

Extracting Brain Disease-Related Connectome Subgraphs by Adaptive Dense Subgraph Discovery

Qiong Wu¹, Xiaoqi Huang², Adam Culbreth³, James Waltz³,
Elliot Hong³, and Shuo Chen^{3*}

¹Department of Mathematics, University of Maryland, College Park, MD, USA

² Department of Mathematics, Johns Hopkins University, Baltimore, MD, USA

³Maryland Psychiatric Research Center, School of Medicine, University of Maryland,
Baltimore, MD, USA

*email: shuochen@som.umaryland.edu

Abstract

Group-level brain connectome analysis has attracted increasing interest in neuropsychiatric research with the goal of identifying connectomic subnetworks (subgraphs) that are systematically associated with brain disorders. However, extracting disease-related subnetworks from the whole brain connectome has been challenging, because no prior knowledge is available regarding the sizes and locations of the subnetworks. In addition, neuroimaging data is often mixed with substantial noise that can further obscure informative subnetwork detection. We propose a likelihood-based adaptive dense subgraph discovery (ADSD) model to extract disease-related subgraphs from the group-level whole brain connectome data. Our method is robust to both false positive and false negative errors of edge-wise inference and thus can lead to a more accurate discovery of latent disease-related connectomic subnetworks. We develop computationally efficient algorithms to implement the novel ADSD objective function and derive theoretical results to guarantee the convergence properties. We apply the proposed approach to a brain fMRI study for schizophrenia research and identify well-organized and biologically

meaningful subnetworks that exhibit schizophrenia-related salience network centered connectivity abnormality. Analysis of synthetic data also demonstrates the superior performance of the ADSD method for latent subnetwork detection in comparison with existing methods in various settings.

Keywords: brain connectome, densest subgraph, likelihood-based criterion, network topology, permutation test, schizophrenia

1 Introduction

Brain connectome analysis has become a powerful tool to understand the neurophysiology and neuropathology of brain diseases at a circuit level. These analyses focused on investigating patterns of functional and/or structural inter-connections between neural populations in the central nervous system associated with symptomatic phenotypes. Mounting evidence has shown that major neuropsychiatric disorders, including schizophrenia, Alzheimer’s disease, and autism among others, are associated with disrupted structural and functional connectivity patterns (Fornito et al., 2012).

Recent advances in neuroimaging statistics have facilitated group-level statistical analysis of structural and functional brain connectome data and the identification of disease-related brain connectome patterns (Biswal et al., 2010, Cao et al., 2019, Chen et al., 2019). In these analyses, the brain is often depicted as a graph (Bullmore and Bassett, 2011), where each node corresponds to a brain region of interest (ROI) and an edge represents the connectivity linking any two nodes. An edge can represent functional connectivity based on functional magnetic resonance imaging (fMRI) data at rest or task, structural connectivity measuring white matter track connections, and weighted connection metric integrating multimodal brain connectivity (Bowman et al., 2012). These multivariate edges are the variables of interest in brain connectome analysis, which are constrained by the nodes in a weighted adjacency matrix and thus exhibit network topological properties (Chen et al., 2020). Statistical inference for multivariate edge variables in an adjacency matrix remains challenging because of the need to account for multiple testing corrections and network topological structures simultaneously. Many statistical graph models have been developed and successfully applied

to brain connectome data analysis yielding important findings (Vogelstein et al., 2012, Durante et al., 2018, Ghoshdastidar and von Luxburg, 2018, Ginestet et al., 2017, Higgins et al., 2019, Kundu et al., 2018, Lukemire et al., 2017, Mejia et al., 2019, Simpson et al., 2019, Xia and Li, 2017, Xia and Li, 2018 among many).

The current study focuses on extracting informative/signal subgraphs that are likely related to brain diseases from the whole brain connectome (Vogelstein et al., 2012). Our overarching goal is to accurately capture underlying signal subnetworks, such that the extracted subnetworks i) cover a high proportion of true positive edges (i.e., high sensitivity); (ii) include a few false positive edges (i.e., low false discovery rate (FDR)); and iii) are composed of highly organized network topological structures (i.e., biologically interpretable). In practice, however, this task is challenging because it is difficult to simultaneously balance the sensitivity and false positive findings while constraining all positive edges in organized subgraphs. The ‘dense’ subnetwork detection then becomes attractive because a subgraph of a small number of nodes in an organized network topological structure covering most signal edges can also lead to low FDR and high sensitivity. Although less discussed in the statistical literature, dense subgraph discovery in the field of computer science research has been carefully worked out (Lee et al., 2010), and thus may suit our needs for statistical analysis of brain connectome data.

Dense subgraph discovery methods are designed to identify a subgraph with a maximal density among all possible subgraphs, in short, the densest subgraph, in a binary graph. These methods rely on the assumption that the overall graph is non-random and there exists some subgraph where the edge ratios are much higher than the rest of the graph. Goldberg (1984) reduces the problem to a sequence of max-flow/min-cut computations, which requires a logarithmic number of min-cut calls. Asahiro et al. (1996) propose a simple and fast greedy algorithm that is showed to have a 2-approximation guarantee by Charikar (2000). Nevertheless, existing dense subgraph discovery algorithms may not be directly applicable to our analysis due to the substantial noise in the group level brain connectome data. As demonstrated in Figure 1a, there may exist an enormous amount of false positive and false negative errors in edge-wise inference results that give rise to the difficulty of detecting dense subgraph using existing methods. Specifically, due to the noise, existing

dense subgraph discovery algorithms tend to either identify over-sized subgraphs that may include a large proportion of false positive edges with low importance levels (high FDR) or detect over-conservative small-sized subgraphs that may not sufficiently cover signal edges (low sensitivity, [Tsourakakis et al., 2013](#)). Moreover, the computational cost of many these dense subgraph density discovery algorithms is expensive, which may lead to intractable computational time for commonly used statistical inference methods for brain connectome analysis (e.g., permutation tests). Hence, we are motivated to integrate modern statistical techniques into dense graph discovery and mitigate these challenges for brain connectivity network analysis.

We propose a likelihood-based adaptive dense subgraph discovery (ADSD) model to extract informative connectomic subnetworks accurately. The new objective function is robust to edge-wise false positive and false negative noise by introducing a tuning parameter to balance the area density and degree density ([Tsourakakis et al., 2013](#)). We optimize the tuning parameter objectively by maximizing the widely used likelihood function in statistical network/graph model (e.g. stochastic block model, [Holland et al., 1983](#), [Zhao et al., 2012](#), [Zhang et al., 2017](#)). We develop efficient algorithms to implement the joint objective function of the ADSD. We further derive theoretical results which guarantee the approximation properties for any fixed graph and consistency for large graphs based on the proposed algorithm. We can perform permutation tests to yield statistical significance of extracted subgraphs ([Ge et al., 2012](#), and [Zalesky et al., 2010](#)). We perform extensive simulation studies to validate the proposed model and theoretical conclusions. The results demonstrate improved accuracy of informative subgraph detection in various settings. Our method is then applied to a resting state fMRI (rfMRI) brain connectomic study for schizophrenia research. The results of our real data analysis reveal for the first time systematic aberrant salience network centered connectivity patterns in schizophrenia patients using whole brain connectome network analysis. Although some of our findings coincide with previous studies using seed voxel-based method, our analysis is more comprehensive and less biased, because it does not require pre-selected seed sets or focuses on exclusively known networks.

2 Methods

2.1 Background: group-level inference for multivariate edges in a graph space

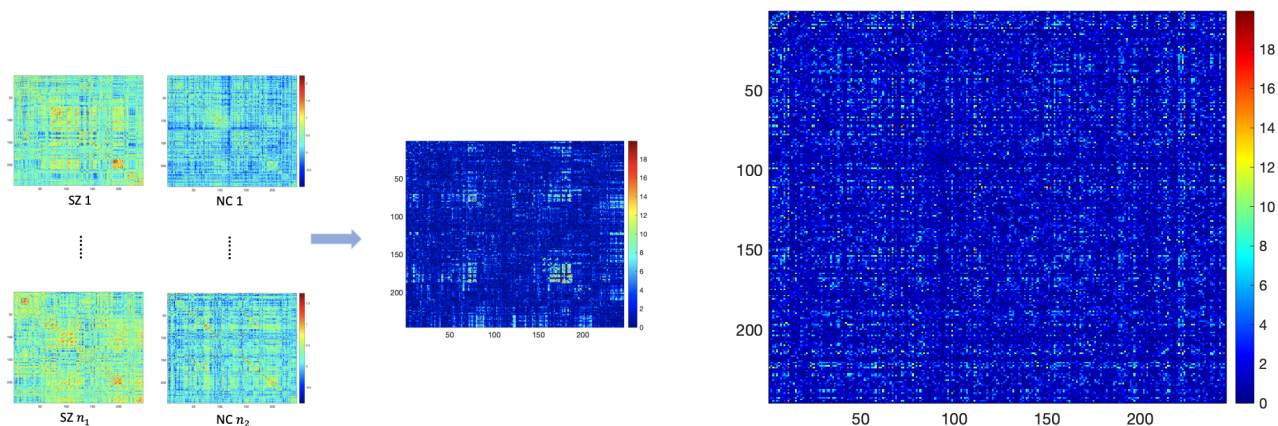
Let $G = (V, E)$ be an undirected graph, where $V = \{v_i\}_{i=1}^n$ is a set of nodes representing brain areas and ROIs, and $E = \{e_{ij}\}_{i<j}^n$ denotes the set of edges between pairs of nodes (i.e., connections between brain areas). $G' = (V', E')$ is a subgraph of G if $V' \subset V$ and $E' \subset E$. Then, $G(S) = (S, E(S))$ is a ‘nodes-induced’ subgraph if $S \subset V$ and $E(S) = \{(u, v) \in E | u, v \in S\}$ being edges in E with endpoints in S . We use $(\mathbf{Y}^k, X^k)_{k=1}^K$ to represent the group-level multivariate edge data in a graph space $G = (V, E)$, where $k = 1, \dots, K$ is the subject index. $\mathbf{Y}_{n \times n}^k$ represents the brain connectome data in a binary/weighted adjacency matrix for subject k , and X^k is the corresponding vector of covariates (clinical and demographic variables). We assume that the location of nodes and edges are identical across subjects after spatial normalization. Thus, our goal is to perform statistical analysis and identify phenotype-related subnetworks with high sensitivity and well-controlled FDR (Lukemire et al., 2017, Kundu et al., 2018, Xia and Li, 2018, Vogelstein et al., 2012, Durante et al., 2018). Figure 1a demonstrates the procedure of group-level inference for brain connectome data.

Let $W = \{w_{ij}\}_{i,j=1}^n$ denote the edge-wise inference matrix based on graph G , where each off diagonal entry w_{ij} represents the edge-wise statistical inference results on edge e_{ij} (e.g., test statistics t_{ij} and p values $-\log(p_{ij})$). For each edge e_{ij} , we denote a corresponding latent indicator variable δ_{ij} such that $\delta_{ij} = 1$ if edge e_{ij} is associated with the phenotype of interest and $\delta_{ij} = 0$ otherwise. We consider the edge-wise inference results as our input data (Chen et al., 2020).

