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

Anonymous Authors[1]

Abstract

This work introduces a novel, weighted-ℓ_1, multi-task graphical model (WELM) for determining functional connectivity. The weighted-ℓ_1 norm induces sparsity while simultaneously imposing a prior that is specified on each edge. Here, applications to brain data with a spatial prior show state-of-the art connectivity determination.

1. Introduction

1.1. Connectivity

Many problems can be formulated as finding an undirected graph over different regions representing conditional correlations in activity between the regions. For example, there has been great interest in mapping the interactions between brain regions, a field known as functional connectomics (Seung, 2011; Smith et al., 2013b). The resulting graphs, or connectomes, have many potential uses such as understanding the differences between those with typically developing brains and those with clinical disorders (Uddin et al., 2013; Milham et al., 2012). With advancements in functional Magnetic Resonance Imaging (fMRI), connectomes are increasingly detailed, creating a need for novel algorithms to interpret this data.

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). Accurately estimating the precision matrix is computationally difficult because of problems with high-order correlations between feature variables (Fan & Liu, 2013). Thus, this problem is well-suited to graphical models, which are able to pick out conditional correlations from high-order data (Koller et al., 2007). Specifically, a sparse Gaussian Graphical Model can be utilized to reliably estimate Ω by limiting the number of nonzero connections in the precision matrix via ℓ_1 penalization (Friedman et al., 2008; Banerjee et al., 2008).

1.2. Weighted-ℓ_1, Multi-task Learning

In several connectivity problems, there is prior knowledge regarding which edges are most likely to be zero. Thus, weights can be associated with edges to penalize their inclusion in the estimated connectivity graph. For fMRI recordings, these weights can be the spatial distances between connections, since the brain favors short connections due to space and energy constraints (Watts & Strogatz, 1998). For problems with few data samples, this can improve robust connectivity estimation.

A common problem in applications is to estimate connectivity for more than one group, *e.g.* a disease group and a control group. In this problem, referred to as multi-task learning, modeling a shared set of parameters between groups improves the overall estimation of the parameters. A variety of recent multi-task graphical models exist, but they have seen limited application to estimating functional connectivity in the brain. Here, several state-of-the-art multi-task models are compared on a real-world dataset. A depiction of weighted-ℓ_1, multi-task learning is shown in Figure 1.

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

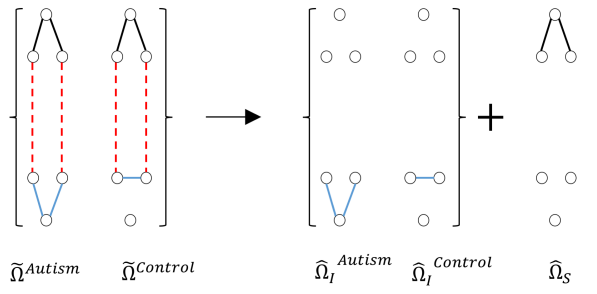

Figure 1. Demonstration of utility of weighted-ℓ_1, multi-task learning. Here, long-distance edges (red dashed lines) are penalized and thus not learned. Black edges are shared between groups and are learned as shared parameters. Blue edges are task-specific and are learned in individual groups.

1.3. Novel Contributions

This study's main contribution is the novel formulation of a weighted-ℓ_1, multi-task graphical model (WELM) which can be applied to a variety of different datasets. Additionally, this study examines a particular resting-state fMRI dataset which serves to compare and validate several recent multi-task learning models. Classification by the novel model here outperforms the state-of-the-art on this dataset.

The organization of the paper is as follows. Section 2 details the novel model, Section 3 reviews previous work, Section 4 shows experiments validating WELM, and Section 5 explains our conclusions.

