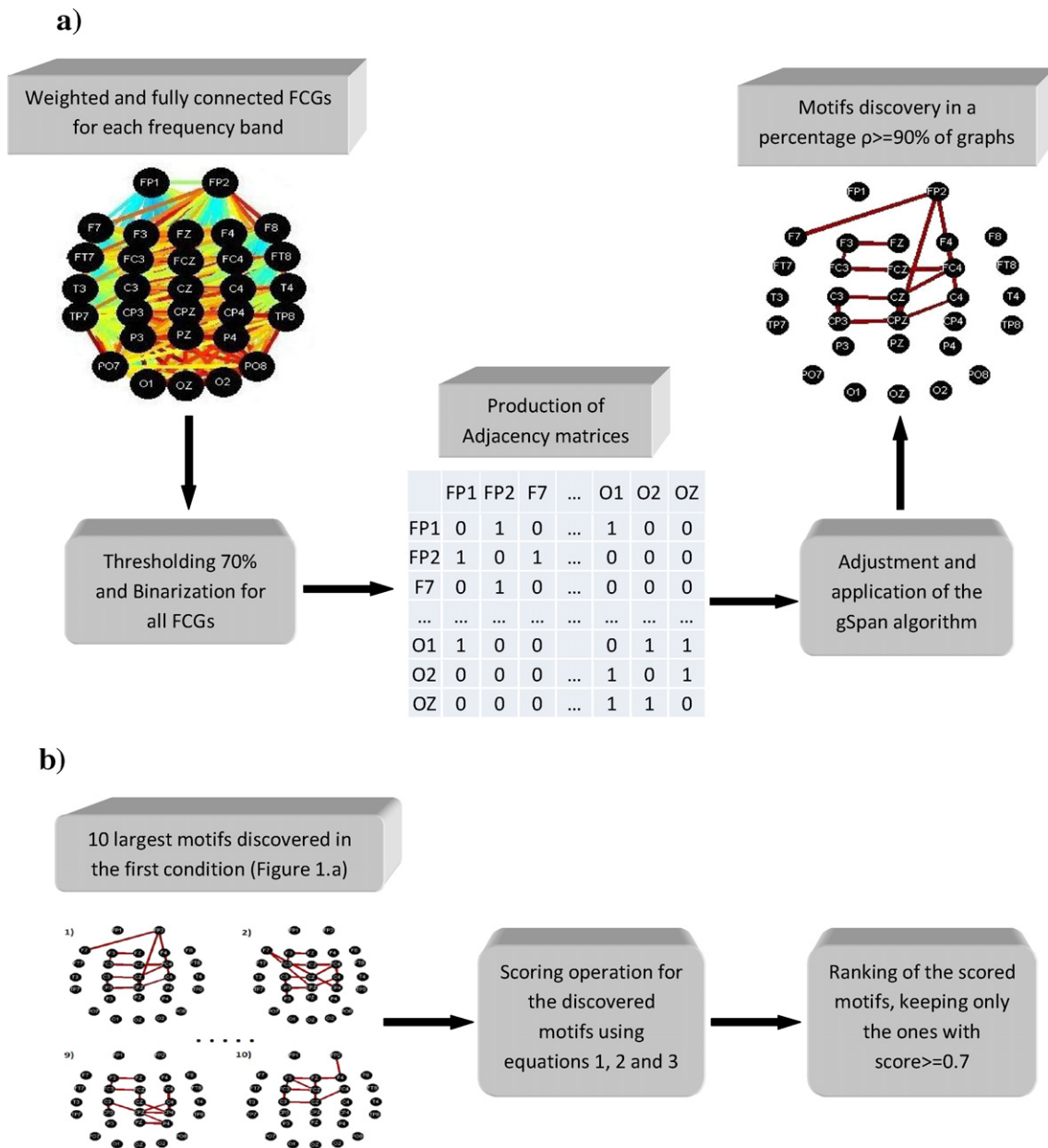
The algorithm proceeds as follows. After mining common subgraphs from the first condition (as described in Section 2.3.1), we identify among the largest detected motifs the ones that appear less frequently in the graphs of the second condition. To this end, we score each large motif  $m_j$  from first condition within three steps that include simple algebraic operations. Let  $A(m_j)$  an adjacency matrix that encodes the edges of a particular motif by means of its nonzero entries. We first apply Hadamard product (denoted below with  $\otimes$  and defined as the elementwise matrix product) between  $A(m_j)$  and each one of the adjacency matrices corresponding to set  $G_2$ . For each derived matrix, we sum all the elements so as to compute a ratio  $R_i$  in  $[0,1]$  that express the matching of motif  $m_j$  with the connectivity structure of graph  $^2G_i$  (as conveyed by the adjacency matrix after thresholding). Finally, we average the ratios  $R_i$  over the whole population of set  $G_2$ , to derive the final differentiability score (Dscore).

$$Q_i = \sum_{\text{row}} \sum_{\text{column}} A(m_j) \otimes A(^2G_i) \quad (1)$$

$$R_i = \frac{\sum_{\text{row}} \sum_{\text{column}} A(m_j) - Q_i}{\sum_{\text{row}} \sum_{\text{column}} A(m_j)} \quad (2)$$

$$\text{Dscore}(m_j) = \frac{\sum_i^{N_2} R_i}{N_2} \quad (3)$$

The Dscore of a detected motif is equal to one if it does not appear in the second condition at all. The largest in size motifs are ranked based on their Dscore, and hence according to the degree of distinctiveness with respect to the first condition. The most idiosyncratic among them are visualized in terms of brain topography.



**Fig. 2.** (a) Single condition. In case of static data, graphs represent subjects and in case of time-varying data, graphs represent time instants. (b) Contrasting conditions.

