A parsimonious statistical method to detect group-wise differentially expressed functional connectivity networks

Shuo Chen [1*], Jian Kang [2], Yishi Xing[1], and Guoqing Wang [1]

[1] Department of Epidemiology and Biostatistics, University of Maryland, College Park, MD 20742, USA

[2] Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA

Abstract

Group level functional connectivity analyses often aim to detect the altered connectivity patterns between subgroups with different clinical or psychological experimental conditions, for example comparing cases and healthy controls. We present a new statistical method to detect differentially expressed connectivity networks with significantly improved power and lower false positive rates. The goal of our method is to capture most differentially expressed connections within networks of constrained numbers of brain regions (by the rule of parsimony). By virtue of parsimony, the false positive individual connectivity edges within a network are effectively reduced, while the informative (differentially expressed) edges are allowed to borrow strength from each other to increase the overall power of the network. We develop a test statistic for each network in light of combinatorics graph theory, and provide p-values for the networks (in the weak sense) by using permutation test with multiple-testing adjustment.We validate and compare this new approach with existing methods including false discovery rate (FDR) and network-based statistic (NBS) via simulation studies and a resting state fMRI case-control study. The results indicate that our method can identify differentially expressed connectivity networks while existing methods are limited.

Keywords: connectivity, family-wise error, fMRI, network, parsimony, statistical power

1 Introduction

Group level region-based whole brain connectivity analyses have been conducted to identify differentially expressed connectivity patterns between cohorts with different clinical or experimental conditions (Craddock *et al*., 2009; Zalesky *et al*., 2012a; Fornito *et al*., 2013; Park and Friston, 2013; Guo *et al*., 2014; Ginestet, *et al*., 2014; Shehzad, *et al*., 2014). However, the high-dimensionality of connectivity features and the complex correlation structure between them pose **difficulties** to detect the truly differentially expressed connectivity features without introducing substantial false positive findings. **Mass-univariate** statistical analyses on connections naturally require multiple testing adjustment methods such as family-wise error rate (FWER) or false discovery rate (FDR) to control false positive findings (Simpson *et al*., 2013a; Varoquaux *et al*., 2013). Different from other types of high-dimensional data (e.g. genomics), the correlation between connectivity features could be affected by an explicit topological structure comprised of brain areas (nodes). Ignoring such topological structure related correlation may lead to overly conservative multiple testing adjustment and a substantial loss of statistical power (i.e. no positive findings) (Fan *et al*., 2012). There have been few attempts to investigate the topological structure of the differentially expressed connectivity features. **Graph/network** based population level connectivity analyses seem to be a good solution to identify differences of connections with topological structures by leveraging both statistical and graph theoretical models (Simpson *et al*., 2012; Zalesky *et al*., 2012a; Guo *et al*., 2014). Graph theoretical models are often used to model brain functional connectivity networks (Rubinov and Sporns, 2010; Braun *et al*., 2009; Bullmore and Sporns, 2009; Sporns, 2011; Simpson *et al*., 2011; Sporns, 2012; Simpson *et al*., 2013b; Simpson *et al*., 2014). The nodes/vertices in the graph represent brain areas/regions and the edges express connections between the brain areas (Sporns, 2011; Rubinov

and Sporns, 2010). Since most connectivity metrics are continuous (e.g. Pearson correlation), all edges are weighted and the overall graph including all nodes is a weighted complete graph (Rubinov and Sporns, 2010; Rubinov and Sporns, 2011; Zalesky *et al*., 2010; Zalesky *et al*., 2012b). The group-wise statistical inferences are conducted based on the weighted complete graphs.

There are mainly two types of commonly used group-wise connectivity graph/network analysis methods: 1) global network metric based **methods** (GNM) which first **calculate** graph theoretical metrics such as 'small-worldness', modularity, and transitivity or cross-entropy/ mutual information for each individual and then **conduct** statistical **testing** or regression analysis on the metrics at **a** group level (Marrelec *et al*., 2008 Rubinov and Sporns, 2010; Sporns, 2011; Sporns, 2012; van den Heuvel *et al*., 2010) and similarly ; 2) differentially expressed network **methods** (DEN) such as network based statistics (NBS) and spatial pairwise clustering (SPC), which first perform mass-univariate statistical analysis for each edge at the group level and next **assembles the edges as a network using optimization algorithms** (Zalesky *et al*., 2010; Zalesky *et al*., 2012a). For example, the NBS method first applies **a** breadth-first method to detect the network and then **conducts** permutation **tests** to adjust **for** multiple **tests** in the weak sense. Both NBS and SPC methods have been successfully applied to neuroimaging studies and yielded many interesting findings (Achard *et al*., 2006; Fornito *et al*., 2012; van den Heuvel *et al*., 2008; van den Heuvel *et al*., 2009; Honey *et al*., 2008; Bassett *et al*., 2011; Bassett *et al*., 2012). In general, the GNM method does not involve multiple testing **corrections** for a single graph theoretical metric, because the edges are combined **into** a single metric. However, the GNM results often only include overall graph theoretical properties without information of localized nodes and edges (Zalesky *et al*., 2010). In contrast, the DEN method can reveal **spatially** specific

information rather than only averaged/summarized metrics, but it requires adjustment for multiple comparisons (Fornito *et al*., 2013). The NBS and SPC **methods** by Zalesky *et al*., 2010 and Zalesky *et al*., 2012a are two widely used DEN methods, which successfully incorporate family wise error control with network detection by applying permutation **testing**. Kim *et al*., 2014 conduct comprehensive comparisons of several differentially expressed network testing/detection **methods** and conclude that **the** NBS method outperforms the others given appropriate threshold values. However, these two methods may still be subject to lack of power when the testing results contain high false positive noises. The noises may cause the detected networks to include many nodes and a small proportion of supra-threshold connections such that the permutation testing results turn out to be not statistically significant.

The main contribution of this paper is to present a novel method to detect differentially expressed networks with greatly improved power and reduced false positive edges by leveraging the concept of parsimony (constraining the number of nodes). The penalized objective function (e.g. 'lasso' or elastic net methods) **has** been widely applied to high dimensional data analysis because the parsimonious selection of features may greatly improve a model's reliability and reproducibility (Hastie *et al*., 2009). The effects of parsimony (of nodes) is more dramatic when the features are edges in networks because the number of edges is power order of the number of nodes. For example, if the detected network including n nodes and $n(n-1)/2$ edges increases its size by adding one more node, then the increased network contains $n+1$ nodes and $n(n+1)/2$ edges with n more edges than the original network. Such increasing **trends** between the number of nodes and the number of edges could give rise to the power loss and high false positive **rates** for network detection, because even adding one more node (to the existing network of n nodes) by mistake could increase $n\alpha$ false positive edges and introduce other noises. Thus, we propose a

parsimonious differential brain connectivity network **d**etection method (Pard) that includes **most** significantly differentially expressed connectivity edges within networks with constrained number of nodes. We construct **an** objective function that maximizes the combined significance levels of the edges in the target networks when using the number of nodes of each network as a penalty term. The objective function can be effectively and efficiently solved by spectral graph theory models (Von Luxburg, 2007). In addition, the detected networks reveal the topological structure of the differentially expressed edges. We apply permutation **tests** to control the family-error rate in the weak sense, which is a similar strategy **to** the NBS method (Zalesky *et al*., 2010). The detailed statistical model is introduced in section 2, and followed by model evaluation and comparison using a simulation study and **an** example of **analyzing** resting state fMRI **data**.

2 Methods

In many **group level** studies, we seek to answer the question whether two groups exhibit differential connectivity patterns. The general statistical test can be described as: the null hypothesis H_0 that the two groups have no difference in connectivity vs. H_a **that** there are differentially expressed connectivity networks between the two groups. To conduct the statistical test, we first define the networks by Pard and then evaluate the probability of the networks **assuming** the null hypothesis is true (by using permutation tests). Clearly, the power and type I error rate are greatly impacted by the network detection method and thus that is our main focus **in** the method section.

