### Title

Smooth graph learning for functional connectivity estimation




### Authors

Gao, Siyuan
Xia, Xinyue
Scheinost, Dustin
et al.



### Publication Date

2021-10-01

### DOI

10.1016/j.neuroimage.2021.118289

Peer reviewed

HHS Public Access

Author manuscript

*Neuroimage*. Author manuscript; available in PMC 2021 November 23.

Published in final edited form as:

*Neuroimage*. 2021 October 01; 239: 118289. doi:10.1016/j.neuroimage.2021.118289.

# Smooth graph learning for functional connectivity estimation

**Siyuan Gao**[a,*], **Xinyue Xia**[b,*], **Dustin Scheinost**[a,c,d,e], **Gal Mishne**[b,f,g]

[a]Department of Biomedical Engineering, Yale University, United States, 06520

[b]Neurosciences Graduate Program, University of California San Diego, United States, 92093

[c]Department of Radiology and Biomedical Imaging, Yale School of Medicine, United States, 06510

[d]Department of Statistics and Data Science, Yale University, United States, 06520

[e]Child Study Center, Yale School of Medicine, United States, 06510

[f]Halıcıoğlu Data Science Institute, University of California San Diego, United States, 92093

[g]Department of Electrical and Computer Engineering, University of California San Diego, United States, 92093

## Abstract

Functional connectivity (FC) estimated from functional magnetic resonance imaging (fMRI) signals is important in understanding neural representation and information processing in cortical networks. However, due to a lack of “ground truth” FC pattern, the reliability and robustness of FC estimates are usually examined in downstream FC analysis tasks, such as performing participant’s identification (also known as “fingerprinting”). In this paper, we propose to learn FC via a smooth graph learning framework. In particular, we treat each time frame of the fMRI time series as a graph signal on an underlying functional brain graph, and estimate the smooth graph functional connectivity (SGFC) by learning the weighted graph adjacency matrix based on graph signal smoothness assumption. We demonstrate that our approach gives rise to a natural and sparse graph representation of FC from which reliable graph measures can be extracted. Reliability of SGFC is evaluated in the context of fingerprinting and compared to correlation FC (CFC). SGFC achieves higher fingerprinting accuracy across several different experiment settings; the improvement is even more significant when a shorter fMRI scanning length is used for FC estimation. In addition to being reliable, we also validate the cognitive relevance of SGFC by using it to predict fluid intelligence. Finally, in evaluating topological measures of the sparse graph, SGFC reveals a more small-world and modular structure compared to CFC. Together, our results suggest that the smooth graph learning framework produces a naturally sparse, reliable, and cognitive-relevant representation of functional connectivity.

gmishne@ucsd.edu (Gal Mishne).

*these authors contributed equally to this work

Conflict of interest: The authors declare no competing financial interests;

**Keywords**

functional connectivity; fMRI fingerprinting; reliability; small-world

## 1. Introduction

Functional magnetic resonance imaging (fMRI), and especially functional connectivity (FC) analyses, have shown success in revealing robust individual differences that are related to neural activity patterns (Dubois and Adolphs, 2016; Finn et al., 2015).

In graph-based FC analyses, an individual's brain is modeled as a graph structure where the affinity matrix of each graph is used in downstream analyses (e.g., graph theoretical analysis, Bassett and Bullmore (2006); Bullmore and Sporns (2009); Fornito et al. (2013), functional connectome fingerprinting, Finn et al. (2015), connectome-based predictive modeling, Shen et al. (2017)). In these models, the graph nodes are defined by the brain regions and the edge weights between nodes are defined using several different methods (Salvador et al., 2005; Mohanty et al., 2020).

The most popular method is to use the Pearson correlation coefficients between the average fMRI time series of the brain regions. The reliability and effectiveness of the Pearson correlation-based FC have been validated in participant's identification (Finn et al., 2015), also known as 'functional connectome fingerprinting'. In FC-based participant identification, individual FCs act as a 'fingerprint' that can be used to identify individual subjects from a large group. More specifically, most subjects' FCs are shown to be the most similar to their own FCs in different runs or task scanning conditions compared with other subjects' FCs.

In the original study by Finn et al., the similarity was calculated by the Pearson correlation between the vectorized FC affinity matrices. Recently, methods that consider the intrinsic structure of the Pearson correlation matrices have also been proposed to calculate the distance between correlation functional connectivity (CFC) matrices to reflect their underlying non-Euclidean geometry (Varoquaux et al., 2010; Venkatesh et al., 2020). CFC lies on a non-linear surface called the positive semidefinite cone. Therefore, the correct distance between CFC matrices should be measured with respect to the cone geometry. More specifically, instead of calculating the similarity between CFC matrices by their Pearson correlation, either 1) projecting the correlation-based matrix onto its tangent space and calculating the distance, or 2) calculating the geodesic distance between matrices leads to a more accurate estimate of the distances between matrices. In another recent paper about geodesic distances between CFC matrices, the authors demonstrated that adding a dataset-dependent regularization to CFC matrices constituted more reliable participant's identification (Abbas et al., 2021). Inspired by the fact that FC can be used for fingerprinting, it is subsequently utilized to successfully predict a wide range of phenotypic measures, including fluid intelligence (Shen et al., 2017), sustained attention (Rosenberg et al., 2018), and creative ability (Beaty et al., 2018). These experiments validate that the FCs, especially Pearson correlation-based FCs, are individually unique in a way that can be used to predict various phenotypic measures.

While correlation is one of the most popular approaches to estimate FC, it still has notable limitations. Firstly, it inherently generates FCs with non-random topological structures (Zalesky et al., 2012), and more specifically, CFCs are inherently more clustered than random networks due to the transitive nature of the correlation coefficient that the indirect paths and their corresponding direct paths are more likely to coexist compared with random networks. Because of the induced intrinsic topological structure, graph measures (e.g., small-worldness) derived from CFCs need to be properly normalized to account for the transitive structure. Moreover, as the CFC, by nature, is fully connected, signed, and weighted, the graph often needs to be thresholded or binarized before one can use graph-theoretic analysis. The thresholding is normally done by setting absolute (correlation-based) or proportional (sparsity-based) thresholds (Garrison et al., 2015). However, it fails to take into account the topological property of the original graph during the process which might result in thresholded graphs that have drastically different topological structures than the original graph (van den Heuvel et al., 2017). More sophisticated graph thresholding methods that preserve some graph structures have been proposed to solve the problem (Bordier et al., 2017; Esfahlani and Sayama, 2018). Additionally, how to interpret the negative correlations when global signal regression (GSR) is performed remains a topic of debate (Fox et al., 2009; Murphy et al., 2009; Saad et al., 2012; Murphy and Fox, 2017). The negative edge weights can be converted to the absolute values or simply ignored, and different choices might lead to different graph properties.

In this paper, we propose learning the FC based on recent advances in graph signal processing (Kalofolias, 2016; Dong et al., 2016). Unlike previous approaches, our graph signal processing perspective of fMRI treats the brain activity at each time point as a graph signal and models the FC as the graph to support these graph signals. The brain graph will then be inferred from the measured signals based on a graph signal smoothness assumption, i.e., the observed signals are smooth with respect to the inferred graph. We call the resulting brain graph ”smooth graph functional connectivity” (SGFC). Moreover, by including a node degree penalty that penalizes nodes with small degrees, it also promotes the FC to be a connected graph. The learned graphs have non-negative edge weights with a hyperparameter to control for the graph sparsity. As dealing with negative edge weights is still a topic of open debate, generating non-negative FC matrices removes the need to choose how to deal with negative edge weights, e.g., makes sparsifying the graph easier and more efficient while taking into account the global structure of the data.

Graph signal processing has been previously proposed in related work on functional brain imaging analysis (Preti and Van De Ville (2019), see Huang et al. (2018) for a review), modeling the brain activity at each time point as a graph signal on the underlying brain graph. However, in Huang et al. (2018) the brain graph (i.e. the weighted adjacency matrix) was provided *a priori*, based on the structural connectivity extracted from diffusion-weighted MRI. Their framework then used the structural connectivity graph as the basis to perform graph signal processing, i.e. applying graph filtering or graph wavelet transformation to the brain activity signals for downstream analysis. The brain graph in our proposed framework, on the other hand, is not given *a priori*, and our goal is to infer it from the fMRI data.

We demonstrate that SGFC shows improved fingerprinting accuracy compared with the CFC with correlation or geodesic distance. Using either edge weights or node degrees, we show that the SGFC is more reliable across repeated scanning sessions. Moreover, the SGFC is also shown to be capable of predicting fluid intelligence, suggesting that the learned FCs from different subjects are unique, and in a meaningful way. Overall, these results show that smooth graph learning generates more reliable and also cognitive-relevant neural activity representations.