The goal of group-level brain network analysis is to identify a set of subnetworks $\{\hat{G}_c\}$ ($\hat{G}_c = (\hat{V}_c, \hat{E}_c)$) that are associated with a phenotype of interest, such that

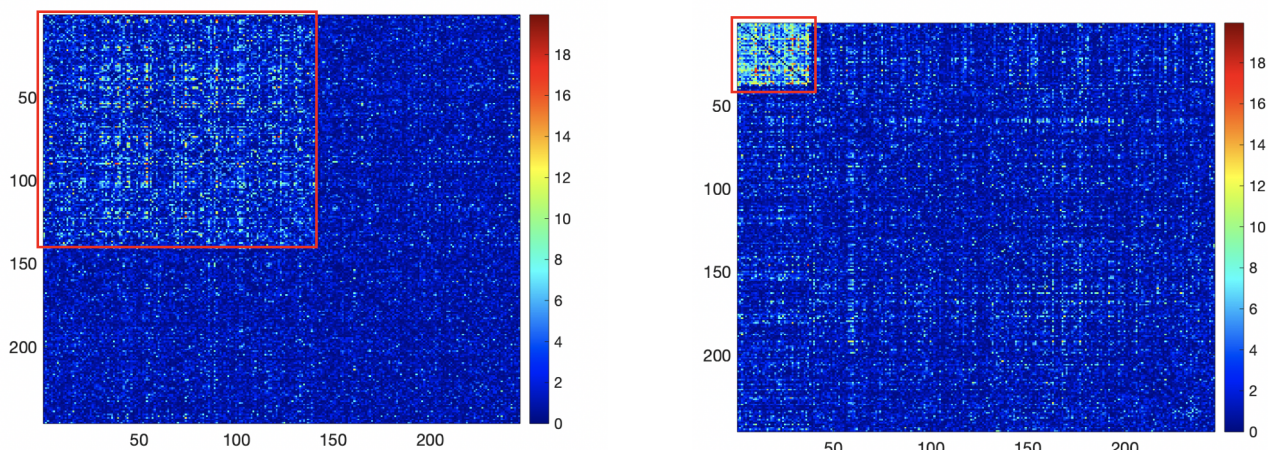
1. $\Pr(\delta_{ij} = 1 | e_{ij} \in \hat{G}_c) > \Pr(\delta_{ij} = 1 | e_{ij} \notin \hat{G}_c)$ (dense subgraph);
2. The false discovery rate (FDR), $\frac{\sum_{i<j} I(\delta_{ij}=0 | e_{ij} \in \hat{G}_c)}{\sum_{i<j} I(e_{ij} \in \hat{G}_c)}$, is low;

Figure 1: Motivation for informative subgraph extraction: (a) demonstrates the process of obtaining edge-wise inference matrix from the population level connectome data; (b) illustrates the commonly used community detection results (e.g. using stochastic block model) cannot detect any informative subgraph; (c) shows the results of existing dense subgraph discovery results; (d) describes a desirable informative subgraph detection procedure which can identify an organized and biologically interpretable topological structure consisting of informative edges. The results in (d) are based on the ADSD method (see details in the Results section).



(a) Edge-wise inference for group-level connectome data

(b) Subgraph extraction by SBM with $K = 30$



(c) Subgraph extraction by conventional dense subgraph discovery

(d) Subgraph extraction by ADSD

3. The sensitivity, $\frac{\sum_{i<j} I(\delta_{ij}=1|e_{ij}\in\hat{G}_c)}{\sum_{i<j} I(\delta_{ij}=1)}$, is high;
4. \hat{G}_c is a well defined community (e.g., a node-induced subgraph that $\hat{G}_c = G(\hat{V}_c)$).

In practice, the task above is challenging. For example, the mass univariate methods including both FDR and family-wise error rate (FWER) controlling models apply an universal threshold on all edges, and yield a set of unrelated ‘significant’ edges. Thus, they can neither address the trade-off between sensitivity and false positive findings by leveraging the information of network or yield findings with an organized and biologically interpretable network topological structure. The network based statistics (NBS) method allows edges borrow strengths from each other, yet it yields an unorganized subgraph (Zalesky et al., 2010, Fornito et al., 2012). Moreover, the signal subnetwork detected by NBS includes all nodes in G almost surely, i.e., $G(V_c) = G$, when n is larger than a handful of nodes (Erdős and Rényi, 1960), and thus less interpretable.

We also notice that the proportion of true positive edges $\frac{\sum I(\delta_{ij}=1)}{|E|}$ in G is often small in our motivated brain connectome data (e.g., around 5%), which may lead to the difficulty of applying the commonly used network models (Chen et al., 2018). Figure 1b shows the results of the application of stochastic block models (SBM) which miss the network topological structure. Therefore, it is highly desirable to extract a ‘dense’ subgraph which is a node-induced subgraph $G(S_0)$, such that the edge density ρ_s is much higher than the overall density ρ :

$$\rho_s > k\rho, k > 1, \text{ with } \rho_s \triangleq \frac{\sum_{i<j, i,j\in S_0} I(\delta_{ij} = 1)}{|E(S_0)|} \text{ and } \rho \triangleq \frac{\sum_{i<j} I(\delta_{ij} = 1)}{|E|},$$

and $G(S_0)$ includes most edges. The detected informative subgraph can either directly become the subnetwork of interest or intermediate results for further refined network analysis (e.g., using SBM).

2.1.1 Dense Subgraph Discovery

The conventional dense subgraph aims to detect a node-induced subgraph with maximized density. Two popular definitions of density function are also referred as average degree and

edge ratio (Asahiro et al., 1996, Charikar, 2000, Lee et al., 2010):

$$f_1 = \frac{\sum_{i < j, i, j \in S_0} I(\delta_{ij} = 1)}{|S_0|} \text{ and } f_2 = \frac{\sum_{i < j, i, j \in S_0} I(\delta_{ij} = 1)}{|E(S_0)|}$$

The edge ratio agrees with our goal for informative subgraph detection. However, the implementation of dense subgraph discovery is not trivial. The direct optimization of edge ratio f_2 tends to detect a high-density subgraph with a tiny size (See Lemma 2 in supplementary for details). Meanwhile, it has been known the optimization of f_1 can lead to the detection of an over-sized subgraph (Tsourakakis et al., 2013), which may cause a high false positive rate for statistical inference. Figure 1c shows the results of conventional dense graph discovery by optimizing f_1 . To address these challenges, we propose a likelihood based method for dense subgraph discovery.

2.2 Adaptive dense subgraph discovery

We consider $G = (V, E, W)$ as our input data that stores edge-wise inference results in a weighted adjacency matrix W . Our goal is to extract a phenotype-related informative subgraph $G(S)$ induced by nodes set S in the sense that $E(w_{ij} | e_{ij} \in E(S)) \gg E(w_{ij} | e_{ij} \notin E(S))$ while maximally reducing false negative findings and improving the sensitivity.

To address the challenges in conventional dense subgraph discovery and improve the balance of the trade-off, we propose an adaptive density function:

$$\arg \max_{S \subset V} f(S; \lambda) := \frac{|W(S)|}{|S|^\lambda} \quad (1)$$

where $\lambda \in [1, 2]$ is a tuning parameter, such that when $\lambda = 1$ and 2, $f(S; \lambda)$ density function reduces to f_1 and f_2 , respectively.

To better illustrate the impact of the tuning parameter λ on the FDR and sensitivity, we transform the objective function (1) to:

$$\arg \max_{S \subset V} \left\{ \log f_2(S) + \gamma \log \frac{|W(S)|}{|W(G)|} \right\} \quad (2)$$

with $\gamma \in [0, 1]$. The optimal solution is approximated by $f(S; \lambda)$ for large graphs with

$\gamma = \frac{2}{\lambda} - 1$. The first term in (2) is the true discovery rate ($1 - \text{FDR}$), while the second term is the sensitivity (power). In that, λ functions similarly to the tuning parameter in the shrinkage methods (e.g., LASSO) since f_1 is related to the loss function and f_2 implements the rule of parsimony. Increasing λ leads to a low FDR, while decreasing λ can improve the sensitivity. Therefore, our objective function is tailored for the four items of our overarching goal in the previous subsection.

In practice, both $G(S)$ and λ need to be estimated, and λ is critical to balance the trade-off between FDR and sensitivity. We propose an iterative procedure to optimize the objective function (1) in subsection 2.2.1 and estimate λ in subsection 2.2.2. We name this new procedure adaptive dense subgraph discovery (ADSD).

2.2.1 Optimization with a known λ

We implement the objective function (1) using a greedy algorithm. The greedy algorithm has been the most commonly used technique to implement objective functions for dense subgraph discovery (Asahiro et al., 1996, Charikar, 2000). Generally, a greedy algorithm removes a node with the minimum-degree at each iteration, and then selects the optimal dense subgraph from the process of node removal. The detailed procedure is described by Algorithm (3).

Algorithm 1 Optimizing objective function (2) with a given λ

```
1: procedure ALGORITHM
    $S_1 \leftarrow V$ 
2:   for  $k=1$  to  $n - 1$  do
3:     let  $v$  be the node in  $G(S_k)$  with smallest degree:  $v = \arg \min_{i \in S_k} \deg_{G(S_k)}(i)$ ;
4:      $S_{k+1} \leftarrow S_k / \{v\}$ ;
5:   end for
6:   Output the subgraph with largest objective function among  $G(S_1), \dots, G(S_{n-1})$ ;
7: end procedure
```

We denote the optimal dense subgraph based on our objective function (1) with a given λ by

$$S_\lambda^* = \arg \max_{S \subset V} f(S; \lambda),$$

and the output of greedy algorithm as:

$$\tilde{S}_\lambda = \arg \max_{S_1, \dots, S_{n-1}} f(S; \lambda).$$

A major advantage of the greedy algorithm is the low computational complexity, which is critical for our application. Although our greedy algorithm may not provide the exact solution, Charikar (2000) proved the greedy algorithm has a 2-approximation as to $f_1(\cdot)$, that is $f_1(\tilde{S}_1) \geq 2f_1(S_1^*)$, where \tilde{S}_1 is the densest subgraph by greedy algorithm and S_1^* is the true maximizer for $f_1(\cdot)$. In Section 3, we prove the theoretical approximation properties of \tilde{S}_λ with regard to the maximization $f(S^*; \lambda)$ for various values of λ .

2.2.2 Likelihood-based method for λ estimation

Clearly, the performance of our greedy algorithm (3) relies on the unknown parameter λ (e.g. $\lambda = 1$ and 2 lead to the optimization $f_1(\cdot)$ and $f_2(\cdot)$ alone respectively). We propose a data-driven approach to automatically determine λ by maximum likelihood estimation. In statistical literature, the likelihood function of network/graph data has been well studied (Zhang et al., 2017). For example, a binary graph with K block can be defined by:

$$A_{ij} | \theta_i = a, \theta_j = b \sim \text{Bernoulli}(\pi_{ab})$$

where $A^{n \times n}$ is a binary adjacency matrix, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$ is a latent vector of node labels, and $\boldsymbol{\pi} = (\pi_{ab})_{a,b=1}^K$ is a $K \times K$ symmetric probability generating matrix for generative for edges within and between blocks/communities.

We adopt the likelihood function of SBM because the dense subgraph structure in our ADSD model can be considered as a special case of the block diagonal structure in SBM. Specifically, in our model the graph $G = (V, E)$ includes an underlying true informative subgraph $G(S_0)$ and all other nodes are singletons. The number of communities of SBM is $K = n - n_s + 1$, where $n = |V|$ and $n_s = |S_0|$. We further assume the planted partition model that the parameters of Bernoulli distributions for edges between blocks are identical in SBM.

To construct the likelihood function for ADSD, we first binarize the input data matrix W using a threshold r and let $A_{ij} = \{W(r)\}_{ij} = I(W_{ij} > r)$. We denote $\boldsymbol{\theta}(S)$ as a vector of node labels concerning the node set S for a dense subgraph $G(S)$, where an element $\theta_i(S) = 1$ if $i \in S$ and $\theta_i(S) = 0$ for $i \in V/S$. Then, the membership of edges regarding the nodes-induced subgraph $G(S)$ can be defined consequently as $\theta_{ij}(S) = \theta_i(S)\theta_j(S)$.

