

Spatially-Inferred Graphical Models for fMRI Data Mining

Pedro Abranches de Carvalho, Hongyu Luo and Roser Viñals Terrés
École polytechnique fédérale de Lausanne, Switzerland

Abstract—Functional Magnetic Resonance Imaging (fMRI) offers the possibility to record neural activity in the human brain. This activity occurs in the shape of spatially organized patterns; networks of brain regions interact dynamically. In particular, some regions cause others to activate or de-activate. In this project, we consider the problem of inferring this causality using graphical Granger modeling. In other words, we want to unravel whether an fMRI signal from a given region can be predicted using signals from other regions. This boils down to the problem of learning causality matrices from a collection of spatio-temporal time-series. We tackle this task by minimizing a functional incorporating a data-fidelity term that fits the signals to a vector-autoregressive (VAR) model and an ℓ_1 -regularization term. Such regularity prior forces the causality matrix to have sparse entries. This is in line with the fact that a given region in the brain is expected to have causal influences from very few other regions. We tested the algorithm on simplistic synthetic data to assess the performance, before applying to real Resting-State fMRI data from healthy subjects. The estimated causality matrix was then used to study causal effects between known large-scale activation networks. The results uncover insightful network-to-network causal couplings.

I. INTRODUCTION

Human brain is a complex system. Intrinsic and spontaneous brain activation can be recorded using functional magnetic resonance imaging (fMRI) through the fluctuations of blood-oxygen-level dependent (BOLD) signals.

Modeling the resting-state fMRI is crucial to understand the regional interactions of temporal brain activation. In this project, we use network modeling and dynamical analysis to infer the temporal-spatial causality from time series activation on different brain regions based on resting state fMRI.

With the purpose of inferring causality between brain regions, we use graphical Granger causality (GC) modeling. GC is based on the fact that when, considering the past temporal samples of a signal improves the predictability of the future time samples of another signal, the first is causing the second. Mathematically, this approach can be formulated using vector autoregressive (VAR) modeling, that accounts multiple linear dependencies between multiple time series.

Granger Causality has been widely used for fMRI [1]–[5]. However, in this project we introduce a novel application of GC for the estimation of brain causality. Our work differs with the existing literature mainly in two aspects. First, we use deconvolved signals which reflect the activity moments, undone from the effect of haemodynamics [6]. Second, we study causality between innovation-driven co-activation patterns (iCAPs), patterns that reveal spatially and temporally

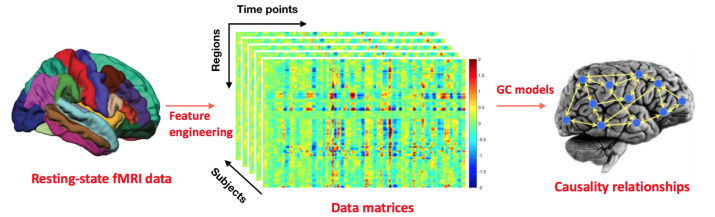


Fig. 1: Project pipeline: from brain regions to causality networks.

overlapping large-scale networks [7]. Figure 1 shows the main pipeline of this project.

The rest of the paper is organized as follows. Section II describes the spatio-temporal modeled and formulates the problem. Section III describes the fMRI data used and the synthetic data generated for verifying our approach. Section IV presents the results and Section V concludes the paper.

II. SPATIO-TEMPORAL CAUSAL MODELING

Modeling the spatio-temporal causality has been widely done using Granger causality. In this section we model the causality between brain regions based on the notion of GC.

Let's consider N nodes, which correspond to brain regions, and T time samples which correspond to the time measurements of fMRI. For any time $t = 0, \dots, T-1$ and node $n = 0, \dots, N-1$, we can denote the fMRI signal at time t and node n as $x_{n,t}$. We can further express the fMRI signal in a matrix \mathbf{X} with $x_{n,t}$ in row n and column t .

