

# ONLINE GRAPH TOPOLOGY INFERENCE WITH KERNELS FOR BRAIN CONNECTIVITY ESTIMATION

Mircea Moscu\*      Ricardo Borsoi<sup>\*,†</sup>      Cédric Richard\*

\* Université Côte d’Azur, CNRS, OCA, Nice, France

† Federal University of Santa Catarina, Florianópolis, Brazil

## ABSTRACT

In graph signal processing, there are often settings where the graph topology is not known beforehand and has to be estimated from data. Moreover, some graphs can be dynamic, such as brain activity supported by neurons or brain regions. This paper focuses on estimating in an online and adaptive manner a network structure capturing the non-linear dependencies among streaming graph signals in the form of a possibly directed, adjacency matrix. By projecting data into a higher- or infinite-dimension space, we focus on capturing non-linear relationships between agents. In order to mitigate the increasing number of data points, we employ kernel dictionaries. Finally, we run a series of tests in order to experimentally illustrate the usefulness of our kernel-based approach on biomedical data, on which we obtain results comparable to state-of-the-art methods.

**Index Terms**— Topology inference, reproducing kernel, graph signal processing, adaptive algorithm, brain connectivity estimation

## 1. INTRODUCTION

Graphs represent powerful mathematical objects able to model and analyse any kind of network. Due to their inherently distributed nature, they are also suited for big data analysis, since a distributed solution can be easily applied on such an object. In applications such as gene regulation systems [1], socio-economical interactions [2], or brain activity [3], information about network structure is paramount. Most graph signal processing algorithms assume the graph structure as known beforehand. Among graph topology inference methods, most assume linear dependencies between the agents (brain regions, genes in a network, sectors of a market economy). Therefore, here resides the need of developing algorithms capable of modeling non-linear relationships, which are naturally present in real-world applications. Using reproducing kernels, through their ability to model non-linear relationships between nodal signals, is hence motivated.

In this article, we consider a setting where online nodal measurements are acquired and used in order to infer the topology of the underlying network. In the developed approach, it is desired to model non-linear dependencies, to be able to adapt to changes in the network structure, and to be distributed over the different agents, aiming at estimating a possibly directed adjacency matrix. Since many real world examples, such as social graphs, show considerable edge sparsity, the block  $\ell_1$ -norm regularisation is used in this paper due to its block sparsity-inducing properties.

**Definitions:** A graph  $\mathcal{G}$  consists of a set  $\mathcal{N}$  of  $N$  nodes, and a set  $\mathcal{E}$  of edges such that if nodes  $m$  and  $n$  are linked, then  $(m, n) \in \mathcal{E}$ . For undirected graphs, these node pairs are unordered. At the node level, we collect a signal  $\mathbf{y} \triangleq [y_1, \dots, y_N]^T$ , assumed to be real-valued,

where  $y_n$  is the sample of the signal  $\mathbf{y}$  at node  $n$ . The adjacency matrix  $\mathbf{A}$  [4, 5], is defined as an  $N \times N$  matrix whose entries  $a_{nm}$  are zero if  $(m, n) \notin \mathcal{E}$  and set to one otherwise.

**Prior works:** In topology identification, several works have been developed. An early proposition is in [6], where a covariance estimation based method of inferring links is introduced. On the same line, in [7] the graphical Lasso is employed in order to estimate the inverse covariance matrix from data. In [8], the authors advocate that connectivity can be recovered from estimated spectral templates, while the authors in [9] propose an adaptive algorithm for learning the topology from streaming graph signals driven by a diffusion process. Under a graph signal smoothness assumption, a so-called pairwise distances matrix whose entries are pairwise squared-norm distances between nodal measurements is introduced in [10] to estimate a weighted adjacency matrix.

In modeling non-linear phenomena, works such as [11, 12] focus on polynomial structural equation models, while the authors of [13] use their non-linear counterparts. They, however, have some limitations, such as assuming knowledge of certain connections or the form of the non-linear functions. Reproducing kernels have seen widespread use in topology inference problems. One of these works is [14] where kernels, chosen to best fit the data, model nonlinear relationships between nodes based on measurements at successive time instants. The authors present an auto-regressive framework that allows to track graph connectivity over time, proving useful in providing insights on brain connectivity. The multi-kernel approach in [15] uses partial correlations to encode graph topology and  $\ell_p$ -norm regression to enhance performance. In [16], a thorough analysis of the kernel-based topology inference problem is given. This work focuses on capturing non-linear and dynamic links. A review of the state of the art methods for graph topology inference is given in [17].

