A Weighted-ℓ_1, Multi-task, Nonparanormal Graphical Model with Applications to Heterogeneous Brain Connectivity

Chandan Singh [1] **Beilun Wang** [1] **Yanjun Qi** [1]

Abstract

Determining functional brain connectivity is crucial to understanding the brain, particularly the connectivity differences underlying disorders (*e.g.* autism). Recent studies have used Gaussian graphical models (GGM) to learn brain connectivity via statistical dependencies across brain regions from neuroimaging. However the normality assumption GGM relies on is rigid for fMRI samples and these models have not taken into account prior properties like spatial smoothness among brain regions. Therefore, this paper proposes to estimate the connectivity between groups using four elements: sparsity, imposing a spatial prior, multi-task learning, and a nonparanormal assumption by introducing a novel, weighted-ℓ_1, multi-task graphical model (WELM). WELM elegantly combines these elements, in a fast, parallelizable formulation. Here, applications to brain data show state-of-the-art functional connectivity determination in terms of (1) log-likelihood, (2) finding edges that distinguish between groups, and (3) classifying different groups based on their connectivity.

1 Introduction

1.1 Brain Connectivity

Recently, there has been great interest in mapping the interactions between brain regions, a field known as functional connectomics (Smith et al., 2013b). The resulting maps, or connectomes, are fundamental to the study of neuroscience, as having a map of the brain allows for understanding neural pathways and systems (Seung, 2011). Furthermore, these connectomes have immediate applications to pathologists trying to understand the neural characteristics underlying clinical disorders (Uddin et al., 2013).

[1]University of Virginia, Charlottesville, Virginia, USA. Correspondence to: Yanjun Qi <yq2h@virginia.edu>.

Here, we focus on the important problem of estimating brain connectivity for more than one group (*i.e.* a disease group and a control group). Many approaches exist to find this connectivity. One approach utilizes time-series methods in whole-brain linear regression (Baldassano et al., 2012; Greicius et al., 2003). Generally studies use pairwise correlations (*i.e.* the pearson correlation coefficient), between the average activity of different areas as markers of a connection (Rogers et al., 2007). However, though many of these studies properly extract correlations from data, they do not extract conditional correlations.

Finding conditional correlations is a task well-suited to Gaussian graphical models, which explicitly seek conditional correlations between nodes (Koller et al., 2007). From a neuroscience perspective, this yields exactly the brain connections that explain patterns in brain connectivity, discarding connections for spurious correlations. Mathematically, determining functional connectivity amounts to first calculating a covariance matrix (Σ) from the data and then estimating the connectivity graph with the precision matrix ($\Omega = \Sigma^{-1}$). Zeros in Ω correspond to conditionally independent nodes, while non-zero values represent conditional edges (Lauritzen, 1996).

1.2 Weighted-ℓ_1, Multi-task Learning

The model proposed here uses a graphical model to combine four elements crucial to brain connectivity: sparsity, imposing a prior, multi-task learning, and a nonparanormal assumption. No previous work merges all these elements, so this study combines them in a novel, weighted-ℓ_1, multi-task graphical model to robustly estimate Ω for each group (see Fig 1). The weighted-ℓ_1 norm enforces both sparsity and prior knowledge regarding which edges are likely to be zero. For fMRI recordings, a good prior is the spatial distances between connections, since space and energy constraints often prevent the existence of long brain connections (Watts & Strogatz, 1998). Additionally, multi-task learning improves parameter estimation by modeling a shared set of edges between groups. Finally, a nonparanormal assumption extends this model to fit the data observed here much better than the often-used Gaussian assumption. On top of these four major components, WELM is highly

parallelizable.

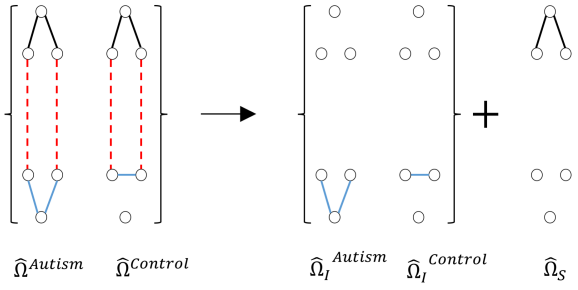

Figure 1. Toy example depicting WELM. Left shows potential edges present in the data and right shows learned edges. Long edges (red) are spatially penalized and discarded, edges that differ between groups (blue) are learned individually, and edges shared between groups (black) are learned in $\widehat{\Omega}_S$.

1.3 Novel Contributions

This study's main contribution is the novel formulation of a weighted-ℓ_1, multi-task graphical model (WELM) which arises naturally from brain connectivity data, but can be applied to a variety of different datasets. Additionally, this study examines a large, resting-state fMRI dataset which serves to compare and validate several recent multi-task learning models. WELM outperforms the state-of-the-art on this dataset in terms of (1) maximizing the log-likelihood of the connectome, (2) picking edges that distinguish between two groups of subjects (Autism and non-Autism), and (3) maximizing classification accuracy between the two groups.

The organization of the paper is as follows: Sec. 2 develops the novel model, Sec. 3 reviews related work, Sec. 4 shows experiments validating WELM, and Sec. 5 explains our conclusions.

2 WELM: A Weighted-ℓ_1 , Multi-task Model

The problem of determining functional brain connectivity concerns how to robustly use the covariance matrix (Σ) to calculate the precision matrix ($\Omega = \Sigma^{-1}$), which represents conditional correlations between brain areas. To do this, WELM takes advantage of 4 properties of brain-imaging data (Sec. 2.1-2.4): sparsity, multi-task learning, a prior, and a nonparanormal assumption. In the work that follows, K is the number of groups, $||\cdot||_1$ is the ℓ_1 norm, $||\cdot||_\infty$ is the ℓ_∞ norm, W is a prior matrix of positive weights, Σ is the covariance matrix, Ω is the precision matrix, and the dot product $(\cdot)$ between two matrices is their elementwise dot product.

