HHS Public Access

Author manuscript

Comput Stat Data Anal. Author manuscript; available in PMC 2021 January 01.

Published in final edited form as:

Comput Stat Data Anal. 2020 January ; 141: 109–122. doi:10.1016/j.csda.2019.06.007.

Detecting and Testing Altered Brain Connectivity Networks with K-partite Network Topology

Shuo Chen[a,b,*], **F. DuBois Bowman**[c], **Yishi Xing**[d]

[a]Division of Biostatistics and Bioinformatics, School of Medicine, University of Maryland, Baltimore, MD, USA

[b]Maryland Psychiatric Research Center, School of Medicine, University of Maryland, Baltimore, MD, USA

[c]Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI, USA

[d]Department of Electrical and Computer Engineering, University of Maryland, College Park, MD, USA

Abstract

Emerging brain connectivity network studies suggest that interactions between various distributed neuronal populations may be characterized by an organized complex topological structure. Many neuropsychiatric disorders are associated with altered topological patterns of brain connectivity. Therefore, a key inquiry of connectivity analysis is to detect group-level differentially expressed connectome patterns from the massive neuroimaging data. Recently, statistical methods have been developed to detect differentially expressed connectivity features at a *subnetwork* level, extending more commonly applied *edge* level analysis. However, the graph topological structures in these methods are limited to community/cliques which may not effectively uncover the underlying complex and disease-related brain circuits/subnetworks. Building on these previous subnetwork detection methods, a new statistical approach is developed to automatically identify the latent differentially expressed brain connectivity subnetworks with k-partite graph topological structures from large brain connectivity matrices. In addition, statistical inferential techniques are provided to test the detected topological structure. The new methods are evaluated via extensive simulation studies and then applied to resting state fMRI data (24 cases and 18 controls) for Parkinson's disease research. A differentially expressed connectivity network with the k-partite graph topological structure is detected which reveals underlying neural features distinguishing Parkinson's disease patients from healthy control subjects.

Keywords

brain network statistics; connectivity; k-partite graph; network topological statistics; Parkinson's disease

[*]Corresponding author: shuochen@som.umaryland.edu (Shuo Chen).

1. Introduction

Brain connectomic research has suggested that neuropsychiatric disorders are associated with altered interactions between distributed neuronal populations and brain regions (Buckner et al, 2009; Craddock et al, 2013; Stam, 2014; Fornito et al, 2015; Chen et al, 2016a). Recent neuroimaging studies have used graph theory as a tool to understand 5 the brain connectivity patterns, which denote brain regions as nodes, and connections between regions as edges (Bullmore and Sporns, 2009; Rubinov and Sporns, 2010; Craddock et al, 2013; Biswal et al, 2010; Yeo et al, 2011; Sporns, 2014; Smith et al, 2015). Such studies have identified connectome-phenotype relationships by leveraging these graph techniques, mainly using network graph descriptive metrics (Bullmore and Sporns, 2009; Seeley et al, 2009; Stam et al, 2009; Rubinov and Sporns, 2010; Achard et al, 2012; van den Heuvel and Sporns, 2013; Crossley et al, 2013; Crossley et al, 2014; Stam, 2014; Fornito et al, 2015).

The overarching goal of brain connectivity network/circuitry research is to enhance understanding of underlying pathophysiological mechanisms and clinically useful predictions concerning disease diagnosis and treatment selection (Fornito et al, 2013; Craddock et al, 2013; Fornito et al, 2015).

However, generally it is challenging to detect such differential connectivity subnetworks that simultaneously i) control false positive rate and obtain sufficient statistical power; ii) reflect complex connectome topological properties; iii) are spatially localized (edge - specific) for explicit clinical interpretation and pathophysiological mechanism discovery; and iv) are reliable and reproducible (van Diessen et al, 2013; Simpson et al, 2015). A driver of these challenges is the nature of connectome data, containing complex topological structure and high-dimensionality.

Most network graph metrics (e.g. small-worldness and modularity) summarize a set of edges as individual measures and lose localized connectivity information (edge - specific). Such measures may lead to difficult clinical interpretability and may lack specificity and sensitivity (Bullmore and Sporns, 2009; Rubinov and Sporns, 2010; Simpson et al, 2015). On the other hand, the mass univariate analyses may keep localized information but are subject to the trade-off between controlling false positive findings and lack of statistical power. Direct application of the family-wise error control (FWER) and false positive discovery rate (FDR) could successfully prohibit spurious positive findings, yet they may be overly conservative and reduce the statistical power and lead to few or no significant findings. To mitigate such trade-offs, many studies pre-define networks of interest to lower the multiple testing burden. But, pre-defined networks are limited and may exclude true signals. Recently, more advanced statistical methods leverage multivariate models to link the edge connection strength and overall topological structure to improve model estimation (Simpson et al, 2012; Simpson and Laurienti, 2015; Narayan and Allen, 2016; Monti et al, 2017). However, these may not allow to automatic detection of differentially expressed subgraphs.

Differentially expressed edges may be distributed in an organized pattern rather than randomly in the brain. Therefore, we consider the union of differentially expressed edges and their organized topological pattern as a potential brain connectivity subnetwork biomarker candidate. Statistical methods have been developed to test the statistical significance of a subset of edges, for example, network based statistic (NBS) and parsimonious differential network detection (Pard) (Zalesky et al, 2010; Chen et al, 2015a). The Pard algorithm can recognizes differentially expressed edges in latent and organized topological structures by applying the principle of parsimony. The detected latent topological structures of differentially expressed subnetworks bring additional information for group-level inference and thus can increase statistical power while effectively controlling multiple testing false positive errors. Hence, automatic detection of latent topological structures in imaging data offers a viable pathway to connectivity network biomarker discovery. In addition, the detected topological structure may help reveal the underlying neurophysiological and neuropathological mechanisms for brain disorders. In graph theory, a clique (or community) refers to a subgraph such that every pair of distinct nodes in the subset are connected. Here, a clique/community refers to a differential network suggesting most edges between the nodes within the subgraph are differentially expressed. The Pard algorithm only detects clique/community differential networks, however more complex topological structure may appear.

In graph theory, a k-partite subgraph is a graph whose nodes could be partitioned into k distinct sets such that the nodes in the same set are not connected but nodes from different sets are connected. For brain connectivity analysis, the quantity of an edge is often continuous (rather than binary) that represents i) the connection strength for one subject (e.g. the Pearson correlation coefficient) at an individual subject level or ii) to what extent the connection is differentially expressed between different clinical groups (e.g. a test statistic). In this paper, we use distinct symbols to denote these two cases. When we try to identify the differentially expressed subnetworks, we use the latter case to detect latent topological structures. Specifically, we refer to a “k-partite phenomenon” (for illustration we let $k = 2$) when: i) there are two sets of nodes and the edges connecting nodes within the same set show non-differential connectivity strengths between clinical groups; ii) the between set connections demonstrate group-wise difference. Many studies have reported disrupted long-range connectivity patterns by neuropsychiatric disorders, for example, Parkinson’s disease (Baggio et al, 2014; Lopes et al, 2016). These altered long-range connections may often be distributed in an organized yet hidden k-partite subgraph as in our data example. If the “k-partite phenomenon” exists in a group-wise connectivity study, we may infer that the brain disease alters the inter-community interactions between two sets of neural populations (nodes) but not the intra-community interactions. One possible explanation is that the within community connections may be linked to basic human brain functions which are more reliable and well-wired for both patients and healthy controls (Göttlich et al, 2013), whereas the inter-community connections are more vulnerable to the pathophysiology of diseases and more likely to show patient-healthy control differences(Baggio et al, 2014; Lopes et al, 2016).

However, in practice, the “k-partite phenomenon” of neuroimaging data is latent and not directly observable. In the statistics and machine learning, uncovering latent k-partite

subgraphs in a graph G has been a longstanding challenge. A few algorithms have been developed to recognize the latent k-partite subgraph (Long et al, 2006; Neubauer and Obermayer, 2009; Hartsperger et al, 2010). However, these methods are not applicable to detect potential brain connectivity networks biomarkers because i) they rely on prior knowledge of k; ii) they do not allow k-partite *subgraph* detection in a relatively large graph; and iii) the computational costs of these iterative algorithms are high. Moreover, our input data is a weighted complete graph (i.e. weighted adjacency matrix) and most algorithms only handle binary input adjacency matrices.

Motivated by these needs of group-level brain connectivity network analysis, we develop novel algorithms to detect k-partite subnetworks from a large weighted adjacency matrix. In addition, we develop a new statistical inferential procedure to determine whether the detected differentially expressed subnetwork has a k-partite topological structure. Our method is characterized by several appealing features as follows: i) it can identify altered (by clinical or experimental conditions) connectivity subnetworks with the k-partite topological structure from whole brain network from the whole brain connectivity network; ii) the results (i.e. altered subnetworks)are at the network level which can simultaneously reveal the complex graph topological structure and specific locations of differentially expressed connectivity features; iii)comparing with the conventional multiple testing control methods (e.g. FDR), our approach increases the statistical power by integrating the spatial constraints and k-partite topological structure into the statistical inference.

The remainder of the paper is organized as follows. In section 2, we present the proposed k-partite graph detection algorithm and statistical tests. We evaluate the performance of the proposed k-partite detection algorithm by simulation studies in Section 3. In Section 4, we apply these methods to our motivating resting-state fMRI Parkinson's disorder (PD) study, and we detect a differentially expressed connectomic subnetwork with a k-partite topological structure and statistical significance. We conclude with a discussion in Section 5.