Encoding the graph topology with the adjacency matrix  $\mathbf{A}$  is ubiquitous in graph signal models. This operator describes the interactions between entities and, by extension, it can be considered as a tool for representing relationships between data. However, to the best of our knowledge, a kernel-based online solution to the topology inference problem has not yet been considered. In this work, we propose an online approach that can sensibly reduce computational stress, as well as adapt to slow changes in the topology that occur in dynamical environments. For the particular case of brain connectivity estimation, the developed method has the advantage of adaptability concerning data availability: due to its online nature, the data acquisition process can be stopped exactly when the estimate is obtained. This, in turn, can render the medical process of signal acquisition less strenuous for both the patient and medical personnel.

**Notations:** Normal font letters denote scalars, while boldface lowercase and uppercase letters stand for column vectors and matrices, respectively. Operator  $\|\cdot\|_0$  returns the number of their non-zero entries. Uppercase calligraphic letters denote sets, of cardinality  $|\cdot|$ .

This work was funded in part by ANR under grant ANR-19-CE48-0002.

## 2. NON-LINEAR MODEL AND CENTRALIZED PROBLEM

Consider an  $N$ -node graph with adjacency matrix  $\mathbf{A}$  which models a system such as the brain network. In this setting, the brain activity in every considered region 1 through  $N$  can be measured at different time instants  $i$ , thus obtaining a signal  $\mathbf{y}(i)$ . Each of these regions influences and is influenced by the other regions, and these links are encoded in the matrix  $\mathbf{A}$ . For the particular case of brain connectivity, the existence of non-linear connections have been reported by studies such as [18, 19]. This motivates modeling certain systems, such as the brain, with non-linear connections. With these remarks, we consider the following data model:

$$\mathbf{y}(i) = \mathbf{A}\mathbf{f}(i) + \mathbf{v}(i), \quad (1)$$

where  $\mathbf{A}$  is the adjacency matrix of the graph. This matrix models how entries of  $\mathbf{f}(i) \triangleq \text{col}\{f_m(\mathbf{y}_{L_m}(i))\}_{m=1}^N$  influence every node. We consider  $f_m : \mathbb{R}^{L_m} \rightarrow \mathbb{R}$  a non-linear function whose argument is  $\mathbf{y}_{L_m}(i) = [y_m(i), \dots, y_m(i - L_m + 1)]^\top$ . The signal  $\mathbf{v}(i)$  models innovation noise. The output at every node  $m$  is therefore non-linearly dependent of all the other signals from the other nodes, including the past. Given nodal measurements  $\mathbf{y}(i)$  acquired online, the goal is estimating the adjacency matrix  $\mathbf{A}$ .

Under the least-mean-squares criterion, the optimization problem can be expressed as:

$$\begin{aligned} \mathbf{A}^* = \underset{\mathbf{A}}{\text{argmin}} \quad & \frac{1}{2} \mathbb{E} \left\| \mathbf{y}(i) - \mathbf{A}\mathbf{f}(i) \right\|^2 + \Psi(\mathbf{A}) \\ \text{subject to } & a_{nm} \in \{0, 1\}, \end{aligned} \quad (2)$$

where  $\Psi(\mathbf{A})$  is a regularization term to account for some prior knowledge of  $\mathbf{A}$  such as symmetry or sparsity. The constraint aims at forcing the entries of the adjacency matrix to be binary. A relaxation of the constraint may be employed, by enforcing  $a_{nm} \in [0, 1]$ .

Focusing on a single node  $n$ , model (1) becomes

$$y_n(i) = \sum_{m \in \mathcal{N}} a_{nm} f_m(\mathbf{y}_{L_m}(i)) + v_n(i) \quad (3)$$

and the local optimization problem is now

$$\begin{aligned} \mathbf{a}_n^* = \underset{\mathbf{a}_n}{\text{argmin}} \quad & \frac{1}{2} \mathbb{E} \left\| y_n(i) - \sum_{m \in \mathcal{N}} a_{nm} f_m(\mathbf{y}_{L_m}(i)) \right\|^2 + \psi(\mathbf{a}_n) \\ \text{subject to } & a_{nm} \in \{0, 1\}, \end{aligned} \quad (4)$$

where  $\mathbf{a}_n$  is the  $n^{\text{th}}$  row of  $\mathbf{A}$ , with entries  $a_{nm}, m = 1, \dots, N$ .

Problem (4) has to be solved based only on local measurements  $\mathbf{y}(\ell), \ell \leq i$  that are available at a certain instant  $i$ .

