

Collective Discovery of Brain Networks with Unknown Groups

Xinyue Liu

Department of Computer Science
Worcester Polytechnic Institute
xliu4@wpi.edu

Xiangnan Kong

Department of Computer Science
Worcester Polytechnic Institute
xkong@wpi.edu

Philip S. Yu

Department of Computer Science
University of Illinois at Chicago
psyu@uic.edu

Abstract—Brain network discovery has attracted much attention in recent years, which aims at inferring a set of cohesive regions (*i.e.*, the network nodes) and the connectivity between these regions (*i.e.*, the network edges) in human brain from the neuroimaging data (*e.g.*, fMRI, PET scans). Previous methods on brain network discovery mainly focus on either estimating the connectivity based on predefined brain regions, or inferring the brain regions and connectivity independently. However, the tasks of discovering brain regions and their connectivity are highly related to each other and should be discovered collectively, instead of independently. In this work, we propose a coherent data-driven method called SGGL (Spectral Group Graphical Lasso) to derive the nodes and edges of a brain network simultaneously. We propose a screening strategy to reduce the time cost of solving the corresponding optimization problem. Extensive experiments are performed on both synthetic data and real data from ADHD-200 project. The results demonstrate the effectiveness of the proposed method.

I. INTRODUCTION

The modern science of graphs and networks has brought significant advances to people’s understanding of complex systems and the interactions within them. One of the most sophisticated systems, human brain, has recently attracted much interest due to the growing availability of high resolution brain imaging data. Brain can be viewed as a network structure, in which neurons are organized into multiple homogeneous regions, and complex interactions exist between neurons within and across different regions. The network representation of human brain as shown in right-hand-side of Figure 1 is useful in many ways. For example, previous studies show that one can employ subgraph selection to build classification models on brain networks to aid disease diagnosis [1]. However, the nodes and edges in brain networks are usually not given and should be derived from the brain imaging data. Thus, the discovery of brain networks is required before one can conduct network analysis on human brain, as illustrated in Figure 1.

Formally, the brain network discovery problem corresponds to infer a set of functionally homogeneous brain regions as the network nodes and the connectivity across these regions as the network edges from the brain imaging data. Previous works in this line usually focus on inferring the edges based on known groups [2] (as shown in Figure 2(a)) or inferring nodes and edges independently (as shown in Figure 2(b)). In Figure 2(a), the network nodes are already given by some predefined brain atlas, which are usually extracted by neurology professionals

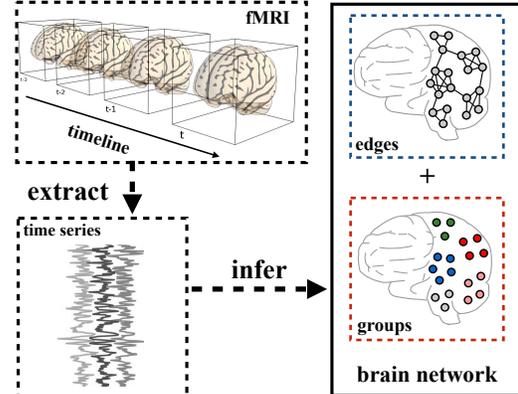


Fig. 1. An illustration of the task of brain network discovery.

using anatomical analysis. Besides, the independent inference framework shown in Figure 2(b) infer the nodes and edges of the brain network separately. On one hand, it derives the groups from the time series data. On the other hand, it infers the edges between these derived groups using the time series data. Although above two methods are straightforward and popular approaches to infer the brain network, they are limited due to following reasons:

- In edge/link discovery, existing methods, such as sparse Gaussian graphical model [3], [4], assume the nodes/groups are given. However, the given groups may be inferred anatomically and contain subregions that are each characterized by different functional connectivity patterns. It may limit the quality and utility of the inferred network.
- In node/group discovery, existing methods, such as k -means [5] and spectral clustering [5], usually infer groups without considering edges/links or assume that the edges/links are given. However, in brain imaging data, the edges/links are usually not given and should be derived from the data. Inferring the group without considering the edges/links may lead to unsatisfactory results.
- In both methods shown in Figure 2(a) and 2(b), once the groups are derived, it is difficult or impossible for one to improve it based on edges/links discovered in latter stages.
- Due to the existence of the log determinant in the group graphical Lasso problem, it is computationally expensive

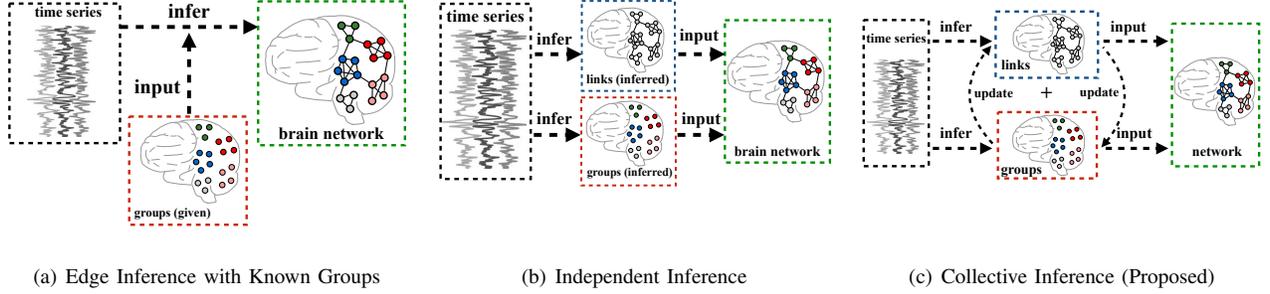


Fig. 2. Comparison of different brain network inference methods. (a) inference with known groups; (b) independent inference; (c) the proposed collective discovery of brain network.

to solve the corresponding maximum likelihood problem.

Thus, a coherent method which can infer both nodes and edges simultaneously is desired to address the aforementioned issues. In this paper, we propose such a method called SGGL, which is illustrated in Figure 2(c), to infer brain network collectively by alternating optimizing the process of node inference and edge inference. In SGGL, the inferred edges (links) are used to further update the groups (nodes). We build an iterative algorithm using alternating optimization. Our proposed method leverages the synergy of group inference and link inference, which can improve the quality of the inferred brain network. The major contributions of this work are summarized as follows:

- We identify the limitations of existing methods for brain network discovery. We show that the node discovery and edges discovery should be considered together and inferred simultaneously.
- We propose a coherent data-driven method that can discover nodes and edges of the brain network collectively.
- We propose a screening strategy to reduce the time cost of solving the group graphical Lasso problem.
- We evaluate the proposed model using both synthetic data and real-world ADHD-200 data. Results demonstrate the effectiveness of the proposed methods.