**2.3.3. Time-varying FCG**

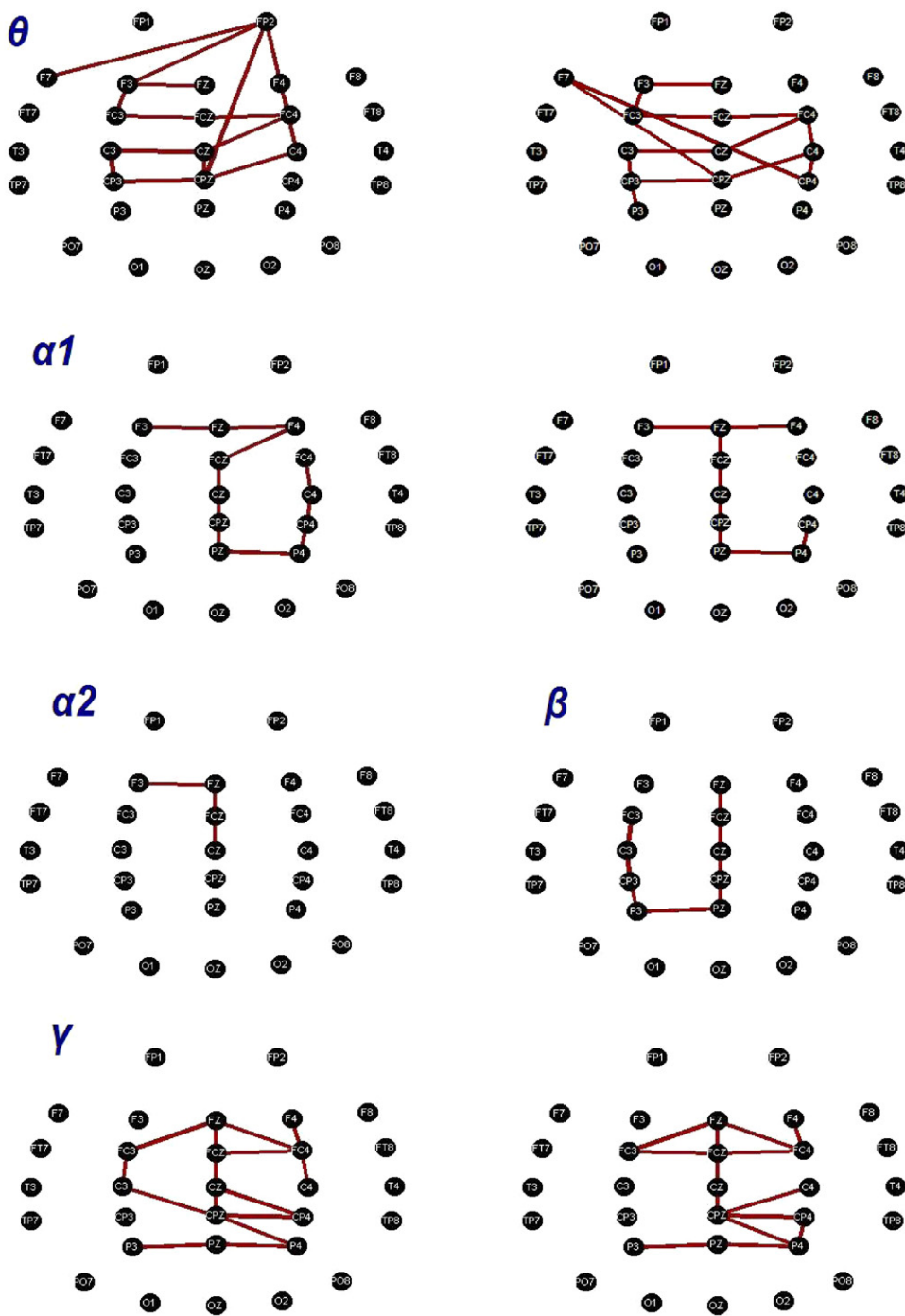
In the setting of time-varying graphs, if we denote by  $tv\_G_i$  the timeseries of functional connectivity graphs from the  $i$ th subject  $\{G_i(t)\}_{t=1:T}$ , the ensemble of  $N \times T$  graphs from  $N$  subjects forms a set  $tv\_G = \{G_1(t), G_2(t), \dots, G_N(t)\}$  that encapsulates the time-dependent variations in network organization. We can apply gSpan algorithm to such a set from a particular recording condition, so as to mine motifs that are not only common across subjects but also persistent in time. In a more elaborate exploration the time-varying graphs from two different recording conditions can be compared (in the notion of contrastive motif extraction of the previous subsection). After extracting the motifs from  $tv\_G_1$ , we rank them according to their frequency of occurrence in  $tv\_G_2$ .

**3. Results**

Fig. 2 provides a schematic outline of the procedure for the analysis of available FCGs. Each frequency band (defined,

conventionally, as  $\theta$  (4–8 Hz),  $\alpha_1$  (8–10 Hz),  $\alpha_2$  (10–13 Hz),  $\beta$  (13–30 Hz) and  $\gamma$  (30–45 Hz)) was treated separately. We first mined motifs from static connectivity graphs associated with baseline condition, then motifs appearing in mental calculations but not in baseline condition and finally motifs appearing in time-varying connectivity graphs while subjects were involved in mental arithmetic. A common thresholding scheme had been performed in all cases before motif extraction. The weights of each graph had been sorted and a threshold was defined so as to null out the 70% of the edges with the lower connectivity strength. The rest edges were considered as the connectivity skeleton and assigned the value of 1. We must mention at this point that we tested various thresholds in the interval [50%, 90%] (Micheloyannis et al., 2006; De Vico Fallani et al., 2007a,b). With threshold values  $> 70\%$  the resultant graphs were very sparse and the algorithm detected only a few motifs with two nodes or no motifs at all. On the other hand, with threshold values near 50% and 60% the graphs were too dense (close to being fully connected) and hence subgraphs could not be detected.





