PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Time-resolved analysis of dynamic graphs: an extended Slepian design

Raphaël Liégeois, Ibrahim Merad, Dimitri Van De Ville

Raphaël Liégeois, Ibrahim Merad, Dimitri Van De Ville, "Time-resolved analysis of dynamic graphs: an extended Slepian design," Proc. SPIE 11138, Wavelets and Sparsity XVIII, 1113810 (9 September 2019); doi: 10.1117/12.2530550



Event: SPIE Optical Engineering + Applications, 2019, San Diego, California, United States

Time-resolved analysis of dynamic graphs: an extended Slepian design

Raphaël Liégeois^{a,b}, Ibrahim Merad^{a,c}, and Dimitri Van De Ville^{a,b}

^aInstitute of Bioengineering, Centre for Neuroprosthetics, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

^bDepartment of Radiology and Medical Informatics, University of Geneva, Geneva, Switzerland ^cDépartement de Mathématiques et Applications, École Normale Supérieure, Paris, France

ABSTRACT

Graphs are extensively used to represent networked data. In many applications, especially when considering large datasets, it is a desirable feature to focus the analysis onto specific subgraphs of interest. Slepian theory and its extension to graphs allows to do this and has been applied recently to analyze various types of networks. One limitation of this framework, however, is that the number of subgraphs of interest is typically limited to one. We introduce an extended Slepian design that allows to consider an arbitrary number of subgraphs of interest. This extension offers the possibility to encode prior information about multiple subgraphs in a two-dimensional plane. As a proof of concept and potential application, we demonstrate that this framework allows to perform time-resolved and spatio-temporal analyses of dynamic graphs.

Keywords: Slepian functions, dynamic graphs, localization, signal processing, graph embedding

1. INTRODUCTION

The development of various research fields ranging from neurosciences to social interactions has brought a lot of attention to the analysis of large networked data. In this context, graphs are extensively used to encode information on irregular domains and spectral graph theory has played a central role to reveal network properties not directly available in the nodal domain¹. For example, the graph Laplacian eigenvector with the smallest nonzero eigenvalue can be used to partition a graph^{2,3}. The Laplacian is also at the center of the graph Fourier transform (GFT) that allows to transpose classical Fourier analysis tools, such as frequency-based filtering, to graphs⁴. Finally, the problem of finding low-dimensional projections that maximally preserve distance relationships between nodes. -a.k.a. graph embedding- also finds its solutions in the Laplacian eigenvectors⁵.

Slepian functions were defined by Slepian and colleagues in a series of seminal papers as bandlimited signals that are maximally concentrated on a temporal interval^{6–9}. Such functions allow to guide the analysis of datasets by enforcing concentration of information in specific subsets of interests. The idea of balancing signal spread in time and frequency can be transposed to functions defined on graphs, thereby allowing to define Slepian functions on graphs^{10,11} or explore graph extensions of uncertainty principles¹². Other variations of the original Slepian design have been proposed, e.g., in order to get rid of the bandlimitedness constraint¹³, but these frameworks only allow to consider a single interval of interest. One way to overcome this limitation is to allow for negative and positive weights in the Slepian selection matrix, thereby allowing to consider two intervals of interest¹⁴.

In this contribution we propose an extended graph Slepian design that allows to consider an arbitrary number of subgraphs. The next Sections provide basic theoretical background and definitions of the Slepian framework (Sections 2.1-2.2). The modified concentration criterion on which the proposed Slepian extension is based is presented in Section 2.3. We finally illustrate the potential benefits of the proposed framework on two -toy and real- datasets, showing that the extended Slepian design allows to guide the analysis of dynamic graphs in both the vertex and temporal domains (Section 3).

Send correspondence to R.L. (Raphael.Liegeois@epfl.ch) or D.V.D.V. (Dimitri.Vandeville@epfl.ch)

Wavelets and Sparsity XVIII, edited by Dimitri Van De Ville, Manos Papadakis, Yue M. Lu, Proc. of SPIE Vol. 11138, 1113810 · © 2019 SPIE · CCC code: 0277-786X/19/\$21 · doi: 10.1117/12.2530550

2. THEORETICAL FRAMEWORK

2.1 Slepian functions on graphs

In seminal work, Slepian, Landau and Pollak solved the problem of optimally concentrating a signal in both space and time^{6–9}. While this original work considered continuous functions and concentration intervals, we here review the graph case that is directly relevant to the extension we introduce and the applications we consider.