2.1 Sparsity

Imposing sparsity is important when determining brain connectivity in order to create interpretable connectomes, as graphs with too many connections yield very litte information. In graphical models, sparsity is generally controlled with an ℓ_1 norm. A simple example of this is the CLIME estimator (Cai et al., 2011), which estimates the precision matrix via constrained-ℓ_1 minimization:

$$\begin{aligned} \widehat{\Omega} &= \underset{\Omega}{\text{argmin}} ||\Omega||_1 \\ \text{subject to: } &||\Sigma\Omega - I||_\infty \leq \lambda, i = 1, ..., K \end{aligned} \tag{1}$$

where λ is a hyperparameter controlling the sparsity of Ω.

2.2 Multi-task learning

Multi-task learning allows the model to extend to estimating more than one group. For example, simply summing the CLIME estimators in Eq. (1) over tasks yields a multi-task formulation:

$$\begin{aligned} \widehat{\Omega}^{(1)}, ..., \widehat{\Omega}^{(K)} &= \underset{\Omega^{(i)}}{\text{argmin}} \sum_i ||\Omega^{(i)}||_1 \\ \text{subject to: } &||\Sigma^{(i)}\Omega^{(i)} - I||_\infty \leq \lambda, i = 1, ..., K. \end{aligned} \tag{2}$$

where $\Omega^{(i)}$ is the precision matrix for a group i.

It is simple to see that multi-task learning will improve performance over single-task models (Evgeniou & Pontil, 2004), especially when there are few samples. However, one must choose between the two multi-task modeling strategies. The first type aims to predict the differences between different groups. We do not use this type as it does not generate full connectomes for each group, as is desired in many neural applications which require understanding whole-brain connectivity patterns. Instead, we share parameters between different groups. Mathematically, we model $\Omega^{(i)}$ as two parts:

$$\Omega^{(i)} = \Omega_I^{(i)} + \Omega_S \tag{3}$$

where $\Omega_I^{(i)}$ is the individual precision matrix for group i and Ω_S is the shared precision matrix between groups. This yields the following formulation:

$$\begin{aligned} \widehat{\Omega}_I^{(1)}, ..., \widehat{\Omega}_I^{(K)}, \widehat{\Omega}_S &= \sum_i \underset{\Omega_I^{(i)}, \Omega_S}{\text{argmin}} ||\Omega_I^{(i)}||_1 + \epsilon K ||\Omega_S||_1 \\ \text{subject to: } &||\Sigma^{(i)}(\Omega_I^{(i)} + \Omega_S) - I||_\infty \leq \lambda, i = 1, ..., K. \end{aligned} \tag{4}$$

2.3 Weighted Prior

Over time, neuroscientists have gathered considerable knowledge regarding the spatial and anatomical priors underlying brain connectivity (*i.e.* short edges and certain

anatomical regions are more likely to be connected). Previous studies (see Sec. 3) enforce these priors via a matrix of weights, W, corresponding to edges. Existing brain connectivity studies enforce a spatial prior using a weighted-ℓ_2 norm (Ng & Abugharbieh, 2011; Grosenick et al., 2011; Baldassano et al., 2012), resulting in the following penalization term: $||W \cdot \Omega||_2$. This weighted-ℓ_2 norm effectively imposes spatial smoothness. Other studies, unrelated to brain connectivity, use a weighted-ℓ_1 norm to enforce a prior in a graphical model: $||W \cdot \Omega||_1$. Here, we opt for an ℓ_1 norm, as it effectively combines the prior with sparsity (Shimamura et al., 2007).

2.4 Nonparanormal Extension.

In addition to Gaussian data, WELM supports nonparanormal Gaussian data (we refer to the nonparanormal version as WELM and the Gaussian version as WELMG). This is implemented by using the Kendall correlation matrix (Σ_N) of the data matrices rather than the sample covariance matrix (Σ). This allows WELM to fit many datasets that violate the often-used Gaussian assumption by fitting a nonparanormal distribution (Liu et al., 2009). Fig 2 shows an example of nonparanormal data that arises in the brain-imaging data studied here.

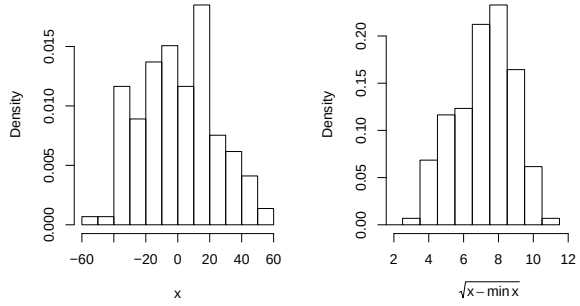

Figure 2. Example of nonparanormal brain-imaging data (described in Sec. 4.1) for one ROI of one subject. After transforming the nonparanormal data (left histogram) via the function $\sqrt{x - \min x}$, it is closer to Gaussian (right histogram).

2.5 WELM: Putting it all together

Combining the elements of sparsity, multi-task learning, and a weighted prior results in the novel formulation of WELM:

$$\widehat{\Omega}_I^{(1)}, ..., \widehat{\Omega}_I^{(K)}, \widehat{\Omega}_S = \sum_i \underset{\Omega_I^{(i)}, \Omega_S}{\text{argmin}} ||W \cdot \Omega_I^{(i)}||_1 + \epsilon K ||W \cdot \Omega_S||_1$$
$$\text{Subject to: } ||\Sigma_N^{(i)}(\Omega_I^{(i)} + \Omega_S) - I||_\infty \leq \lambda, i = 1, ..., K. \quad (5)$$

WELM has three hyperparameters (W, λ, and ϵ) that make it incredibly general. Using a different W can enforce a different prior or change how strongly a prior is enforced. Next, changing the hyperparameter λ controls the total sparsity of the resulting precision matrices. Finally, changing the hyperparameter ϵ allows for controlling how strongly the group penalty is imposed, *i.e.* the relative sparsities between the shared parameters and the individual parameters.