**Fig. 3.** Significant ( $P$ -value  $< 0.001$ ) motifs detected from control condition graphs in a percentage  $\rho \geq 90\%$  of subjects.

We finally report the results from thresholding at 70%, since the emerged motifs were characterized by the highest significance (see next section). We also verified the absence of disconnected nodes in the binarized graphs.

### 3.1. Resting state motifs from static FCGs

The study of resting state connectivity has recently received enormous attention (Damoiseaux 2006). Inferring resting-state connectivity patterns from neuroimaging data (fMRI, positron emission tomography (PET), etc.) is a challenging task and of great importance for the delineation of the so-called default-mode network. It was therefore plausible, as a first demonstration of gSpan

application, to seek motifs from the control condition FCGs. The most important results (in terms of motif-size and frequency of occurrence) of this exploration are shown in Fig. 3. Of course, given that the total number of motifs found in all bands is more than 50,000, it is impossible to show all these motifs in the current paper. Nevertheless, we depict in Fig. 4 some smaller motifs detected in  $\theta$  band, that either are subsets of the motifs shown in Fig. 3 and consequently included within them, or/and they have a higher  $P$ -value ( $\geq 0.01$ ) and so they do not share the same level of significance with the motifs shown in Fig. 3 (the computation of significance is described later in this subsection).

A first important common observation is that the detected motifs constitute connected subgraphs, which is an expected