Model background

The connectivity network in neuroimaging studies **is** often represented by a graph with a set of nodes and edges $G=\{V,E\}$, the set of nodes V denote a set of distinct brain areas and the edges E

are the connections between those nodes. To investigate the differential connectivity expressions between two groups of subjects (e.g. controls vs. cases), two sample tests are often conducted for all edges. For example, p_{ij} is the test *p*-value between a pair of nodes *i* and *j*. Based on all these testing results, we obtain a |*V*| by |*V*| testing significance weight matrix $W_0=\{w_{ij}\}$ with $w_{ij}=-\log(p_{ij})$. We utilize the "*–log*" transformation of *p* to express that the edges of small *p*-values may contain important information and hence are highly weighted. In addition, we note that the empirical distribution of $-\log(p_{ij})$ often follows a Gamma distribution (with both parameters equal to 1 based on maximum likelihood estimation).

Objective function of network detection

The primary goal of differential connectivity detection is to identify the significantly differentially expressed edges with well controlled false positive discovery rates. In contrast to other high-throughput genomic or proteomic expression features, the brain connectivity features (edges) are spatially constrained by nodes and thus are not independent. The NBS and SPC methods have wisely used this property to select features with more power (Zalesky *et al*., 2010 and Zalesky *et al*., 2012a). We also leverage this property to construct our objective function and furthermore add **a** penalty term of **the** number of nodes. The main objective function is to search the *C*-component clustering of the whole graph, denoted, $\{A_c\}_{c=1}^{C}$ where $\cap_{c=1}^{C}A_c=V$ and $A_c \cap A_{c'}=\emptyset$, that allocates most significant edges within networks of small number of nodes:

$$\operatorname{argmin}_{\{A_c\}_{c=1}^{C}} \sum_{c=1}^{C} \frac{\sum_{i\in A_c, j\notin A_c} -\log(p_{ij})}{|A_c|} \quad (1)$$

where $|A_c|$ represents the size (number of nodes) of the detected network cluster A_c. The objective function minimizes the weights of edges between the selected networks and the rest of G, which ensures the edges of heavy weights (more significant) are included in some networks rather than left between networks. Only the edges in the detected networks are included as biomarkers, the optimization process can be intuitively considered to cover more informative/supra-threshold edges by using small-sized networks.

The first step of the optimization is the screening step which thresholds the noisy edges of larger p-values, for example to let $w_{ij}=0$ if $p_{ij} > p_0$ and we refer **to** the thresholded weight matrix as W. Then, the Laplacian matrix based on the thresholded W matrix is

$$L=D-W, \tag{2}$$

where the degree matrix D is defined as the diagonal matrix with diagonal element $d_1,\cdots,d_n$ and $d_i = \sum_{j=1}^{|V|} w_{ij}$.

Next, we investigate how many disconnected components/subgraphs in the overall graph G with the thresholded W matrix. We denote G_q as a disconnected subgraph/subset of G ($q=1,\cdots,Q$ and $G=\cup_{q=1}^{Q} G_q$) such that $G_q \subset G$ and **there is** no edge with weight >0 connecting between G_q and its complement subset $G \setminus G_q$. To identify the disconnected subgraphs, we conduct the eigen-decomposition on the Laplacian matrix L, and the number of zero-valued eigenvalues equals the number of disconnected subgraphs (Von Luxburg, 2007). The corresponding eigenvectors of zero-valued eigenvalues exhibit the allocation of nodes to the disconnected

subgraphs. This step is equivalent to the network detection step in NBS, but **a** spectral graph model is used rather **than the** breadth first search by NBS. The objective function in formula (1) is at minimum and equals zero, if C the total number of network clusters in formula (1) is the same as Q (Q>1) the number of disconnected subgraphs in G. However, rather than stopping at this step and performing family-wise error control, we further conduct parsimonious network detection within each unconnected component to identify smaller networks **with a** higher proportion of significant edges. Thus, the overall objective function becomes parsimonious network detection within each unconnected subgraph :

$$\operatorname{argmin}_{\{\tilde{A}_k\}_{k=1}^{K_q}} \sum_{k=1}^{K_q} \frac{\sum_{i\in \tilde{A}_k, j\notin \tilde{A}_k} -\log(p_{ij})}{|\tilde{A}_k|} \qquad (3)$$

where K_q is the number of clusters in a disconnected subgraph G_q and $C = \sum_{q=1}^{Q} K_q$ which links between formulae (2) and (3). However, the direct optimization of formula (3) is a NP problem. We seek the solution by using spectral graph models. After discretization relaxation, it turns into the RatioCut spectral clustering problem which has been well developed by Hagen and Kahng, 1992. The details of the implementation of the RatioCut algorithm are illustrated in the following detailed algorithm. Then, the only tuning parameter for each disconnected subgraph is K_q, which **chooses the** number of clusters for K-means clustering. Rather than applying the conventional methods such as silhouette criteria, we develop a novel and connectivity network specific criteria to choose K_q objectively by maximizing the of product of 1) the ratio of the total number of significant edges in all K_q clusters to the total number of non-zero edges in the

disconnected subgraph (quantity) and 2) the ratio of the total number of significant edges to the number of edges within all K_q clusters (quality):

$$\frac{\sum_{k=1}^{K_q}\sum_{i\in\tilde{A}_k, j\in\tilde{A}_k} I(W_{ij}>0)}{\sum_{i<j} I(W_{ij}>0)} \cdot \frac{\sum_{k=1}^{K_q}\sum_{i\in\tilde{A}_k, j\in\tilde{A}_k} I(W_{ij}>0)}{\sum_{k=1}^{K_q}\sum_{i\in\tilde{A}_k, j\in\tilde{A}_k} 1}. \quad (4)$$

The criteria in formula (4) provides a data-driven and objective pathway to select tuning parameters K_q that tries to maximize the proportion of significant edges in the selected networks and to include most significant edges of *W* in the detected differential networks.

Moreover, we provide an approach to automatically select p_0 by **a** grid search algorithm. We search p_0 in the range of (0.05, 0.1) by **increments of** 0.005 and select p_0 that maximizes the criteria below:

$$\frac{\sum_{k=1}^{K_q}\sum_{i\in\tilde{A}_k, j\in\tilde{A}_k} -\log(p_{ij})}{\sum_{k=1}^{K_q}\sum_{i\in\tilde{A}_k, j\in\tilde{A}_k} 1} \div \frac{\sum_{k=1}^{K_q}\sum_{i\in\tilde{A}_k, j\notin\tilde{A}_k} -\log(p_{ij})}{\sum_{k=1}^{K_q}\sum_{i\in\tilde{A}_k, j\notin\tilde{A}_k} 1}, \quad (5)$$

which is the ratio of the average intensity of $-\log(p_{ij})$ (information intensity) within selected networks and the average intensity of $-\log(p_{ij})$ outside of selected networks. Note that K_q is selected by formula (4). Overvall, formulae (4) and (5) ensure that most of **the** information differentiating the two groups of subjects **is** contained in the selected networks while minimizing

the sizes of the networks **needed** (for higher concentration). Rather than applying **a** penalty term to control the network sizes, we implement the rule of parsimony by optimizing tuning parameters for objective functions. Thus, our approach is not only computationally convenient but also less ad-hoc (to provide more reproducible results).

Last, we apply **a** permutation test to provide **the** p-value of selected networks while controlling family error **rates**, which is similar to the family error control in NBS (Zalesky *et al*., 2010).