II. NETWORK DISCOVERY

Throughout this paper, \mathbb{R} denotes the set of all real numbers, \mathbb{R}^n stands for the n -dimensional euclidean space. The set of all $m \times n$ matrices with real entries is denoted as $\mathbb{R}^{m \times n}$. All matrices are written in bold format. All sets are written in calligraphical format. We write $\mathbf{X} \succ 0$ to denote that \mathbf{X} is positive definite. We write $\text{tr}(\cdot)$ to refer the trace of a matrix, which is defined to be the sum of the elements on the main diagonal of the matrix. We use $\det(\mathbf{X})$ to denote the determinant of a real square matrix \mathbf{X} . Suppose \mathbf{X} is a square matrix, $\text{DiagMat}(\mathbf{X})$ denotes the matrix formed by retaining the elements on the main diagonal of \mathbf{X} and set other elements to 0. The important notations used in this paper are also summarized in Table I.

Assume we are given a p -variate normal observations $\mathbf{X} \in \mathbb{R}^{n \times p}$, where n denotes the number of samples and p

denotes the number of variables. The n samples have mean of $\boldsymbol{\mu}$ and covariance of $\boldsymbol{\Sigma}$. Thus, we have the formulation $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. We can assume $\boldsymbol{\mu} = \mathbf{0}$ without losing generality, the problem of estimating the inverse of covariance matrix $\boldsymbol{\Theta} \approx \boldsymbol{\Sigma}^{-1}$ from \mathbf{X} can be cast as follows [3], [4]

$$\arg \min_{\boldsymbol{\Theta} \succ 0} -\log \det \boldsymbol{\Theta} + \text{tr}(\mathbf{S}\boldsymbol{\Theta}) + \lambda \|\boldsymbol{\Theta}\|_1 \quad (1)$$

Eq. (1) is also known as Graphical Lasso (GLasso), where $\mathbf{S} = \frac{1}{n} \mathbf{X}\mathbf{X}^T$ is the empirical covariance matrix and ℓ_1 regularization is posed to encourage sparsity (λ is the parameter to control the sparseness of estimated matrix). Note that Eq. (1) is convex and can be solved by various dual methods [3], [4], [6] and primal methods [7], [8]. In this paper, we consider a variation of GLasso, which desires block-wise sparse estimation of $\boldsymbol{\Sigma}^{-1}$:

$$\arg \min_{\boldsymbol{\Theta} \succ 0} -\log \det \boldsymbol{\Theta} + \text{tr}(\mathbf{S}\boldsymbol{\Theta}) + \sum_{i,j} \lambda_{ij} \|\{\boldsymbol{\Theta}_{G_i, G_j}\}\|_F \quad (2)$$

where G_i is the set of indices for all the elements in group i . And we use $\mathcal{D} = \{G_1, \dots, G_L\}$ to denote the set of all such groups, where L is the total number of groups. Without losing generality, we pose separate regularizer λ_{ij} for each group pair (G_i, G_j) .

A. Link Inference

We first assume that the groups \mathcal{D} are given and to solve Eq. (2). Let $f(\boldsymbol{\Theta}) = -\log \det \boldsymbol{\Theta} + \text{tr}(\mathbf{S}\boldsymbol{\Theta})$. We can replace the non-differentiable regularization term with a linear function. To do this, we introduce an additional variable h_{ij} for each group pair (G_i, G_j) , then Eq. (2) is equivalent to

$$\arg \min_{\boldsymbol{\Theta} \succ 0, \mathbf{h}} f(\boldsymbol{\Theta}) + \sum_{i,j} \lambda_{ij} h_{ij}, \text{ s.t. } h_{ij} \geq \|\boldsymbol{\Theta}_{G_i, G_j}\|_F \quad (3)$$

where $\mathbf{h} = (h_{11}, h_{12}, \dots, h_{LL})$. We note that each constrain in (3) is actually an ℓ_2 norm cone [9]. Thus, Eq. (3) can be solved by the spectral projected-gradient (SPG) method [10]. To solve Eq. (3), SPG needs to compute the minimizer iteratively as follows

$$\boldsymbol{\Theta}_{t+1} = \mathcal{P}_{\mathcal{C}}(\boldsymbol{\Theta}_t - \alpha \nabla f(\boldsymbol{\Theta})) \quad (4)$$

TABLE I
IMPORTANT NOTATIONS.

Symbol	Definition
p	the number of variables or objects
n	the number of samples
$\mathbf{X} \in \mathbb{R}^{n \times p}$	n observations of p -variate Gaussian distribution
$\mathbf{\Sigma} \in \mathbb{R}^{p \times p}$	The covariance of p -variate Gaussian distribution
$\mathbf{S} \in \mathbb{R}^{p \times p}$	The empirical covariance matrix
$\mathbf{\Theta} \in \mathbb{R}^{p \times p}$	The true precision matrix of a network
$\hat{\mathbf{\Theta}} \in \mathbb{R}^{p \times p}$	The estimated precision matrix of a network
L	the number of groups
G_i	the set of indices in group i
$\mathcal{D} = \{G_1, \dots, G_k\}$	the set of group indices sets
$\text{DiagMat}(\cdot)$	the matrix obtain by only retaining the elements on the diagonal of \cdot
λ, λ_{ij}	the regularization terms

where $\tilde{f}(\mathbf{\Theta}) = f(\mathbf{\Theta}) + \sum_{i,j} \lambda_{ij} h_{ij}$, α is the maximal step size selected by non-monotonic Armijo backtracking line search [11] and \mathcal{P}_C is the Euclidean projection onto a closed convex set \mathcal{C} :

$$\mathcal{P}_C(\mathbf{x}) = \arg \min_{\mathbf{y} \in \mathcal{C}} \|\mathbf{x} - \mathbf{y}\|_F \quad (5)$$