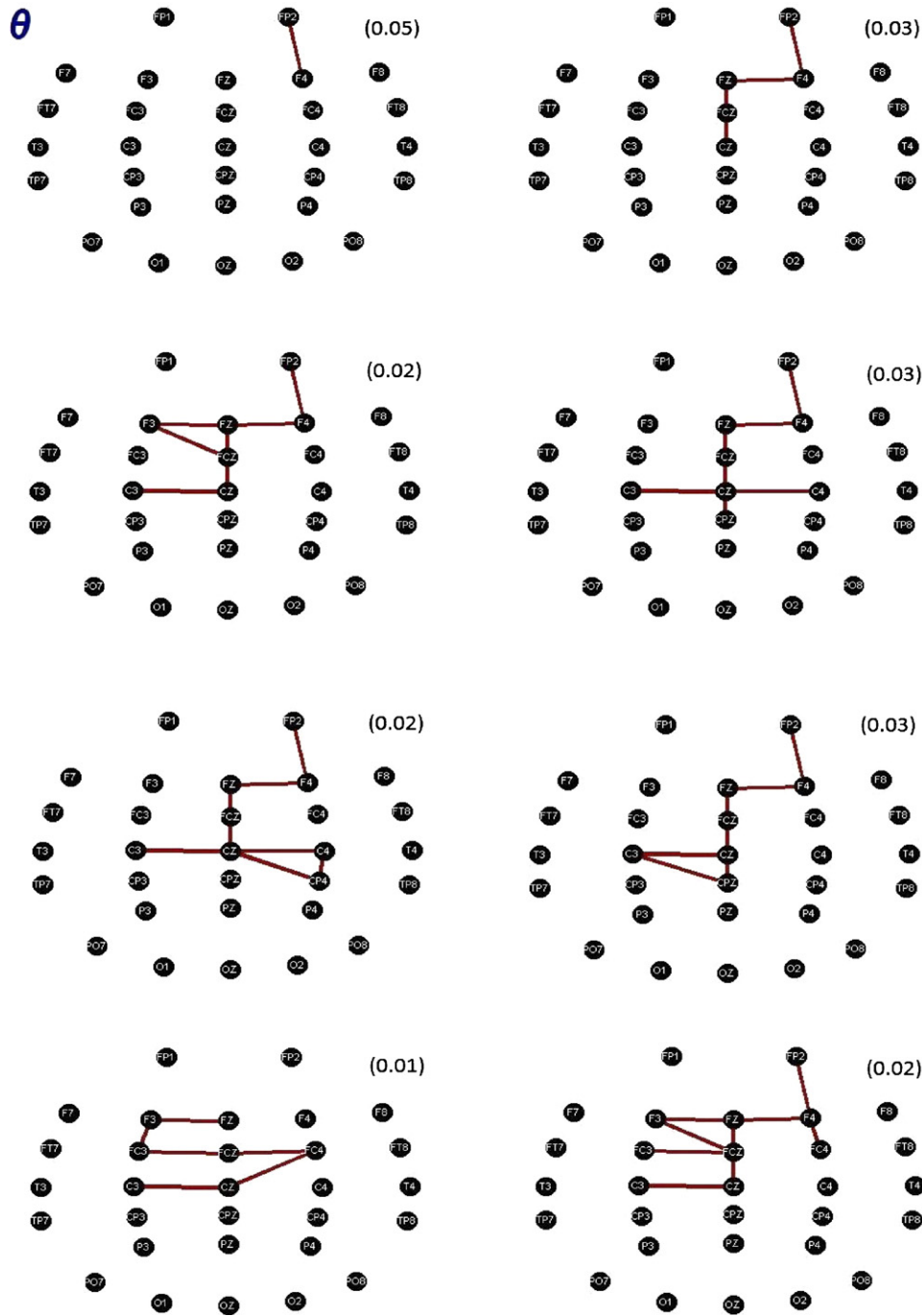


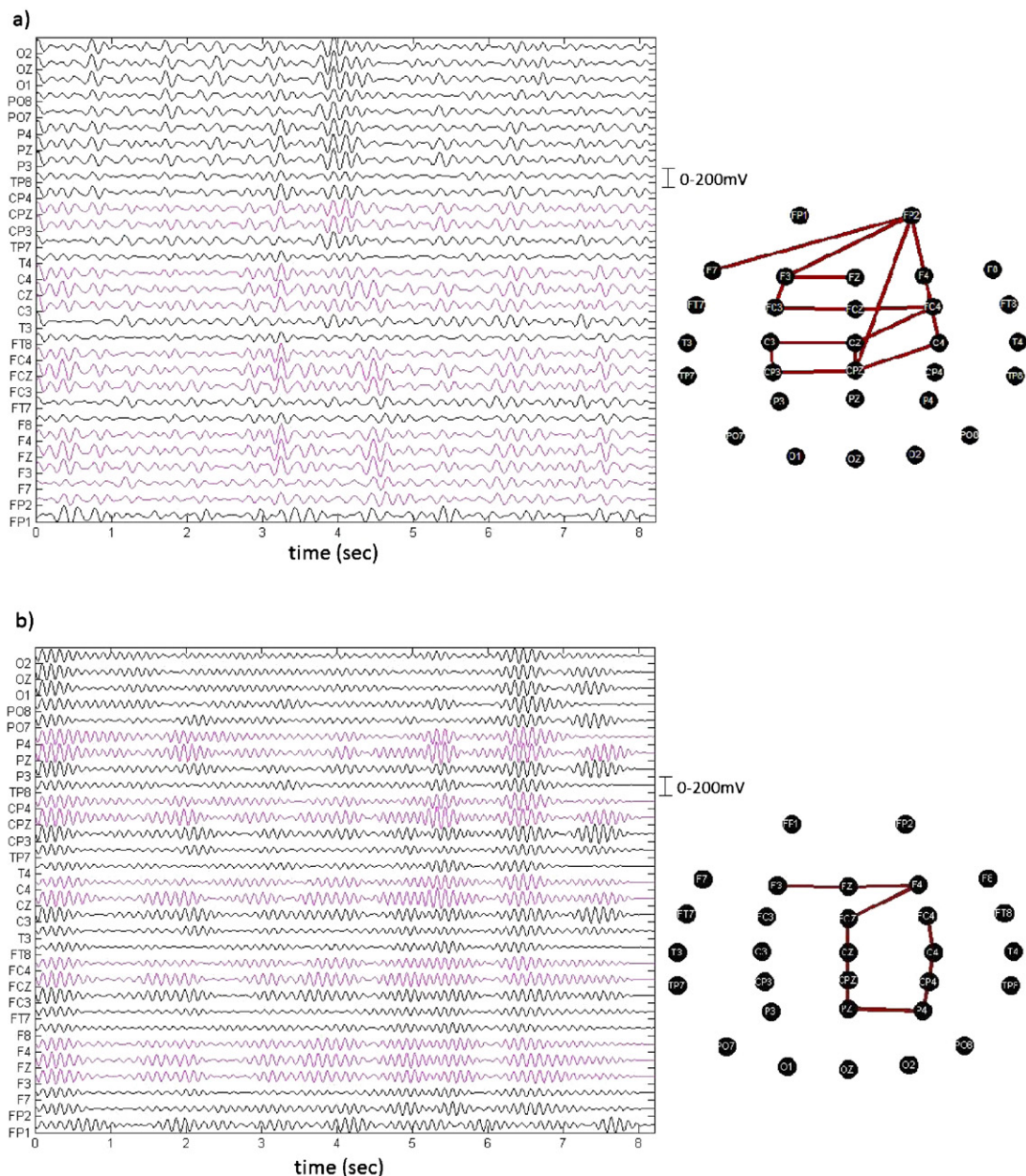
Fig. 4. Various, non-significant, motifs detected from control condition graphs in a percentage  $\rho \geq 90\%$  of subjects. Their  $P$ -values are shown in parenthesis.

property for motifs (Milo et al., 2002). In  $\theta$ -band, motifs include extended fronto-central brain areas. In  $\alpha_1$ -band the motifs are extending over fronto-central areas with a lateralization in the right hemisphere. In  $\alpha_2$ -band only a single motif was detected over the frontal midline. The characteristic motif for  $\beta$ -band was located over centro-parietal areas with a lateralization to the left-hemisphere. Finally, the two  $\gamma$ -band related motifs were associated with frontal-central-parietal brain areas and extended in both hemispheres.

To indicate the association of detected motifs with the original EEG traces, we have included in Fig. 5 the multichannel signal counterpart (the signals filtered in the corresponding frequency band from a randomly selected subject) for two of the above

mentioned motifs. Since motif extraction is based on static graphs and a phase synchrony estimator (the PLV), the brain activity signals from the sensors included in the detected motif (highlighted waveforms) show coherent oscillations for the whole time period.

In order to justify the significance of each motif extracted via the gSpan algorithm, we incorporated a surrogate-assisted analysis step (Ansmann and Lehnertz, 2012) by generating 1000 random adjacency matrices with the following iterative procedure. At each iteration, we first selected at random one of the 18 adjacency matrix  $A^{\text{subj.id}}$  corresponding to resting state connectivity and then shuffled its entries. Adopting the strategy of permutation tests (Good 2000), the count of perfect match between a motif and the set of random graphs (i.e. how many times the motif was contained in



**Fig. 5.** Using single subject data to indicate the grouping of time series (highlighted traces) corresponding to the nodes that participated in a motif. (a)  $\theta$  band (left motif in Fig. 3) and (b)  $\alpha 1$  band (left motif in Fig. 3).

the surrogates) can be interpreted as a  $P$ -value. In all cases shown in Fig. 3, the results indicated significant deviations ( $P < 0.001$ ) from a random process. On the contrary, the motifs included in Fig. 4 are characterized by lower significance.

### 3.2. Contrastive motif extraction for the multiplication task using static FCGs

Task specific motifs with score  $\geq 0.7$  (as depicted in Fig. 2b) related to the cognitive processes involved in the multiplication were detected in  $\theta$ ,  $\beta$  and  $\gamma$  bands. In  $\theta$ -band, motifs were extended in the fronto-central and in the fronto-central-temporal-parietal with a lateralization to the right hemisphere while in  $\beta$ -band motifs were distributed in the fronto-central and in the fronto-central-parietal with a lateralization to the right hemisphere. Finally, in  $\gamma$ -band, characteristic motifs were located in the fronto-central brain areas with a lateralization to the right hemisphere (see Fig. 6).

