Graph combinatorics based group-level network inference for brain connectome analysis

Shuo Chen
Department of Epidemiology and Biostatistics,
University of Maryland, College Park
and
Division of Biostatistics and Bioinformatics, School of Medicine,
University of Maryland, Baltimore
Qiong Wu
Department of Mathematics, University of Maryland, College Park
and
L. Elliot Hong
Maryland Psychiatric Research Center, Department of Psychiatry,
School of Medicine, University of Maryland, Baltimore

September 4, 2019

Abstract

We consider group-level statistical inference for networks, where outcomes are multivariate edge variables constrained in an adjacency matrix. The graph notation is used to represent the network outcome variables, where nodes are identical biological units (e.g. brain regions) shared across subjects and edge-variables indicate the strengths of the interactive relationships between nodes. The edge-variables vary

across subjects and may be associated with covariates of interest. The statistical inference for multivariate edge-variables is challenging because both localized inference on individual edges and the joint inference of a combinatorial of edges (network-level) are desired. We develop a group-level network inference model to integrate graph theory and combinatorics into group-level network statistical inference. We first propose an objective function with ℓ_0 norm regularization to capture latent subgraphs/subnetworks accurately by suppressing false positive edges. We next statistically test each detected subnetwork using graph combinatorics based statistical inferential procedure. We apply the proposed approach to a brain connectome study to identify latent brain functional subnetworks that are associated with brain disorders and verify the findings using an independent replicate data set. The results demonstrate the proposed method outperform existing multivariate statistical methods by simultaneously reducing false positive and false negative discovery rates and increasing replicability.

Keywords: combinatorics, graph theory, graph topology, l_0 norm regularization, network statistics

1 Introduction

There has been an increased interest in statistical literature to model group-level network data. For example, the population level brain connectome studies often aim to investigate whether brain functional and/or structural networks are related to behavioral and symptomatic phenotypes, and the microbiome network studies focus on whether microbial networks are influenced by the clinical status (Lukemire et al., 2017; Xia and Li, 2017; Cai et al., 2018; Simpson et al., 2019; Warnick et al., 2018; Shaddox et al., 2018). In these applications, the data structure of each subject can be represented by a graph notation, where a node represents a well-defined biological unit (e.g. a brain area) and an edge indicates the interactive relationship between a pair of nodes. We consider all nodes are identical across subjects because brain regions are shared across all human subjects.

The outcome variables are edge-variables quantifying the strengths of interactions between nodes (the values of edge-variables), which vary across subjects and can be associated with external phenotypes (e.g. the clinical diagnosis and treatment response). In that, edges are (weighted/continuous or binary) multivariate outcomes that are constrained by a set of nodes in an adjacency matrix. The constraint of graph space distinguishes the statistical inferential procedure of multivariate edge-variables from the conventional high-throughput statistical methods, for example, the commonly used false positive discovery rate (FDR) and LASSO (Benjamini and Hochberg, 1995; Efron et al., 2004) because edges are intrinsically linked with graph theory and combinatorics.

For example, a set of phenotype-related edges can comprise subgraphs/subnetworks with organized topological structures rather than randomly distributed in a graph/network space, and thus the statistical inference should take into account the graph combinatorics besides the effect sizes on individual edges. We define a subnetwork as an induced subgraph

with an organized topological structure (e.g. community) in the graph/network space (here 'subnetwork' and 'subgraph' are exchangeable). Each localized edge exclusively belongs to one subnetwork and each subnetwork is explicitly defined by a set of nodes and edges and the corresponding topological structure. The goal of our group-level network inference to extract and test each latent subnetwork (a combinatorial set of edges) that is interrogatively related to the phenotypes of interest as an object.

In the recent statistical literature, group level network models have been developed for brain network inference using advanced statistical techniques (Craddock et al., 2013; Kim, 2014; Narayan et al., 2015; Ginestet et al., 2017; Zhang et al., 2017; Durante et al., 2018; Xia and Li, 2018; Kundu et al., 2018; Wang et al., 2019; Higgins et al., 2018; Cao et al., 2019 among ohters). These methods have greatly improved the accuracy of statistical inference and yield numerous meaningful biological findings.

In this paper, we focus on integrating graph theory and combinatorics into the statistical inference of multivariate edges by capturing latent subnetworks. The results can reveal which set of localized edges comprising subnetworks are associated with the phenotype of interest. We propose a new statistical framework for **g**roup **le**vel network inference (GLEN) consisting of two steps (Figure 1). First, GLEN implements l_0 norm shrinkage based optimization to extract subgraphs that cover the maximal number of phenotype-related edges by the minimum size (i.e. minimizing the number of edges in the subgraphs). The ℓ_0 norm regularization is critical for accurate subgraph detection because it can suppress the noise of false positive edges outside the latent subnetworks and thus can lead to more accurate subnetwork discovery. The ℓ_0 norm regularization for edge variables are implemented by computationally convenient algorithms. GLEN can efficiently estimate the number of subgraphs, the compositions (nodes and edges) and the topological structures

of the subnetworks. Second, we propose statistical tests based on graph combinatorics for subnetwork-level inference. Intuitively, the combinatoric probability of false positive edges composing of subgraphs in organized topological structures converges to zero almost surely. Therefore, the statistical inference should take into account both effect sizes of edges in the extracted subgraphs and graph combinatorics of the organized topological structures of the subgraphs. The joint application of ℓ_0 norm regularization and graph combinatorics based statistical inference fully capitalizes on the graph theory properties of edge variables and thus improve the accuracy of statistical inference. In addition, GLEN can also be a complement to the existing methods. For example, the locations and topological structures of phenotype-related subgraphs detected by GLEN can become prior information for existing network analysis models (Xia and Li, 2017; Simpson et al., 2019; Xia et al., 2019).

We apply the proposed method to an example data of brain connectome study for schizophrenia research with a primary data set and an independent validation data set. The proposed method along with other existing methods are applied to both data sets separately. We find that the subnetwork identified by GLEN in the primary data set is almost identical to the validation data set, and thus the findings are highly replicable. In contrast, the competitive methods only detect none or a small proportion of edges that are shared by both data sets. These results can be further supported by our simulation results that the false positive and false negative discovery rates are simultaneously reduced by GLEN when phenotype related edges consist of dense subgraphs with organized topological structures. The biological findings using GLEN may also provide insights into neurophysiology and neuropathology.

For example, GLEN can detect that the interconnections between three well-known brain networks (the default mode network, executive network, and salience network) are

associated with a brain disease. In previous studies, these networks have been studied separately because no prior knowledge is available to interconnect the three networks (i.e. which subset of each network are interconnected). These questions are intrinsically interesting, yet, computationally challenging. GLEN may provide a viable solution to investigate the unknown brain subnetworks that are related to phenotypes and further our understanding of the most complex organ, the human brain.

2 Methods

2.1 Background

We consider a group of networks $\boldsymbol{A}^1_{n\times n}, \cdots, \boldsymbol{A}^S_{n\times n} \sim \mathcal{P}$, where S is the number of subjects and all networks share an identical graph representation $G = \{V, E\}$ with $|V| = n$ nodes and $|E| = n(n-1)/2$ edges. For subject s, the multivariate outcome $\boldsymbol{A}_s = \{z^s_{ij}|e_{ij}\}_{i,j=1,\ldots,n}$ (e_{ij} denotes an edge connecting nodes i and j) can be a binary or weighted adjacency matrix and a vector of covariates $\boldsymbol{x}_s$ are also observed (e.g. clinical and demographic variables). We assume that $\boldsymbol{A}_s$ follows a distribution $\mathcal{P}$ with parameters related to $\boldsymbol{x}_s$.

We let z^s_{ij} $(i < j \in 1, \cdots, n)$ denote an off-diagonal edge entry in $\boldsymbol{A}_s$, and let z^s_{ij} follow a conditional distribution of the exponential family (e.g. a Bernoulli distribution for the binary adjacency matrix $\boldsymbol{A}_s$ and a Gaussian distribution for the weighted adjacency matrix $\boldsymbol{A}_s$). The conditional models $z^s_{ij}|z^s_{-(ij)} \sim f(\mu^s_{ij}, \phi_{ij})$ have been widely used for imaging analysis, where $g(\mu^s_{ij}) = \boldsymbol{x}_s\boldsymbol{\beta}_{ij}$ denotes the mean parameter and ϕ_{ij} for variance parameter (Besag, 1974; Bowman, 2005; Bowman et al., 2008; Derado et al., 2010; Risk et al., 2018). Recently, advanced statistical methods have been developed to provide localized (edge-wise) statistical inference while accounting for the dependence between edges for group-level

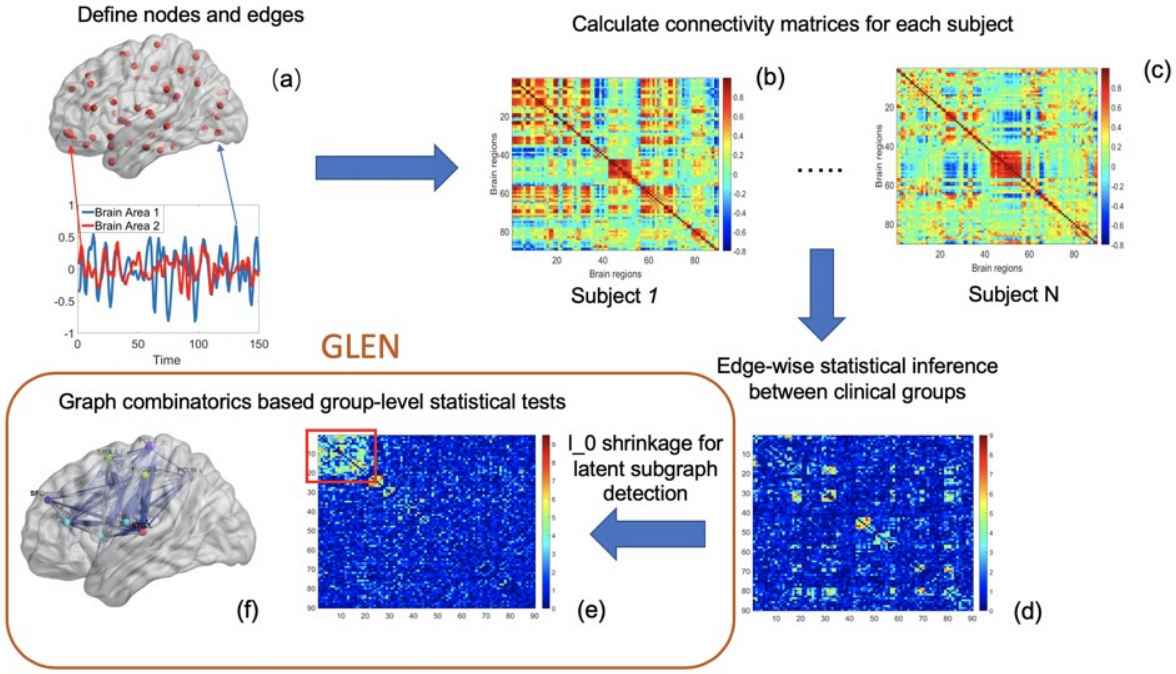

Figure 1: a) Define brain regions as nodes and connectivity metrics between each pair of nodes as edges. b c) Calculate the connectivity matrix for each single subject in a study where each off-diagonal element in the matrix represents the connectivity strength between two nodes; then identify differential connectivity patterns between clinical groups. d) Plot the edge-wise statistical inference where each off-diagonal element is a log transformed p-value (e.g. two sample test p-value per edge between clinical groups and a hotter point in the heatmap suggests larger group-wise difference). e) Reveal the disease related subnetwork detected by GLEN. f) Shows the corresponding 3D brain image. Note that e) is obtained by re-ordering the nodes in d) by listing detected subnetwork first (i.e. these two graphs are isomorphic).

network inference (Xia and Li, 2017; Xia and Li, 2018; Chen et al., 2018). These methods often yield improved statistical inferential results (e.g. testing statistics and p-values) on individual edges by taking into account the covariance structure between edges. However, the results of these methods still face the issue of multiple testing for multivariate edges to identify $E_1 = \{e_{ij}|\boldsymbol{\beta}_{ij} \neq 0\} \subset E$. If edges in E_1 are randomly distributed in whole brain connectome G, conventional methods for multivariate statistical inference (e.g. FDR and FWER) are applicable to the $n(n-1)/2$ vector because the order of edges can be randomly shuffled with no impact on the results. However, in practice, phenotypes rarely influence brain connections (edges) that are randomly distributed in the brain, instead, most times systematically.