where $\mathcal{C} = \{\mathbf{\Theta} | h_{ij} \geq \|\mathbf{\Theta}_{G_i, G_j}\|_F\}$ for our problem. It is usually computationally expensive to solve problem (5) [12]. Fortunately, since groups in Eq. (3) are non-overlapping with each other, so each projection can be solved independently. For any given variable h_{ij} of group pair (G_i, G_j) , the projection has closed form solution [13], [9] as follows

$$\mathcal{P}_C(\mathbf{x}, h_{ij}) = \begin{cases} (\mathbf{x}, h_{ij}), & \text{if } \|\mathbf{x}\|_F \leq h_{ij}, \\ (\hat{\mathbf{x}}, \hat{h}_{ij}), & \text{if } \|\mathbf{x}\|_F > h_{ij}, \\ & \text{if } \|\mathbf{x}\|_F + h_{ij} > 0, \\ (\mathbf{0}, 0), & \text{if } \|\mathbf{x}\|_F + h_{ij} \leq 0. \end{cases} \quad (6)$$

where $\hat{\mathbf{x}} = \frac{\mathbf{x}}{\|\mathbf{x}\|_F} \frac{\|\mathbf{x}\|_F + h_{ij}}{2}$ and $\hat{h} = \frac{\|\mathbf{x}\|_F + h_{ij}}{2}$. Accordingly, the sub-problem Eq. (4) can be solved in $\mathcal{O}(|\mathbf{\Theta}|)$ using Eq. (6). Note that in Eq. (3), $\mathbf{\Theta}$ is constrained to be positive definite, it is proved that one can always find such step size α satisfying the constraint [7]. The algorithm is summarized in Alg. 1.

Algorithm 1 SPG Method for Solving Eq. (3)

Require: $\mathbf{S}, \mathcal{D} = \{G_1, \dots, G_L\}, \boldsymbol{\lambda} = \{\lambda_{ij} | i, j = 1, \dots, L\}, iter_{max}$

- 1: Initialize $\mathbf{\Theta}_0 \leftarrow (\text{DiagMat}(\mathbf{S}))^{-1}$, $iter \leftarrow 0$
- 2: Project the initial estimation $\mathbf{\Theta}_0 \leftarrow \mathcal{P}_C(\mathbf{\Theta}_0)$
- 3: $\tilde{f}_t \leftarrow \tilde{f}(\mathbf{\Theta}_0)$, $\tilde{g}_t \leftarrow \nabla \tilde{f}(\mathbf{\Theta}_0)$
- 4: **repeat**
- 5: Initialize α using Barzilai-Borwein step size
- 6: Choose the α by performing the non-monotonic Armijo backtracking line search
- 7: Compute the new projection $\mathbf{\Theta}_{t+1} \leftarrow \mathcal{P}_C(\mathbf{\Theta}_t - \alpha \tilde{g}_t)$ using Eq. (6).
- 8: Compute the new objective function $\tilde{f}_{t+1} \leftarrow \tilde{f}(\mathbf{\Theta}_{t+1})$
- 9: Compute the new gradient $\tilde{g}_{t+1} \leftarrow \nabla \tilde{f}(\mathbf{\Theta}_{t+1})$
- 10: **until** $iter = iter_{max}$ or convergence
- 11: **Return** $\mathbf{\Theta}_{t+1}$

B. Group Inference

Group inference can be done by applying clustering methods such as k-means directly on the p -variate normal observations \mathbf{X} , but it is difficult for one to refine the groups afterwards. We notice that the output of the Gaussian graphical model can be viewed as a similarity matrix or *affinity matrix*, which can be used as the input for spectral clustering [14]. The similarity matrix is defined as a symmetric matrix $\tilde{\mathbf{S}}$, where $\tilde{S}_{ij} \geq 0$ measures the similarity between variable i and variable j .

Spectral clustering requires all elements in the affinity matrix to be non-negative. Thus, an thresholding can be done after we obtain the estimated precision matrix $\hat{\mathbf{\Theta}}$, one simple thresholding function is

$$\Omega(\hat{\mathbf{\Theta}}) = \text{abs}(\hat{\mathbf{\Theta}}) \quad (7)$$

where $\text{abs}(\cdot)$ produces the absolute value for each element in $\hat{\mathbf{\Theta}}$.

C. Collective Network Discovery

Algorithm 2 SGGL (Spectral Group Graphical Lasso)

Require: $\mathbf{S}, k, \lambda, \boldsymbol{\lambda}, iter_{max}$

- 1: Initialize $\mathbf{\Theta}_0 \leftarrow \text{GLasso}(\mathbf{S}, \lambda)$
 - 2: Initialize $\mathcal{D}_0 \leftarrow \text{SpectralClus}(\Omega(\mathbf{\Theta}_0), k)$
 - 3: Initialize $iter \leftarrow 0$
 - 4: **repeat**
 - 5: Update $\mathbf{\Theta}_{t+1} \leftarrow \text{SPG}(\mathbf{S}, \mathcal{D}_t, \boldsymbol{\lambda})$ using Algorithm 1 with screening
 - 6: Update $\mathcal{D}_{t+1} \leftarrow \text{SpectralClus}(\Omega(\mathbf{\Theta}_{t+1}), k)$
 - 7: **until** $iter = iter_{max}$ or convergence
 - 8: **Return** $\mathbf{\Theta}_{t+1}, \mathcal{D}_{t+1}$
-

In this section, we present the proposed SGGL method. Without the group information given a priori, we first utilize the Gaussian graphical model to infer an initial network in which each variable itself is a group. We do this by assuming the network is sparse, so it is equivalent to solve Eq. (1). Given the empirical covariance matrix \mathbf{S} and the regularization parameter λ_1 , the estimated precision matrix $\mathbf{\Theta}_0$ can be derived by applying existing GLasso algorithm. In the next step, we first transform $\mathbf{\Theta}_0$ using thresholding function $\Omega(\cdot)$,

and then apply widely used spectral clustering on $\Omega(\Theta_0)$ to obtain the initial group \mathcal{D}_0 . In the iterative part, SGGL first updates the estimated precision matrix by solving Eq. 2 with group information given by \mathcal{D}_0 . Then it updates the estimated group by applying spectral clustering on the updated precision matrix. This process is repeated until the algorithm converges or the maximum number of iterations is reached. Generally, one can set $\lambda_{ij} = \lambda$ to assign the same regularization term for every group pairs in \mathcal{D} . However, one can penalize each group separately by assigning a regularization vector $\lambda = \{\lambda_{ij}|i, j = 1, \dots, L\}$ to the SGGL method. This can be useful in various circumstances, for example, one may want to assign different magnitude of regularization on the off-diagonal blocks and diagonal blocks to encourage different sparsity between groups and within. The convergence check can be performed in several ways. We choose to compute the update of the precision matrix in each iteration as

$$u_p = \|\Theta_{t+1} - \Theta_t\|_F \quad (8)$$

In our implementation, the iterative procedure stops when $u_p \leq 10^{-5}$. The SGGL algorithm is summarized in Alg. 2.

