Machine learning with brain graphs

Jonas Richiardi, *Member, IEEE,* Sophie Achard, Horst Bunke, and Dimitri Van De Ville, *Member, IEEE*

I. Introduction

A. Functional MRI as a fundamental tool for neuroscience

The observation and description of the living brain has attracted a lot of research over the past centuries. Many non-invasive imaging modalities have been developed, such as topographical techniques based on the electromagnetic field potential (i.e., electroencephalography and magnetoencephalography), and tomography approaches including positron emission tomography and magnetic resonance imaging (MRI). Here we will focus on functional MRI (fMRI) since it is widely deployed for clinical and cognitive neurosciences today, and it can reveal brain function due to neurovascular coupling (see BOX). It has led to a much better understanding of brain function, including the description of brain areas with very specialised functions such as face recognition. These neuroscientific insights have been made possible by important methodological advances in MR physics, signal processing, and mathematical modelling.

B. A network perspective on the brain

Early analysis of fMRI data looked for correlational evidence of brain regions being related to specific functions (known as *functional segregation*). However, it became obvious that the brain operates as a global complex system with many interactions (*functional integration*). While a large body of work in the literature focuses on structural connectivity of the brain, that is, how white matter interconnects brain regions, some authors have attempted to model and characterise the brain as a network using functional connectivity (i.e., temporal correlation between remote brain regions) [1]. This network-centric perspective has led to fundamental insights in terms of the organisation of the healthy and diseased brain [2], how its resilient network architecture allows it to withstand injury [3], and how evolutionary arguments can be advanced for the distributed information processing it performs [4].

This relatively recent trend towards formalising integration and segregation of brain function borrowed many tools and concepts from statistical physics, graph theory, sociology, and statistics, and led naturally to the adoption of graphs as an essential mathematical tool. Indeed, the popularity of graph-based approaches in contemporary neuroscience is easily understood: graphs offer a proper language to describe whole-brain patterns and inter-regional interactions. Besides, neuroimagery provides us with time series, associated with voxels, which reflect information processing by a brain region in time. These data have brought to light the dynamic nature of the brain, which follows complex temporal patterns. The complexity of fMRI data (including low signal-to-noise ratio, spatial correlations, long-range temporal dependencies, and high dimensionality) and the importance of capturing spatio-temporal dependencies make it very desirable to find a level of abstraction at which inference can be performed. Graphs have the desirable property of being able to represent data at many spatial (and temporal) resolutions, meaning that the same mathematical models and algorithms can be applied at different spatial and temporal scales. Moreover, the semantics associated with a graph, i.e., the meaning of its nodes and edges, is flexible and can be chosen depending on the underlying application.

JR is jointly affiliated with Stanford University, USA, and the University of Geneva, Switzerland.

DVDV is with the Medical Image Processing Laboratory at Ecole Polytechnique Fédérale de Lausanne and the University of Geneva, Switzerland

SA is with the GIPSA-lab at Centre National de la Recherche Scientifique, Grenoble, France

HB is with the Institute of Computer Science and Applied Mathematics at the University of Bern, Switzerland

C. Machine learning on brain graphs

In parallel with the rise of interest in brain networks, there has been an increase in the use and development of machine learning techniques in neuroscience [5]. Indeed, the high-dimensional nature of functional magnetic resonance imaging (fMRI) data hinders the application of many multivariate methods from classical statistics, prompting an increasing number of researchers to rely on regularisation methods common in machine learning and signal processing in addition to well-established mass-univariate analysis techniques. Furthermore, inference at the level on single subjects is gaining prominence, with many new developments in the field of neuroimaging marker development. Predictive modelling using machine learning techniques are therefore particularly suitable to the field, and are now commonly applied to cognitive, clinical, affective, and social neuroscience. The interest for these techniques is evident in practitioners, and most neuroimaging conferences have special sessions on machine learning. Concurrently, workshops on the topic are regularly held at machine learning conferences, and dedicated meetings are emerging.

The intersection of statistical machine learning techniques and graph representations has been of interest for several years in fields such as computer vision, pattern recognition, and data mining (as evidenced by regular workshops such as GbR (Graph-Based Representations in Pattern Recognition), SSPR (Structural and Syntactic Pattern Recognition), or MLG (Mining and Learning with Graphs)), but has only relatively recently started to be exploited in the context of brain networks, and formalising neuroscientific questions as graph classification problems is a very recent trend. We believe that applying machine learning techniques to brain connectivity data, for example, by following the scheme in Figure 1, has unique potential. Given the current appeal of graphs for brain data representation and the simultaneous enthusiasm for machine learning approaches in the neuroimaging community, we expect this emerging approach to see increasing adoption. In particular, clinical applications were among the first to appear; in cognitive neuroscience, previously unseen relationships can be uncovered and the hypothesis of no effect can be more convincingly rejected. From a methodological point of view, because BOLD fMRI data is particularly challenging to work with for the reasons mentioned previously, there is also large prospective advancement in signal processing and machine learning.

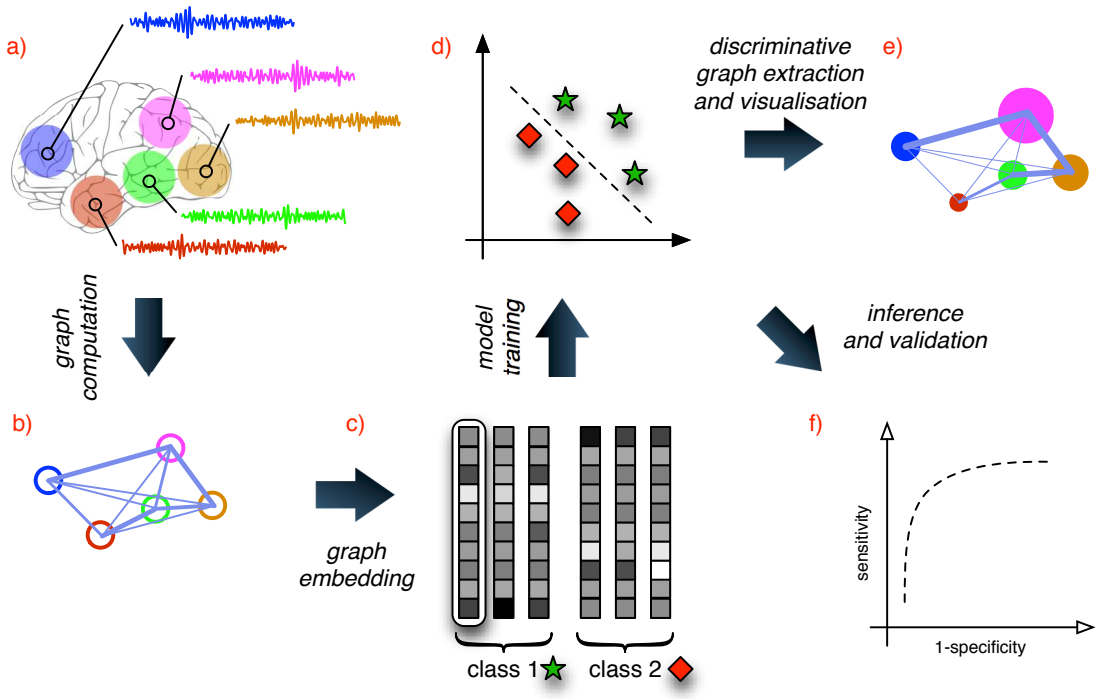

Fig. 1. Overall scheme for predictive modelling with brain graphs. (a) Imaging data is first preprocessed, then the brain is divided into regions, and each region is assigned a regional representative time series. (b) A labelled simple graph is computed from the regional time series, where edge labels correspond to statistical dependency between brain regions, and brain regions are mapped to graph vertices. (c) The graph is embedded into a vector space, after which (d) statistical machine learning can be used. (e) Brain-space visualisation of the discriminative pattern used by the classifier is critical for interpretation. (f) Statistics and confidence intervals can be obtained on inference results, allowing validation of the techniques when used e.g. to elicit imaging markers in clinical applications.

BOX: FROM BRAIN IMAGES TO FMRI TIME SERIES

Neuronal clusters involved in brain activity consume more oxygen compared to their baseline state. Due to neurovascular coupling, blood flow and volume are increased and lead to a significant overcompensation of the oxygen demands; i.e., the ratio of oxygenated and deoxygenated haemoglobin is altered. Deoxygenated haemoglobin is paramagnetic and acts as an endogenous contrast agent since it alters the T2*-weighted MR images. This gives rise to the blood-oxygen-level-dependent (BOLD) signal, discovered in the 1990s, which has allowed MRI to become functional (fMRI) and to observe the brain at work.