In this paper, we focus on group-level network statistical inference which statistically assesses the impact of $\boldsymbol{x}_s$ on $\boldsymbol{A}_s$ at the subnetwork level. We consider $G_c \subset G$ as a subnetwork with high density of phenotype-related edges, where $\sum I(\beta_{ij} \neq 0|e_{ij} \in G_c)/|E_c| >> |E_1|/|E|$. Since edges in E_1 are not randomly distributed in G, G_c may have an organized topological structure. We define the lower bound of the subnetwork size (see subsection 2.3). The underlying topological structures of $\{G_c\}$ are inherently related to the graph theory and combinatorics which has a significant impact on network level inference. However, in practice, both E_1 and the subnetworks $\{G_c\}$ are latent and often overwhelmed by false positive noises.

Input Data of GLEN. We denote the resulting edge-wise statistical inference matrix as $\mathbf{W}_0$, where w_{ij}^0 is an entry in $\mathbf{W}_0$ (e.g. test statistics t_{ij} and p values $-\log(p_{ij})$). We first perform Sure Independence Screening (SIS) on $\mathbf{W}_0$ (Fan and Lv, 2008). We denote screened matrix $\mathbf{W}$, where the off-diagonal entry $w_{ij} = w_{ij}^0$ if $|w_{ij}^0| > r_0$ and $w_{ij} = 0$ if $|w_{ij}^0| \leq r_0$, where r_0 is the cutoff. $\mathbf{W}$ can effectively exclude most non-informative false positive edges

while maintaining a high proportion of true positive edges (Fan and Lv, 2008; Li et al., 2012). We consider $\mathbf{W}$ and the graph notation $G = \{V, E, \mathbf{W}\}$ as the input data of our method. The goal of our analysis is to accurately identify the true phenotype related latent subnetworks that are maximally composed of edge set E_1.

2.2 Detecting subgraphs via ℓ_0 norm regularization

If $G = \{V, E, \mathbf{W}\}$ is a non-random graph, then there exist subgraphs $\cup_{c=1}^{C} G_c \subset G$ and $G_c = \{V_c, E_c\}$ with higher density of phenotype-related edges than the rest of G. In other words, edges in G_c are more likely to be non-null: $\mathbb{P}(\beta_{ij} \neq 0 | e_{ij} \in G_c) > \mathbb{P}(\beta_{ij} \neq 0 | e_{ij} \notin G_c)$, which can be reflected by $\mathbb{E}(w_{ij} | e_{ij} \in G_c) > \mathbb{E}(w_{ij} | e_{ij} \notin G_c)$ from $\mathbf{W}$ of the sample data. We provide details of statistically testing whether $G = \{V, E, \mathbf{W}\}$ represents a random weighted graph in the supplementary materials.

We extract subnetworks from the input matrix $\mathbf{W}$ that can cover informative edges parsimoniously. We denote $\mathbf{U}_{n \times n}$ as a subgraph based matrix to link subgraphs $G = \cup_{c=1}^{C} G_c$ with the input matrix $\mathbf{W}$. The block diagonal matrix $\mathbf{U} = \text{Diag}(\mathbf{U}_c)$ or $\mathbf{U} = \cup_{c=1}^{C} \mathbf{U}_c$ is determined by the subgraphs $G = \cup_{c=1}^{C} G_c$, where an entry $u_{ij} = w_{ij}$ if the corresponding edge $e_{ij} \in G_c, \forall c$ and $u_{ij} = 0$ if $e_{ij} \notin G_c$. G_c can be a singleton that V_c=1 and $|E_c| = 0$. Thus, a diagonal submatrix $\mathbf{U}_c$ can be mapped to the corresponding subgraph G_c and $\mathbf{W}$. Note that our goal is to estimate subgraphs $G = \cup_{c=1}^{C} G_c$ instead of $\mathbf{U}$. Since G_c often shows an organized topological structure, we further denote the latent organized topological pattern of G_c by $\mathcal{T}(G_c)$. Our goal is to identify the phenotype related subnetworks/subgraphs $\cup_{c=1}^{C} G_c \subset G$ from $\mathbf{W}$, and we are also interested in learning $\mathcal{T}(G_c)$.

Detecting subgraphs with organized topological structures in $\mathbf{W}$, however, remains a major challenge because neither of the number C, E_1, nor $\mathcal{T}(G_c)$ is known. The existing

algorithms for subgraph/subnetwork detection can be substantially impacted by false positive noise edges randomly distributed between subnetworks (Newman and Girvan, 2004 and Rohe et al., 2011). This could be more challenging for a weighted matrix $\mathbf{W}$. To overcome these computational challenges, we propose an ℓ_0 norm regularization based objective function that minimizes the sizes of the phenotype-related topological structures (subgraphs) $\sum_{c=1}^{C} |E_c|$ and thus suppresses the impact of random false positive noise and better reveal the latent subnetworks.

The objective function is:

$$\underset{\hat{\mathbf{U}}=\cup_{c=1}^{C}\hat{\mathbf{U}}_c, C}{\arg\max} \log||\hat{\mathbf{U}}||_1 - \lambda_0 \log||\hat{\mathbf{U}}||_0 \tag{1}$$

where $1 \leq C \leq n$, and each $\hat{\mathbf{U}}_c$ corresponds to a subgraph $G_c = \{V_c, E_c\}$ that $\cup_{c=1}^{C} V_c = n$, $\cup_{c=1}^{C} E_c \subset E$, and $V_c \cap V_{c'} = \emptyset$.

We maximize the first term $||\hat{\mathbf{U}}||_1$ to cover more high-weight edges in $\mathbf{W}$ by the set of subgraphs $G = \cup_{c=1}^{C} G_c$. Maximizing the first term can reduce the chance to miss the true positive edges. The second term is defined by $||\hat{\mathbf{U}}||_0 = \sum_{c=1}^{C} |E_c|$, which is a penalty term. The first term is greater when the size $||\hat{\mathbf{U}}||_0$ is larger because a larger $||\hat{\mathbf{U}}||_0$ can cover more edges. However, a larger size of $||\hat{\mathbf{U}}||_0$ tends to include more false positive edges. Therefore, the objective function aims to maximize the signals (true positive findings) while penalizing the ℓ_0 norm of $\hat{\mathbf{U}}$ (false positive findings). λ_0 is a tuning parameter. By default, we let $\lambda_0 = 0.5$ and to balance false positive and negative findings. When $\lambda_0 = 0.5$, the proposed objective function is equivalent to the well-known problem of dense subgraph discovery in the field of graph theory and computer science and thus can be conveniently computed by linear programming and greedy algorithms (Khuller and Saha, 2009; Tsourakakis et al., 2013; Miyauchi and Kakimura, 2018). The density of G_c

is defined by $\sum_{e_{ij} \in G_c}(w_{ij}|e_{ij} \in G_c)/|V_c|$ and the density in G_c is higher than all other possible induced subgraphs (Gionis and Tsourakakis, 2015). However, the group level network inference for our applications may involve more complex scenarios because a set of subgraphs with unknown sizes and topological structures, and thus the existing dense subgraph discovery algorithms may not be directly applicable. Therefore, we resort it to the ℓ_0 norm regularization for subgraph extraction.

The ℓ_0 norm regularization has been attracting interest in the field of statistics (Shen et al., 2012; Li et al., 2018; Hazimeh and Mazumder, 2018). Our optimization is distinct from these methods because we focus on edge-variables in a graph space. Interestingly, we note that the penalty term $||\hat{\mathbf{U}}||_0$ is related to the number of subgraphs C, which links the network/graph detection with ℓ_0 norm regularization. This property is unique for edge-variables in the graph space and provides a computationally efficient heuristic of ℓ_0 norm regularization for edge variables. The increase of C generally leads to smaller sizes of organized topological structures, and thus reduced ℓ_0 norm but larger loss of $||\hat{\mathbf{U}}||_1$. If $C = n$ then $||\hat{\mathbf{U}}||_0 = 0$, and $C = 1$ then $||\hat{\mathbf{U}}||_0 = n \times (n-1)/2$.

To illustrate this point, the objective function can also be re-written as (similar to the counterpart of expression for LASSO):

$$\underset{\hat{\mathbf{U}}=\cup_{c=1}^{C}\hat{\mathbf{U}}_c, C}{\arg\max} \ \log||\hat{\mathbf{U}}||_1 = \log\sum_{i<j}(w_{ij}|e_{ij} \in G_c)$$

with subject to

$$\log||\hat{\mathbf{U}}||_0 = \log\sum_{i<j} I(e_{ij} \in G_c, \forall c) \leq t.$$

We note that C can provide an upper bound for $||\hat{\mathbf{U}}||_0$.

Lemma 1. *For a given value of C, we have the upper bound* $\sup ||\hat{\boldsymbol{U}}||_0 = (n-C+1)(n-C))/2$.

Proof. For $C = 2$,

$$\begin{aligned} 2\sup ||\hat{\mathbf{U}}||_0 &= \sup\{n_1(n_1-1) + n_2(n_2-1) : n_1+n_2 = n, n_1, n_2 \in \mathbf{Z}^+\} \\ &= -n + \sup\{n_1^2 + n_2^2 : n_1+n_2 = n, n_1, n_2 \in \mathbf{Z}^+\} \\ &= -n + (n-1)^2 + 1 \end{aligned}$$

where $n_1 = |V_1|, n_2 = |V_2|$ are number of vertices for the two communities. Hence, $2\sup ||\hat{\mathbf{U}}||_0 = (n-C+1)(n-C)$ for C=2.

Inductively, assume it's true for $C = k-1$, such that $2\sup ||\hat{\mathbf{U}}||_0 = (n-k+2)(n-k+1)$. Then, for $C = k$,

$$\begin{aligned} 2\sup ||\hat{\mathbf{U}}||_0 &= -n + \sup\{n_1^2 + ... + n_k^2 : \sum_{i=1}^{k} n_i = n, n_1, ..., n_k \in \mathbf{Z}^+\} \\ &= -n + \sup\{(n-n_k-k+2)^2 + (k-2) + n_k^2 : n_k \in \{1, 2, ..., n-k+1\}\} \\ &= -(n-k+2) + \sup\{n_1^2 + n_2^2 : n_1+n_2 = n-k+2, n_1+n_2 \in \mathbf{Z}^+\} \\ &= -(n-k+2) + (n-k+2-1)^2 + 1 \\ &= (n-C+1)(n-C) \end{aligned} \tag{2}$$

Therefore, the result is true for $C = k$. Hence, the claim is proved. □

Lemma 1 shows that increasing C can effectively shrink the size of $||\hat{\mathbf{U}}||_0$. Given the value of $||\hat{\mathbf{U}}||_1$, a larger C is favored by the objective function. We develop the algorithm to optimize (1) based on this property. Specifically, the algorithm optimizes the objective function (1) with a given C and the corresponding upper bound of $||\hat{\mathbf{U}}||_0 = (n-C+1)(n-C)$

to identify the subgraphs can cover most edges with penalized sizes. Then, we exhaust C from 2 to $n-2$ and perform grid search to estimate $\hat{C}$. The combination of $\hat{C}$ and $\hat{\mathbf{U}}$ that optimizes (1) become the final estimates. The details of the algorithm, the derivation, and the software package are provided in provided in the supplementary materials. Comparing with the algorithms for dense subgraph detection, the proposed algorithm can automatically determine the number and membership of unconnected dense subgraphs. Interestingly, we find that detecting dense subnetworks by increasing C is equivalent to the steps of pealing algorithms (to remove the less connected nodes) in dense graph discovery (Charikar, 2000). When there is one subgraph, the two methods have similar performance.