2. Methods

2.1. Background

We consider data from S subjects, representing distinct subgroups (e.g. based on the presence of clinical diagnosis). fMRI data from each subject undergoes preprocessing and is used to calculate functional connectivity between all pairs of n nodes or brain regions, with whole-brain connectivity for a subject s represented by an $n \times n$ symmetric matrix $\mathbf{M}^s$. Therefore, the overall data set is denoted by $\mathbb{M} = \left\{\mathbf{M}^2, \cdots, \mathbf{M}^S\right\}$. The spatial location (3D coordinates) of a node i ($i \in 1, \cdots, n$) is identical for all subjects. The whole brain connectome network can be represented by a graph $G = \{V, E\}$, where V is the set of nodes ($|V| = n$) and E is the set of edges. We denote M_{ij}^s as the connectivity metric between node i and node j for subject s. For instance, Pearson correlation coefficient is the most commonly used metric for the functional connectivity (FC), and each edge represents the correlation between two fMRI time courses from the respective pair of nodes (Zalesky et al, 2012; Kim et al, 2015a).

Next, we perform statistical analysis (e.g. two sample test or regression for a case-control study) on each edge M_{ij} across subjects while accounting for the covariance between edges (Chen et al, 2018), and record a p-value p_{ij} between nodes i and j. Although direct inference on all $\{p_{ij}\}$ or the corresponding test statistics by using FWER and FDR could be performed, the results are often overly conservative due to ignoring the spatial and topological dependence structure between edges (Craddock et al, 2013).

Therefore, instead of making inference by p_{ij} on each edge we aim to detect differentially expressed subgraphs ($G_c \subset G$) fusing edge-wise inference and graph combinatoric probability. We use a weighted matrix **W** with elements $W_{ij} = -\log(p_{ij})$, as our initial input weighted adjacency matrix (see Figure 2a). W_{ij} serves as a metric to delineate how differentially expressed two clinical groups are with respect to a corresponding edge. We refer to edges as informative if there is a high degree of differential expression between groups. **W** is the input matrix for both NBS and Pard algorithms, and the outputs of these algorithms are clique/community differential subnetworks $\{G_c\}$ ($c = 1, \cdots, C$ and C is the number of detected communities) with significance levels (p-values). In this article, we develop a new **k**-**p**artite **g**raph **d**etection (KPGD) algorithm to identify the latent k-partite structure of differentially expressed connectivity networks.

2.2. k-partite graph detection algorithm

We let G_c be a k-partite subgraph. Note that, as illustrated in Figure 1a, the k-partite pattern of $G_c = \{V_c, E_c\}$ is latent in G. To observe the k-partite graph explicitly as shown in Figure 1b, we need to reorder and group the nodes. In our algorithm, we seek an edge preserving bijective mapping (node reordering) function $\pi : G_c \rightarrow H_c$ or $H_c = \pi(G_c)$. Under such mapping π, G_c is isomorphic to H_c ($G_c \simeq H_c$), where $H_c = \{V_c', E_c'\}$. The bijective mapping function π permutes the order of nodes (in columns and rows of the adjacency matrix) simultaneously. If nodes i and j are connected, denoted $i \sim j$, then after permutation $\pi(i) \sim \pi(j)$. Figure 1b illustrates a desirable mapping function π because in $H_c = \pi(G_c)$, the bipartite structure is directly apparent.

One challenge in determining k-partite graphs is that the number of all possible permutations is massive with $n!$ permutations for a graph with n nodes. For example, when n=100 there are more than 10×10^{157} possible permutations. Therefore, it is impractical to search all permutations, and we need an algorithm to seek an appropriate mapping π that reveals the k-partite structure. Note that edges in our input matrix $\mathbf{W}_c$ are weighted. The key heuristic of our algorithm is that the target mapping function π allocates more informative edges to off-diagonal blocks and less informative edges along the diagonal blocks. The diagonal blocks represent independent sets of a k-partite graph, where edges within the same independent set are less differentially expressed between groups.

We define the objective function of the KPGD algorithm as:

$$\arg\max_{C,\{G_c\}} \sum_{c=1}^{C} \exp\Big[\log\Big\{\sum (w_{i,j} \,\Big|\, e_{i,j} \in G_c)\Big\} - \lambda_0 \log(|E_c|)\Big], \quad (1)$$

with following definitions and conditions:

1. $\cup_{c=1}^{C} V_c = V$, $\cap_{c=1}^{C} V_c = \varnothing$ and $\cup_{c=1}^{C} E_c \subseteq E$.
2. the size of a subgraph G_c is determined by the number of edges $|G_c| = |E_c|$;
3. G_c ($c = 1, \cdots, C$) is a k-partite subgraph that $G_c = \{V_c, E_c\}$ and $|V_c| \geq 1$;
4. Within each G_c, there are $1 \leq K_c \leq |V_c|$ independent sets $\{A_k^c\}$ and $A_k^c = \{V_k^c, E_k^c\}$ that $\cup_{kc=1}^{Kc} V_c^k = V_c$ and $e_{i,j} = \varnothing$ if both $i, j \in A_k^c$

Formula 1 maximizes the informative edges in a set of k-partite subgraphs, while penalizing the sizes of the subgraphs. λ_0 is the tuning parameter for the penalty term, and a larger λ_0 often leads to detected subnetworks with higher proportion of more informative edges and smaller sizes whereas a smaller λ_0 often produces larger networks including more informative edges in G. Generally, the default λ_0 is set as 0.5 and the algorithm performs well when λ_0 ranges between 0.35 to 0.65 balancing precision and recall rates. For each G_c, the subgraph size $|E_c| = \prod_{k<k', \leq K_c} |V_k^c||V_{k'}^c|$ as there are no edges connecting nodes within an independent set A_k^c. When $K_c = |V_c|$, the k-partite subgraph becomes a clique. We also estimate parameters in a subgraph G_c including K_c and A_k^c to specify the k-partite topological structure. The penalty term is crucial for accurately detecting the latent topological patterns because: i) the objective function often selects a relatively large $\hat{C}$ and include many G_c as singletons to minimize the subgraph sizes, which can suppress false positive noises; ii) the more advanced topological structures (e.g. k-partite and rich-club) are preferred than the community/clique structure to further reduce the sizes of the subgraphs. For instance, in Figure 2 the penalized objective function can identify the altered connectivity network G_c and further specify the bipartite topological structure. The organized network topological structure may reveal the underlying neurophysiopathological mechanisms.

However, the direct optimization of the non-convex objective function 1 is difficult. We develop a multi-level iterative algorithm to optimize C, $\{G_c\}$, K_c, and A_k^c in the objective function for the latent k-partite structure detection. We provide detailed derivation and optimization algorithms for detecting the latent k-partite in the Supplementary Materials and sample code online.

2.3. Statistical test for k-partite structure

In section 2.1, by applying the KPGD algorithm, we detect independent sets $\{A_k^c\}$ from G_c, where the intra-set edges are less informative than inter-set edges. We propose a statistical test to verify the k-partite structure (i.e. the organized pattern of informative edges) and provide a p-value for statistical inference. With G_c denoting a differentially expressed subgraph, by default the our goal is to test whether the informative edges of G_c are distributed in a k-partite structure $H_c = \pi(G_c) = \cup_{k=1}^{K^c} A_k^c$, as opposed to a random ordering. Therefore, our null and alternative hypotheses are:

H_0: The informative edges are distributed randomly in G_c and the topological structure of G_c is a community.

H_1: The informative edges are distributed as a k-partite pattern in G_c.

We develop a 'graph edge permutation' testing strategy to determine the statistical significance of the non-randomness of the k-partite topological structure. The non-parametric permutation test for statistical inferences is appealing here, given the challenges posed for determining appropriate asymptotic distributions of the test statistic based on the complex object G_c and multiple testing issues (Nichols and Holmes, 2002; Zalesky et al, 2010; Winkler et al, 2014).

The edge permutation test shuffles the order/location of each edge in G_c. We first transform the input matrix $\mathbf{W}_c$ of G_c into a vector of edges E_c, with the length $|V_c| \times (|V_c| - 1)/2$ ($|V_c|$ is the number of nodes in G_c), so that $E_c = \left\{W^c_{1,2}, W^c_{1,3}, \cdots, W^c_{|V_c|-1,|V_c|}\right\}$. The edge permutation results from a mapping function ϕ that projects $\mathbf{W}_c$ to $\mathrm{W}^m_c = \phi(\mathrm{W_c})$ for the mth permutation. But, the mapping is not edge preserving as $\mathbf{W}_c(i, j) \neq \mathbf{W}^m(i, j)$. Then the edge permutation can map a subgraph G_c with an organized topological structure, e.g. k-partite structure, to a random subgraph $\phi(G_c)$. Here, we use a test statistic

$$T_c 0 = 1/|E_c^{off}| \sum_{k=1}^{K_c} \sum_{i \in A_k^c, j \notin A_k^c} \{-\log(p_{ij})\} - 1/|E_c{}^{diag}| \sum_{k=1}^{K_c} \sum_{i \in A_k^c, j \in A_k^c} \{-\log(p_{ij})\},$$

to contrast the differential levels of edges between the independent sets (off-diagonal blocks) and within these sets (diagonal blocks). Since the proportion of informative edges in G_c is high as a selected subgraph, if the informative edges are distributed in an organized pattern, the test statistic of $\mathbf{W}_c$ should be greater than those $\{T_c^m\}$ calculated for most of the permutations $\mathbf{W}_c^m (m = 1, \cdots, M)$. T_c^0 is the contrast between empirical estimates of information entropy (Paninski, 2003) of the detected k-partite subgraph and the rest of the clique. If there exists a latent and organized k-partite structure in G_c, T_c^0 is a large positive number. When differentially expressed edges are randomly distributed in $\mathbf{W}_c$, T_c^0 is close to zero and the probability of T_c^0 being a large number is very small. When $K_c = |V_c|$, we set $T_c^0 = 0$. We leverage the edge permutation and graph combinatorics to examine the hypotheses H_0 and H_1, that is whether differentially expressed edges are concentrated in the detected topological pattern of G_c. The detailed algorithm is described in Algorithm 1. The goal of the GEP test is model comparison, which statistically examines whether a more sophisticated topological structure (H_1) fits better than the default community structure (H_0). The GEP test functions similarly to the likelihood ratio test regarding the model comparison because both methods compare the test metrics under H_1 vs. H_0 after optimizing the objective function (the likelihood function and (1)). Therefore, the proposed testing scheme is general and applicable to test any organized graph topological structure of G_c against the default community structure, for example, rich-club and overlapped communities (van den

Heuvel and Sporns, 2011) versus the default community/clique structure. A more precise topological structure of the latent phenotype-related subnetwork can reduce false positive rates while maintaining similar true positive rates. Last, we perform the commonly used permutation strategy (e.g. Nichols and Holmes, 2002; Zalesky et al, 2010; Chen et al, 2016b) to examine whether the subgraph is differentially expressed between clinical groups. The proposed statistical testing strategy is similar to the commonly used spatial scan statistics in spatial statistics, which first identifies the 'incidence' edges (differential edges) and next examines the distribution patterns of the differential edges in the whole brain connectome. Therefore, the issues of multiple testing and selection bias can be solved by the above procedure (Waller and Gotway, 2004).