MRI allows sampling a 3D volume of the brain at millimetric spatial resolution every 1–3 seconds (or faster with recent sequences). This way, we obtain multivariate time series of brain activity. Raw fMRI signals suffer from low signal-to-noise ratio, and need to be processed heavily to be amenable to analysis. Several pre-existing open source software packages allow reliable results to be obtained rapidly[1]. The main preprocessing steps, illustrated in Figure 2, are to realign the volumes to compensate for subject motion and ensure voxel-to-voxel correspondence across time, to coregister functional images to a high-resolution structural image, and to normalise the data into a common reference space so that subjects can be compared and existing anatomical knowledge can be leveraged. Once this is achieved, representative time-series can be extracted from different brain regions and serve as a basis for brain graph construction (see Sections II-B and II-C).

We point out that this is only one possible pipeline, and there is not necessarily a consensus in the field [6]. Our guiding principle here is to avoid over-processing the functional data. For example, we advocate avoiding upsampling the functional data to structural resolution, which in typical settings (1 mm isotropic structural voxels, 3mm isotropic functional voxels) would results in a close to 30-fold increase in the amount of data with no additional information gain. A principled choice of the optimal preprocessing steps and their order for the application of interest can be guided by several objectives, for example a pattern reproducibility / model generalisability compromise as advocated by the NPAIRS approach to pipeline evaluation [7].

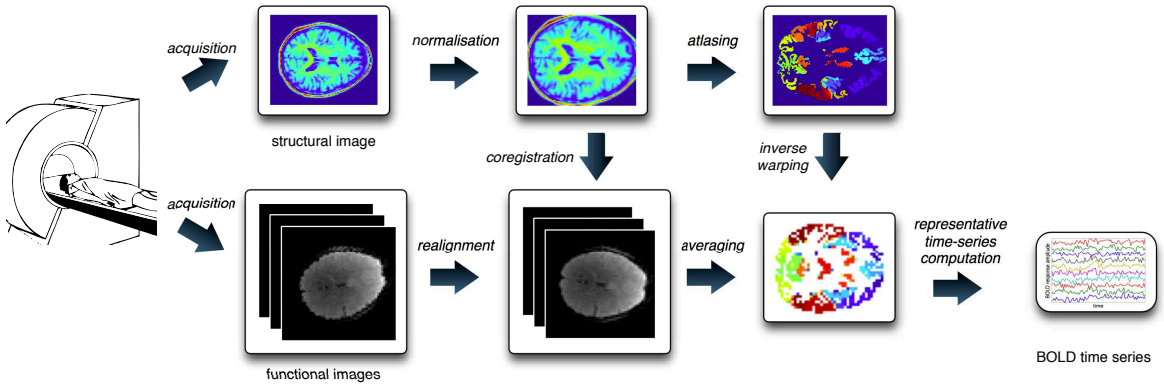

Fig. 2. Overview of preprocessing for time series extraction from fMRI data, including atlas-based parcellation of the brain.

II. FROM IMAGING DATA TO CONNECTIVITY GRAPHS

With imaging data preprocessed, several additional steps are necessary to obtain "brain graphs". In particular, a mapping between brain space and vertices must be defined, a representative time series per vertex chosen, and graph edges labelled. First, however, proper mathematical formalism has to be put in place.

A. Mathematical definition

Formally, a graph $g = (V, E)$ consists of a finite set V of vertices and a finite set of edges $E \subseteq V \times V$. We say there is an edge from vertex i to vertex j if $(i, j) \in E$. Graphs can be directed or undirected. In the first case, the direction of an edge matters, while in the second case we assume that for each edge (i, j) there exists an edge (j, i) in the opposite direction, such that the direction of an edge is not important any longer. A slight generalisation of

[1]SPM: http://www.fil.ion.ucl.ac.uk/spm/, FSL: http://www.fmrib.ox.ac.uk/fsl/, Freesurfer: http://surfer.nmr.mgh.harvard.edu/

this definition is achieved by multi-graphs, where several edges $e_1, e_2, \ldots$ all pointing from the same node i to the same node j can exist. Graphs that have at most one edge between any pair of nodes are also called *simple* graphs.

For the subject at hand, *labelled simple graphs* (where labels are defined over both vertices and edges, and are members of the sets L_V and L_E respectively) are expressive enough to represent many properties of interest in brain connectivity graphs. For a particular graph g, representing either a subject's functional connectivity or a particular brain state, we can write [8]:

$$g = (V_g, E_g, \alpha_g, \beta_g), \tag{1}$$

where V_g is the set of vertices, E_g is the set of edges, and $\alpha_g : V_g \to L_V$ and $\beta_g : E_g \to L_E$ are respectively the vertex labelling and edge labelling functions.

The labelling functions are essential to learning and inference on brain graphs — for example by letting $L_E = \mathbb{R}$ we can obtain a scalar weight on each edge, which can encode the strength of the statistical dependency between brain regions, and provide more information than the fact that $(i, j) \in E_g$. Likewise, $L_V = \mathbb{R}^+$ could be used to label vertices with graph-theoretical attributes such as the centrality of a vertex [9]

By further restricting the vertex set V_g to have a fixed ordering (to be a sequence), and all graphs in the class to have the same number of vertices (to have a fixed-cardinality vertex sequence, so $\forall s, |V_g| = R$), a considerably simpler graph comparison and analysis problem results. In particular, when comparing two graphs, the vertex assignment problem, which has exponential complexity in the general case, is avoided. Graphs with unique node labels [8] benefit from the same simplification. This allows the engineering effort to be spent on defining vertices, vertex labels, edges, and edge labels, as well as vector space embedding techniques. From a neuroscience perspective, having the same set of vertices for all subjects allows easier inter-subject comparisons, and is a way of abstracting away the important individual anatomical variability that exists in human brains.

B. Vertices in brain space

Understanding inference results on brain graphs requires a link with the underlying neural substrate. Indeed, choosing how image voxels map to graph vertices will have a large influence on the meaning of the resulting graphs, and edges will represent interactions between these brain systems. Broadly speaking, methods can be anatomy-driven or data-driven, and yield contiguous or non-contiguous sets of voxels, which can have empty or non-empty intersections. Thus, each vertex $v_i \in V_g$ is mapped to a set of image voxels $\mathcal{V}_i \subseteq \mathcal{V}$. This choice also dictates the graph size and the method that can be used to elicit edge labels.

Figure 3 shows three commonly used choices for mapping voxels to vertices. The first approach consists of assigning one vertex per voxel, leading to $|V_g| = N, \forall i\ |\mathcal{V}_i| = N$, where N is the number of voxels. Because this approach is often used to study interactions between one particular region of the brain and the rest of the brain and uses temporal correlation to assess dependencies between voxel timecourses, it is often called *seed-based correlation* [10]. In this case we have $\forall i, j\ \mathcal{V}_i \cap \mathcal{V}_j = \emptyset$ and no spatially non-contiguous subsets exist.

The second approach consists of using anatomical knowledge to divide the brain into R regions of contiguous voxels (see Figure 2 for an example of using a brain atlas for this) [3], [11]. In this case, there is one vertex per region $|V_g| = R \ll N$, leading to smaller graphs than for seed-based approaches (typically in the low hundreds of vertices). We also have $\forall i, j\ \mathcal{V}_i \cap \mathcal{V}_j = \emptyset$, and regions are contiguous. A similar type of graph is obtained by localising spherical regions of interest at coordinates reported in the literature.

The third commonly used approach is to use a data-driven procedure such as spatial independent component analysis (ICA) [12] or clustering [13] to define regions of interest. For example, an ICA decomposition could yield around 20 components, and the voxels included in a thresholded spatial component would serve as the spatial extent of a graph vertex. The temporal dependency between these components can then be examined [14]. This approach yields voxel sets that are spatially disjoint and overlapping — depending on thresholding, a voxel can be claimed by several spatial ICA components, which may considerably complicate interpretation if vertices are to be considered an abstraction of independent voxels. In a clustering approach, a suitable similarity measure between voxel time courses has to be defined (which may or may not be the same as used later to establish dependencies between clusters), and a consistency threshold is chosen above which voxels are said to belong to the same cluster. A representative voxels is then chosen for each cluster, which could be the centroid time course. If the consistency