Furthermore, for a given C, the proposed objective function can provide a consistent estimate for the community topological structure (the node-induced subgraph). Considering that C is optimized by grid search, the objective function (1) yields a consistent estimate of the set of subgraphs.

Theorem 1. *Let* $\boldsymbol{W}_{n\times n}$ *be a weighted adjacency matrix for a graph* $G = \{V, E\}$ *such that there is a set of subgraphs* $\{G_c = \{V_c, E_c\} : c = 1, ..., C, \cup_{c=1}^{C} G_c \subset G\}$ *with* $w_{ij}|e_{ij} \in \cup_{c=1}^{C} G_c \overset{i.i.d}{\sim} f_1$ *and* $w_{ij}|e_{ij} \notin \cup_{c=1}^{C} G_c \overset{i.i.d}{\sim} f_0$ *where* f_1 *and* f_0 *are continuous densities in* $[0, 1]$ *with mean and variance* (μ_1, σ_1^2) *and* (μ_0, σ_0^2), $\mu_1 > \mu_0$, *respectively. Then* $\boldsymbol{P} = \mathbb{E}(\boldsymbol{W})$ *can be expressed as* $\boldsymbol{P} = \boldsymbol{\Theta}\boldsymbol{B}\boldsymbol{\Theta}^T$, *where* $\boldsymbol{\Theta} \in \mathbb{R}^{n\times C}$ *is a membership matrix indicating the community index of each node, and* $\boldsymbol{B} \in \mathbb{R}^{C\times C}$ *equals* μ_1 *in diagonal and* μ_0 *for others. Assume* $\boldsymbol{P}$ *of rank* C *has smallest absolute nonzero eigenvalue* ξ *and* $(\mu_1 \vee \sigma_1^2 \vee \sigma_0^2) \leq d$ *for* $d \geq c_0 \log n$ *and* $c_0 > 0$. *Then, if* $(2+\varepsilon)\frac{Cnd}{\xi^2} < \tau$ *for some* $\tau, \varepsilon > 0$, *the output* $\hat{\boldsymbol{\Theta}}$ *is consistent up to a permutation in the sense that with probability at least* $1 - n^{-1}$, *the*

clustering correction is guaranteed outside subsets $S_c \subset G_c$ *for* $c = 1, ..., C$ *and*

$$\sum_{c=1}^{C} \frac{|S_c|}{|V_c|} \leq \tau^{-1}(2+\varepsilon)\frac{Cnd}{\xi^2}.$$

Proof. The proof is included in supplemental material. Note that the theorem is also true for a general setting of weighted stochastic block model with $(\max_{ij} B_{ij} \vee \max_{ij} \sigma_{ij}^2) \leq d$.

The ℓ_0 norm regularization is critical to suppress false positive noise and thus improve the accuracy of subgraph detection. Let $w_{ij} \neq 0 | \beta_{ij} = 0$ be an false positive edge that $i \in G_c$ and $j \notin G_c, \forall c$. Adding e_{ij} to G_c will increase $||\hat{\mathbf{U}}||_0$ by $|V_c|$. However, the increase of $||\hat{\mathbf{U}}||_1$ is small because there is only a small proportion of $w_{i'j} > 0, i' \in G_c$ and $i' \neq i$. In another scenario, $w_{ij} \neq 0 | \beta_{ij} = 0$ be an false positive edge that $i \in G_c$ and $j \in G_{c'}, \forall c \neq c'$. The inclusion of a false positive edge e_{ij} can connect two subgraphs false positively and increase $||\hat{\mathbf{U}}||_0$ by $|V_c| \times |V_{c'}|$. In both cases, the inclusion of false positive edges will lead to a high cost. Therefore, subnetwork detection by objective function (1) is robust to the false positive noise/edges.

Extracting complex topological structures $\mathcal{T}(G_c)$ *by the objective function (1)*: we consider the community/clique structure as the default topological structure, although the objective function (1) can be further optimized if the phenotype related edges can be covered more subgraphs with more sophisticated topological structures. For example, there are two potential subgraphs, one is a clique G_c and the other is a bipartite subgraph $G_{c'}$with the same number of nodes $V_c = V_{c'}$. The number of edges of the clique is greater than the bipartite subgraph $E_c > E_{c'}$. If the edges with $\beta_{ij} \neq 0$ are equivalently covered by G_c and $G_{c'}$, the bipartite subgraph is preferred by (1). The detailed algorithms for more sophisticated topological structures (e.g. k-partite structure, rich-club, and interconnected subgraphs) and model selection procedure are described by Chen et al. (2019) and Wu 2019.

Therefore, the complex topological structures are favored by the ℓ_0 norm regularization if the objective function is further optimized.

2.3 Graph combinatorics for network-level test

We optimize the objective function (1) to discover that a large proportion of informative edges can be allocated into organized dense subgraphs, which has naturally provoked the properties of graph combinatorics. Next, we perform the group-level network inference to examine whether the combinatorial set of edges (instead of individual edges) in a subnetwork is significantly related to the phenotype. In contrast to conventional multivariate inference on individual edges, the subnetwork based statistical inference is inherently linked with graph combinatorics as

$$\{e_{ij}|\beta_{ij} \neq 0\} \xrightarrow{\text{Graph combinatorics}} G_c.$$

We first specify the null and alternative hypotheses of the subnetwork test. Let $\rho = \sum_{i<j} I(\beta_{ij} \neq 0)/|E|$ and G be a graph with n nodes and the connection probability of ρ. The null can be categorized into two cases: the stronger case ($\mathbf{H}_0^s$) that no edge is related to the phenotype ($\rho = 0$); and the weaker case that a small proportion ($\rho \geq 0$) of phenotype-related edges are randomly distributed in G. Clearly, the weaker case is more general and commonly used and the stronger case is a special case of the weaker case (Benjamini and Hochberg, 1995), and additional tests could be conducted to distinguish the two nulls as follows. Thus, we focus on the weaker case of the null. The null and alternative hypotheses

are:

$$\begin{aligned} &\mathbf{H}_0 : G(n,\rho) \text{ is a random graph.} \\ &\mathbf{H}_a : \exists G_c \subset G, \text{ that } \rho_c = \sum_{e_{ij}\in G_c} I(\beta_{ij} \neq 0)/|E_c| > k\rho, \text{ with constant } k > 1. \end{aligned} \tag{3}$$

If we fail to reject the null (weaker case) that G is a random graph, then the conventional multi-testing methods (e.g. FDR and FWER) can be applied to identify individual edges that are related to the phenotypes. The stronger case of null is rejected if at least one edge passed the cut-off of the multiple testing methods.

In the follows, we show that from the perspective of graph combinatorics, testing a combinatorial set of multivariate edges constrained in a subnetwork (subnetwork level inference) can lead to low false positive and false negative error rates. We consider the permutation test for the network level inference because the set of subgraphs G_c is the power set of G that imposes difficulty for asymptotic inference and multiple testing (Zalesky et al., 2010; Chen et al., 2015; Chen et al., 2019). Briefly, the permutation tests generate multiple simulations (e.g. 10^4 times) of data under the null by shuffling the labels, and calculate test statistic T_c^m for each extracted subnetwork G_c and record the maximum statistic T_{max}^m for each simulation. The percentile of observed test statistic among the maximum test statistics of all simulations becomes the p-value. We include the detailed permutation test procedure in the supplementary materials. Specifically, we assume that each simulation of the permutation test generates a matrix $\mathbf{W}^m$ under the null and the false positive edges are randomly distributed in G (G is a random graph), where $m = 1, \cdots, M$. The test statistic in the permutation test statistic is a function of $\bar{w}_c = \mathbb{E}(w_{ij}|e_{ij} \in G_c)$ and we assume $\bar{w}_c = \bar{w}_{c'}$. Then, the test statistic is related to the size of the subnetwork and $\mathbb{P}(|T_c| < |T_{max}^m|) = \mathbb{P}(|G_c| < |G_{max}^m|)$ since $\bar{w}_c = \bar{w}_{c'}$. We demonstrate the impact of graph combinatorics on network-level inference in terms of Type I and Type II errors.

2.3.1 Type I error of network-level inference

Type I error rate is the probability to false positively identify a significant subgraph when the null is true. For subnetwork level inference, the subgraph can be false positive only if it is detected. We first consider the probability of a (latent) dense induced subgraph existing in G under the null. We assume a random graph $G^R = \{V_c^R, E_c^R\}$ is generated with probability p. $\forall e_{ij} \in G^R$, we have $\delta_{ij} \sim \text{Bern}(p)$. We further let $w_{ij}|\delta_{ij} = 1 \sim f_1$ and $w_{ij}|\delta_{ij} = 0 \sim f_0$ as entries of $\mathbf{W}^m$. Therefore, the weighted adjacency matrix is generated from a random graph $G(n, p)$ with $\mathbb{E}(w_{ij}|\delta_{ij} = 1) = \mu_1 > \mu_0 = \mathbb{E}(w_{ij}|\delta_{ij} = 0)$. We prove that under the null the probability of a subgraph with a detectable size and a density $> \mu_0$ from a random graph (null) converges to zero.

Theorem 2. *Let G_c^R be a detectable community structure under the null, i.e. the average weight inside structure satisfies $\sum[w_{ij}|e_{ij} \in G_c^R]/|E_c^R| > \gamma$ with some $\gamma \in (\sum_{i<j} w_{ij}/|E^R| \wedge \mu_0, \mu_1)$. Then, we have*

$$\mathbb{P}(|G_c^R| \geq \frac{1}{2}\eta^2(\ln n)^3) \to 0,$$

with

$$\eta = \eta(\tilde{\gamma}) = 2\left\{\ln\left[\left(\frac{\tilde{\gamma}}{p}\right)^{\tilde{\gamma}}\left(\frac{1-\tilde{\gamma}}{1-p}\right)^{1-\tilde{\gamma}}\right]\right\}^{-1} \text{ and } \tilde{\gamma} = \frac{\gamma - \mu_0}{\mu_1 - \mu_0}.$$

Proof. For any random graph $G_S = (S, E_S)$, let $\overline{G}_S = (S, \overline{E}_S) = (\{S, \{uv|u, v \in S, u \neq v, \text{ and } uv \notin E_S\}\})$ be the complement of graph G_S. Assume $|E_S|/(|E_S|+\overline{E}_S) = |E_S|/\binom{|V_S|}{2} \to q$, and $\mathbf{W}_S$ is the associated weighted adjacency matrix generated from the random graph with $\mathbb{E}(w_{ij}|\delta_{ij} = 1) = \mu_1 > \mu_0 = \mathbb{E}(w_{ij}|\delta_{ij} = 0)$. Then, from law of large numbers, the

average weights for $\mathbf{W}_S$ will have in probability

$$\frac{\sum_{i<j} w_{ij}}{\binom{|V_S|}{2}} = \frac{\sum_{(i,j)\in E_S} w_{ij} + \sum_{(i,j)\in \overline{E}_S} w_{ij}}{\binom{|V_S|}{2}}$$
$$\to q\mu_1 + (1-q)\mu_0$$

as $|S| \to \infty$. Hence, if the average weight is no less than γ, i.e. $\sum_{i<j}[w_{ij}|e_{ij} \in G_S]/\binom{|V_S|}{2} > \gamma$, we will have $|E_S|/\binom{|V_S|}{2} \geq \tilde{\gamma}$ asymptotically. Otherwise, for any $0 < \epsilon < 1$, if $|E_S|/\binom{|V_S|}{2} \to \tilde{\gamma} - \epsilon$, from law of large numbers, $\mathbb{P}(\frac{\sum_{i<j} w_{ij}}{\binom{|V_S|}{2}} < \gamma - \epsilon/2) \to 1$ which contradicts the assumption.