## 2. Methods

### 2.1. Graph learning from smooth data

For each subject $s$ and under each task condition $c$, we are given an observed regional time series matrix $\mathbf{X} \in \mathbb{R}^{N \times T}$, where $N$ is the number of functionally coherent regions, and $T$ is the task-dependent time-series length. We seek to compute a $N \times N$ FC matrix that captures inter-regional functional relationships for each observed $\mathbf{X}$. The most common approach calculates the FC matrix by computing the pairwise Pearson correlation coefficients between rows of $\mathbf{X}$, and thus views $\mathbf{X}$ from the perspective of $N$ brain regional time-series signals (Figure 1 top). We propose a change of perspective: viewing $\mathbf{X}$ in terms of its columns, and treating each column of $\mathbf{X}$ as a graph signal supported on a common latent functional brain graph $\mathcal{G}$ (Figure 1 bottom). Processing of such graph signals has been developed in the recent field of graph signal processing, to address data that is observed on an irregular grid, such as sensor networks, traffic flow, brain activity and tensor-based data (Shuman et al., 2013; Huang et al., 2018; Mishne et al., 2016; Ortega et al., 2018; Stanley et al., 2020) and extend classical signal processing to the domain of networks and graphs. Consequently, we seek to infer the FC matrix by learning the functional brain graph, using recent techniques from graph learning and graph signal processing (Kalofolias, 2016; Dong et al., 2016; Mateos et al., 2019).

Our framework is illustrated in Figure 1. Formally, we characterize the unknown latent brain connectivity as a weighted undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{W})$, where $\mathcal{V} = \{1, 2, \ldots, N\}$ is the set of nodes corresponding to the set of brain regions, $\mathcal{E} \in \mathcal{V} \times \mathcal{V}$ is the set of edges connecting the nodes, and $\mathbf{W}$ is the weighted adjacency matrix for $\mathcal{G}$. Notice that since the graph is undirected, $\mathbf{W}$ is symmetric, and we further assume that the graph only takes non-negative edge weights, with $\mathbf{W}[i, j] \geq 0$ capturing the degree of similarity between nodes $i$ and $j$. Let $\mathbf{x}_t = (\mathbf{X}[1, t], \mathbf{X}[2, t], \ldots, \mathbf{X}[N, t])^\mathsf{T}$ be the $t$-th column of $\mathbf{X}$; each of the $T$ columns of $\mathbf{X}$ can be viewed as a graph signal on the same graph $\mathcal{G}$. This is visualized in the bottom panel of Figure 1. Graph signals are plotted overlaying the brain nodes, and the value and sign of the graph signal at a certain time point for each node is indicated by the height and direction of the arrow at that node. In the graph signal modeling context, the most widely used assumption is that signals supported on a graph change smoothly between connected nodes of the graph (Zhu et al., 2003; Kalofolias, 2016; Dong et al., 2016; Egilmez et al., 2017). Specifically, a graph signal is considered smooth if the signal values of well-connected node pairs are similar. A common way to quantify the smoothness of a graph signal $\mathbf{x} \in \mathbb{R}^N$ with respect to the graph is through the Laplacian quadratic form,

$\mathbf{x}^\top \mathbf{L}\mathbf{x} = \sum_{(i,j)\in\mathcal{E}} \mathbf{W}[i,j](\mathbf{x}[i] - \mathbf{x}[j])^2$, where $\mathbf{L} = \mathbf{D} - \mathbf{W}$ is the graph Laplacian, and $\mathbf{D}[i,i] = \sum_j \mathbf{W}[i,j]$ is the diagonal weighted degree matrix. This term measures smoothness because it will be small if the entries of $\mathbf{x}$ have similar values on nodes connected by large edge weights. In the case where we have multiple graph signals, $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_T]$, we simply use the matrix form of the Laplacian quadratic form, $\mathrm{tr}\left(\mathbf{X}^\top \mathbf{L}\mathbf{X}\right) = \sum_{t=1}^{T} \mathbf{x}_t^\top \mathbf{L}\mathbf{x}_t$, to measure the overall smoothness of the signals.

The Laplacian quadratic form has been popular in semi-supervised learning and manifold embedding (Zhu et al., 2003; Belkin et al., 2006; Belkin and Niyogi, 2002; Stanley et al., 2020) tasks to regularize the learned feature representation. More recently, it has been used in approaches that try to learn the underlying graph from data. Given the data matrix $\mathbf{X}$ (whose columns are graph signals), Dong et al. (2016) and Hu et al. (2015) both proposed to learn the graph through the graph Laplacian $\mathbf{L}$ via solving an optimization problem of the form:

$$\begin{aligned} \min_{\mathbf{L}} \quad & \mathrm{tr}\left(\mathbf{X}^\top \mathbf{L}\mathbf{X}\right) + \gamma \parallel \mathbf{L} \parallel_F^2 \\ \text{s.t.} \quad & \mathrm{tr}(\mathbf{L}) = N \\ & \mathbf{L}_{ij} = \mathbf{L}_{ji} \leq 0, i \neq j \\ & \mathbf{L} \cdot \mathbf{1} = 0, \end{aligned} \tag{1}$$

where $\mathbf{1} = [1, \ldots, 1]^\top$ is the vector of all ones. The minimization objective in the above optimization problem consists of two terms. The first trace term is the Laplacian quadratic form. Finding a graph Laplacian matrix $\mathbf{L}$ that yields a small value for this term means that we are searching for a graph such that the given graph signals in the columns of $\mathbf{X}$ are smooth on it (in terms of its graph Laplacian matrix $\mathbf{L}$). The second Frobenius term in the objective function is a regularization term that controls the sparsity and distribution of edge weights in the learned graph. The constraints in this problem ensure that the learned matrix $\mathbf{L}$ is a valid graph Laplacian matrix (symmetric with negative off-diagonal values and zero degree).

However, this formulation is not scalable when learning large graphs from data. Kalofolias (2016) resolved the scalability issue by reformulating problem (1) into an optimization problem with respect to the weighted adjacency matrix $\mathbf{W}$. By recognizing that $\mathrm{tr}(\mathbf{X}^\top \mathbf{L}\mathbf{X}) = \mathrm{tr}(\mathbf{W}\mathbf{Z})$, where $\mathbf{Z}[i,j] = \|\mathbf{X}[i,:] - \mathbf{X}[j,:]\|^2$ is the squared pairwise Euclidean distance matrix between rows of $\mathbf{X}$, Kalofolias (2016) proposed to transform the graph Laplacian learning problem of (1) into a more general form:

$$\begin{aligned} \min_{\mathbf{W}} \quad & \mathrm{tr}(\mathbf{W}\mathbf{Z}) + \mathscr{R}(\mathbf{W}) \\ \text{s.t.} \quad & \mathbf{W} = \mathbf{W}^\top \\ & \mathbf{W} \geq 0 \\ & \mathrm{diag}(\mathbf{W}) = 0 \end{aligned} \tag{2}$$

The minimization objective again consists of two terms: first a Laplacian quadratic term (now written in terms of the weighted adjacency matrix $\mathbf{W}$) that promotes the data to be

smooth with respect to the learned graph, and a regularization term for the graph's weighted adjacency matrix. The constraints in the problem ensure that the learned matrix **W** is a valid weighted graph adjacency matrix (symmetric and non-negative). Two important differences can be noticed between the above graph learning problem (2) and the problem in (1). First, instead of learning the graph in terms of the graph Laplacian matrix **L**, here the graph is learned directly in terms of its weighted adjacency matrix **W**. This entails a simplified search space for the problem and makes graph learning scalable (Kalofolias, 2016). Second, we now have a more general expression for the regularization term, where $\mathscr{R}(\cdot)$ can be any user-defined regularization function for the graph weights. This generalization allows us to impose structure on the learned graph beyond the given data being smooth on the learned graph.For $\mathscr{R}(\mathbf{W}) = \gamma\left( \| \mathbf{W1} \|_2^2 + \| \mathbf{W} \|_F^2 \right)$, where an $l_2$-norm based regularizer acts on the weighted degree, problem (2) becomes equivalent to problem (1), so the regularization controls sparsity and distribution of edge weights. When replacing the $l_2$-norm regularizer by a log-barrier function as proposed in Kalofolias (2016), the new regularizer becomes $\mathscr{R}(\mathbf{W}) = -\alpha \mathbf{1}^\mathsf{T} \log(\mathbf{W1}) + \beta \| \mathbf{W} \|_F^2$. The logarithmic barrier term controlled by $\alpha$ forces the degree of each node to be positive, and thus promotes the overall connectivity of the graph. The Frobenius term controlled by $\beta$ regularizes the sparsity of the learned graph in the sense that larger $\beta$ favors the formation of dense small edges over sparse large edges, leading to an overall denser graph. Hence by using proper regularization, we can learn graphs that are both sparse and connected, which are desirable properties for FC estimation.

In this paper, we adopted the general formulation of problem (2) to learn brain connectivity graphs (in terms of **W**) using both kinds of regularization for each subject and under each task condition. The optimization was solved using the primal-dual algorithm as described in Kalofolias (2016) and Komodakis and Pesquet (2015).

### 2.2. Experiment setup

**Dataset—**We applied all algorithms (see Competing methods below) to the Human Connectome Project (HCP) 900 Subject Release (Van Essen et al., 2013). These data releases were the only releases available at the time that this work began. From this dataset, we restricted our analyses to those individuals who participated in all nine fMRI conditions (7 task, 2 rest), whose mean frame-to-frame displacement was less than $0.1mm$ and whose maximum frame-to-frame displacement was less than $0.15mm$, and for whom fluid intelligence measures were available ($n = 506$; 237 males; ages 22–37). This conservative threshold for exclusion due to motion was used to mitigate the substantial effects of motion on functional connectivity.