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

2.1. Notation

In the work that follows, k is the number of groups, $||\cdot||_1$ represents the ℓ_1 norm, $||\cdot||_\infty$ represents the ℓ_∞ norm, W is a prior matrix of positive weights, Σ is the covariance matrix, and we define the dot product $(\cdot)$ between two matrices to be their elementwise dot product. Furthermore, we define $\Omega^{(i)}$ as the precision matrix for a group i and model it as two parts:

$$\Omega^{(i)} = \Omega_I^{(i)} + \Omega_S \tag{1}$$

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

2.2. Formulation

The model starts with a simple ℓ_1 penalty, which yields a precision matrix for each group.

$$\hat{\Omega}^{(1)}, ..., \hat{\Omega}^{(k)} = \sum_i \underset{\Omega^{(i)}}{\text{argmin}} ||\Omega^{(i)}||_1$$
$$\text{Subject to: } ||\Sigma^{(i)}(\Omega^{(i)} - I)||_\infty \leq \lambda, i = 1, ..., k. \tag{2}$$

Next, explicitly modeling the shared and individual parameters of each precision matrix separately yields the formulation of SIMULE (Wang et al., 2016), an effective multi-task model.

$$\hat{\Omega}_I^{(1)}, ..., \hat{\Omega}_I^{(k)}, \hat{\Omega}_S = \sum_i \underset{\Omega_I^{(i)}, \Omega_S}{\text{argmin}} ||\Omega_I^{(i)}||_1 + \epsilon k ||\Omega_S||_1$$
$$\text{Subject to: } ||\Sigma^{(i)}(\Omega_I^{(i)} + \Omega_S) - I||_\infty \leq \lambda, i = 1, ..., k. \tag{3}$$

Finally, the ℓ_1 norm of each precision matrix is weighted by the prior matrix W, resulting in the novel formulation of WELM:

$$\hat{\Omega}_I^{(1)}, ..., \hat{\Omega}_I^{(k)}, \hat{\Omega}_S = \sum_i \underset{\Omega_I^{(i)}, \Omega_S}{\text{argmin}} ||W \cdot \Omega_I^{(i)}||_1 + \epsilon k ||W \cdot \Omega_S||_1$$
$$\text{Subject to: } ||\Sigma^{(i)}(\Omega_I^{(i)} + \Omega_S) - I||_\infty \leq \lambda, i = 1, ..., k. \tag{4}$$

WELM has a flexible prior (W) and two hyperparameters (λ, and ϵ) that allow it to be incredibly flexible. 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, $e.g.$ the relative sparsities between the shared parameters and the individual parameters.

2.3. Model Benefits

The model poses numerous benefits in addition to its ability to generate a sparse, prior-enforced, multi-task graph. The model can support more than just Gaussian data. If the kendall correlation matrix is used in place of the covariance matrix, WELM can learn nonparanormal Gaussian data (referred to as NWELM). Next, the formulation can be solved in parallel. 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. Previous work

3.1. Weighted-ℓ_1 Models

ℓ_1 norms effectively induce sparsity in graphical models (Friedman et al., 2008). Importantly, by weighting the

Table 1. Summary of relevant previous work.

METHOD	CONDITIONAL INDEPENDENCE	MULTI-TASK	COLUMN-WISE PARALLELIZABLE	IMPOSES PRIOR
WELM	√	√	√	√
CLIME	√	√	×	×
GLASSO	√	√	×	×
JGL	√	√	×	×
DPM	√	√	×	×
SPATIAL REGULARIZATION	×	×	×	√
SGGGM	√	√	×	×
MNS	√	√	×	×
SIMONE	√	√	×	×
SIMULE	√	√	√	×

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 while maintaining sparsity. For example, one previous model uses reweighted-ℓ_1 norms to maintain sparsity while reducing penalties on nodes high degree, thus encouraging the appearance of "hub" nodes with high degree (Liu & Ihler, 2011). Another study uses weighted-ℓ_1 optimization for gene network estimation (Shimamura et al., 2007). However, none of these weighted-ℓ_1 studies extends to brain connectivity or the multi-task setting.