D. Screening

Due to the existence of the log determinant, it is computationally expensive to solve the penalized log likelihood model in Eq. (2) or Eq. (3) by applying Algorithm 1 directly. The screening strategy has commonly been used to reduce the size of optimization problems as well as the computational time for solving the problems. In [15], Kolar *et al.* proposed a sufficient condition for the solution of (2) to be block diagonal:

$$\|\mathbf{S}_{G_i, G_j}\|_F \leq \lambda, \forall G_i \subseteq C_l^{(\lambda)}, G_j \subseteq C_{l'}^{(\lambda)}, l \neq l', \quad (9)$$

where $C_1^{(\lambda)}, \dots, C_N^{(\lambda)}$ are the block structures of the optimal solution to (2) with regularization term equals to λ , \mathbf{S}_{G_i, G_j} denotes the sub-matrix of the empirical covariance matrix \mathbf{S} , and G_i is the i -th group of features. We utilize the sufficient condition to derive the screening rule for (2) as follows. We first compute the empirical covariance matrix \mathbf{S} from the data. Then we define the group-based F-norm matrix $\mathbf{S}^{(F)}$, where each entry $\mathbf{S}_{ij}^{(F)} = \|\mathbf{S}_{G_i, G_j}\|_F$. Given $\lambda = \{\lambda_{ij}|i, j = 1, \dots, L\}$, we perform a thresholding on the entries of $\mathbf{S}^{(F)}$ as

$$\mathbf{E}_{ij}^{(\lambda)} = \begin{cases} 1, & \text{if } |\mathbf{S}_{ij}^{(F)}| > \lambda_{ij} \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

to obtain the graph edge skeleton $\mathbf{E}^{(\lambda)}$. Similarly, $\mathbf{E}^{(\lambda)}$ defines a symmetric graph on nodes $\mathcal{V}' = \{1, \dots, L\}$ given by $\mathcal{G}^{(\lambda)} = (\mathcal{V}', \mathbf{E}^{(\lambda)})$, where each node actually stands for a group of features. In order to find the connected components of $\mathcal{G}^{(\lambda)}$, we apply DFS algorithm to decompose it into N' isolated parts: $\mathcal{G}^{(\lambda)} = \bigcup_{\ell=1}^{N'} \mathcal{G}_\ell^{(\lambda)}$. Since each component $\mathcal{G}_\ell^{(\lambda)}$ is corresponding to a block structure $C_l^{(\lambda)}$ in the optimal solution Θ^λ of (2), we can conclude that $\Theta_{C_l^{(\lambda)}, C_{l'}^{(\lambda)}} = \mathbf{0}$ for all $l \neq l'$. Thus, one can instead solve (2) by deriving $\Theta_{C_1^{(\lambda)}, C_1^{(\lambda)}}, \dots, \Theta_{C_{N'}^{(\lambda)}, C_{N'}^{(\lambda)}}$ independently, which are much smaller problems compared to the original one.

III. EXPERIMENTAL EVALUATION

A. Synthetic Data with Ground-Truth

We first evaluate our model using synthetic data, where ground-truth is available. We follow the approach in [16] to generate the synthetic precision matrix. Particularly, we first generate block diagonal matrix Θ with p features and L diagonal blocks (groups), each block Θ_{G_i, G_i} is of size $p/L \times p/L$ and has random sparsity structures. We control the density of each block on diagonal to be 0.7 ± 0.1 , and then we add off-diagonal blocks to Θ as follows to simulate the interconnections between groups. We first select $\beta L(L-1)/2$ pairs of groups randomly, where β is the parameter of interconnection density. We control the density of each off-diagonal block to be about 0.3 ± 0.05 . Given the precision matrix Θ , we draw r samples from the Gaussian distribution to compute the empirical covariance matrix.

Three synthetic precision matrices and empirical covariance matrices are generated using the parameters as follows

- **Dataset 1(Weak Interconnections)**: $p = 50, L = 5, \beta = 0, r = 40, 60, \dots, 200$.
- **Dataset 2(Moderate Interconnections)**: $p = 50, L = 5, \beta = 0.15, r = 40, 60, \dots, 200$.
- **Dataset 3(Strong Interconnections)**: $p = 50, L = 5, \beta = 0.45, r = 40, 60, \dots, 200$.

In all three datasets, the same group index is used:

$$G_i = \{10i - 9, \dots, 10i\}, i = 1, \dots, 5 \quad (11)$$

For each ground-truth precision matrix, we randomly draw r samples 20 times, where r varies from 40 to 200 with a step size of 20.

On edge detection, we compare SGGL with two baseline methods:

- **GLasso**: The graphical lasso method [3].
- **k-means + GGL**: A pipeline method, which first performs k -means clustering on the time series data to derive the groups, then applies group graphical Lasso with the inferred groups to obtain the estimated precision matrix.

On group detection, we compare SGGL with two clustering methods: k -means and spectral clustering. Both methods are widely used baselines in the literature [5].

To evaluate the quality of edge detection. We follow [16] to define the accuracy and F1 score of edge detection as $Accuracy = \frac{n_d}{n_g}, F1 = \frac{2n_d^2}{n_a n_d + n_g n_d}$ where n_d is the number of true edges detected by the algorithm, n_g is the number of true edges and n_a is the total number of edges detected. We control the number of edges detected by all compared methods to be similar.

To evaluate the quality of group inference, we follow [17] to use normalized mutual information (NMI) score and purity score. Higher NMI score or higher purity score indicates better quality of group detection.