Therefore, the community structure G_c^R is generated from $\tilde{\gamma}$-dense communities in random graph in the sense that $|E_c^R|/\binom{|V_c^R|}{2} > \tilde{\gamma}$. From Lemma 7 in the supplementary material, if G_c^R is constructed by $\ln n$ communities, $|G_c^R|$ will be bounded by $\frac{1}{2}\eta^2(\ln n)^3$ almost surely. □

In general, we consider G_c as a potential phenotype-related subnetwork if $|G_c| > \frac{1}{2}\eta^2(\ln n)^3$ because a smaller subnetwork can be observed by chance. Therefore, the probability of false positive edges comprising a detectable dense subgraph is approximately zero with a threshold level r_0 under the null. The non-detectable subgraph leads to no false positive report of phenotype-related subnetwork and no Type I error for network-level inference when H_0 is true.

2.3.2 Type II error of network level inference

We next consider the power for network-level inference, which is the probability to reject null given that the alternative hypothesis is true and $G_c \in G$ is significantly associated with the phenotype. The power is determined by the values of test statistics from all simulations

of the null of the permutation test. In each simulation of the permutation test, we assume a random graph G^R is generated as above. Suppose G_c is a detectable subgraph from the data and $|G_c| \geq \frac{1}{2}\eta^2(\ln n)^3$ when the alternative hypothesis is true. The probability to reject the null converges to 1 when the each simulation in the permutation test yields a maximum subgraph G^m_{max} that $\mathbb{P}(|G^m_{max}| \geq \frac{1}{2}\eta^2(\ln n)^3) \to 0$ and thus $\mathbb{P}(T^m_{max} > T_c) \to 0$.

Theorem 3. *Assume that for mth simulation in the permutation test, the weighted adjacency can be regarded as a generation from random graph $G^R(n,p)$ with $\mathbb{E}(w_{ij}|\delta_{ij}=1) = \mu_1 > \mu_0 = \mathbb{E}(w_{ij}|\delta_{ij}=0)$. Let G^m_{max} be the maximum detectable community structure with average weight no less than γ, then we will have*

$$\mathbb{P}(|G^m_{max}| \geq \frac{1}{2}\eta^2(\ln n)^3) \to 0,$$

with $\eta = \eta(\tilde{\gamma})$ and $\tilde{\gamma} = \frac{\gamma-\mu_0}{\mu_1-\mu_0}$ defined in Theorem 2.

Proof. The claim is automatically true from Theorem 2. □

The above theorem shows that the Type II error of network-level inference is low when the phenotype related subnetworks present. In summary, the graph combinatorics plays an important role in network-level statistical inference because it is rare for an organized subnetwork related to the phenotype by combinatorics. Given the ℓ_0 norm regularization based algorithm can capture the dense subgraphs with a density $\gamma \gg \mu_0$, the following network-level statistical tests enjoy simultaneously reduced false positive and false negative discovery rates. Besides network-level inference, we can also draw statistical inference on individual edges inside and outside G_c adaptively based on an empirical Bayes framework (Chen et al., 2018). Since this topic is out of the scope of this article, we include brief descriptions in the supplementary materials. We also prove that the edge-wise false positive

and false negative error rates given $\{G_c\}$ are also simultaneously improved compared with the conventional methods that apply a universal cut-off like in FDR, local fdr, and FWER.

3 Simulations

In the simulation analysis, we validate whether 1) whether the ℓ_0 norm shrinkage based objective function can capture the latent phenotype-related subnetworks; 2) whether the graph combinatorics based network level inference can reduce false positive and false negative error rates. We simulate group-level connectome data sets $\boldsymbol{A} = \{\boldsymbol{A}^1, ..., \boldsymbol{A}^S\}$ and the corresponding graph whole brain connectome G, we define a community subnetwork $G_c \subset G$ where edges are differentially expressed between controls and cases. We let $|V| = 100$, $|V_c| = 20$, and G_c to be a clique (and $|E_c| = 190$). The transformed (e.g. Fishers Z) connectivity metric of each edge is set to marginally follow a normal distribution with μ_0 and σ_0^2 (for controls) and μ_1 and σ_1^2 (for cases). For differentially expressed edges (i.e. $e_{i,j} \in E_c$), $\mu_0 = \mu_1 + \theta$ and otherwise $\mu_0 = \mu_1$. Also, we let $\sigma_0^2 = \sigma_1^2 = \sigma^2$. $\theta = 0.5$ is used with varying values of σ^2 for different signal-to-noise ratios (SNRs)/standardized effect sizes. By shuffling the order of nodes in G, the altered connectivity subnetwork becomes latent in the sample data. Two sample sizes (60 and 120) are used to represent the commonly observed sample size from a single study. Each setting is simulated with different θ, σ and the number of subjects for each group for 100 times.

The subnetwork detection algorithm and graph combinatorics based tests of GLEN are applied to each simulated data set. The permutation tests are performed by shuffling the group labels and edge orders to simulate the null, denoted by GLEN1 and GLEN2 (see details in the supplementary materials). The false positive discovery rates (FP) and false

negative discovery rates (FN) are evaluated at both subnetwork and edge level rates. Note that edge-wise power can be further calculated as $1 - FN$. Our method is compared with other multiple testing methods including BenjaminiHochberg FDR and local false discovery rate control (fdr). The false positive findings (number of FP edges in mean and standard deviation across 100 repetitions) and false negative (FN) edges are shown in Table 1.

GLEN shows improved performance on network level inference by identifying the latent and differentially expressed subnetwork with 0 FP and FN rates. Next, GLEN (based on the selected subnetwork) is compared to FDR and local fdr at individual edge inference using $q = 0.2$ as cut-off for both FDR and fdr. The results show that generally FDR has higher FP but lower FN rates compared with fdr (i.e. fdr is more conservative). Importantly, GLEN outperforms FDR and fdr when jointly considering FP and FN rates, see Table 1. Finally, we apply NBS method to the simulated data, but no subnetwork is detected by NBS when tuning the cutoff parameter from 3 to 6 for all settings (not shown in Table 1).

On Type I error rate, we count the number of false positive significant subnetworks for the data sets with no differentially expressed connectome networks (e.g. $\theta = 0$). Based on simulation of 1000 iterations, the network level false positive rates of GLEN1 is 1.2% and GLEN2 is 2.9%. Therefore, the network level Type I error is well controlled and below the subnetwork claimed level of 5%.

4 Data example

The primary data set D^1 includes 70 individuals with schizophrenia (age $= 40.80 \pm 13.63$ years) and 70 control subjects (age $= 41.79 \pm 13.44$ years) frequency-matched on age ($t =$

[1]FDR: Benjamini-Hochberg false discovery rate control (FDR)

[2]fdr: local false discovery rate control.

Number of Subjects: 30 vs 30

	σ	0.5	1	2
GLEN1	FP	10.7(28.53)	11.8(30.36)	9.25(18.38)
	FN	4.95(22.13)	3.75(16.77)	0(0)
	Network FP	0(0)	0(0)	0(0)
	Network FN	0(0)	0(0)	0(0)
GLEN2	FP	7.15(15.25)	8.25(18.31)	16(36.13)
	FN	0(0)	2.7(12.07)	11.3(42)
	Network FP	0(0)	0(0)	0(0)
	Network FN	0(0)	0(0)	0(0)
FDR[1]	FP	45.98(9.07)	43.37(9.57)	31.11(8.01)
	FN	8.23(3.02)	24.99(5.37)	70.67(8.12)
fdr[2]	FP	1.01(1.05)	0.74(1.05)	0.11(0.4)
	FN	57.26(12.56)	101.58(16.11)	175.95(14.52)

Number of Subjects: 60 vs 60

	σ	0.5	1	2
GLEN1	FP	2(6.16)	2.05(9.17)	8.1(13.86)
	FN	0(0)	0(0)	0(0)
	Network FP	0(0)	0(0)	0(0)
	Network FN	0(0)	0(0)	0(0)
GLEN2	FP	3.05(9.99)	2.05(9.17)	11.8(30.13)
	FN	0(0)	0(0)	0(0)
	Network FP	0(0)	0(0)	0(0)
	Network FN	0(0)	0(0)	0(0)
FDR[1]	FP	54.55(8.35)	51.6(11.77)	47.85(7.53)
	FN	0(0)	0.25(0.44)	5.85(2.39)
fdr[2]	FP	0.1(0.31)	0.5(0.61)	0.8(1.32)
	FN	5.6(2.54)	17.55(5.4)	56.8(7.63)

0.62, $p = 0.54$) and sex ratio ($\chi^2 = 0$, $p = 1$). In the validation data set D^2, another 30 individuals with schizophrenia (age = 39.73 ± 13.79 years) and 30 control subjects were recruited (age = 39.73 ± 14.16 years) matched on age ($t = 0.27$, $p = 0.78$) and sex ratio ($\chi^2 = 0.09$, $p = 0.77$), following the initial sample. The recruitment procedures, inclusion and exclusion criteria, and imaging acquisition and preprocessing procedure were kept the same. The details of subject recruitment, imaging acquisition, and preprocessing procedures are described in Supplementary Materials. The nodes of the connectome graph G are defined based on the commonly used automated anatomical labeling (AAL) regions. Time courses of all voxels within a 10 mm sphere around the centroid of each region are pre-processed as region-wise signals, followed by calculating 4005 Pearson correlation coefficients between the time courses of the 90 AAL regions. Fishers Z transformation and normalization are then applied to obtain connectivity matrices. We apply GLEN to both data sets separately and hen compare D^1 and D^2 results. We also compare the results by GLEN with the traditional edge-wise and the commonly used network based statistic (NBS) methods.

4.1 Network-level results of D^1

We first apply GLEN to D^1. Let symmetric matrix $\mathbf{W}^1$ be the whole brain graph edge-wise testing result matrix (Fig 3a), where the element is $w_{ij} = -\log(p_{ij})$ where i and j are two distinct brain regions and p_{ij} is the corresponding test p-value for the edge between i and j. The graph combinatorics based testing results show that one subnetwork in D^1 is significant ($p < 0.001$). The significant subnetwork ($\mathbf{R}^1$ denotes the subnetwork from D^1) includes 22 nodes, 231 altered edges, and a well-organized topological structure. The detected topology is a community structure with 22 nodes including the left medial frontal

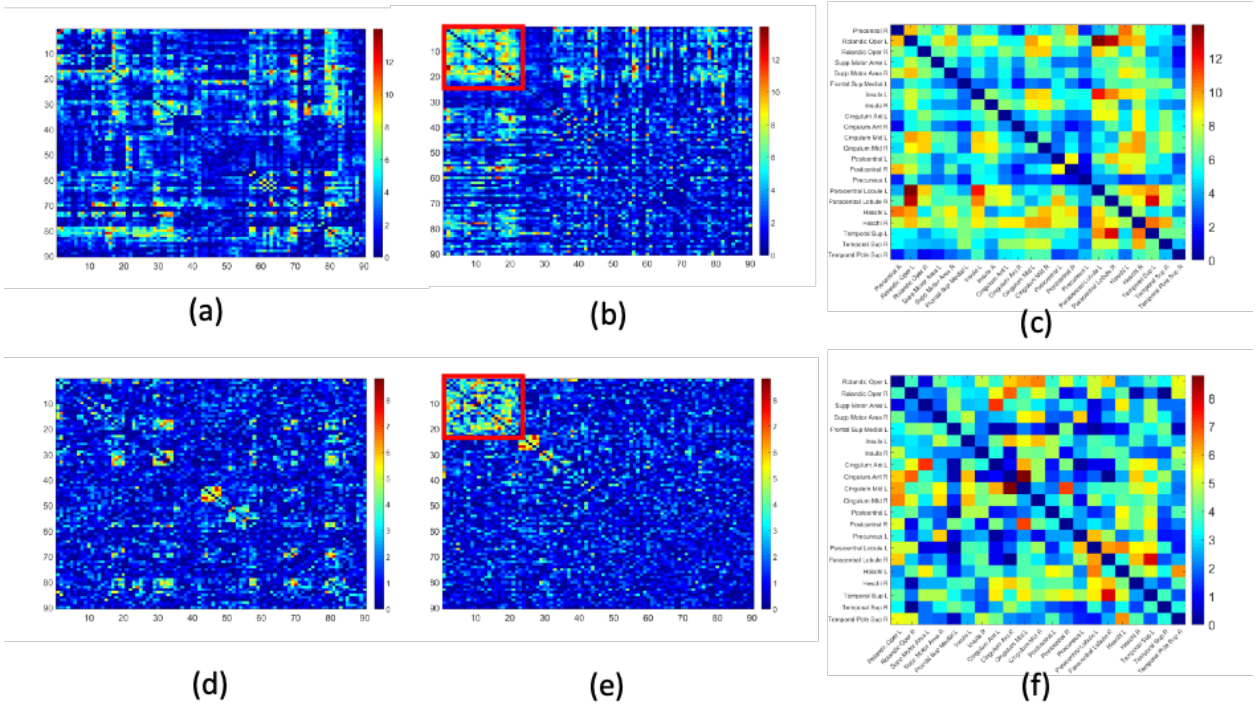

Figure 2: Applying GLEN to clinical data D^1 (a-c) and replication data D^2 (d-f). a) A heatmap of $\log(p)$ of the first data set (D^1). A hotter pixel indicates more differential edge between cases and controls. There is no apparent topological pattern of these hot edges. b) We then perform GLEN in D^1 and find a significant subnetwork (red square, which is magnified in c). c) The enlarged disease-relevant subnetwork in D^1 with region names. d) A heatmap of log(p) of the second data set (D^2). e) The detected disease-relevant subnetwork by using D^2 alone. f) The enlarged network in D^2 with region names.

cortex, bilateral insula, bilateral anterior and middle cingulate cortices, bilateral Heschl gyrus and superior temporal cortices, bilateral paracentral and postcentral cortices, right precentral cortex, and the precueous (Fig 3a to 3c) (full list of region names in Table 1 of supplementary materials).

