Brain Connectivity Biomarkers

Shuo Chen

Department of Epidemiology and Biostatistics
School of Public Health, University of Maryland, College Park

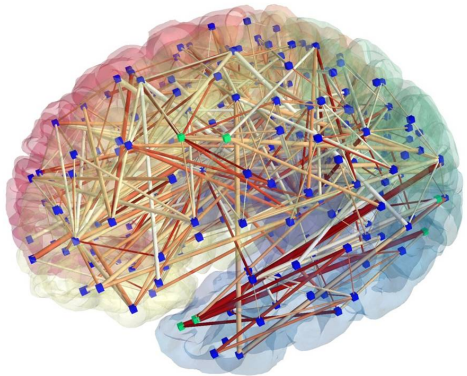
Feb 4th, 2016
Mathematical and Statistical Challenges in Neuroimaging Data Analysis

UNIVERSITY OF MARYLAND

Outline

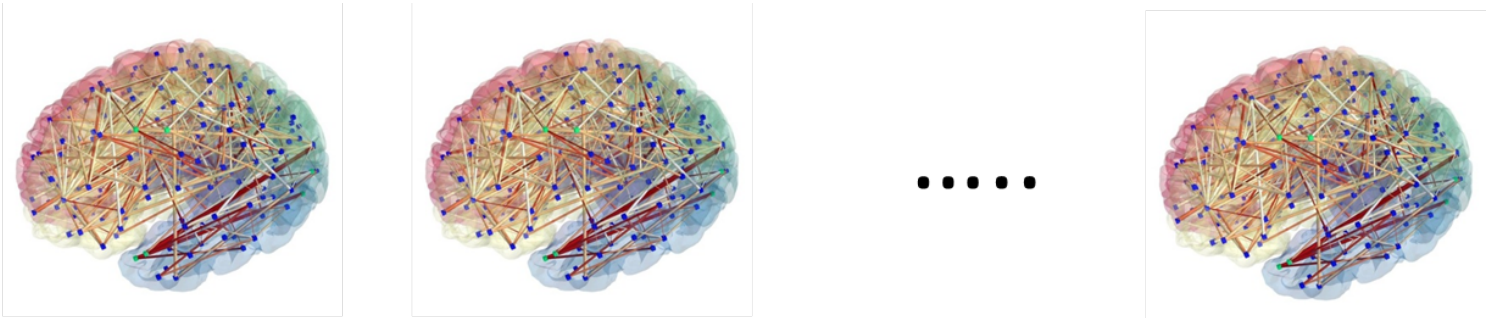
Brain Connectivity

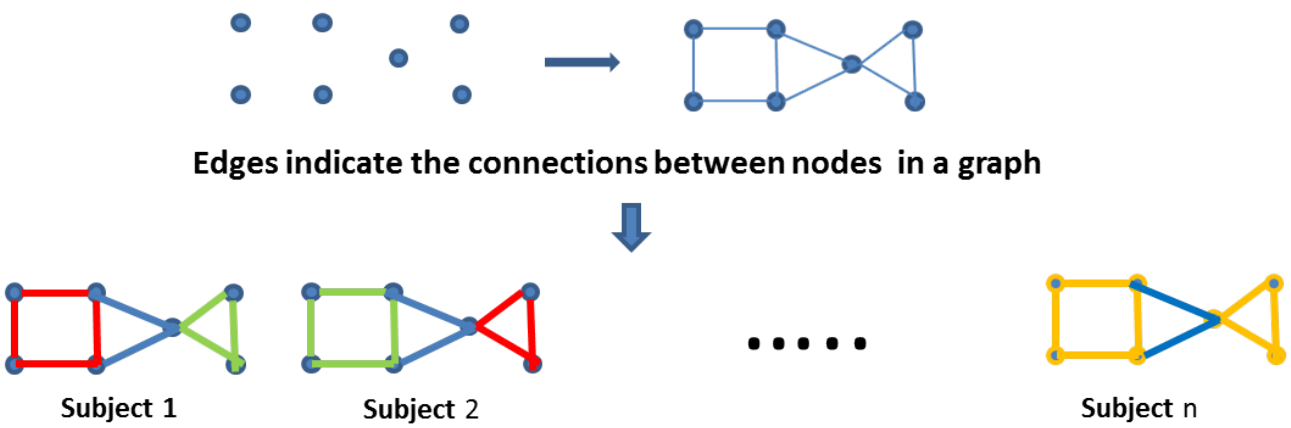
- Inter-links between neuronal processing units, which is formed by chemical/electronic signal pathways or neural fibre pathways.
- The brain connectivity networks are responsible for our perception to outer world, behaviours, and emotions; mental illness can also alter brain connectivity patterns.
- fMRI technology provides a means to detect the **functional brain connectivity** by measuring the coherence of temporal profiles from spatially distinct locations (e.g. correlation or MIC). Different clinical groups or experimental conditions show differences in connectivity network patterns.



http://blog.enthought.com/enthought-tool-suite/visualizing-brain-connectivity-with-ets

Brain Networks and Graphs





For one subject, V nodes, and $|E|{=}V \times (V-1)/2$ edges, $G = \{V, E\}$.