We summarize the overall parsimonious differential brain connectivity network detection (Pard) algorithm as **follows**:

1. Conduct statistical tests on all edges E and calculate the weight matrix W by screening (e.g. thresholding values p_0):

$$w_{ij} = \begin{cases} -\log(p_{ij}) & \text{if } p_{ij} \le p_0; \\ 0 & \text{else.} \end{cases}$$

2. Detect disconnected subgraphs in G: first eigen decompose the Laplacian matrix $L=D-W$ and **the** number of zero eigenvalues of L equals the number of disconnected subgraphs, and the allocation of nodes to disconnected subgraphs is **based** on the eigenvectors with zero eigenvalues.

3. Within each disconnected subgraph G_q, search networks that include most informative/significant edges with constrained **numbers** of nodes for each network. Although the direct optimization of this step is NP, it can be solved by the RatioCut algorithm after discretization relaxation:

(a) Compute the first K_q eigenvectors $[u_1,\cdots,u_{K_q}]$ of *L*, with eigenvalues ranked from the smallest.

(b) Let $U=[u_1^T,\cdots,u_{K_q}^T]$ be a $|V|\times K_q$ matrix containing all K_q eigenvectors.

(c) Perform K-means clustering algorithm on U with $K=K_q$ to cluster |V| nodes into K_q networks: $\tilde{A}_{K_q}^1,...,\tilde{A}_{K_q}^{K_q}$.

4. Try all possible K_q for each disconnected subgraph and select the optimum number of networks by formula (4).

5. Select p_0 by using the criteria of formula (5).

6. Perform permutation **testing** to control family-wise error **rates** for each detected network A_c:

(a) Shuffle the group labels for each subject *T* times (e.g. *T*=5000) and calculate *W* with the same threshold p_0 at each shuffling *t*.

(b) Obtain the most significant test statistic (e.g. Fisher's combination test) of A_c as m_t in each permutation and let m_0 represent the test statistic with $w_{ij}>0$ for original labeling.

(c) Calculate the permutation *p*-value as how many m_t are larger than m_0 divided by *T*, $P_{A_c}=\frac{\#(m_t>m_0)}{T}$ and determine whether the network is significant **at a** predetermined α level.

7. Output the significant networks with permutation test *p* values.

3 Simulations

In this section, we simulate a case-control connectivity study including 30 subjects of healthy controls and 30 subjects with neural disorders to evaluate the performance of our *Pard* algorithm. We generate a overall graph G_s of 90 ROIs as nodes and 4005 edges for a subject *s* (*s*=1,⋯60) to represent the widely used first 90 Automated Anatomical Labeling (AAL) regions in functional connectivity analysis (Tzourio-Mazoyer *et al*., 2002; Zalesky *et al*., 2010; Zalesky *et al*., 2012a).

We assume that the normalized connectivity metrics (e.g. correlations after Fisher's transformation and z-score normalization) follow a standard normal distribution. Within the overall graph *G*, we generate a truly differentially expressed connectivity network G_D of size 10 (10 nodes and 45 edges). Thus, we simulate the connectivity metrics for all subjects by

$$Z_{ij}^{s} \sim \begin{cases} N(0,\sigma^2)+d & \text{if } i,j \in G_D \text{ and subject } s \text{ is from the healthy control group;} \\ N(0,\sigma^2) & \text{otherwise.} \end{cases}$$

Then we conduct two sample t **tests** to obtain *p*-values and weight matrix W_0. The simulated data is summarized in Figure 1: Figure 1a illustrates the truth: the truly differentially expressed network by the red color; Figure 1b is the heatmap of $-\log(p_{ij})$ based on the *p*-values of two sample t tests between the two cohorts of the simulated data; and Figure 1c is the shuffled version of Figure 1b (i.e. the labels of all nodes are permuted) which may better reflect the real spatial distribution of significant edges in practice (Figure 1c is the input data for the differential

network detection algorithm). We repeat the procedures above to obtain 100 simulated data sets by using each set of parameters.

Figure1

(a)

(b) (c)

We then perform our *Pard* algorithm on W_0 to identify differentially expressed networks. After thresholding (p_0), there are no disconnected subgraphs (i.e. *G* is connected) due to false positive significant edges. Thus, the only network size tuning parameter is *C*, the number of clusters for the overall graph *G*. The optimum number *C* ranges from 49 to 62 and most p_0 are between 0.08 and 0.10 across the 100 simulated sets. The individual edges or networks with a few nodes are rarely detected as significant based on the permutation test results. Figure 2 shows how the tuning parameter selection criteria function **changes** with **an** increasing number of clusters for one simulated data set, and the score is highest at *C*=59 (for a simualtion data set). Then, we perform our algorithm with *C*=59, and the differentially expressed network G_D is successfully detected. We then conduct permutation **testing** based on 10,000 times permutation and P_{G_D} <0.001, which indicates that the detected network is significant after controlling **for the** familywise error rate. The final results are demonstrated by Figure 2b, which reveals the true differentially expressed network accurately.

Figure2

(a) (b)

For comparison, we apply the NBS algorithm for differential network detection by default parameter (t statistic =3.1) and several other threshold values for example: 2, 2.7 and 4. We also apply **it** to detect the differentially expressed edges by using FDR as a reference for false positive rates and negative rates without considering networks as output biomarkers, and we use q=0.1 as **a** cut-off. In comparisons, we consider many senarios by using different sets of parameters including network sizes (5, 10, 15) and different significance levels of trully differentially expressed edges. The significance levels of trully differentially expressed edges are generally determined by three factors: effect sizes (d), noise levels (σ^2), and sample sizes, and we only tune noise levels (σ^2) because it is redundant to tune all three factors (the same p-values). We let d=0.8, and sample sizes for cases and controls are both 30. Table 1 summarizes the means and standard errors of false postive (FP) and false negative (FN) findings under different settings.

Table 1. Simulation results under different settings

The true differentially expressed network is detected and tested as significant in 100 of the 100 data sets by using the Pard algorithm for different network sizes and most noise levels, though there is a small chance that false postive nodes (the number of nodes ranges 2~4) could be included. As a contrast, the FDR method misses most of the true postives while effectively controlling the false postive rates. The power increase of the Pard method relies on both the combined significance levels of all edges in the network and the size of **the** detected network, because a network with more significant edges and smaller number of nodes **is** more likely to be significant based on permutation **testing**. In addition, the NBS method could not detect the

differentially expressed **networks** in most settings, and we apply different thresholds (ranging from 2 to 4) and report the results (of the threshold value) with the best performance. One possible reason could be the false positive edges connecting a large number of nodes and thus a large network is detected by breadth first search, but within the detected large network there **is** only a small proportion of edges **that** are significant and the number of significant edges is similar to those of the networks from permutations. Therefore, by applying the rule of parsimony (constraining the number of nodes of the detected networks) our method increases the power substantially and excludes false positive edges effectively. In summary, the simulation study indicates that our proposed *Pard* algorithm is effective for differentially expressed connectivity network detection and less affected by noises (false positive edges).

4 Data example

This data set was collected at the Yale child study center in Yale school of medicine, one of the data collecting sites in the Autism Brain Imaging Data Exchange (ABIDE) (Di Martino *et al*., 2014). The imaging was performed on Siemens magneto Trio scanners. The imaging data was obtained using a gradient echo T2*-weighted echo planar imaging sequence, echo time TE = 25ms, repetition time TR = 2000ms, 64 × 64 matrix with 34 slices 4.0 mm tick, skip 0 mm, resulting in whole brain coverage with a voxel size of 3.4mm × 3.4mm ×4.0 mm. The **publically** available data set includes 28 participants (typical controls, TC) and 28 patients with Autism spectrum disorders (ASD), and the two groups exhibit no signifcantly different demographics (e.g. age and gender). During the MRI scanning, all subjects were asked to lie as still as possible, keep their eyes open, try not to fall asleep, and think about whatever they wanted. A black background with a gray central fixation cross was presented during the resting state scan,

although participants were not asked to fixate, **it was** verified that they had not fallen asleep at the end of the scan.