2.6 Optimization

Eq. (5) can be solved in parallel for each column j:

$$\underset{\beta^{(i)}, \beta^s}{\text{argmin}} \sum_i ||W_{,j} \cdot \beta^{(i)}||_1 + \epsilon K ||W_{,j} \cdot \beta^s||_1 \quad (6)$$
$$\text{Subject to: } ||\Sigma^{(i)}(\beta^{(i)} + \beta^s) - e_j||_\infty \leq \lambda, i = 1, \ldots, K$$

where $W_{,j}$ is the j-th column of W, $\beta^{(i)}$ is the j-th column of $\Omega_I^{(i)}$ of i-th graph (we take out the subscript I in $\beta^{(i)}$ to simplify notations), and β^s is the corresponding column in the shared part Ω_S. Simplifying Eq. (6) yields

$$\underset{\theta}{\text{argmin}} ||\theta||_1 \quad (7)$$
$$\text{Subject to: } |\mathbf{A}^{(i)}\theta - b|_\infty \leq c, \ i = 1, \ldots, K$$
$$\text{Where } \mathbf{A}^{(i)} = [0, \ldots, 0, \Sigma^{(i)}, 0, \ldots, 0, \frac{1}{\epsilon K}\Sigma^{(i)}],$$
$$\theta = [W_{,j} \cdot \beta^{(1)^T}, \ldots, W_{,j} \cdot \beta^{(K)^T}, \epsilon K (W_{,j} \cdot \beta^s)^T]^T,$$
$$\boldsymbol{b} = \mathbf{e}_j, c = \lambda$$

To solve Eq. (7), we follow the simplex method (Pang et al., 2014), which is empirically faster than the primal dual interior method (Cormen, 2009). The final formulation becomes

$$\underset{\theta^+, \theta^-}{\text{argmin}} \, \theta^+ + \theta^-$$
$$\text{Subject to :}$$
$$\begin{pmatrix} \mathbf{A}^{(i)} & -\mathbf{A}^{(i)} \\ -\mathbf{A}^{(i)} & \mathbf{A}^{(i)} \end{pmatrix} \begin{pmatrix} \theta^+ \\ \theta^- \end{pmatrix} \leq \begin{pmatrix} c + b \\ c - b \end{pmatrix} \quad (8)$$
$$\begin{pmatrix} \theta^+ \\ \theta^- \end{pmatrix} \geq 0$$

where θ^+ and θ^- refer to the positive and negative parts of θ, respectively.

WELM is summarized in Algorithm 1. Following CLIME, then we apply the same symmetric operators on $\{\Omega^{(i)} = \Omega_S + \Omega_I^{(i)}\}$ obtained from Algorithm 1.

Algorithm 1 A Weighted-ℓ_1, Multi-task Graphical Model (WELM)

Input: Data matrices $\mathbf{X}^{(i)}, ..., \mathbf{X}^{(K)}$, regularization hyperparameter λ, hyperparameter ϵ, and linear programming solver, **LP(.)** which solves Eq. (7)

Output: Shared graph Ω_S and individual graphs $\Omega_I^{(1)}, ..., \Omega_I^{(K)}$

1: **for** $i = 1$ **to** K **do**
2: Initialize $\Sigma^{(i)}$ as the sample cov. matrix of $\mathbf{X}^{(i)}$
3: Initialize $\Omega_I^{(i)} = \mathbf{0}_{p \times p}$
4: Initialize $\mathbf{A}^{(i)} = [0, \ldots, 0, \Sigma^{(i)}, 0, \ldots, 0, \frac{1}{\epsilon K}\Sigma^{(i)}]$
5: **end for**
6: Initialize $\Omega_S = \mathbf{0}_{p \times p}$
7: **for** $j = 1$ **to** p **do**
8: $\theta = \mathbf{LP}(\mathbf{A}^{(i)}, \boldsymbol{b} = \mathbf{e}_j, c = \lambda)$ where $i = 1, \ldots, K$
9: **for** $i = 1$ **to** K **do**
10: ${\Omega_I}_{,j}^{(i)} = \theta_{((i-1)p+1):ip}$
11: **end for**
12: $\Omega_{S,j} = \theta_{(Kp+1):(K+1)p}$
13: **end for**

Parallelization. Each column of WELM can be solved in parallel. Algorithm 1 can be revised into a parallel version by modifying the "for loop" of step 8 in Algorithm 1 into, a parallel for loop over columns.

Convergence. The model's convergence is evident from proofs of the similar formulation of SIMULE (Wang et al., 2016) combined with the fact that the positive weights in W yield a convex norm.

3 Related work

Table 1. Summary of related work.

Method	Conditional Independence	Multi-task	Column-wise Parallelizable	Imposes Prior
WELM	√	√	√	√
CLIME	√	√	×	×
GLASSO	√	√	×	×
JGL	√	√	×	×
DPM	√	√	×	×
Spatial Regularization	×	×	×	√
Weighted-ℓ_1 GGMs	√	×	×	√
sGGGM	√	√	×	×
MNS	√	√	×	×
SIMONE	√	√	×	×
SIMULE	√	√	√	×

3.1 Weighted-ℓ_1 Models

ℓ_1 norms effectively induce sparsity in graphical models (Friedman et al., 2008). Importantly, by weighting the ℓ_1 norm with a prior, the norm can induce sparsity while simultaneously penalizing the selection of certain edges. This differs from the reweighted-ℓ_1 minimization commonly used in compressed sensing, which typically equips a general linear model to robustly impose sparsity with very few samples (Candes et al., 2008). Some recent studies use a weighted-ℓ_1 norm to enforce a prior on a Gaussian graphical model. For example, one model uses reweighted-ℓ_1 norms to maintain sparsity while reducing penalties on nodes with high degree, thus encouraging the appearance of "hub" nodes with a large number of connections (Liu & Ihler, 2011). Spatial penalization, as is done here doesn't necessarily give rise to hub nodes, but rather disincentivizes all nodes from making long connections. Another study uses weighted-ℓ_1 optimization to improve neighborhood selection for gene network estimation (Shimamura et al., 2007). However, current weighted-ℓ_1 studies do not extend to brain connectivity or multi-task learning.