3.2. Brain Connectivity Priors

WELM here requires choosing and enforcing a prior for functional brain connectivity. Spatial distance is a strong candidate, as spatial constraints are sufficient to capture a significant amount of connectivity via generative modeling (Vértes et al., 2012). Previous studies have utilized spatial penalization, but use it for smoothing rather than feature selection (Baldassano et al., 2012; Ng & Abugharbieh, 2011; Grosenick et al., 2011). That is, weights on edges are fed into an ℓ_2 norm to induce spatial smoothness.

3.3. Multi-task Models

Two recent studies apply multi-task learning to brain connectivity determination. Mixed Neighborhood Selection (Monti et al., 2015) is an algorithm designed for learning population and subject-specific connectivity in brain networks, but isn't meant for discerning between two large classes, as is done here. Another recent study applies sparsity in a multi-task setting to functional connectivity determination (Ng et al., 2013).

Here, comparisons are made 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). 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. 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. This includes fMRI recordings acquired from several international sites. 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) with no global signal correction and no band-pass filtering. After successful preprocessing with this pipeline, 871 individuals remain. Signals for the 160 ROIs in the Dosenbach Atlas (Dosenbach et al., 2010) are examined.

To weight the ℓ_1 norm, two separate priors were derived from the Dosenbach atlas (Dosenbach et al., 2010). The first gave each ROI one of forty well-known, anatomic labels (e.g. "basal ganglia", "thalamus", "lateral cerebellum"). Weights were then given a low constant value if two ROIs had the same label, and a high constant value if two ROIs had different labels. The second prior was the spatial distance, in MNI space, between the ROIs. The graphical method here is used to generate connectomes at varying levels of sparsity by sweeping over the regularization parameter λ, and the resulting average edge length is plotted against the total number of edges in each graph. When the weight matrix is raised to a power, thus increasing the

spread of the distances (all distances are greater than 1), the prior is enforced more strongly. Experimental designed summarized in Table 2.

Table 2. Experimental design summary.

	DATA	
NUMBER AUTISM SUBJECTS	468	
NUMBER CONTROL SUBJECTS	403	
TOTAL SUBJECTS	871	
PREPROCESSING	CPAC	NO GLOBAL SIGNAL CORRECTION / BAND-PASS FILTERING
ROIS EXAMINED	160	DOSENBACH ATLAS

PRIOR NAME	PRIOR DESCRIPTION	EDGE WEIGHT
DISTANCE	MNI DISTANCE BETWEEN EDGES	CONTINUOUS VALUE
DISTANCE^2	MNI DISTANCE BETWEEN EDGES SQUARED	CONTINUOUS VALUE
ANATOMICAL1	ONE OF 40 LABELS (E.G. "BASAL GANGLIA", "THALAMUS")	0.1 IF SAME LABEL 0.9 OTHERWISE
ANATOMICAL2	ONE OF 40 LABELS (E.G. "BASAL GANGLIA", "THALAMUS")	0.2 IF SAME LABEL 0.8 OTHERWISE

4.2. Model Performance

The most often-used metric for comparing graphs generated by graphical models is the log-likelihood. Here, for each sparsity level, graphs are generated using 871 subjects, and the resulting log-likelihoods are plotted (Figure 2A,B). Unsurprisingly, as the number of total edges included in the graphs increases, the log-likelihood of the model increases. In all cases, an intertwined version of the model did not improve results and thus these results are omitted.

Additionally, the study aims to use classification scores to compare how well multiple sparse graphical model algorithms are able to select edges that are most representative of the dataset. Several recent studies perform classification on the dataset with varying accuracies (summarized in Table S1). Notably, few studies classify all of these subjects in the ABIDE dataset. In fact, no study achieving over 63% classification accuracy uses more than 700 subjects.

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. The graphical model produces graphs for the autism group and the control group using the training set. Then, the difference graph is calculated by subtracting the autism graph from the control graph. Then, the nonzero edges in the difference graph 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 general linear ridge regressor (ridge regression outperformed lasso and elastic net regression). This method evaluates the learned structure of the graphs; in all cases, self-edges are disallowed. 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 graphical model examined here.