What is a connectivity biomarker?

1. In a group level connectivity study, we are interested in detecting genuinely differentially expressed connectivity metrics (recall the early stage of micro-array data analysis): univariate or individual edge.

   However, inferences on individual edges (e.g. FWER, local fdr, or lasso type shrinkage techniques) are subject to two major flaws:

   - Universal cut-off causes trade-off betweenloss of power and false positive discovery by multiple testing correction.
   - It is fine to detect the difference on each edge, but the decision rule should account for the dependence of edges.
   - Loss of spatial and topological patterns: network structure or graph structure.

2. Graph descriptive methods (e.g. modularity or small-worldness) which is lack of location details and prediction power.

3. Predefine networks and compare them across subjects, which could not guarantee the differential expression of edges and faces selection bias and multiple-testing control.

What is a connectivity biomarker?

The challenge is to improve the power when controlling the multiple testing.

3. Network based statistics (NBS, Zalesky et al, 2010) is the first method to detect a set of edges as a biomarker.
    - Set a suprathreshold for example $p < 0.1$ and binarize edges.
    - Detect breadth first network detection.
    - Test the significance of the 'structure' by permutation test (using the number of edges as the statistic).

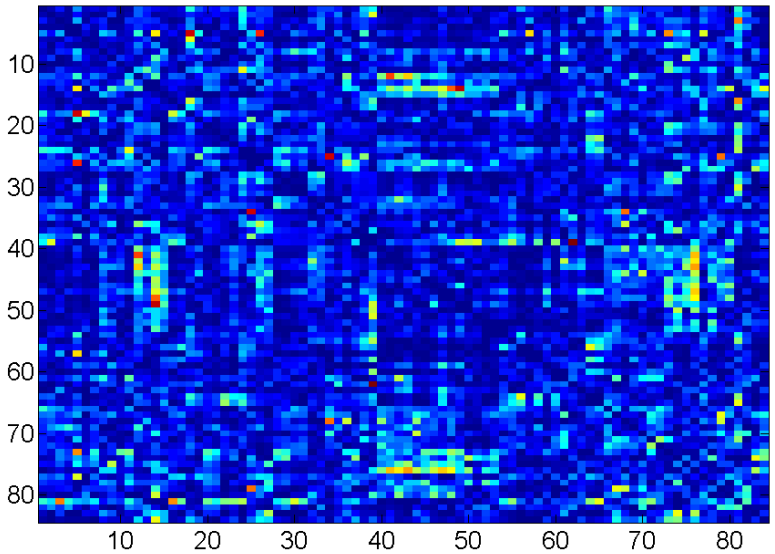
But...
But, NBS is very sensitive to noises and lack of power because adding a false positive node will bring n_k FP edges (Chen et al, 2015).
The detected 'structure' has no explicit topology.
Hard to interpret the results.

Our Beliefs

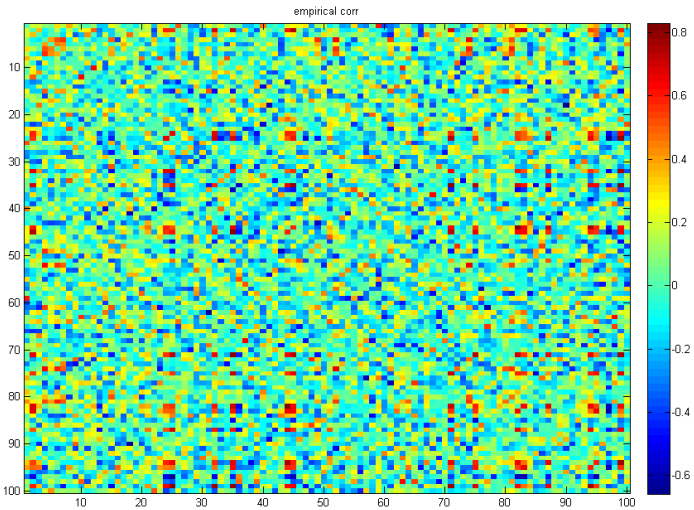
- Differentially expressed edges are not distributed randomly.
- They follow an organized and complex, yet latent pattern. But it could be challenging to identify the pattern as the permutation of 100 nodes is 10^{159}.
- Perform statistical tests by accounting for such topology is equivalent to adjusting the correlation between edges (across subjects) because edges can borrow strength from each other (similar to the spatial dependence adjustment).