Figure 3 shows the comparison between SGGL and graphical lasso in terms of edge detection. The left column shows the sparsity patterns of ground truth of Dataset 1 and Dataset 2

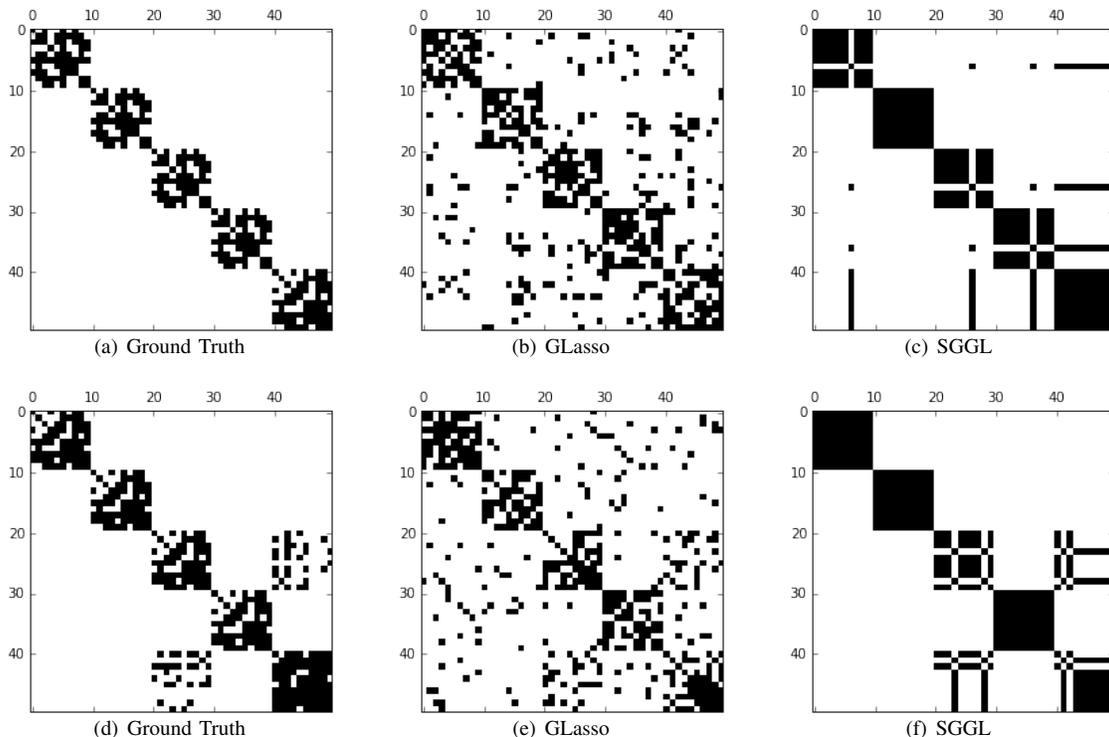


Fig. 3. Comparison between SGGL and Graphical Lasso in terms of edge detection. Left: the ground truth precision matrices; middle: the precision matrices estimated by Graphical Lasso; right: the precision matrix estimated by SGGL. Upper: the precision matrices inferred for Dataset 1; Bottom: the precision matrices inferred for Dataset 2

described above. The middle column denotes the sparsity patterns detected by graphical lasso. The right column presents the sparsity patterns detected by the proposed SGGL method. For each dataset, λ is adjusted to make sure that both graphical lasso and SGGL generates similar number of nonzero entries in the estimated precision matrix. We can observe that compared to the graphical lasso method, the proposed SGGL method generates a more interpretable results with much less noises. These results demonstrate that SGGL outperforms graphical lasso in terms of detecting true edges in the precision matrix.

Figure 4 shows the results of the accuracy and F1 scores on edge detection. We can observe that in terms of accuracy, the proposed SGGL method outperform other two baselines on all datasets consistently, and in terms of F1 score, SGGL achieves competitive or better performance compared to the baselines. SGGL does especially well on data set 3, when the connection is denser. These results demonstrate that SGGL outperforms the compared methods in terms of edge detection. Figure 5 shows the comparison of SGGL, k -means and spectral clustering in terms of NMI score and purity score. We can observe from Figure 5 that for dataset 2 and dataset 3, SGGL achieves higher or equal NMI score and purity score than two baselines consistently. For dataset 1 with weak interconnections, SGGL performs slightly worse than k -means, but the results are very close. It indicates that SGGL may not have much advantages compared to k -means when

the components in the network are very isolated, but such case is relatively rare in real-world dataset such as social networks and brain networks. Thus, these results demonstrate that SGGL outperforms state-of-art clustering baselines in terms of group detection.

B. Real World ADHD-200 Data

We evaluate our method using real world fMRI data from ADHD-200 project¹. ADHD (Attention Deficit Hyperactivity Disorder) is a chronic condition that happens on more than 5% - 10% of school-age children. The annual costs on treating ADHD exceeds 36 billion in the United States. The dataset we used is distributed by nilearn². There are in total 40 subjects in the dataset, 20 of which are labeled as ADHD, and the other 20 subjects are labeled as TDC. The rsfMRI scan of each subject in the dataset is a series of snapshots of 3D brain images of size $61 \times 73 \times 61$ over ~ 180 time steps.

The first set of experiments are performed on the default mode network (DMN) of human brain. DMN is a network of interacting brain regions known to have activity highly correlated with each other and distinct from other networks in the brain. In details, we extract time series from four pre-defined regions in DMN: Posterior Cingulate Cortex, Left Temporoparietal junction, Right Temporoparietal junction and Medial prefrontal cortex. We regard each DMN region as a