We let all edges in A follow a Bernoulli distribution with parameters π_{ij} that

$$\pi_{ij} = \begin{cases} \pi_s & \text{both } i, j \in S_0 \\ \pi_0 & \text{o.w.} \end{cases}$$

Using the mixture model representation, $\pi_{ij} = \theta_{ij}(S_0)\pi_s + (1 - \theta_{ij}(S_0))\pi_0$. The likelihood function based on $A = W(r)$ is:

When the membership of informative subgraph is given, the MLE of the edge probabilities can be obtained by:

$$\hat{\pi}_s^{MLE} = \frac{|A(S_0)|}{|E(S_0)|} \text{ and } \hat{\pi}_0^{MLE} = \frac{|A| - |A(S_0)|}{|E| - |E(S_0)|}.$$

In practice, S_0 is unknown and can be estimated by \tilde{S}_λ from the Algorithm (3) with a given λ . The likelihood function based on \tilde{S}_λ is in the form:

$$\begin{aligned} L_\lambda(\hat{\pi}_s^{MLE}, \hat{\pi}_0^{MLE}; \boldsymbol{\theta}(\tilde{S}_\lambda), A) &= \prod_{i < j, i, j \in \tilde{S}(\lambda)} (\hat{\pi}_s^{MLE})^{a_{ij}} (1 - \hat{\pi}_s^{MLE})^{1-a_{ij}} \\ &\times \prod_{i < j, i \in V/\tilde{S}(\lambda) \text{ or } j \in V/\tilde{S}(\lambda)} (\hat{\pi}_0^{MLE})^{a_{ij}} (1 - \hat{\pi}_0^{MLE})^{1-a_{ij}} \end{aligned}$$

where $\boldsymbol{\theta}(\tilde{S}_\lambda)$ is the node label vector associated with \tilde{S}_λ .

Therefore, λ can be estimated by two steps. First, for any $\lambda \in [1, 2]$, we can extract a dense subgraph \tilde{S}_λ by the greedy algorithm (3). Next, $\hat{\lambda}$ is determined by the combination of λ and \tilde{S}_λ that maximizes the likelihood function:

$$\hat{\lambda} = \arg \max_{\lambda} L_\lambda(\hat{\pi}_s^{MLE}, \hat{\pi}_0^{MLE}; \boldsymbol{\theta}(\tilde{S}_\lambda), A),$$

The final result of dense subgraph discovery based on the MLE determined $\hat{\lambda}$ is $\hat{\theta} = \theta(\tilde{S}_{\hat{\lambda}})$.

We further consider the threshold r in $A_{ij} = \{W(r)\}_{ij} = I(W_{ij} > r)$ as a random variable following a distribution $g(r)$ rather than a fixed value in order to avoid an arbitrary selection. We integrate the likelihood with respect to r based on the prior distribution $g(r)$, and thus our optimization is invariant to the selection of r . $g(r)$ can be a discrete distribution with a support $\{r_1, \dots, r_m\}$ and corresponding probability $\{g(r_1), \dots, g(r_m)\}$. In practice, the performance of our algorithm is robust to the prior distribution, given the reasonable support of r is used. By integrating r out, the likelihood function becomes:

$$L_{\lambda}(\hat{\pi}_s^{MLE}, \hat{\pi}_0^{MLE}; \theta(\tilde{S}_{\lambda}), W) = \int L_{\lambda}(\hat{\pi}_s^{MLE}, \hat{\pi}_0^{MLE}; \theta(\tilde{S}_{\lambda}), W(r)) g(r) dr$$

The general algorithm for ADSD is described in the Algorithm (2). Since Algorithm (3) is nested within the overall Algorithm (2), the low computational cost of Algorithm (3) is critical for the overall computational efficiency of ADSD. The complexity of the ADSD algorithm is $O(Mn^2)$ where M is a sufficient searching range of λ . The resulting subgraph $G(\hat{S}_{\hat{\lambda}})$ from our ADSD model can be further investigated for more delicate latent topological structures and statistically tested by permutation tests with family-wise error rate control (Zalesky et al., 2010, Chen et al., 2015) and we include the details in the supplementary materials.

Algorithm 2 The complete ADSD algorithm

- 1: **procedure** ALGORITHM $\lambda \leftarrow 1$
 - 2: **while** $\lambda \leq 2$ **do**
 - 3: return the densest subgraph $\tilde{S}(\lambda)$ of W from Algorithm (3)
 - 4: **for** $r = r_1$ to r_m **do**
 - 5: calculate the likelihood: $L_{\lambda}(\hat{\pi}_s^{MLE}, \hat{\pi}_0^{MLE}; \theta(\tilde{S}_{\lambda}), W(r))$
 - 6: **end for**
 - 7: integrate w.r.t. λ :
 $L_{\lambda}(W) = \sum_{i=1}^m L_{\lambda}(\hat{\pi}_s^{MLE}, \hat{\pi}_0^{MLE}; \theta(\tilde{S}_{\lambda}), W(r_i))g(r_i)$
 - 8: **end while**
 - 9: Output $\hat{\lambda}$ and $\tilde{S}_{\hat{\lambda}}$ with maximized $L_{\lambda}(W)$
 - 10: **end procedure**
-

3 Theoretical results

The theoretical work for conventional dense graph discovery has been well-established (Lee et al., 2010). For example, Charikar (2000) showed that the commonly used greedy algorithm proposed by Asahiro et al. (1996) has a 2-approximation bound. In this article, we aim to extend the theoretical results for our new ADSD algorithms in 2 which generalizes the traditional objective function by introducing the parameter λ . Specifically, we discuss the approximation bounds for ADSD with a full range of λ values in the following theorem 1.

Theorem 1 (Exact property of Algorithm 3). *For a given graph $G = (V, E)$, with S_λ^* and \tilde{S}_λ defined in section 2, the Algorithm (3) has a $\rho(\lambda, n)$ -approximation, especially $f(S_\lambda^*; \lambda) \leq \rho(\lambda, n)f(\tilde{S}_\lambda; \lambda)$ with*

$$\rho(\lambda, n) = \begin{cases} c(\lambda) & \text{if } \lambda \geq 2 \\ 2c'(\lambda)n^{(\lambda-1)(2-\lambda)} & \text{if } 1 < \lambda < 2 \\ 2n^{1-\lambda} & \text{if } 0.5 < \lambda < 1 \\ 2n^\lambda, & \text{if } 0 < \lambda \leq 0.5 \end{cases}$$

where $c(\lambda) = 2^{\lambda-1}$ and $c'(\lambda) = 1 \vee 2^{1-\lambda}$.

The theorem 1 provides the performance of Algorithm (3) by guaranteeing the closeness of objective function in S_λ^* and \tilde{S}_λ . However, an optimal optimization may not result from a perfect recovery of informative subgraph for randomness (i.e. $S_\lambda^* \neq S_0$ for all λ). Hence, we further prove the asymptotic consistency of \tilde{S}_λ from Algorithm (3) in following theorem 2. When the observed graph is generated from some underlying model with true informative subgraph S_0 , there exist an λ such that \tilde{S}_λ tends to S_0 with probability 1 asymptotically.

Theorem 2 (Asymptotic property of Algorithm 3). *Assume the graph $G = (V, E)$ including an informative subgraph $G(S_0) = (S_0, E(S_0))$ is generated from the special SBM we defined in section 2.2.2, such that the edges are drawn from independent Bernoulli distributions with parameter $\pi_{ij} = \pi_{ij}(S_0) = \theta_{ij}(S_0)\pi_s + (1 - \theta_{ij}(S_0))\pi_0$, where $\theta_{ij}(S) = \theta_i(S)\theta_j(S)$, $\theta_i(S) = I(i \in S)$ and $\pi_s > \pi_0$. Let $|S_0| = O(|V|^{1/2+\epsilon})$ as $n \rightarrow \infty$ for any $\epsilon > 0$.*

Then, there exist some λ such that we will get exact recovery with probability 1 in Algorithm (3), i.e. as $n \rightarrow \infty$,

$$\mathbb{P}(\forall i, \theta_i(\tilde{S}_\lambda) = \theta_i(S_0)) \rightarrow 1.$$

Theorem 2 provides the existence of parameter λ for a consistent estimator as the size of graph goes to infinity. We use the following Theorem 3 to demonstrate the performance of Algorithm (2) by illustrating the selected λ based on our likelihood-based criterion will lead to an estimator with negligible proportion of incorrect assignment for nodes.

Theorem 3. Assume the graph $G = (V, E)$ includes an informative subgraph $G(S_0) = (S_0, E(S_0))$, such that the edges generate from independent Bernoulli distributions with parameter $\pi_{ij} = \pi_{ij}(S_0) = \theta_{ij}(S_0)\pi_s + (1 - \theta_{ij}(S_0))\pi_0$, where $\theta_{ij}(S) = \theta_i(S)\theta_j(S)$, $\theta_i(S) = I(i \in S)$ and $\pi_s > \pi_0$. Let $|S_0| = O(|V|^{1/2+\epsilon})$ as $n \rightarrow \infty$ for any $\epsilon > 0$.

Then, as $n \rightarrow \infty$, the adaptive greedy algorithm with likelihood-based criterion results in an estimate $\hat{\theta} = \theta(\tilde{S}_\lambda)$ with:

$$\hat{\lambda} = \arg \max_{\lambda} \sup_{\pi_s, \pi_0} L(\pi_s, \pi_0; \theta(\tilde{S}_\lambda); A)$$

has incorrect assignment with probability converging to zero, i.e.

$$N_e(\hat{\theta}) = \sum_{i=1}^n I(\hat{\theta}_i \neq \theta_i(S_0)) = o_p(n).$$

The detailed derivations and proofs for the above three theorems are provided in the supplementary materials.

4 Results

We apply the proposed ADSD method to the neuroimaging data collected from patients with schizophrenia and healthy controls. This data set includes 104 patients with schizophrenia (SZ) (age 36.88 ± 14.17 , 62 males and 41 females, 1 other) and 124 healthy controls (HC)

(age 33.75 ± 14.22 , 61 males and 63 females). There are no systematic differences in age (test statistic 1.64, p value 0.10) or gender (test statistic 1.67, p value 0.10) between the two groups. The imaging acquisition and preprocessing details are described in [Adhikari et al. \(2019\)](#). A brain connectivity-based atlas is used to denote 246 regions of interest (ROIs) as nodes in a brain connectome graph ([Fan et al., 2016](#)). The functional connection (edge) between a pair of nodes for each subject is calculated by the covariation between averaged time series from the two corresponding brain ROIs. The Fisher's Z transformed Pearson correlation coefficient then is applied for each edge. We perform non-parametric group level testing on each edge, although alternative inference methods can be used as well.

We focus on the input matrix W reflecting the importance levels ($-\log(p_{ij})$) on all edges, as demonstrated in [Figure 2a](#). We first apply the greedy algorithm (e.g., Charikar's method that is equivalent to the proposed greedy algorithm with an ad-hoc $\lambda = 1$) for dense subgraph extraction. The results in [Figure 2b](#) seem to be an over-inflated subnetwork without clear biological interpretation and a large set of false positive edges. We also applied other popular subgraph detection methods, for example, breadth first search in network-based statistics, stochastic block model, and various community detection methods ([Zalesky et al., 2010](#), [Amini et al., 2013](#), [Newman and Girvan, 2004](#)). However, these algorithms either detect a subgraph including all brain regions or yield no findings. In contrast, by implementing our ADSD method ([2](#)), we obtain a subnetwork $\hat{S} = \tilde{S}_{\hat{\lambda}}$ with $\hat{\lambda} = 1.2$. We note that the detected subgraph is robust to the prior distribution of $G(r)$ as long as a reasonable support is used.