## 3. RKHS AND KERNEL DICTIONARIES

### 3.1. Reproducing Kernel Hilbert Spaces

Several solutions in modeling non-linearities exist, such as non-linear and polynomial Structural Equation Models [20], as well as function selection from an existing function set [21]. The focus in this work is on kernel methods, which are able to deal with non-linearities in classification or regression problems by applying linear algorithms over a high-dimensional representation of the input data on a Reproducing Kernel Hilbert Space (RKHS)  $\mathcal{H}_k$  associated with a positive definite reproducing kernel  $k(\cdot, \cdot)$ .

For the sake of clarity, consider  $L_m = 1$  from now on, making  $y_1(i) \equiv y_m(i)$ . Let us denote  $\phi_{nm} = a_{nm} f_m$ , which allows us to

incorporate the binary variable  $a_{nm}$  and turn (4) into a problem that is linear in  $\phi_{nm}$ . Assuming that  $\phi_{nm}$  belongs to a RKHS  $\mathcal{H}_{k_m}$ , for  $m \in \mathcal{N}$ , and approximating the expected value in (4) by empirical averages computed over the available measurements for  $\ell \leq i$ , a non-parametric version of the local optimization problem for node  $n$  and time instant  $i$  can be written as:

$$\begin{aligned} \{\phi_{nm}^*\}_{m=1}^N = \underset{\phi_{nm} \in \mathcal{H}_{k_m}}{\text{argmin}} \quad & \frac{1}{2i} \sum_{\ell=1}^i \left\| y_n(\ell) - \sum_{m \in \mathcal{N}} \phi_{nm}(y_m(\ell)) \right\|^2 \\ & + \sum_{m \in \mathcal{N}} \psi_{\mathcal{H}_{k_m}}(\|\phi_{nm}\|_{\mathcal{H}_{k_m}}), \end{aligned} \quad (5)$$

where  $\psi_{\mathcal{H}_{k_m}} : \mathbb{R} \rightarrow [0, \infty[$  are non-decreasing functions. Since (5) employs a convex loss function, the conditions of the linear representer theorem are satisfied [22]. Thus, the solution to (5) admits a finite-dimensional representation of the form:

$$\phi_{nm}^*(\cdot) = \sum_{p=1}^i \alpha_{nmp} k_m(\cdot, y_m(p)), \quad m = 1, \dots, N, \quad (6)$$

where  $\alpha_{nmp} \in \mathbb{R}$  and  $\{y_m(i), \dots, y_m(1)\}$  is the set of data available at node  $m$  and time instant  $i$ .

An important question is how to introduce sparsity in the graph connections now that the problem is formulated in terms of  $\phi_{nm}$ . Since  $\phi_{nm} = a_{nm} f_m$ ,  $a_{nm} = 0$  implies that function  $\phi_{nm} \equiv 0$ . Thus, promoting sparsity over  $\mathbf{A}$  is equivalent to promoting sparsity over the functions  $\phi_{nm}$ , for  $m = 1, \dots, N$ . Fortunately, the coefficient-based representation (6) means that this can be performed equivalently by promoting sparsity of groups of variables  $\{\alpha_{nmp}\}_{p=1}^i$ , for  $m \in \mathcal{N}$ . This can be done very efficiently by using a block-sparse regularization over the coefficients, which leads to the following optimization problem for each time instant  $i$ :

$$\mathbf{\alpha}_n^* = \underset{\mathbf{\alpha}_n}{\text{argmin}} \quad \frac{1}{2i} \sum_{\ell=1}^i \left\| y_n(\ell) - \mathbf{\alpha}_n^\top \tilde{\mathbf{k}}(\ell) \right\|^2 + \eta_n \|\mathbf{\alpha}_n\|_{\mathcal{B},1}, \quad (7)$$

where the  $Ni \times 1$  block vectors  $\mathbf{\alpha}_n$  and  $\tilde{\mathbf{k}}(\ell)$  are defined as:

$$\begin{aligned} \tilde{\mathbf{k}}(\ell) &= [\mathbf{k}_1^\top(\ell), \dots, \mathbf{k}_N^\top(\ell)]^\top, \\ \mathbf{\alpha}_n &= [\tilde{\alpha}_{n1}^\top, \dots, \tilde{\alpha}_{nN}^\top]^\top, \end{aligned} \quad (8)$$

with

$$\begin{aligned} \mathbf{k}_m(\ell) &= \text{col}\{k_m(y_m(\ell), y_m(p))\}_{p=1}^i, \\ \tilde{\alpha}_{nm} &= \text{col}\{\alpha_{nmp}\}_{p=1}^i. \end{aligned} \quad (9)$$

Constant  $\eta_n > 0$  is a regularization parameter. Also, a block sparsity-inducing regularisation on  $\mathbf{\alpha}_n$  was added through the term  $\|\mathbf{\alpha}_n\|_{\mathcal{B},1}$ , the block-wise  $\ell_1$ -norm, i.e.,  $\|\mathbf{\alpha}_n\|_{\mathcal{B},1} = \sum_{m=1}^N \|\tilde{\alpha}_{nm}\|_2$ . This norm is known to promote group sparsity [23], favoring solutions with entire blocks of variables  $\tilde{\alpha}_{nm}$  equal to zero, from which it can be inferred that  $a_{nm} = 0$  and thus there will be no connection from node  $m$  towards  $n$ .