### 3.3. Contrastive motif extraction for the multiplication task using time-varying FCGs

The extracted motifs from *tv*-FCGs were different from the ones mined using static FCG from the same subjects. A motif example is shown in Fig. 7a. It is associated to the  $\theta$ -band and emerged during multiplication (in more than 50% of the *tv*-FCGs) but not during resting state. Its structure indicates connections from the left-temporal brain areas (areas responsible for language functions) to the right-frontal ones (see Fig. 7a). The particular motif was appearing or disappearing in time (see Fig. 7b). This dynamic behavior is subject dependent and could be interpreted as signaling the access to language brain areas. The frequency of appearance differs across subjects (see Fig. 7c), and this may relate to the difficulty of the task for each subject (the higher the frequency, the more frequently language areas are accessed, hence the subject conceives the task as more demanding).

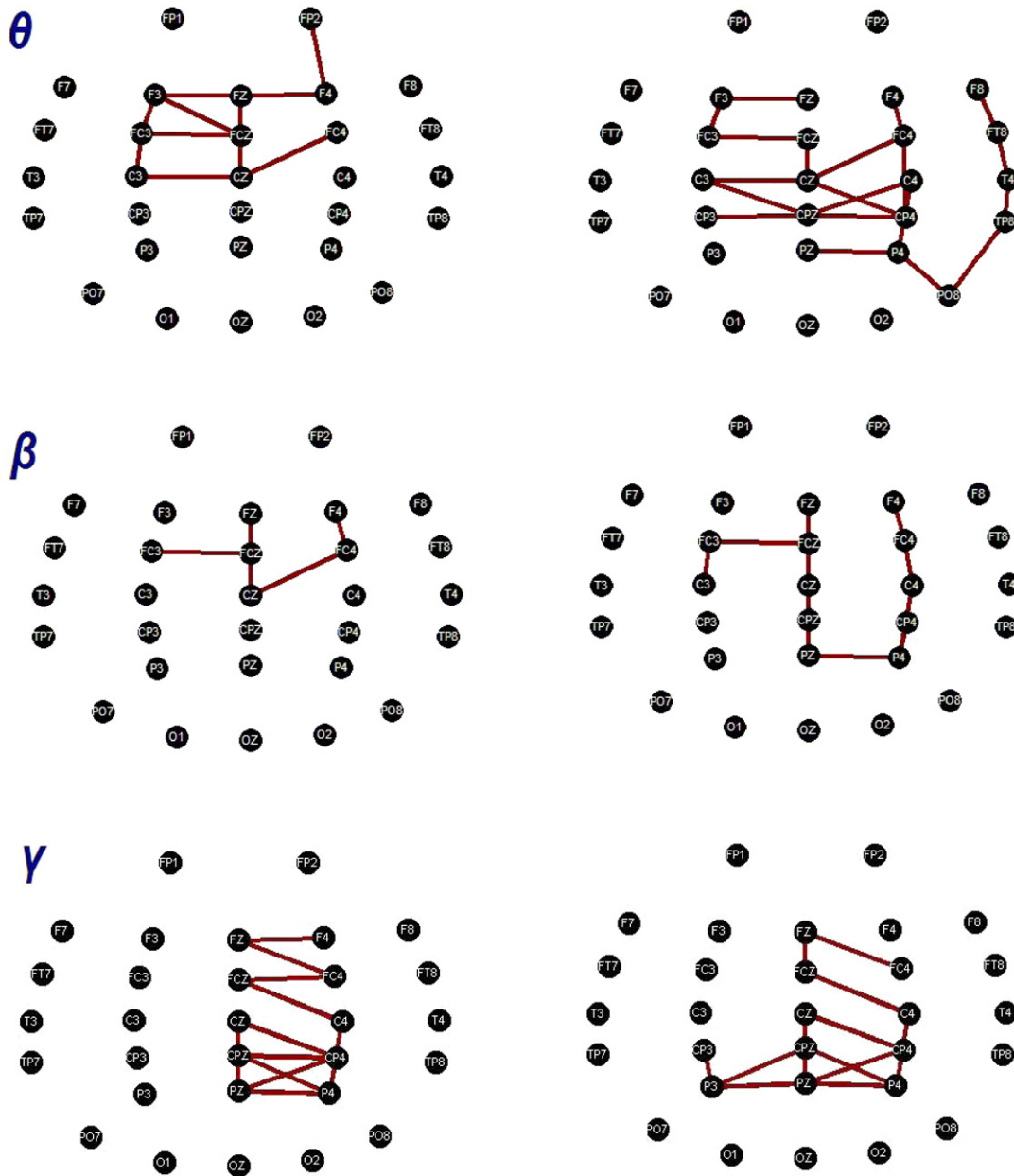


Fig. 6. Multiplication specific motifs extracted via contrastive-learning with respect to resting state, according to Fig. 2b.