The resulting accuracy AUC for each level of sparsity is shown in Figure 2 C,D. The new model outperforms all baselines, particularly at low sparsities which are most biophysically plausible.

4.3. Parameter Searching

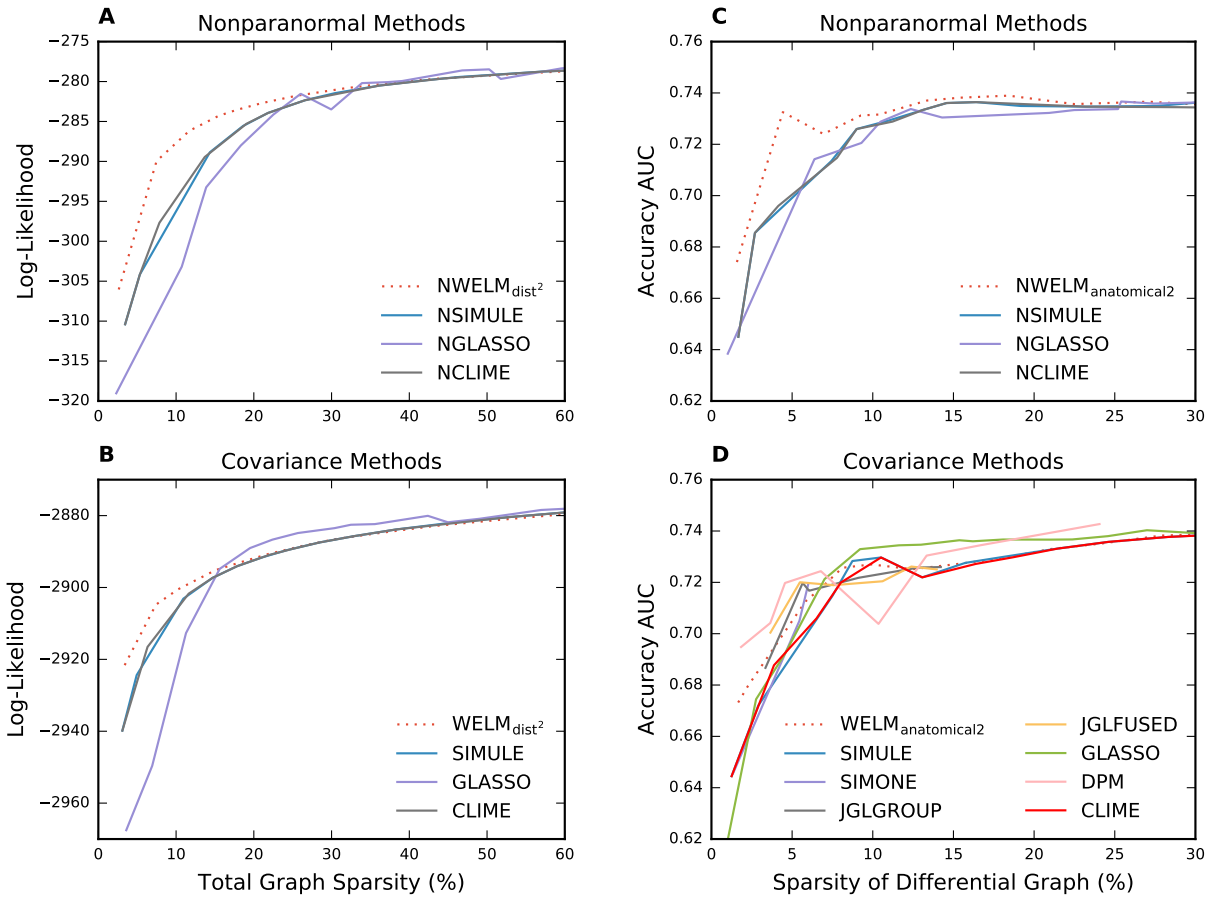

Figure 2. Model performace measured by log-likelihood and classification accuracy. A and B show the log-likelihood versus number edges included in the model. Note that A and B are not directly comparable, as A takes as input the Kendall correlation matrix while B uses the sample covariance matrix. Simone is not shown as it yields a log-likelihood below -10^{-6} 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 included in the model. Edges represent features that were fed into a binary ridge regression classifier. C shows the edges using nonparanormal methods while D shows the edges using covariance methods. The same process was carried out using elastic net and lasso regression, but these classifiers perform more poorly than ridge regression.

Table 3. Variations of prior and epsilon. All methods are nonparanormal.

CHANGING PRIOR ($\epsilon = 1$)