¹http://fcon_1000.projects.nitrc.org/indi/adhd200

²<http://nilearn.github.io/>

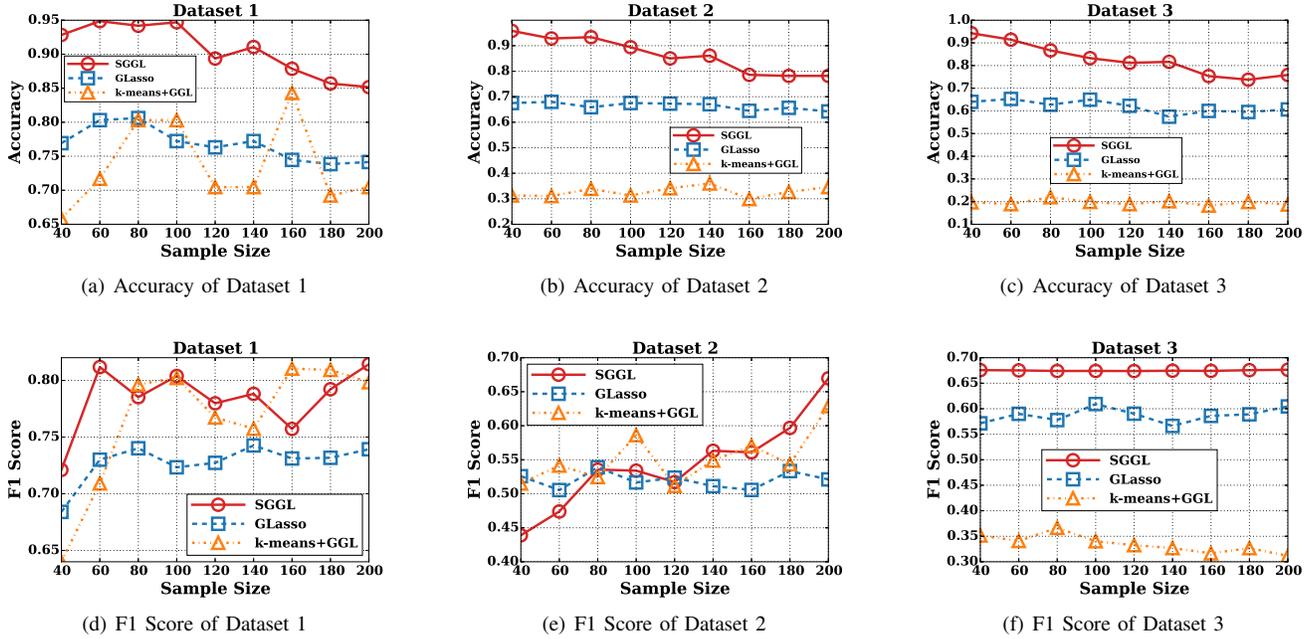


Fig. 4. Comparison in terms of accuracy and F1 score on detecting true edges. Sample size varies from 40 to 200 with a step size of 20.

sphere and the radius is set as 8 mm for all regions. We apply SGGL, k -Means and Spectral clustering on the time series data without the group information, to see how well can each method recover the known groups. In Figure 6, we show the comparison of clustering performance in term of NMI score. We compare the proposed SGGL method with k -Means and spectral clustering using RBF affinity. One can observe that SGGL achieves better NMI score compared to the other two methods, indicating SGGL improves group detection in human brain. In Figure 7, we present the connectivity pattern of ADHD group derived by SGGL method. The network discovered for TDC group is very similar, indicating that the strength of DMN is not affected by ADHD, which may be due to the simplicity of the DMN itself. From Figure 7, we can observe that all four regions in DMN are strongly connected to each other, which is consistent with the essences of DMN. Thus, we can see the effectiveness of the proposed SGGL method in discovering the groups and connectivity of human brain.

The second set of experiments aims at discovering the networks of the entire human brain. We use the AAL brain-shaped mask (provided by neurology professionals) to extract the voxels that are considered parts of the brain. We follow [18] to use a middle slice of these scan for the ease of presentation. Each of the scans can be represented by about 3200 voxels. In Figure 8, we present the comparison of the group inference results between SGGL and spectral clustering. Figure 8(a) and 8(b) show the groups inferred by the proposed SGGL for TDC and ADHD respectively; Figure 8(c) and 8(d) show the groups inferred by spectral clustering. We can observe that the results of spectral clustering are very scattered and it is difficult to capture the difference between the results

of TDC and ADHD from the figures. However, the proposed SGGL method presents a much interpretable results compared to the ones of spectral clustering.

In Figure 9, we show the comparison of time cost on running SGGL with and without using screening strategy. We observe that the proposed screening strategy can achieve about 40% time gain on ADHD-200 dataset, while time cost of screening itself is negligible.

IV. RELATED WORKS

This work is related to brain parcellation and brain connectivity analysis, we discuss them briefly in this section.

It is an important and challenging task to infer the brain parcellation. Early work in this direction have focused on anatomical atlases. Although one can learn much from these anatomical brain mapping, no functional or structural connectivity information was used to construct them. Thus, it is highly possible that regions in anatomical atlases contain subregions characterized by different functional and structural patterns. Recent studies in brain parcellation have mainly focused on data-driven methods, which aimed at obtaining brain mapping directly from the neuroimaging data. [19] have studied the problem of identifying brain regions that are related to Alzheimer's disease from multi-modality neuroimaging data. [20] have focused on parcellating the brain into a set of regions that are functionally homogeneous. Specifically, the voxels within a region should share similar time courses or generating similar functional connectivity patterns [21]. On the other hand, researchers also have focused on preserving the spatial contiguity of the parcellated regions to ensure the interpretability of them [22], [23], [5]. [24] have summarized the commonly used clustering techniques for inferring brain mapping and makes comparison among them with various