In the case of *tv*.FCGs, it is important to scrutinize the time course of each motif and test whether its appearance corresponds to a random process or can be assigned to systematic mechanisms. To this end, we generated 1000 random time courses by

shuffling the entries of each row in Fig. 7b (where 1/0 denotes appearance/disappearance of a motif in time). Then we compute the correlation of the original time course with each one of the randomized ones and counted the number of times the correlation

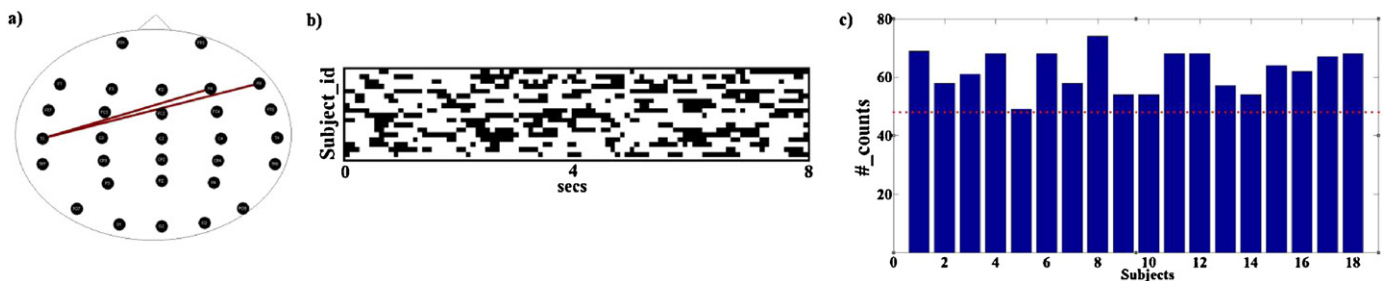


Fig. 7. (a) Multiplication specific motif extracted from  $\theta$ -band, via contrastive-learning with respect to resting state, based on *tv*.FCGs. (b) Visualization of the time course of motif appearance for all 18 subjects (white indicates that the motif is 'on'). (c) Total number of appearances per subject for the motif of (a).



value was greater than 0.9. This count was transformed to a  $P$ -value, in accordance to permutation tests (Good 2000). Again in all cases, the results indicated significant deviations ( $P < 0.001$ ) from a random process.

#### 4. Discussion

We have introduced a new approach to detect brain sub-network patterns using ensembles of either static or time-varying graphs of functional connectivity. The discovery of group-consistent motifs proceeds so as to either characterize a particular brain state (or recording-condition) or a brain state with respect to a reference state (or baseline condition). “Our motifs” refer to a common connected set of labeled nodes and their functional role can be directly deduced due to the topographic configuration of the included nodes.

The key idea is the employment of an algorithm that detects intelligently, size-free network motifs instead of exhaustively enumerating the number of occurrences of each subnetwork of a given size  $k$  (Milo et al., 2002). In previous attempts (Sporns and Kötter, 2004; De Vico Fallani et al., 2007a,b), the investigators had to limit the search space by setting the parameter  $k$  to values up to 4 in order to avoid unreasonable execution times. On the contrary, our approach does not force any restriction. Two recently proposed studies introduced motif discovery algorithms that incorporated the notion of vertex labels and also the detection of significantly presented motifs in a group of graphs (Kashani et al., 2009; Schmidt et al., 2012). However, the proposed algorithms suffer from need of enumerating all the possible subnetworks of a certain size and hence its use is feasible only in the case of moderate-sized sets of graphs and for detecting motifs of small size.

An important methodological contribution in this paper is the contrastive learning scheme for motif extraction, which includes a procedure that scores and finally ranks the detected motifs according to their importance. Another advantage is that within the same methodological framework not only static FCGs, but also dynamic FCGs can be analyzed. This is considered an important option especially in the case of fast-recording modalities (EEG/MEG) and, occasionally, fMRI (Chang and Glover, 2010). Additionally, the proposed methodology can handle the analysis of brain networks of considerable size, while enabling the detection of consistent (within a population or a condition) motifs of variable size.

While resting state EEG has been studied extensively based on power spectral features (Barry et al., 2007; Chen et al., 2008), studies of connectivity have just started to appear (Dimitriadis et al., 2012; Schmidt et al., 2012). The application of our motif extraction procedure to an eyes-open recording condition revealed consistent motifs for the entire set of frequency bands under investigation (see Fig. 3). The motifs extracted in  $\theta$  and  $\gamma$  bands were more spatially extended compared to the rest ones (see Fig. 3). This experimental observation is consistent with the correlation of  $\theta$  and  $\gamma$  bands with the internal processing of personalized information. Specifically, it is considered that  $\theta$ -band during resting state reflects switching, dreamlike thoughts, while  $\gamma$ -band reflects thinking and integrated thoughts (Von Stein and Sarnthein, 2001).

In the case of task-related data, multiplication-specific motifs were revealed in  $\theta$ ,  $\beta$  and  $\gamma$  bands (Fig. 6). The  $\theta$ -motifs were located over fronto-central and fronto-central-temporal-parietal regions with a right hemisphere lateralization, which can be interpreted as an integration of cognitive functions involved in working memory (Sauseng et al., 2010) and in fact retrieval from long-term memory (Dehaene et al., 1999; Sauseng et al., 2002; Delazer et al., 2003). The  $\beta$ -band motifs had a distribution that could be characterized as fronto-central and fronto-central-parietal with right hemisphere lateralization. This band has been suggested to play

an important role during attention (Bekisz and Wrobel, 2003) or higher cognitive functions (Razumnikova, 2004). The characteristic motifs of  $\gamma$ -band were located over fronto-central brain areas with a right hemisphere lateralization as well. These motifs most likely reflect retrieval of information independent of sensory modality but related to the cognitive mental task and memory processes (Tallon-Baudry and Bertrand, 1999; Hermann et al., 2004; Kaiser and Lutzenberger, 2005; Sauseng and Klimesch, 2008). Finally, the absence of  $\alpha_{1,2}$ -band related motifs (due to low score), probably means that there is no difference in attentional (Von Stein and Sarnthein, 2001) and memory processes developed via  $\alpha$ -band activity (Klimesch et al., 2005). A recent study (Micheloyannis et al., 2005), analyzing a two-digits multiplication task revealed a widespread pattern in four lobes of the brain reflecting demanding cognitive functions (working memory, retrieval)

The analysis of tv\_FCGs in  $\theta$ -band during the multiplication task revealed an interesting motif that associates left-temporal brain areas with right frontal areas (Fig. 7a). The left temporal lobe is known to be critical for language comprehension and production. Additionally, this fronto-temporal motif possibly reflects the engagement of a network of areas responsible for language functions, including long-term and verbal working memory (Dehaene et al., 1999; Delazer et al., 2003). By tracking the appearance of this motif along time, we revealed an important dynamic pattern that was masked in the study of static FCGs. The chessboard pattern seen in Fig. 7b can be interpreted as a time-course reflecting the access of language brain areas. It is evident that the particular motif (reflecting mostly the performance of mental calculations) disappears for short time periods.