### 3.2. Optimization and kernel dictionaries

Solving (7) in batch mode is, however, impractical and computationally costly. This is why we propose a stochastic gradient descent-based solution in order to update  $\mathbf{\alpha}_n$  every instant  $i$ . Remark that a subgradient of the block- $\ell_1$  regularization  $\|\mathbf{\alpha}_n\|_{\mathcal{B},1} = \sum_{m \in \mathcal{N}} \|\tilde{\alpha}_{nm}\|_2$  is given by the block vector

$\mathbf{\Gamma}_n = [\mathbf{\Gamma}_{n1}^\top, \dots, \mathbf{\Gamma}_{nN}^\top]^\top$  [24, 25], where each block  $\mathbf{\Gamma}_{nm}$  is:

$$\mathbf{\Gamma}_{nm} = \begin{cases} \frac{\tilde{\alpha}_{nm}}{\|\tilde{\alpha}_{mn}\|_2} & \text{if } \|\tilde{\alpha}_{mn}\|_2 \neq 0 \\ \mathbf{0} & \text{if } \|\tilde{\alpha}_{mn}\|_2 = 0 \end{cases}. \quad (10)$$

This entails the use of group zero-attracting LMS (GZA-LMS) [24], leading to the following update rule, for every time instant  $i$ :

$$\hat{\alpha}_n(i+1) = \hat{\alpha}_n(i) + \mu_n[\mathbf{r}_{\tilde{k}_y} - \mathbf{R}_{\tilde{k}\tilde{k}}\hat{\alpha}_n(i) - \eta_n\mathbf{\Gamma}_n(i)], \quad (11)$$

where  $\mathbf{r}_{\tilde{k}_y} = \mathbb{E}\{\tilde{\mathbf{k}}(i)y_n(i)\}$ ,  $\mathbf{R}_{\tilde{k}\tilde{k}} = \mathbb{E}\{\tilde{\mathbf{k}}(i)\tilde{\mathbf{k}}^\top(i)\}$ . Estimating these second order moments can however prove to be unattainable or computationally costly. This warrants for the use of approximations, such as those based on instantaneous realisations:

$$\mathbf{r}_{\tilde{k}_y} \approx y_n(i)\tilde{\mathbf{k}}(i), \quad \mathbf{R}_{\tilde{k}\tilde{k}} \approx \tilde{\mathbf{k}}(i)\tilde{\mathbf{k}}^\top(i). \quad (12)$$

The use of (12) leads to the stochastic GZA-LMS update:

$$\hat{\alpha}_n(i+1) = \hat{\alpha}_n(i) + \mu_n\tilde{\mathbf{k}}(i)[y_n(i) - \tilde{\mathbf{k}}^\top(i)\hat{\alpha}_n(i)] - \mu_n\eta_n\mathbf{\Gamma}_n(i). \quad (13)$$

An immediate observation concerning update (13) is that the size of  $\tilde{\mathbf{k}}(i)$  can become prohibitive as  $i$  increases, since each acquired measurement increases the number of kernel functions. A solution to this problem are kernel dictionaries which admit a new candidate kernel function only if the candidate function passes a certain sparsification rule. Under this framework, each node  $m$  in the network creates, updates, and stores a dictionary of kernel functions,  $\mathcal{D}_m = \{k_m(\cdot, y_m(\omega_j)) : \omega_j \in \mathcal{I}_m^i \subset \{1, \dots, i-1\}\}$ , where  $\mathcal{I}_m^i$  represents the set of time indices of elements selected for the dictionary, before instant  $i$ . A candidate kernel function  $k_m(\cdot, y_m(i))$  is added in  $\mathcal{D}_m$  if the following sparsification condition holds [26]:

$$\max_{\omega_j \in \mathcal{I}_m^i} |k_m(y_m(i), y_m(\omega_j))| \leq \xi_m, \quad (14)$$

where  $\xi_m \in [0, 1[$  determines the level of sparsity and coherence of the dictionary [26]. Vector  $\tilde{\mathbf{k}}(i)$  now only stores functions which satisfy (14), leading to a rewriting of (9):

$$\begin{aligned} \tilde{\mathbf{k}}(i) &= [\mathbf{k}_1^\top(i), \dots, \mathbf{k}_N^\top(i)]^\top, \\ \mathbf{k}_m(i) &= \text{col}\{k_m(y_m(i), y_m(\omega_j))\}_{\omega_j \in \mathcal{I}_m^i}. \end{aligned} \quad (15)$$