**fMRI processing—**The HCP minimal preprocessing pipeline was used (Glasser et al., 2013), which included artifact removal, motion correction, and registration to standard space. All subsequent preprocessing was performed using image analysis tools available in BioImage Suite and included standard preprocessing procedures (Finn et al., 2015). Several covariates of no interest were regressed from the data including linear and quadratic drifts, mean cerebral-spinal-fluid (CSF) signal, mean white-matter signal, and mean gray matter signal. For additional control of possible motion-related confounds, a 24-parameter motion

model (including six rigid-body motion parameters, six temporal derivatives, and these terms squared) was regressed from the data. The data were temporally smoothed with a Gaussian filter (approximate cutoff frequency=0.12Hz). Mean frame-to-frame displacement yielded seven motion values per subject, which were used for subject exclusion and motion analyses. fMRI data were parcellated into 268 nodes using a whole-brain functional atlas defined in a separate sample (Shen et al., 2013). These 'raw' time series were then used to generate functional connectivity based on different methods, and no regression of task-evoked activity was performed. The global mean was not regressed from the data. However, we also discuss the influence of global signal regression in the supplementary materials (Figure S4)

**Pearson correlation-based functional connectivity—**Pearson correlation matrices were generated for both the LR and RL phase encoding runs in the HCP data, resulting in two 268×268 connectivity matrices per individual per task condition. Nine nodes (node index: {249, 239, 243, 129, 266, 109, 115, 118, 250}) which had missing coverage during any individual's scan were excluded from all individuals. These matrices were used to perform fingerprinting or generate cross-validated predictive models of fluid intelligence.

**Fingerprinting analysis—**In the HCP dataset, we performed fingerprinting using two approaches: within-task fingerprinting and cross-task fingerprinting. Within-task fingerprinting was done by matching each individual's LR scan FC with all the available RL scan FCs in the same scan condition. The RL scan FC with the smallest distance was considered as the match to the individual. If the individual's LR scan was matched to the same individual's RL scan, it would be counted as a successful fingerprinting. Fingerprinting accuracy was then calculated as the number of successful identifications divided by the number of individuals. To test the within-task fingerprinting accuracy on the HCP dataset, we randomly sampled 100 individuals (experiment results with other numbers of sampled individuals are included in supplementary materials) from 506 individuals without repetition. The sampling process was repeated 20 times. In each repetition, we calculated the fingerprinting accuracy independently. Instead of matching the LR scan to the RL scan under the same task condition, cross-task fingerprinting was performed by matching the individual's LR scan to all the available LR scans under a different task condition. No repetition of subsampling was performed in cross-task fingerprinting. All 506 subjects were used to test the cross-task fingerprinting accuracy.

**Tangent space projection and geodesic distance—**Covariance matrices lie on a positive semidefinite (PSD) cone where the geometry is not Euclidean. Vectorizing these covariance matrices directly and using these vectors as features is not ideal as the features will not be independent because of this PSD constraint. Moreover, using the correlation or Euclidean distance to compare two vectorized covariance matrices is also inaccurate because these matrices lie on the cone structure. In order to compare different covariance matrices in a way that reflects the geometry of the underlying cone, tangent space parametrization is utilized to project the matrices to the Euclidean space. More specifically, the PSD cone is treated as a Riemannian manifold and the matrices are projected onto the tangent space of a selected reference point $\mathbf{C}_R$ on the cone. We denote the covariance matrix by $\mathbf{C}_i$, which lies

on the PSD cone. The Euclidean mean of all $\mathbf{C}_i$ in the same fMRI scanning condition (e.g., gambling task, working memory task) is chosen to be the reference point $\mathbf{C}_R = \frac{1}{N}\sum_i \mathbf{C}_i$ as it was shown that it generated the most stable performance across different reference point estimation methods (Pervaiz et al., 2020). Then for any covariance matrix $\mathbf{C}_i$ on the PSD cone, the tangent space projected matrix is calculated as $\overrightarrow{\mathbf{C}_i} = \log_m\left(\mathbf{C}_R^{-1/2}\mathbf{C}_i\mathbf{C}_R^{-1/2}\right)$, where $\log_m$ represents the matrix logarithm. The matrix logarithm is the generalization of the scalar logarithm, such that the matrix exponential of the matrix logarithm equals the original matrix. The projected matrices $\overrightarrow{\mathbf{C}_i}$ can then be vectorized and used as features that are no longer inter-correlated. Geodesic distances on the PSD cones adopt similar ideas to tangent space projection, to compare two covariance matrices. To calculate the geodesic distance between two covariance matrices $\mathbf{C}_1$ and $\mathbf{C}_2$, one of the two matrices, for example $\mathbf{C}_2$, can be used as the reference point. Then the projected $\mathbf{C}_1$ on $\mathbf{C}_2$'s tangent space is calculated as $\overrightarrow{\mathbf{C}_1} = \log_m\left(\mathbf{C}_2^{-1/2}\mathbf{C}_1\mathbf{C}_2^{-1/2}\right)$. Their geodesic distance can then be computed as (Pennec et al., 2006)

$$d_G(\mathbf{C}_1, \mathbf{C}_2) = \sqrt{\mathrm{tr}\left(\log_m^2\left(\overrightarrow{\mathbf{C}_1}\right)\right)} = \sqrt{\sum_{i=1}^{n}(\log(\lambda_i))^2}$$

where $\{\lambda_i\}_{i=1}^{n}$ are the $n$ eigenvalues of $\overrightarrow{\mathbf{C}_1}$ of size $n \times n$. It is also worth noting that the geodesic distance is symmetric. Both the tangent space and geodesic distance require that the matrix that is used as the reference point is invertible, which calls for the need of regularizing the matrices before calculation. In our experiment, we adopted the suggested regularization of adding a scaled identity matrix $\lambda\mathbf{I}$ (Venkatesh et al., 2020; Abbas et al., 2021). We set $\lambda = 14$ for all the task CFCs with short scan length in within-task fingerprinting (details can be found in supplementary materials), and we set $\lambda = 1$ is for all the CFCs in cross-task fingerprinting.

**Connectome-based Predictive Modeling (CPM)—**CPM (Shen et al., 2017) is a validated method for extracting and pooling the most relevant features from connectivity data in order to construct linear models to predict behavioral measures. Briefly, on the training dataset, edges of connectivity matrices that are significantly correlated with the behavior of interest are first selected (thresholding by $p$-value). The selected features are then fit to learn the regression model coefficients via ridge regression (Gao et al., 2019). The regularization strength of the ridge regression is determined by another inner 10-fold cross-validation on the training dataset. The model can then be used to predict the behavior of novel participants. CPM was tested on four kinds of FCs (correlation, tangent, $l_2$-penalized and log-penalized). For the correlation-based FC, we used a $p$-value threshold of 0.5 to filter out edges that were not correlated with fluid intelligence in the training dataset (the 0.5 threshold was chosen based on the previous paper, Gao et al. (2019)). For the other three FCs, we did not filter out edges as we do not expect these FCs to have CFC's feature dependence issue. FCs from each condition's LR and RL runs were averaged before building the model.

**Phenotypic measure—**In the HCP dataset, a 24-item version of the Penn Progressive Matrices test—a measure of fluid intelligence—was used as the phenotypic measure for prediction. This test is an abbreviated form of Raven's standard progressive matrices (Bilker et al., 2012).

**Prediction evaluation—**10-fold cross-validation was used to evaluate all models. In our experiment, the whole sample was randomly divided into 10 approximately equal-sized groups; on each fold, the model was trained on 9 groups and tested on the excluded 10th group. This process was iteratively repeated 10 times, with each group excluded once. Model performance was evaluated by the cross-validated $R^2$,

$$R^2_{\mathrm{CV}} = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

where $n$ is the number of data points, $y_i$ is the observed data, $\hat{y}_i$ is the prediction value and $\bar{y}$ is the mean of the observed data. $R^2_{\mathrm{CV}}$ could be negative and those negative values were set to 0. In this paper, $\sqrt{R^2_{\mathrm{CV}}}$ is reported as it is comparable to, but less biased than, the most commonly used Pearson correlation between observed and predicted measures when using cross-validation (Scheinost et al., 2019). This whole 10-fold cross-validation was repeated 20 times.

**Task decoding—**To decode the task conditions, we used a multi-class support vector machine (SVM) and validated the performance of each kind of FC by 10-fold cross-validation. Matlab's fitcecoc() multi-class SVM function was adopted, which fitted $K(K-1)/2$ binary SVM models using the one-versus-one coding design for multi-class classification, where $K$ was the number of unique class labels. Default SVM model parameters were used including linear kernel and $l_1$ regularized hinge loss. For each fold, training subjects' FCs in all scanning conditions were used to train the model which was then tested on the testing subjects' FCs to decode the scanning condition. A high task decoding accuracy suggests a more reliable topological structure that is sensitive to the condition of the graph.

**Graph Metrics—**We calculated graph metrics on both CFC and SGFC to examine their network typologies. Common measures of network organizations including average clustering coefficient, characteristic path length, global efficiency, assortativity and modularity were computed using the Brain Connectivity Toolbox (see Rubinov and Sporns (2010) for detailed description), and are presented in the main text and supplementary. In addition to these common measures, we also computed the small world propensity (SWP; Muldoon et al. (2016)), which quantifies the extent to which a network displays the structure of small-worldness, an important property that is ubiquitous across many real-world networks. Formally, SWP reflects the deviation of a network's clustering coefficient $C_{\mathrm{obs}}$ and characteristic path length $L_{\mathrm{obs}}$, from both lattice ($C_{\mathrm{latt}}$, $L_{\mathrm{latt}}$) and random

($C_{rand}$,$L_{rand}$) networks of the same number of nodes and the same degree distributions:

$$SWP = 1 - \sqrt{\frac{1}{2}\left(\left(\frac{C_{latt} - C_{obs}}{C_{latt} - C_{rand}}\right)^2 + \left(\frac{L_{obs} - L_{rand}}{L_{latt} - L_{rand}}\right)^2\right)}.$$

**Data availability—**The HCP data used in this study are publicly available from the ConnectomeDB database (https://db.humanconnectome.org). MATLAB scripts to run the SGFC analyses can be found at (https://github.com/carricky/graph_learning_FC).