The computation is efficient, and it takes 2.21 seconds to implement the ADSD algorithm on a Mac with CPU Core i5 and memory 8GB. We further calculate the p-value of the network based on permutation test ([Zalesky et al., 2010](#), [Chen et al., 2015](#), [Chen et al., 2020](#)). The p-value for the network is significant $p < 0.001$ with family wise error rate adjustment.

The results show a subnetwork with reduced functional connectivity in patients with schizophrenia compared to healthy controls (see in [Figures 3](#)), which is consistent with the current knowledge that schizophrenia is possibly a degenerative disorder and associated with hypoconnectivity ([Li et al., 2019](#)). This subnetwork is centered around the well-known salience network (SN) which is primarily composed of bilateral insular gyri (INS) and anterior cingulate cortices (ACC). The salience network contributes to complex and integrative

brain functions including emotions, cognition, and self-awareness (Uddin, 2015). Numerous previous studies have reported that decreased functional connectivity in the salience network is related to several core symptoms of schizophrenia using seed voxel methods (Palaniyappan et al., 2012). Our findings are well aligned with these established results. In addition to SN, our subnetwork extracted by ASD involves several other brain regions including bilateral superior temporal gyri (STG), superior frontal gyri (SFG), precentral gyri (PCL), inferior parietal lobe left (IPL), and orbitofrontal cortex right (OrG). These regions have been identified to associate with auditory perceptual abnormalities (STG, IPL), voluntary movement (PCL), and sensory and cognition (SFG, OrG) (Fornito et al. (2012)). Jointly, our detected subnetwork reveals a comprehensive and systematic brain connectivity aberrance in patients with schizophrenia, which is related to the impaired capability to integrate and comprehend information (e.g., multiple external stimuli) and to respond appropriately. The detected schizophrenia-related brain connectome subnetwork is biologically plausible. It provides evidence to combine prior isolated findings, and thus enhances our understanding of the complex brain connectomic patterns and clinical symptoms.

Thus, our novel analytic approach revealed a neural sub-network that has been previously shown to both differentiate healthy controls and patients with schizophrenia and has been critically linked to core symptoms of the disorder. Since our results do not depend on the arbitrary selection of seed voxels and pre-specified networks of interest, our results are subject to less selection bias and thus more reliable and comprehensive.

5 Simulation

In the simulation study, we generate multiple brain connectivity data sets under several settings. We consider a graph G with $|V| = 100$, where an informative subgraph in a community structure with two possible sizes $|S_0| = 15$ and 30. We generate 60 connectivity matrices for 30 controls and 30 cases. We assume that most edges in the informative subgraph are differentially expressed between cases and healthy controls. We let the connectivity weights of edges inside the informative subgraph follow a normal distribution with mean μ_1 and variance σ^2 , while all other edges have normal μ_0 and σ^2 for the case group. In the

Figure 2: Results of data example: (a) is the input matrix W ; (b) shows the results of existing dense graph discovery; (c) demonstrates the results by applying ADSD; (d) illustrates refined topological structure based on results of ADSD.

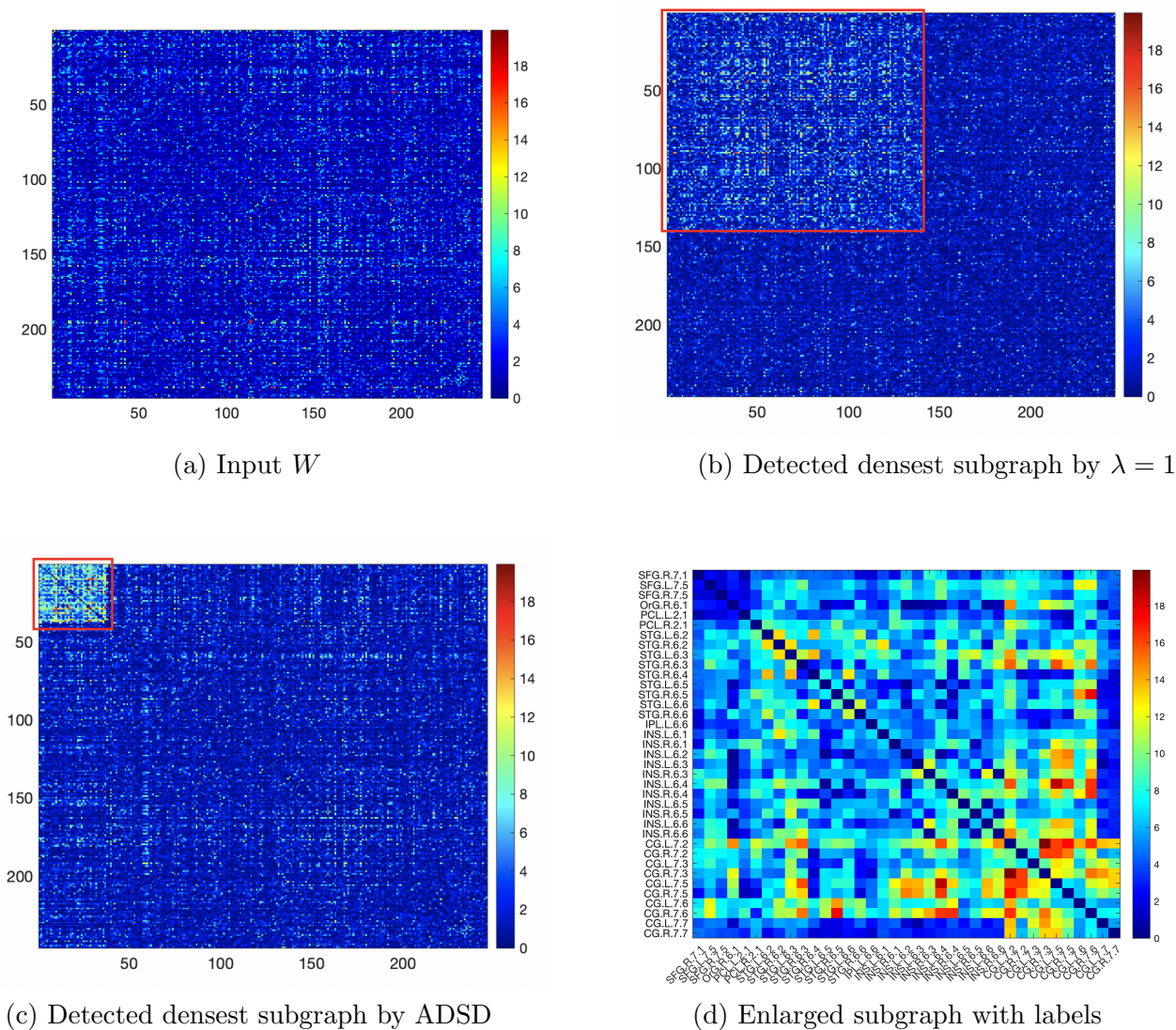
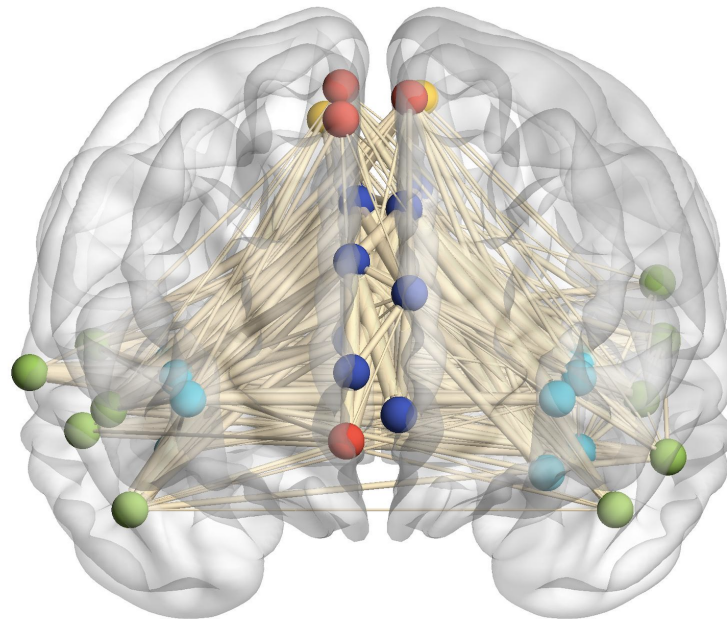
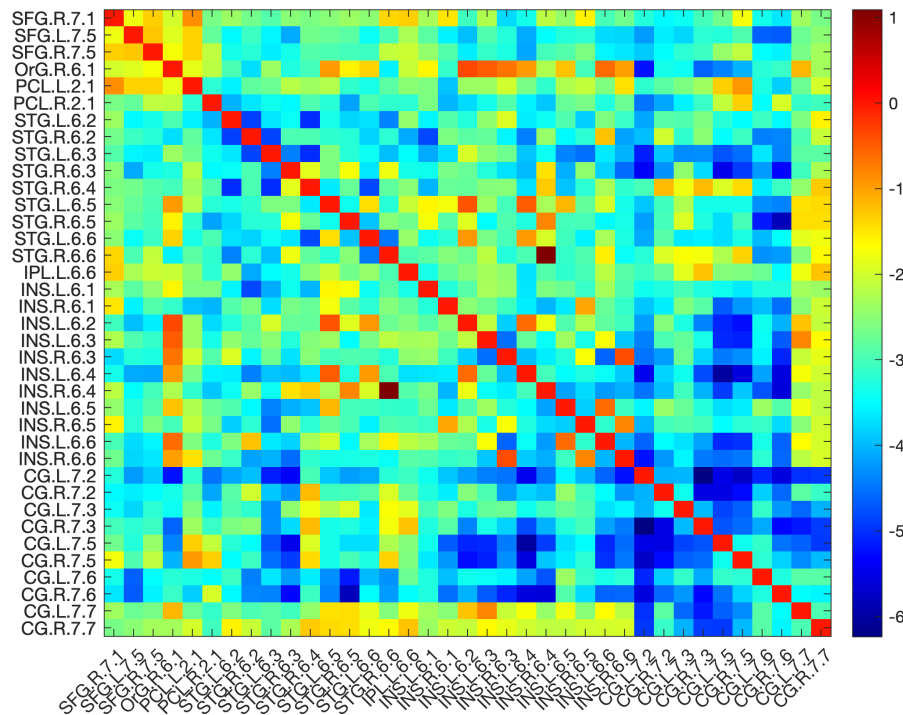


Figure 3: (a) illustrates the enlarged and labeled informative subgraph in t-statistics detected by ASD which indicates decreased functional connectivity of SZ. (b) is a 3D demonstration of the subgraph: red nodes represent superior frontal gyrus (SFG) + orbitofrontal cortex right (OrG); yellow nodes are precentral gyri (PCL); green nodes are superior temporal gyrus (STG)+inferior parietal lobe left (IPL); blue nodes represent insular gyrus (INS); navy nodes represents cingulate cortex (CG).



control group, we let all edges follow a normal distribution of μ_0 and σ^2 . Specifically,

$$x_{ij(s)}^{\text{case}}|\{i < j, i, j \in S_0\} \sim N(\mu_1, \sigma^2), \quad x_{ij(s)}^{\text{case}}|\{i < j, i \text{ or } j \notin S_0\} \sim N(\mu_0, \sigma^2),$$

and

$$x_{ij(s')}^{\text{control}}|\{i < j, i, j \in V\} \sim N(\mu_0, \sigma^2),$$

where $x_{ij(s)}^{\text{case}}$ represents the edge linking node i and j for the s th subject in case group, and $x_{ij(s')}^{\text{control}}$ defines the edge weight for the s' th subject in control group.