In order to predict the causality of the time-series we define a lag, L , which indicates the number of previous samples used for the estimation. Thus, given a node i , we would like to estimate $x_{i,t}$ for $t = L, \dots, T-1$ as

$$\sum_{t=L}^{T-1} x_{i,t} \approx \sum_{l=1}^L \beta_l x_{i,t-l}, \quad (1)$$

being β_l the coefficients that weight the contribution of the previous time samples. Note that equation (1) is only considering the time causality for a given location while we are interested on the time-spatial causality. In order to take into account the impact of other nodes, equation (1) can be reformulated as

$$\sum_{t=L}^{T-1} x_{i,t} \approx \sum_{n=0}^{N-1} \sum_{l=1}^L \beta_{i,n,l} x_{n,t-l}, \quad (2)$$

where the coefficients not only weight the previous samples impact, but also the nodes. These coefficients are a direct

measure of the time-spatial causality and its estimation can be performed by a regression problem.

For each node i , the coefficients that measure the causality from other nodes to node i can be obtained through solving the following problem:

$$\hat{\beta}_i = \underset{\beta_i}{\operatorname{argmin}} \sum_{t=L}^{T-1} (x_{i,t} - \sum_{n=0}^{N-1} \sum_{l=1}^L \beta_{i,n,l} x_{n,t-l})^2 + \lambda \sum_{n=0}^{N-1} \sum_{l=1}^L |\beta_{i,n,l}|, \quad (3)$$

with λ being the regularization term. It is important mentioning that we are using L1-regularization since we are interested in sparseness, since, for each node, very few others are causal. Note that other techniques such as least-squares or ridge regression would not allow us to control the sparsity of our solution. Equation (3) can be reformulated taking into account more than one data matrices, \mathbf{X} as:

$$\hat{\beta}_i = \underset{\beta_i}{\operatorname{argmin}} \sum_{d=0}^{D-1} \sum_{t=L}^{T-1} (x_{i,t}^d - \sum_{n=0}^{N-1} \sum_{l=1}^L \beta_{i,n,l} x_{n,t-l}^d)^2 + \lambda \sum_{n=0}^{N-1} \sum_{l=1}^L |\beta_{i,n,l}|, \quad (4)$$

where D is the total amount of data matrices and $x_{n,t}^d$ the signal coming from the data matrix d . Problem (3) can be further expressed as

$$\hat{\beta}_i = \underset{\beta_i}{\operatorname{argmin}} \|\mathbf{y} - \tilde{\mathbf{X}}\tilde{\beta}_i\|_2^2 + \lambda \|\tilde{\beta}_i\|_1, \quad (5)$$

with

$$\mathbf{y} \triangleq [x_{i,L}^0, x_{i,L+1}^0, \dots, x_{i,T-1}^0, x_{i,L}^1, \dots, x_{i,T-1}^1, \dots, x_{i,T-1}^{D-1}]^T, \\ \mathbf{y} \in \mathbb{R}^{(T-L)D \times 1}, \text{ and } \tilde{\mathbf{X}} \triangleq [\tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_1, \dots, \tilde{\mathbf{X}}_{N-1}], \\ \tilde{\mathbf{X}} \in \mathbb{R}^{(T-L)D \times LN}, \text{ where } \tilde{\mathbf{X}}_j \text{ is a matrix defined as}$$

$$\tilde{\mathbf{X}}_j = \begin{bmatrix} x_{j,L-1}^0 & x_{j,L-2}^0 & x_{j,L-3}^0 & \dots & x_{j,L-L}^0 \\ x_{j,L+1-1}^0 & x_{j,L+1-2}^0 & x_{j,L+1-3}^0 & \dots & x_{j,L+1-L}^0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{j,T-1-1}^0 & x_{j,T-1-2}^0 & x_{j,T-1-3}^0 & \dots & x_{j,T-1-L}^0 \\ x_{j,L+1-1}^1 & x_{j,L+1-2}^1 & x_{j,L+1-3}^1 & \dots & x_{j,L+1-L}^1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{j,T-1-1}^{D-1} & x_{j,T-1-2}^{D-1} & x_{j,T-1-3}^{D-1} & \dots & x_{j,T-1-L}^{D-1} \end{bmatrix}, \quad (6)$$

and $\tilde{\beta}_i \triangleq [\beta_{i0,1}, \dots, \beta_{i0,L}, \beta_{i1,1}, \dots, \beta_{i1,L}, \dots, \beta_{iN-1,L}]^T \in \mathbb{R}^{LN \times 1}$.