The structure of a weighted undirected graph containing N nodes is typically encoded in an $N \times N$ adjacency matrix \mathbf{A} , where $A_{i,j}$ indicates the (positive) edge weight between nodes i and j for all i, j = 1, ..., N. The graph Laplacian is defined as $\mathbf{L} = \mathbf{D} - \mathbf{A}$, where \mathbf{D} is the diagonal degree matrix with elements $D_{i,i} = \sum_{j=1}^{N} A_{i,j}$. Beside the graph and its structure, a graph signal is defined as a mapping from the N nodes to a length-N vector associating a scalar value with every node. Based on the spectral decomposition of \mathbf{L} : $\mathbf{L} = \mathbf{U}\Lambda\mathbf{U}^{-1}$, the graph Fourier transform (GFT) of a graph signal \mathbf{f} has been defined by analogy with the the classical discrete Fourier transform as $\hat{\mathbf{f}} = \mathbf{U}^{-1}\mathbf{f}$, and the entries of Λ are referred to as graph frequencies 4 .

The graph Slepian problem is expressed as finding a bandlimited length-N graph signal \mathbf{g} with maximal concentration in a subgraph \mathcal{S} . In other words, the aim is to maximize

$$\mu = \frac{\sum\limits_{k \in \mathcal{S}} g[k]\bar{g}[k]}{\sum\limits_{k=1}^{N} g[k]\bar{g}[k]}$$
(1)

for the bandlimited graph signal \mathbf{g} . The spectral constraint is expressed as $\mathbf{g} = \mathbf{U}\mathbf{W}\hat{\mathbf{g}}$, where \mathbf{U} is an $N \times N$ matrix encoding the inverse GFT and \mathbf{W} is an $N \times N_W$ matrix that selects the first N_W Fourier components $\hat{\mathbf{g}}$ with smallest graph frequencies¹¹. Denoting \mathbf{S} the $N \times N$ matrix with $\mathbf{S}_{k,k} = 0/1$ indicating the absence or presence of k in \mathcal{S} , Eq. (1) rewrites

$$\mu = \frac{\hat{\mathbf{g}}^H \mathbf{C} \hat{\mathbf{g}}}{\hat{\mathbf{g}}^H \hat{\mathbf{g}}},\tag{2}$$

2.2 Augmented Slepians

One limitation of the classical Slepian framework is that it is not possible to define more than one interval of interest. A first step to overcome this limitation was proposed by Demesmaeker and colleagues who introduced augmented Slepians that allow for the presence of positive (1), zero, or negative (-1) entries in the selection matrix S^{14} . This extension was proposed in the continuous domain and the corresponding criterion to be maximized for a graph signal g writes:

$$\mu = \frac{\sum_{k \in \mathcal{S}^{+}} g[k]\bar{g}[k] - \sum_{k \in \mathcal{S}^{-}} g[k]\bar{g}[k]}{\sum_{k=1}^{N} g[k]\bar{g}[k]},$$
(3)

where S^+ (S^-) is the subset corresponding to positive (negative) entries of **S**. As in the classic case, problem (3) can be turned into an eigendecompostion problem in the spectral domain as in Eq. (2), the only difference being that the selection matrix **S** has positive *and* negative entries. The eigenvalues of the concentration matrix $C = W^T U^T SUW$ now range between -1 for Slepian vectors fully concentrated in S^- and 1 for Slepian vectors fully concentrated in S^+ .

2.3 Towards an arbitrary number of intervals

Building upon the extension proposed by Demesmaeker and colleagues¹⁴, we propose to populate the selection matrix S with complex-valued entries, thereby allowing to consider an arbitrary number of subgraphs. More precisely, when considering M subgraphs of interest S^1, \ldots, S^M , the diagonal entries of the selection matrix are such that $S_{kk} = \rho_i \cdot e^{\phi_i}$ if and only if $k \in S^i$, and $\rho_i \cdot e^{\phi_i}$ is the complex weight associated to S^i .