3.2 Brain Connectivity Priors

WELM requires choosing a prior to enforce. For fMRI data, spatial distance is a strong candidate, as spatially distant regions are less likely to be connected in the brain (Watts & Strogatz, 1998; Vértes et al., 2012). Previous studies have utilized spatial regularization, but use it for smoothing rather than feature selection (Ng & Abugharbieh, 2011; Grosenick et al., 2011). Notably, Baldassano et al. use weighted-ℓ_2 regularization to generate ROIs for brain connectivity (2012).

3.3 Multi-task Models

Brain studies Two recent studies apply multi-task learning to brain connectivity determination. MNS (Monti et al., 2015) learns population and subject-specific connectivity in brain networks, but can't effectively discern between two large classes, as is done here. Another recent model, sGGGM (Ng et al., 2013), applies sparsity in a multi-task setting to functional connectivity determination.

Baselines. WELM is compared to the two most-cited graphical models for multi-task learning: JGL (Danaher et al., 2014) and SIMONE (Chiquet et al., 2011), and two more recent models with formulations closer to the one here: CLIME (Cai et al., 2011) and SIMULE (Wang et al., 2016). As a baseline, all models are compared against the extremely popular graphical lasso (GLASSO) (Friedman et al., 2008). Since previous weighted-ℓ_1 GGMs are not multi-task, comparisons to these models are made by lowering the parameter ϵ to eliminate WELM's multi-task component. A final comparison is made to DPM, a model that directly predicts the differences between two classes (Zhao et al., 2014). All studies are summarized in Table 1.

4 Experiments

4.1 Experimental Setup

Data. The data examined here comes from the Autism Brain Imaging Data Exchange (ABIDE) (Di Martino et al., 2014), a publicly available resting-state fMRI dataset. The ABIDE data was released as part of the Human Connectome Project, with the goal of understanding human brain connectivity and how it reflects neural disorders (Van Essen et al., 2013). The ABIDE Data was retrieved from the Preprocessed Connectomes Project (Craddock, 2014), where preprocessing was performed using the Configurable Pipeline for the Analysis of Connectomes (CPAC) (Craddock et al., 2013) without global signal correction or bandpass filtering. After preprocessing with this pipeline, 871 individuals remain (468 diagnosed with autism). Signals for the 160 regions of interest (ROIs) in the often-used Dosenbach Atlas (Dosenbach et al., 2010) are examined. Using an Atlas with more ROIs could yield more detailed information, but this atlas effectively validates WELM, as it divides the whole brain into just 160 ROIs.

Spatial Priors. To weight the ℓ_1 norm, two separate spatial priors were derived from the Dosenbach atlas. The first, referred to as *anatomical*^{i}, gives each ROI one of 40 well-known, anatomic labels (*e.g.* "basal ganglia", "thalamus"). Weights take the low value i if two ROIs have the same label, and the high value $10 - i$ otherwise. The second prior, referred to as *dist*^{i}, sets the weight of each edge to its spatial length, in MNI space[1], raised to the power i.

Classification Setup. Classification is performed using 3-fold cross validation. The data is randomly partitioned into 3 equal sets: a training set, a validate set, and a test set. Each baseline produces $\widehat{\Omega}^{\text{control}}$ and $\widehat{\Omega}^{\text{autism}}$ using the training set. Then, the nonzero edges in the difference graph ($\widehat{\Omega}^{\text{control}}$- $\widehat{\Omega}^{\text{autism}}$) are used for feature selection. Namely, for every edge between ROI x and ROI y, the mean value of x·y over time was selected as a feature. These features are fed to a linear ridge regressor.[2] The regressor is trained via cross-validation using only the validate set. Finally, accuracy for the classifier is reported on the test set. This full process is performed and averaged over 3 folds for each baseline. Importantly, this method evaluates the learned structure of the graphs.

4.2 Experimental Results

Log-likelihood. The most often-used metric for comparing graphs generated by graphical models is the log-likelihood. Here, connectomes are generated for various sparsity levels and their resulting log-likelihoods are plotted (Fig 3A-B).[3] Unsurprisingly, as the number of total edges included in the graphs increases, the log-likelihood of the model increases. WELM ($\epsilon = 1$,W =*dist*^{2}), shown as the dotted red line, outperforms all of the relevant baselines, especially at low sparsities, which are most biophysically plausible. Intuitively, this suggests that WELM is better able to find the few edges that explain correlations than other existing methods. Without the nonparanormal assumption, WELMG only outperforms other baselines at low sparsities. These results suggest that

Table 2. Classification accuracy obtained on ABIDE dataset by various studies. In general, classifiers significantly improve over randomness (50%). These studies employ different preprocessing, training, and validation schemes. Smaller subsets of the data are generally able to achieve better performance, as correlations can be found in these small datasets that overestimate the accuracy of a classifier (Haar et al., 2014).

Study	Method	Total Subjects	Autism Subjects	Control Subjects	Accuracy (%)
Ghiassian et al. 2013	MRMR	1111	403	468	63
Nielsen et al. 2013	GLM	964	538	573	60
Haar et al. 2014	LDA/QDA	906	447	517	∼50
WELM	WELM	871	453	453	**93**
Iidaka 2015	PNN	640	328	312	90
Haar et al. 2014	LDA/QDA	590	295	295	60
Chen et al. 2015	RF	252	126	126	91
Chen et al. 2016	SVM	240	112	128	79
Plitt et al. 2015	L2LR	178	89	89	71

Accuracy AUC. Additionally, WELM outperforms state-of-the-art multi-task graphical models at correctly identifying the edges which distinguish between groups. Fig 3C-D shows the accuracy AUC of a ridge regressor trained on the edges differing between Ω^{Autism} and Ω^{Control} for various sparsity levels. Again, WELM ($\epsilon = 1$,W =*anatomical*^{2}) outperforms all baselines, particularly at low sparsities. After WELM, the next strongest baseline appears to be DPM (pink line in Fig 3D). This is unsurprising, as DPM aims to directly predict the edges that differ between two classes.