4.2 Network-level results of D^2

One subnetwork is significant in D^2 ($p < 0.001$). $\mathbf{R}^2$ (subnetwork from D^2) includes 21 nodes, 210 edges in a clique/community topological structure. The majority of the nodes of $\mathbf{R}^2$ are similar to $\mathbf{R}^1$ with some minor differences, and include left medial superior frontal gyrus, bilateral insula, bilateral anterior and middle cingulate cortices, bilateral Heschl gyrus, Rolandic operculums, supplementary motor areas, paracentral lobules, postcentral lobules, and left precuneus (Fig 3d to 3e) (Table 2 in supplementary materials). Similar to $\mathbf{R}^1$, most (206 of the 210) edges showed reduced connectivity in patients with schizophrenia in $\mathbf{R}^2$.

4.3 Comparing subnetworks in D^1 and D^2

Remarkably, we note that $\mathbf{R}^2 \subset \mathbf{R}^1$: GLEN reports the altered subnetwork in D^1 that can be rediscovered when analyzing D^2 independently with only one node in difference. Based on the combinatorics, $\mathbb{P}(\mathbf{R}^2 \subset \mathbf{R}^1) < 2 \times 10^{-16}$ given 21 nodes are included in $\mathbf{R}^2$. Although the sample size of D^2 is smaller and the testing p-values are larger, the subnetwork detection in D^2 is not affected by the sample size and other sources of noises. While the statistical inference on individual edges is subject to numerous variation and false positive-negative findings, we found that the latent network topological patterns of differentially expressed edges are stable across independent samples. The new network size regularization term

is critical to recognize the organized patterns from the noisy background. The organized (latent) subnetwork topological structure can reduce the false positive and false negative findings simultaneously. As a result, the subnetwork level findings are reproducible and verified by D^2.

4.4 Comparisons with existing methods

For comparison purpose, we also perform edge-wise multiple testing analysis and NBS. Wilcoxon rank sum tests are then used to assess patient-control differences in the normalized correlation coefficients for all edges. In D^1, 430 of 4005 edges are $p < 0.005$. $p = 0.005$ is commonly used for uncorrected threshold in neuroimaging literature (Derado biometrics 2010). After FDR correction for multiple comparison, none of 4005 edges is found significant by using the threshold $q = 0.05$. In D^2, 22 of 4005 edges are $p < 0.005$, and none of edges are found significant after applying FDR correction with the threshold $q = 0.05$. Two edges among the 430 edges in D^1 and 22 edges in D^2 overlap, which indicates a very low replicability between the two data sets. In addition, the conventional network method NBS shows no differentially expressed subnetwork in D^1 and D^2 by using various thresholds (tuning threshold values from 2 to 6). This is likely due to the loss of power by the impact of false positive edges (without ℓ_0 norm shrinkage).

Finally, the positive agreement is used to compare reproducibility of features between D^1 and D^2 using GLEN vs. individual edge based statistics. The network approach is significantly better than individual edge based method ($p < 0.001$). In summary, by utilizing an independent replication data set collected posteriorly we can conclude that the findings identified by the GLEN are more reproducible.

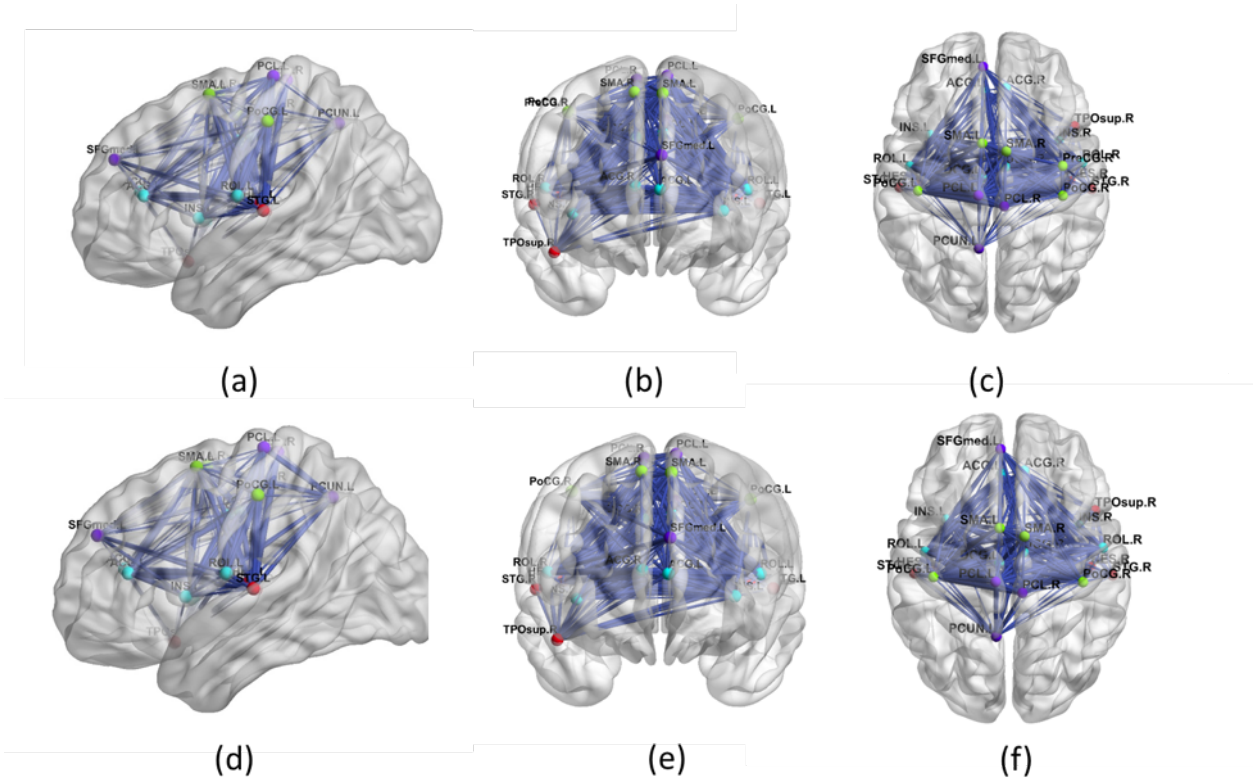

Figure 3: The edges in the subnetwork using 3D demonstration for data set 1 (a)-(c) and data set 2 (d)-(f). The width of length is proportional to the effect size. The disease-relevant network involves with the salience network, part of default mode network, and part of executive network, and more importantly the interconnections between these three networks are revealed. (e)-(f) show 3D brain subnetwork for data set 2, which shows a highly replicable brain subnetwork as seen in data set 1 with one brain region (precentral R) less.

4.5 Biological interpretation of the brain subnetwork

The brain region constellation of the detected subnetwork consists of many well-known brain regions involved with the schizophrenic disorder, which included inferior frontal, superior temporal, insula, cingulate, and paracentral areas (Fig 3). This altered subnetwork is composed of approximately the salience network (SN), part of default mode network (DMN), and part of central executive network (CEN), which have been frequently associated with abnormalities in schizophrenia during attention, working memory, and executive control, and resting functional imaging studies. Interestingly, the detected subnetwork reveals not only that the SN, DMN, and CEN are altered but also that the interconnections between these three networks are disrupted. Of the 231 differentially expressed edges, all edges show decreased or equivalent connections in patients with schizophrenia. This aligns with findings suggesting that schizophrenia is a 'dysconnectivity' disorder with primarily reduced functional connectivity across brain regions (Lynall et al., 2010), although medication effects cannot be completely ruled out.

5 Discussion

The multiple testing problem has been a staple of neuroimaging statistics. The goal of the neuroimaging study is to investigate which subset of the brain imaging features are related to the phenotypes of interest. The multiple testing problem neuroimaging data is more complicated than other types of high-throughput data (e.g. gene expression data) because the neuroimaging data are involved with the intrinsic constraints of spatial, temporal, and network structures (Lazar, 2008; Lindquist, 2008). In this paper, we present a graph theory and combinatorics based statistical approach for group-level network inference

which is motivated by brain connectome analysis. We aim to identify which subnetwork (a combinatorial set of edges in an organized subgraph) is related to the phenotype. It can lead to an important and interesting biological question investigating whether and which brain circuitries are systematically altered by the phenotype. Clearly, the subnetwork-level inferential results are more interpretable than graph theoretical summary statistics (e.g. small-worldness and clustering coefficients) by clearly specifying localized edges, and provide more biological insights than multiple univariate (localized) edge inference by indicating how brain networks are systematically influenced instead of a set of individual edges separately. Therefore, our development has focused on detecting and statistically testing latent (phenotype related) organized subnetworks from group level network data.

We propose a new objective function to extract latent subnetworks via the ℓ_0 norm shrinkage. The ℓ_0 norm regularization ensures that detected subnetworks are dense subgraphs where informative edges are highly concentrated in organized topological structures. Our algorithm bears resemblance to the peeling algorithms of dense subgraph discovery which reveals the targets by increasing the number of isolated vertices but is more flexible for multiple subgraphs and subgraphs with topological structures beyond cliques. We statistically test each subnetwork by graph combinatorics based on the fact that the chance of informative edges comprising an organized subnetwork converges to zero under the null. The subnetwork level test results are particularly useful for brain network research as they may provide insights of unknown subnetworks related to the phenotypes and thus can better help to understand the underlying mechanisms of neuropathology and neurophysiology.

The reproducibility and replicability have been key challenges for neuroscience research (Eklund et al., 2016). One potential cause for the low replicability is that a universal threshold (e.g. the primary threshold for cluster-extent test and FDR for all edges) is applied to

all imaging features. The multivariate brain imaging data are often impacted by substantial noises from various sources, and hence it is necessary to rigorously to control the false positive findings. However, the prohibition of false positive findings may also significantly reduce the true positive findings, and thus the results of each study can only include a small proportion of true signals. Therefore, the small proportions of true signals across studies rarely overlap with each other, which leads to low replicability. Our results demonstrate that findings by GLEN are highly replicable because both false positive and false negative findings are simultaneously reduced via our graph combinatorics based statistical methods. The true signals can unite with each other in a detected organized topological structure to avoid false negative errors while false positive signals randomly distributed the graph space can be effectively suppressed by ℓ_0 shrinkage. In contrast, the traditional statistical methods show almost no overlapped findings. These jointly provide evidence that graph theory and combinatorics are critical for group-level network inference. Last, the graph combinatorics based statistical method is naturally related to human visual recognition. For example, when we look at a photograph, we often focus on a combinatorial set of pixels in well-organized patterns and jointly recognize them rather than scanning individual pixels. Therefore, GLEN may provide a new 'natural' pathway to investigate the large scale and complexly structured network data.