We perform rs-fMRI data prepocessing based on the Configurable Pipeline for the Analysis of Connectomes (C-PAC, http://fcp-indi.github.io). The resting-state fMRI data **was** first slice time and motion **corrected**. The data **was next** registered to a standard MNI space with voxel size $2mm^3$ and normalized to be percent signal change. The masks of the white matter (WM), the gray matter (GM) and the cerebrospinal fluid (CSF) **were created** in the standard MNI space. The mean time series from the WM and the CSF **was** calculated. **The mean time series of the WM, CSF and the six movement parameters were regressed from the GM.** A linear trend **was** removed from all the signal. The fMRI time series were filtered using a bandpass with passing band (0.009-0.08 Hz) and spatially smoothed with **a** 6mm FWHM Gaussian kernel. We then use the first 90 AAL ROIs as nodes, and take the weighted average **of** all voxels' temporal profiles within each ROI as **the** region level signal for all subjects. The Pearson correlation coefficients **were** calculated between the 90 nodes. In this analysis, we focus on the differential connectivity network detection between TC and ASD.

We first **conducted** two sample t tests to obtain *p*-values and $-\log(p_{ij})$ for all edges between TC and TSD (Figure 3a), and then **calculated** the weight matrix W_0. Next, we **applied** the parsimonious differential connectivity network detection method to the W_0 matrix. We **excluded** singleton nodes in *G*, which have all edges $p_{ij} > p_0$ connected to the rest of **the** nodes. Then, there was no disconnected subgraph in *G*. We **implemented** the optimization algorithm for network detection and **selected** the tuning parameter based on the criteria function. Based on the equation (5) in section 2, we **selected** p_0 as 0.1. Figure 3b shows the relationship between the

tuning parameter selection criteria function and the number of clusters, and the maximum value is reached at C=31. Therefore, the final **results were** detected by using the tuning parameter C=31. The results are summarized in Figure 3c, and we note that all significant edges tend to be along the diagonal because of the shrinkage effect. Two networks are detected and tested as significant by permutation tests: the first network includes 15 nodes (P<0.001) and the second network includes 10 nodes (P<0.001).

Figure3

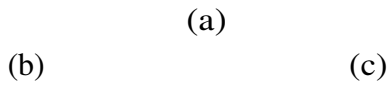

Figure 4 and Figure 5 show the differentially expressed edges within **the two detected** networks (figures are generated by using BrainNet Viewer by Xia *et al*., 2013). Many differentially expressed edges have been found in previous studies (Cherkassky *et al*., 2006, Tyszka *et al*., 2013, Di Martino *et al*., 2014). The first cluster mainly **exhibits** altered connectivity expressions between pre-frontal cortex, parietal cortex, middle inferior temporal cortex, and basal ganglia. The second cluster mainly **shows** differences between superior frontal cortex, limbic system, and occipital cortex. The details are included in Supplementary Tables 1 and 2 in the Appendix.

For comparison, we applied both **the** NBS method (with several threshold values from 2 to 4) for differentially expressed network detection and the false positive discovery rate (FDR) control for individual differentially expressed edge detection. Neither of these two methods detect significant results, which may be caused by the **noise** of false positives (NBS) and ignorance of correlation between edges (FDR) (similar atlas-based results by Tyszka *et al*., 2013).

Fig4

Figure 4: Cluster 1: 3D plots of the differentially expressed edges. The width of the edges reflects the significance level, and the color is coded as: red (TC<TSD) and blue (TC>TSD).

Fig5

Figure 5: Cluster 2: 3D plots of the differentially expressed edges. The width of the edges reflects the significance level, and the color is coded as: red (TC<TSD) and blue (TC>TSD).

5 Discussion

Group-wise whole brain connectivity analyses using atlas regions have been facing trade-offs between false positive findings and lack of statistical power (false negatives). Traditional multiple testing adjustment methods often could not detect truly differentially expressed features when trying to avoid false positive findings. Some studies conduct group-wise connectivity analyses within predefined regions rather than the whole brain in order to lower **the stringent level required for** multiple testing adjustment and to increase the likelihood of detecting statistically significant findings. Clearly, such **procedures** may lead to limited and inaccurate **results**. The DEN type network based methods such as NBS provide a pathway to improve the statistical power while controlling the FWER. DEN methods select a significant edge not only by the criteria of the test p-value but also the distribution of p-values of its neighborhood edges. Therefore, the DEN methods naturally incorporate the topological structure of the edges for

significant connectivity detection and improve statistical power. Note that our method only controls the FWE in the weak sense and thus we can only make inferences on networks rather than individual edges.

The statistical power of the existing DEN methods (e.g. NBS) depends on the proportion of supra-threshold edges within the connected subgraphs, and a smaller proportion may lead to a insignificant permutation test result. From the aspect of graph combinatrics, the probability of all edges with small p-values clustering in a small network is extremely low and therefore the permutation test p-value is very small, and the organized structure **of such p values** yields important topological information of differentially expressed connectivity networks. Hence, if the detected network includes truly differentially expressed edges but the proportion of significant edges is low, the statistical power to detect these trully signficant edges is very low because the detected network is very likely to be tested as non-singificant using a permutation test. Therefore, the objective function of our Pard algorithm aims: i) to include most significant edges in the detected networks; ii) by constraining the number of nodes of the networks to increase the proportion of significant edges within the detected networks. The constraint of the network size in the objective function is critical to reduce the (false positive) noise to improve statistical power. Therefore, our Pard algorithm improves the statistical power of network detection by allowing edges to borrow power between each other; and meanwhile effectively controls the false positive findings because false positive edges are more likely to be randomly distributed rather than concentrated within a small network. The detected networks in turn reveal the topological structures of the significant edges, and the parsimonious networks are more informative because the shrinkage procedure removes substantial noises.

We implement the optimization step by using the RatioCut algorithm. Although most spectral clustering algorithms primarily aim to allocate similar nodes to the same cluster, our objective function is to capture most significant edges within constrained networks. Fortunately, the algorithms have been well developed to implement the optimization of our objective function without intensive computational load. However, for most spectral clustering **algorithms**, the selection of number of clusters can be an arbitrary and ad-hoc procedure (Von Luxburg, 2007). We **developed** a new tuning parameter selection criteria function specifically for brain connectivity analysis to choose the number of clusters objectively. In addition, we provide a similar procedure to choose p_0. We express the importance (weight) of a edge by using –log transformation of **the** test p-value rather than raw p-value or t statistic, because the scale is more appropriate to differentiate the small p-values (e.g. 0.001 and 0.0001) and is naturally linked to Fisher's combined probability test (that has been used in cluster activity intensity analysis by Hayasaka and Nichols, 2004). From the **computational** statistics point of view, we **developed** a novel procedure to fuse network size shrinkage and ad-hoc tuning parameter selection, which avoids **use of penalty terms** (e.g. *lasso* and eleastic nets methods) and reduces **computational** cost. Further asymptotic properties of such procedure will be studied.

In the simulation study, the truly differentially expressed network can only be accurately detected and tested as significant by using our Pard algorithm. The ABIDE data provides another example of increased statistical power of our method **where** the differential networks can only be detected by **our** Pard algorithm. The edges can borrow power **from** each other within the network, and the high proportion of small p-value edges lead to significant permutation results. The detected networks exhibit many significantly differentially expressed edges that have been found in previous studies. As we focus on methods and models in this article, due to the space limit we do

not intend to discuss the results in more detail from the neurophysiological aspect. **We** provide the list the edges with p-values less than 0.05 for the two clusters in tables 1 and 2 in the Appendix. We plan to further verify the results by applying our methods to several data sets including resting fMRI data sets from other sites of the ABIDE project.