Our approach is applicable to data from other neuroimaging techniques as well (e.g., MEG, MRI/fMRI, diffusion tensor imaging (DTI)). The network size varies for different modalities ranging from tens (EEG, fMRI, DTI) to hundreds (~300) nodes (MEG) (Frye et al., 2010), a parameter that increases the computational load for motif detection. However, such network sizes are computationally affordable for the gSpan algorithm (Yan and Han, 2002). With an appropriate estimator of the functional connectivity that can be pair-wise or  $n$ -way, undirected or directed, important graph patterns can be revealed that will help us understand the speed and nature of self-organization during cognitive states (Bassett and Bullmore, 2009).

Finally, the study of binary graphs is the first step in this research direction. Our primary goal was to ascertain that the introduced technique can provide promising and interesting results. The full exploitation of connectivity weights (based on recent methods for weighted motif detection (Jiang et al., 2010; Yang et al., 2012)) will be considered in the future in order to avoid any arbitrariness induced by the binarization step and also to enhance the quality of the obtained results.

#### References

- Achard S, Salvador R, Whitcher B, Suckling J, Bullmore E. A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *J Neurosci* 2006;26:63–72.
- Ansmann G, Lehnertz K. Surrogate-assisted analysis of weighted functional brain networks. *J Neurosci Methods* 2012;208(2):165–72.
- Barry RJ, Clarke AR, Johnstone SJ, Magee CA, Rushby JA. EEG differences between eyes-closed and eyes-open resting conditions. *Clin Neurophysiol* 2007 December;118(12):2765–73.
- Bassett DS, Bullmore E. Small-world brain networks. *Neuroscientist* 2006;12:512–23.
- Bassett DS, Bullmore ET. Human brain networks in health and disease. *Curr Opin Neurol* 2009;22:340–7.
- Bekisz M, Wrobel A. Attention-dependent coupling between beta activities recorded in the cat's thalamic and cortical representations of the central visual field. *Eur J Neurosci* 2003;17:421–6.
- Bullmore E, Sporns O. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci* 2009;10:186–98.