Algorithm 1

Graph Edge Permutation (GEP) Test

1: **procedure** GEP –ALGORITHM

2: Apply the KPGD algorithm to $\mathbf{W}_c$, and calculate a statistic $T_c^0 = 1/|E_c^{off}| \sum_{k=1}^{K_c} \sum_{i \in A_k^c, j \notin A_k^c} -\log(p_{ij}) - 1/|E_c^{diag}| \sum_{k=1}^{K_c} \sum_{i \in A_k^c, j \in A_k^c} -\log(p_{ij})$, where $\left|E_c^{off}\right|$ is the number of edges between the independent sets and $\left|E_c^{diag}\right|$ is the number of edges within the independent sets.

3: List all edges in G_c as vectors in the original order, $vec(\mathbf{W}_c) = \left\{W_{1,2}^c, \cdots, W_{|V_c|-1,|V_c|}^c\right\}$, where $|V_c|$ is the number of nodes and $W_{i,j}^c = -\log(p_{i,j})$.

4: **for** each permutation iteration $m = 1 : M$ **do**

5: Shuffle the order of edges in $vec(\mathbf{W}_c)$, and obtain an edge reordered graph G_c^m with a weighted edge matrix $\mathbf{W}_c^m$;

6: Apply the KPGD algorithm on G_c^m (or $\mathbf{W}_c^m$) and obtain K_c^m independent sets, where K_c^m is determined by the 'quantity and quality' criteria;

7: Calculate the test statistic T_c^m as described in line 2.

8: **end for**

9: If T_c^0 is greater than the top 5th percentile of T_c^m, we reject the null hypothesis, and thus G_c has a k-partite topological structure.

10: **end procedure**

3. Numerical Results

We conduct simulation studies i) to evaluate the performance of the proposed approach and to compare it with existing methods by analyzing group-wise functional connectivity data; ii) to examine whether the KPGD algorithm can optimally determine the number of independent sets and recognize the latent k-partite structure of G_c; iii) to evaluate the computational cost of the KPGD algorithm.

3.1. Simulations of group-wise functional connectivity data analysis

We simulate group-wise functional connectivity data, which have properties that mimic those of our experimental data (forthcoming) from a resting-state fMRI study of Parkinson's disease (PD). We consider data from 30 PD patients and 30 healthy control subjects, and for

each subject we generate (preprocessed) time series of 150 data points for 90 brain regions of interest (ROIs), which correspond to the first 90 regions from a commonly used Automated Anatomical Labeling (AAL) brain parcellation (Tzourio-Mazoyer et al, 2002). For subject s in the control group, a vector of 90 elements is sampled by $x_t^s \sim MVN(0, \Omega_0)$ and $\mathbf{X}^s = \left\{x_1^2, \cdots, x_T^s\right\}$ (T=150) and for subject from the patient group Ω_0 is replaced by Ω_1 with $\Delta = \Omega_1 - \Omega_0$ (Kim et al, 2015b). The non-zero (truly differential) elements of Δ are around 0.5 and are distributed in a latent bi-partite subgraph. Ω_0 and Ω_1 are smoothed to be semi-positive definite. We introduce white noises to the time series with signal to noise ratios (SNR) around 2.5 (the mean contrast to standard deviation). We obtain each M^s of $\mathbb{M}$ by calculating the Fisher's Z transformed correlation coefficients between the time series of two ROIs, and then perform statistical testing on each edge. The test produces p-values p_{ij}, which are stored in $\mathbf{W}$ with elements $W_{i,j} = -\log(p_{i,j})$ (Figure 2a). We use different sizes of k-partite subgraphs ranging from 10 to 30, with signal to noise ratio (SNR) of 2.5, and increase the noise levels by tuning SNR to 1.5, 1, and 0.7 for the size 20 k-partite subgraph. In addition, we increase the number of subjects per group to 60 and 100 and increase the number of regions to 250. For larger sample sizes, we use the size 20 k-partite subgraph because the performance is similar for other sizes.

Figure 2 demonstrates the FC analysis pipeline of our approach. We apply the KPGD algorithm and the edge permutation test to detect the differentially expressed subnetwork (Figure 2b) and identify the k-partite topological structure with formal statistical inferential results (Figure 2c). We compare our method with some existing network testing methods including adaptive sum of powered score (aSPU) and NBS (Zalesky et al, 2010; Pan et al, 2014; Kim et al, 2014; Kim et al, 2015b) on simulated $\mathbb{M}$. We repeat the data simulation 100 times for each setting considered and record the percent of times that aPSU, NBS, and KPGD identify the differentially expressed subnetwork. For the NBS and KPGD results, we further compare the detected network $\widehat{G_c}$ with 'true' network G_c by measuring the size difference as $|(|V_c| - |\widehat{V_c}|)|$.

When the noise level is low, all methods have similar power to detect that there exist differentially expressed edges/subnetworks between the brain connectomes of healthy controls and PD patients. However, only the KPGD method can correctly identify the latent k-partite structure (from Figure 2a to 2c). Although NBS can identify the existence of differentially expressed subnetworks, the accuracy to capture the network is relatively low in comparison to KPGD because the principle of parsimony in KPGD effectively controls falsely discovered edges (Chen et al, 2015a). The KPGD method could be a good complement to the omnibus tests of aSPU and/or aSPUw to further identify the latent and organized differentially expressed subnetworks. When the noise level is higher, KPGD is more robust to noise and false positive edges since the false positive edges tend to appear randomly in G without organized topological structure and do not impact GEP inferential procedure based on graph combinatorics. We also tune p_0 of the KPGD algorithm from 0.01 to 0.1, and across all simulation settings, the maximum network size deviation $|(|V_c| - |\widehat{V_c}|)|$ caused by p_0 is two (in only 6 simulations) and the optimal K_c is not affected. The network sizes (from 10 to 30) seem not to affect the performance of the KPGD algorithm.

In addition, we let $\Delta = \Omega_1 - \Omega_0$ follow a clique structure (i.e. $K_c = 20$) with SNR equal to 2.5, and we examine the false positive rates of the KPGD and GEP methods. The permutation tests show significant k-partite structure three times (on average 97 failures), and thus the false positive rate is 3% for k-partite structure detection. We also investigate the false positive discovery rates when $\Delta = \Omega_1 - \Omega_0 = 0$ (SNR=0 for all edges) with a size of 20, and all methods show similar performance with appropriate false positive rates. In contrast to applying community detection, the KPGD results can more accurately identify the latent differentially expressed subnetwork.

In summary, the power of KPGD method is mainly affected by the SNR and sample size which are critical to distinguish differentially expressed edges from false positive edges. When SNR and sample size are moderate (e.g. SNR=0.7 and n=60), the power is around 80% to accurately capture the latent network structure. The false positive network detection error rate is below 5% for all settings because the combinatorial probability of false positive edges to compose an organized topological structure is close to zero.

3.2. k-partite structure detection and automatic selection of k

We generate a 20×20 $\mathbf{W}_c$ matrix with each edge given by $-\log(p_{ij})$. Next, we let the number of independent components (k) of the k-partite graph equal 2, 4, and 10, and we use larger values of $-\log(p_{ij})$ for off-diagonal edges. For example, we set the edges within the diagonal blocks around $-\log(0.35) = 1.05$, and edges within the off-diagonal blocks around $-\log(0.02) = 4.0$. We set the center difference between diagonal and off-diagonal block edges as a parameter δ. We permute the order of the nodes, and the k-partite structures are not directly observable (subfigures of the left column, Figure 3). We apply the KPGD algorithm to detect the topological structure and the GEP test for statistical inferences. For all settings, our algorithm successfully identifies the correct numbers of independent sets k and reveals the k-partite graph. We evaluate the performance of our new methods by simulating 100 data sets with different values of both k and δ.

We summarize our results in Table 2, specifically reporting the average numbers of false positive (FP) and false negative (FN) edges across the 100 repetitions. We observe that in general the KPGD algorithm performs well, and the GEP testing strategy successfully recognizes k-partite structure. When k is smaller, the numbers of FP and FN edges decrease, because the pattern is closer to a complete subgraph (clique) when k is large. Moreover, our 'quantity and quality' rule successfully determines the optimal k, and we consistently estimate the true k correctly for each simulated data set (more than 99%). Also, our methods demonstrate robustness to the choice of δ. The selection of p_0 between 0.01 and 0.1 has no impact for the results above. We also test the KPGD and GEP algorithms for $\delta = 0$, and the GEP report false positive fining two of the 100 simulated data sets.