Note that the vector $\tilde{\beta}_i$ can be reshaped into a matrix $\mathbf{B}_i \in \mathbb{R}^{L \times N}$, that contains the causality coefficients from other nodes to node i for $l = 1, \dots, L$. It is important mentioning that we are not interested in the coefficients $\forall l$, instead we want to obtain the relation between nodes. Thus, the causality weight from node j to i we can be computed for instance by $a_{j,i} = \frac{1}{L} \sum_{l=1}^L \beta_{i,j,l}$ or $a_{j,i} = \max_l \beta_{i,j,l}$, giving rise to the vector $\mathbf{a}_i \triangleq [a_{0,i}, \dots, a_{j,i}, \dots, a_{N-1,i}]^T$. From performing the estimation of the coefficients for all the nodes, we can easily obtain an adjacency matrix, $\mathbf{A} = [\mathbf{a}_0, \mathbf{a}_1, \dots, \mathbf{a}_{N-1}] \in \mathbb{R}^{N \times N}$, which is our ultimate goal.

Moreover, we are neither interested in the self-causality. Thus, problem (6) for node i , can be further reduced by compacting the matrices to $\tilde{\mathbf{X}} \triangleq [\tilde{\mathbf{X}}_0, \dots, \tilde{\mathbf{X}}_{i-1}, \tilde{\mathbf{X}}_{i+1}, \dots, \tilde{\mathbf{X}}_{N-1}] \in \mathbb{R}^{(T-L)D \times L(N-1)}$ and

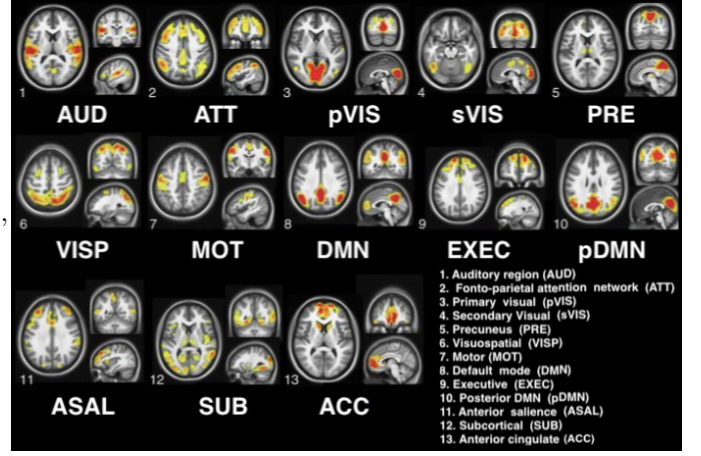


Fig. 2: iCAPs considered. Adapted from [7].

$\tilde{\beta}_i \triangleq [\beta_{i0,1}, \dots, \beta_{i0,L}, \dots, \beta_{iN-1,L}]^T \in \mathbb{R}^{L(N-1) \times 1}$, reducing the overall computation cost and giving rise to a adjacency matrix whose diagonal elements are 0.

III. DATA

A. Brain data

The data used for this project was obtained from 12 different subjects. It contains the fMRI signal taken in 12462 voxels and 440 timepoints. Note that the signals are deconvolved. Since our goal is to obtain the causality between brain regions, the data was parcelled to 90 regional timecourses using the Anatomical Automatic Labeling (AAL) ATLAS [8]. Additionally, the lab also provided us 13 innovation-driven co-activation patterns (iCAPs) in the brain from resting-state fMRI. fMRI signals can be decomposed in iCAPs, which reveal spatially and temporally overlapping large-scale networks [7]. The iCAPs considered are the same than the ones used in [9], that correspond to Figure 2.

It is important mentioning that the interpretation of the results when using these databases was complex. In order to proof and optimize our results, we generated synthetic data as it is going to be explained in III-B.

B. VAR model

In order to make sure our model is working correctly and to allow the possibility of systematic tests we have created our own synthetic data using a VAR model. In each iteration the next row of the matrix is computed using the L rows before that account for the lag. The first L rows are just white Gaussian noise. Let $\mathbf{x}_{s,t}$ denote the row of the matrix of patient S at time t . A diagonal of zeros was also imposed to prevent same node causality. We use the following generative model:

$$\mathbf{x}_{s,t} = \sum_{l=1}^L \mathbf{A}_{s,l} \mathbf{x}_{s,t-l} + \boldsymbol{\eta} \quad (7)$$

The matrices $\mathbf{A}_{s,l}$ are generated using an adjacency matrix, concerning the relationships between all the nodes, where $A[i,j] = 1$ if there is causality between those nodes in a way that x_i causes x_j , and $A[i,j] = 0$ if the nodes are independent.