In summary, we have presented a novel parsimonious differential brain connectivity network detection method to discover differentially expressed connectivity features at the group level for fMRI data. The simulation study and data example have shown that the statistical inferences based on our Pard method are more powerful and reliable (lower false positive discovery rate). We are also optimistic that the Pard method **is** ready to be applied to connectivity analyses for task-induced fMRI data and structural connectivity network analyses.

Acknowledgement

The authors would thank the ABIDE project for sharing the resting state fMRI and clinical data at http://fcon_1000.projects.nitrc.org/indi/abide/. Chen's research is supported in part by UMD Graduate School's faculty Research and Scholarship Awards. Kang's research was partially supported by NIH grant 1R01MH105561. The authors also want thank Dr. Luiz Pessoa from University of Maryland, College Park and Dr. F. DuBois Bowman from Columbia University for constructive discussions.

Appendix

Supplementary Table 1: The supra-threshold edges in cluster 1
Table 1

Supplementary Table 2: The supra-threshold edges in cluster 2

Table 2

*Note: Since we only make inferences in the weak sense the p-values of the edges are not for inferences, nevertheless we use them to exhibit how individual edges are differentially expressed.

References

Achard, S., Salvador, R., Whitcher, B., Suckling, J., & Bullmore, E. D. (2006). A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *J. Neurosci,* 26(1), 63-72.

Bassett, D. S., Wymbs, N. F., Porter, M. A., Mucha, P. J., Carlson, J. M., & Grafton, S. T. (2011). Dynamic reconfiguration of human brain networks during learning. *Proc. Natl. Acad. Sci.* U. S. A, 108(18), 7641-7646.

Bassett, D. S., Nelson, B. G., Mueller, B. A., Camchong, J., & Lim, K. O. (2012). Altered resting state complexity in schizophrenia. *NeuroImage*, 59(3), 2196-2207.

Bennett, C. M., Baird, A. A., Miller, M. B., & Wolford, G. L. (2011). Neural correlates of interspecies perspective taking in the post-mortem atlantic salmon: an argument for proper multiple comparisons correction. *Journal of Serendipitous and Unexpected Results*, 1, 1-5.

Braun, U., Plichta, M. M., Esslinger, C., Sauer, C., Haddad, L., Grimm, O., ... & Meyer-Lindenberg, A. (2012). Test–retest reliability of resting-state connectivity network characteristics using fMRI and graph theoretical measures. *NeuroImage*, 59(2), 1404-1412.

Bullard, J. H., Purdom, E., Hansen, K. D., Dudoit, S. (2010). Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics*, 11(1), 94.

Bullmore, E., & Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3), 186-198. .

Cherkassky, V. L., Kana, R. K., Keller, T. A., Just, M. A. (2006). Functional connectivity in a baseline resting-state network in autism. *Neuroreport*, 17(16), 1687-1690.

Craddock, R. C., Holtzheimer, P. E., Hu, X. P., & Mayberg, H. S. (2009). Disease state prediction from resting state functional connectivity. *Magn. Reson*. Med. 62(6), 1619-1628.

Di Martino, A., Yan, C. G., Li, Q., Denio, E., Castellanos, F. X., Alaerts, K., ... Milham, M. P. (2014). The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular Psychiatry*, 19(6), 659-667.

Efron, B. (2012). Large-scale simultaneous hypothesis testing. *Journal of the American Statistical Association*.

Fornito, A., Zalesky, A., Pantelis, C., Bullmore, E. T. (2012). Schizophrenia, neuroimaging and connectomics. *NeuroImage*, 62(4), 2296-2314.

Fornito, A., Zalesky, A., Breakspear, M. (2013). Graph analysis of the human connectome: promise, progress, and pitfalls. *NeuroImage*, 80, 426-444.

Fox, M. D., Zhang, D., Snyder, A. Z., Raichle, M. E. (2009). The global signal and observed anticorrelated resting state brain networks. *Journal of Neurophysiology*, 101(6), 3270-3283.

Fox, M. D., Buckner, R. L., White, M. P., Greicius, M. D., & Pascual-Leone, A. (2012). Efficacy of transcranial magnetic stimulation targets for depression is related to intrinsic functional connectivity with the subgenual cingulate. *Biological Psychiatry*, 72(7), 595-603. .

Ginestet, C. E., Fournel, A. P., & Simmons, A. (2014). Statistical network analysis for functional MRI: Summary networks and group comparisons. *Frontiers in Computational Neuroscience*, 8(MAY)

Guo, S., Kendrick, K. M., Yu, R., Wang, H. L. S., Feng, J. (2014). Key functional circuitry altered in schizophrenia involves parietal regions associated with sense of self. *Hum. Brain Mapp*, 35(1), 123-139.

Hagen, L., & Kahng, A. B. (1992). New spectral methods for ratio cut partitioning and clustering. IEEE Trans. *Computer-Aided Design*, 11(9), 1074-1085. .

Hastie, T., Tibshirani, R., Friedman, J., (2009). The elements of statistical learning (Vol. 2, No. 1). New York: Springer.

Hayasaka, S. and Nichols, T. E. (2004). Combining voxel intensity and cluster extent with permutation test framework. *NeuroImage*, 23(1), 54-63.

Honey, C. J., & Sporns, O. (2008). Dynamical consequences of lesions in cortical networks. *Hum. Brain Mapp*, 29(7), 802-809. .

Kim, J., Wozniak, J. R., Mueller, B. A., Shen, X., Pan, W. (2014). Comparison of statistical tests for group differences in brain functional networks. *NeuroImage*, 101, 681-694.

Marrelec, G., Bellec, P., Krainik, A., Duffau, H., Pélégrini-Issac, M., Lehéricy, S., Doyon, J. (2008). Regions, systems, and the brain: hierarchical measures of functional integration in fMRI. *Medical Image Analysis*, 12(4), 484-496.

Park, H. J., Friston, K. (2013). Structural and functional brain networks: from connections to cognition. *Science*, 342(6158), 1238411.

Robinson, M. D., Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol*, 11(3), R25.

Rubinov, M., Sporns, O., 2011. Weight-conserving characterization of complex functional brain networks. *NeuroImage* 56 (4), 2068–2079.

Rubinov, M., Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *NeuroImage*, 52(3), 1059-1069.

Sato, J. R., Takahashi, D. Y., Hoexter, M. Q., Massirer, K. B., & Fujita, A. (2013). Measuring network's entropy in ADHD: A new approach to investigate neuropsychiatric disorders. *NeuroImage*, 77, 44-51.

Shehzad, Z., Kelly, C., Reiss, P. T., Cameron Craddock, R., Emerson, J. W., McMahon, K., . . . Milham, M. P. (2014). A multivariate distance-based analytic framework for connectome-wide association studies. *NeuroImage*, 93(P1), 74-94.

Schwartzman, A., Dougherty, R. F., Lee, J., Ghahremani, D., Taylor, J. E. (2009). Empirical null and false discovery rate analysis in neuroimaging. *NeuroImage*, 44(1), 71-82.

Shi, J., and Malik, J. (2000). Normalized cuts and image segmentation. IEEE Trans. *Pattern Analysis and Machine Intelligence*, 22(8), 888-905.

Simpson, S. L., Bowman, F. D., & Laurienti, P. J. (2013b). Analyzing complex functional brain networks: fusing statistics and network science to understand the brain. *Statistics Surveys*, 7, 1.

Simpson, S. L., Hayasaka, S., & Laurienti, P. J. (2011). Exponential random graph modeling for complex brain networks. *PLoS One*, 6(5), e20039.

Simpson, S. L., & Laurienti, P. J. (2015). A two-part mixed-effects modeling framework for analyzing whole-brain network data. *NeuroImage*, 113, 310-319.