3.3. Computational costs

We evaluate the scalability and computational costs of our method. We apply our approach to larger k-partite subgraphs, for example varying k-partite subgraph sizes with 100, 200, 500, and 1000 nodes. We simulate 10 data sets for each setting and evaluate the average computational time of KPGD on a computer with i7 CPU and 32G memory without and

with parallel computing (four threads). The results show all k-partite structures can be detected and the computational times ranged roughly from a few seconds to about 24 minutes, without parallel computing. The detailed results are summarized in Table 3. For FC analysis and general k-partite subgraph identification, we consider a community subnetwork of 1000 nodes as relatively large. Therefore, our method is scalable for larger networks and the computational time seems to be affordable.

4. Data example

We apply our method to data from an fMRI Parkinson's disease study (Bowman et al, 2016). The fMRI data were acquired using a multi-slice ZSAGA sequence, yielding 30 axial slices (4mm thick) covering the entire cerebrum. Subjects lay supine with eyes open, maintaining attention to a visual fixation point on the computer screen, without other explicit tasks. The data include 42 subjects, with 24 Parkinson's patients (PD) patients and 18 healthy control subjects. The 42 subjects include 21 males and 21 females, and are at the mean age of 65.0 ±9.0 (standard deviation). The mean age is 61.9±8.7 years for the PD group and 71.4±5.8 years for the control group. Among the 24 patients 13 are females (46.4%), and 8 of 14 controls are females (57.1% of controls) with p = 0.74. The mean Unified Parkinson's Disease Rating Scale (UPDRS) Part III (motor) score for these patients is 19.4 ±10.2.

The data were preprocessed in AFNI using several steps commonly applied to fMRI neuroimaging data, including slice-timing correction, co-registration, spatial normalization, and regional parcellation using AAL. For the region level signal, we summarize the temporal signals from all voxels in each of the 90 AAL regions (Tzourio-Mazoyer et al, 2002), separately for each subject, by averaging the preprocessed time courses.

We calculate 4005 Pearson correlation coefficients between the time courses of all pairs of 90 AAL regions. We then perform statistical tests on Fisher's Z transformed correlation coefficients while adjusting the dependence between edges (Chen et al, 2018) and calculate the graph matrix **W**, with the (i,j)th entry of $W_{ij} = -\log(p_{ij})$. Age and sex of subjects are corrected in the regression analysis. The overall network detection procedure is demonstrated in Figure 4. Figure 4a displays our input data **W**, which is a 90 × 90 matrix of testing results $-\log(p_{ij})$, and the original distribution of differentially expressed edges (hot color) is also shown in the heatmap. We apply the KPGD algorithm on **W** for differential brain connectivity subnetwork detection, which seeks to capture the most significantly differentially expressed connectivity edges within minimal-sized k-partite subnetworks. Figure 4b shows the detected disease -related community subnetwork and the resulting k-partite structure is revealed by Figure 4c. Based on the GEP testing results, this subnetwork is differentially expressed and the k-partite structure ($\hat{k} = 2$) is significant (with $p < 0.001$ based on 5000 iterations). The k-partite structure includes 23 AAL regions such as orbito-frontal cortex, parietal region, basal ganglia, and limbic gyrus (see Table 4). We also apply NBS to the data with multiple tuning parameters, however, no significant network is detected.

We enlarge the detected k-partite network in Figure 4d. There are two large independent sets: set one mainly includes insular cortices, occipital lobes and frontal lobes, and set two

mainly includes central frontal lobes and temporal lobes. In set one, the altered connectivity of regions from the occipital lobe of the patients with PD which are well documented in Burton et al, 2004 and Emre et al, 2007. Similarly, the functions of insular cortex are linked with many symptoms of the Parkinson's disease (Kikuchi et al, 2001; Mattay et al, 2002; and Wu et al, 2005). The findings regarding temporal lobes in set two have also been identified in previous studies (e.g. Tam et al, 2005; Moody et al, 2004). Interestingly, many of our long-range differentially expressed edges are also reported by recent literature (e.g. Baggio et al, 2014; Lopes et al, 2016). Yet, we first identify the underlying organized graph topological structures of these findings. Overall, most differentially expressed edges in the detected k-partite subgraph based on data analysis coincide with findings of numerous precedent studies (Prodoehl et al, 2014). The 132 differentially expressed inter-set edges are summarized in Supplementary Table 1.

In this article, we identify a differentially expressed connectivity network with a latent bipartite topological structure. The network is not predefined but rather is automatically detected by our proposed methods. In the detected differential network with a bipartite topological structure, most connections within each independent set are high for most subjects, yet they have no significant difference in connectivity between normal controls and PD patients. Nevertheless, the connections between the two sets are differentially expressed between the two groups. We illustrate the k-partite structure in a 3D brain image (Figure 5). We note that the normal control group shows stronger connections for most differentially expressed edges (around 85%) in the detected k-partite subgraph (edge color code yellow in Figure 5). These results concur with the fact that Parkinson's disease is a neurodegenerative disorder. In addition, the normal control group exhibits stronger connections for most long-range differentially expressed edges such as edges from occipital lobes and inferior temporal lobes to insular and superior temporal lobes. There is only a small proportion of the differentially expressed edges in which the PD patients express hyper-connections relative to the normal controls, mainly including edges connected with the nodes of insular(R) or superior frontal gyrus orbital part. The structure apparent in our detected "k-partite phenomenon" may reflect neuropathology of Parkinson's disease.

5. Discussion and Conclusion

In this article, we have introduced algorithms to detect a specific yet pervasive network topological structure: k-partite graph topology along with new statistical testing techniques. The KPGD algorithm provides a convenient and fast solution to detect k-partite subgraphs in a large weighted graph and to identify the differentially expressed brain connectivity subnetworks more precisely. Besides the application to brain connectivity network analysis for neuroimaging data, the algorithm may also be used to uncover interesting findings in other applications (e.g. protein-protein interactions Singh et al, 2008; Liao et al, 2009). The low computational cost of our topological structure detection algorithms allows the implementation of the permutation test for statistical inference.

Functional integration has been a longstanding principle underlying complex human brain function, enabling interactions between different brain areas (Bowman, 2014). From a network perspective, the interactions are assumed to be highly organized, rather than

distributed randomly. Neurological disorders, such as PD, and psychiatric disorders may reveal patterns of dysregulation in such system level characterizations of brain function. Our work provides an illustration of how differentially expressed connectome features (between PD patients and normal controls) are also distributed in an organized topology. Our methods provide a pathway to reveal the phenotype of connectome features, along with the latent topological structure, and to conduct statistical tests about this organized topology. Detected connectivity patterns may serve as useful features, e.g. as network biomarkers, in the diagnosis and treatment of brain disorders.

Brain connectivity analyses naturally involve high dimensionality. Numerous statistical methods have been developed aiming to select differentially expressed features from high-dimensional biomedical data that are based on the test statistics or p-values of univariate statistical inferences, for example, the FWER and FDR (and *fdr*) rate controlling methods (Benjamini and Yekutieli, 2002; Storey, 2002; Fan et al, 2012; Efron et al, 2012). Yet, these methods do not address the graph topology between the massive edge features (that each edge is constrained by a pair of nodes). Often, the graph topology is highly informative for the population level analysis of ‘edge’ type features (Bullmore and Sporns, 2009), and tailored statistical methods are needed. Rather than drawing statistical inferences on individual edges, methods including NBS, Pard, and KPGD are developed to detect and test altered ‘subnetworks’ at the group level. Our approach is a good addition to these methods.

The detected graph topology structure could reveal important underlying disease mechanisms. For example, the community detection algorithm detects a general clique subgraph structure revealing a differentially expressed ‘community’ of edges between clinical subgroups. We develop more advanced statistical tools to further examine whether the detected clique subgraph includes a k-partite graph structure. The k-partite topological structure may uncover new important characteristics of brain disorders, e.g. revealing systematic disrupted connections between neural network communities (say posterior cortex to frontal and parietal cortices Baggio et al, 2014; Lopes et al, 2016).

Although we use functional connectivity brain network data for demonstration, the ideas underlying our method are applicable to all types of weighted adjacency matrix including functional connectivity (FC, e.g. from EEG and fMRI data) and structural connectivity (SC, e.g. from diffusion-weighted imaging data) if the connectivity matrix represents an undirected graph. In addition, we choose correlation coefficients as connectivity metrics for our data example based on our empirical analysis and suggestions by previous works (Zalesky et al, 2012; Kim et al, 2015a). However, when outliers exist users are free to choose other connectivity metrics, for example, maximum information coefficient or rank-based coherence and perform normalization (Chen et al, 2015b). Similarly, when parametric testing assumptions are not met, nonparametric methods could be adopted. Our study results are based on a moderate sample size. Further verification analysis will be performed with additional subjects recruited for the study. As with many high-dimensional biomedical data, the data may contain sample heterogeneity, machine noises, different options of preprocessing procedure (e.g. the global regression for resting state fMRI preprocessing). These may jointly impact the following statistical analysis. Our method is robust to many of these factors because these false positive/negative edges are less likely to appear in an

organized graph topological structure. Therefore, when graph combinatorics is incorporated in the inferential procedure, the false false positive/negative error rates may be decreased. Moreover, here we only consider direct connections for group-wise inference because comparing indirect connections between clinical groups could introduce numerous complexities due to varying indirect connectivity paths (e.g. the shortest path between two nodes) across subjects and the variation of selected connectivity metrics. Finally, our proposed methods are currently based on region level connectivity metric. One may consider extensions to our approach that model voxels and regions in a hierarchical fashion, e.g. similar to ideas presented by (Chen et al, 2016a).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Bowman's research work was funded by a grant from the NINDS (U18 NS082143) at NIH as part of the Parkinson's Disease Biomarker Program.