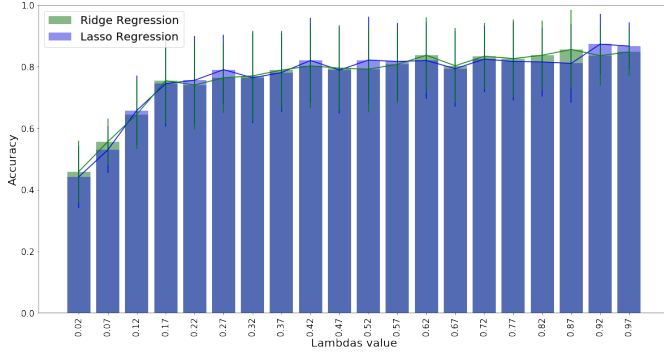


Fig. 3: Test different lambdas with 4 nodes and a lag of 4.

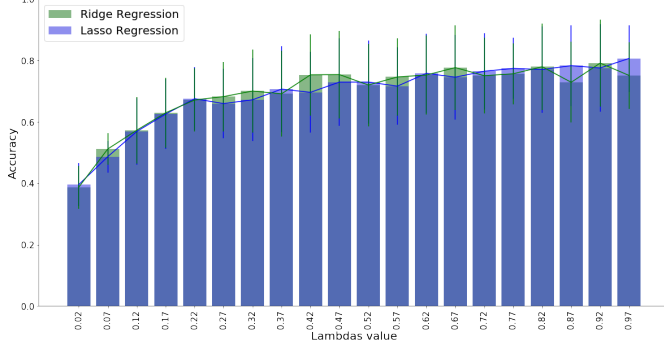


Fig. 4: Test different lambdas with 6 nodes and a lag of 4.

The values for the adjacency matrix are chosen randomly regarding a binomial distribution, where the probability of equaling 0 is 0.8.

After obtaining the adjacency matrix, we then multiply it by a \mathbf{C} matrix, that is created by sampling from a uniform distribution between -0.8 and 0.8. This is done in order to give a random coefficient weight to the nodes that do participate in the causality.

C. Training the model

Now that we have both the model and the VAR synthetic data we can begin to test different parameters in order to see which one suits us the best. When using synthetic data the number of nodes is not as high as when using real human data due to the fact that the artificial data being more noisy and random. Therefore, the test that showed were performed using 4 and 6 nodes with a lag of 4. The values demonstrate the mean and standard deviation of 40 test performed for each lambda.

The way the model chooses which nodes have causality is by computing the max value of the different coefficients matrices. After obtaining this final matrix, we change every value that is positively different from 0 to 1, indicating the causality between nodes. The accuracy was calculated, comparing this matrix to the original created adjacency matrix.

All of the lambdas and lags tested are not here depicted as we performed a grid search on these parameters. Nevertheless, these images show that we do get good results using lasso and ridge regression, obtaining accuracies of 0.87 and 0.85 for lasso regression using 4 and 6 nodes respectively. Ridge

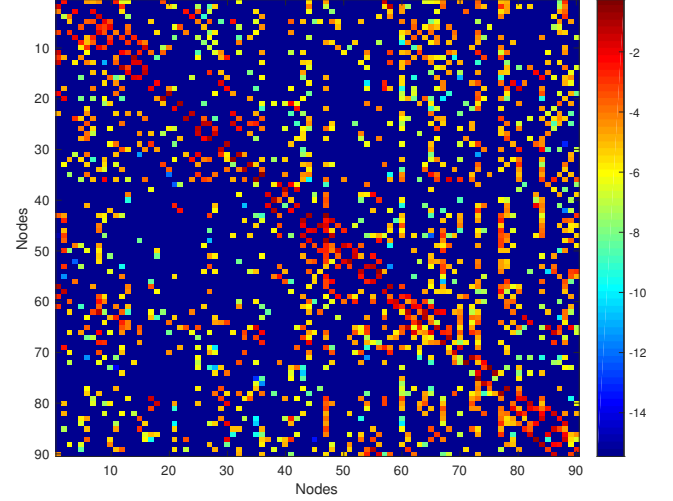


Fig. 5: Coefficients matrix with $L = 1$.

regression does not behave much worse, getting 0.84 and 0.79 using 4 and 6 nodes respectively. You can see that only changing 2 nodes already makes a difference in the accuracy, however our main goal of making sure our model works was achieved with satisfying results.