We apply various standard effect sizes (i.e., signal-to-noise ratios - SNRs) by setting $\sigma = 1$, and $\mu_0 = 0$, $\mu_1 = 0.4, 0.6, 0.8$. We further consider a more realistic scenario by letting the proportion q_1 of edges inside informative-subgraph be non-differentially expressed (i.e. $N(\mu_0, \sigma^2)$ for both cases and controls). Similarly, we set a q_2 proportion of edges outside informative-subgraph are differentially expressed (i.e. $N(\mu_1, \sigma^2)$ for cases and $N(\mu_0, \sigma^2)$ for controls). (q_1, q_2) represent the practical non-perfect distribution of informative edges in the overall graph. In the simulation data, two sets of parameters $(q_1, q_2) = (0.8, 0.1)$ and $(0.9, 0.05)$ are used.

We compare the ADSD method with the two most popular dense subgraph discovery methods including Greedy algorithm with $\lambda = 1$ and Goldberg's algorithm. The results are evaluated by node-assignment accuracy in terms of true positive rate (TP) and true negative rate (TN) defined as follows:

$$TP = \frac{\sum_{i=1}^n I(\theta_i = \hat{\theta}_i = 1)}{\sum_{i=1}^n I(\theta_i = 1)}, \quad TN = \frac{\sum_{i=1}^n I(\theta_i = \hat{\theta}_i = 0)}{\sum_{i=1}^n I(\theta_i = 0)}$$

The mean and standard errors of TP and TN for three methods across 30 replicates for all settings are displayed in the following Tables 1 and 2.

Table 1 demonstrates results with $(q_1, q_0) = (0.8, 0.1)$. When the subnetwork size is 15, the Charikar's greedy algorithm and Goldberg's algorithm have extremely high values of true positive rate and low true negative rate for all small and large effect sizes/SNR. This indicates that the conventional algorithms tend to detect over-sized dense subgraphs consisting of a high proportion of false positive edges when the subgraph is small and SNR is medium-

large. This phenomenon is common in brain connectome data analysis in practice: the size of disease-related subnetwork is relatively small (the proportion of non-null distribution is small) with medium edge-wise effect sizes which can be disturbed by false positive and positive signals. If a universal threshold is used for all edges (e.g., BH-FDR and FWER cutoffs), then the most true positive edges will be missed along with the topological structure due to the control of false positive findings. In contrast, the proposed ADSD approach is robust to the noise and can accurately extract the informative subgraph. With increasing strengths of signal (i.e., larger informative subnetwork size and SNR), the traditional dense graph discovery algorithms are comparable to ADSD and they all outperform edge-wise inference (e.g., FDR and FWER). Table 2 shows the results with higher accuracy of subnetworks $(q_1, q_0) = (0.9, 0.05)$. The ADSD can more accurately recover the dense subgraph when the SNR is small-medium and robust to false positive and false negative edges. Similarly, all three methods show comparably perfect performance when the subnetwork size and/or SNR are large.

In summary, the simulation results clearly show that likelihood-based ADSD approach is more robust to both false positive and false negative noise and can better capture smaller subnetworks with a high sensitivity and a low false positive rate. These properties are critical for the brain connectome analysis in practice because the real data sets are often mixed with substantial noise and include a small proportion of signal edges.

6 Discussions

In this article, we compare brain connectome matrices between diagnostic groups (e.g. schizophrenia and healthy subjects) to understand connectivity patterns altered by psychiatric illness. As in our motivation data example, however, disease-related subnetworks can be overwhelmed by substantial noise in the connectome data and thus difficult to extract. The noise heavily influences statistical inference by introducing enormous edge-wise false positive and negative errors that are constrained in a weighted adjacency matrix, and thus impose difficulty in understanding the network topology of disease-related brain circuits and in yielding valid statistical inference.

Table 1: The node-assignment accuracy of three methods under varied SNRs and subgraph size with $(q_1, q_0) = (0.8, 0.1)$

		ADSD	Greedy	Goldberg	
15	0.4	TP	0.7400 (0.1604)	0.9900 (0.0238)	0.9767 (0.0382)
		TN	0.5812 (0.2151)	0.1029 (0.0455)	0.1194 (0.0450)
	0.6	TP	0.8833 (0.1152)	1 (0)	0.9900 (0.0238)
		TN	0.9129 (0.1500)	0.0729 (0.0301)	0.0865 (0.0339)
	0.8	TP	0.9633 (0.0536)	1 (0)	0.9900 (0.0238)
		TN	0.9929 (0.0114)	0.0647 (0.0325)	0.0776 (0.0346)
30	0.4	TP	0.9167 (0.0695)	0.9967 (0.0100)	0.9833 (0.0197)
		TN	0.8900 (0.0700)	0.3286 (0.1700)	0.3307 (0.1736)
	0.6	TP	0.9967 (0.0100)	0.9967 (0.0100)	0.9833 (0.0197)
		TN	0.9914 (0.0139)	0.9864 (0.0245)	0.9807 (0.0474)
	0.8	TP	1 (0)	1 (0)	0.9867 (0.0163)
		TN	1 (0)	1 (0)	1 (0)

Table 2: The node-assignment accuracy of three methods under varied SNRs and subgraph size with $(q_1, q_0) = (0.9, 0.05)$

		ADSD	Greedy	Goldberg	
15	0.4	TP	0.7900 (0.1387)	0.9967 (0.0145)	0.9867 (0.0267)
		TN	0.7929 (0.2046)	0.1653 (0.0607)	0.1924 (0.0666)
	0.6	TP	0.9767 (0.0382)	0.9933 (0.0200)	0.9833 (0.0357)
		TN	0.9935 (0.0102)	0.7435 (0.3866)	0.7929 (0.3515)
	0.8	TP	1 (0)	0.9933 (0.0200)	0.9833 (0.0289)
		TN	0.9994 (0.0026)	1 (0)	1 (0)
30	0.4	TP	0.9500 (0.0489)	0.9633 (0.0378)	0.9600 (0.0374)
		TN	0.9586 (0.0686)	0.9557 (0.0686)	0.9443 (0.0868)
	0.6	TP	1 (0)	1 (0)	0.9867 (0.0163)
		TN	0.9979 (0.0051)	1 (0)	1 (0)
	0.8	TP	1 (0)	1 (0)	0.9867 (0.0163)
		TN	1 (0)	1 (0)	1 (0)

To overcome these challenges, we develop a novel ADSD method to reliably and robustly identify signal subgraphs (related to the phenotypes of interest) from the whole brain connectome network. The overall brain connectome inference network is often over-sized with a small proportion of signal edges which are not compatible with existing statistical network models. Therefore, it is desirable to detect a dense subnetwork maintaining most signal edges in a clique with a much smaller number of nodes (nodes induced subnetwork) and discarding a large proportion of false positive edges from the overall network. Dense graph discovery has been a popular research topic in network analysis for a couple of decades. Dense graph discovery methods are distinct from existing statistical methods for network analysis (e.g. various versions of community detection) because they focus on a network with a far fewer number of connections than a highly connected network consisting of communities. The dense graph discovery method is well suited for our application because the number of edges from the non-null distribution is relatively small (Efron, 2012). A key limitation of the current dense graph discovery methods is sensitive to noise. Due to the substantial noise in brain connectome data, the existing dense graph discovery methods tend to extract over-sized dense subgraphs which can lead to a high FDR, potentially incorrect biological findings, and low replicability. The proposed ADSD method integrates the concept of shrinkage into dense graph discovery by introducing a balance parameter to include the most informative edges into the subgraph (high sensitivity) while maintaining a low FDR. The balance parameter can be estimated based on the likelihood function which is commonly used in network statistics. We develop efficient algorithms to implement the objective function that is compatible with computationally intensive inference methods (e.g., permutation tests and bootstraps) . In the current research, we apply permutation test based statistical inference on the dense subgraph. Both the simulation and data example results show that the proposed method is robust to the false negative and positive edges and can accurately detect the target dense subgraph with high sensitivity and low false positive rates. Therefore, our goal of brain connectome analysis can be well met by applying ADSD.

Our work makes several contributions to the field: first, the ADSD objective function and algorithms provide new dense subgraph detection tools for noisy, weighted, large, and less dense graphs, which may have wide applications in data mining and knowledge discovery.

Secondly, we derive theoretical results to provide the bounds for the approximation of ADSD algorithms in a full range of the balance parameter. The asymptotic property of subgraph detection and balance parameter estimation are also developed. Last, the biological findings are novel, integrative, and clinically meaningful. Although part of these findings has been found in previous studies, only edge-wise results (i.e. links between regions to a fixed seed) are reported without fully investigating the interactive nature of network-level inference.

In this article, the hypo-connections in the salience network centered subnetwork groups in patients with schizophrenia are detected for the first time by whole brain connectome network analysis with explicit network topology. The reported network reveals the novel links between aberrant functional connectivity networks and impaired capability to integrate information from multiple sources (cognition deficits) in patients with schizophrenia, which may assist to further understand the underlying biological mechanism for multiple schizophrenic disorder symptoms.

In summary, we develop a likelihood-based adaptive dense graph detection method to extract the dense subgraph from a large and noisy network (weighted and/or binary). Our ADSD method outperforms existing dense subgraph discovery methods when the overall graph includes a small proportion of edges with high importance levels, and thus is well-suited for group-level brain connectome analysis. ADSD can also serve as a screening step for group level network analysis to effectively extract a dense subnetwork from a large overall network for further analysis. In addition, ADSD can be applied to other biological network data (e.g. interactive networks of genomics and proteomics data) and yield findings revealing latent and complex co-expression subnetworks. Therefore, ADSD can become a new useful tool for statistical analysis of large and less dense networks.

References

Adhikari, B. M., L. E. Hong, H. Sampath, J. Chiappelli, N. Jahanshad, P. M. Thompson, L. M. Rowland, V. D. Calhoun, X. Du, S. Chen, et al. (2019). Functional network connectivity impairments and core cognitive deficits in schizophrenia. *Human brain mapping* 40(16), 4593–4605.