Simpson, S. L., Lyday, R. G., Hayasaka, S., Marsh, A. P., & Laurienti, P. J. (2013a). A permutation testing framework to compare groups of brain networks. *Frontiers in Computational Neuroscience*, 7.

Simpson, S. L., Moussa, M. N., & Laurienti, P. J. (2012). An exponential random graph modeling approach to creating group-based representative whole-brain connectivity networks. *NeuroImage*, 60(2), 1117-1126.

Smith, S. M. (2012). The future of FMRI connectivity. *NeuroImage*, 62(2), 1257-1266.

Sporns, O. (2011). The human connectome: a complex network. *Ann. N. Y. Acad. Sci*, 1224(1), 109-125.

Sporns, O. (2012). From simple graphs to the connectome: networks in neuroimaging. *NeuroImage*, 62(2), 881-886.

Strimmer, K. (2008). A unified approach to false discovery rate estimation. *BMC Bioinformatics*, 9(1), 303.

Tyszka, J. M., Kennedy, D. P., Paul, L. K., & Adolphs, R. (2014). Largely typical patterns of resting-state functional connectivity in high-functioning adults with autism. *Cerebral Cortex*, 24(7), 1894-1905.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., ... & Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, 15(1), 273-289.

van den Heuvel, M. P., Stam, C. J., Boersma, M., & Pol, H. H. (2008). Small-world and scale-free organization of voxel-based resting-state functional connectivity in the human brain. *NeuroImage*, 43(3), 528-539.

van den Heuvel, M. P., Mandl, R. C., Stam, C. J., Kahn, R. S., & Pol, H. E. H. (2010). Aberrant frontal and temporal complex network structure in schizophrenia: a graph theoretical analysis. *J. Neurosci*, 30(47), 15915-15926.

van den Heuvel, M. P., Stam, C. J., Kahn, R. S., & Pol, H. E. H. (2009). Efficiency of functional brain networks and intellectual performance. *J. Neurosci*, 29(23), 7619-7624.

Varoquaux, G., & Craddock, R. C. (2013). Learning and comparing functional connectomes across subjects. *NeuroImage,* 80, 405-415.

Von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and Computing*, 17(4), 395-416.

Xia, M., Wang, J., He, Y. (2013). BrainNet Viewer: a network visualization tool for human brain connectomics. *PLoS One*, 8(7), e68910.

Zalesky, A., Fornito, A., and Bullmore, E. T. (2010). Network-based statistic: identifying differences in brain networks. *NeuroImage*, 53(4), 1197-1207.

Zalesky, A., Cocchi, L., Fornito, A., Murray, M. M., Bullmore, E. (2012a). Connectivity differences in brain networks. *NeuroImage*, 60(2), 1055-1062.

Zalesky, A., Fornito, A., Bullmore, E. (2012b). On the use of correlation as a measure of network connectivity. *NeuroImage*, 60(4), 2096-2106.

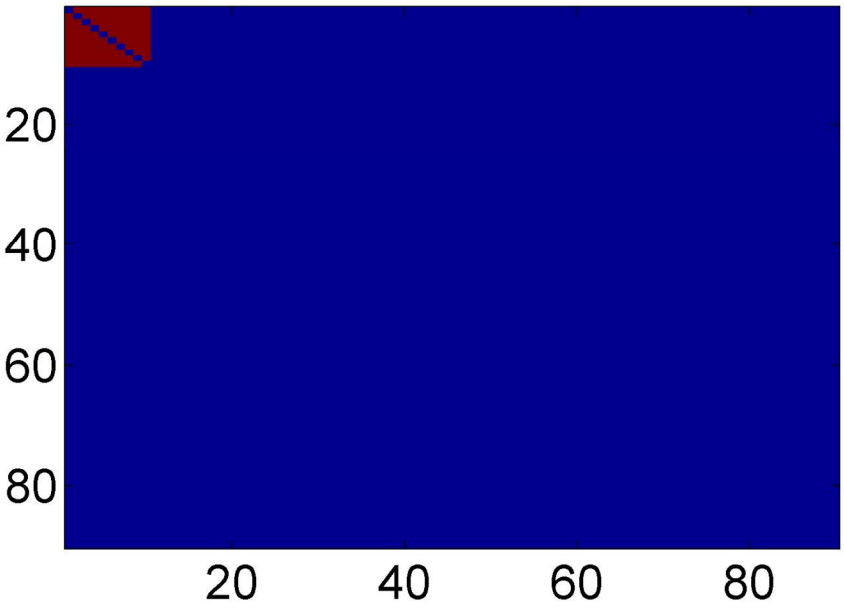

Fig 1(a)the heatmap of truth the connectivity between the first 10 nodes are differentially expressed between the two groups;
150x100mm (300 x 300 DPI)

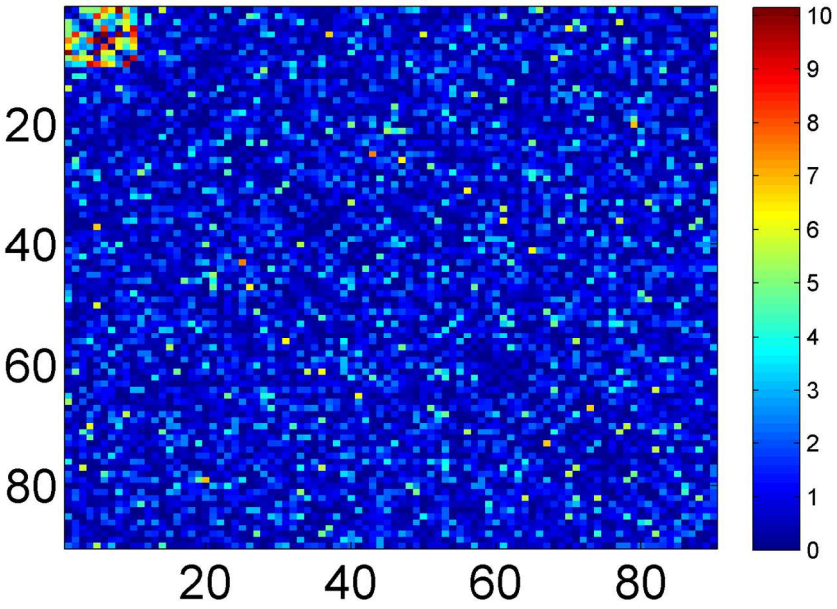

Fig 1(b) heatmap of two sample t test -log(p) values of the simulated connectivity based on 60 subjects (30 cases vs. 30 controls);
150x100mm (300 x 300 DPI)

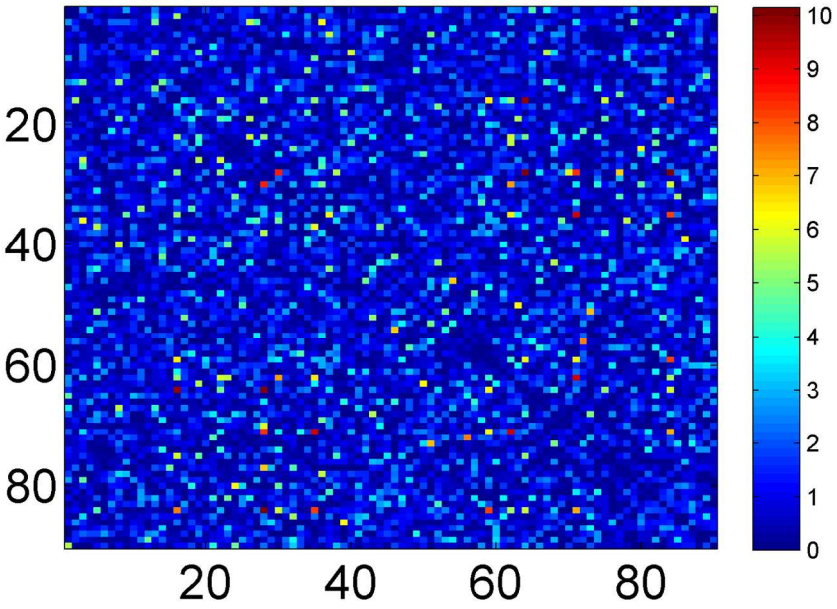

Fig 1(c) heatmap with shuffled region number of (b) is used the input of our method.
150x100mm (300 x 300 DPI)