## 3. Results

### 3.1. Smooth graph functional connectivities improve fingerprinting accuracy

Four different methods to perform FC-based fingerprinting were compared. Among these four methods, correlation and geodesic distance-based fingerprinting were all based on the CFC matrices. The other two methods were based on graph learning. We labeled the method that imposed an $l_2$-norm node degree regularizer as '$l_2$-penalized' while the other method that imposed log regularizer as 'log-penalized'.

Fingerprinting were tested both within the same scanning condition/task or cross distinct conditions/tasks. In most of the tasks, within-task fingerprinting was the most accurate by the $l_2$-penalized or the log-penalized SGFCs (Figure 2a). Only in one of the resting-state scanning sessions, geodesic distance-based fingerprinting outperformed SGFC. Among the two graph learning methods with different regularizers, there was no significant difference between their performances (corrected resampled $t$-test, Bouckaert and Frank (2004), $p > 0.1$) and both achieved above 90% accuracy in most of the conditions. In terms of cross-task fingerprinting, scanning runs corresponding to the LR direction were used. Similar to the within-task fingerprinting, the two graph-learning methods still had consistent high accuracy across all tasks (Figure 2b). In comparison, the fingerprinting accuracy of the correlation and geodesic distances all dropped. In particular, the geodesic distance-based fingerprinting showed inconsistency in the accuracy of different tasks. Fingerprinting was more accurate when the two resting-state scans were used as the matching dataset for the row scanning conditions. However, the opposite was not true. Moreover, as we observed the importance of matrix regularization on geodesic distances, we added an identity matrix ($\lambda = 1$) to regularize all scanning conditions regardless of the scanning length so that the cross-task conditions can be properly compared. We set $\lambda = 1$ to keep the same experiment setting as the original paper (Venkatesh et al., 2020).

### 3.2. Topological measures derived from smooth graphs are more reliable

As the graph-theoretical metrics, e.g., node degree, of the functional connectivity is often used as a powerful feature in the downstream analyses (Wang et al., 2010), it is worthwhile to test if these metrics are also reliable, or more specifically, if they can be used for functional fingerprinting as well. As shown in previous research (Salehi et al., 2020), parcel sizes of the brain (number of voxels assigned to each parcel) could be used to predict which task condition the subject was in. More specifically, using the parcel sizes from individualized state-dependent functional parcellations as features and the functional state as output, the authors were able to build predictive models that predicted the functional state

of each unseen sample solely based on the size of parcels in that parcellation. We adopted a similar concept here to validate if the topological measures were also reliable in the 'task dimension'. More specifically, we viewed reliability testing of fMRI scans in two orthogonal dimensions: subject dimension and task dimension. The within-task functional fingerprinting in the previous subsection assessed the reliability by fixing the task and varying different subjects to check if the subject could be matched correctly. In task decoding, the subject was fixed and for each testing subject, and we predicted which task (e.g., gambling task, working memory task) each scan belonged to by a classifier trained on other subjects. A higher task decoding accuracy will suggest that the FCs are reliable under the same scanning condition and potentially have information that is only related to the specific task.

Three different kinds of FCs were involved in our experiments: CFC, $l_2$-regularized SGFC and log-regularized SGFC. Node degrees derived from these three kinds of FCs were calculated. As mentioned earlier in the manuscript, CFCs have signed and weighted edge weights. We used the absolute values to calculate the node degree. Moreover, as the SGFCs were sparse by nature, to eliminate the potential influence of sparsity, CFCs were also sparsified to match the sparsity level of the subject's corresponding log-penalized graph. The sparsification was done by retaining the edges with the top absolute value of edge weights.

As the task decoding results show (Figure 3a), the two SGFC variants generate FCs with unique and identifiable topological structures in each scanning conditions.

As generating an identifiable topological structure for each scanning condition was orthogonal to having high fingerprinting accuracy, we also tested fingerprinting using the node degree instead of edge weight vectors. Again, the two SGFC variants generated significantly higher fingerprinting accuracy compared to the CFC (Figure 3b), suggesting a more identifiable and reliable topological pattern within the same subject.

### 3.3. Smooth graph fingerprinting accuracy is less affected by short scanning time length

Scanning length is a key factor determining the reliability of FC estimates (Horien et al., 2018). While most existing fMRI studies adopt a scanning length between 5–7 minutes of resting-state data to estimate correlation-based connectivity, reports have suggested increasing this duration to improve reliability (Birn et al., 2013; Noble et al., 2019). In contrast, our result shows that when using the log-penalized SGFC as our FC estimate, we were able to achieve decent fingerprinting accuracy on resting-state data with only 200 TR (2.4 minutes). The graph learning method surpassed the other two methods by a large margin on resting-state fingerprinting accuracy for all scanning durations less than 400 TRs (Figure 4a). This suggests that when given limited time points to estimate individual FC, SGFC is more robust than CFC. An intuitive explanation is that the dense FC matrix computed using correlation-based methods may contain many noisy edge weight estimates due to the limited measurements, whereas in the graph learning method, through proper sparsity regularization, the learned FC graph contains fewer noisy edges, thus making it robust to limited samples.

To test the shorter scanning length's influence on all the other conditions, the fMRI time series from each task were truncated to 176 TRs by taking the middle part of the scan

so that the length matched the task with the shortest scanning length (emotion task). As shown in the results (Figure 4b), most of the task condition's SGFC-based fingerprinting accuracy retained similar accuracy using the full scanning length (Figure 2b). Resting-state conditions show decreased fingerprinting performance as it was previously shown to be less reliable for identifying subjects compared with task conditions (Finn et al., 2017). In comparison, correlation or geodesic distance-based fingerprinting was still worse than graph learning-based fingerprinting and the difference was more obvious compared with the results using the full scanning length, suggesting that the graph learning-based method is also more robust when a shorter scanning length is available.

### 3.4. Smooth graphs predict cognitive measures

We showed that the FCs are unique and reliable within the same subject. However, the immediate next question to ask is if the differences between subject's unique FCs are cognitively meaningful. We performed CPM with FCs generated with different methods. There were no statistically significant differences between the prediction using plain CFC and the two SGFC methods in all scanning conditions ($p > 0.1$, corrected resampled $t$-test, Bouckaert and Frank (2004)). However, there were differences between the tangent space projected CFC and the two SGFC methods. More specifically, in resting-state based prediction, SGFC-based prediction was significantly worse than the prediction with CFC tangent space projection (Figure 5). The good prediction performance using tangent space projection for resting-state data aligns with previous research (Dadi et al., 2019). However, prediction with tangent space projection worsened significantly and under-performed the two SGFC methods when using the task-based data. Our hypothesis is that tangent space projection's dependency on matrix inversion makes it unstable for task-based prediction as resting-state has a longer scanning length in the HCP dataset (resting-state: 1200 TRs, longest task, WM, 405TRs). Achieving higher behavioral measure prediction accuracy with CPM is not the goal of this paper as CPM is an established prediction framework for CFC data. Instead, these behavior measure prediction results serve as evidence that SGFCs are also cognitively meaningful and that SGFC can be used with CPM. Moreover, it is worth mentioning that no CPM edge selection was performed with SGFC prediction while the optimal $p < 0.5$ threshold (based on Gao et al. (2019)) was used for CFC. The results that SGFC still performed similarly with CFC suggests that removing unreliable edges may not be necessary for SGFC-based CPM prediction due to the induced sparse graph structure.

### 3.5. Sparsification of SGFC entails a biologically meaningful brain graph

As discussed in the Methods section, the regularization terms in the graph learning framework promote the learned graph to be sparse and connected. In particular, by adjusting the parameter in the graph learning objective, we can obtain graphs of different densities. In comparison, post hoc sparsification can always be performed on the dense CFC to sparsify it to a certain density level. As there is no 'ground truth' density for brain FCs, we want to study the log-penalized SGFC and CFC at different graph densities. To this end, we computed the log-penalized brain graph with different sparsity levels by varying $\beta$ in the graph learning objective (SGFC can be sparsified effectively by only varying $\beta$, more details can be found in the supplementary materials), and correspondingly sparsifying the CFC to the same sparsity level by keeping the same number of edges sorted by the absolute value

of their weights in descending order. It is also worth noting that the sparsification process was performed on the individual level, so although all the subjects' FCs will have the same sparsity, the sets of edges that were set as zero were different across subjects.

We first investigated the change of fingerprinting accuracy with different graph densities (Figure 6). Although a decrease in fingerprinting accuracy with fewer nonzero edges might be expected, we found that sparser log-penalized SGFCs had higher fingerprinting accuracy, whereas CFC's fingerprinting accuracy kept almost constant across different sparsity levels. When the FCs were very sparse (3% density, ~ 1000 edges), fingerprinting was able to be done with almost perfect accuracy, suggesting that a robust topological structure was learned within each subject as the nonzero edges all have similarly high edge weights.