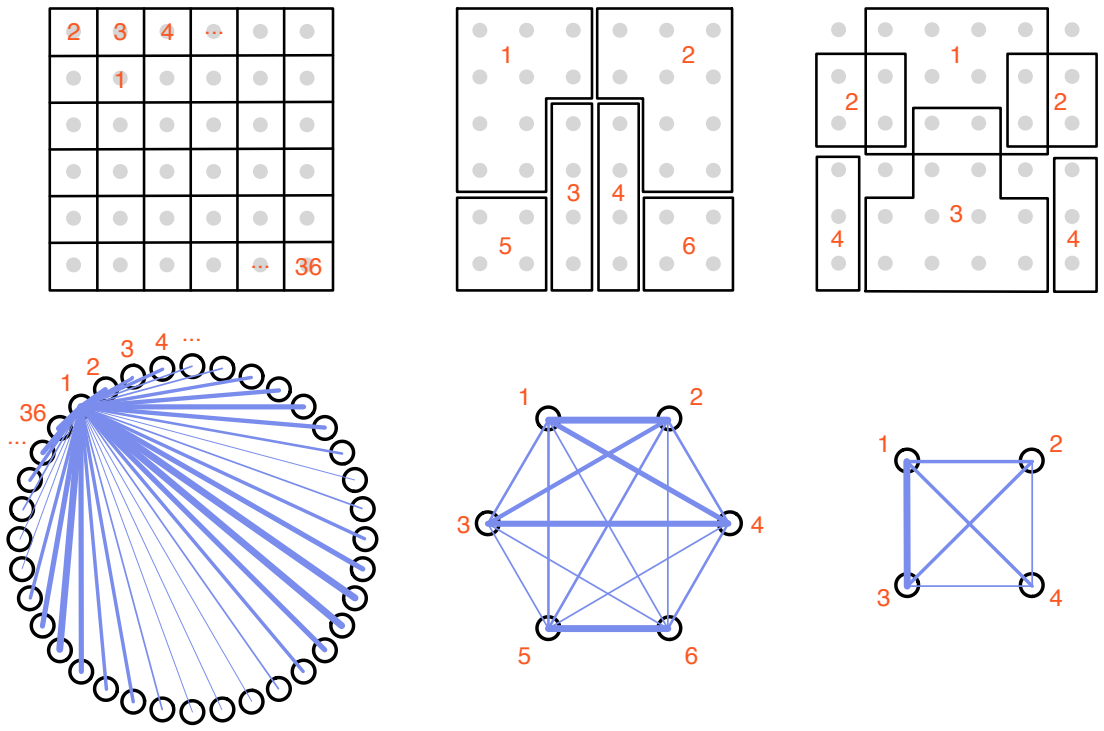

Fig. 3. Defining graph vertices in brain space, two dimensional example. Top row (brain space parcellation): gray dots correspond to voxel center of mass. Black divisions correspond to the brain space definition of a vertex, that is, identify members of each voxel set $\mathcal{V}_i$. Numbers indicate region identifier, corresponding to vertex labels. Bottom row (corresponding graphs): graphs corresponding to the spatial division outlined above. Vertices are shown as black circles, and each vertex maps one-to-one to a spatial region. Edges are represented by blue lines, where line width is shown proportional to the edge weight (encoded as an edge label). From left to right: seed-based voxel-wise approach, atlas-based approach, and ICA-based approach.

threshold can be set arbitrarily high, then clusters can degenerate to single voxels and the approach reverts to seed-based correlation.

In all three cases, it is generally possible to obtain the fixed-cardinality vertex sequence property (see Section II-A): for seed correlation, normalisation and realignment can ensure that this is the case; for atlas-based methods, using the same atlas for all subjects and cognitive states guarantees that the property holds; for data-driven methods such as ICA, a multi-subject technique (e.g. Group ICA [12]) can ensure that the spatial definition of voxels sets mapping to vertices is the same for all subjects and vertices, assuring the property holds there too.

C. Vertex time series

For each fMRI voxel, we obtain a time series that correspond to the BOLD signal recorded inside the voxel. Different strategies have been used in the recent literature to extract representative time series corresponding to regions in brain space and vertices in a graph (as per Figure 3). These strategies depend on the spatial assignment of voxels to vertices.

With seed-based approaches, spatial smoothing is typically used as an attempt to improve the signal-to-noise ratio. The regional representative for each region (voxel) is then a linear mixture of the neighbouring timecourses.

For atlas-based approaches, the two dominant approaches are to compute the temporal mean timecourse of all the voxels within a region [11], and to use this as a representative (an aggressive form of smoothing), or to use the first eigenvariate of the region as a representative. The former is optimal in terms of root mean-squared error, while the latter maximises the explained variance. A generalisation of these approaches is offered by canonical correlation analysis (CCA), where the weights of the voxels' contribution to a regional representative are optimised so that correlation between atlas regions is maximised [15]. In this case, the vertex time series computation and edge label assignment (see section below) are a single step.

Finally, for the data-driven analysis using spatial ICA, the representative time series are computed directly from the functional data. Because spatial independence (rather than temporal independence) is sought in the decomposition,

the representative time series may exhibit significant correlation with representative time series from other spatial components.

As first noted in electroencephalography, the temporal dynamics observed while the brain is functioning usually can be divided in different frequency bands that are related to the rhythms of the brain [16]. This can be done using bandpass filtering with Fourier basis functions or wavelet transforms. While the use of Fourier basis functions is dominant in the literature, the presence of long memory or $1/f$ properties in the cortical fMRI time series [17], make wavelets well suited in this context [18], [19]. In particular, the discrete wavelet transform (DWT) at scales $j = 1, 2 \ldots, J$ for the time series $\mathbf{X}$ is written as

$$\mathbf{X} = \sum_{k \in \mathbb{Z}} c_{J,k} \phi_{J,k} + \sum_{j \leq J} \sum_{k \in \mathbb{Z}} d_{j,k} \psi_{j,k} \tag{2}$$

where $\phi_{j,k}(t) = 2^{-j/2}\phi(2^{-j}t - k)$, $\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t - k)$, $c_{J,k}$ is the approximation coefficient at scale J located at time point k, and $d_{j,k}$ is the detail coefficient at scale j and time point k. In practice, the mother wavelet function $\psi(\cdot)$ needs to have a sufficient number of vanishing moments so that low-order polynomial trends are removed (e.g., to deal with MRI magnet gradient heating effects during acquisition sessions). We opt for the redundant transform to afford shift-invariance, a useful property because the haemodynamic lag is known to differ between brain regions. Specifically, we choose the commonly used redundant 3rd degree Battle-Lemarié wavelet transform.

D. Assigning edge labels to the graph

In fMRI brain graphs, edge labels are typically taken to represent dependencies between the brain regions underlying the connected vertices. Many different techniques have been proposed and continue to be proposed to estimate dependencies between brain regions, which can be organised along several axes, in particular measures yielding directed vs non-directed graphs, the domain where the dependency is computed (frequency, time, phase), whether the dependency is linear or non-linear, and whether a zero-lag or a lagged estimate of dependency is used [20].

There has been lots of debate on the choice of adequate measure of dependence in fMRI, but the zero-lag Pearson product-moment linear correlation is a popular choice. If the vertex time series has been decomposed using a wavelet transform, the scale-dependent correlation between two fMRI regional representative time-series $\mathbf{X}$ and $\mathbf{Y}$ in the wavelet domain[2] [3], [21] is given by

$$\rho_{\mathbf{X},\mathbf{Y}}(j) = E[\widehat{\rho}_{\mathbf{X},\mathbf{Y}}(j)] = E\left[\frac{\mathbf{y}_j^T \mathbf{x}_j}{((\mathbf{x}_j^T \mathbf{x}_j)(\mathbf{y}_j^T \mathbf{y}_j))^{1/2}}\right], \tag{3}$$

where $\mathbf{x}_j = (d_{j,1}^{(\mathbf{X})}, \ldots, d_{j,K}^{(\mathbf{X})})$ (likewise for $\mathbf{y}_j$), and $d_{j,k}^{(\mathbf{X})}$ are the wavelet coefficients at scale j and time point k for $\mathbf{X}$ (likewise for $\mathbf{Y}$). The use of wavelets has the advantage of taking into account the long-memory properties of the fMRI time series, and produce correlation estimation that is unbiased at each wavelet scale [19] with a known variance depending on the number of points in the time series at a given scale. This means that using wavelets, as long as the number of vanishing moments is sufficiently large, there is no need to take into account any temporal dependences between the time points in the fMRI time series. However, if centered, filtered time series instead of wavelet coefficient time series are used in $\mathbf{x}_j, \mathbf{y}_j$, Equation 3 corresponds directly to the Pearson product-moment correlation used in the majority of the fMRI literature.

For lagged correlation, we can redefine a circularly shifted version of the second wavelet coefficient time-series as $\mathbf{x}_j = (d_{j,1+\Delta}^{(\mathbf{X})}, \ldots, d_{j,K}^{(\mathbf{X})}, d_{j,1}^{(\mathbf{X})}, \ldots, d_{j,\Delta}^{(\mathbf{X})})$, where Δ is an integer lag, and in the same way define a lagged version of a second filtered time-series if no wavelet decomposition is used. The commonly-applied "functional network connectivity" approach of Jafri and colleagues [14] can be computed in this way, with the time series $\mathbf{X}$ and $\mathbf{Y}$ obtained from an ICA decomposition. Although lagged measures of dependence are often used with electro- or magneto-encephalography [22], they are comparatively less frequent with fMRI data mainly because of the low sampling rate and the hemodynamic response; consequently authors often use a very small or 0 lag.