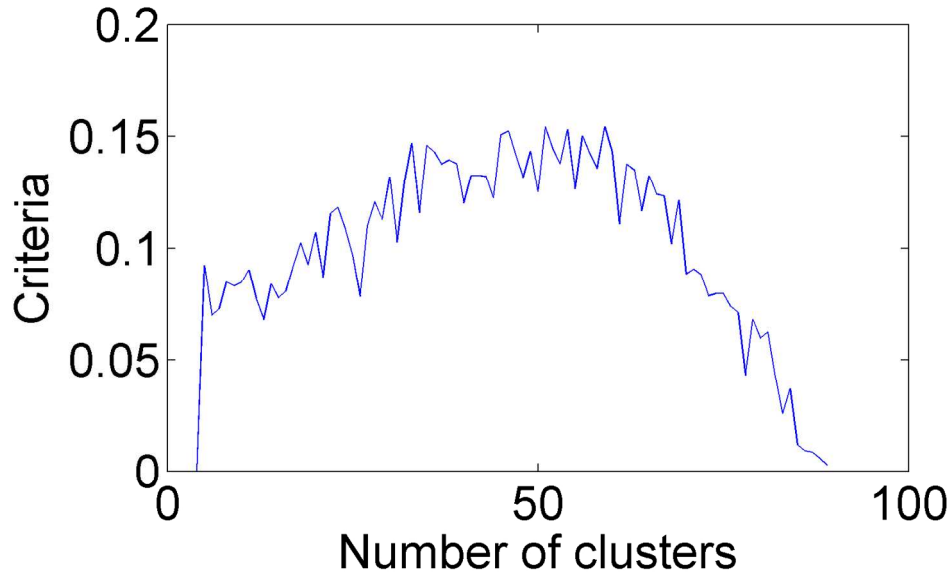

Fig 2(a) the number of cluster selection criteria function: scores of the function vs. the number of clusters; 150x90mm (300 x 300 DPI)

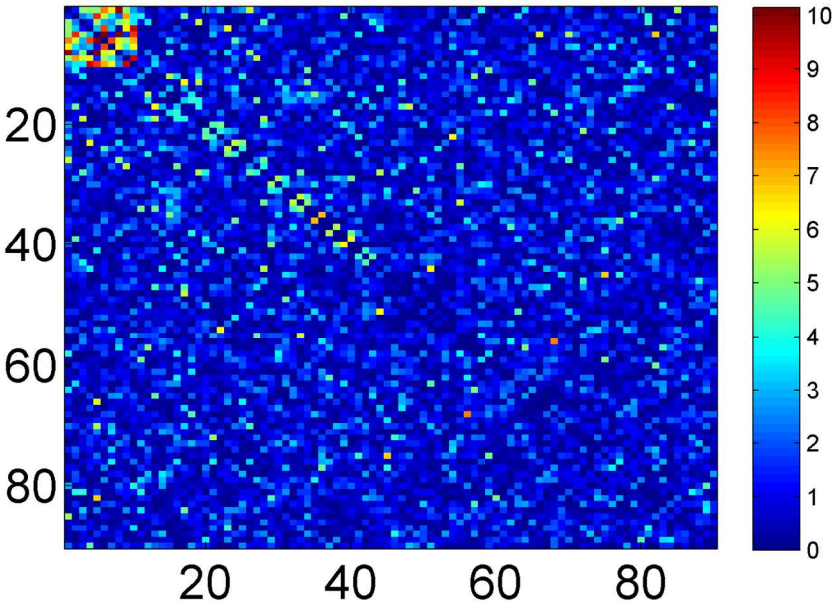

Fig 2(b) The resulting heatmap of –log(p): the detected network is at the left-top corner.
150x100mm (300 x 300 DPI)

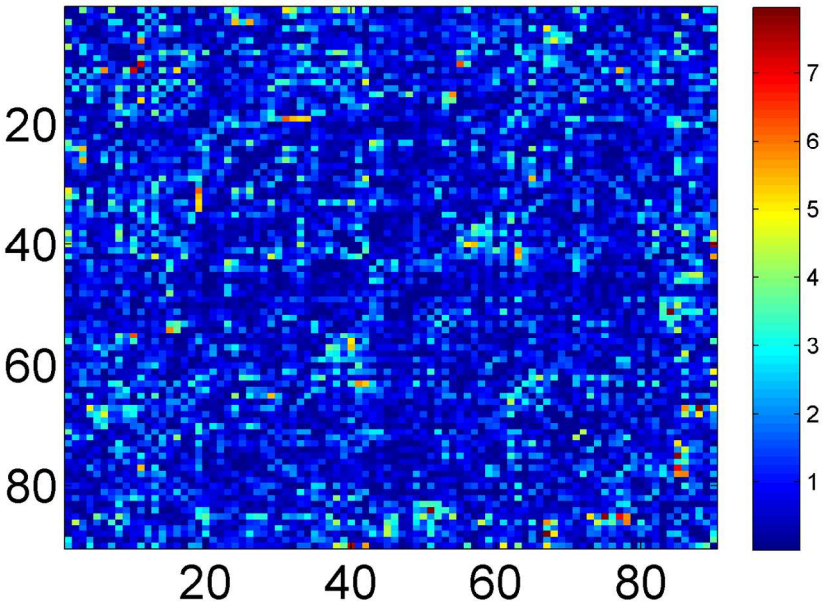

Fig 3(a) heatmap of -log p values for all edges between TC and TSD; 150x100mm (300 x 300 DPI)

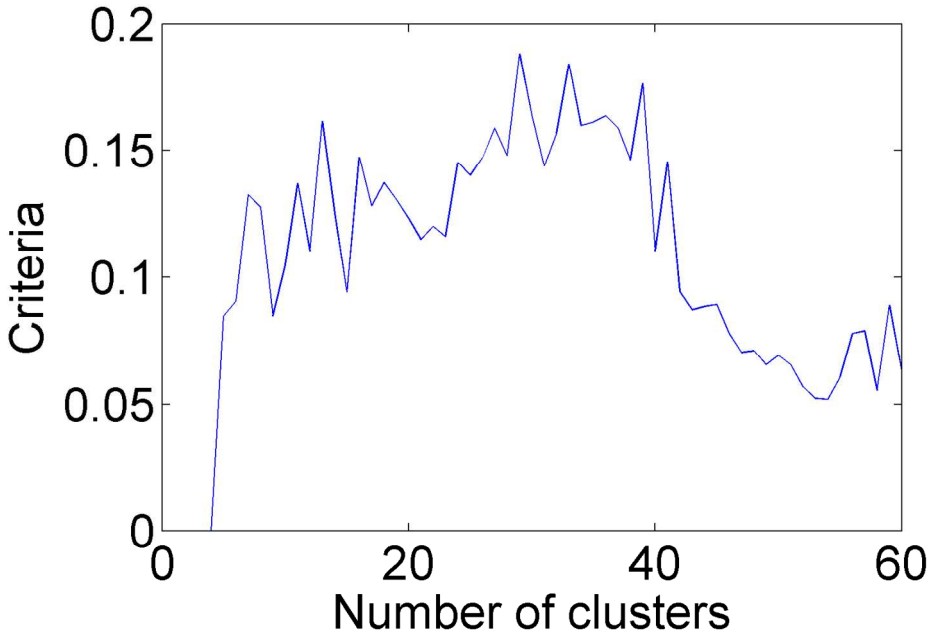

Fig 3(b) tuning parameter selection criteria function;
150x100mm (300 x 300 DPI)