ABIDE Classification Accuracy. Utilizing the best results from the accuracy curves (WELM,$\epsilon = 1$,W =*anatomical*^{2},$\lambda = 0.1$) yields a classification accuracy of 93% between the autism and control groups. This outperforms recent studies which perform classification on the dataset (see Table 2). Importantly, studies use different numbers of subjects when classifying. No previous study

[1]MNI space is a coordinate system used to refer to analagous points on different brains.

[2]Ridge regression outperformed lasso and elastic-net regression.

[3]All models were also run with intertwined covariances (covariances generated in a multi-task setting), but the results did not improve and are omitted.

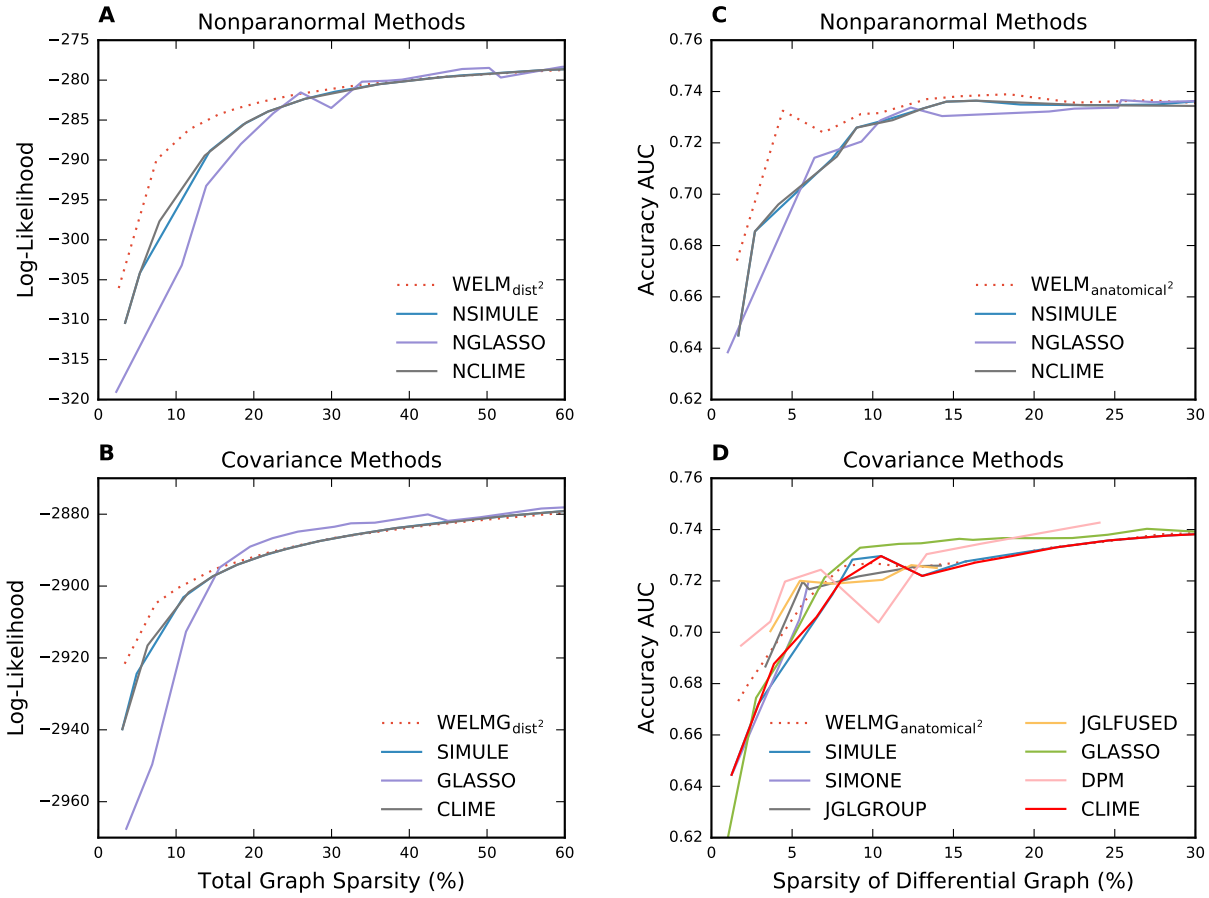

Figure 3. Model performance measured by log-likelihood and classification accuracy. **A** and **B** show the log-likelihood versus number edges included in the model. Note that **A** and **B** are not directly comparable, as **A** uses the nonparanormal versions of the models in **B**. Simone is not shown as it yields a log-likelihood below -10^{-4} and JGL is omitted because it fails to converge when run on the entire dataset. **C** and **D** show the accuracy AUC versus number of edges differing between Ω^{Autism} and Ω^{Control}. Each point represents an AUC (calculated by varying threshold) for a ridge regression classifier. **C** shows the edges using nonparanormal methods while **D** shows the edges using covariance methods.

achieving over 63% classification accuracy uses more than 700 subjects. In contrast, WELM achieves 93% accuracy using 871 subjects. Additionally, WELM produces a more interpretable connectome than many previous studies that aim to classify without learning the connectivity structure (*e.g.* with a probabilistic neural network).