References

Abe Y, Kachi T, Kato T, Arahata Y, Yamada T, Washimi Y, Sobue G (2003). Occipital hypoperfusion in Parkinson's disease without dementia: correlation to impaired cortical visual processing. Journal of Neurology, Neurosurgery & Psychiatry, 74(4), 419–422.

Achard S, Delon-Martin C, Vértes PE, Renard F, Schenck M, Schneider F, … Bullmore ET (2012). Hubs of brain functional networks are radically reorganized in comatose patients. Proceedings of the National Academy of Sciences, 109(50), 20608–20613.

Baggio HC, Sala-Llonch R, Segura B, Marti MJ, Valldeoriola F, Compta Y, … & Junque C (2014). Functional brain networks and cognitive deficits in Parkinson's disease. Human brain mapping, 35(9), 4620–4634. [PubMed: 24639411]

Benjamini Y, and Yekutieli D (2001). The control of the false discovery rate in multiple testing under dependency. Annals of statistics, 1165–1188.

Biswal BB, Mennes M, Zuo XN, Gohel S, Kelly C, Smith SM, … Dogonowski AM (2010). Toward discovery science of human brain function. Proceedings of the National Academy of Sciences, 107(10), 4734–4739.

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

Bowman FD, Caffo B, Bassett SS, Kilts C (2008). A Bayesian hierarchical framework for spatial modeling of fMRI data. Neuroimage, 39(1), 146–156. [PubMed: 17936016]

Bowman FD (2014). Brain Imaging Analysis. Annual Review of Statistics and Its Application, vol 1: 61–85. DOI: 10.1146/annurev-statistics-022513-115611

Bowman D, Drake D and Huddleston D (2016). Multimodal Imaging Signatures of Parkinson's Disease. Front. Neurosci. 10:131. doi: 10.3389/fnins.2016.00131 [PubMed: 27147942]

Buckner RL, Sepulcre J, Talukdar T, Krienen FM, Liu H, Hedden T, … Johnson KA (2009). Cortical hubs revealed by intrinsic functional connectivity: mapping, assessment of stability, and relation to Alzheimer's disease. The Journal of Neuroscience, 29(6), 1860–1873. [PubMed: 19211893]

Bullmore E, Sporns O (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. Nature Reviews Neuroscience, 10(3), 186–198. [PubMed: 19190637]

Burton EJ, McKeith IG, Burn DJ, Williams ED, …, O'Brien JT (2004). Cerebral atrophy in Parkinson's disease with and without dementia: a comparison with Alzheimer's disease, dementia with Lewy bodies and controls. Brain, 127(4), 791–800. [PubMed: 14749292]

Chen S, Kang J, Xing Y, Wang G (2015). A parsimonious statistical method to detect groupwise differentially expressed functional connectivity networks. Human brain mapping , 36(12), 5196–5206. [PubMed: 26416398]

Chen S, Kang J, Wang G (2015). An empirical Bayes normalization method for connectivity metrics in resting state fMRI. Frontiers in neuroscience, 9, 316. [PubMed: 26441493]

Chen S, Bowman FD, Mayberg HS (2016b). A Bayesian hierarchical framework for modeling brain connectivity for neuroimaging data. Biometrics, 72(2) 596–605. [PubMed: 26501687]

Chen S, Xing Y, and Hong E (2016a). On testing differentially expressed functional connectivity networks. arXiv preprint arXiv:1609.00360

Chen S, Xing Y, Kang J, Kochunov P, Hong EL (2018). Bayesian Modeling of Dependence in Brain Connectivity Data. Biostatistics, in process.

Craddock RC, Jbabdi S, Yan CG, Vogelstein JT, Castellanos FX, Di Martino A, … , Milham MP (2013). Imaging human connectomes at the macroscale. Nature methods, 10(6), 524–539. [PubMed: 23722212]

Crossley NA, Mechelli A, Vertes PE, Winton-Brown TT, Patel AX, Ginestet CE, … Bullmore ET (2013). Cognitive relevance of the community structure of the human brain functional coactivation network. Proceedings of the National Academy of Sciences, 110(28), 11583–11588.

Crossley NA, Mechelli A, Scott J, Carletti F, Fox PT, McGuire P, Bullmore ET (2014). The hubs of the human connectome are generally implicated in the anatomy of brain disorders. Brain, 137(8), 2382–2395. [PubMed: 25057133]

Derado G, Bowman FD, and Kilts C (2010). Modeling the spatial and temporal dependence in fMRI data. Biometrics, 66: 949–957. [PubMed: 19912175]

Efron B (2012). Large-scale inference: empirical Bayes methods for estimation, testing, and prediction (Vol. 1). Cambridge University Press.

Emre M, Aarsland D, Brown R, Burn DJ, Duyckaerts C, Mizuno Y, …, Goldman J (2007). Clinical diagnostic criteria for dementia associated with Parkinson's disease. Movement Disorders, 22(12), 1689–1707. [PubMed: 17542011]

Fan J, Han X, & Gu W (2012). Estimating false discovery proportion under arbitrary covariance dependence. Journal of the American Statistical Association, 107(499), 1019–1035. [PubMed: 24729644]

Fornito A, Zalesky A Breakspear M (2013) Graph analysis of the human connectome: promise, progress, and pitfalls. NeuroImage 80, 426–444. [PubMed: 23643999]

Fornito A, Zalesky A, Breakspear M (2015). The connectomics of brain disorders. Nature Reviews Neuroscience, 16(3), 159–172. [PubMed: 25697159]

Göttlich M, Munte TF, Heldmann M, Kasten M, Hagenah J, Kramer UM (2013). Altered resting state brain networks in Parkinson's disease. PloS one, 8(10), e77336. [PubMed: 24204812]

Hartsperger ML, Blochl F, Stümpflen V, Theis FJ (2010). Structuring heterogeneous biological information using fuzzy clustering of k-partite graphs. BMC bioinformatics, 11(1), 522. [PubMed: 20961418]

Horn RA, Johnson CR (2012). Matrix analysis. Cambridge university press.

Gerster S, Qeli E, Ahrens CH, Buhlmann P (2010). Protein and gene model inference based on statistical modeling in k-partite graphs. Proceedings of the national academy of sciences, 107(27), 12101–12106.

Kikuchi A, Takeda A, Kimpara T, Nakagawa M, Kawashima R, Sugiura M, …, Takase S (2001). Hypoperfusion in the supplementary motor area, dorsolateral prefrontal cortex and insular cortex in Parkinson's disease. Journal of the neurological sciences, 193(1), 29–36. [PubMed: 11718747]

Kim J, Wozniak JR, Mueller BA, Shen X, & Pan W (2014). Comparison of statistical tests for group differences in brain functional networks. Neuroimage, 101, 681–694. [PubMed: 25086298]

Kim J, Wozniak JR, Mueller BA, & Pan W (2015). Testing group differences in brain functional connectivity: using correlations or partial correlations?. Brain connectivity, 5(4), 214–231. [PubMed: 25492804]

Kim J, Pan W, & Alzheimer's Disease Neuroimaging Initiative. (2015). Highly adaptive tests for group differences in brain functional connectivity. Neuroimage: Clinical, 9, 625–639. [PubMed: 26740916]

Liao CS, Lu K, Baym M, Singh R, Berger B (2009). IsoRankN: spectral methods for global alignment of multiple protein networks. Bioinformatics, 25(12), i253–i258. [PubMed: 19477996]

Long B, Wu X, Zhang ZM, Yu PS (2006). Unsupervised learning on k-partite graphs. In Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 317–326). ACM.

Lopes R, Delmaire C, Defebvre L, Moonen AJ, Duits AA, Hofman P, Leentjens AFG and Dujardin K (2016), Cognitive phenotypes in parkinson's disease differ in terms of brain-network organization and connectivity. Hum. Brain Mapp.. doi:10.1002/hbm.23474

Mattay VS, Tessitore A, Callicott JH, Bertolino A, Goldberg TE, Chase TN, …, Weinberger DR (2002). Dopaminergic modulation of cortical function in patients with Parkinson's disease. Annals of neurology, 51(2), 156–164. [PubMed: 11835371]

Monti RP, Anagnostopoulos C, Montana G (2017). Learning population and subject-specific brain connectivity networks via mixed neighborhood selection. The Annals of Applied Statistics, 11(4), 2142–2164.

Moody TD, Bookheimer SY, Vanek Z, Knowlton BJ (2004). An implicit learning task activates medial temporal lobe in patients with Parkinson's disease. Behavioral neuroscience, 118(2), 438. [PubMed: 15113271]

Narayan M, and Allen GI (2016). Mixed effects models for resampled network statistics improves statistical power to find differences in multi-subject functional connectivity. Frontiers in neuroscience, 10, 108. [PubMed: 27147940]

Ng AY, Jordan MI, Weiss Y (2002). On spectral clustering: Analysis and an algorithm. Advances in neural information processing systems, 2, 849–856.

Nichols TE, and Holmes AP (2002). Nonparametric permutation tests for functional neuroimaging: a primer with examples. Human brain mapping, 15(1), 1–25. [PubMed: 11747097]

Neubauer N, Obermayer K (2009, 12). Towards community detection in k-partite k-uniform hypergraphs. In Proceedings of the NIPS 2009 Workshop on Analyzing Networks and Learning with Graphs (pp. 1–9).

Pan W, Kim J, Zhang Y, Shen X, Wei P (2014). A powerful and adaptive association test for rare variants. Genetics, 197(4), 1081–1095. [PubMed: 24831820]

Paninski L (2003). Estimation of entropy and mutual information. Neural computation, 15(6), 1191–1253.