The corresponding concentration criterion to be maximized writes

$$\mu = \frac{\sum_{i=1}^{M} \left(\rho_i \cdot e^{\phi_i} \sum_{k \in \mathcal{S}^i} g[k] \bar{g}[k] \right)}{\sum_{k=1}^{N} g[k] \bar{g}[k]}.$$

$$(4)$$

It can be seen from Eq. (4) that Slepian vectors associated with eigenvalues of \mathbf{C} distributed along the $[0, \rho_i \cdot e^{\phi_i}]$ complex segment are getting more and more concentrated in \mathcal{S}^i as the eigenvalue approaches $\rho_i \cdot e^{\phi_i}$. Therefore, the interval capturing the concentration of a given Slepian vector is determined from the phase of the corresponding eigenvalue: if $\angle(\mu_k) \in [\phi_i - \delta, \phi_i + \delta]$, where δ is a phase tolerance parameter to be defined, then the corresponding Slepian vector is associated to the interval \mathcal{S}^i . Slepian vectors associated to the same interval are then ranked based on their magnitude. We denote \mathbf{s}_{kl} the Slepian vector associated to interval \mathcal{S}^k having the l^{th} largest magnitude. For example, \mathbf{s}_{31} is the Slepian vector that is maximally concentrated in \mathcal{S}^3 and \mathbf{s}_{42} is the Slepian vector with the second highest concentration in \mathcal{S}^4 . Finally, the concentration of a vector \mathbf{g} in \mathcal{S}^i is given by $\sum_{k \in \mathcal{S}^i} g[k] \bar{g}[k] / \sum_{k=1}^N g[k] \bar{g}[k]$.

In the applications presented in the next Section, we consider the simple case where the weights of the selection matrix are the roots of unity, i.e., $\rho_i = 1$ and $\phi_i = \frac{2(i-1)j\pi}{M}$ for i = 1, ..., M. We note that in that case, the criterion of Eq. (4) reduces to the case of augmented Slepians (Eq. (3)) when considering only two subgraphs (M = 2).

3. MATERIALS AND METHODS

3.1 Application to the chain graph

We first consider a chain graph which consists of a graph representation of a discrete timeline. Each time point is represented by a node, and successive time points (as well as the first and the last time points) are connected by edges in the graph adjacency matrix \mathbf{A}_C :

$$\mathbf{A}_{C} = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 1 \\ 1 & 0 & 1 & 0 & \dots & 0 \\ 0 & 1 & 0 & 1 & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \dots & 0 & 1 & 0 & 1 \\ 1 & \dots & 0 & 0 & 1 & 0 \end{pmatrix}.$$
 (5)

The Laplacian of the chain graph is $\mathbf{L} = \mathbf{D} - \mathbf{A}_C$, where \mathbf{D} is the degree matrix. We consider a chain graph of length 2000 (i.e., \mathbf{L} is a square matrix of size 2000) and three intervals: $\mathcal{S}^1 = [400, 799]$, $\mathcal{S}^2 = [800, 1199]$ and $\mathcal{S}^3 = [1200, 1599]$. The weights for these intervals were $e^{\frac{2\cdot 0\cdot j\pi}{3}} = 1$ for the entries of \mathbf{S} corresponding to \mathcal{S}^1 , $e^{\frac{2j\pi}{3}}$ for the entries corresponding to \mathcal{S}^3 . We use a bandlimit $N_W = 20$ and the phase tolerance $\delta = 0.3$.

3.2 Application to a multivariate dynamic graph

The chain graph is univariate because a single node of the graph is associated to each time point. We now consider a multivariate dynamic graph representing yearly commercial exchanges within the EU countries plus Switzerland over 145 years, from 1870 to $2014^{15,16}$. For a given year, these exchanges are encoded in a symmetric adjacency matrix \mathbf{A}_i of size 29 (28 EU countries plus Switzerland). The overall graph adjacency matrix \mathbf{A} is then built by (i) stacking the 145 successive yearly adjacency matrices \mathbf{A}_i on the main diagonal of \mathbf{A} and (ii) filling the secondary diagonals of \mathbf{A} by identity matrices of size 29 in order to encode the graph temporal structure:

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{1} & \mathbb{I} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbb{I} & \mathbf{A}_{2} & \mathbb{I} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbb{I} & \mathbf{A}_{3} & \mathbb{I} & \dots & \mathbf{0} \\ \vdots & & \ddots & & \vdots \\ \mathbf{0} & \dots & \mathbf{0} & \mathbb{I} & \mathbf{A}_{144} & \mathbb{I} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbb{I} & \mathbf{A}_{145} \end{pmatrix}.$$
(6)

The selection matrix \mathbf{S} has the same size as \mathbf{A} (29 · 145 = 4205) and is built as follows. We define 5 temporal intervals: $\mathcal{S}^1 = [1870, 1900]$, $\mathcal{S}^2 = [1901, 1930]$, $\mathcal{S}^3 = [1931, 1960]$, $\mathcal{S}^4 = [1961, 1990]$ and $\mathcal{S}^5 = [1991, 2014]$. We also focus our analysis on six selected countries with largest GDP among the 29*: Switzerland, Spain, France, Netherlands, United Kingdom (UK) and Italy. The entries of the selection matrix are $\mathbf{S}_{k,k} = 0$ when k corresponds to an unselected country and $\mathbf{S}_{k,k} = e^{j\Phi_{s(k)}}$ otherwise, with s(k) the index of the temporal interval associated to k and $\Phi_i = \frac{2(i-1)\pi}{5}$ for $i = 1, \ldots, 5$. We use a bandlimit $N_W = 500$ and the phase tolerance $\delta = 0.3$.

4. RESULTS AND DISCUSSION

4.1 Proof of concept

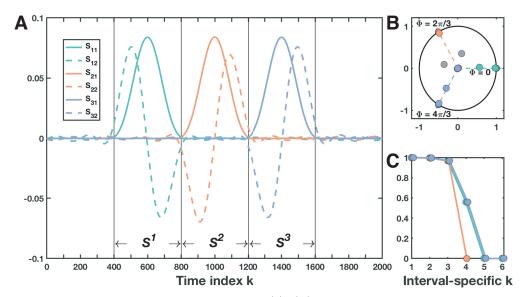


Figure 1. Extended Slepian design on the chain graph of Eq. (5). (A) Top-2 Slepian vectors in each selected interval (\mathcal{S}^1 , \mathcal{S}^2 or \mathcal{S}^3). (B) Eigenvalues of the extended concentration matrix are attributed to an interval and color-coded based on their phases: green for \mathcal{S}^1 , orange for \mathcal{S}^2 and blue for \mathcal{S}^3 . Unclassified eigenvalues are shown in grey. (C) Concentration curves of the eigenvalues attributed to the three intervals with same color-coding as in (A-B).

^{*}We did not select Germany because its different partitions in the last two centuries did not allow a consistent analysis through the time period covered by the dataset.

Figure 1 shows the results on the chain graph described in Eq. (5) with three intervals of interest. The $N_W = 20$ complex eigenvalues of \mathbf{C} are shown in Figure 1B and attributed to \mathcal{S}^1 , \mathcal{S}^2 or \mathcal{S}^3 based on their phases. The Slepian vectors corresponding to the two eigenvalues with largest magnitude in each interval are shown in Figure 1A. It can be seen that the Slepian vectors are concentrated in the interval selected by their corresponding eigenvalue. We note that only the real part of the Slepian vectors are shown because in this case the imaginary part is negligible. Figure 1C shows the concentration curve of the successive Slepian vectors in each interval. As in the original Slepian decomposition 11, we observe that in each interval a sharp phase transition occurs between highly concentrated Slepian vectors and poorly concentrated Slepian vectors.

Figure 2 shows the results of the extended Slepian decomposition of the multidimensional dynamic graph of Eq. (6). Figure 2A shows the $N_W = 500$ complex eigenvalues of C and are distributed along 5 segments corresponding to the 5 temporal intervals described in Section 3.2. In each temporal interval, the two corresponding Slepian vectors with highest magnitudes are averaged and used to build the embeddings of Figure 2B-F. Within each of these panels, selected countries -Switzerland, Spain, France, Netherlands, United Kingdom (UK) and Italy- are represented with red dots while remaining 23 countries are represented with blue dots. We only represent the amplitudes of Slepian vector entries in order to simplify visualization of the results.