It is worth noting that using this approach, every time one kernel function is added to a  $\mathcal{D}_m$ , all the blocks  $\hat{\alpha}_{nm}$  increase in size by one new entry,  $\forall n \in \mathcal{N}$ . Also, at each instant  $i$ ,  $\hat{\alpha}_n$  and  $\tilde{\mathbf{k}}(i)$  are of size  $\sum_{m \in \mathcal{N}} |\mathcal{I}_m^i| \times 1$ . Most importantly, the number of entries in the dictionary satisfy  $|\mathcal{I}_m^i| < \infty$  when  $i \rightarrow \infty$  [26].

Algorithm 1 summarizes the developed method. In the last step,  $\tau_n$  acts as an edge identification threshold. This parameter is used in order to identify the topology from the estimated coefficients  $\hat{\alpha}_n(i)$ , determining whether there exist links from each node  $m \in \mathcal{N}$  towards  $n$ . When processing real data,  $\tau_n$  can be set as to obtain an estimated topology which realistically explains the studied process.

## 4. EXPERIMENTS

Multiple experiments have been conducted, with the goal of showcasing different characteristics of the developed algorithm. Firstly, a simple 3-node graph is considered, where non-linearities are present in node interactions. Secondly, on real biomedical data, the obtained results are coherent with the results obtained in other works. Thirdly, we take into consideration a dynamic graph setting in order to test the adaptive capabilities of the algorithm. Along our experiments, we either used the Gaussian kernel  $k_n^G(\mathbf{a}, \mathbf{b}) = \exp(-\|\mathbf{a} - \mathbf{b}\|^2 / 2\sigma_n^2)$  or the exponential kernel  $k_n^E(\mathbf{a}, \mathbf{b}) = \exp(-\|\mathbf{a} - \mathbf{b}\| / \sigma_n)$ .

---

### Algorithm 1: Kernel-based online topology inference

---

For every node  $n$ :

**Inputs:** Parameters  $\mu_n, \eta_n, k_n(\cdot, \cdot), \xi_n$ , and  $\tau_n$

**Initialization:** Set all entries of  $\hat{\alpha}_n(0)$  to 0

**Algorithm:** At each time instant  $i \geq 1$

Introduce  $k_n(\cdot, y_n(i))$  in  $\mathcal{D}_n$  if (14) is verified

Receive and store the updated dictionaries  $\mathcal{D}_m, m \neq n$

Compute  $\tilde{\mathbf{k}}(i)$  defined in (15)

Update  $\hat{\alpha}_n(i)$  using (13)

Set  $\hat{a}_{nm}(i)$  to 1 if  $\|\hat{\alpha}_{nm}(i)\| \geq \tau_n$ , to 0 otherwise

---

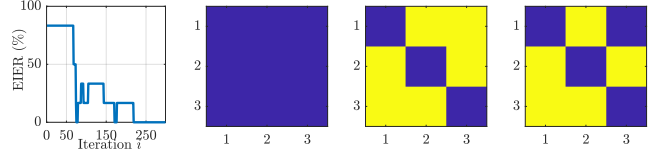


Fig. 1. Performance, as well as estimates at  $i = 50, 150, 250$ .

### 4.1. Discretized Lorenz Attractor

Consider the discretized version of the Lorenz attractor [27, 28]:

$$\begin{bmatrix} y_1^{i+1} \\ y_2^{i+1} \\ y_3^{i+1} \end{bmatrix} = \begin{bmatrix} y_1^i \\ y_2^i \\ y_3^i \end{bmatrix} + 0.01 \begin{bmatrix} 10(y_2^i - y_1^i) \\ y_1^i(28 - y_3^i) - y_2^i \\ y_1^i y_2^i - \frac{8}{3}y_3^i \end{bmatrix}, i \geq 0, \quad (16)$$

where exponents, for ease of notation, denote time instants, with initial conditions  $[y_1^0, y_2^0, y_3^0]^\top = [0.01, 0.01, 0.01]^\top$ . We used the Gaussian kernel and set  $\mu_n = 0.1, \sigma_n = 8, \xi_n = 0.8$ . Parameters  $\eta_n$  and  $\tau_n$  were set as to achieve the best performance in terms of edge identification rate, EIER  $\triangleq \|\mathbf{A}_{gt} - \hat{\mathbf{A}}\|_0 / (N(N-1)) \cdot 100\%$ , where  $\mathbf{A}_{gt}$  is the ground truth, whose binary entries encode direct influence, based on (16), between node couples, excluding self-loops.