Prodoehl J, Burciu RG, Vaillancourt DE (2014). Resting state functional magnetic resonance imaging in Parkinson's disease. Current neurology and neuroscience reports, 14(6), 1–8.

Rubinov M, Sporns O (2010). Complex network measures of brain connectivity: uses and interpretations. Neuroimage, 52(3), 1059–1069. [PubMed: 19819337]

Seeley WW, Crawford RK, Zhou J, Miller BL Greicius MD (2009). Neurodegenerative diseases target large-scale human brain networks. Neuron 62, 42–52 [PubMed: 19376066]

Simpson SL, Moussa MN, Laurienti PJ. (2012). An exponential random graph modeling approach to creating group-based representative whole-brain connectivity networks. Neuroimage, 60(2):1117–1126. [PubMed: 22281670]

Simpson SL, Burdette JH, Laurienti PJ (2015). The brain science interface. Significance, 12(4), 34–39. [PubMed: 27441037]

Simpson SL, Laurienti PJ. (2015). A two-part mixed-effects modeling framework for analyzing whole-brain network data. Neuroimage. 113:310–319. [PubMed: 25796135]

Singh R, Xu J, Berger B (2008). Global alignment of multiple protein interaction networks with application to functional orthology detection. Proceedings of the National Academy of Sciences, 105(35), 12763–12768.

Smith SM, Nichols TE, Vidaurre D, Winkler AM, Behrens TE, Glasser MF, … Miller KL (2015). A positive-negative mode of population covariation links brain connectivity, demographics and behavior.Nature neuroscience, 18(11), 1565–1567. [PubMed: 26414616]

Sporns O (2014). Contributions and challenges for network models in cognitive neuroscience. Nature neuroscience, 17(5), 652–660. [PubMed: 24686784]

Stam CJ, De Haan W, Daffertshofer ABFJ, Jones BF, Manshanden I, Van Walsum AVC, … Berendse HW (2009). Graph theoretical analysis of magnetoencephalographic functional connectivity in Alzheimer's disease. Brain, 132(1), 213–224. [PubMed: 18952674]

Stam CJ (2014). Modern network science of neurological disorders. Nature Reviews Neuroscience, 15(10), 683–695. [PubMed: 25186238]

Storey JD (2002). A direct approach to false discovery rates. Journal of the Royal Statistical Society: Series B (Statistical Methodology),?64(3), 479–498.

Tam CWC, Burton EJ, McKeith IG, Burn DJ, O'brien JT (2005). Temporal lobe atrophy on MRI in Parkinson disease with dementia A comparison with Alzheimer disease and dementia with Lewy bodies. Neurology, 64(5), 861–865. [PubMed: 15753423]

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, and Joliot M (2002), "Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain," , 15: 273–289.

van Diessen E, Diederen SJH, Braun KPJ, Jansen FE Stam CJ (2013).Functional and structural brain networks in epilepsy: what have we learned? Epilepsia 54, 1855–1865 [PubMed: 24032627]

Van Den Heuvel MP, & Sporns O (2011). Rich-club organization of the human connectome. The Journal of neuroscience, 31(44), 15775–15786. [PubMed: 22049421]

van den Heuvel MP, Sporns O (2013). Network hubs in the human brain. Trends in cognitive sciences, 17(12), 683–696. [PubMed: 24231140]

Waller LA, & Gotway CA (2004). Applied spatial statistics for public health data (Vol. 368). John Wiley & Sons.

Winkler AM, Ridgway GR, Webster MA, Smith SM, Nichols TE (2014). Permutation inference for the general linear model. Neuroimage, 92, 381–397. [PubMed: 24530839]

Wu T, Hallett M (2005). A functional MRI study of automatic movements in patients with Parkinson's disease. Brain, 128(10), 2250–2259. [PubMed: 15958505]

Yeo BT, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M, … Fischl B (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. Journal of neurophysiology, 106(3), 1125–1165. [PubMed: 21653723]

Yu SX, Shi J (2003). Multiclass spectral clustering. In Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on (pp. 313–319).

Zalesky A, Fornito A, Bullmore ET (2010). Network-based statistic: identifying differences in brain networks. Neuroimage, 53(4), 1197–1207. [PubMed: 20600983]

Zalesky A, Fornito A, & Bullmore E (2012). On the use of correlation as a measure of network connectivity. Neuroimage, 60(4), 2096–2106. [PubMed: 22343126]

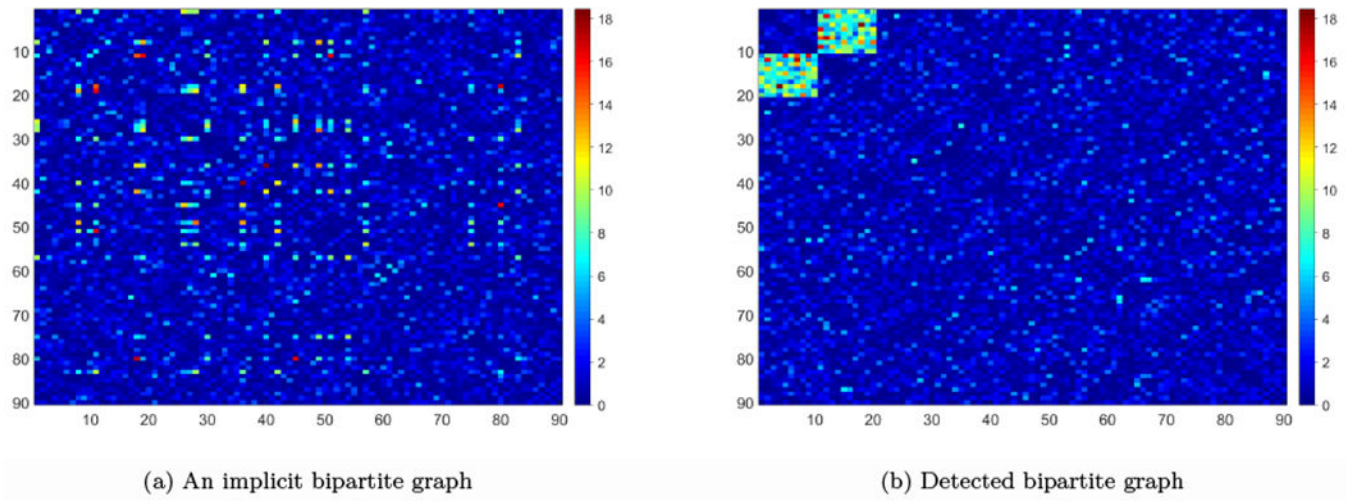


Figure 1:

An example of k-partite (k=2) graph G_c with 20 nodes in a weighted adjacency matrix (heatmap) G with 90 nodes: (a) is the k-partite graph G_c with the original order of nodes that is comparable to our input data with a latent k-partite topological structure; and (b) is output graph after applying KPGD algorithm that is an isomorphic graph of G_c with reordered nodes and an explicit topological structure.