PRIOR NAME	LOG-LIKELIHOOD AUC (SPARSITY $\leq$ 7%)	ACCURACY AUC (SPARSITY $\leq$ 7%)	LOG-LIKELIHOOD AUC (SPARSITY $\leq$ 15%)	ACCURACY AUC (SPARSITY $\leq$ 15%)
DISTANCE	-261.91	0.68	-266.84	0.70
DISTANCE^2	-262.32	0.65	-249.93	0.68
ANATOMICAL1	-102.10	0.68	-269.54	0.72
ANATOMICAL2	-119.54	0.68	-267.53	0.72

CHANGING ϵ (PRIOR = ANATOMICAL2)

EPSILON	LOG-LIKELIHOOD AUC (SPARSITY $\leq$ 7%)	ACCURACY AUC (SPARSITY $\leq$ 7%)	LOG-LIKELIHOOD AUC (SPARSITY $\leq$ 15%)	ACCURACY AUC (SPARSITY $\leq$ 15%)
0.5	-128.84	0.70	-296.49	0.57
0.1	-134.12	0.62	-235.93	0.70
10^{-6}	-136.34	0.63	-240.15	0.69

Figure S1 shows the effectiveness of the distance prior at reducing average edge length.

4.4. Connectome

The new model yields connectomes at varying levels of sparsity. One such overall connectome, generated from the entire dataset, is shown in Figure 3.

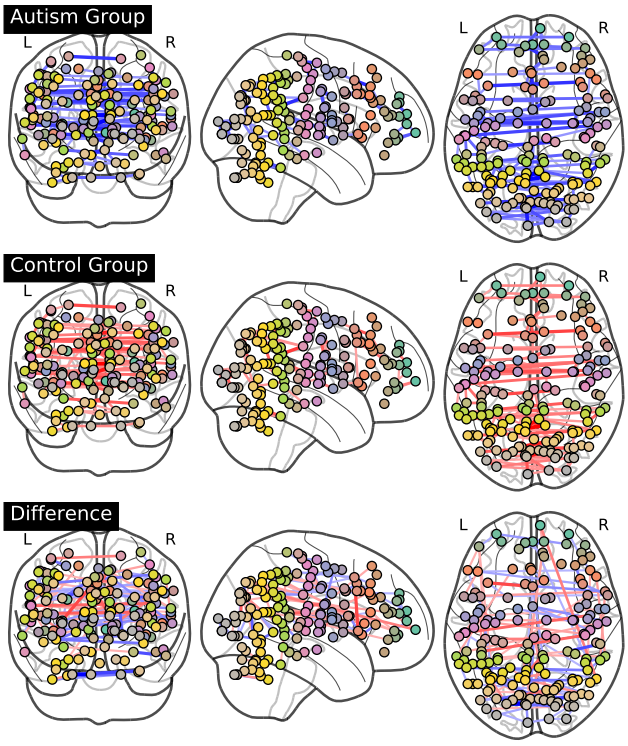

Figure 3. Example connectome generated by NWELM. The context-specific graphs for the autism, control, and their difference (e.g. control subtracted from autism) are shown. Visualization performed with nilearn (Abraham et al., 2014).