Inspired by the reliable subject-wise topological structure, we also computed the small world propensity (SWP, Muldoon et al. (2016)) and maximum modularity score for the log-penalized SGFC and CFC at different sparsity levels (Figure 7, see supplementary Figures S8–S9 for other graph metrics). For a dense FC, the log-penalized SGFC and CFC had similar SWP and modularity scores. However, when the two kinds of FCs were sparsified to the same density level, the log-penalized graphs achieved higher SWP and modularity scores than the corresponding CFCs. This suggests that with sparsity regularization, the sparse graph obtained from SGFC retains a more biologically meaningful structure than the post hoc sparsification of the CFC. Moreover, in our experiment, as the FCs were sparsified, the SWP and modularity score change of the log-penalized SGFCs was smoother compared with the CFC, i.e., every step of sparsification led to an effective change in the SWP and modularity score, suggesting that the natural sparsification by the graph learning approach is topologically more effective and meaningful. In comparison, the SWP and modularity score did not change much when the CFC was sparsified initially (density from 1 to 0.7). As the brain is well-known as a small-world network (Bassett and Bullmore, 2017), the more 'small-worldness' of the sparse log-penalized SGFC also suggests that its sparsification better reveals the brain's intrinsic topological structure.

## 4. Discussion

In this paper, we proposed a smooth graph learning-based FC estimation framework to estimate reliable and cognitively relevant brain graphs. While most FC evaluations are based on the subsequent analysis and there is no ground truth to evaluate and compare different kinds of FCs, SGFCs were shown to be more reliable and have the same or better utility as CFCs. First, we used fingerprinting to investigate the reliability of the SGFCs compared with both the commonly-used correlation-based fingerprinting and the more recent geodesic distance-based fingerprinting. Next, we validated the SGFC by showing its advantage in both task decoding and fingerprinting using the node degree calculated from the graph. Lastly, we illustrated SGFC's utility by showing that it has comparable performance with CFC in successfully predicting fluid intelligence for all task-based scans.

### 4.1. Choosing between $l_2$ and log penalty for smooth graph learning

We tested two variants of the graph learning framework: $l_2$-penalized and log-penalized. Overall, we found the two methods performed similarly in most experiments while the

log-penalized graph was sometimes better (e.g., fingerprinting with node degree, Figure 3b, also see Supplementary for comparisons of the two variants on synthetic data). This is expected as the log-penalized graph better promotes the overall connectivity of the graph and thus is our recommended graph learning method to use. However, the performance margin between the two SGFC methods is small.

### 4.2. Computation time

One of the disadvantages we noticed when using geodesic distances for fingerprinting was the increased computation time. Specifically, to compute the geodesic distance between two $N \times N$ matrices, the time complexity is $O(N^3)$ as a matrix inversion, two matrix multiplications, and an SVD decomposition are needed. Even if the matrix inversion can be precomputed and stored for each individual, the matrix multiplication and SVD of the multiplication product are inevitable. However, for the distance calculated with Pearson correlation, the time complexity is $O(N^2)$. In practice, the time needed to finish computing the pairwise distances between 100 subjects in one task is around 86s using geodesic distance and less than 0.1s using the correlation distance. The test was run on a machine with an 8-Core Intel i9 CPU @ 2.3Ghz and 32GB of RAM and the algorithm implemented on MATLAB R2020a 64-bit version. This computation burden made tasks involving more subject pairs harder to evaluate (e.g., cross-task fingerprinting).

### 4.3. Strengths and limitations

The present study shows that applying SGFC as a FC estimation framework has the benefit of reducing noisy spurious connections, and the resulting FC matrices serve naturally as the weighted adjacency matrix of a latent brain graph. Its sparsity is induced by the algorithm's regularization design and gives rise to a biologically meaningful functional brain graph that has high SWP and modularity scores. It is also more reliable (higher fingerprinting accuracy with the whole graph or derived graph measures) and more robust (fingerprinting is less affected by shorter scanning length and number of participants, see supplementary materials). Given the popularity and utility of spectral graph theory in the analysis of fMRI data, explicitly learning a weighted graph facilitates the direct application of such analyses. Furthermore, as opposed to CFC, which is constructed using only pairwise information, the learned graph via the solution of the regularized optimization problem, better reflects the global structure, as is evident in our sparsification experiments. However, both our current graph-learning based approach and correlation-based approach suffer from the limitation that the FC matrices inferred are symmetric, and thus directed functional dependencies cannot be inferred. Future research would therefore extend the current graph learning framework from undirected graphs to directed graphs (Marques et al., 2020) and apply to FC analysis.

### 4.4. Conclusions

Estimating FC from fMRI time series has always been an important step in understanding fMRI data. In this paper, we proposed to use the smooth graph learning framework to learn naturally sparse and weighted brain graphs. We compared the fingerprinting accuracy using the smooth graph with the traditionally-used correlation graph. In our results, not only was the fingerprinting accuracy improved, the accuracy was also less affected by

shortened scanning length. FCs estimated from smooth graph learning were also able to predict fluid intelligence. Another important advantage of smooth graph learning compared with correlation graph is that it is naturally sparse, induced by the regularization design. Compared with the sparse graph generated from post hoc thresholding of the correlation graph, the sparse smooth graph tends to be more small-world and modular under the varying graph densities, suggesting that this natural sparsification better reveals the intrinsic topological structure. Although the framework was only tested on fingerprinting and behavioral measure prediction, future research will focus on other applications of SGFC such as cohort classification. Moreover, to utilize SGFC's natural sparsity and reliability, more topological measure-based FC analyses will be examined to provide new insights about our brain's topological structure compared with CFC-based analyses.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements:

Data were provided in part by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657 funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University). This study was funded in part by R01MH121095 and R01EB026936.

## References

Abbas K, Liu M, Venkatesh M, Amico E, Kaplan AD, Ventresca M, Pessoa L, Harezlak J, Goñi J, 2021. Geodesic distance on optimally regularized functional connectomes uncovers individual fingerprints. Brain Connectivity.

Bassett DS, Bullmore E, 2006. Small-world brain networks. The Neuroscientist 12, 512–523. URL: 10.1177/1073858406293182, doi:10.1177/1073858406293182, [PubMed: 17079517]

Bassett DS, Bullmore ET, 2017. Small-world brain networks revisited. The Neuroscientist 23, 499–516. [PubMed: 27655008]

Beaty RE, Kenett YN, Christensen AP, Rosenberg MD, Benedek M, Chen Q, Fink A, Qiu J, Kwapil TR, Kane MJ, et al. , 2018. Robust prediction of individual creative ability from brain functional connectivity. Proceedings of the National Academy of Sciences 115, 1087–1092.

Belkin M, Niyogi P, 2002. Laplacian eigenmaps and spectral techniques for embedding and clustering, in: Advances in neural information processing systems, pp. 585–591.

Belkin M, Niyogi P, Sindhwani V, 2006. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. Journal of machine learning research 7, 2399–2434.

Bilker WB, Hansen JA, Brensinger CM, Richard J, Gur RE, Gur RC, 2012. Development of abbreviated nine-item forms of the raven's standard progressive matrices test. Assessment 19, 354–369. [PubMed: 22605785]

Birn RM, Molloy EK, Patriat R, Parker T, Meier TB, Kirk GR, Nair VA, Meyerand ME, Prabhakaran V, 2013. The effect of scan length on the reliability of resting-state fmri connectivity estimates. Neuroimage 83, 550–558. [PubMed: 23747458]

Bordier C, Nicolini C, Bifone A, 2017. Graph analysis and modularity of brain functional connectivity networks: searching for the optimal threshold. Frontiers in neuroscience 11, 441. [PubMed: 28824364]

Bouckaert RR, Frank E, 2004. Evaluating the replicability of significance tests for comparing learning algorithms, in: Pacific-Asia Conference on Knowledge Discovery and Data Mining, Springer. pp. 3–12.

Bullmore E, Sporns O, 2009. Complex brain networks: graph theoretical analysis of structural and functional systems. Nature reviews neuroscience 10, 186–198. [PubMed: 19190637]

Dadi K, Rahim M, Abraham A, Chyzhyk D, Milham M, Thirion B, Varoquaux G, Initiative ADN, et al. , 2019. Benchmarking functional connectome-based predictive models for resting-state fmri. NeuroImage 192, 115–134. [PubMed: 30836146]

Dong X, Thanou D, Frossard P, Vandergheynst P, 2016. Learning laplacian matrix in smooth graph signal representations. IEEE Transactions on Signal Processing 64, 6160–6173.

Dubois J, Adolphs R, 2016. Building a science of individual differences from fmri. Trends in cognitive sciences 20, 425–443. [PubMed: 27138646]

Egilmez HE, Pavez E, Ortega A, 2017. Graph learning from data under laplacian and structural constraints. IEEE Journal of Selected Topics in Signal Processing 11, 825–841.

Erdős P, Rényi A, 1960. On the evolution of random graphs. Publ. Math. Inst. Hung. Acad. Sci 5, 17–60.

Esfahlani FZ, Sayama H, 2018. A percolation-based thresholding method with applications in functional connectivity analysis, in: International Workshop on Complex Networks, Springer. pp. 221–231.