SUPPLEMENTARY MATERIALS

A1. Testing the nonrandom patterns of differentially expressed edges

Let r_0 be a threshold for multiple testing correction, and $E_1(r_0) = \{e_{i,j}|\ |w_{ij}| > r_0\}$ and the corresponding edge-induced subgraph $G_1(r_0) \subset G$. $d_i = \sum_{j \neq i} I(|w_{ij}| > r_0)$ represents the degree of node i in G_1. We examine $H_0 : G_1$ is a random graph vs. $H_1 : G_1$ is not random. However, the set of edges with $\{\boldsymbol{\beta}_{ij} \neq 0\}$ is unknown and thus needs to be estimated. Under H_0, the multiple correction based thresholding is valid because $h_{ij} = h_{i'j'}|w_{ij} = w_{i'j'}$and the inference is irrelevant to the shuffling order of edges.

Under H_0 that G_1 is a random graph, d_i follows a Poisson distribution (Newman 2002). Thus, we reject H_0 if the distribution d_i statistical significantly deviate from the Poisson distribution. We perform permutation test to assess the significance of the deviation. Since the true r_0 is unknown, we assume it follows a distribution $f(r_0)$. $P(d_i)$ denotes the empirical sample distribution of d_i and $Q(d_i) = \text{Poisson}(d_i)$ with parameter estimated by the random graph model. The Kullback-Leibler divergence based statistic $\int D_{KL}(P(G1)||Q(G1))h(r_0)dr_0 = \int \sum P(d_i) \log(\frac{P(d_i)}{Q(d_i)})h(r_0)dr_0$ is used to measure the deviation. $h(r_0)$ is the prior distribution of the cutoff. In each permutation, $G_1(r_0)$ is randomized by shuffling the order of edge is shuffled (Hanhijarvi 2009). We reject the null if the the observed testing statistic is among the top 5 percentile of statistics from permutations. If we fail to reject H_0, the conventional methods like FDR can be used. If H_0 is rejected, our next goal is to identify the and test latent topological structure of G_1

A2. Derivation of the algorithm for objective function (1)

The optimization of objective function (1) is implemented by exhaustive search for C and estimating $\hat{\mathbf{U}}$ at each C. With a given C,

$$\begin{aligned} &\underset{\hat{\mathbf{U}}=\cup_{c=1}^{C}\hat{\mathbf{U}}_c}{\arg\max}\ \log||\hat{\mathbf{U}}||_1 - \lambda_0 \log||\hat{\mathbf{U}}||_0 \\ =&\underset{\hat{\mathbf{U}}=\cup_{c=1}^{C}\hat{\mathbf{U}}_c}{\arg\max}\ \log\left(\frac{||\hat{\mathbf{U}}||_1}{||\hat{\mathbf{U}}||_0^{\lambda_0}}\right) \doteq \underset{\hat{\mathbf{U}}=\cup_{c=1}^{C}\hat{\mathbf{U}}_c}{\arg\max}\ f(\hat{\mathbf{U}}) \end{aligned} \tag{4}$$

By default $\lambda_0 = 0.5$ reflects balanced covering quality and quantity of true positive edges, and the objective function $\underset{\hat{\mathbf{U}}=\cup_{c=1}^{C}\hat{\mathbf{U}}_c}{\arg\max}\ f(\hat{\mathbf{U}})$ then becomes the well-known problem of k dense subgraph discovery, where $f(\cdot)$ is the density function. The problem has been solved in polynomial time by Goldberg's min-cut algorithm (Goldberg, 1984) and a greedy algorithm with 1/2 approximation by Charikar (2000). In addition, the default topological community structure can be considered as quasi-cliques and the problem can be solved by additive approximation algorithms and local-search heuristics (Tsourakakis et al., 2013). Alternatively, with the mild spatially-invariant assumptions that $\frac{E(w_{ij}|e_{ij})\in G_c}{|E_c|} = \rho_1, \forall c,\ 0 \le c \le C$, and $\frac{E(w_{ij}|i\in V_c, j\in V_{c'})}{|V_c||V_{c'}|} = \rho_0, \forall c,\ 0 \le c \le C$ the primary objective function is equivalent to

$$\begin{aligned} &\underset{\hat{\mathbf{U}}=\cup_{c=1}^{C}\hat{\mathbf{U}}_c}{\arg\min}\ \log\frac{\sum_{c=1}^{C}\sum_{i<j}(w_{ij}|e_{ij}\notin G_c)}{[\sum_{c=1}^{C}\sum_{i<j} I(e_{ij}\notin G_c)]} \\ \doteq &\underset{\hat{\mathbf{U}}=\cup_{c=1}^{C}\hat{\mathbf{U}}_c}{\arg\min}\ \log\sum_{c=1}^{C}\frac{\sum_{i<j}(w_{ij}|e_{ij}\notin G_c)}{|V_c|}, \text{with spatially invariant } \rho_0 \end{aligned} \tag{5}$$

Although the objective function (5) is not convex, the issue of local optima in the discrete optimization can be solved by restarting the algorithm for several times with different initializations and/or through orthonormal transforms (Stella and Shi, 2003 and Bolla, 2013). The proposed algorithm may better extract multiple weighted dense subgraphs (with an unknown number and unknown sizes of dense subgraphs) than the existing algorithms of dense subgraph discovery (Chen et al., 2018). We then choose the optimal C^* by grid searching that maximizes the following criteria:

$$\arg\max_{C^*} \left(\frac{\sum_{c=1}^{C^*} \sum_{i<j} (w_{i,j} | e_{i,j} \in G_c)}{\sum_{c=1}^{C^*} |E_c|} \right)^{\lambda_0} \left(\sum_{c=1}^{C^*} \sum_{i<j} (w_{i,j} | e_{i,j} \in G_c) \right)^{1-\lambda_0}. \quad (6)$$

The criteria (6) can be directly derived from the our primary objective function that $\log(\sum_{c=1}^{C^*} \sum_{i<j} (w_{i,j} | e_{i,j} \in G_c)) - \lambda_0 \log ||\mathbf{U}||_0$. The first term in (6) indicates the 'qaultiy' (the area density) of the extracted subgraphs, while the second term represents the 'quantity' of edges covered by the subgraphs. C^* is selected with optimal quality and quantity in terms of covering informative edges. λ_0 can be tuned to either extract subgrahs with higher area density (i.e. low false positive rates) or covering more high-weight edges using subgraphs with larger sizes (i.e. low false negative rates). In general, C^* selection is robust for λ_0 in the range of 0.4 to 0.7.

In summary, the above procedure can extract latent organized topological structures containing most high-weight edges while controlling the sizes of the topological structures by ℓ_0 norm regularization. Furthermore, we have recently developed more flexible algorithms to extract subgraphs beyond the default community structure, for example, k-partite/rich club and interconnected induced subugrahs can be further detected based on detected quasi-cliques (Chen et al., 2019 and Wu 2019). These more sophisticated topological structures can further improve the objective function

by preserving the high-weight edges inside of more parsimonioulsy-sized subgraphs. The sample codes can be found at https://github.com/shuochenstats/Network_program/tree/master/NTS_package.

A3. Proof of theorem 1: To proof theorem 1, we first establish the following lemma based on the results of theorem 5.2 in Lei et al. (2015) by extending the binary edges to weighted edges, and the proof of Lemma 2 is provided in A4.

Lemma 2. *: Let $\boldsymbol{W}$ be the weighted adjacency matrix of a random graph with entries follow independent distributions defined in [0,1]. Set $\mathbb{E}(\boldsymbol{W}) = \boldsymbol{P} = (p_{ij})_{i,j=1,..,n}$ and $\mathbb{V}(\boldsymbol{W}) = (\mathbb{V}(w_{ij}))_{i,j=1,..,n} = \boldsymbol{V} = (\sigma_{ij}^2)_{i,j=1,..,n}$. Assume that $n(p_{\max} \vee \sigma_{\max}^2) \leq d$ for $d \geq c_0 \log n$ and $c_0 > 0$, where $p_{\max} = \max_{ij} p_{ij}$ and $\sigma_{\max}^2 = \max_{ij} \sigma_{ij}^2$. Then, for any $r > 0$, there exists a constant $C_0 = C_0(r, c_0)$ such that*

$$\|\boldsymbol{W} - \boldsymbol{P}\| \leq C_0\sqrt{d}$$

Lemma 2.1, 5.1 and 5.3 in Lei et al. (2015) are automatically true for the weighted stochastic model. Combining with the Lemma 2, the claim follows from the same calculation as Theorem 3.1 in Lei et al. (2015).

A4. Proof of Lemma 2: In order to bound the supreme of $x^T(\boldsymbol{W} - \boldsymbol{P})y$, we first discretize all pairs of x, y to a finite number of grid points T in the unit ball S defined as follows.

For $t > 0$, let $S_t = \{v \in \mathbb{R}^n : \|v\|_2 \leq t\}$ be the Euclidean ball of radius t and $S = S_1$. For $\delta \in (0, 1/2)$, define

$$T = \left\{x = (x_1, \ldots, x_n) \in S : \sqrt{n}x_i/\delta \in \mathbb{Z}, \forall i\right\}$$

be a set of all grid points within unit ball in $\mathbb{R}^n$ of size $\delta/\sqrt{n}$.

Split all pairs (x, y) in T into light pairs

$$\mathcal{L} = \mathcal{L}(x, y) := \left\{(i, j) : |x_i y_j| \leq \sqrt{d}/n\right\},$$

and heavy pairs

$$\overline{\mathcal{L}} = \overline{\mathcal{L}}(x, y) := \left\{(i, j) : |x_i y_j| > \sqrt{d}/n\right\}.$$

Then the union bound are considered separately in light and heavy pairs. Set $\boldsymbol{W}' = (w'_{ij})_{i,j=1,..,n} = \boldsymbol{W} - \boldsymbol{P}$.

Lemma 3. *(Bounding light pairs)*

$$\mathbb{P}\left(\sup_{x,y\in T}\left|\sum_{(i,j)\in\mathcal{L}(x,y)} x_i y_j w'_{ij}\right| \geq c_0\sqrt{d}\right) \leq \exp\left[-\left(\frac{c_0^2}{4+\frac{4c_0}{3}} - 2\log\left(\frac{7}{\delta}\right)\right)n\right]$$

Proof. Let $u_{ij} = x_i y_j \mathbf{1}(|x_i y_j| \leq \sqrt{d}/n) + x_j y_i \mathbf{1}(|x_j y_i| \leq \sqrt{d}/n)$, then $|u_{ij}| \leq 2\sqrt{d}/n$, and $x^T W' y$ can be written as

$$\sum_{1\leq i<j\leq n} w'_{ij} u_{ij}.$$

For zero-mean independent random variables, apply Bernstein inequality,

$$\begin{aligned}
\mathbb{P}\left[\left|\sum_{i<j} w'_{ij} u_{ij}\right| \geq c_0\sqrt{d}\right] &\leq 2\exp\left(-\frac{\frac{1}{2}c_0^2 d}{\sum_{i<j}\sigma_{ij}^2 u_{ij}^2 + \frac{1}{3}\frac{2\sqrt{d}}{n}c_0\sqrt{d})}\right) \\
&\leq 2\exp\left(-\frac{\frac{1}{2}c_0^2 d}{\sigma_{\max}^2\sum_{i<j} u_{ij}^2 + \frac{2c_0}{3}\frac{d}{n}}\right) \\
&\leq 2\exp\left(-\frac{c_0^2}{4+\frac{4c_0}{3}}n\right)
\end{aligned}$$

□

Lemma 4. *(Bounding heavy pairs) For any given $c_0 > 0$, there exists a constant $C_0(c_0)$, such that with probability $1 - 2n^{-c_0}$*

$$\sup_{x,y\in T}\left|\sum_{(i,j)\in\overline{\mathcal{L}}} x_i y_j w'_{ij}\right| \le C_0\sqrt{d}$$

Proof. To bound the heavy pairs, it suffices to show

$$\sum_{(i,j)\in\overline{\mathcal{L}}} x_i y_j w_{ij} = O(\sqrt{d})$$

with high probability.