[2]We provide a Matlab implementation and useful related code at http://miplab.epfl.ch/richiardi/software.php and an R implementation at http://cran.r-project.org/web/packages/brainwaver/.

Partial correlation [23] is a variant which has been shown experimentally to yield good sensitivity in picking up existing correlations and offers robustness to various processing parameters [20]. The goal in partial correlation is to estimate the "direct" correlation between two regions while removing the influence of all other regions. Given a matrix of (filtered) regional time series $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{K \times R}$, one way of computing it is from the (possibly regularised) inverse $\mathbf{P}$ of the empirical covariance matrix of $\boldsymbol{\mathcal{X}}$: with $\mathbf{P} = \boldsymbol{\Sigma}^{-1} = (\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}^T)^{-1}$, we can compute each partial correlation between region i and j as $\pi_{ij} = -\mathbf{P}_{ij}(\mathbf{P}_{ii}\mathbf{P}_{jj})^{-1/2}$. We are not aware of this approach being applied to wavelet coefficient time-series.

Many other dependence measures exist[3] (see e.g. the review by Smith and colleagues [20]). To cite one frequency-domain measure yielding a directed graph and using lags, partial directed coherence [25], a popular method originally proposed on electrophysiological data, computes edge directions in the frequency domain, and has been applied to fMRI [26]. However because of the haemodynamic smoothing and quasi-Gaussian distribution of BOLD signals, measures yielding directed graphs and those based on fine frequency information must be used with caution [20]. In addition, several authors have reported that non-linear measures of dependence such as mutual information may not be necessary for fMRI data [27].

After all dependencies have been computed and edge labels assigned, it is possible to perform hypothesis tests on the edge labels to assert whether they are significantly different from zero. In wavelet correlation, the estimators of correlation are associated with variance [19] and it is then possible to construct hypothesis tests to select only edge labels greater than a given threshold [3]. More generally, the $R \times R$ correlation matrix computed on all pairs of regions can be approximately Gaussianised using Fisher's R-to-Z nonlinear transform, after which different hypothesis tests can be applied – typically a one-sample t-test with a multiple comparison correction such as false discovery rate. Edges labels that survive the hypothesis test are kept, and some authors then fix their label to 1, yielding an unweighted graph. Edges whose labels do not survive are removed from the edge set, or their label is fixed to 0. This procedure can be seen as a filter-type feature selection approach in the subsequent learning procedure. If no threshold is applied, the result is an undirected complete graph with edge labels.

III. Learning and inference on brain graphs

The goal of inference on functional brain graphs is to classify and characterise changes in brain dynamics due to pathology, or due to a cognitive state change (within a single subject or across subjects), possibly related to experimental stimulation. A predictive modelling framework is therefore well suited, because being able to consistently form predictions from brain graphs of unseen subject samples or unseen cognitive state samples provides good evidence for the fact that the model captures patterns with good generalisation ability.

As with any other type of graphs used in machine learning, several algorithms can be used for predictive modelling with brain graphs. However, certain properties of brain graphs as defined here should guide the choice of algorithms. Most importantly, the fixed-cardinality vertex sequence property (see Section II-A) means that no vertex correspondence problem has to be solved. Thus, algorithms designed for more general graphs that do not possess this property may not be suitable, as they may focus on changes in the vertex set. Secondly, brain graphs are noisy, and some graph-theoretical properties such as isomorphism or subgraph isomorphism are not necessarily useful to measure the (dis)similarity between graphs. Lastly, interpretability of results is paramount, and a link with classical statistics is always appreciated in the neuroscience community, where such tools are in common use.

Satisfying these three key requirements, approaches based on graph embedding have started to appear in the neuroimaging literature (including unwittingly). In graph embedding, one defines a mapping that associates each graph of a given graph population to a point in the n-dimensional real space. This relatively recent approach to learning with graphs has made available a very wide variety of statistical machine learning algorithms [28]. The engineering effort is then spent on finding a vector space representation of graphs that is amenable to the learning task at hand. In the sequel, we will present several embeddings that have been or could be used to analyse brain graphs.

A. Graph and vertex properties as features

From physical sciences to social sciences through biological sciences, the representation of data with complex networks has attracted much interest. Although these representations can be used to visually summarise the

[3] several of which are implemented in the Conn software [24] (http://www.nitrc.org/projects/conn/), a full-featured toolbox to compute functional connectivity graphs.

information in two dimensions, it may be difficult to compare different networks with more than around one hundred vertices. Therefore, an important focus of analysis in neuroscience has been large-scale graph organisation measures (not necessarily graph invariants), such as clustering coefficients, or graph efficiencies. These topological measures can be used to extract one or more features that characterise each vertex in a graph, or a graph as a whole. Several topological measures[4] have been considered for neuroimaging data [1], [29], which have particular intepretation in terms of integration and segregation of brain activity.

For example, the *strength* of a vertex i captures the number and weight of connections between i and other nodes of graph g:

$$S_i = \sum_{j=1}^{R} \beta_g(i,j).$$

If edge labels are binary (unweighted graph), vertex strength corresponds to vertex degree. Another example of an often used property is the *clustering* coefficient, which can be regarded as a measure of information transfer or connectedness in the immediate neighbourhood of each vertex [9]:

$$Clust_i = \frac{1}{|V_{g_i}|(|V_{g_i}|-1)} \sum_{j,k \in V_{g_i}} \frac{1}{L_{jk}},$$

where $g_i = (E_{g_i}, V_{g_i}, \alpha_g, \beta_g)$ is a subgraph of g defined by the set of nodes that are the immediate neighbours of the i^{th} node, and L_{jk} is the minimum path length between vertex $j \in V_{g_i}$ and vertex $k \in V_{g_i}$ in the subgraph.

Because vertices in brain graphs following our definition have a fixed ordering, these properties can be arranged into a vector, which can then be used for machine learning. Additionally, computing the average or the median of one particular vertex property over all vertices of a graph results in an abstract measure that characterizes the graph as a whole. Hence, applying this process to a number of properties eventually yields a feature vector that describes a particular graph. Thus, these graph- or vertex-level properties can be seen as the result of sophisticated feature extraction, and can systematically be used directly as input features to statistical machine learning algorithms [30], [31], or in a mass-univariate test setting (see, e.g., [32]), which was the first approach used in order to discriminate between populations.

B. Edge labels and properties as features

Rather than presuming to know which topological property might be of interest to the discrimination task at hand, it may be advantageous to extract a simple representation of graphs, and let the learning algorithm find a function of the representation that yields the best discriminative performance. In this regard, one approach that has brought very competitive results experimentally is to model the edge label distributions, for example in an undirected graph, by using the lexicographically ordered entries of the upper-triangular part of the weighted adjacency matrix $\mathbf{A}$ (or sequence of edge labels) as a feature vector [33]–[36]. In this case the embedding is formed by $\phi(g) = (\ldots, \beta(i,j), \ldots)_{i,j \in \{1,\ldots,R\}, j>i}$.

As a drawback, this procedure leads to high-dimensional feature vectors of order $\mathcal{O}(|V|^2)$, and suffers from the curse of dimensionality. Hence, various feature selection techniques have been tried to address the problem [36], [37]. Mass-univariate analysis of edge labels (e.g., using two-sample t-tests) is very common in the neuroimaging literature, and here the dimensionality problem translates directly into a multiple comparisons problem.

Midway between the direct edge label embedding mentioned above and graph properties, one can also define topological edge properties such as the edge betweenness, related to the number of geodesics (shortest paths) going through an edge, and form a similar high-dimensional embedding from these properties. Edge properties can yield interesting insight into how different "communities" of the network are connected together, although interpretation differs depending on the type of correlation measure used (e.g. partial versus full), as well as imaging modality. In particular, this type of measure might be most interesting for structural connectivity, where an edge can be mapped to a white matter fiber pathway. For example, a significant correlation between the edge betweenness of a white matter fiber tract and grasping skill in stroke patients has been observed [38].

[4]Many toolboxes exist to compute these properties on brain graphs, for example the BCT toolbox (https://sites.google.com/a/brain-connectivity-toolbox.net/bct/) in Matlab, or the iGraph and brainwaver packages using R.

C. Spectral embedding

Another well-known and widely used family of graph embedding algorithms is spectral embedding. The main idea is to perform an eigen-decomposition on the adjacency or the Laplacian matrix of a graph and then use the eigenvectors, possibly after application of some suitable dimensionality reduction algorithms, to derive feature vectors that represent the given graphs in the new vector space [39].