Finn ES, Scheinost D, Finn DM, Shen X, Papademetris X, Constable RT, 2017. Can brain state be manipulated to emphasize individual differences in functional connectivity? Neuroimage 160, 140–151. [PubMed: 28373122]

Finn ES, Shen X, Scheinost D, Rosenberg MD, Huang J, Chun MM, Papademetris X, Constable RT, 2015. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. Nature neuroscience 18, 1664–1671. [PubMed: 26457551]

Fornito A, Zalesky A, Breakspear M, 2013. Graph analysis of the human connectome: promise, progress, and pitfalls. Neuroimage 80, 426–444. [PubMed: 23643999]

Fox MD, Zhang D, Snyder AZ, Raichle ME, 2009. The global signal and observed anticorrelated resting state brain networks. Journal of neurophysiology 101, 3270–3283. [PubMed: 19339462]

Gao S, Greene AS, Constable RT, Scheinost D, 2019. Combining multiple connectomes improves predictive modeling of phenotypic measures. Neuroimage 201, 116038. [PubMed: 31336188]

Garrison KA, Scheinost D, Finn ES, Shen X, Constable RT, 2015. The (in) stability of functional brain network measures across thresholds. Neuroimage 118, 651–661. [PubMed: 26021218]

Glasser MF, Sotiropoulos SN, Wilson JA, Coalson TS, Fischl B, Andersson JL, Xu J, Jbabdi S, Webster M, Polimeni JR, et al. , 2013. The minimal preprocessing pipelines for the human connectome project. Neuroimage 80, 105–124. [PubMed: 23668970]

van den Heuvel MP, de Lange SC, Zalesky A, Seguin C, Yeo BT, Schmidt R, 2017. Proportional thresholding in resting-state fmri functional connectivity networks and consequences for patient-control connectome studies: Issues and recommendations. Neuroimage 152, 437–449. [PubMed: 28167349]

Holland PW, Laskey KB, Leinhardt S, 1983. Stochastic blockmodels: First steps. Social networks 5, 109–137.

Horien C, Noble S, Finn ES, Shen X, Scheinost D, Constable RT, 2018. Considering factors affecting the connectome-based identification process: Comment on waller et al. NeuroImage 169, 172–175. URL: http://www.sciencedirect.com/science/article/pii/S1053811917310698, doi:https://doi.org/10.1016/j.neuroimage.2017.12.045. [PubMed: 29253655]

Hu C, Cheng L, Sepulcre J, Johnson KA, Fakhri GE, Lu YM, Li Q, 2015. A spectral graph regression model for learning brain connectivity of alzheimer's disease. PloS one 10, e0128136. [PubMed: 26024224]

Huang W, Bolton TA, Medaglia JD, Bassett DS, Ribeiro A, Van De Ville D, 2018. A graph signal processing perspective on functional brain imaging. Proceedings of the IEEE 106, 868–885.

Kalofolias V, 2016. How to learn a graph from smooth signals, in: Artificial Intelligence and Statistics, pp. 920–929.

Komodakis N, Pesquet JC, 2015. Playing with duality: An overview of recent primal? dual approaches for solving large-scale optimization problems. IEEE Signal Processing Magazine 32, 31–54.

Li J, Kong R, Liegeois R, Orban C, Tan Y, Sun N, Holmes AJ, Sabuncu MR, Ge T, Yeo BT, 2019. Global signal regression strengthens association between resting-state functional connectivity and behavior. NeuroImage 196, 126–141. [PubMed: 30974241]

Marques AG, Segarra S, Mateos G, 2020. Signal processing on directed graphs: The role of edge directionality when processing and learning from network data. IEEE Signal Processing Magazine 37, 99–116.

Masuda N, Sakaki M, Ezaki T, Watanabe T, 2018. Clustering coefficients for correlation networks. Frontiers in neuroinformatics 12, 7. [PubMed: 29599714]

Mateos G, Segarra S, Marques AG, Ribeiro A, 2019. Connecting the dots: Identifying network structure via graph signal processing. IEEE Signal Processing Magazine 36, 16–43.

Mishne G, Talmon R, Meir R, Schiller J, Lavzin M, Dubin U, Coifman RR, 2016. Hierarchical coupled-geometry analysis for neuronal structure and activity pattern discovery. IEEE Journal of Selected Topics in Signal Processing 10, 1238–1253. doi:10.1109/JSTSP.2016.2602061.

Mohanty R, Sethares WA, Nair VA, Prabhakaran V, 2020. Rethinking measures of functional connectivity via feature extraction. Scientific reports 10, 1–17. [PubMed: 31913322]

Muldoon SF, Bridgeford EW, Bassett DS, 2016. Small-world propensity and weighted brain networks. Scientific reports 6, 22057. [PubMed: 26912196]

Murphy K, Birn RM, Handwerker DA, Jones TB, Bandettini PA, 2009. The impact of global signal regression on resting state correlations: are anti-correlated networks introduced? Neuroimage 44, 893–905. [PubMed: 18976716]

Murphy K, Fox MD, 2017. Towards a consensus regarding global signal regression for resting state functional connectivity mri. Neuroimage 154, 169–173. [PubMed: 27888059]

Noble S, Scheinost D, Constable RT, 2019. A decade of test-retest reliability of functional connectivity: A systematic review and meta-analysis. NeuroImage 203, 116157. URL: http://www.sciencedirect.com/science/article/pii/S1053811919307487, doi:https://doi.org/10.1016/j.neuroimage.2019.116157. [PubMed: 31494250]

Ortega A, Frossard P, Kovačević J, Moura JM, Vandergheynst P, 2018. Graph signal processing: Overview, challenges, and applications. Proceedings of the IEEE 106, 808–828.

Pennec X, Fillard P, Ayache N, 2006. A riemannian framework for tensor computing. International Journal of computer vision 66, 41–66.

Pervaiz U, Vidaurre D, Woolrich MW, Smith SM, 2020. Optimising network modelling methods for fmri. NeuroImage 211, 116604. [PubMed: 32062083]

Preti MG, Van De Ville D, 2019. Decoupling of brain function from structure reveals regional behavioral specialization in humans. Nature communications 10, 1–7.

Rosenberg MD, Casey B, Holmes AJ, 2018. Prediction complements explanation in understanding the developing brain. Nature communications 9, 1–13.

Rubinov M, Sporns O, 2010. Complex network measures of brain connectivity: uses and interpretations. Neuroimage 52, 1059–1069. [PubMed: 19819337]

Saad ZS, Gotts SJ, Murphy K, Chen G, Jo HJ, Martin A, Cox RW, 2012. Trouble at rest: how correlation patterns and group differences become distorted after global signal regression. Brain connectivity 2, 25–32. [PubMed: 22432927]

Salehi M, Greene AS, Karbasi A, Shen X, Scheinost D, Constable RT, 2020. There is no single functional atlas even for a single individual: Functional parcel definitions change with task. NeuroImage 208, 116366. [PubMed: 31740342]

Salvador R, Suckling J, Coleman MR, Pickard JD, Menon D, Bullmore E, 2005. Neurophysiological architecture of functional magnetic resonance images of human brain. Cerebral cortex 15, 1332–1342. [PubMed: 15635061]

Scheinost D, Noble S, Horien C, Greene AS, Lake EM, Salehi M, Gao S, Shen X, O'Connor D, Barron DS, et al. , 2019. Ten simple rules for predictive modeling of individual differences in neuroimaging. NeuroImage 193, 35–45. [PubMed: 30831310]

Shen X, Finn ES, Scheinost D, Rosenberg MD, Chun MM, Papademetris X, Constable RT, 2017. Using connectome-based predictive modeling to predict individual behavior from brain connectivity. nature protocols 12, 506–518. [PubMed: 28182017]

Shen X, Tokoglu F, Papademetris X, Constable RT, 2013. Groupwise whole-brain parcellation from resting-state fmri data for network node identification. Neuroimage 82, 403–415. [PubMed: 23747961]

Shuman DI, Narang SK, Frossard P, Ortega A, Vandergheynst P, 2013. 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.

Stanley JS, Chi EC, Mishne G, 2020. Multiway graph signal processing on tensors: Integrative analysis of irregular geometries. IEEE Signal Processing Magazine 37, 160–173. doi:10.1109/MSP.2020.3013555. [PubMed: 33473243]

Van Essen DC, Smith SM, Barch DM, Behrens TE, Yacoub E, Ugurbil K, Consortium WMH, et al. , 2013. The wu-minn human connectome project: an overview. Neuroimage 80, 62–79. [PubMed: 23684880]

Varoquaux G, Baronnet F, Kleinschmidt A, Fillard P, Thirion B, 2010. Detection of brain functional-connectivity difference in post-stroke patients using group-level covariance modeling, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 200–208.

Venkatesh M, Jaja J, Pessoa L, 2020. Comparing functional connectivity matrices: A geometry-aware approach applied to participant identification. NeuroImage 207, 116398. [PubMed: 31783117]

Wang J, Zuo X, He Y, 2010. Graph-based network analysis of resting-state functional mri. Frontiers in systems neuroscience 4, 16. [PubMed: 20589099]

Watts DJ, Strogatz SH, 1998. Collective dynamics of ‘small-world’networks. nature 393, 440–442. [PubMed: 9623998]

Zalesky A, Fornito A, Bullmore E, 2012. On the use of correlation as a measure of network connectivity. Neuroimage 60, 2096–2106. [PubMed: 22343126]

Zhu X, Ghahramani Z, Lafferty JD, 2003. Semi-supervised learning using gaussian fields and harmonic functions, in: Proceedings of the 20th International conference on Machine learning (ICML-03), pp. 912–919.