WLOG, consider the heavy pairs in set,

$$\overline{\mathcal{L}}_1 = \left\{(i,j) \in \overline{\mathcal{L}} : x_i > 0, y_j > 0\right\}.$$

Use Lei's notation in a continuous version as follows:

- $I_1 = \left\{i : \frac{\delta}{\sqrt{n}} \le x_i \le \frac{2\delta}{\sqrt{n}}\right\}, I_s = \left\{i : \frac{\delta}{\sqrt{n}}2^{s-1} < x_i \le \frac{\delta}{\sqrt{n}}2^s\right\}$ for $s = 2, 3, \ldots, \left\lceil \log_2 \frac{\sqrt{n}}{\delta}\right\rceil$;$J_1 = \left\{j : \frac{\delta}{\sqrt{n}} \le y_j \le \frac{2\delta}{\sqrt{n}}\right\}, J_t = \left\{j : \frac{\delta}{\sqrt{n}}2^{t-1} < y_j \le \frac{\delta}{\sqrt{n}}2^t\right\}$ for $t = 2, 3, \ldots, \left\lceil \log_2 \frac{\sqrt{n}}{\delta}\right\rceil$.

- Let $s(I, J)$ be the set of all possible distinct edges (i, j) with index $i \in I$ and $j \in J$. $e(I, J)$ to be the summation of edge weights in node sets I and J: $e(I, J) = \sum_{(i,j)\in s(I,J)} w_{ij}$.
- Let $\mu(I, J) = \mathbb{E}e(I, J), \overline{\mu}(I, J) = p_{\max}|I||J|$.
- $\lambda_{st} = e\left(I_s, J_t\right)/\overline{\mu}_{st}$, where $\overline{\mu}_{st} = \overline{\mu}\left(I_s, J_t\right)$
- $\alpha_s = |I_s|\, 2^{2s}/n, \beta_t = |J_t|\, 2^{2t}/n, \sigma_{st} = \lambda_{st}\sqrt{d}2^{-(s+t)}$

Based on the following lemma 5 and 6 and using exactly the same calculations as Lei et al. (2015) supplemental materials part 4, we obtain the claim.

□

Lemma 5. *For $c_0 > 0$, there exists constant $c_1 = c_1(c_0)$ such that with probability at least $1 - n^{-c_0}$, $\sum_{j=1}^{n} w_{ij} \le c_1 d$.*

Proof. Using Bernstein inequality:

$$\mathbb{P}\left(\sum_{j=1}^{n} w_{ij} \ge c_1 d\right) \le \mathbb{P}\left(\sum_{j=1}^{n} w'_{ij} \ge (c_1 - 1)d\right) \le \exp\left[-\frac{\frac{1}{2}(c_1-1)^2 d^2}{\sum_{j=1}^{n}\sigma_{ij}^2 + \frac{1}{3}(c_1-1)d}\right]$$
$$\le \exp\left[-\frac{\frac{1}{2}(c_1-1)^2 d^2}{n\sigma_{\max}^2 + \frac{1}{3}(c_1-1)d}\right] \le \exp\left[-\frac{\frac{1}{2}(c_1-1)^2 d}{1 + \frac{1}{3}(c_1-1)}\right] \le n^{-\frac{3c_0(c_1-1)^2}{2c_1+4}}$$

□

Lemma 6. *For $c_0 > 0$, there exists constants $c_2 = c_2(c_0)$ and $c_3 = c_3(c_0)$, both larger than 1, such that with probability at least $1 - 2n^{-c_0}$, for any I, $J \subset [n]$ with $|I| \le |J|$, at least one of the following holds:*

1. $\frac{e(I,J)}{\bar{\mu}(I,J)} \le ec_2$
2. $e(I,J)\log\frac{e(I,J)}{\bar{\mu}(I,J)} \le c_3|J|\log\frac{n}{|J|}$

Proof. If $|J| \ge n/e$, then from Lemma 3, we can get with probability at least $1 - n^{-c_0}$

$$\frac{e(I,J)}{\bar{\mu}(I,J)} = \frac{e(I,J)}{p_{\max}|I||J|} \le \frac{|I|c_1 d}{p_{\max}|I||J|} \le \frac{|I|c_1 d}{p_{\max}|I|n/e} \le c_1 e$$

If $|J| > n/e$, let $k > 1$, then from Chernoff Bound:

$$\begin{aligned}\mathbb{P}[e(I,J) \geq k\overline{\mu}(I,J)] &= \mathbb{P}\left[\sum_{(i,j)\in s(I,J)} w_{ij} \geq k\bar{\mu}(I,J)\right] \\ &\leq \exp(-\bar{\mu}(I,J)(k\ln k - (k-1))) \\ &\leq \exp\left[-\frac{1}{2}(k\ln k)\bar{\mu}\right]\end{aligned}$$

The last inequality holds for $k \geq 8$. For a given number $c_3 > 0$, define $t(I,J)$ as the unique value of t such that $t\log t = \frac{c_3|J|}{\bar{\mu}(I,J)}\log\frac{n}{|J|}$. Let $k(I,J) = \max\{8, t(I,J)\}$. Then, we have

$$\mathbb{P}\left[\exists(I,J) : |I| \leq |J| \leq \frac{n}{e}, e(I,J) \geq k(I,J)\overline{\mu}(I,J)\right] \leq n^{-\frac{1}{2}(c_3-12)}.$$

Hence, with probability $1 - n^{-\frac{1}{2}(c_3-12)}$ we have $e(I,J) \leq k(I,J)\bar{\mu}(I,J)$, and further implies

$$e(I,J)\log\frac{e(I,J)}{\overline{\mu}(I,J)} \leq c_3|J|\log\frac{n}{|J|}.$$

Therefore, the claim is true for $c_2 = \max\{c_1, 8/e\}$ and $c_3 = 2c_0 + 12$. □

A5. Lemma 7 For a graph $G(V,E)$, a subgraph indexed by $S \subseteq V$ is said to be a $\gamma-$quasi-clique, if the subgraph has at least $\gamma\begin{pmatrix}|S|\\2\end{pmatrix}$ edges, where $\gamma \in [0,1]$ is a parameter. For a random graph $G(n,p)$ and $1 > \gamma > p > 0$, let $A_n^\gamma(k)$ be the maximum number of edges that can be included in at most k $\gamma-$quasi-clique. Then, the maximum number of edges that can be included in at most $\ln n$ $\gamma-$quasi-clique is $\mathcal{O}((\ln n)^3)$, in other words,

$$A_n^\gamma(\ln n) \leq \frac{1}{2}\eta^2(\ln n)^3 \quad a.s. \quad \text{with } \eta = \eta(\gamma) = \frac{2}{\ln\left[\left(\frac{\gamma}{p}\right)^\gamma\left(\frac{1-\gamma}{1-p}\right)^{1-\gamma}\right]}.$$

Proof. If the number of edges included in at most k $\gamma-$quasi-clique is $A_n^\gamma(k)$, there must be at least one $\gamma-$quasi-clique which include edges no less than $\lfloor A_n^\gamma(k)/k \rfloor$. From theorem 1 in Veremyev et al. (2012),

$$M_n^\gamma \leq \frac{2}{\ln\left[\left(\frac{\gamma}{p}\right)^\gamma \left(\frac{1-\gamma}{1-p}\right)^{1-\gamma}\right]} \ln n \quad a.s.$$

where M_n^γ is the maximum number of vertices in a $\gamma-$quasi-clique. Hence,

$$A_n^\gamma(1) \leq \frac{1}{2}\eta^2(\ln n)^2 \quad a.s. \quad \text{with } \eta = \frac{2}{\ln\left[\left(\frac{\gamma}{p}\right)^\gamma \left(\frac{1-\gamma}{1-p}\right)^{1-\gamma}\right]}$$

Therefore, $A_n^\gamma(k) \leq k\eta^2(\ln n)^2/2$ almost surely.

A6. Edge-level inference by GLEN

GLEN provides network level inference by reporting phenotype related subnetwork G_c instead of $e_{ij} \in G_c$. We argue that the detected subnetworks can also assist edge-level inference by applying an empirical Bayes based adaptive thresholding strategy. The details of method is introduced in section 2.2 Chen et al. (2018). When phenotype related edges consist of organized topological structures, the false positive and false negative discovery rates for individual edges are lower than conventional universal threshold multiple testing methods.

Theorem 4. *Let* $F_0(x) = P(w_{ij} \leq x|\delta_{ij} = 0)$ *and* $F_1(x) = P(w_{ij} \leq x|\delta_{ij} \neq 0)$. *Under the conditions,*

$$\frac{F_0(z_0) - F_0(z_{0,\ out})}{F_0(z_{0,in}) - F_0(z_0)} > \frac{E(\sum_{i<j} I(\delta_{ij} = 0|e_{ij} \in G_c))}{E(\sum_{i<j} I(\delta_{ij} = 0|e_{ij} \in \overline{G}_c))}$$

and

$$\frac{F_1\left(z_0\right)-F_1\left(z_{0,\ out}\right)}{F_1\left(z_{0,in}\right)-F_1\left(z_1\right)}>\frac{E(\sum_{i<j} I(\delta_{ij}\neq 0|e_{ij}\in G_c))}{E(\sum_{i<j} I(\delta_{ij}\neq 0|e_{ij}\in \overline{G}_c))},$$

where z_0 is a universal threshold value and $z_{0,in}$, $z_{0,out}$ are threshold values inside and outside the community structure, we have

(1) the expected false positively thresholded edges by using the network level inference (GLEN) are less than the universal thresholding method

$$E\left(\sum_{i<j} I\left(\widehat{\delta}_{ij}^{GLEN}=1|\delta_{ij}=0\right)\right)\leq E\left(\sum_{i<j} I\left(\widehat{\delta}_{ij}^{Univ}=1|\delta_{ij}=0\right)\right);$$

(2) the expected false negatively thresholded edges by using GLEN are less than the universal thresholding method

$$E\left(\sum_{i<j} I\left(\widehat{\delta}_{ij}^{GLEN}=0|\delta_{ij}=1\right)\right)\leq E\left(\sum_{i<j_t} I\left(\widehat{\delta}_{ij}^{Univ}=0|\delta_{ij}=1\right)\right)$$

See the proof of theorem 1 in Chen et al. (2018).

A7. Permutation test

Graph edge permutation vs. graph node permutation

There are two sets of elements in a graph: the set of vertices and the set of edges as in $G=\{V,E\}$. Correspondingly, there are two options of permutation: permuting nodes or edges. First, we consider the permutation of nodes as a reordering processπ, which is an 'edge-preserving bijection. If two nodes a and b are connected in graph$G$, then in the node-permuted graph$H=\pi(G)=\{\pi(V),F\}$, then $\pi(a)$ and $\pi(b)$ are connected:$E_{ab}=1\Leftrightarrow F_{\pi(a)\pi(b)}=1$. G and H are isomorphic graphs $G\simeq H$ (see Figs 1d and 1e). The GLEN subnetwork detection algorithms reorder and group the

nodes to uncover these latent topological patterns. In contrast, the edge permutation is different because it permutes the order of edges and is not 'edge-preserving bijection. For example, two nodes a and b are connected in a graph$G$; the edge-permuted graph L(e.g. by permutationμ) that $L = \mu(G) = \{V, F = \mu(E)\}$ and in L,a and b are only connected with probability ofp_G, where p_G is number of connected edges in G divided by$n \times n/2$. Therefore, the above two events are independent: $\{E_{ab} = 1\} \perp \{F_{ab} = 1\}$. Hence, though there is an organized pattern in$G$, the edge permuted graph $L = \mu(G)$ becomes a random graph without any organized patterns. The connectivity testing p-value matrix after edge permutation represents a random graph where each edge has the identical probability such that$p_{i,j} < p_0$. Therefore, the edge permutation can be used to test the organized topological pattern. The GLEN1 permutation test simulates the null by shuffling the group labels (i.e. the order of the covariate of interest) while GLEN2 permute the order of edge to simulate a random graph.