Parameter variation. Different priors and values for ϵ can change the results (see Table 3). The effect of changing the prior seems to have a fairly small effect on the log-likelihood of the model. This is likely because all the priors here penalize picking large edges to some degree (even the anatomical labels tend to be spatially localized). This agrees with the neuroscience perspective that picking short edges should increase the a model's likelihood at low sparsities. Changing ϵ (the strength of the multi-task component) significantly changes the log-likelihood. As ϵ gets larger, thus increasing the importance of the shared parameters between groups, the log-likelihood increases. This emphasizes the importance of the multi-task term of WELM. Since the total number of subjects is limited, strengthening the multi-task component effectively doubles the sample size (including both classes) and allows for better picking edges. The accuracy AUC is fairly constant throughout these parameter sweeps, indicating that any of these priors supplies enough information to properly identify the edges that distinguish between groups.

4.3 Validation

Edge-lengths. Fig 4 shows the effectiveness of the distance prior at reducing average edge length. As the *dist* prior is raised to a higher power, it (thereby increasing the

Table 3. Variations of the prior and multi-task component. All values are generated with WELM in a manner similar to Fig 3.

Changing Prior ($\epsilon = 1$)

Prior	Log-Likelihood Sparsity=7%	Accuracy AUC Sparsity=7%	Log-Likelihood Sparsity=15%	Accuracy AUC Sparsity=15%
dist	-300.08	0.73	-283.70	0.74
dist^2	-296.47	0.72	-282.97	0.74
anatomical^1	-301.68	0.73	-283.92	0.74
anatomical^2	-302.49	0.73	-283.97	0.74

Changing ϵ (Prior = *anatomical*^{2})

Epsilon	Log-Likelihood Sparsity=7%	Accuracy AUC Sparsity=7%	Log-Likelihood Sparsity=15%	Accuracy AUC Sparsity=15%
0.5	-313.03	0.73	-299.71	0.74
0.1	-335.24	0.73	-325.97	0.74
10^{-6}	-342.04	0.72	-333.78	0.74

spread of the weights in the matrix W), the prior is enforced more strongly. This results in a lower average edge length at every sparsity level.

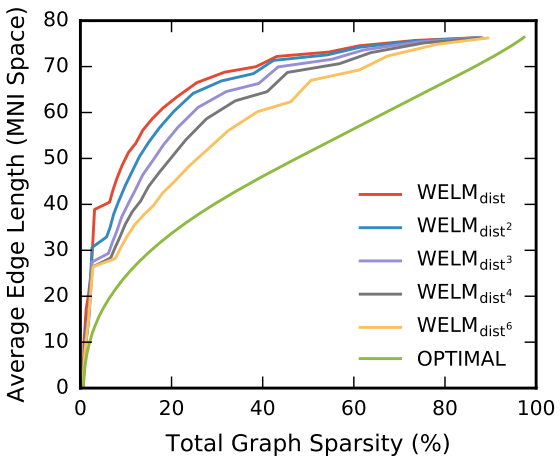

Figure 4. WELM effectively enforces the prior. Average edge-length, as a function of the graph sparsity, decreases as the weight matrix is raised to a larger power (this increases the spread of the distances since all distances are greater than 1). "Optimal" line shows the lowest possible average edge length as a function of graph sparsity.

Connectome. WELM yields different connectomes depending on its hyperparameters. Here, one such connectome, generated from the entire dataset, is shown in Fig 5. A very sparse connectome is shown, in order to be easily interpretable. This connectome, which uses the *dist*^{2} prior ($\epsilon = 1, \lambda = 0.15$), displays the desired selection of short edges. Additionally, many edges are shared between the groups, emphasizing the need for multi-task learning.

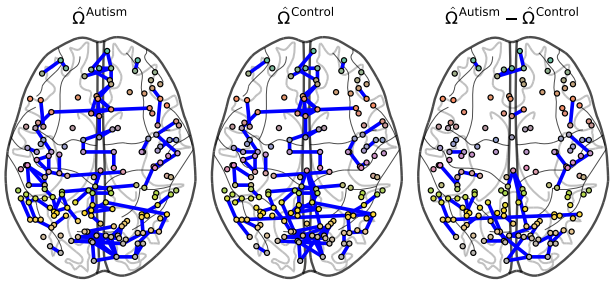

Figure 5. Sparse connectome generated by WELM using the *dist*^{2} prior. The autism graph, control graph, and their difference are shown. The distance prior effectively selects short edges. Many edges are shared between the groups. Visualized with nilearn (Abraham et al., 2014).

Autism-specific areas. Again using the $dist^2$ prior and sweeping over different values of λ yields connectomes at various sparsity levels. At each sparsity level, we calculate the difference between the autism and control group, as seen in Fig. 5. On average, the edges that connect to the following four areas are most affected (in decreasing order): the precuneus, the basal ganglia, the anterior cingulate cortex, and the medial frontal cortex. These results are consistent with the findings of previous brain-imaging studies. Cherkassky et al. find significant underconnectivity to and from the precuneus in autistic subjects (2006). Together, the medial frontal cortex and anterior cingulate cortex have been linked to the the neural basis of social impairments in autism (Mundy, 2003). Finally, the changes in the basal ganglia due to autism reflect changes in gait of autism patients (Rinehart et al., 2006). These results serve as validation for WELM, but note that the analysis here yields no information about the neural connections within these fairly large areas. In order to find more detailed information within one of these areas, one could run WELM using an atlas more refined than the Dosenbach atlas, which divides the brain into just 160 ROIs.

5 Conclusions

WELM effectively enforces a prior while imposing sparsity and taking advantage of multi-task learning. Connectomes generated by WELM selectively highlight connections that are important for distinguishing between autism and control groups; this can help researchers to pinpoint the neural connectivity basis of autism or other disorders with large fMRI datasets (Milham et al., 2012; Smith et al., 2013a).