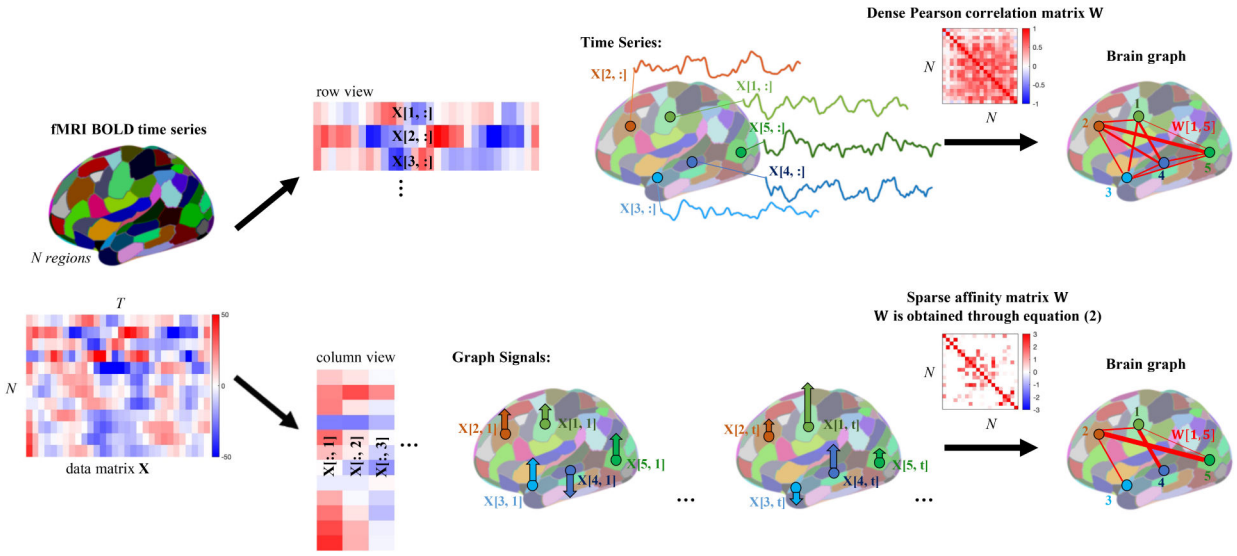


**Figure 1: Estimation of functional connectivity.**

The traditional approach (top) takes a "row-wise" perspective, and calculates pairwise Pearson correlation coefficients between the time-series of the $N$ brain regions (i.e. the rows of the data matrix **X**) to estimate a dense functional connectivity matrix. The smooth graph learning framework (bottom) takes a "column-wise" perspective, treating each time frame of the regional time series (i.e. each column of the data matrix **X**) as a graph signal lying on the brain graph. The graph signal at a given time point $t$ is plotted as arrows at each brain region, where the height (and direction) of the arrow indicate the value (and sign) of the signal at that brain region. This approach learns a sparse connected functional connectivity matrix via smooth graph learning. The two approaches generate brain graphs with different properties.

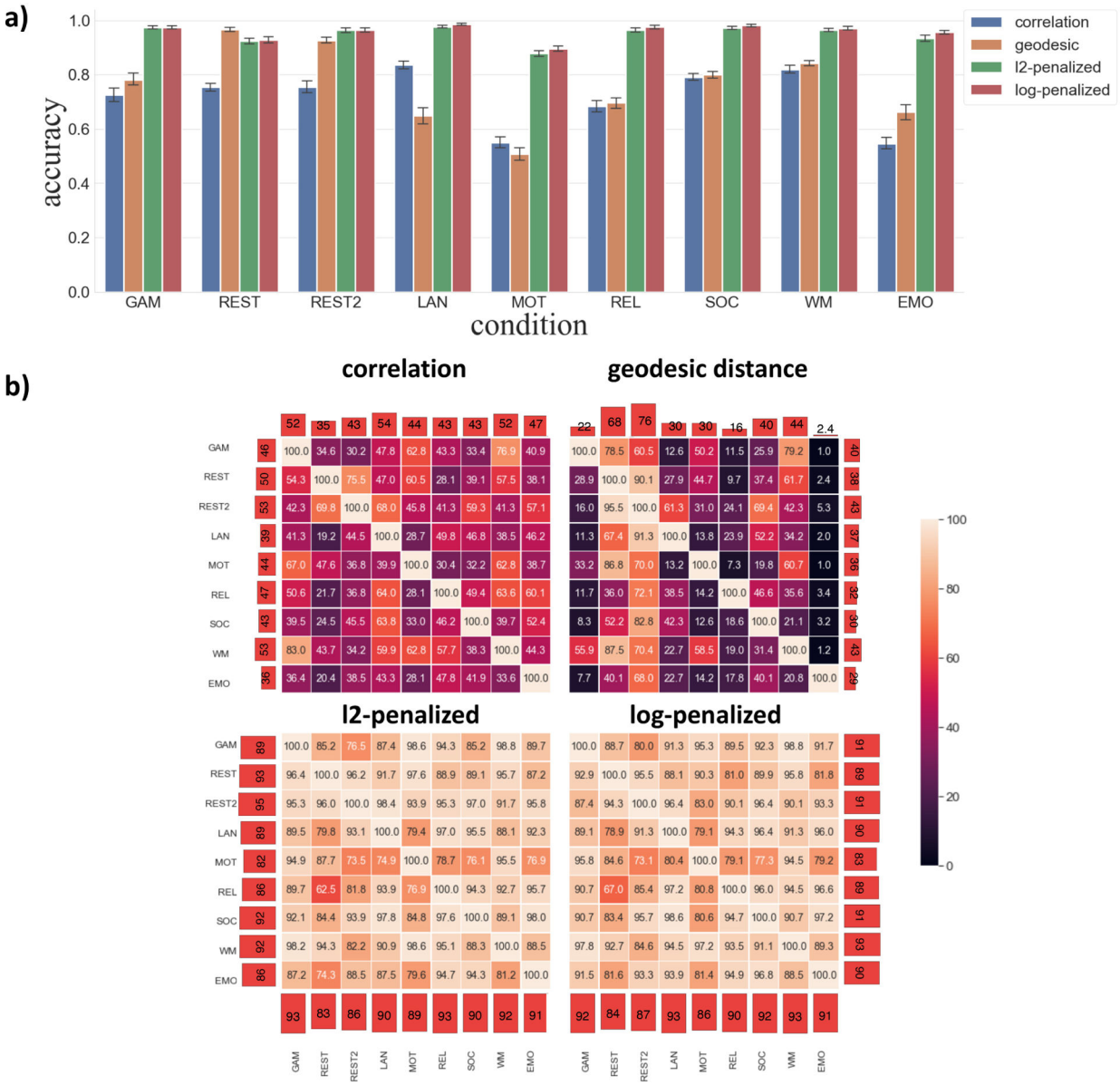


**Figure 2: Fingerprinting accuracy comparison of different methods.**

**a)** Fingerprinting within the same scanning condition. Error bars indicate 95% confidence intervals of the mean across repetitions. Other errors bars in the figure represent the same quantity. **b)** Fingerprinting cross scanning condition. Each subject's scan in the row condition is matched with all the subjects' scans in the column condition. Scanning run corresponds to LR direction was used. $\lambda = 1$ regularization was applied to all the correlation matrices before calculating geodesic distances. The average accuracy of each column or row (not including the diagonal) is plotted as the red bars around the heatmap.

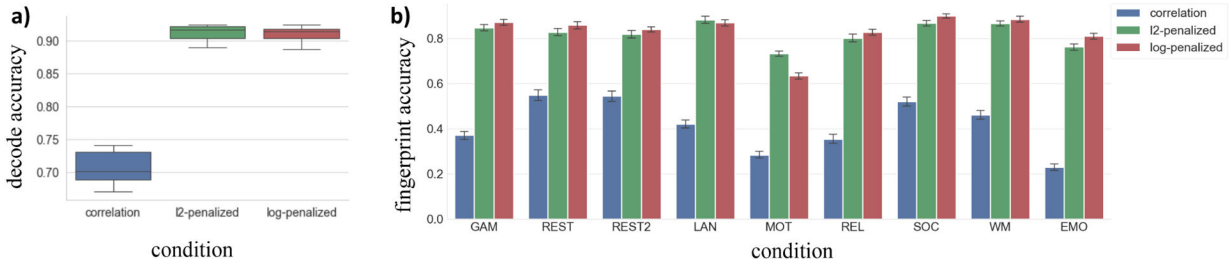


**Figure 3: Task decoding and functional fingerprinting using the node degree.**

**a)** Task decoding accuracy using the node degree. There is no significant difference between the performance of log-regularized graph and $l_2$-regularized graph and they are both significantly better than the correlation graph. **b)** Fingerprinting accuracy using the node degree. In most methods and tasks (except MOT and LAN), the log-penalized SGFC generates the most reliable node degree measures for fingerprinting.