Fig. 1 shows the EIER, and the estimates at iterations 50, 150, 250. These results, obtained after less than 300 samples with the dictionaries containing only between 6 and 8 kernel functions per node, show that the proposed method is able to infer links in a distributed and online manner, even if they are based on non-linear interactions.

### 4.2. Tests on epilepsy seizure data

In this experiment, we aimed to show how the estimated topology using the presented method is consistent with results presented in other works. The used data come from a 39-year-old female subject suffering from intractable epilepsy. The data acquisition and pre-processing information is provided in [3]. The data-set contains 8 instances of electrocorticography (ECoG) time series, each instance representing one seizure and contains voltage measurements from 76 different regions on and inside the brain, during the 10 seconds before the epilepsy seizure (*preictal* interval) and the first 10 seconds during the seizure (*ictal* interval). The Gaussian kernel was used, and we set  $\mu_n = 5 \cdot 10^{-5}, \sigma_n = 90, \xi_n = 0.9$  for each  $n$ . Concerning the choice of the  $\eta_n$  and  $\tau_n$ , since there was no ground truth, they were set as to obtain coherent results.

In Fig. 2 we show the estimated connectivity of the brain, during both these intervals, averaged over the 8 instances. Fig. 3 depicts the degree (sum of in- and out-degree), encoded in the radii of the circles, relative to each interval. Interestingly, our online estimate reveals roughly the same behavior before and during the seizure as

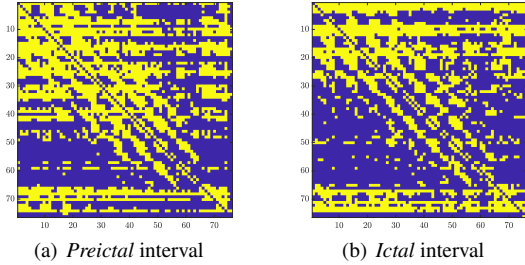


Fig. 2. Estimated adjacency matrices for each interval

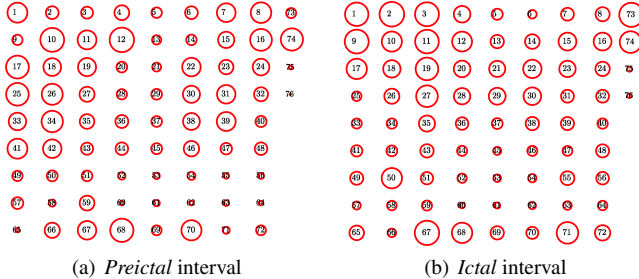


Fig. 3. Summed in- and out-degrees for the estimated graphs for both *preictal* and *ictal* intervals. The larger the radius corresponding to node  $n$ , the larger the summed degree of node  $n$ .

the estimate obtained using the method developed in [29]. More precisely, the number of total connections decreases from one interval to the other, especially due to the variation of in-degrees for nodes 30 to 50. Further analyzing the connections, nodes 75 and 76 have a small in-degree, however they present a more important out-degree. Observe the decrease of the degree of node 26 or the major increase for node 73. This behavior is consistent with the findings of the aforementioned paper. The algorithm is therefore able to obtain results similar to those obtained in previous works, while based on kernel dictionaries. For reference, the number of kernel functions inserted in the dictionaries varied between 15 to 30, after 4000 samples. These results show how only a reduced number of kernel functions are actually needed in order to obtain a topology estimate. This fact, alongside the online approach, can translate in reduced computational complexity, depending on the solver, due to the drastically reduced number of needed kernel functions.

### 4.3. Real dynamic setting

The goal of this experiment is to analyze how the proposed method adapts to slight dynamic changes in topology. Once again, we use real data whose details are found in [30]. They represent electroencephalography (EEG) measurements taken from a group of 81 subjects in total, some of which are healthy and some of which suffer from schizophrenia. A simple button-pressing task is set up, in three separate settings where subjects either: 1 - pressed the button and a tone was immediately played, 2 - listened to the tone without the button press, and 3 - pressed the button and the tone was not played. The goal of the experiment was to check how the subjects' brains respond to sensory consequences of their own actions, in healthy and unhealthy subjects. This behavior arises when, for example, one voluntarily moves their eyes from side to side and their brain knows that the environment is not actually shifting. Patients suffering from schizophrenia have difficulties in differentiating between internally