- Amini, A. A., A. Chen, P. J. Bickel, E. Levina, et al. (2013). Pseudo-likelihood methods for community detection in large sparse networks. *The Annals of Statistics* 41(4), 2097–2122.
- Asahiro, Y., K. Iwama, H. Tamaki, and T. Tokuyama (1996). Greedily finding a dense subgraph. In *Scandinavian Workshop on Algorithm Theory*, pp. 136–148. Springer.
- Biswal, B. B., M. Mennes, X.-N. Zuo, S. Gohel, C. Kelly, S. M. Smith, C. F. Beckmann, J. S. Adelstein, R. L. Buckner, S. Colcombe, et al. (2010). Toward discovery science of human brain function. *Proceedings of the National Academy of Sciences* 107(10), 4734–4739.
- Bowman, F. D., L. Zhang, G. Derado, and S. Chen (2012). Determining functional connectivity using fmri data with diffusion-based anatomical weighting. *NeuroImage* 62(3), 1769–1779.
- Bullmore, E. T. and D. S. Bassett (2011). Brain graphs: graphical models of the human brain connectome. *Annual review of clinical psychology* 7, 113–140.
- Cao, X., B. Sandstede, and X. Luo (2019). A functional data method for causal dynamic network modeling of task-related fmri. *Frontiers in neuroscience* 13.
- Charikar, M. (2000). Greedy approximation algorithms for finding dense components in a graph. In *International Workshop on Approximation Algorithms for Combinatorial Optimization*, pp. 84–95. Springer.
- Chen, G., P.-C. Bürkner, P. A. Taylor, Z. Li, L. Yin, D. R. Glen, J. Kinnison, R. W. Cox, and L. Pessoa (2019). An integrative bayesian approach to matrix-based analysis in neuroimaging. *Human brain mapping*.
- Chen, S., F. D. Bowman, and Y. Xing (2020). Detecting and testing altered brain connectivity networks with k-partite network topology. *Computational Statistics & Data Analysis* 141, 109–122.
- Chen, S., J. Kang, Y. Xing, and G. Wang (2015). A parsimonious statistical method to detect groupwise differentially expressed functional connectivity networks. *Human brain mapping* 36(12), 5196–5206.
- Chen, S., J. Kang, Y. Xing, Y. Zhao, and D. K. Milton (2018). Estimating large covariance matrix with network topology for high-dimensional biomedical data. *Computational Statistics & Data Analysis* 127, 82–95.
- Chen, S., Y. Xing, J. Kang, P. Kochunov, and L. E. Hong (2020). Bayesian modeling of

- dependence in brain connectivity data. *Biostatistics* 21(2), 269–286.
- Choi, D. S., P. J. Wolfe, and E. M. Airoidi (2012). Stochastic blockmodels with a growing number of classes. *Biometrika* 99(2), 273–284.
- Durante, D., D. B. Dunson, et al. (2018). Bayesian inference and testing of group differences in brain networks. *Bayesian Analysis* 13(1), 29–58.
- Efron, B. (2012). *Large-scale inference: empirical Bayes methods for estimation, testing, and prediction*, Volume 1. Cambridge University Press.
- Erdős, P. and A. Rényi (1960). On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci* 5(1), 17–60.
- Fan, L., H. Li, J. Zhuo, Y. Zhang, J. Wang, L. Chen, Z. Yang, C. Chu, S. Xie, A. R. Laird, et al. (2016). The human brainnetome atlas: a new brain atlas based on connectional architecture. *Cerebral cortex* 26(8), 3508–3526.
- Fornito, A., A. Zalesky, C. Pantelis, and E. T. Bullmore (2012). Schizophrenia, neuroimaging and connectomics. *Neuroimage* 62(4), 2296–2314.
- Ge, T., J. Feng, D. P. Hibar, P. M. Thompson, and T. E. Nichols (2012). Increasing power for voxel-wise genome-wide association studies: the random field theory, least square kernel machines and fast permutation procedures. *Neuroimage* 63(2), 858–873.
- Ghoshdastidar, D. and U. von Luxburg (2018). Practical methods for graph two-sample testing. In *Advances in Neural Information Processing Systems*, pp. 3019–3028.
- Ginestet, C. E., J. Li, P. Balachandran, S. Rosenberg, E. D. Kolaczyk, et al. (2017). Hypothesis testing for network data in functional neuroimaging. *The Annals of Applied Statistics* 11(2), 725–750.
- Goldberg, A. V. (1984). *Finding a maximum density subgraph*. University of California Berkeley, CA.
- Grimmett, G. R. and C. J. McDiarmid (1975). On colouring random graphs. In *Mathematical Proceedings of the Cambridge Philosophical Society*, Volume 77, pp. 313–324. Cambridge University Press.
- Higgins, I. A., S. Kundu, K. S. Choi, H. S. Mayberg, and Y. Guo (2019). A difference degree test for comparing brain networks. *Human brain mapping* 40(15), 4518–4536.
- Holland, P. W., K. B. Laskey, and S. Leinhardt (1983). Stochastic blockmodels: First steps.

- Social networks* 5(2), 109–137.
- Kundu, S., J. Ming, J. Pierce, J. McDowell, and Y. Guo (2018). Estimating dynamic brain functional networks using multi-subject fmri data. *NeuroImage* 183, 635–649.
- Lee, V. E., N. Ruan, R. Jin, and C. Aggarwal (2010). A survey of algorithms for dense subgraph discovery. In *Managing and Mining Graph Data*, pp. 303–336. Springer.
- Li, S., N. Hu, W. Zhang, B. Tao, J. Dai, Y. Gong, Y. Tan, D. Cai, and S. Lui (2019). Dysconnectivity of multiple brain networks in schizophrenia: a meta-analysis of resting-state functional connectivity. *Frontiers in psychiatry* 10, 482.
- Lukemire, J., S. Kundu, G. Pagnoni, and Y. Guo (2017). Bayesian joint modeling of multiple brain functional networks. *arXiv preprint arXiv:1708.02123*.
- Mejia, A. F., M. B. Nebel, Y. Wang, B. S. Caffo, and Y. Guo (2019). Template independent component analysis: Targeted and reliable estimation of subject-level brain networks using big data population priors. *Journal of the American Statistical Association*, 1–27.
- Newman, M. E. and M. Girvan (2004). Finding and evaluating community structure in networks. *Physical review E* 69(2), 026113.
- Palaniyappan, L., T. P. White, and P. F. Liddle (2012). The concept of salience network dysfunction in schizophrenia: from neuroimaging observations to therapeutic opportunities. *Current topics in medicinal chemistry* 12(21), 2324–2338.
- Simpson, S. L., M. Bahrami, and P. J. Laurienti (2019). A mixed-modeling framework for analyzing multitask whole-brain network data. *Network Neuroscience* 3(2), 307–324.
- Tsourakakis, C., F. Bonchi, A. Gionis, F. Gullo, and M. Tsiarli (2013). Denser than the densest subgraph: extracting optimal quasi-cliques with quality guarantees. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 104–112. ACM.
- Uddin, L. Q. (2015). Salience processing and insular cortical function and dysfunction. *Nature Reviews Neuroscience* 16(1), 55–61.
- Vogelstein, J. T., W. G. Roncal, R. J. Vogelstein, and C. E. Priebe (2012). Graph classification using signal-subgraphs: Applications in statistical connectomics. *IEEE transactions on pattern analysis and machine intelligence* 35(7), 1539–1551.
- Xia, Y. and L. Li (2017). Hypothesis testing of matrix graph model with application to brain

- connectivity analysis. *Biometrics* 73(3), 780–791.
- Xia, Y. and L. Li (2018). Matrix graph hypothesis testing and application in brain connectivity alternation detection. *Statistica Sinica*, to appear.
- Zalesky, A., A. Fornito, and E. T. Bullmore (2010). Network-based statistic: identifying differences in brain networks. *Neuroimage* 53(4), 1197–1207.
- Zhang, Y., E. Levina, and J. Zhu (2017). Estimating network edge probabilities by neighbourhood smoothing. *Biometrika* 104(4), 771–783.
- Zhao, Y., E. Levina, J. Zhu, et al. (2012). Consistency of community detection in networks under degree-corrected stochastic block models. *The Annals of Statistics* 40(4), 2266–2292.

7 Appendix A

7.1 Permutation Tests

We could evaluate the significance of detected subgraph by permutation testing (Zalesky et al., 2010). For an undirected graph $G = (V, E)$, the edge permutation permutes the order of edges. The edge-permuted graph according to a permutation π is given by $G_\pi = (V, \pi(E))$, with $\{e_{ij} \in E\} = \{e_{\pi(i), \pi(j)} \in E_\pi\}$. For the randomness of permutation, we have $\{e_{ij} \in E\} \perp \{e_{ij} \in E_\pi\}$. Therefore, the organized pattern of G does not exist for G_π , and the edge-permuted graph G_π becomes a random graph with probability being the proportion of connected edges in G (i.e. $\rho = \frac{\sum_{i < j} I(\delta_{ij}=0)}{|E|}$). Hence, the significance of the topological pattern of G can be evaluated among random graphs with identical proportion of connected edges G_π .

In a weighted graph $G = (V, E, W)$ with $|V| = n$, let $\phi(\cdot)$ be the half-vectorization of a symmetric matrix. Denote π as a permutation of $\binom{n}{2}$ elements, and P_π is the corresponding permutation matrix. Then, the weight matrix of the edge-permuted graph $G_\pi = (V, E_\pi, W_\pi)$ is given by $W_\pi = \phi^{-1}(P_\pi \phi(W))$.

A true informative subgraph includes edges with higher average level of importance (i.e. p-values) which can not be extract from a random graph. Hence, we propose a network level statistic such that it prefers subgraphs with more informative edges (larger subgraphs that can not be generated by random) and averagely higher level of significant edges (smaller subgraphs by excluding less non-significant edges). For a graph G with informative subgraph $G(S)$, let

$$\bar{z}_G = -2 \sum_{i,j \in G(S)} \log p_{ij} / 2|E(S)|.$$

We propose the test statistic based on Fisher's combination test and Chernoff bound of χ^2 the cumulative distribution function:

$$T_G = |E(S)|(\bar{z}_G - 1 - \log(\bar{z}_G))$$

The whole procedure of the permutation testing is as follows:

Algorithm 3 Permutation testing of extracted subgraph

```

1: procedure ALGORITHM
2:   calculate the test statistic on  $G$  with informative subgraph  $G(S)$  and denote as:  $T_0$ 
3:   for  $m = 1$  to  $M$  do
4:     generate permutation matrix  $P_m$  on  $\binom{n}{2}$  elements
5:     observe weight matrix of edge-permuted graph  $G_m$ :  $W_m = \phi^{-1}(P_m\phi(W))$ 
6:     perform Algorithm 2 to detect a informative subgraph and denote the nodes set
       as  $S_m$ 
7:     calculate the test statistic on  $G_m$  with informative subgraph  $G_m(S_m)$  as:  $T_m$ 
8:   end for
9:    $p_0$  is the percentile of  $T_0$  in  $\{T_m\}_{m=1}^M$ ;
10:  if  $p_0$  is smaller than  $\alpha$  then
11:     $G(S)$  is a significant subgraph of  $G$ .
12:  end if
13: end procedure

```

7.2 Related work in densest subgraph

Due to noise, dense graph discovery algorithms tend to either identify over-sized subgraphs that may include a large proportion of false positive edges with low importance levels (high false positive discovery rate) or detect over-conservative small-sized subgraphs that may not sufficiently cover edges with high importance levels (low sensitivity, [Tsourakakis et al., 2013](#)). The two following lemmas are constructed to demonstrate the cases where conventional dense discovery methods may lead to detecting trivial-sized subgraphs. The generative mechanism is considered to be the same as stated previously.

Lemma 1 (Densest subgraph regarding average degree f_1). *Assume that a graph $G = (V, E)$ with an induced dense subgraph $G(S_0) = (S_0, E(S_0))$ is generated by the same joint distribution as Theorem 3 with $n^{1/2+\epsilon} \lesssim n_s \leq cn$. Let $\pi_s > \pi_0 > \epsilon_1$ and $c\pi_0 - c^2(\pi_s - \pi_0) > 2c_1n^{-\epsilon}$ be some positive constant ϵ_1 and c_1 . Then, denoting the densest subgraph as \tilde{S}_1 , we have*

$$\mathbb{P}\left(|\tilde{S}_1| > n - n^{2\epsilon}\right) \rightarrow 1$$

Lemma 2 (Densest subgraph regarding edge ratio f_2). *Assume that a graph $G = (V, E)$ with an induced dense subgraph $G(S_0) = (S_0, E(S_0))$ is generated by the same joint distribution as Theorem 3 with $n^{1/2+\epsilon} \lesssim n_s \leq cn$. Let $\pi_s > \pi_0 > \epsilon_1$. Then, for the densest subgraph \tilde{S}_2 ,*

we have

$$\mathbb{P}\left(|\tilde{S}_2| < \frac{2\log(n)}{\log(1/\pi_s)}\right) \rightarrow 1$$

almost surely as $n \rightarrow \infty$.