Network level test statistic. We propose a new test statistic in the permutation tests, which is specifically tailored for the network as an object. Our goal is to select disease-related subnetwork that is both informative (including the largest number of informative edges that is possible) and efficient (most edges in each subnetwork are significant). Generally, these two factors are conflicting: including more informative edges (resulting in larger networks) can reduce the average significance of edges in the subnetwork whereas restricting only very significant edges in the subnetwork can shrink the network size and overall include less informative edges in G. We develop a new test statistic to integrate these two aspects of the network object. The new test statistic

$$T^m_{max,NTS} = \max_{k \in K^m} \{|E^m_k|(\bar{x}^m_k - 1 - \log(\bar{x}^m_k))\}$$

$$\bar{x}_k^m = -2 \sum_{i,j \in G_k^m} \log p_{i,j} / 2|E_k^m|$$

is derived based on Fishers combination test and Chernoff bound of χ^2 the cumulative distribution function. The test statistic is negative (and generally the network is not significant) when it includes a small proportion of informative edges. Therefore, the test statistic jointly evaluates the effect size of edges in the subnetwork and the size of each selected subnetwork which together are an index of the information of a subnetwork.

A8. fMRI data acquisition and pre-processing

All participants provided written informed consent that had been approved by the University of Maryland Internal Review Board. All participants were evaluated using the Structured Clinical Interview for the DSM-IV diagnoses. We recruited medicated patients with an Axis I diagnosis of schizophrenia through the Maryland Psychiatric Research Center and neighboring mental-health clinics. We recruited control subjects, who did not have an Axis I psychiatric diagnosis, through media advertisements. Exclusion criteria included hypertension, hyperlipidemia, type 2 diabetes, heart disorders, and major neurological events, such as stroke or transient ischemic attack. Illicit substance and alcohol abuse and dependence were exclusion criteria. Data were acquired using a 3-T Siemens Trio scanner equipped with a 32 channel head coil at the University of Maryland Center for Brain Imaging Research. A T1-weighted structural image (MP-RAGE: 1 mm isotropic voxels, 256 x 256 mm FOV, TR/TE/TI = 1900/3.45/900ms) was acquired for anatomical reference. Fifteen minutes of rfMRI was collected on each subject. During the resting scans, subjects were given a simple instruction to rest and keep their eyes closed. Head motion was mini-

mized using foam padding, foam molding, and tapes. RfMRI were acquired over 39 axial, interleaving slices using a gradient-echo EPI sequence (450 volumes, TE/TR = 27/2000 ms; flip angle = 90^o; FOV = 220x220 mm; image matrix = 128x128; in-plane resolution 1.72x1.72mm. Following the previously published procedures, data were preprocessed in AFNI and MATLAB (MathWorks, Inc., Natick, MA). Volumes were slice-timing aligned and motion corrected to the base volume that minimally deviated from other volumes using an AFNI built-in algorithm. After linear detrending of the time course of each voxel, volumes were spatially normalized and resampled to Talairach space at 3x3x3 mm3, spatially smoothed (FWHM 6 mm), and temporally low-pass filtered (fcutoff = 0.1 Hz). For functional connectivity analyses, the six rigid head-motion parameter time courses and the average time course in white matter were treated as nuisance covariates. A white matter mask was generated by segmenting the high-resolution anatomical images and down-gridding the obtained white matter masks to the same resolution as the functional data. These nuisance covariates regress out fluctuations unlikely to be relevant to neuronal activity.

A9. Tables of brain regions In the following tables, we list the region names and coordinates of subnetworks from D^1 and D^2.

Table 1: Region names and coordinates in the subnetwork of D^1

Region Name	x	y	z
Precentral R	-39	-6	51
Rolandic Oper L	-47	-8	14
Rolandic Oper R	53	-6	15
Supp Motor Area L	-5	5	61
Supp Motor Area R	9	0	62
Frontal Sup Medial L	-5	49	31
Insula L	-35	10	3
Insula R	39	6	2
Cingulum Ant L	-4	35	14
Cingulum Ant R	8	37	16
Cingulum Mid L	-5	-15	42
Cingulum Mid R	8	-9	40
Postcentral L	-42	-23	49
Postcentral R	41	-25	53
Precuneus L	-7	-56	48
Paracentral Lobule L	-8	-25	70
Paracentral Lobule R	7	-32	68
Heschl L	-42	-19	10
Heschl R	46	-17	10
Temporal Sup L	-53	-21	7
Temporal Sup R	58	-22	7
Temporal Pole Sup R	48	15	-17

Table 2: Region names and coordinates in the subnetwork of D^2

Region Name	x	y	z
Rolandic Oper L	-47	-8	14
Rolandic Oper R	53	-6	15
Supp Motor Area L	-5	5	61
Supp Motor Area R	9	0	62
Frontal Sup Medial L	-5	49	31
Insula L	-35	10	3
Insula R	39	6	2
Cingulum Ant L	-4	35	14
Cingulum Ant R	8	37	16
Cingulum Mid L	-5	-15	42
Cingulum Mid R	8	-9	40
Postcentral L	-42	-23	49
Postcentral R	41	-25	53
Precuneus L	-7	-56	48
Paracentral Lobule L	-8	-25	70
Paracentral Lobule R	7	-32	68
Heschl L	-42	-19	10
Heschl R	46	-17	10
Temporal Sup L	-53	-21	7
Temporal Sup R	58	-22	7
Temporal Pole Sup R	48	15	-17

References

Benjamini, Y. and Y. Hochberg (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological) 57*(1), 289–300.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological) 36*(2), 192–225.

Bolla, M. (2013). *Spectral clustering and biclustering: Learning large graphs and contingency tables*. John Wiley & Sons.

Bowman, F. D. (2005). Spatio-temporal modeling of localized brain activity. *Biostatistics 6*(4), 558–575.

Bowman, F. D., B. Caffo, S. S. Bassett, and C. Kilts (2008). A bayesian hierarchical framework for spatial modeling of fmri data. *NeuroImage 39*(1), 146–156.

Cai, T., H. Li, J. Ma, and Y. Xia (2018). Differential markov random field analysis with an application to detecting differential microbial community networks. *Biometrika*.

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, S., F. D. Bowman, and Y. Xing (2019). Detecting and testing altered brain connectivity networks with k-partite network topology. *Computational Statistics & Data Analysis*.

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 (2018). Bayesian modeling of dependence in brain connectivity data. *Biostatistics*.

Craddock, R. C., S. Jbabdi, C.-G. Yan, J. T. Vogelstein, F. X. Castellanos, A. Di Martino, C. Kelly, K. Heberlein, S. Colcombe, and M. P. Milham (2013). Imaging human connectomes at the macroscale. *Nature methods 10*(6), 524.

Derado, G., F. D. Bowman, and C. D. Kilts (2010). Modeling the spatial and temporal dependence in fmri data. *Biometrics 66*(3), 949–957.

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., T. Hastie, I. Johnstone, R. Tibshirani, et al. (2004). Least angle regression. *The Annals of statistics 32*(2), 407–499.

Eklund, A., T. E. Nichols, and H. Knutsson (2016). Cluster failure: Why fmri inferences for spatial extent have inflated false-positive rates. *Proceedings of the national academy of sciences 113*(28), 7900–7905.

Fan, J. and J. Lv (2008). Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 70*(5), 849–911.

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.

Gionis, A. and C. E. Tsourakakis (2015). Dense subgraph discovery: Kdd 2015 tutorial. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 2313–2314. ACM.

Goldberg, A. V. (1984). *Finding a maximum density subgraph.* University of California Berkeley, CA.

Hazimeh, H. and R. Mazumder (2018). Fast best subset selection: Coordinate descent and local combinatorial optimization algorithms. *arXiv preprint arXiv:1803.01454*.

Higgins, I. A., Y. Guo, S. Kundu, K. S. Choi, and H. Mayberg (2018). A differential degree test for comparing brain networks. *arXiv preprint arXiv:1809.11098*.

Khuller, S. and B. Saha (2009). On finding dense subgraphs. In *International Colloquium on Automata, Languages, and Programming*, pp. 597–608. Springer.

Kim, Y. (2014). Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.

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.

Lazar, N. (2008). *The statistical analysis of functional MRI data*. Springer Science & Business Media.

Lei, J., A. Rinaldo, et al. (2015). Consistency of spectral clustering in stochastic block models. *The Annals of Statistics 43*(1), 215–237.

Li, R., W. Zhong, and L. Zhu (2012). Feature screening via distance correlation learning. *Journal of the American Statistical Association 107*(499), 1129–1139.

Li, X., S. Xie, D. Zeng, and Y. Wang (2018). Efficient 0-norm feature selection based on augmented and penalized minimization. *Statistics in medicine 37*(3), 473–486.

Lindquist, M. A. (2008). The statistical analysis of fmri data. *Statistical science 23*(4), 439–464.

Lukemire, J., S. Kundu, G. Pagnoni, and Y. Guo (2017). Bayesian joint modeling of multiple brain functional networks. *arXiv preprint arXiv:1708.02123*.

Lynall, M.-E., D. S. Bassett, R. Kerwin, P. J. McKenna, M. Kitzbichler, U. Muller, and E. Bullmore (2010). Functional connectivity and brain networks in schizophrenia. *Journal of Neuroscience 30*(28), 9477–9487.

Miyauchi, A. and N. Kakimura (2018). Finding a dense subgraph with sparse cut. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pp. 547–556. ACM.

Narayan, M., G. I. Allen, and S. Tomson (2015). Two sample inference for populations of graphical models with applications to functional connectivity. *arXiv preprint arXiv:1502.03853*.

Newman, M. E. and M. Girvan (2004). Finding and evaluating community structure in networks. *Physical review E 69*(2), 026113.

Risk, B. B., M. C. Kociuba, and D. B. Rowe (2018). Impacts of simultaneous multislice acquisition on sensitivity and specificity in fmri. *NeuroImage 172*, 538–553.

Rohe, K., S. Chatterjee, B. Yu, et al. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics 39*(4), 1878–1915.

Shaddox, E., C. B. Peterson, F. C. Stingo, N. A. Hanania, C. Cruickshank-Quinn, K. Kechris, R. Bowler, and M. Vannucci (2018). Bayesian inference of networks across multiple sample groups and data types. *Biostatistics*.

Shen, X., W. Pan, and Y. Zhu (2012). Likelihood-based selection and sharp parameter estimation. *Journal of the American Statistical Association 107*(497), 223–232.

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.

Stella, X. Y. and J. Shi (2003). Multiclass spectral clustering. In *null*, pp. 313. IEEE.

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.

Veremyev, A., V. Boginski, P. A. Krokhmal, and D. E. Jeffcoat (2012, 12). Dense percolation in large-scale mean-field random networks is provably explosive. *PLOS ONE 7*(12), 1–14.

Wang, W., X. Zhang, and L. Li (2019). Common reducing subspace model and network alternation analysis. *Biometrics*.

Warnick, R., M. Guindani, E. Erhardt, E. Allen, V. Calhoun, and M. Vannucci (2018). A bayesian approach for estimating dynamic functional network connectivity in fmri data. *Journal of the American Statistical Association 113*(521), 134–151.

Xia, C. H., Z. Ma, Z. Cui, D. Bzdok, D. S. Bassett, T. D. Satterthwaite, R. T. Shinohara, and D. M. Witten (2019). Multi-scale network regression for brain-phenotype associations. *bioRxiv*, 628651.

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, T., Q. Yin, B. Caffo, Y. Sun, D. Boatman-Reich, et al. (2017). Bayesian inference of high-dimensional, cluster-structured ordinary differential equation models with applications to brain connectivity studies. *The Annals of Applied Statistics 11*(2), 868–897.