In neuroimaging, the related eigenvector centrality is often used to characterise brain graphs (see, e.g., [40]), albeit only in group-level statistics, although recent work has sought to use singular value decomposition to generate embedding vectors from brain graphs [41].

D. Kernels and dissimilarity techniques

Another kind of graph embedding is dissimilarity embedding. The basic idea is to define a set of prototypical graphs $p_1, \ldots, p_n$ and measure the dissimilarity, or distance, of a given graph g to each of the prototypes. Thus n distances $d(g, p_1), \ldots, d(g, p_n)$ are obtained, which can be concatenated to a vector $\phi(g) = (d(g, p_1), \ldots, d(g, p_n))$ that serves as the representation of g in the embedding space. One crucial question in this approach is the underlying graph dissimilarity function $d(g, g')$. In [28] the authors have proposed to use the graph edit distance, which is a well-established concept in graph-based machine learning. Dissimilarity embedding has recently been applied to brain graphs with an adapted graph edit distance [35]. However, there is no guarantee other than empirical that this is a good choice, and the crucial aspect in dissimilarity algorithms is the design of a pairwise dissimilarity function. This concern is shared by kernel methods.

Kernel methods, originally designed to operate on feature vectors, can be extended so as to include symbolic data structure, in particular graphs [42]. The basic idea of using similarity between pairs of objects can be adapted to graphs in a straightforward way by using, for example, the number of common labels, common subgraphs, common walks, or similar common substructures. This has become a very active area of research, with applications in diverse fields including computer vision, biology, or chemistry, and the relationship between seemingly different graph kernels is increasingly being understood and formalised [43]. While graph embedding methods allow one to get access to the full repository of machine learning methods, graph kernels are restricted to kernel machines. Interestingly, assembling the embedding vectors constructed by a dissimilarity embedding procedure into a square matrix and normalising it can yield a valid kernel matrix, that is, a positive semidefinite matrix (although certain dissimilarity measures may generate an indefinite kernel matrix) [44]. Even in the indefinite case, some SVM solvers are able to converge and these dissimilarity functions can then also be used with SVM-type learning algorithms [45].

There are currently very rare applications of graph kernels to brain graphs [46], although given a suitable kernel, it is probable that this could yield competitive results.

IV. Applications

While the dominant approach for using graphs in neuroimaging consists of group-level statistics on graph and vertex properties, or statistics on edge labels, there has recently been some interest in using graphs for predictive modelling. We will discuss only a few papers here, but it is clear that the techniques presented for machine learning with brain graphs are useful in real applications, and that there is much room for improvement.

A. Clinical neuroscience

Perhaps due to the large amount of evidence showing that brain graphs are affected by disease and that these alterations could form the basis for imaging biomarkers [2], the first application of machine learning for brain graphs has been in clinical neurosciences. Indeed, the predictive nature of machine-learning tools makes them a perfect fit for diagnosis and prognosis of neurological diseases and disorders. It should be noted that data are often acquired in the "resting state", meaning subjects are not asked to perform a specific action. In these circumstances, the mean activity level is typically statistically not different between groups (the absolute magnitude of the BOLD signal is meaningless), and computing a brain graph is a way to provide prior information to learning algorithms that the dependency structure between brain regions is of interest.

In what we believe to be the earliest use of machine learning on brain graphs in clinical neuroscience, an atlas was used to extract 22 brain regions, from which a brain graph was extracted by using a mean regional representative

and linear correlation [33]. They then generated a feature space from each graph using direct embedding, (See Section III-B)), after which they used a PCA-based version of Fisher Linear Discriminant Analysis (FLDA) to predict patient or healthy control status. Significantly worse results are reported when a whole-brain graph is used. Adding several feature selection steps, [34] also used direct embedding to predict depression status. In a similar vein, [37] used univariate filter feature selection, and locally linear embedding (LLE) for dimensionality reduction, before performing classification of schizophrenic patients. Although these papers did not identify their technique as graph embedding, our own recent experiments confirm that the direct approach works well for other diseases with very heterogeneous presentation such as multiple sclerosis [47], with high sensitivity and specificity, by using a brain atlas containing 90 regions. Figure 4 shows a visualisation of the discriminative graph for this task (where we see that connections to and from the occipital lobe, whose regions are shown in yellow, have low discriminative weights, but that those in the temporal lobe, in red, have high discriminative weights), as well as a low-dimensional representation of the discriminant function.

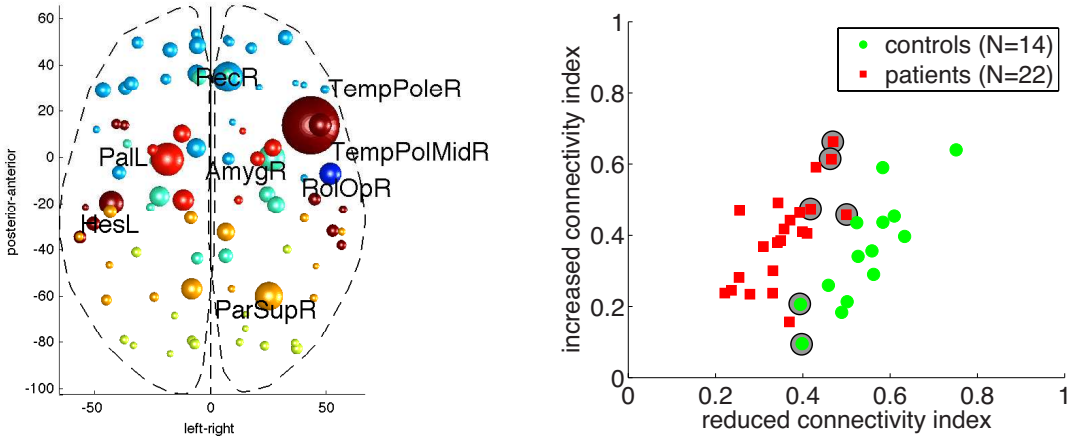

Fig. 4. (left) Discriminative graph for multiple sclerosis patients-versus-controls classification from brain graphs using an ensemble of functional tree classifiers. The feature space is obtained by direct edge label embedding, thus each edge can enter the decision function. We can compute how often edges are picked by the trees in the ensemble, and at which level, across all cross-validation folds, and obtain a measure of the relative importance of each edge in the discrimination task. Then, we can aggregate edge importance on the vertices to which they are connected. Here, the size of the spheres is proportional to the sum of edge discriminative importances, and the colour represents the brain lobe. (right) Post-hoc index of discriminative connectivity alterations. This is obtained by a sum of edge labels (correlation value), each weighted by the (normalised) discriminative importance of the edge, and suggests that the high-dimensional discriminant function is learning a useful combination of edges. Reprinted from [47] Copyright (2012), with permission from Elsevier.

Other approaches have also been used, for example graphs properties (amongst other features) were used to classify schizophrenic patients versus controls via Markov Random Field, SVM, and Naïve Bayes classifiers [30]. Statistical testing was used to identify the set of edge labels that are significantly different between groups [48], from which summary indices were extracted by linear combination of edge label values. These summary indices were then used as input features to an FLDA classifier and allowed high sensitivity and specificity for classifying Alzheimer's disease patients versus controls and Mild Cognitive Impairment (MCI) patients (thought to be an early stage of Alzheimer's) versus controls.

While the patient versus control diagnosis based on brain graphs is interesting as a proof-of-concept and can bring new insights in the disease (i.e., to reveal disease-specific patterns of changes that remained undetected before), it is not yet an accepted tool for clinical practice. Further refinement of the methodology, and greater availability of datasets are required to go towards differential diagnosis and identifying confounds or to predict fine-grained scales of clinical prognosis. Nevertheless, the clinical possibilities opened by such an approach are exciting because it uses a simple, non-invasive test that requires very minimal patient collaboration, using MRI hardware that already exists in most hospitals.

Lastly, it is also important to note that graphs computed by estimating brain connectivity may have very different structure depending on which imaging modality is used (indeed, this also applies to imaging parameters within a particular modality). For example, it was shown that very different graphs are obtained from MEG and fMRI data, even when the graph structure learning algorithm is the same [49]. This indicates that interpretation of brain

graphs and their properties must always consider the limitations of the modality used, but also that multimodal graph analysis methods might bring additional insight [50].

B. Cognitive neuroscience

The application to cognitive neuroscience has focused on how brain regions interact during specific brain states. Predictive modelling on brain connectivity graphs has very recently also enabled brain state decoding, by which "inverse inference" can be performed: The current brain state of the subject is predicted from the connectivity pattern of the brain.

Our paper [36] was among the first to propose decoding brain states from brain graphs; i.e., we showed that rest and movie-watching can be classified with very high accuracy using ensembles of classifiers, both within frequency subbands and across frequency subbands, when direct edge label embedding is used. The same embedding approach (but with different classifiers) was used in [51] with a linear SVM to show that the brain graph is significantly altered by visuomotor task preparation — impressively, the task can be predicted before it even is performed, because the brain "prepares" for the task by altering its connectivity. [52] also used direct embedding and classification to classify sleep stages.