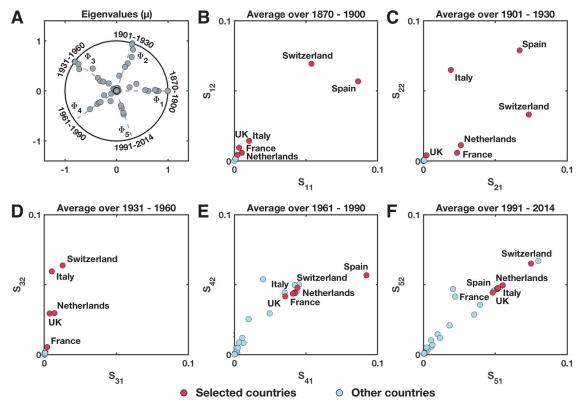


Figure 2. Extended Slepian design on the multivariate dynamic graph of Eq. (6). (A) Eigenvalues of the extended concentration matrix. (B-F) 2-D embedding of the 29 countries using the two Slepian vectors with highest magnitude in each interval, averaged within each interval. Selected countries are represented with red dots and unselected countries are represented with blue dots.

It can be seen from Figure 2A that the eigenvalues of the extended concentration matrix \mathbf{C} are distributed in the complex plane along radial segments with phases corresponding to the weights of the 5 selected temporal intervals in the selection matrix \mathbf{S} . This suggests that the proposed decomposition allows to specifically guide the concentration of the Slepian vectors within these temporal intervals. The time-resolved 2-D embedding shown in Figures 2B-F offers an original representation of the evolution of the commercial organization between the 29 countries considered here, with a focus on the 6 selected countries. For example, we observe that in the first

temporal segments (late 19th - early 20th century) the selected countries are scattered in the space spanned by the two first Slepian vectors. In contrast, the 5 selected EU countries (Spain, France, Netherlands, UK and Italy) seem to cluster apart from Switzerland (non-EU country) in later temporal intervals which might be explained by the development of EU commercial agreements.

4.2 Conclusion and future work

We introduced an extension of the graph Slepian framework that theoretically allows to consider an arbitrary number of subgraphs. The potential of this extension is illustrated on two dynamic datasets. For example we show that after building an adjacency matrix encoding the temporal structure of the data (e.g., Eq. (6)), different subgraphs can be chosen so as to capture portions of the adjacency matrix corresponding to successive temporal intervals. Different graph spectral embeddings in these temporal intervals are then computed, thereby providing a novel way to perform time-resolved analyses of dynamic graphs. More generally, this versatile tool could be adapted to guide graph analyses in multiple dimensions but further work is needed to assess its potentialities and limitations.

First, the nature of the graph Slepian eigenvectors associated to eigenvalues that were not attributed to any interval should be clarified (e.g., grey dots in Figure 1B). Second, in the two toy examples that we present, the temporal resolution is based on the partition of the data along its temporal dimension into segments of similar lengths. In order to zoom into specific time periods, we might explore the possibility of changing this temporal partition. Alternatively, an irregular distribution of the weights of the selection matrix S on the complex unit circle might also be considered to achieve the desired temporal partitioning. In general, these approaches should be compared to other types of dynamic graphs analyses such as the ones based on sliding windows, autoregressive models, multilayer networks or topological properties^{17–21}. Third, we restricted our analyses to the real parts (Figure 1) or amplitudes (Figure 2) of the results and future work is needed to explore the complex nature of the extended graph Slepian vectors. Finally, one could consider applying the proposed complex weighting of the selection matrix on the combined Slepian criterion introduced by Petrović and colleagues¹³. This criterion combines maximization of energy concentration as defined in Eq. (1) with minimization of a modified embedded distance¹¹, thereby leading to a stable spectral decomposition with no bandlimitedness constraint. One challenge of this extension will be to deal with the negative eigenvalues of the combined Slepian criterion that might make the match between eigenvalues of the concentration matrix and selected intervals harder than in the case presented here.

ACKNOWLEDGMENTS

This research was supported in parts by the CHIST-ERA IVAN project (20CH21 174081), the Center for Biomedical Imaging (CIBM) of the Geneva - Lausanne Universities and the EPFL.