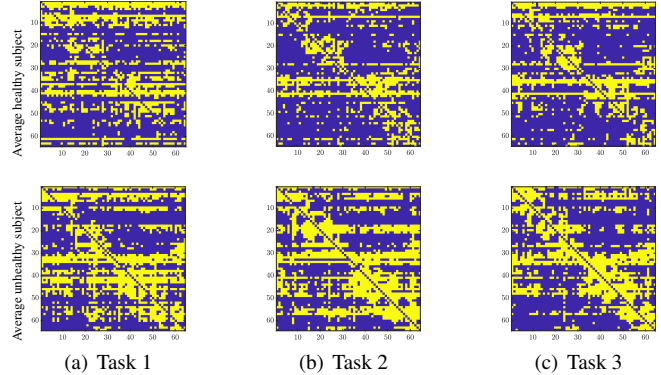


Fig. 4. Estimated topologies per task, averaged per group.

and externally generated stimuli. During our experiment, we used the measurements pertaining to three of the healthy subjects, namely subjects 1, 2, 3, and three of the schizophrenia-suffering, namely 67, 68, 69. A total of 5000 measurements per task and per subject were selected and fed to our algorithm, in order: task 1, task 2 and task 3, as if done one after the other. For both the healthy and unhealthy sets we used the exponential kernel with  $\sigma_n = 1$ ,  $\mu_n = 9 \cdot 10^{-2}$ ,  $\xi_n = 0.1$ . In order to obtain comparable results, the same manner of choosing parameters  $\eta_n$  and  $\tau_n$  was used in both cases.

The estimated topologies, averaged over the three subjects in the healthy and unhealthy groups respectively, are depicted in Fig. 4 for each of the three tasks. For both the healthy subjects (first row) and the unhealthy schizophrenia-suffering (bottom row), no important changes appear in topology while moving from task to task. However, comparing the average healthy and unhealthy subjects, a rather different network structure arises for each case. Some graph metrics are given in Table 1. For each subject in the healthy group, 30 to 80 kernel functions were chosen for the dictionary, while for each subject in the unhealthy group between 20 and 60.

Table 1. Metrics for the estimated topologies concerning the average healthy and unhealthy subject

Metric	Task	Healthy	Unhealthy
Network density	1	0.356	0.410
	2	0.286	0.451
	3	0.310	0.395
Average in- (and out-)degree	1	22.41	25.84
	2	17.98	28.44
	3	19.50	24.86

## 5. CONCLUSION

An online, kernel-based, and distributive graph topology inference method was devised, which advocates the use of kernel dictionaries as a sparsifying solution. The use of kernels allows for the inference of connections when non-linear links are presumed. While most state of the art methods rely on batch methods, which usually come with a high computational cost, the developed online algorithm comes with advantages such as adaptability. On the considered biomedical data, the method proved effective, paving the way to further work and research, such as a thorough theoretical analysis and the use of a multi-kernel approach.