Recent work [53] showed that it is even possible to use brain graphs to discriminate between seemingly similar brain states, namely different conditions such as remembering events of the day, "singing" music internally, and performing mental arithmetic. Here, a data-driven approach (i.e., ICA) was used to define voxel sets corresponding to graph vertices.

Regression techniques have also been proposed, for example support vector regression was used to predict age from resting-state brain scans [54]. Here, meta-analyses were used to define 160 regions of interest, yielding graphs with 12270 different edge labels. They first reduced dimensionality to 200 edges by using univariate filter feature selection (correlation of edge label with age), on a separate dataset. Then, a radial basis function kernel was used with an SVM solver to predict age from these 200 edges. Figure 5 illustrates the results, including a brain-space map of the relatively more predictive edges.

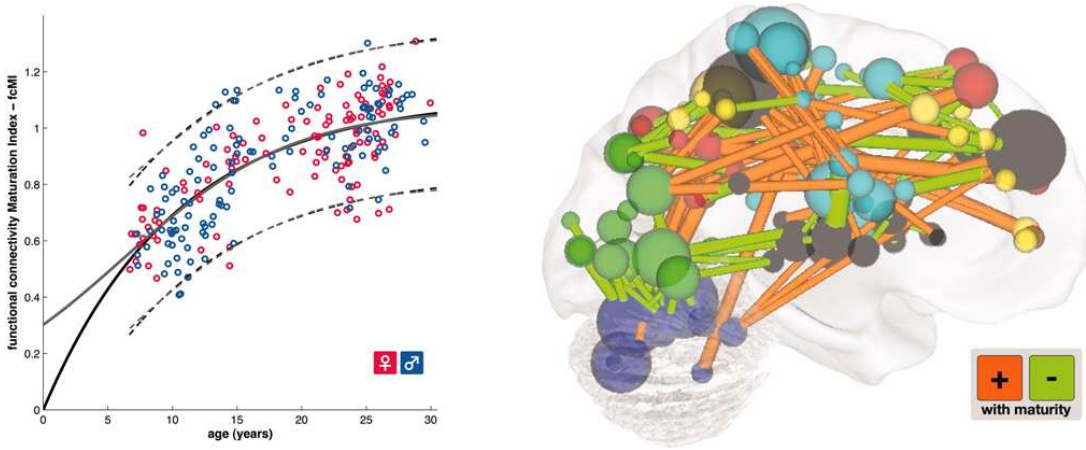

Fig. 5. (left) Support vector regression (SVR) prediction of age from a subset of edge labels from a resting-state brain graph. The y axis shows the normalised prediction value. (right) Importance mapping of edges: the average SVR weight of 156 edges that were selected in all cross-validation folds is shown in a sagittal projection. The width of edges is proportional to their weight in the regression coefficients vector. Edges whose label value increases with age are shown in red, while green indicates the reverse. As in Figure 4, spheres indicate the relative importance of vertices. Sphere colour indicates the functional subnetwork the vertex is part of. Figure from [54]. Reprinted with permission from AAAS.

C. Visualisation

Visualising the results of inference on brain graphs is challenging, but a consensus is slowly emerging for two- or three-dimensional view of the relative discriminative importance of edges and vertices (see Figure 5), where each edge is scaled in proportion to its relative importance in the discriminant function of the classifier or regression algorithm trained on the graph's embedding (e.g., weight vector component for a linear SVM). A challenging

aspect is that the regularisation term added to the learning algorithm may not yield a sparse weight vector, meaning that many graph edges may have non-zero weights. Given the high number of potential edges, this can lead to a confusing display. One attempt to address the issue is to plot each vertex's "strength" in the discriminative graph to represent how much connections to and from this region of the brain contribute to the discrimination. Another approach adopted by some authors, is to use class label permutation testing to see how significant the discriminative weight on each edge is [47]. However, this technique is mass-univariate and does not truly reflect the dependencies in the weight vector.

Specific tools have been written for brain graph visualisation[5], but network visualisation tools from other fields[6] can also be used to provide a variety of two-dimensional layouts (e.g., force-directed layouts).

V. Open issues and future trends

A. Large graphs

As schematised on Figure 3, the current atlas-based approaches typically result into about 100 regions, mapping to 100 vertices, while data-driven approaches typically produce graphs with around 20–30 vertices. Resolution and quality of functional MRI data will further increase with better acquisition sequences and higher magnetic field strengths, and, therefore, brain connectivity graphs with many more vertices will be defined, based on more fine-grained structural or functional regions.

A general problem with graph representations is that number of edges grows like $\mathcal{O}(|V|^2)$. Typical contemporary algorithms of "moderate" complexity applied in machine learning are of cubic time complexity (e.g., inversion of a general matrix). If we apply those algorithms to graphs, then we are facing a complexity that is not $\mathcal{O}(|V|^3)$, but $\mathcal{O}(|V|^6)$ as soon as edge data is to be taken into account. For this reason, many state-of-the-art graph algorithms can deal only with graphs including some hundred up to a few thousand nodes at maximum. In particular, the well-performing direct edge label embedding technique is ill-equipped for dealing with anything other than small-scale graphs (around 100 vertices).

Potential ways out of the dilemma are currently a topic of intensive investigation. Possibly, one could resort to sparse graphs (where the number of edges is not $\mathcal{O}(|V|^2)$, but only $\mathcal{O}(|V|)$) and use approximate algorithms or regularisation techniques. However, using an ℓ_1-type regulariser might promote sparsity in a way that impedes interpretation (e.g., what is the neuroscientific meaning of a discriminant function based on a single brain graph edge?). A recent effort in this direction was to use mixed-norm regularisation to derive sparse models of functional connectivity followed by vertex property computation and subsequent classification to discriminate between Mild Cognitive Impairment patients and healthy controls [55] While experimental results are for a small number of vertices (116), they show an improvement over learning a full graph, which suggests that this could hold for larger graphs as well.

B. Robustness of statistical dependency estimators

The estimation of dependencies is difficult especially when the number of vertices is large in comparison to the number of points in time. Recent work [56] showed exactly the ratio between the number of vertices and the number of points in time so as to ensure the values of correlation are statistically truly different from zero. Moreover, in the context of fMRI acquisition, estimates of dependencies between brain regions underlying vertices are often biased by region size, noise with spatial characteristics, and physiological confounds [57]. Specific robust statistical procedures [58] and denoising procedures [59] are being developed to cope with such challenges. While physiological denoising techniques are gaining acceptance and are routinely included in recent work, issues related to spatial statistics of regions are much less recognised.

C. Beyond the steady-state assumption

To conclude, we should keep in mind that the human brain is a complex system with a high degree of adaptability, which is achieved by dynamical reorganization at different temporal scales. For instance, within a single run of fMRI,

[5]e.g. Connectome Viewer (http://connectomeviewer.org/viewer) or Brain Connectivity Toolbox (http://sites.google.com/site/bctnet/visualization)

[6]e.g. Graphviz (http://www.graphviz.org/), Gephi (http://gephi.org/), or Cytoscape (http://cytoscape.org)

brain activity observed during "rest" shows a high degree of non-stationarity as it involves continuous switching between attention, memory recall, sensory awareness, and so on. This is the neurological reason that functional connectivity of resting state only becomes a stable measure as longer runs are considered (i.e., average behavior over several minutes). On larger timescales, brain network organization gets shaped and reconfigured by learning experiences; e.g., changes in network modularity have been reported at the timescales of minutes and hours [60]. For those reasons, techniques that would properly consider non-stationarity of brain states should lead to more sensitive measures. Recent work in machine learning, for example casting the problem as regularised high-dimensional covariance learning with non-IID data [61], are of particular interest, provided modelling assumptions (i.e. slowly-varying changes) are compatible with the experimental paradigm and existing neurophysiological knowledge.

It is an open challenge to adapt machine learning techniques and dynamical models in particular to properly take into account the non-stationary behavior of the brain.

ACKNOWLEDGMENTS

This work was supported in part by the European Commission (Marie Curie IOF 299500); Swiss National Science Foundation (grant number PP00P2-123438); the Société Académique de Genève (FOREMANE fund); the Swiss Society for Multiple Sclerosis; the Center for Biomedical Imaging (CIBM) of the Geneva and Lausanne Universities, EPFL, and the Leenaards and Louis-Jeantet foundations. SA was partly funded by the ANR project (grant number ANR JCJC 0302 01). The brain illustration of figure 1 is from the Open Clipart Library (http://openclipart.org).