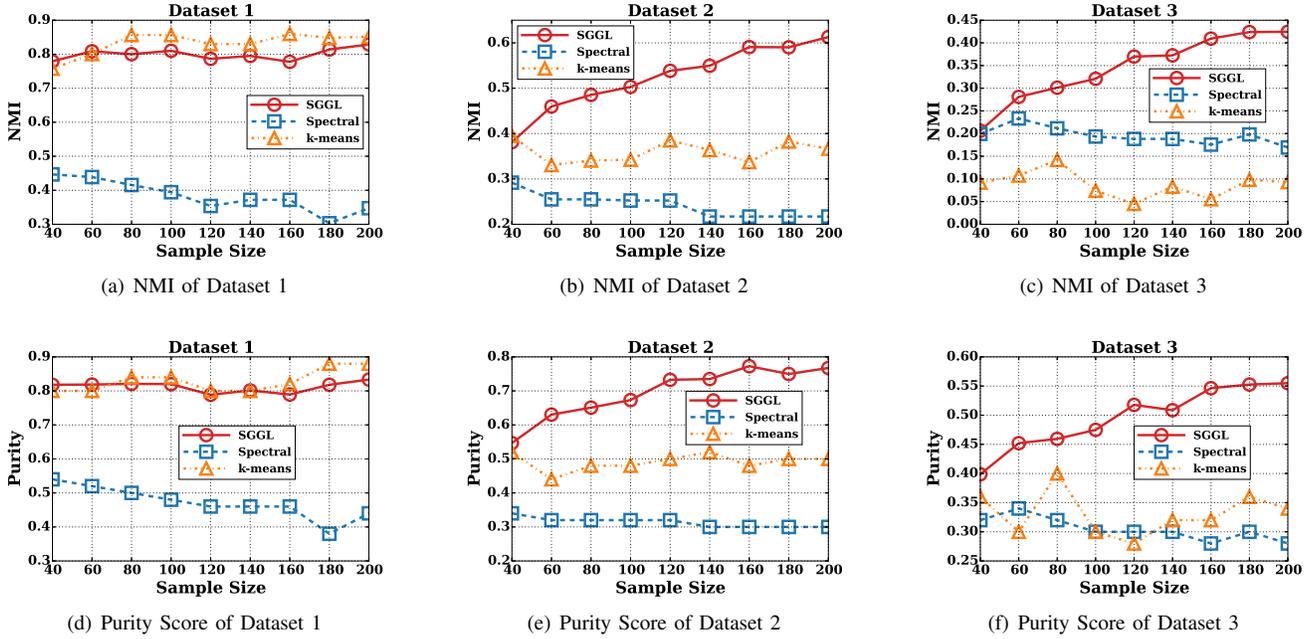


Fig. 5. Comparison in term of NMI and Purity on detecting groups. Sample size varies from 40 to 200 with a step size of 20.

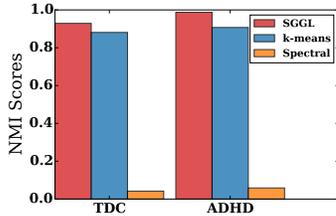


Fig. 6. Comparison of NMI scores on the DMN (Default Mode Network) of ADHD-200 Data.

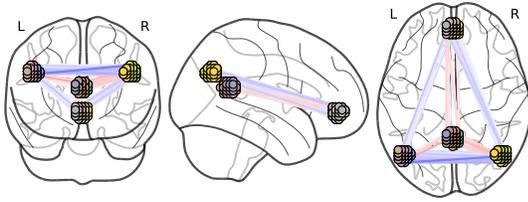


Fig. 7. The connectivity of DMN of ADHD group discovered by SGGL. All regions in the DMN are strongly connected to each other, which is consistent with the essence of DMN.

evaluation metrics. Especially, [5] have proposed to employ spatially constrained spectral clustering to build brain atlas from fMRI data.

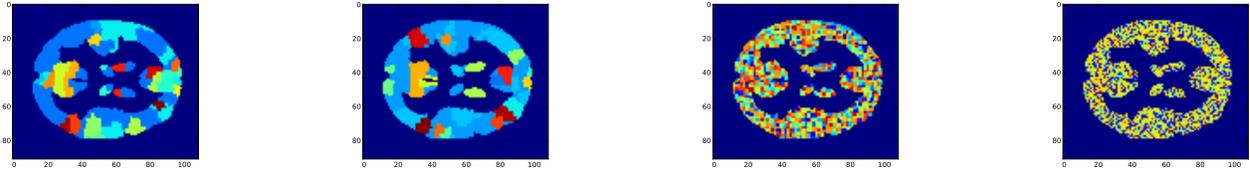
The task of brain connectivity analysis has two major branches: (1) effective connectivity estimation; (2) functional connectivity estimation. For the first type of connections, many of researchers have focused on using structure learning method for Bayesian Networks to obtain a directed network from fMRI data [25]. As to the estimation of functional connectivity, there are a few simple approaches such as hierarchical clustering, pairwise correlations and independent component analysis

(ICA), a comprehensive survey in these directions can be found in [26]. Sparse Gaussian graphical models (sGGM) [3], [4], [27], [28], [29] have been very popular for discovering large-scale brain connectivity recently due to their solid probabilistic foundation for distinguishing direct connections from indirect connections (*i.e.* conditional independent).

Deriving the mapping and connectivity of brain is crucial, but the brain network system itself is complexed structured. For instance, neurons are usually organized into multiple much larger regions and complex interconnections are existed between and within those regions. In this paper, we propose to use spectral clustering to discover the underlying cohesive regions, and the group constrained graphical model is employed to reconstruct the connectivity. Our model is different from [30] in several ways. Specifically, the node discovery methods proposed in [30] is semi-supervised and supervised, while our SGGL method is unsupervised. In [30], they discovered the edges within a network by estimating the correlations between nodes instead of inferring direct connections as we do in Sparse Gaussian Graphical Model (sGGM).

V. CONCLUSION

In this paper, we study the *brain network discovery* problem. By using the inferred networks as the input of group inference, we propose an iterative method, SGGL, to discover groups and links in the brain network simultaneously. Empirical experiments on both synthetic data and ADHD-200 dataset demonstrate that SGGL is promising in discovering meaningful brain networks.



(a) Groups of TDC inferred by SGGL ($k = 20$) (b) Groups of ADHD inferred by SGGL ($k = 20$) (c) Groups of TDC inferred by spectral clustering ($k = 20$) (d) Groups of ADHD inferred by spectral clustering ($k = 20$)

Fig. 8. Comparison between SGGL and spectral clustering on ADHD-200 Brain Dataset.

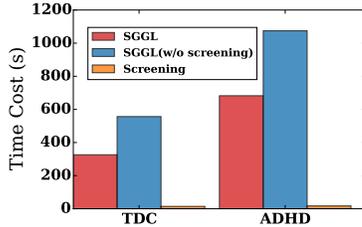


Fig. 9. Comparison of time costs on ADHD-200 Data.