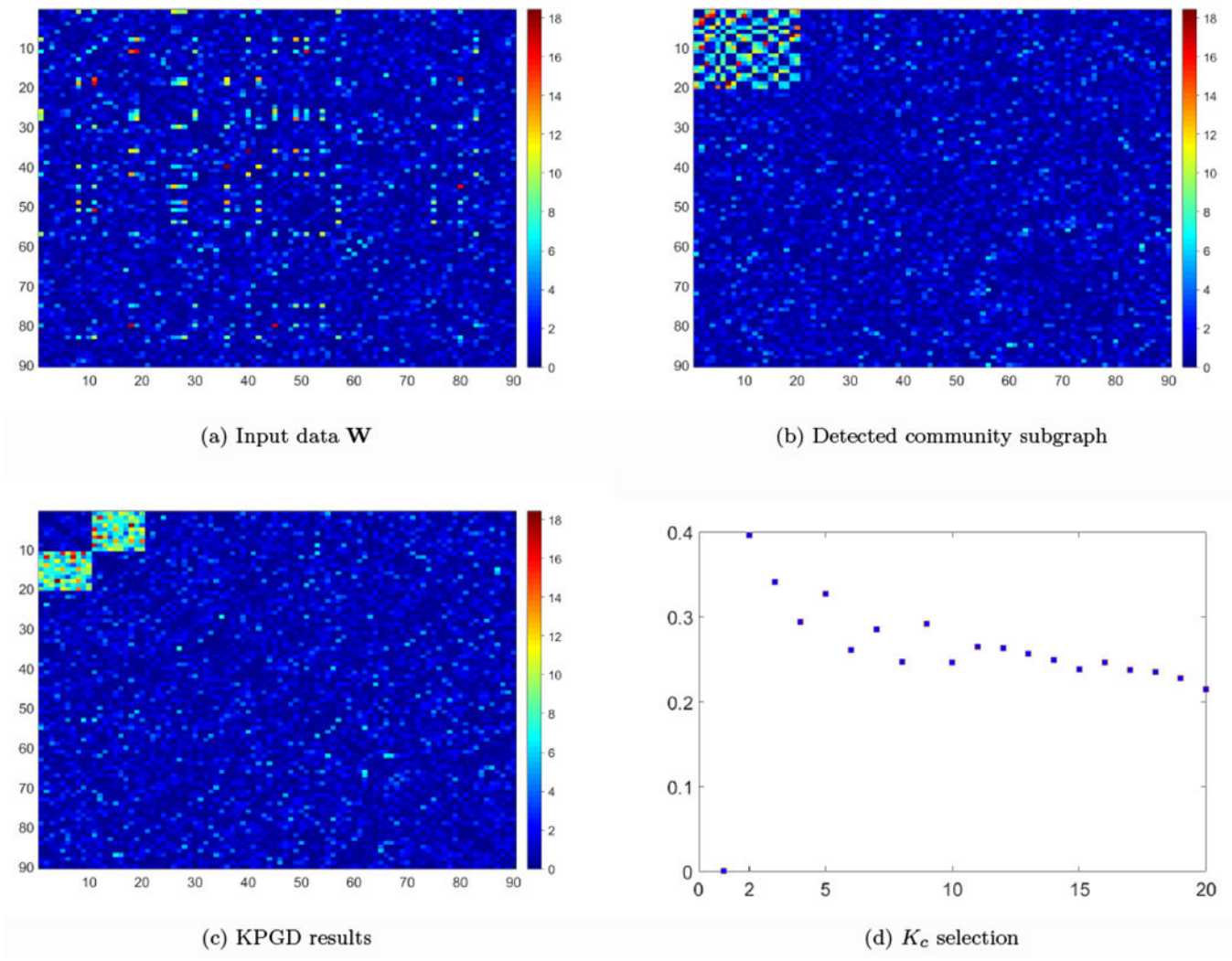
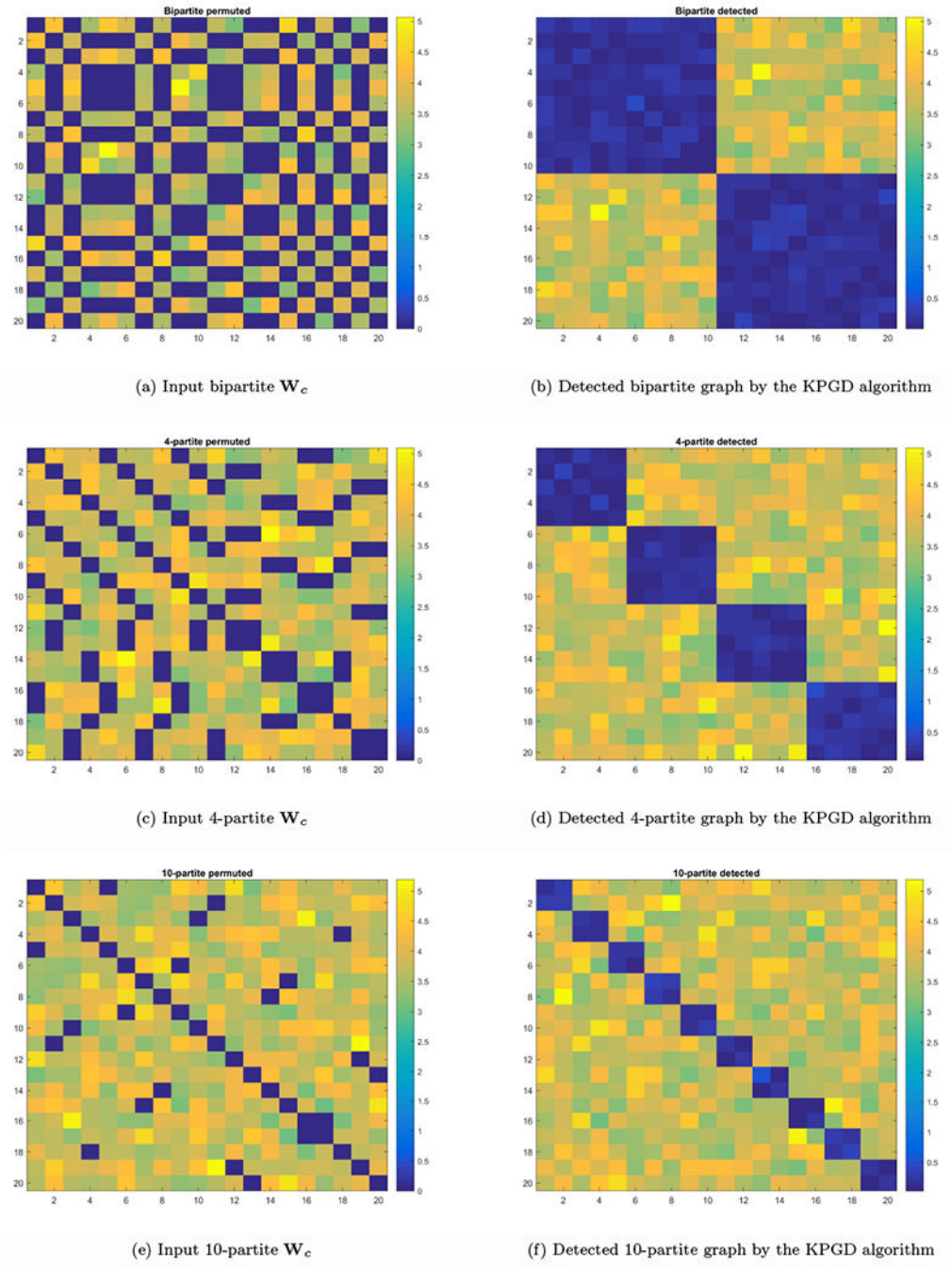


Figure 2:

Demonstration for simulation study of FC analysis: (a) **W** is the input data; (b) KPGD first identifies a community showing a differentially expressed clique subnetwork; (c) KPGD algorithm detects the K-partite structure within the detected community; (d) the scatter plot shows the K selection criteria score across different values , with K = 2 yielding the optimum.



(a) Input bipartite $\mathbf{W}_c$ (b) Detected bipartite graph by the KPGD algorithm

(c) Input 4-partite $\mathbf{W}_c$ (d) Detected 4-partite graph by the KPGD algorithm

(e) Input 10-partite $\mathbf{W}_c$ (f) Detected 10-partite graph by the KPGD algorithm

Figure 3:

Applying our approach to simulated data sets: the left side figures are input G_c with latent k-partite structure and the right side figures are the results of the KPGD algorithm with the apparent k-partite structure.

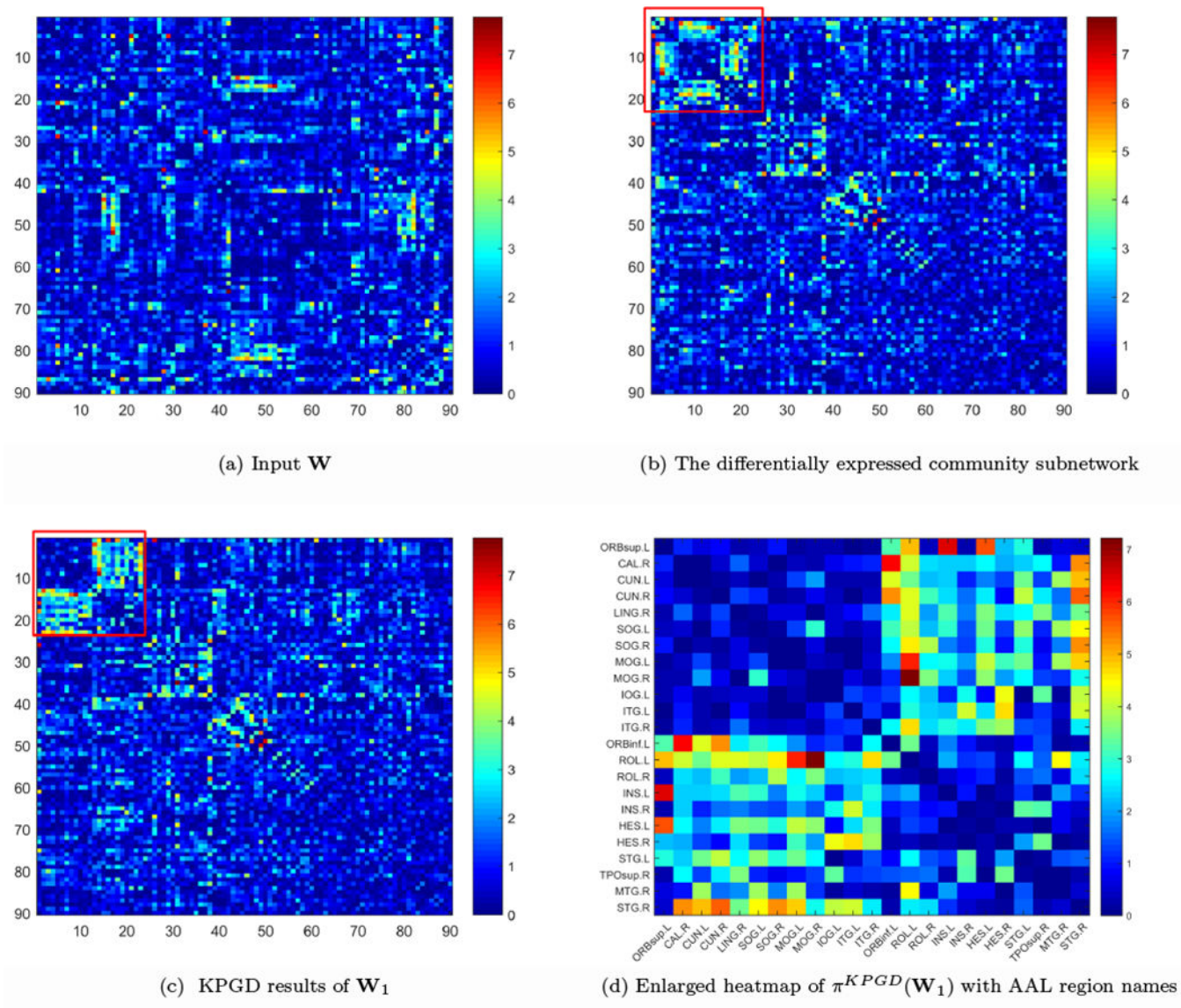


Figure 4:

We apply the KPGD algorithm to first obtain (b) from (a), and then identify a subgraph with k-partite graph topology in (c). We enlarge the heatmap of the k-partite differentially expressed subnetwork with AAL region names in (d)

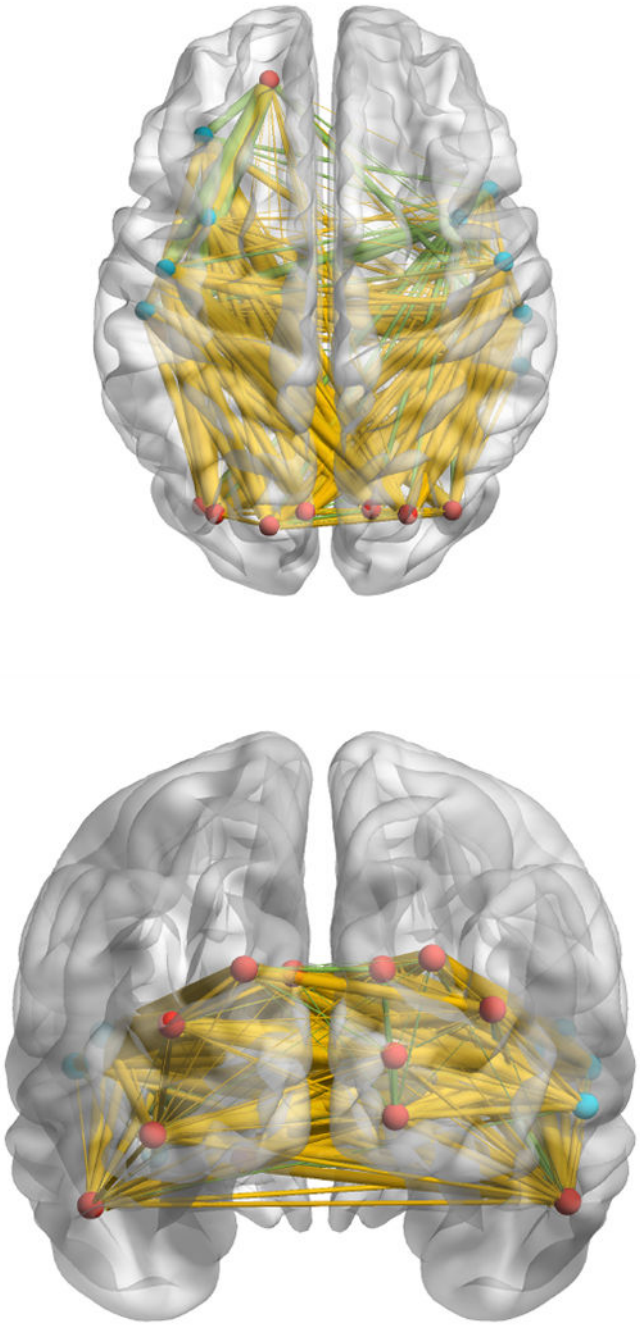


Figure 5:
3D demonstration of the differentially expressed connectivity network with k-partite graph topology: red nodes are brain regions from set one and blue nodes for set two; yellow edges indicate controls > cases and green edges indicate controls < cases, the width of an edge represent the difference between the two groups. The control group shows hyper-connections for most differentially expressed edges in the k-partite structure.

Table 1:

Simulation results: the columns of aSPU, NBS, and KPGD show the percentages to detect the differentially expressed edges/networks; the columns of NBS size-diff and KPGD size-diff report $|(|V_c| - |\widehat{V_c}|)|$ by using NBS and Pard; and the last column demonstrates the times KPGD fails to detect a k-partite subgraph within a detected clique subgraph G_c. The first five rows represent different sizes of $|V_c|$ at the SNR of 2.5; the 6-8th rows are the scenarios of lower SNRs with $|V_c| = 20$; the row $K_c = 20$ indicates the community has no k-partite structure; and the last row shows the false positive discovery rates when no edges are differentially expressed between clinical groups. We repeat this setting for results based on different sample sizes.

	aSPU	NBS	NBS size-diff	KPGD	KPGD size-diff	k-partite missing
				n=30 per group		
Size10	100%	72%	26.33±5.65	100%	0.45±1.03	0
Size15	99%	76%	35.99±4.11	100%	0.43±1.95	2
Size20	100%	89%	39.48±4.34	100%	0.12±0.41	0
Size25	97%	95%	41.12±2.89	100%	0.11±0.37	1
Size30	100%	97%	43.36±3.12	100%	0.1±0.33	0
SNR=1.5	79%	13%	28.13±3.60	92%	0.36±3.99	0
SNR=1	34%	8%	27.99±3.63	84%	0.12±2.72	0
SNR=0.7	13%	3%	31.22±4.73	62%	6.38±7.62	21
SNR=0.5	4%	6%	42.15±7.26	29%	18.96±15.18	19
G250	100%	89%	76.36±14.15	100%	0.91±2.82	0
$K_c = 20$	100%	100%	14.52±2.38	100%	0.03±0.01	97
Δ=0	7%	2%	NA	6%	NA	NA
				n=60 per group		
Size20	100%	93%	37.16±5.93	100%	0.27±0.69	0
SNR=1.5	100%	72%	29.65±4.12	100%	0.08±1.83	0
SNR=1	92%	31%	32.51±3.52	100%	0.15±1.97	0
SNR=0.7	87%	5%	41.87±4.89	79%	5.36±9.14	7
SNR=0.5	59%	8%	36.59±3.18	52%	7.92±11.53	23
G250	100%	93%	118.19±56.83	100%	0.36±0.95	0
$K_c = 20$	100%	100%	18.17±4.28	100%	0.11±0.01	99
Δ=0	3%	6%	NA	4%	NA	NA
				n=100 per group		
Size20	100%	98%	23.51±4.62	100%	0.05±0.32	0
SNR=1.5	100%	77%	31.78±5.19	100%	0.11±0.59	0
SNR=1	97%	73%	25.41±3.59	98%	0.09±0.24	0
SNR=0.7	94%	29%	39.72±6.11	100%	0.36±0.49	0
SNR=0.5	85%	12%	41.81±6.64	83%	0.29±2.55	6
G250	100%	88%	97.62±33.25	100%	1.27±0.75	0
$K_c = 20$	100%	100%	5.39±15.81	100%	0.01±0.02	95
Δ=0	6%	1%	NA	3%	NA	NA

Table 2:

K-partite simulation study to evaluate k optimally: in each cell we summarize the mean number (across 100 simulations) of falsely grouped edges and its standard deviation. Each row shows the results based on a δ that is the difference of differentially and non-differentially expressed edges and in each column. In each column, we set K equal to 2, 4, or 10. The choice of K is robust to noise and the number of independent sets.

Difference δ	$K = 2$		$K = 4$		$K = 10$	
	FP	FN	FP	FN	FP	FN
2.50	0.08±0.94	0.75±1.59	0.29±1.52	0.92±1.20	1.21±1.85	1.50±1.24
3.00	0.02±0.14	0.78±1.52	0.47±2.01	0.82±1.16	1.03±1.82	1.36±1.10
3.50	0.08±1.05	0.81±1.69	0.42±1.89	0.94±1.25	0.96±1.70	1.36±1.11
4.00	0.06±0.71	0.58±1.32	0.22±1.41	0.88±1.11	1.07±1.87	1.32±1.21
4.50	0.09±0.10	0.84±1.66	0.41±1.85	0.77±1.11	1.10±1.89	1.46±1.21
5.00	0.16±1.50	0.91±1.68	0.45±1.91	0.86±1.20	0.89±1.84	1.46±1.13

Table 3:

Computational times of the k-partite subgraph of different sizes

size	time(mean) seconds	Parallel time(mean) seconds
100	2.07	2.63
200	9.19	3.29
500	104.83	58.42
1000	1401.01	836.74

Table 4:

AAL regions in the k-partite graph

AAL region ame	abbrevation	index	x	y	z	Set
Superior frontal gyrus, orbital part, Left	ORBsup.L	5	−17	47	−13	1
Inferior frontal gyrus, orbital part, Left	ORBinf.L	15	−36	30	−12	2
Rolandic operculum, Left	ROL.L	17	−47	−8	14	2
Rolandic operculum, Right	ROL.R	18	53	−6	15	2
Insula, Left	INS.L	29	−35	7	3	2
Insula, Right	INS.R	30	39	6	2	2
Calcarine fissure and surrounding cortex, Right	CAL.R	44	16	−73	9	1
Cuneus, Left	CUN.L	45	−6	−80	27	1
Cuneus, Right	CUN.R	46	14	−79	28	1
Lingual gyrus, Right	LING.R	48	16	−67	−4	1
Superior occipital gyrus, Left	SOG.L	49	−17	−84	28	1
Superior occipital gyrus, Right	SOG.R	50	24	−81	31	1
Middle occipital gyrus, Left	MOG.L	51	−32	−81	16	1
Middle occipital gyrus, Right	MOG.R	52	38	−80	19	1
Inferior occipital gyrus, Left	IOG.L	53	−36	−78	−8	1
Heschl gyrus, Left	HES.L	79	−42	−19	10	1
Heschl gyrus, Right	HES.R	80	46	−17	10	2
Superior temporal gyrus, Left	STG.L	81	−53	−21	7	2
Superior temporal gyrus, Right	STG.R	82	58	−22	7	2
Temporal pole: superior temporal gyrus, Right	TPOsup.R	84	48	15	−17	2
Middle temporal gyrus, Right	MTG.R	86	57	−37	−1	2
Inferior temporal gyrus, Left	ITG.L	89	−50	−28	−23	1
Inferior temporal gyrus, Right	ITG.R	90	54	−31	−22	1