REFERENCES

- [1] Newman, M. E. J., "Networks: an introduction," Oxford University Press (2010).
- [2] Fiedler, M., "Laplacian of graphs and algebraic connectivity," Combinatorics and Graph Theory 25, 57–70 (1989).
- [3] Merris, R. R. R., Merris, and Merris, R., "Laplacian matrices of graphs: a survey," *Linear algebra and its applications* **197**, 143–176 (1994).
- [4] Shuman, D. I., Narang, S. K., Frossard, P., Ortega, A., and Vandergheynst, P., "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Processing Magazine* **30**, 83–98 (May 2013).
- [5] Goyal, P. and Ferrara, E., "Graph embedding techniques, applications, and performance: A survey," *Knowledge-Based Systems* **151**, 78 94 (2018).
- [6] Slepian, D. and Pollak, H. O., "Prolate spheroidal wave functions, fourier analysis and uncertainty i," Bell System Technical Journal 40(1), 43–63 (1961).
- [7] Landau, H. J. and Pollak, H. O., "Prolate spheroidal wave functions, fourier analysis and uncertainty—ii," Bell System Technical Journal 40(1), 65–84 (1961).

- [8] Landau, H. J. and Pollak, H. O., "Prolate spheroidal wave functions, fourier analysis and uncertainty—iii: The dimension of the space of essentially time- and band-limited signals," Bell System Technical Journal 41(4), 1295–1336 (1962).
- [9] Slepian, D., "Prolate spheroidal wave functions, fourier analysis and uncertainty iv: Extensions to many dimensions; generalized prolate spheroidal functions," *Bell System Technical Journal* 43(6), 3009–3057 (1964).
- [10] Tsitsvero, M., Barbarossa, S., and Di Lorenzo, P., "Signals on graphs: Uncertainty principle and sampling," *IEEE Transactions on Signal Processing* **64**, 4845–4860 (Sep. 2016).
- [11] Van De Ville, D., Demesmaeker, R., and Preti, M. G., "When slepian meets fiedler: Putting a focus on the graph spectrum," *IEEE Signal Processing Letters* **24**, 1001–1004 (July 2017).
- [12] Agaskar, A. and Lu, Y. M., "A spectral graph uncertainty principle," *IEEE Transactions on Information Theory* **59**, 4338–4356 (July 2013).
- [13] Petrovic, M., Bolton, T. A. W., Preti, M. G., Liégeois, R., and Van De Ville, D., "Guided graph spectral embedding: Application to the c. elegans connectome," *Network Neuroscience*, 1–20 (Mar 2019).
- [14] Demesmaeker, R., Preti, M. G., and Van De Ville, D., "Augmented slepians: Bandlimited functions that counterbalance energy in selected intervals," *IEEE Transactions on Signal Processing* 66, 4013–4024 (Aug 2018).
- [15] Barbieri, K., Keshk, O. M., and Pollins, B. M., "Trading data: Evaluating our assumptions and coding rules," *Conflict Management and Peace Science* **26**(5), 471–491 (2009).
- [16] Barbieri, K. and Keshk, O. M., "Correlates of war project trade data set codebook, version 4.0.," *Online:* http://correlatesofwar.org. (2016).
- [17] Mucha, P., Richardson, T., Macon, K., Porter, M., and Onnela, J.-P., "Community structure in time-dependent, multiscale, and multiplex networks," *Science* **328**(5980), 876–878 (2010).
- [18] Liégeois, R., Mishra, B., Zorzi, M., and Sepulchre, R., "Sparse plus low-rank autoregressive identification in neuroimaging time series," *Proc.* 54th IEEE Conf. Decision and Control, 3965–3970 (Dec. 2015).
- [19] Bianconi, G., "Multilayer networks: structure and function," Oxford University Press (2018).
- [20] Khambhati, A. N., Sizemore, A. E., Betzel, R. F., and Bassett, D. S., "Modeling and interpreting mesoscale network dynamics," *Neuroimage* **180**, 337–349 (10 2018).
- [21] Saggar, M., Sporns, O., Gonzalez-Castillo, J., Bandettini, P. A., Carlsson, G., Glover, G., and Reiss, A. L., "Towards a new approach to reveal dynamical organization of the brain using topological data analysis," *Nat Commun* 9, 1399 (04 2018).