5. Conclusions

WELM effectively enforces a prior while imposing sparsity and taking advantage of multi-task learning. The connectomes generated here demonstrate the expected connectivity within anatomic regions, as expected. They also selectively highlight the connections that are important for distinguishing between autism and control gorups; this can help researchers in neuroscience to pinpoint the neural connectivity basis of autism. This model's application to neuroscience is very important and it can be readily applied to several other functional connectivity datasets (Milham et al., 2012; Smith et al., 2013a). As datasets become more complex and include more structural (e.g. MRI data) coupled with functional (e.g. fMRI data), this model will become increasingly important. Studies with small sample sizes, such as task-specific studies require strong priors in order to reasonably compute connectivity (Real et al., 2017).

Additionally, as the spatial resolution of fMRI increases, spatial penalization will become more imporant to construct increasingly accurate ROIs and brain connections. Determining how to construct these ROIs is an active area of research (Craddock et al., 2015; Thirion et al., 2014),

and using different groupings as input to this model can help to evaluate them.

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.

Banerjee, Onureena, Ghaoui, Laurent El, and dAspremont, Alexandre. Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data. *Journal of Machine Learning Research*, 9(Mar): 485–516, 2008.

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.

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

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.

Fan, Jianqing and Liu, Han. Statistical analysis of big data on pharmacogenomics. *Advanced drug delivery reviews*, 65(7):987–1000, 2013.

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.

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, 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.

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. 2013.

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.

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.

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.

Supplement

Table S1. Classification accuracy obtained on ABIDE dataset by previous studies. In general, there is a significant classification improvement over randomness (50%). Many of 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 can overestimate the accuracy of a classifier (Haar et al., 2014).

Study	**This Study**	Ghiassian et al. (2013)	Nielsen et al. (2013)	Haar et al. (2014)	Iidaka (2015)	Haar et al. (2014)	Chen et al. (2015)	Chen et al. (2016)	Plitt et al. (2015)
Total Subjects	871	1111	964	906	640	590	252	240	178
Autism Subjects	403	538	447	453	328	295	126	112	89
Control Subjects	468	573	517	453	312	295	126	128	89
Method	See Model	MRMR selection of HOG features	GLM	LDA and QDA	Probabilistic neural network	LDA and QDA	Random forest	S VMs with slow frequency range	L2LR
Accuracy (%)	**X**	63	60	∼50	90	60	91	79	71

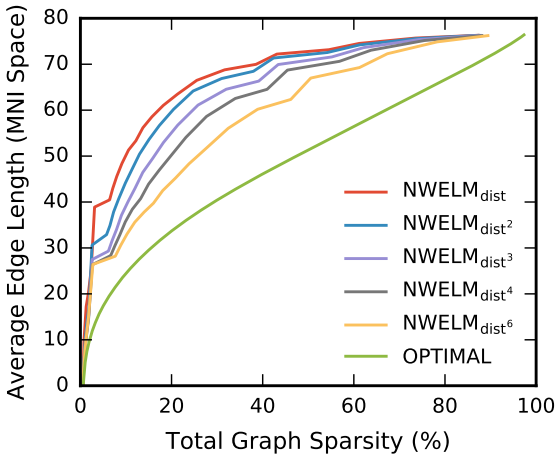

Figure S1. WELM effectively enforces the prior. Average edge-length, as a function of the number of edges for different models decreases as the spread between the prior weights increases. "Optimal" line shows the lowest possible average edge length for different numbers of edges.

Table S2. Range of parameters generating the graphs shown in the Experiments section.

Method	λ_1	λ_2
CLIME	[0.1,0.5]	
GLASSO	[0.1,0.5]	
JGL	[0.1,0.5]	[0.4,0.5]
DPM	[0.1,0.5]	
SIMONE	[0.1,0.5]	
WEIGHTED SIMULE	[0.1,0.5]	[0.4,0.5]
SIMULE	[0.1,0.5]	[0.4,0.5]