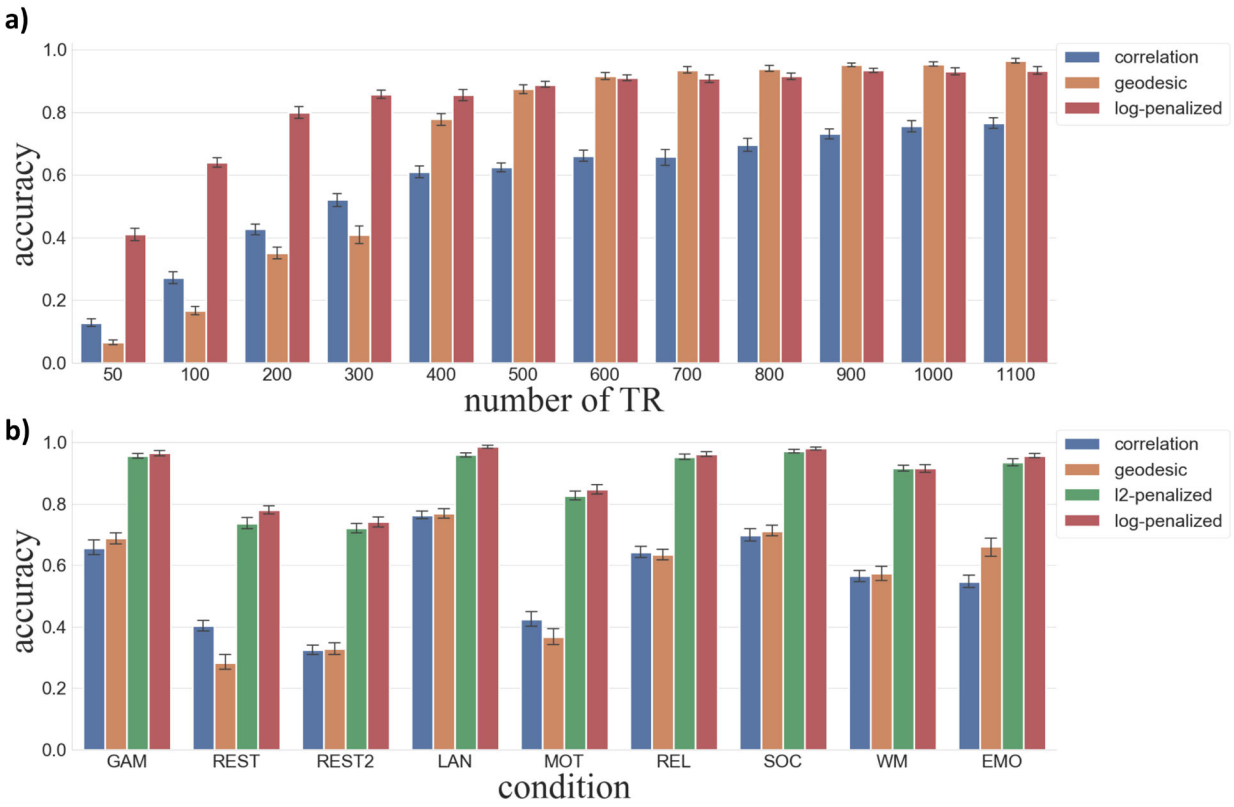


**Figure 4: Influence of scanning length on fingerprinting.**

**a)** Fingerprinting accuracy with varying time length to generate the graph. Rest1 scan's performance are shown. The log-penalized graph was less impacted by the change of scanning time length. **b)** Fingerprinting accuracy in all tasks with shorter (176) scan length.

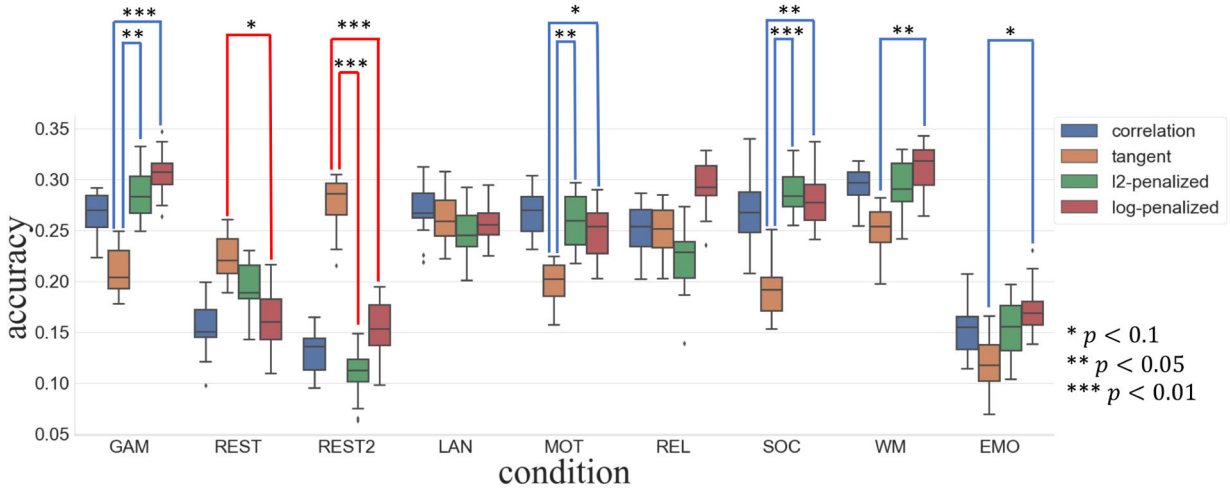


**Figure 5: Predicting fluid intelligence from graphs from different methods.**

Averaged $\sqrt{R^2_{\text{CV}}}$ over 10-fold cross validation is shown. Boxes show $25^{th}$ to $75^{th}$ percentile and whiskers show min and max $\sqrt{R^2_{\text{CV}}}$ over 20 repetitions. Statistically significant differences are labeled by the red or blue brackets where red brackets indicate better CFC-based prediction performance and blue brackets indicate better SGFC-based prediction performance. No statistical comparison is shown within the same graph type (e.g., correlation vs tangent).

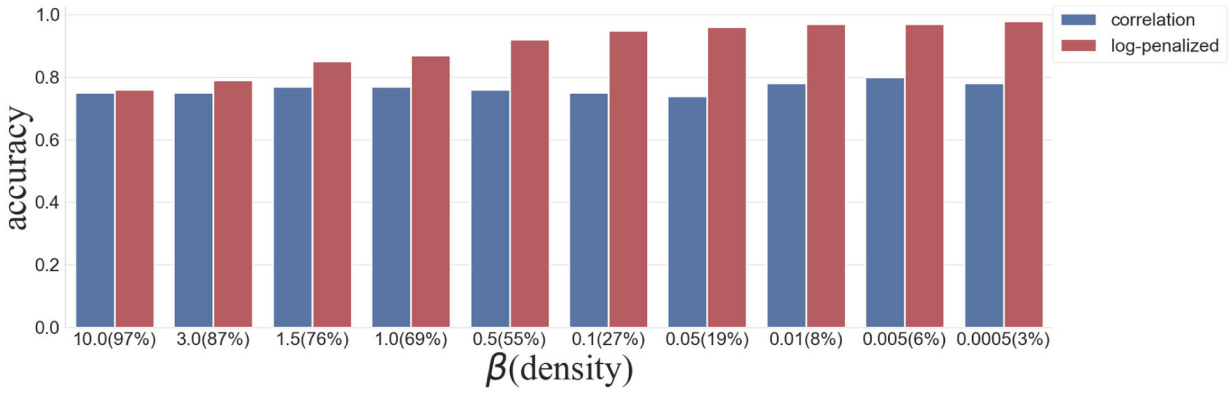


**Figure 6: Regularization strength's influence on fingerprinting accuracy and density of the graph.**

Sparsified CFC with the same density as the SGFC is shown as comparison. Fingerprinting is done on 100 individuals. From left to right, the FCs are sparser.

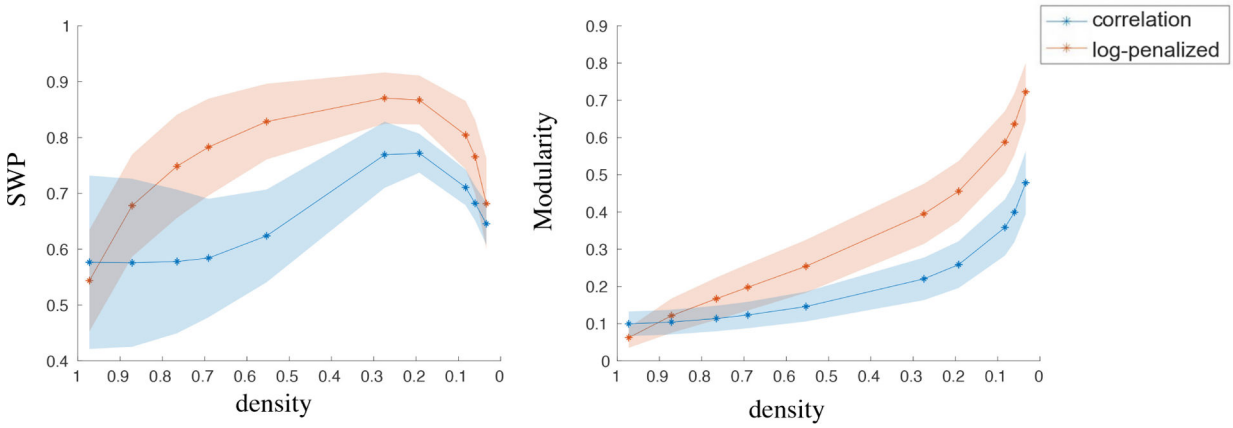


**Figure 7: Small-word propensity and maximum modularity vs. sparsity.**

SWP score (left) and maximum modularity score (right) are shown at different graph densities. Sparsified CFC (blue) with the same density as the log-penalized SGFC (red) is shown as the comparison. Solid line represents the mean and shaded region represents standard deviation calculated across 100 randomly chosen subjects' FCs under resting-state.