Note for lemma 2, the upper bound is relatively small compared with the size of graph. For example, for $n = 1000$ and $\pi_s = 0.5$, $2\log n/\log(1/\pi_s) \approx 19.93$, and for $n = 10000$ and $\pi_s = 0.7$, $2\log n/\log(1/\pi_s) \approx 51.65$. The proofs for these two lemmas are included in the appendix. Therefore, the direct application of conventional dense subgraph discovery algorithms to our network data may miss the underlying important findings of disease-related subnetworks.

7.3 Proofs

Before we prove the Theorem 1, we first establish the following Lemma 3 by applying the idea of proof in Charikar (2000) with density function $f_1(S)$ directly.

Lemma 3. *The greedy algorithm will give a $2n^{|\lambda-1|}$ approximation, i.e. $f(S^*) \leq 2n^{|\lambda-1|}v$.*

Proof. For an undirected graph, we assign the edge e_{ij} to either node i or j . Denote the number of edges assigned to node i as $d(i)$. Then,

$$\sum_{i \in S^*} d(i) \geq |E(S^*)|,$$

since all edges in $E(S^*)$ will be assigned to node i or j which will be included in $\sum_{i \in S^*} d(i)$.

Consider a specific way to assign edges. Let the node gets assigned when it is removed in the greedy algorithm. In other words, $d(i)$ equals the degree of node i in the iteration that it is removed in greedy algorithm. Assume the subgraph at the iteration that i is removed is $G(S')$ such that at this iteration the nodes set changes from $S' \rightarrow S'/\{i\}$. Therefore, since the node i is deleted for the smallest degree, we will have the degree of i is smaller than the average degree in this iteration:

$$d(i) \leq 2 \frac{|E(S')|}{|S'|}.$$

Case 1: For $\lambda < 1$,

$$d(i) \leq 2 \frac{|E(S')|}{|S'|} \leq 2 \frac{|E(S')|}{|S'|^\lambda} \leq 2 \frac{|E(\tilde{S})|}{|\tilde{S}|^\lambda} = 2v,$$

$$\Rightarrow f(S^*) = \frac{|E(S^*)|}{|S^*|^\lambda} \leq \frac{\sum_{i \in S^*} d(i)}{|S^*|^\lambda} \leq 2v \frac{|S^*|}{|S^*|^\lambda} \leq 2vn^{1-\lambda}.$$

Case 2: For $\lambda > 1$,

$$d(i) \leq 2 \frac{|E(S')|}{|S'|} \leq 2 \frac{|E(S')|}{|S'|^\lambda} \times |S'|^{\lambda-1} \leq 2 \frac{|E(\tilde{S})|}{|\tilde{S}|^\lambda} \times |S'|^{\lambda-1} = 2vn^{\lambda-1},$$

$$\Rightarrow f(S^*) = \frac{|E(S^*)|}{|S^*|^\lambda} \leq \frac{\sum_{i \in S^*} d(i)}{|S^*|^\lambda} \leq 2vn^{\lambda-1} \frac{|S^*|}{|S^*|^\lambda} \leq 2vn^{\lambda-1}.$$

□

7.3.1 Proof of theorem 1

Case 1: $\lambda \geq 2$. The subgraph with only two nodes has $f(S; \lambda) = 1/2^\lambda$, otherwise all elements in this 2×2 matrix should be zero and inductively the raw graph have no edges (all elements in the corresponding matrix should be zero). Thus, $f(\tilde{S}; \lambda) \geq 1/2^\lambda$.

On the other hand, when $\lambda \geq 2$,

$$f(S^*; \lambda) = \frac{|E(S^*)|}{|S^*|^\lambda} \leq \frac{|S^*|^2}{2|S^*|^\lambda} \leq \frac{1}{2}.$$

Hence, $f(\tilde{S}; \lambda) \geq 2^{1-\lambda} f(S^*; \lambda)$ and $c(\lambda) = 2^{\lambda-1}$.

Case 2: $1 < \lambda < 2$. Assume there exists $\lambda > 0$ such that $f(S^*; \lambda) \leq 2n^\lambda f(\tilde{S}; \lambda)$ and we want to find such λ . It suffices to show

$$f(S^*; \lambda) = \frac{|E(S^*)|}{|S^*|^\lambda} \leq \frac{|S^*|^2}{2|S^*|^\lambda} \leq 2 \frac{1}{2^\lambda} n^\lambda$$

since from case 1, $f(\tilde{S}; \lambda)$ has a lower bound $1/2^\lambda$. It's automatically true for $|S^*| \leq (2^{2-\lambda} n^\lambda)^{1/(2-\lambda)} = 2n^{\lambda/(2-\lambda)}$.

For $|S^*| > 2n^{a/(2-\lambda)}$, from the proof of case 2 in lemma 3, we have

$$\begin{aligned} f(S^*; \lambda) &= \frac{E(S^*)}{|S^*|^\lambda} \leq \frac{\sum_{i \in S^*} d(i)}{|S^*|^\lambda} \leq 2f(\tilde{S}; \lambda)n^{\lambda-1}|S^*|^{1-\lambda} \\ &\leq 2f(\tilde{S}; \lambda)n^{\lambda-1}(2n^{a/(2-\lambda)})^{1-\lambda} \end{aligned}$$

Let $2f(\tilde{S}; \lambda)n^{\lambda-1}(2n^{a/(2-\lambda)})^{1-\lambda} = 2f(\tilde{S}; \lambda)n^a$, we will get $a = (\lambda - 1)(2 - \lambda)$.

Hence, $f(S^*; \lambda) \leq 2n^{(\lambda-1)(2-\lambda)}(1 \vee 2^{1-\lambda})f(\tilde{S}; \lambda)$, then $c'(\lambda) = 1 \vee 2^{1-\lambda}$.

Case 3: For $0.5 < \lambda < 1$, the claim is automatically true from lemma 3.

Case 4: $0 < \lambda < 0.5$

$$\begin{aligned} f(S^*; \lambda) &= \frac{|E(S^*)|}{|S^*|^\lambda} \leq \frac{|E(V)|}{|S^*|^\lambda} = \frac{|E(V)|}{|S^*|^\lambda |V|^\lambda} \times |V|^\lambda \\ &\leq f(\tilde{S}; \lambda) \frac{|V|^\lambda}{|S^*|^\lambda} \leq 2f(\tilde{S}; \lambda)n^\lambda \end{aligned}$$

7.3.2 Proof of theorem 2

Part 1. Let $\mathcal{C} = \{S_1, S_2, \dots, S_{n-1}\}$ be the sequence of subgraphs generated by deleting the smallest-degree node in Algorithm 1. We first prove the true subgraph S_0 is included in \mathcal{C} up to a permutation with high probability as $n \rightarrow \infty$. Denote $n = |V|$ and $n_s = |S_0|$.

At stage $k (< n - n_s)$ of Algorithm 1, for $i \in S_0$, and $i' \in V/S_0$, we have

$$d_i = \sum_{j=1}^{n_s} X_{ij} + \sum_{j=n_s+1}^{n-k+1} Y_{ij} \quad \text{and} \quad d_{i'} = \sum_{j=1}^{n-k+1} Y_{i'j}$$

where $X_{ij} \sim \text{Bernouli}(\pi_s)$, $Y_{ij} \sim \text{Bernouli}(\pi_0)$, and $Y_{i'j} \sim \text{Bernouli}(\pi_0)$

From chernoff bound, for $i \in S_0$, $i' \in V/S_0$ and $\delta \in (0, 1)$

$$\mathbb{P}(d_i \leq (1 - \delta)(n_s\pi_s + (n - k - n_s)\pi_0)) \leq \exp \left[-\frac{\delta^2}{2}(n_s\pi_s + (n - k - n_s)\pi_0) \right],$$

and

$$\mathbb{P}(d_{i'} \geq (1 + \delta)(n - k)\pi_0) \leq \exp \left[-\frac{\delta^2}{2 + \delta^2}(n - k)\pi_0 \right].$$

Hence, there exist $\epsilon_0(\delta) > 0$, such that

$$\mathbb{P}(d_i - d_{i'} \geq \epsilon_0(\delta)) \geq 1 - \exp \left[-\frac{\delta^2}{2}(n_s \pi_s + (n - k - n_s) \pi_0) \right] - \exp \left[-\frac{\delta^2}{2 + \delta^2}(n - k) \pi_0 \right].$$

At stage $k (< n - n_s)$ of Algorithm 1, the event of deleting a node outside S_0 is:

$$\bigcap_{i \in S_0} \bigcap_{i' \in S_k/S_0} \{d_i \geq d_{i'}\} = \bigcap_{i \in S_0} [\bigcup_{i' \in S_k/S_0} \{d_i < d_{i'}\}]^c$$

Then,

$$\begin{aligned} \mathbb{P} \left(\bigcap_{i \in S_0} \bigcap_{i' \in S_k/S_0} \{d_i \geq d_{i'}\} \right) &\geq 1 - \sum_{i \in S_0} \mathbb{P} \left(\bigcup_{i' \in S_k/S_0} \{d_i < d_{i'}\} \right) \\ &\geq 1 - \sum_{i \in S_0} \sum_{i' \in S_k/S_0} \mathbb{P}(d_i < d_{i'}) \\ &\geq 1 - n_s(n - k - n_s) \left\{ \exp \left[-\frac{\delta^2}{2}(n_s \pi_s + (n - k - n_s) \pi_0) \right] + \exp \left[-\frac{\delta^2}{2 + \delta^2}(n - k) \pi_0 \right] \right\} \end{aligned}$$

Using similar argument, the class \mathcal{C} includes the true subgraph up to a permutation Q :

$$\begin{aligned} \mathbb{P}(Q(S_0) \in \mathcal{C}) &= \mathbb{P} \left(\bigcap_{k=1}^{n-n_s-1} \bigcap_{i \in S_0} \bigcap_{i' \in S_k/S_0} \{d_i \geq d_{i'}\} \right) \\ &\geq 1 - (n - n_s - 1)n_s(n - k - n_s) \times \\ &\quad \left\{ \exp \left[-\frac{\delta^2}{2}(n_s \pi_s + (n - k - n_s) \pi_0) \right] + \exp \left[-\frac{\delta^2}{2 + \delta^2}(n - k) \pi_0 \right] \right\} \\ &\rightarrow 1 \text{ as } n \rightarrow \infty \text{ and } n_s = O(n^{1/2+\epsilon}). \end{aligned}$$

Part 2. We then establish the true subgraph can be selected from \mathcal{C} by density function $f(S; \lambda)$ for some λ . For $k < n - n_s$, from the proof of part 1, $S_0 \subset S_k$ with high probability. From Chernouff bounds,

$$\begin{aligned} \mathbb{P} \left(2|A(S_k)| - 2|A(S_0)| \leq (1 + \delta)[(n - k - 1)^2 - (n - n_s - k - 1)^2] \pi_0 \right) \\ \geq 1 - \exp \left[-\frac{\delta^2}{2 + \delta^2} [(n - k - 1)^2 - (n - n_s - k - 1)^2] \pi_0 \right] \end{aligned}$$

and

$$\mathbb{P}\left(2|A(S_0)| > (1 - \delta)n_s^2\pi_s\right) \geq 1 - \exp\left[-\frac{\delta^2}{2}n_s^2\pi_s\right].$$

Thus, with high probability,

$$\frac{|A(S_k)| - |A(S_0)|}{|A(S_0)|} < \frac{1 + \delta}{1 - \delta} * \frac{[(n - k - 1)^2 - (n - n_s - k - 1)^2]\pi_0}{n_s^2\pi_s}$$

On the other hand,

$$\frac{|A(S_k)|}{(n - k - 1)^\lambda} \leq \frac{|A(S_0)|}{(n_s)^\lambda} \Leftrightarrow \frac{|A(S_k)| - |A(S_0)|}{|A(S_0)|} \leq \frac{(n - k - 1)^\lambda}{(n_s)^\lambda} - 1,$$

hence, it suffices to have

$$\frac{1 + \delta}{1 - \delta} * \frac{[(n - k - 1)^2 - (n - n_s - k - 1)^2]\pi_0}{n_s^2\pi_s} \leq \frac{(n - k - 1)^\lambda}{(n_s)^\lambda} - 1.$$

For $\pi_s > \pi_0$, there exist corresponding δ and λ to make the inequality holds.