REFERENCES

- [1] X. Kong, A. Ragin, X. Wang, and P. Yu, "Discriminative feature selection for uncertain graph classification," in *Proc. 13th SIAM Int. Conf. Data Mining (SDM'13)*, Austin, TX, 2013, pp. 82–93.
- [2] S. Yang, Q. Sun, S. Ji, P. Wonka, I. Davidson, and J. Ye, "Structural graphical lasso for learning mouse brain connectivity," in *Proc. 21st ACM SIGKDD Conf. Knowledge Discovery and Data Mining (KDD'15)*, 2015, pp. 1385–1394.
- [3] J. Friedman, T. Hastie, and R. Tibshirani, "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics*, vol. 9, no. 3, pp. 432–441, 2008.
- [4] O. Banerjee, L. E. Ghaoui, and A. d'Aspremont, "Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data," *Journal of Machine Learning Research (JMLR)*, vol. 9, pp. 485–516, 2008.
- [5] R. Craddock, G. James, P. Holtzheimer, X. Hu, and H. Mayberg, "A whole brain fMRI atlas generated via spatially constrained spectral clustering," *Human Brain Mapping*, vol. 33, no. 8, pp. 1914–1928, 2012.
- [6] J. Duchi, S. Gould, and D. Koller, "Projected subgradient methods for learning sparse gaussians," in *Proc. 24th Conf. Uncertainty in Artificial Intelligence (UAI'08)*, 2008, pp. 145–152.
- [7] C. Hsieh, I. Dhillon, P. Ravikumar, and M. Sustik, "Sparse inverse covariance matrix estimation using quadratic approximation," in *Advances in Neural Information Processing Systems 24 (NIPS'11)*, 2011, pp. 2330–2338.
- [8] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 1, pp. 49–67, 2006.
- [9] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [10] E. Birgin, J. Martínez, and R. Raydan, "Nonmonotone spectral projected gradient methods on convex sets," *SIAM Journal on Optimization*, vol. 10, no. 4, pp. 1196–1211, 2000.
- [11] L. Grippo, F. Lampariello, and S. Lucidi, "A nonmonotone line search technique for newton's method," *SIAM Journal on Numerical Analysis*, vol. 23, no. 4, pp. 707–716, 1986.
- [12] J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra, "Efficient projections onto the ℓ_1 -ball for learning in high dimensions," in *Proc. 25th Int. Conf. Machine Learning (ICML'08)*, 2008, pp. 272–279.
- [13] M. Schmidt, "Graphical model structure learning with ℓ_1 -regularization," Ph.D. dissertation, University Of British Columbia, 2010.
- [14] U. V. Luxburg, "A tutorial on spectral clustering," *Statistics and computing*, vol. 17, no. 4, pp. 395–416, 2007.
- [15] M. Kolar, H. Liu, and E. Xing, "Markov network estimation from multi-attribute data," in *Proc. 30th Int. Conf. Machine Learning (ICML'13)*, Atlanta, GA, June 2013, pp. 73–81.
- [16] S. Yang, Z. Lu, X. Shen, P. Wonka, and J. Ye, "Fused multiple graphical lasso," *SIAM Journal on Optimization*, vol. 25, no. 2, pp. 916–943, 2015.
- [17] Y. Sun, J. Han, P. Zhao, Z. Yin, H. Cheng, and T. Wu, "RankClus: integrating clustering with ranking for heterogeneous information network analysis," in *Proc. 12th Int. Conf. Extending Database Technology (EDBT'09)*, 2009, pp. 565–576.
- [18] C. Kuo, X. Wang, P. Walker, O. Carmichael, J. Ye, and I. Davidson, "Unified and contrasting cuts in multiple graphs: Application to medical imaging segmentation," in *Proc. 21st ACM SIGKDD Conf. Knowledge Discovery and Data Mining (KDD'15)*, 2015, pp. 617–626.
- [19] S. Huang, J. Li, J. Ye, T. Wu, K. Chen, A. Fleisher, and E. Reiman, "Identifying Alzheimer's disease-related brain regions from multi-modality neuroimaging data using sparse composite linear discrimination analysis," in *Advances in Neural Information Processing Systems 24 (NIPS'11)*, 2011, pp. 1431–1439.
- [20] B. Thirion, G. Flandin, P. Pinel, A. Roche, P. Ciuciu, and J. Poline, "Dealing with the shortcomings of spatial normalization: Multi-subject parcellation of fMRI datasets," *Human Brain Mapping*, vol. 27, no. 8, pp. 678–693, 2006.
- [21] Y. Zang, T. Jiang, Y. Lu, Y. He, and L. Tian, "Regional homogeneity approach to fmri data analysis," *Neuroimage*, vol. 22, no. 1, pp. 394–400, 2004.
- [22] P. Bellec, V. Perlbarg, S. Jbabdi, M. Péligrini-Issac, J. Anton, J. Doyon, and H. Benali, "Identification of large-scale networks in the brain using fMRI," *Neuroimage*, vol. 29, no. 4, pp. 1231–1243, 2006.
- [23] S. Smith, P. Fox, K. Miller, D. Glahn, P. Fox, C. Mackay, N. Filippini, K. Watkins, R. Toro, A. Laird *et al.*, "Correspondence of the brain's functional architecture during activation and rest," *Proceedings of the National Academy of Sciences*, vol. 106, no. 31, pp. 13 040–13 045, 2009.
- [24] B. Thirion, G. Varoquaux, E. Dohmatob, and J. Poline, "Which fMRI clustering gives good brain parcellations?" *Frontiers in Neuroscience*, vol. 8, no. 167, p. 13, 2014.
- [25] S. Huang, J. Li, J. Ye, A. Fleisher, K. Chen, T. Wu, and E. Reiman, "Brain effective connectivity modeling for Alzheimer's disease by sparse gaussian bayesian network," in *Proc. 17th ACM SIGKDD Conf. Knowledge Discovery and Data Mining (KDD'11)*, San Diego, CA, August 2011, pp. 931–939.
- [26] K. Friston, "Functional and effective connectivity: a review," *Brain Connectivity*, vol. 1, no. 1, pp. 13–36, 2011.
- [27] B. Marlin and K. Murphy, "Sparse gaussian graphical models with unknown block structure," in *Proc. 26th Int. Conf. Machine Learning (ICML'09)*, 2009, pp. 705–712.
- [28] R. Mazumder and T. Hastie, "Exact covariance thresholding into connected components for large-scale graphical lasso," *The Journal of Machine Learning Research*, vol. 13, no. 1, pp. 781–794, 2012.
- [29] L. Sun, R. Patel, J. Liu, K. Chen, T. Wu, J. Li, E. Reiman, and J. Ye, "Mining brain region connectivity for Alzheimer's disease study via sparse inverse covariance estimation," in *Proc. 15th ACM SIGKDD Conf. Knowledge Discovery and Data Mining (KDD'09)*, Paris, France, August 2009.
- [30] I. Davidson, S. Gilpin, O. Carmichael, and P. Walker, "Network discovery via constrained tensor analysis of fMRI data," in *Proc. 19th ACM SIGKDD Conf. Knowledge Discovery and Data Mining (KDD'13)*, Chicago, IL, August 2013, pp. 194–202.