REFERENCES

[1] E. Bullmore and O. Sporns, "Complex brain networks: graph theoretical analysis of structural and functional systems," *Nature Reviews Neuroscience*, vol. 10, no. 3, pp. 186–198, 2009.

[2] M. D. Fox and M. Greicius, "Clinical applications of resting state functional connectivity," *Front Syst Neurosci*, vol. 4, p. 19, 2010. [Online]. Available: http://dx.doi.org/10.3389/fnsys.2010.00019

[3] S. Achard, R. Salvador, B. Whitcher, J. Suckling, and E. Bullmore, "A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs," *The Journal of Neuroscience*, vol. 26, no. 1, pp. 63–72, Jan. 2006.

[4] O. Sporns, *Networks of the brain*. MIT Press, 2011.

[5] F. Pereira, T. Mitchell, and M. Botvinick, "Machine learning classifiers and fMRI: A tutorial overview," *NeuroImage*, vol. 45, no. 1, Supplement 1, pp. S199–S209, Mar. 2009. [Online]. Available: http://www.sciencedirect.com/science/article/B6WNP-4TYYSKX-5/2/1cc423c98274625360b74dd18930fc61

[6] S. C. Strother, "Evaluating fMRI preprocessing pipelines," *IEEE Engineering in Medicine and Biology Magazine*, vol. 25, no. 2, pp. 27–41, 2006.

[7] S. Strother, S. La Conte, L. Kai Hansen, J. Anderson, J. Zhang, S. Pulapura, and D. Rottenberg, "Optimizing the fMRI data-processing pipeline using prediction and reproducibility performance metrics: I. a preliminary group analysis," *NeuroImage*, vol. 23, Supplement 1, no. 0, pp. S196–S207, 2004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1053811904003945

[8] P. Dickinson, H. Bunke, A. Dadej, and M. Kraetzl, "Matching graphs with unique node labels," *Pattern Analysis & Applications*, vol. 7, no. 3, pp. 243–254, Sep. 2004. [Online]. Available: http://dx.doi.org/10.1007/BF02683991

[9] M. E. J. Newman, *Networks: An Introduction*. Oxford University Press, 2010.

[10] B. Biswal, F. Z. Yetkin, V. M. Haughton, and J. S. Hyde, "Functional connectivity in the motor cortex of resting human brain using echo-planar MRI," *Magnetic Resonance in Medicine*, vol. 34, no. 4, pp. 537–541, 1995.

[11] R. Salvador, J. Suckling, M. Coleman, J. Pickard, D. Menon, and E. Bullmore, "Neurophysiological architecture of functional magnetic resonance images of human brain," *Cerebral Cortex*, vol. 15, no. 9, pp. 1332–1342, 2005.

[12] V. D. Calhoun, T. Adali, G. D. Pearlson, and J. J. Pekar, "A method for making group inferences from functional MRI data using independent component analysis," *Human Brain Mapping*, vol. 14, no. 3, pp. 140–151, Nov 2001.

[13] D. Cordes, V. Haughton, J. D. Carew, K. Arfanakis, and K. Maravilla, "Hierarchical clustering to measure connectivity in fMRI resting-state data," *Magnetic Resonance Imaging*, vol. 20, no. 4, pp. 305–317, May 2002.

[14] M. J. Jafri, G. D. Pearlson, M. Stevens, and V. D. Calhoun, "A method for functional network connectivity among spatially independent resting-state components in schizophrenia," *NeuroImage*, vol. 39, no. 4, pp. 1666–1681, Feb 2008. [Online]. Available: http://dx.doi.org/10.1016/j.neuroimage.2007.11.001

[15] F. Deleus and M. M. Van Hulle, "Functional connectivity analysis of fMRI data based on regularized multiset canonical correlation analysis." *J. of Neuroscience Methods*, vol. 197, no. 1, pp. 143–157, Apr 2011. [Online]. Available: http://dx.doi.org/10.1016/j.jneumeth.2010.11.029

[16] G. Buzsáki, *Rhythms of the brain*. Oxford University Press, 2006.

[17] V. Maxim, L. Şendur, M. J. Fadili, J. Suckling, R. Gould, R. Howard, and E. T. Bullmore, "Fractional Gaussian noise, functional MRI and Alzheimer's disease," *NeuroImage*, vol. 25, no. 1, pp. 141–158, March 2005.

[18] E. Bullmore, J. Fadili, V. Maxim, L. Şendur, B. Whitcher, J. Suckling, M. Brammer, and M. Breakspear, "Wavelets and functional magnetic resonance imaging of the human brain," *NeuroImage*, vol. 23, Supplement 1, no. 0, pp. S234–S249, 2004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1053811904003775

[19] B. Whitcher, P. Guttorp, and D. B. Percival, "Wavelet analysis of covariance with application to atmospheric time series," *Journal of Geophysical Research*, vol. 105, no. D11, pp. 14 941–14 962, 2000.

[20] S. Smith, K. Miller, G. Salimi-Khorshidi, M. Webster, C. Beckmann, T. Nichols, J. Ramsey, and M. Woolrich, "Network modelling methods for fMRI," *NeuroImage*, vol. 54, no. 2, pp. 875–891, january 2011.

[21] R. Salvador, J. Suckling, M. R. Coleman, J. D. Pickard, D. Menon, and E. Bullmore, "Neurophysiological architecture of functional magnetic resonance images of human brain." *Cerebral Cortex*, vol. 15, no. 9, pp. 1332–1342, Sep 2005. [Online]. Available: http://dx.doi.org/10.1093/cercor/bhi016

[22] J.-P. Lachaux, M. Chavez, and A. Lutz, "A simple measure of correlation across time, frequency and space between continuous brain signals," *J Neurosci Methods*, vol. 123, no. 2, pp. 175–188, Mar 2003.

[23] G. Marrelec, A. Krainik, H. Duffau, M. Pélégrini-Issac, S. Lehéricy, J. Doyon, and H. Benali, "Partial correlation for functional brain interactivity investigation in functional MRI," *NeuroImage*, vol. 32, no. 1, pp. 228–237, Aug 2006. [Online]. Available: http://dx.doi.org/10.1016/j.neuroimage.2005.12.057

[24] S. Whitfield-Gabrieli and A. Nieto-Castanon, "Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks." *Brain Connectivity*, vol. 2, no. 3, pp. 125–141, 2012. [Online]. Available: http://dx.doi.org/10.1089/brain.2012.0073

[25] L. A. Baccalá and K. Sameshima, "Partial directed coherence: a new concept in neural structure determination." *Biol Cybern*, vol. 84, no. 6, pp. 463–474, Jun 2001.

[26] J. R. Sato, D. Y. Takahashi, S. M. Arcuri, K. Sameshima, P. A. Morettin, and L. A. Baccalá, "Frequency domain connectivity identification: an application of partial directed coherence in fMRI." *Human Brain Mapping*, vol. 30, no. 2, pp. 452–461, Feb 2009. [Online]. Available: http://dx.doi.org/10.1002/hbm.20513

[27] J. Hlinka, M. Palus, M. Vejmelka, D. Mantini, and M. Corbetta, "Functional connectivity in resting-state fMRI: is linear correlation sufficient?" *NeuroImage*, vol. 54, no. 3, pp. 2218–2225, Feb 2011. [Online]. Available: http://dx.doi.org/10.1016/j.neuroimage.2010.08.042

[28] K. Riesen and H. Bunke, *Graph Classification and Clustering Based on Vector Space Embedding*. World Scientific, 2010.

[29] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: Uses and interpretations," *NeuroImage*, vol. 52, no. 3, pp. 1059–1069, September 2010.

[30] G. A. Cecchi, I. Rish, B. Thyreau, B. Thirion, M. Plaze, M.-L. Paillere-Martinot, C. Martelli, J.-L. Martinot, and J.-B. Poline, "Discriminative network models of schizophrenia," in *Proc. Neural Information Processing Systems (NIPS)*, Vancouver, Canada, 2009, pp. 252–260.

[31] D. S. Bassett, B. G. Nelson, B. A. Mueller, J. Camchong, and K. O. Lim, "Altered resting state complexity in schizophrenia," *NeuroImage*, vol. 59, no. 3, pp. 2196–2207, Feb 2012. [Online]. Available: http://dx.doi.org/10.1016/j.neuroimage.2011.10.002

[32] S. Achard and E. Bullmore, "Efficiency and cost of economical human brain functional networks," *PLoS Computational Biology*, vol. 3, no. 2, pp. 0174–0183, February 2007.

[33] K. Wang, T. Jiang, M. Liang, L. Wang, L. Tian, X. Zhang, K. Li, and Z. Liu, "Discriminative analysis of early Alzheimer's disease based on two intrinsically anti-correlated networks with resting-state fMRI," in *Proc. Int. Conf. on Medical Image Computing and Computer Assisted Intervention*, 2006, pp. 340–7.