For $k > n - n_s$, we could use similar argument and the claim is true.

7.3.3 Proof of theorem 3

We implement the proof of SBM under maximum likelihood fitting by [Choi et al. \(2012\)](#) that constrain stochastic block model with $K \lesssim n^{1/2}$ communities and the average degree $M \gtrsim (\log(n))^{3+\delta}$. The growth restriction on K is utilized to bound the number of possible choices of assignment. Although the restriction is not satisfied in our model, the assumption such that all edges outside the dense subgraph are considered as singletons makes the number of possible assignments being the same as $K = 2$. The restriction on average degree is also automatically satisfied for fixed π_s and π_0 .

Next, we prove the conclusion is valid when the assignment is maximized over a smaller class of possible solutions, which is generated by different values of λ .

The Theorems 1 and 2 in [Choi et al. \(2012\)](#) also hold in our model because the maximization over a subset of parameter space is smaller than over the whole space. For their theorem 3, from the proof of our theorem 2, the true assignment is in the subset that we

maximized with probability converging to 1, i.e. $\boldsymbol{\theta}(S_0) \in \{\boldsymbol{\theta}(\tilde{S}_\lambda), \lambda \in (0, 2)\}$. Then theorem 3 holds with probability converging to 1, i.e. $\bar{L}_P(\boldsymbol{\theta}(S_0)) - \bar{L}_P(\hat{\boldsymbol{\theta}}) = o_p(M)$. Hence, our claim is true. \square

7.3.4 Proof of Lemma 1

It's equivalent to show

$$\mathbb{P}(\forall m < n - n^{2\varepsilon}, f_1(S_{n-m}) > f_1(G)) \leq 2 \exp(-n^{2\varepsilon}).$$

Consider the case $n_s = cn$ for some constant $0 < c < 1$. We first establish $\forall m < n - n^{2\varepsilon}$,

$$\mathbb{P}\left(\frac{|S_{n-m}|}{m} > \frac{|G|}{n}\right) \leq 2 \exp(-n^{2\varepsilon}).$$

Case 1: Consider subgraph S_{n-m} with $m > n_s$. We have

$$\mathbb{E}(|S_{n-m}|) = \binom{n_s}{2} \pi_s + \left(\binom{m}{2} - \binom{n_s}{2} \right) \pi_0,$$

$$\text{and } \mathbb{E}(|S_n|) = \binom{n_s}{2} \pi_s + \left(\binom{n}{2} - \binom{n_s}{2} \right) \pi_0.$$

Then, from chernoff bound, we have

$$\mathbb{P}(|S_{n-m}| > (1 + \delta)\mathbb{E}(|S_{n-m}|)) \leq \exp\left(-\frac{\delta^2}{2 + \delta^2} \mathbb{E}(|S_{n-m}|)\right),$$

$$\text{and } \mathbb{P}(|S_n| < (1 - \delta)\mathbb{E}(|S_n|)) \leq \exp\left(-\frac{\delta^2}{2} \mathbb{E}(|S_n|)\right),$$

where the second probability is even smaller if we consider the correlation between $|S_{n-m}|$ and $|G|$.

Hence, if we choose $\delta = n^\varepsilon/n$, each of the above probability is bounded by $\exp(-n^{2\varepsilon})$, in this case, when

$$\frac{n\mathbb{E}(|S_{n-m}|)}{m\mathbb{E}(|G|)} \times \frac{1 + n^\varepsilon/n}{1 - n^\varepsilon/n} \leq 1,$$

we have

$$\mathbb{P}\left(\frac{|S_{n-m}|}{m} > \frac{|G|}{n}\right) \leq 2 \exp(-n^{2\varepsilon}).$$

Since $\frac{1+n^\varepsilon/n}{1-n^\varepsilon/n} \approx 1 + c_1 n^{-1+\varepsilon}$, it's equivalent to have

$$\frac{m\mathbb{E}(|G|)}{n\mathbb{E}(|S_{n-m}|)} - 1 \geq c_1 n^{-1+\varepsilon}.$$

Then,

$$\begin{aligned} LHS &\geq \frac{1}{n^3} [m\mathbb{E}(|G|) - n\mathbb{E}(|S_{n-m}|)] \\ &= \frac{1}{n^3} \left\{ m \left[\binom{n_s}{2} \pi_s + \left(\binom{n}{2} - \binom{n_s}{2} \right) \pi_0 \right] - n \left[\binom{n_s}{2} \pi_s + \left(\binom{m}{2} - \binom{n_s}{2} \right) \pi_0 \right] \right\} \\ &= \frac{1}{2n^3} [n_s(n_s - 1)(\pi_s - \pi_0)(m - n) + \pi_0(n - m)mn] \\ &= \frac{1}{2n^3} (n - m) [\pi_0 mn - (\pi_s - \pi_0)n_s(n_s - 1)]. \end{aligned}$$

In order to have $LHS \geq c_1 n^{-1+\varepsilon}$, under condition $n - m > n^{2\varepsilon}$, it's sufficient to have $c\pi_0 - c^2(\pi_s - \pi_0) > 2c_1 n^{-\varepsilon}$.

Case 2: Consider $m \leq n_s$. If

$$\frac{m(m-1)}{2} < \frac{1}{4} \frac{n(n-1)}{2} \pi_0,$$

in other words, $m < c_2 n$ for some c_2 , we have $\max |S_{n-m}| \leq \frac{1}{4} \mathbb{E}(|G|)$, then let $\delta = 3/4$,

$$\mathbb{P}\left(\frac{|S_{n-m}|}{m} > \frac{|G|}{n}\right) \leq \exp(-n^\varepsilon).$$

If $c_2 n \leq m \leq n_s = cn$, using similar argument as Case 1, the claim is true.

The result is true if we take union among all possible m . The case with $n_s = o(n)$ is automatically true. Hence, the claim is proved.

7.3.5 Proof of Lemma 2

Since $\mathcal{C} = \{S_1, S_2, \dots, S_{n-2}\}$ is the sequence of subgraphs selected by Greedy Algorithm, for any subgraph in the sequence S_i , the next subgraph S_{i-1} is generated by deleting the node with smallest degree. Without loss of generality, we assume the node deleted from S_i is node u . Then, $\deg_{S_i}(u) \leq 2\frac{|A(S_i)|}{|S_i|}$, and also, $|A(S_{i-1})| = |A(S_i)| - \deg_{S_i}(u)$. Hence, $|A(S_{i-1})| \geq (1 - 2/|S_i|)|A(S_i)|$, and then

$$f_2(S_{i-1}) = 2\frac{|A(S_{i-1})|}{|S_{i-1}|(|S_{i-1}| - 1)} \geq 2\frac{|S_i| - 2}{|S_i|} \frac{|A(S_i)|}{(|S_i| - 1)(|S_i| - 2)} = f_2(S_i).$$

Therefore, the sequence $\{S_1, S_2, \dots, S_{n-2}\}$ has nondecreasing values of objective function f_2 , i.e. $f_2(S_{n-2}) \leq \dots \leq f_2(S_2) \leq f_2(S_1)$. The maximizer of $f_2(\cdot)$ has the same value of objective function for all smaller subgraphs in the sequence $\{S_1, S_2, \dots, S_{n-2}\}$.

On the other hand, the smallest subgraph S_{n-2} can only be two cases: $(0, 0; 0, 0)$ or $(0, 1; 1, 0)$, which has $f_2(S_{n-2})$ being 0 or 1, respectively. We don't consider the case with $f_2(S_{n-2}) = 0$ in general, since from the nondecreasing property, we conclude all subgraphs in the sequence has objective function being 0. For $f_2(S_{n-2}) = 1$, the nondecreasing property results that the maximizer of $f_2(\cdot)$ and all smaller subgraphs have objective function being 1.

Therefore, the densest subgraph must be a clique generated from the graph G . Based on the results in [Grimmett and McDiarmid \(1975\)](#), for any random graph $G(n, p)$, the size of largest clique $K_n(p)$ in this random graph satisfy,

$$\frac{K_n(p)}{\log(n)} \rightarrow \frac{2}{\log(1/p)},$$

almost surely as $n \rightarrow \infty$

Since the largest clique generated from the graph G with informative subgraph $G(S_0)$ is smaller than generated from random graph $G(n, \pi_s)$ with high probability. The densest subgraph returned by Greedy Algorithm in f_2 is bounded by $2\log(n)/\log(1/\pi_s)$.

8 Appendix B: Region Names

8.1 Region names

In the following tables, we list the region names and coordinates of subgraph extracted from the data of schizophrenia research.

Web Table 1: Region names and coordinates for the nodes in the detected subgraph

Abbreviation	Region Names	x	y	z
SFG.R.7.1	Superior Frontal Gyrus	7	16	54
SFG.L.7.5	Superior Frontal Gyrus	-6	-5	58
SFG.R.7.5	Superior Frontal Gyrus	7	-4	60
OrG.R.6.1	Orbital Gyrus	6	47	-7
PCL.L.2.1	Paracentral Lobule	-8	-38	58
PCL.R.2.1	Paracentral Lobule	10	-34	54
STG.L.6.2	Superior Temporal Gyrus	-54	-32	12
STG.R.6.2	Superior Temporal Gyrus	54	-24	11
STG.L.6.3	Superior Temporal Gyrus	-50	-11	1
STG.R.6.3	Superior Temporal Gyrus	51	-4	-1
STG.R.6.4	Superior Temporal Gyrus	66	-20	6
STG.L.6.5	Superior Temporal Gyrus	-45	11	-20
STG.R.6.5	Superior Temporal Gyrus	47	12	-20
STG.L.6.6	Superior Temporal Gyrus	-55	-3	-10
STG.R.6.6	Superior Temporal Gyrus	56	-12	-5
IPL.L.6.6	Inferior Parietal Lobule	-53	-31	23
INS.L.6.1	Insular Gyrus	-36	-20	10
INS.R.6.1	Insular Gyrus	37	-18	8
INS.L.6.2	Insular Gyrus	-32	14	-13
INS.L.6.3	Insular Gyrus	-34	18	1
INS.R.6.3	Insular Gyrus	36	18	1
INS.L.6.4	Insular Gyrus	-38	-4	-9
INS.R.6.4	Insular Gyrus	39	-2	-9
INS.L.6.5	Insular Gyrus	-38	-8	8
INS.R.6.5	Insular Gyrus	39	-7	8
INS.L.6.6	Insular Gyrus	-38	5	5
INS.R.6.6	Insular Gyrus	38	5	5
CG.L.7.2	Cingulate Gyrus	-3	8	25
CG.R.7.2	Cingulate Gyrus	5	22	12
CG.L.7.3	Cingulate Gyrus	-6	34	21
CG.R.7.3	Cingulate Gyrus	5	28	27
CG.L.7.5	Cingulate Gyrus	-5	7	37
CG.R.7.5	Cingulate Gyrus	4	6	38
CG.L.7.6	Cingulate Gyrus	-7	-23	41
CG.R.7.6	Cingulate Gyrus	6	-20	40
CG.L.7.7	Cingulate Gyrus	-4	39	-2
CG.R.7.7	Cingulate Gyrus	5	41	6