## 6. REFERENCES

- [1] The International HapMap Consortium, “A second generation human haplotype map of over 3.1 million snps,” *Nature*, vol. 449, pp. 851 EP–, 10 2007.
- [2] R. H. Heiberger, “Predicting economic growth with stock networks,” *Physica A: Statistical Mechanics and its Applications*, vol. 489, pp. 102–111, 2018.
- [3] M. Kramer, E. D. Kolaczyk, and H. Kirsch, “Emergent network topology at seizure onset in humans,” *Epilepsy research*, vol. 79, pp. 173–86, 2008.
- [4] N. Biggs, *Algebraic Graph Theory*, Cambridge University Press, 1993.
- [5] A. Sandryhaila and J. M. F. Moura, “Big data analysis with signal processing on graphs: Representation and processing of massive data sets with irregular structure,” *IEEE Signal Processing Magazine*, vol. 31, no. 5, pp. 80–90, 2014.
- [6] A. P. Dempster, “Covariance selection,” *Biometrics*, vol. 28, no. 1, pp. 157–175, 1972.
- [7] J. Friedman, T. Hastie, and R. Tibshirani, “Sparse inverse covariance estimation with the graphical Lasso,” *Biostatistics*, vol. 9, no. 3, pp. 432–41, 2008.
- [8] S. Segarra, A. G. Marques, G. Mateos, and A. Ribeiro, “Network topology inference from spectral templates,” *IEEE Transactions on Signal and Information Processing over Networks*, vol. 3, no. 3, pp. 467–483, 2017.
- [9] S. Vlaski, H. P. Maretic, R. Nassif, P. Frossard, and A. H. Sayed, “Online graph learning from sequential data,” in *Proc. IEEE Data Science Workshop*, Lausanne, Switzerland, 2018, pp. 190–194.
- [10] V. Kalofolias, “How to learn a graph from smooth signals,” in *Proc. International Conference on Artificial Intelligence and Statistics*, 2016, pp. 920–929.
- [11] J. Harring, B. Weiss, and J.-C. Hsu, “A comparison of methods for estimating quadratic effects in nonlinear structural equation models,” *Psychological methods*, vol. 17, pp. 193–214, 2012.
- [12] W. Holmes Finch, “Modeling nonlinear structural equation models: A comparison of the two-stage generalized additive models and the finite mixture structural equation model,” *Structural Equation Modeling: A Multidisciplinary Journal*, vol. 22, no. 1, pp. 60–75, 2015.
- [13] N. Lim, F. D’Alché-Buc, C. Auliac, and G. Michailidis, “Operator-valued kernel-based vector autoregressive models for network inference,” *Machine learning*, vol. 99, no. 3, pp. 489–513, 2015.
- [14] Y. Shen, B. Baingana, and G. B. Giannakis, “Topology inference of directed graphs using nonlinear structural vector autoregressive models,” in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 6513–6517.
- [15] L. Zhang, G. Wang, and G. B. Giannakis, “Going beyond linear dependencies to unveil connectivity of meshed grids,” in *Proc. IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Curaçao, Dutch Antilles, 2017, pp. 1–5.
- [16] G. B. Giannakis, Y. Shen, and G. V. Karanikolas, “Topology Identification and Learning over Graphs: Accounting for Non-linearities and Dynamics,” *Proceedings of the IEEE*, vol. 106, no. 5, pp. 787–807, 2018.
- [17] X. Dong, D. Thanou, M. Rabbat, and P. Frossard, “Learning graphs from data: A signal representation perspective,” *arXiv preprint arXiv:1806.00848*, 2018.
- [18] W. J. Freeman, “EEG analysis gives model of neuronal template-matching mechanism for sensory search with olfactory bulb,” *Biological cybernetics*, vol. 35, no. 4, pp. 221–234, 1979.
- [19] J. A. de Zwart, P. van Gelderen, J. M. Jansma, M. Fukunaga, M. Bianciardi, and J. H. Duyn, “Hemodynamic nonlinearities affect BOLD fMRI response timing and amplitude,” *Neuroimage*, vol. 47, no. 4, pp. 1649–1658, 2009.
- [20] K. G. Jöreskog, F. Yang, G. Marcoulides, and R. Schumacker, “Nonlinear structural equation models: The Kenny-Judd model with interaction effects,” *Advanced structural equation modeling: Issues and techniques*, , no. 3, pp. 57–88, 1996.
- [21] X.-Y. Song, Z.-H. Lu, J.-H. Cai, and E. H.-S. Ip, “A bayesian modeling approach for generalized semiparametric structural equation models,” *Psychometrika*, vol. 78, no. 4, pp. 624–647, 2013.
- [22] B. Schölkopf, R. Herbrich, and A. J. Smola, “A generalized representer theorem,” in *International conference on computational learning theory*. Springer, 2001, pp. 416–426.
- [23] M. Yuan and Y. Lin, “Model selection and estimation in regression with grouped variables,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 1, pp. 49–67, 2006.
- [24] D. Jin, J. Chen, C. Richard, and J. Chen, “Adaptive parameters adjustment for group reweighted zero-attracting LMS,” in *Acoustics, Speech and Signal Processing (ICASSP), 2018 IEEE International Conference on*, 2018.
- [25] D. Jin, J. Chen, C. Richard, and J. Chen, “Model-driven online parameter adjustment for zero-attracting LMS,” *Signal Processing*, vol. 152, pp. 373–383, 2018.
- [26] C. Richard, J.-C. M. Bermudez, and P. Honeine, “Online prediction of time series data with kernels,” *Signal Processing, IEEE Transactions on*, vol. 57, no. 3, pp. 1058–1067, 2009.
- [27] E. N. Lorenz, “Computational chaos—a prelude to computational instability,” *Physica D: Nonlinear Phenomena*, vol. 35, no. 3, pp. 299 – 317, 1989.
- [28] F. A. Tobar, S. Kung, and D. P. Mandic, “Multikernel least mean square algorithm,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 2, pp. 265–277, 2014.
- [29] Y. Shen, B. Baingana, and G. B. Giannakis, “Nonlinear structural vector autoregressive models for inferring effective brain network connectivity,” *arXiv preprint arXiv:1610.06551*, 2016.
- [30] J. M. Ford, V. A. Palzes, B. J. Roach, and D. H. Mathalon, “Did I do that? Abnormal predictive processes in schizophrenia when button pressing to deliver a tone,” *Schizophrenia bulletin*, vol. 40, no. 4, pp. 804–812, 2013.