The choice between lasso and ridge regression and the different lags does not focus solemnly in this results but also from a biological perspective. Regarding the method it makes more sense to use lasso as it imposes sparsity and it is a prior that makes sense in our case as we know that few nodes are causal. After trying different lags for the real data, the lag of 1 was chosen due to the nature of the data that we are working with, which was sampled every 1 second, being slow compared to neuronal activity and we do not desire to go back in time greatly.

IV. RESULTS

The coefficients matrix obtained for $L = 1$ is shown in Figure 5. Combining the coefficients results and by taking into account the iCAPs, we computed for each node in a certain iCAP the number of nodes causing it from another iCAP. As a result, a matrix that relates the 13 different iCAPs by quantifying the causality between them was obtained as shown in Figure 6.

As explained in Section III-A, we are not only interested in causality between nodes, but also between iCAPs. Figure 7a shows the nodes that correspond to the primary visual iCAP and the nodes that are known to cause it. Figure 7b represents the estimated causality obtained by combining the obtained coefficients matrix shown in Figure 5 with the iCAPs nodes in Figure 7a.

For visualization purposes, we have applied a threshold to the matrix in Figure 6 and represented the causality relations in Figure 8. Note that we are obtaining similar results than in [9], proving the high performance of our approach.

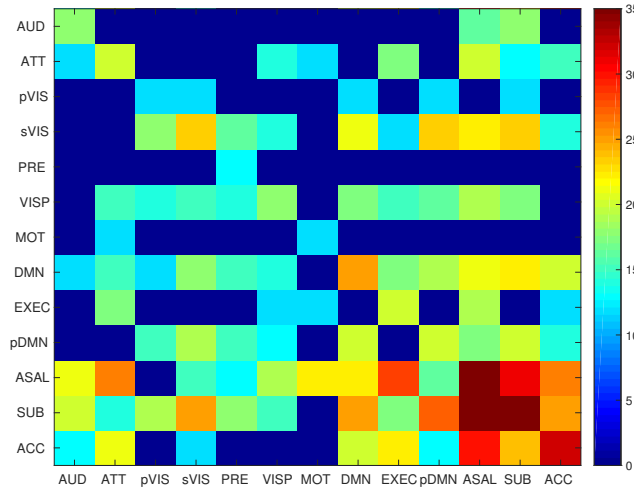


Fig. 6: Estimated causality matrix for the iCAPs in Figure 2.

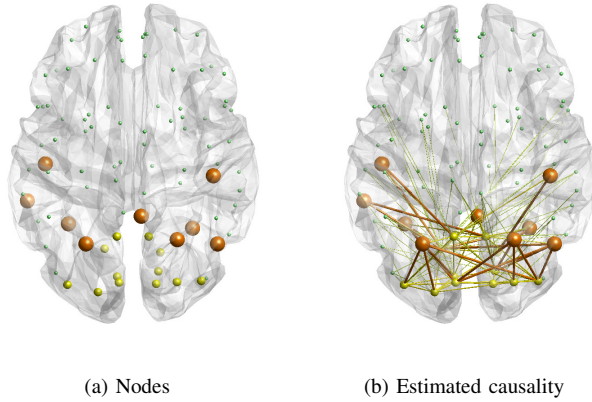


Fig. 7: a) iCAPs nodes for primary visual in yellow and nodes that are known to cause it in orange. b) Estimated causality with our approach using lag 1.

V. SUMMARY

In this paper, we have tackled the estimation of causality between brain regions from resting-state fMRI by applying the concept of Granger causality and formulating a regression problem. First, synthetic data generated with VAR modeling was used to test and optimize our approach. Then, the brain connectivity between 90 regions of the brain was estimated by using deconvolved fMRI signals. Apart from obtaining the causality between these regions, the results were extended to causality between innovation-driven co-activation patterns, being an innovative approach to studying causality in the brain. As further improvement, we have started including neighborhood information to account also for the spatial dimension.