[34] R. C. Craddock, P. E. Holtzheimer, X. P. Hu, and H. S. Mayberg, "Disease state prediction from resting state functional connectivity," *Magn Reson Med*, vol. 62, no. 6, pp. 1619–1628, Dec 2009. [Online]. Available: http://dx.doi.org/10.1002/mrm.22159

[35] J. Richiardi, D. Van De Ville, K. Riesen, and H. Bunke, "Vector space embedding of undirected graphs with fixed-cardinality vertex sequences for classification," in *Proc. 20th Int. Conf. on Pattern Recognition (ICPR)*, Istanbul, Turkey, 2010, pp. 902–905.

[36] J. Richiardi, H. Eryilmaz, S. Schwartz, P. Vuilleumier, and D. Van De Ville, "Decoding brain states from fMRI connectivity graphs," *NeuroImage (Special Issue on Multivariate Decoding and Brain Reading)*, vol. 56, no. 2, pp. 616–626, May 2011.

[37] H. Shen, L. Wang, Y. Liu, and D. Hu, "Discriminative analysis of resting-state functional connectivity patterns of schizophrenia using low dimensional embedding of fMRI," *NeuroImage*, vol. 49, no. 4, pp. 3110–3121, Feb 2010. [Online]. Available: http://dx.doi.org/10.1016/j.neuroimage.2009.11.011

[38] E. R. Buch, A. Modir Shanechi, A. D. Fourkas, C. Weber, N. Birbaumer, and L. G. Cohen, "Parietofrontal integrity determines neural modulation associated with grasping imagery after stroke," *Brain*, vol. 135, no. Pt 2, pp. 596–614, Feb 2012. [Online]. Available: http://dx.doi.org/10.1093/brain/awr331

[39] B. Luo, R. C. Wilson, and E. R. Hancock, "Spectral embedding of graphs," *Pattern Recognition*, vol. 36, no. 10, pp. 2213–2230, 2003.

[40] G. Lohmann, D. S. Margulies, A. Horstmann, B. Pleger, J. Lepsien, D. Goldhahn, H. Schloegl, M. Stumvoll, A. Villringer, and R. Turner, "Eigenvector centrality mapping for analyzing connectivity patterns in fMRI data of the human brain," *PLoS One*, vol. 5, no. 4, p. e10232, 2010. [Online]. Available: http://dx.doi.org/10.1371/journal.pone.0010232

[41] J. Richiardi, H. Eryilmaz, and D. Van De Ville, "Low-dimensional embedding of functional connectivity graphs for brain state decoding," in *Proc. 1st Workshop on Brain Decoding: pattern recognition challenges in neuroimaging*, 2010, pp. 21–24.

[42] T. Gaertner, *Kernels for structured data*. World Scientific, 2008.

[43] S. Vishwanathan, N. N. Schraudolph, R. Kondor, and K. M. Borgwardt, "Graph kernels," *Journal of Machine Learning Research*, vol. 11, pp. 1201–1242, April 2010.

[44] E. Pekalksa, P. Paclik, and R. P. W. Duin, "A generalized kernel approach to dissimilarity-based classification," *Journal of Machine Learning Research*, vol. 2, pp. 175–211, December 2001.

[45] B. Haasdonk, "Feature space interpretation of SVMs with indefinite kernels," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, no. 4, pp. 482–492, 2005.

[46] F. Mokhtari and G.-A. Hossein-Zadeh, "Decoding brain states using backward edge elimination and graph kernels in fMRI connectivity networks," *Journal of Neuroscience Methods*, vol. 212, no. 2, pp. 259–268, 2013.

[47] J. Richiardi, M. Gschwind, S. Simioni, J.-M. Annoni, B. Greco, P. Hagmann, M. Schluep, P. Vuilleumier, and D. Van De Ville, "Classifying minimally-disabled multiple sclerosis patients from resting-state functional connectivity," *NeuroImage*, vol. 62, pp. 2021–2033, 2012. [Online]. Available: http://dx.doi.org/10.1016/j.neuroimage.2012.05.078

[48] G. Chen, B. D. Ward, C. Xie, W. Li, Z. Wu, J. L. Jones, M. Franczak, P. Antuono, and S.-J. Li, "Classification of Alzheimer disease, mild cognitive impairment, and normal cognitive status with large-scale network analysis based on resting-state functional MR imaging," *Radiology*, vol. 259, no. 1, pp. 213–221, Apr 2011. [Online]. Available: http://dx.doi.org/10.1148/radiol.10100734

[49] S. M. Plis, M. P. Weisend, E. Damaraju, T. Eichele, A. Mayer, V. P. Clark, T. Lane, and V. D. Calhoun, "Effective connectivity analysis of fMRI and MEG data collected under identical paradigms." *Computers in Biology and Medicine*, vol. 41, no. 12, pp. 1156–1165, Dec 2011. [Online]. Available: http://dx.doi.org/10.1016/j.compbiomed.2011.04.011

[50] X. Lei, D. Ostwald, J. Hu, C. Qiu, C. Porcaro, A. P. Bagshaw, and D. Yao, "Multimodal functional network connectivity: an EEG-fMRI fusion in network space." *PLoS One*, vol. 6, no. 9, p. e24642, 2011. [Online]. Available: http://dx.doi.org/10.1371/journal.pone.0024642

[51] J. Heinzle, M. A. Wenzel, and J.-D. Haynes, "Visuomotor functional network topology predicts upcoming tasks," *Journal of Neuroscience*, vol. 32, no. 29, pp. 9960–9968, Jul 2012. [Online]. Available: http://dx.doi.org/10.1523/JNEUROSCI.1604-12.2012

[52] E. Tagliazucchi, F. von Wegner, A. Morzelewski, S. Borisov, K. Jahnke, and H. Laufs, "Automatic sleep staging using fMRI functional connectivity data," *NeuroImage*, vol. 63, no. 1, pp. 63–72, Jun 2012. [Online]. Available: http://dx.doi.org/10.1016/j.neuroimage.2012.06.036

[53] W. R. Shirer, S. Ryali, E. Rykhlevskaia, V. Menon, and M. D. Greicius, "Decoding subject-driven cognitive states with whole-brain connectivity patterns," *Cerebral Cortex*, vol. 22, no. 1, pp. 158–165, May 2012. [Online]. Available: http://dx.doi.org/10.1093/cercor/bhr099

[54] N. U. F. Dosenbach, B. Nardos, A. L. Cohen, D. A. Fair, J. D. Power, J. A. Church, S. M. Nelson, G. S. Wig, A. C. Vogel, C. N. Lessov-Schlaggar, K. A. Barnes, J. W. Dubis, E. Feczko, R. S. Coalson, J. R. Pruett, D. M. Barch, S. E. Petersen, and B. L. Schlaggar, "Prediction of individual brain maturity using fMRI," *Science*, vol. 329, no. 5997, pp. 1358–1361, Sep. 2010.

[55] C.-Y. Wee, P.-T. Yap, D. Zhang, L. Wang, and D. Shen, "Constrained sparse functional connectivity networks for MCI classification," in *Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, ser. Lecture Notes in Computer Science, N. Ayache, H. Delingette, P. Golland, and K. Mori, Eds. Springer Berlin / Heidelberg, 2012, vol. 7511, pp. 212–219. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-33418-4_27

[56] A. Hero and B. Rajaratnam, "Large scale correlation screening," *Journal of the American Statistical Association (JASA)*, vol. 106, no. 496, pp. 1540–1552, 2011.

[57] A. Zalesky, A. Fornito, I. H. Harding, L. Cocchi, M. Yücel, C. Pantelis, and E. T. Bullmore, "Whole-brain anatomical networks: does the choice of nodes matter?" *NeuroImage*, vol. 50, pp. 970–983, 2010.

[58] S. Achard, J. Coeurjolly, R. Marcillaud, and J. Richiardi, "fMRI functional connectivity estimators robust to region size bias," in *Proc. IEEE Workshop on Statistical Signal Processing (SSP)*, Nice, France, June 2011, pp. 813–816.

[59] H. J. Jo, Z. S. Saad, W. K. Simmons, L. A. Milbury, and R. W. Cox, "Mapping sources of correlation in resting state fmri, with artifact detection and removal." *NeuroImage*, vol. 52, no. 2, pp. 571–582, Aug 2010. [Online]. Available: http://dx.doi.org/10.1016/j.neuroimage.2010.04.246

[60] D. S. Bassett, N. F. Wymbs, M. A. Porter, P. J. Mucha, J. M. Carlson, and S. T. Grafton, "Dynamic reconfiguration of human brain networks during learning," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 18, pp. 7641—-7646, April 2011.

[61] S. Zhou, J. Lafferty, and L. Wasserman, "Time varying undirected graphs," *Machine Learning*, vol. 80, pp. 295—319, 2010.