As brain-imaging datasets become more complex and include more structural data (*e.g.* MRI) coupled with functional data (*e.g.* fMRI), WELM will become increasingly important to neuroscience. This is especially true for studies with small sample sizes, such as task-specific studies,

which require strong priors and multi-task learning in order to robustly determine connectivity (Real et al., 2017). Additionally, as the spatial resolution of fMRI increases, spatial penalization will become more important in constructing accurate ROIs and brain connections (Craddock et al., 2015; Thirion et al., 2014).

Finally, many problems outside of neuroscience can benefit from WELM; it can utilize diverse priors to find conditional independence between nodes in any multi-task setting. Thus, WELM can be readily applied to gene-network estimation (Shimamura et al., 2007), computer vision (where physical distance could be used as a prior in images), speech recognition (where a variety of textual priors are avaiable), and many other problems that currently utilize Gaussian graphical models.

References

Abraham, Alexandre, Pedregosa, Fabian, Eickenberg, Michael, Gervais, Philippe, Muller, Andreas, Kossaifi, Jean, Gramfort, Alexandre, Thirion, Bertrand, and Varoquaux, Gäel. Machine learning for neuroimaging with scikit-learn. *arXiv preprint arXiv:1412.3919*, 2014.

Baldassano, Christopher, Iordan, Marius Cătălin, Beck, Diane M, and Fei-Fei, Li. Voxel-level functional connectivity using spatial regularization. *Neuroimage*, 63(3): 1099–1106, 2012.

Cai, Tony, Liu, Weidong, and Luo, Xi. A constrained 1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106(494):594–607, 2011.

Candes, Emmanuel J, Wakin, Michael B, and Boyd, Stephen P. Enhancing sparsity by reweighted 1 minimization. *Journal of Fourier analysis and applications*, 14(5-6):877–905, 2008.

Chen, Colleen P, Keown, Christopher L, Jahedi, Afrooz, Nair, Aarti, Pflieger, Mark E, Bailey, Barbara A, and Müller, Ralph-Axel. Diagnostic classification of intrinsic functional connectivity highlights somatosensory, default mode, and visual regions in autism. *NeuroImage: Clinical*, 8:238–245, 2015.

Chen, Heng, Duan, Xujun, Liu, Feng, Lu, Fengmei, Ma, Xujing, Zhang, Youxue, Uddin, Lucina Q, and Chen, Huafu. Multivariate classification of autism spectrum disorder using frequency-specific resting-state functional connectivitya multi-center study. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 64:1–9, 2016.

Cherkassky, Vladimir L, Kana, Rajesh K, Keller, Timothy A, and Just, Marcel Adam. Functional connectivity in a baseline resting-state network in autism. *Neuroreport*, 17(16):1687–1690, 2006.

Chiquet, Julien, Grandvalet, Yves, and Ambroise, Christophe. Inferring multiple graphical structures. *Statistics and Computing*, 21(4):537–553, 2011.

Cormen, Thomas H. *Introduction to algorithms*. MIT press, 2009.

Craddock, C, Sikka, S, Cheung, B, Khanuja, R, Ghosh, SS, Yan, C, Li, Q, Lurie, D, Vogelstein, J, Burns, R, et al. Towards automated analysis of connectomes: The configurable pipeline for the analysis of connectomes (c-pac). *Front Neuroinform*, 42, 2013.

Craddock, Cameron. Preprocessed connectomes project: open sharing of preprocessed neuroimaging data and derivatives. In *61st Annual Meeting*. AACAP, 2014.

Craddock, R Cameron, Tungaraza, Rosalia L, and Milham, Michael P. Connectomics and new approaches for analyzing human brain functional connectivity. *GigaScience*, 4(1):13, 2015.

Danaher, Patrick, Wang, Pei, and Witten, Daniela M. The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(2): 373–397, 2014.

Di Martino, Adriana, Yan, Chao-Gan, Li, Qingyang, Denio, Erin, Castellanos, Francisco X, Alaerts, Kaat, Anderson, Jeffrey S, Assaf, Michal, Bookheimer, Susan Y, Dapretto, Mirella, et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular psychiatry*, 19(6):659–667, 2014.

Dosenbach, Nico UF, Nardos, Binyam, Cohen, Alexander L, Fair, Damien A, Power, Jonathan D, Church, Jessica A, Nelson, Steven M, Wig, Gagan S, Vogel, Alecia C, Lessov-Schlaggar, Christina N, et al. Prediction of individual brain maturity using fmri. *Science*, 329 (5997):1358–1361, 2010.

Evgeniou, Theodoros and Pontil, Massimiliano. Regularized multi–task learning. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 109–117. ACM, 2004.

Friedman, Jerome, Hastie, Trevor, and Tibshirani, Robert. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.

Ghiassian, Sina, Greiner, Russell, Jin, Ping, and Brown, M. Learning to classify psychiatric disorders based on fmr images: Autism vs healthy and adhd vs healthy. In *Proceedings of 3rd NIPS Workshop on Machine Learning and Interpretation in NeuroImaging*, 2013.

Greicius, Michael D, Krasnow, Ben, Reiss, Allan L, and Menon, Vinod. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences*, 100 (1):253–258, 2003.

Grosenick, L, Klingenberg, B, Knutson, B, and Taylor, JE. A family of interpretable multivariate models for regression and classification of whole-brain fmri data. *arXiv preprint arXiv:1110.4139*, 2011.

Haar, Shlomi, Berman, Sigal, Behrmann, Marlene, and Dinstein, Ilan. Anatomical abnormalities in autism? *Cerebral Cortex*, pp. bhu242, 2014.

Iidaka, Tetsuya. Resting state functional magnetic resonance imaging and neural network classified autism and control. *Cortex*, 63:55–67, 2015.

Koller, D, Freedman, N, Getoor, L, and Taskar, B. Graphical models in a nutshell in introduction to statistical relational learning. *Stanford*, 2007.

Lauritzen, Steffen L. *Graphical models*, volume 17. Clarendon Press, 1996.