REFERENCES

- [1] G. Deshpande, K. Sathian, and X. Hu, "Effect of hemodynamic variability on granger causality analysis of fmri," *Neuroimage*, vol. 52, no. 3, pp. 884–896, 2010.
- [2] A. Roebroeck, E. Formisano, and R. Goebel, "Mapping directed influence over the brain using granger causality and fmri," *Neuroimage*, vol. 25, no. 1, pp. 230–242, 2005.

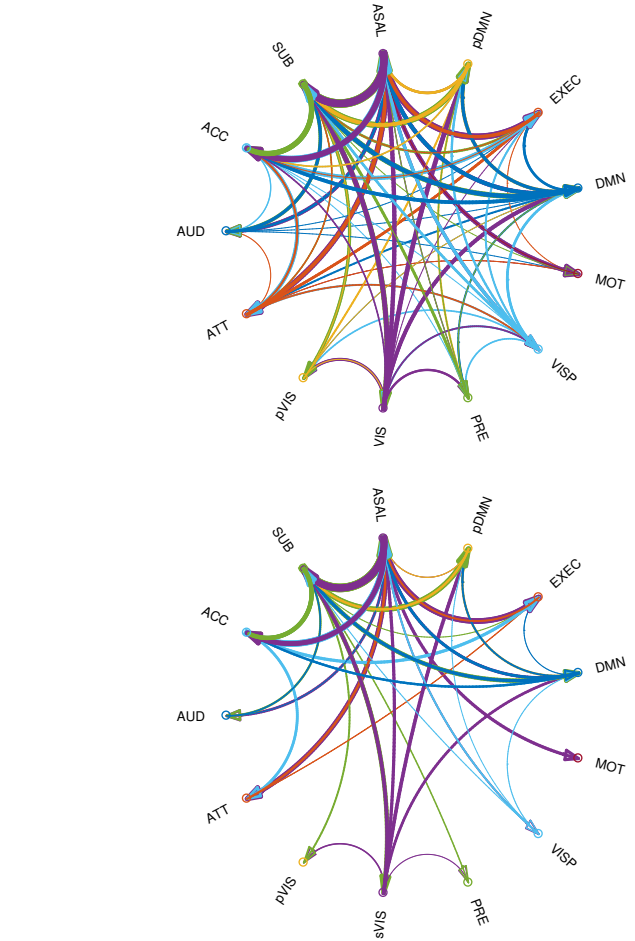


Fig. 8: Estimated causality between iCAPs by applying two thresholds, 10 and 15, respectively, to the results in Figure 6.

- [3] J. P. Hamilton, G. Chen, M. E. Thomason, M. E. Schwartz, and I. H. Gotlib, "Investigating neural primacy in major depressive disorder: multivariate granger causality analysis of resting-state fmri time-series data," *Molecular psychiatry*, vol. 16, no. 7, p. 763, 2011.
- [4] X. Wen, G. Rangarajan, and M. Ding, "Is granger causality a viable technique for analyzing fmri data?" *PloS one*, vol. 8, no. 7, p. e67428, 2013.
- [5] G. Deshpande, S. LaConte, G. A. James, S. Peltier, and X. Hu, "Multivariate granger causality analysis of fmri data," *Human brain mapping*, vol. 30, no. 4, pp. 1361–1373, 2009.
- [6] F. I. Karahanoğlu, C. Caballero-Gaudes, F. Lazeyras, and D. Van De Ville, "Total activation: fmri deconvolution through spatio-temporal regularization," *Neuroimage*, vol. 73, pp. 121–134, 2013.
- [7] F. I. Karahanoğlu and D. Van De Ville, "Transient brain activity disentangles fmri resting-state dynamics in terms of spatially and temporally overlapping networks," *Nature Communications*, vol. 6, pp. 7751 EP –, 07 2015.
- [8] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliet, "Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain," *NeuroImage*, vol. 15, no. 1, pp. 273 – 289, 2002.
- [9] T. A. Bolton, A. Tarun, V. Sterpenich, S. Schwartz, and D. Van De Ville, "Interactions between large-scale functional brain networks are captured by sparse coupled hmms," *IEEE transactions on medical imaging*, vol. 37, no. 1, pp. 230–240, 2018.