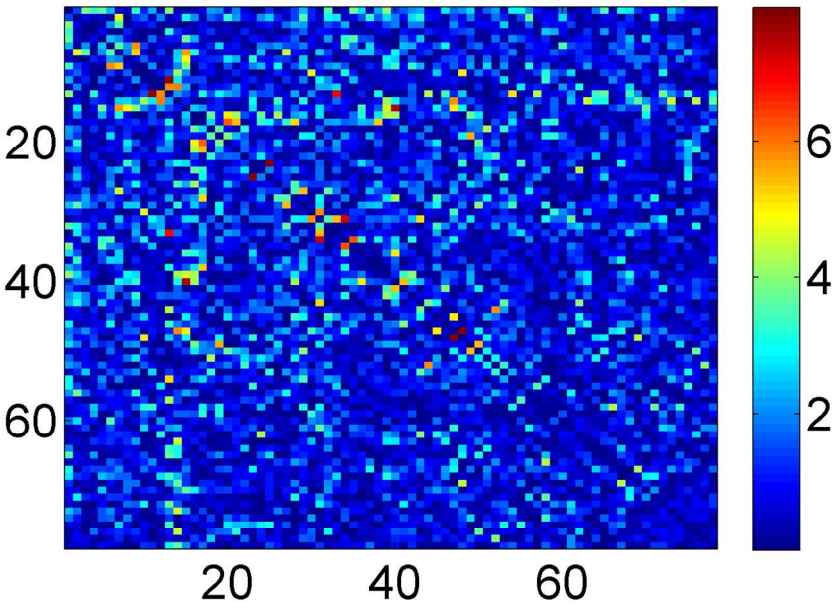

Fig 3(c) the resulting heatmaps: detected networks along the diagonal.
150x100mm (300 x 300 DPI)

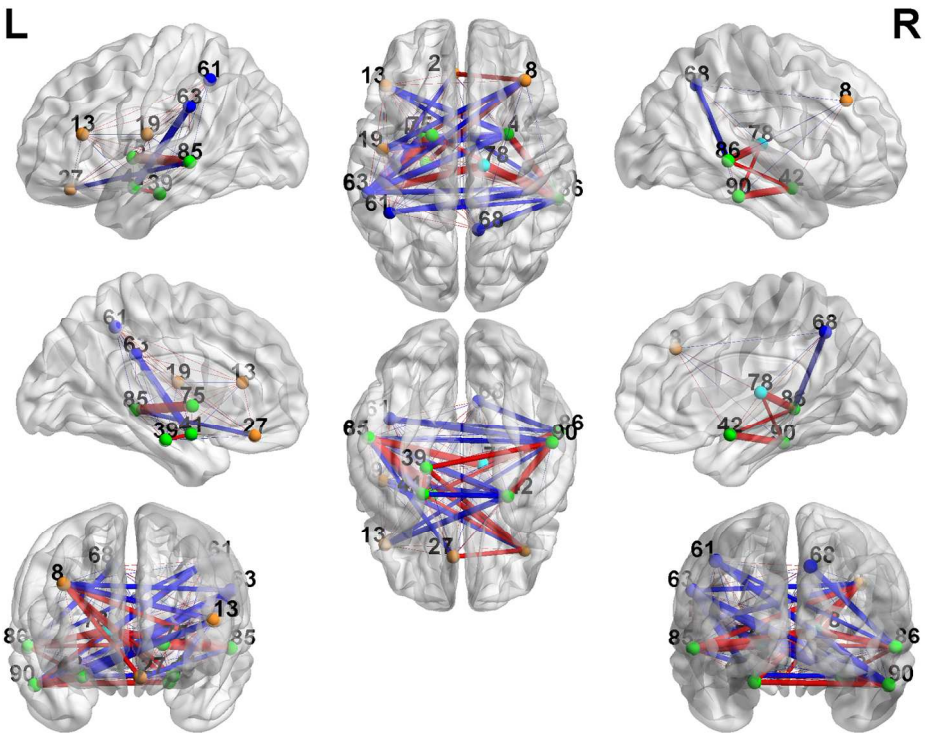

Figure 4 Cluster 1 3D plots of the differentially expressed edges. The width of the edges reflects the significance level, and the color is coded as: red (TC < TSD) and blue (TC > TSD)
149x111mm (300 x 300 DPI)

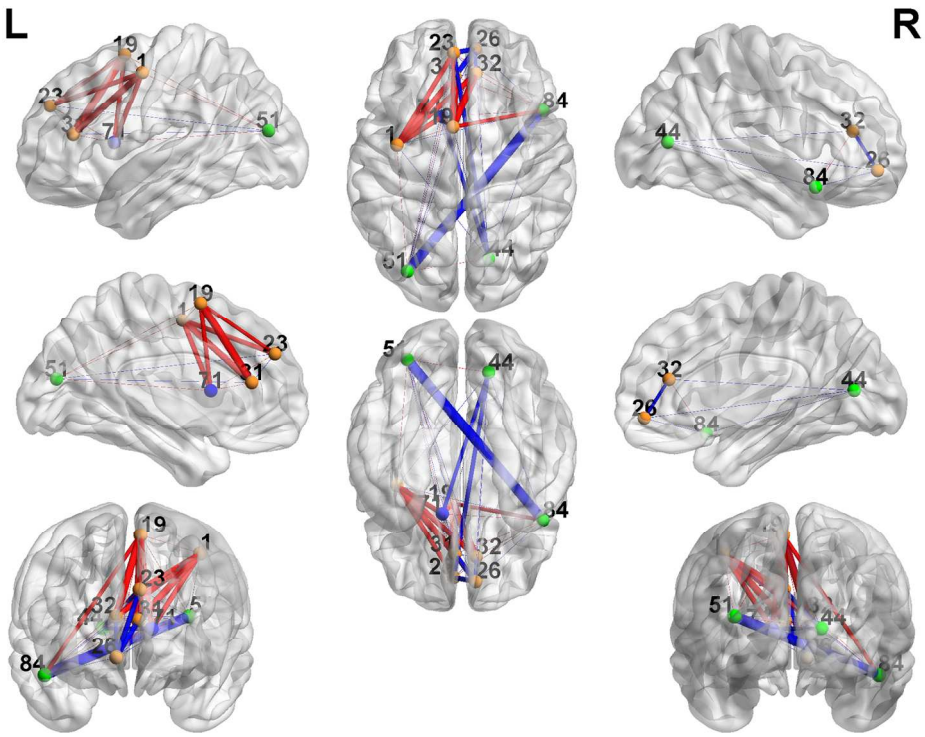

Figure 5: Cluster 2 3D plots of the differentially expressed edges. The width of the edges reflects the significance level, and the color is coded as: red (TC < TSD) and blue (TC > TSD).
149x111mm (300 x 300 DPI)

Table 1. Simulation results under different settings

	Pard			FDR			NBS		
	FP	FN	Network	FP	FN	Network	FP	FN	Network
Size=10 σ^2=1	0.3 ±0.17	0	Yes	0.29 ±0.06	40.42 ±0.35	No	0	45	No
Size=5 σ^2=1	4.519 ±0.47	0	Yes	0.19 ±0.05	9.45 ±0.10	No	0	10	No
Size=20 σ^2=1	1.33 ±0.51	0	Yes	3.21 ±0.18	134.43 ±1.12	No	0	190	No
Size=10 σ^2=0.25	0	0	Yes	1.04 ±0.11	31.91 ±0.49	No	6.37± 1.17	27.28± 2.20	Yes
Size=10 σ^2=5	19.64 ±2.12	16.28 ±1.44	Yes	0.05 ±0.02	44.88 ±0.04	No	0	45	No