Liu, Han, Lafferty, John, and Wasserman, Larry. The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *Journal of Machine Learning Research*, 10(Oct):2295–2328, 2009.

Liu, Qiang and Ihler, Alexander T. Learning scale free networks by reweighted l1 regularization. In *AISTATS*, pp. 40–48, 2011.

Milham, Michael P, Fair, Damien, Mennes, Maarten, Mostofsky, Stewart HMD, et al. The adhd-200 consortium: a model to advance the translational potential of neuroimaging in clinical neuroscience. *Frontiers in systems neuroscience*, 6:62, 2012.

Monti, Ricardo Pio, Anagnostopoulos, Christoforos, and Montana, Giovanni. Learning population and subject-specific brain connectivity networks via mixed neighborhood selection. *arXiv preprint arXiv:1512.01947*, 2015.

Mundy, Peter. Annotation: The neural basis of social impairments in autism: the role of the dorsal medial-frontal cortex and anterior cingulate system. *Journal of Child Psychology and psychiatry*, 44(6):793–809, 2003.

Ng, Bernard and Abugharbieh, Rafeef. Generalized sparse regularization with application to fmri brain decoding. In *Biennial International Conference on Information Processing in Medical Imaging*, pp. 612–623. Springer, 2011.

Ng, Bernard, Varoquaux, Gaël, Poline, Jean Baptiste, and Thirion, Bertrand. A novel sparse group gaussian graphical model for functional connectivity estimation. In *International Conference on Information Processing in Medical Imaging*, pp. 256–267. Springer, 2013.

Nielsen, Jared A, Zielinski, Brandon A, Fletcher, P Thomas, Alexander, Andrew L, Lange, Nicholas, Bigler, Erin D, Lainhart, Janet E, and Anderson, Jeffrey S. Multisite functional connectivity mri classification of autism: Abide results. *Frontiers in Human Neuroscience*, 7, 2013.

Pang, Haotian, Liu, Han, and Vanderbei, Robert J. The fastclime package for linear programming and large-scale precision matrix estimation in r. *Journal of Machine Learning Research*, 15(1):489–493, 2014.

Plitt, Mark, Barnes, Kelly Anne, and Martin, Alex. Functional connectivity classification of autism identifies highly predictive brain features but falls short of biomarker standards. *NeuroImage: Clinical*, 7:359–366, 2015.

Real, Esteban, Asari, Hiroki, Gollisch, Tim, and Meister, Markus. Neural circuit inference from function to structure. *Current Biology*, 2017.

Rinehart, Nicole J, Tonge, Bruce J, Iansek, Robert, McGinley, Jenny, Brereton, Avril V, Enticott, Peter G, and Bradshaw, John L. Gait function in newly diagnosed children with autism: cerebellar and basal ganglia related motor disorder. *Developmental Medicine & Child Neurology*, 48(10):819–824, 2006.

Rogers, Baxter P, Morgan, Victoria L, Newton, Allen T, and Gore, John C. Assessing functional connectivity in the human brain by fmri. *Magnetic resonance imaging*, 25(10):1347–1357, 2007.

Seung, H Sebastian. Neuroscience: towards functional connectomics. *Nature*, 471(7337):170–172, 2011.

Shimamura, Teppei, Imoto, Seiya, Yamaguchi, Rui, and Miyano, Satoru. Weighted lasso in graphical gaussian modeling for large gene network estimation based on microarray data. *Genome Informatics*, 19:142–153, 2007.

Smith, Stephen M, Beckmann, Christian F, Andersson, Jesper, Auerbach, Edward J, Bijsterbosch, Janine, Douaud, Gwenaëlle, Duff, Eugene, Feinberg, David A, Griffanti, Ludovica, Harms, Michael P, et al. Resting-state fmri in the human connectome project. *Neuroimage*, 80:144–168, 2013a.

Smith, Stephen M, Vidaurre, Diego, Beckmann, Christian F, Glasser, Matthew F, Jenkinson, Mark, Miller, Karla L, Nichols, Thomas E, Robinson, Emma C, Salimi-Khorshidi, Gholamreza, Woolrich, Mark W, et al. Functional connectomics from resting-state fmri. *Trends in cognitive sciences*, 17(12):666–682, 2013b.

Thirion, Bertrand, Varoquaux, Gaël, Dohmatob, Elvis, and Poline, Jean-Baptiste. Which fmri clustering gives good brain parcellations? *Frontiers in neuroscience*, 8:167, 2014.

Uddin, Lucina Q, Supekar, Kaustubh, Lynch, Charles J, Khouzam, Amirah, Phillips, Jennifer, Feinstein, Carl, Ryali, Srikanth, and Menon, Vinod. Salience network–based classification and prediction of symptom severity in children with autism. *JAMA psychiatry*, 70(8):869–879, 2013.

Van Essen, David C, Smith, Stephen M, Barch, Deanna M, Behrens, Timothy EJ, Yacoub, Essa, Ugurbil, Kamil, Consortium, WU-Minn HCP, et al. The wu-minn human connectome project: an overview. *Neuroimage*, 80:62–79, 2013.

Vértes, Petra E, Alexander-Bloch, Aaron F, Gogtay, Nitin, Giedd, Jay N, Rapoport, Judith L, and Bullmore, Edward T. Simple models of human brain functional networks. *Proceedings of the National Academy of Sciences*, 109(15):5868–5873, 2012.

Wang, Beilun, Singh, Ritambhara, and Qi, Yanjun. A constrained l1 minimization approach for estimating multiple sparse gaussian or nonparanormal graphical models. *arXiv preprint arXiv:1605.03468*, 2016.

Watts, Duncan J and Strogatz, Steven H. Collective dynamics of small-worldnetworks. *nature*, 393(6684):440–442, 1998.

Zhao, Sihai Dave, Cai, T Tony, and Li, Hongzhe. Direct estimation of differential networks. *Biometrika*, pp. asu009, 2014.