- Chang C, Glover GH. Time-frequency dynamics of resting-state brain connectivity measured with fMRI. *Neuroimage* 2010;50:81–98.
- Chen AC, Feng W, Zhao H, Yin Y, Wang P. EEG default mode network in the human brain: spectral regional field powers. *Neuroimage* 2008;41:561–74.
- Dehaene H, Spelke E, Pinel P, Stanescu R, Tsivkin S. Sources of mathematical thinking: behavioral and brain-imaging evidence. *Science* 1999;284:970–4.
- Delazer M, Domahs F, Bartha L, Breneis C, Lochy A, Trieb T, et al. Learning complex arithmetical fMRI study. *Cogn Brain Res* 2003;18:76–88.
- Dimitriadis SI, Laskaris NA, Del Rio-Portilla Y, Koudounis GC. Characterizing dynamic functional connectivity across sleep stages from EEG. *Brain Topogr* 2009;22(September (2)):119–33.
- Dimitriadis SI, Laskaris NA, Tsirka V, Vourkas M, Micheloyannis S, Fotopoulos S. Tracking brain dynamics via time-dependent network analysis. *J Neurosci Methods* 2010a;193(October (1)):145–55.
- Dimitriadis SI, Laskaris NA, Tsirka V, Vourkas M, Micheloyannis S. What does delta band tell us about cognitive processes: a mental calculation study. *Neurosci Lett* 2010b;483(October (1)):11–5.
- Dimitriadis SI, Laskaris NA, Tsirka V, Vourkas M, Micheloyannis S. An EEG study of brain connectivity dynamics at the resting state. *Nonlinear Dynamics Psychol Life Sci* 2012;16(January (1)):5–22.
- De Vico Fallani F, Latora V, Astolfi L, Cincotti F, Mattia D, Marciani MG, et al. Persistent patterns of interconnection in time-varying cortical networks estimated from high-resolution EEG recordings in humans during a simple motor act. *J Phys A Math Theor* 2007a;41:224014.
- De Vico Fallani F, Astolfi L, Cincotti F, Mattia D, Marciani MG, Salinari S, et al. Cortical functional connectivity networks in normal and spinal cord injured patients: evaluation by graph analysis. *Hum Brain Mapp* 2007b;28:1334–46.
- De Silva E, Stumpf M. Complex networks and simple models in biology. *J R Soc Interface* 2005;22:419–30.
- Faloutsos M, Faloutsos P, Faloutsos C. On power-law relationships of the internet topology. *Comput Commun Rev* 1999;29:251–62.
- Frye R, Wu M, Liederman J, Fisher J. Greater pre-stimulus effective connectivity from the left inferior frontal area to other areas is associated with better phonological decoding in dyslexic readers. *Front Syst Neurosci* 2010;4.
- Hermann CS, Munk MHJ, Engel AK. Cognitive functions of gamma activity: memory match and utilization. *Trends Cogn Sci* 2004;8:347–55.
- Jiang C, Coenen F, Zito M. Frequent sub-graph mining on edge weighted graphs. In: Proceedings of the 12th international conference on data warehousing and knowledge discovery (DaWaK'10). Berlin, Heidelberg: Springer-Verlag; 2010. p. 77–88.
- Joyce KE, Laurienti PJ, Burdette JH, Hayasaka S. A new measure of centrality for brain networks. *PLoS ONE* 2010;5(8):e12200. <http://dx.doi.org/10.1371/journal.pone.0012200>.
- Kaiser J, Lutzenberger W. Human gamma-band activity: a window to cognitive processing. *Neuroreport* 2005;16:207–11.
- Kashani RM, Ahrabian Z, Elahi H, Nowzari-Dalini E, Saberi A, Asadi E, et al., Schreiber S. Kavosh: a new algorithm for finding network motifs. *BMC Bioinform* 2009;10:318.
- Klimesch W, Schack B, Sauseng P. The functional significance of theta and upper alpha oscillations for working memory: a review. *Exp Psychol* 2005;52:99–108.
- Kolaczyk ED. Statistical analysis of network data: methods and models (Springer series in statistics) [Hardcover]. 1st ed. New York: Springer; 2009 March.
- Lachaux JP, Rodriguez E, Martinerie J, Varela FJ. Measuring phase synchrony in brain signals. *Hum Brain Mapp* 1999;8:194–208.
- Micheloyannis S, Sakkalis V, Vourkas M, Stam CJ, Simos PG. Neural networks involved in mathematical thinking: evidence from linear and non-linear analysis of electroencephalographic activity. *Neurosci Lett* 2005;373:212–7.
- Micheloyannis S, Pachou E, Stam CJ, Vourkas M, Erimaki S, Tsirka V. Using graph theoretical analysis of multi channel EEG to evaluate the neural efficiency hypothesis. *Neurosci Lett* 2006;402(July (3)):273–7.
- Micheloyannis S, Sakkalis V, Vourkas M, Tsirka V, Karakonstanti E, Kanatsouli K, et al. The influence of ageing on complex brain networks: a graph theoretical analysis. *Hum Brain Mapp* 2009;30(1):200–8.
- Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U. Network motifs: simple building blocks of complex networks. *Science* 2002;298:824–7.
- Milo R, Itzkovitz S, Kashtan N, Levitt R, Shen-Orr S, Ayzenshtat I, et al. Superfamilies of evolved and designed networks. *Science* 2004;303:1538–42.
- Razumnikova OM. Gender differences in hemispheric organization during divergent thinking: an EEG investigation in human subjects. *Neurosci Lett* 2004;362:193–5.
- Rubinov M, Knock SA, Stam CJ, Micheloyannis S, Harris AW, Williams LM, et al. Small world properties of nonlinear brain activity in schizophrenia. *Hum Brain Mapp* 2009;30:403–16.
- Rubinov M, Sporns O. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* 2010;52:1059–69.
- Sauseng P, Klimesch W, Gruber W, Doppelmayr M, Stadler W, Schabus M. The interplay between theta and alpha oscillations in the human electroencephalogram reflects the transfer of information between memory systems. *Neurosci Lett* 2002;324:121–4.
- Sauseng P, Klimesch W. What does phase information of oscillatory brain activity tell us about cognitive processes? *Neurosci Biobehav Rev* 2008;32:1001–13.
- Sauseng P, Griesmayr B, Freunberger R, Klimesch W. Control mechanisms in working memory: a possible function of EEG theta oscillations. *Neurosci Biobehav Rev* 2010;34:1015–22.
- Schmidt C, Weiss T, Komusiewicz C, Witte H, Leistriz L. An analytical approach to network motif detection in samples of networks with pairwise different vertex labels. *Comput Math Methods Med* 2012;2012:12. <http://dx.doi.org/10.1155/2012/910380>, Article ID 910380.
- Simpson SL, Moussa MN, Laurienti PJ. An exponential random graph modeling approach to creating group-based representative whole-brain connectivity networks. *Neuroimage* 2012;60(April (2)):1117–26 [Epub 2012 January 17].
- Sporns O, Tononi G, Edelman GM. Theoretical neuroanatomy: relating anatomical and functional connectivity in graphs and cortical connection matrices. *Cereb Cortex* 2000;10:127–41.
- Sporns O, Kötter R. Motifs in brain networks. *PLoS Biol* 2004;2(11):e369. <http://dx.doi.org/10.1371/journal.pbio.0020369>.
- Stam CJ, Reijneveld JC. Graph theoretical analysis of complex networks in the brain. *Nonlinear Biomed Phys* 2007;1:3.
- Tallon-Baudry C, Bertrand O. Oscillatory gamma activity in humans and its role in object representation. *Trends Cogn Sci* 1999;3:151–62.
- Valencia M, Martinerie J, Dupont S, Chavez M. Dynamic small-world behavior in functional brain networks unveiled by an event-related networks approach. *Phys Rev E* 2008;77, 050905(R).
- van Wijk BCM, Stam CJ, Daffertshofer A. Comparing brain networks of different size and connectivity density using graph theory. *PLoS ONE* 2010;5(10):e13701. <http://dx.doi.org/10.1371/journal.pone.0013701>.
- Vanetik N, Shimony SE, Gudes E. Support measures for graph data. *Data Min Knowl Disc* 2006;13(2):243–60.
- Vértes PE, Alexander-Bloch AF, Gogtay N, Giedd JN, Rapoport JL, Bullmore ET. Simple models of human brain functional networks. *Proc Natl Acad Sci* 2012. <http://dx.doi.org/10.1073/pnas.1111738109>.
- Von Stein A, Sarnthein J. Different frequencies for different scales of cortical integration: from local gamma to long-range alpha/theta synchronization. *Int J Psychophysiol* 2001;38:301–13.
- Yan X, Han J. gSpan: graph-based substructure pattern mining. In: ICDM 2002; 2002. p. 721–4.
- Yang J, Su W, Li S, Dalkılıç M. WIGM: discovery of subgraph patterns in a large weighted graph. In: SDMSIAM. Anaheim, California, USA: Omnipress; 2012. p. 1083–94.