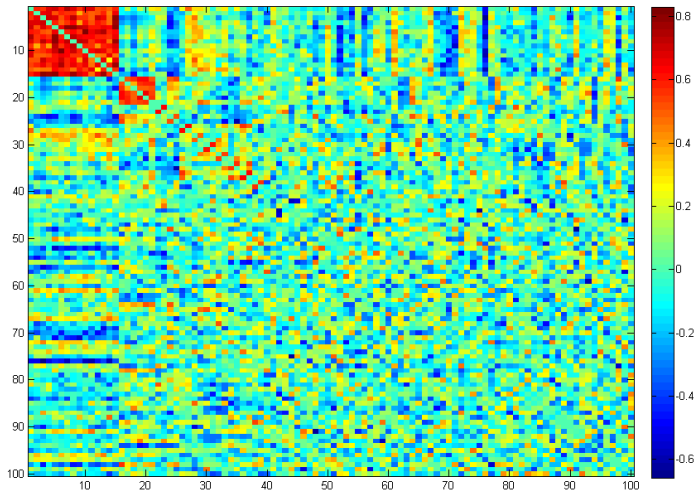
Input Data

- $G = \{V, E\}$, and edges are features of interest.
- $w_{ij} = -log(p_{ij})$ (closely linked to information entropy), where p_{ij} is the test p-value (could be permutation test) for connectivity between nodes i and j .
- Goal: capturing most differentially expressed connections within networks of constrained numbers of brain regions (nodes).
- The rule of parsimony is the key as it is very rare to observe a subgraph with a high proportion of edges with large $-log(p_{ij})$.



Detect a Mixture of SBM and Random graph from input data





Parsimonious differential network **d**etection (Pard)

- Definition: **object oriented connectivity network biomarker**. If our data is $\mathbb{M} = \{\mathbf{M}_i, \cdots, \mathbf{M}_S\}$, the biomarker $S_k = \{V_k, E_k, \mathcal{T}_k\}$ is a graph based strongly non-Euclidean object oriented statistic.
- Objective function: capturing most differentially expressed connections within networks of constrained numbers of brain regions (nodes).
- A stochastic block model and random graph mixture (graph object oriented biomarker version of local fdr).

Parsimonious differential network detection (Pard)

- Step 1: Build the weighted adjacency matrix **W**, perform screening, and identify the disconnected components using Laplacian matrix,

$$W_{ij} = \begin{cases} -\log(p_{ij}) & \text{if } p_{ij} \leq p_0; \\ 0 & \text{if } p_{ij} > p_0. \end{cases}$$

- Step 2: within a connected component, capture as many informative edges as you can along the diagonal blocks (K tunes the parsimony level):

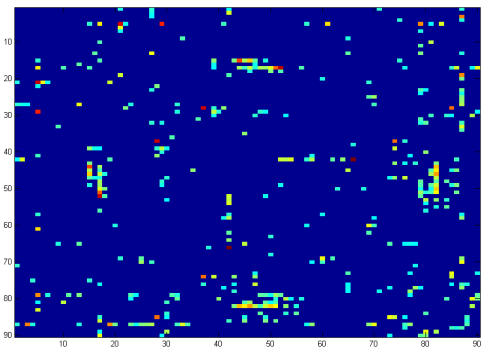
$$\arg\min_{\{\widetilde{A}_k\}_{k=1}^{K_q}} \sum_{k=1}^{K_q} \frac{\sum_{i\in\widetilde{A}_k, j\notin\widetilde{A}_k} -\log(p_{ij})}{|\widetilde{A}_k|} \tag{1}$$

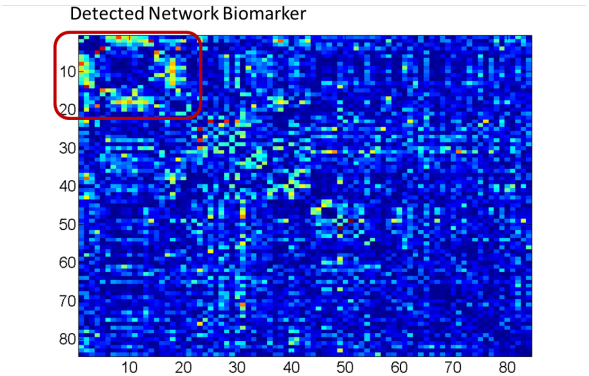
- Step 3: applying the 'quantity and quality' standard.

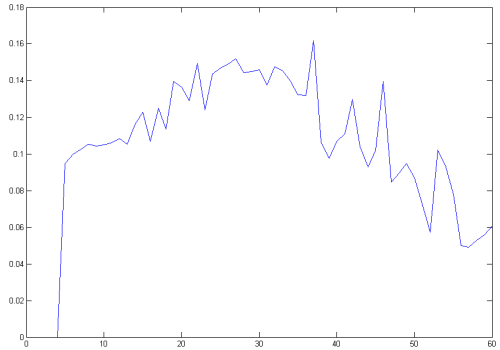
$$\frac{\sum_{k=1}^{K_q}\sum_{i\in\widetilde{A}_k, j\in\widetilde{A}_k} I(W_{ij}>0)}{\sum_{i<j} I(W_{ij}>0)} \cdot \frac{\sum_{k=1}^{K_q}\sum_{i\in\widetilde{A}_k, j\in\widetilde{A}_k} I(W_{ij}>0)}{\sum_{k=1}^{K_q}\sum_{i\in\widetilde{A}_k, j\in\widetilde{A}_k} 1}. \tag{2}$$

Implementation

- Step 1: Laplacian matrix $L = D - W$ and number of zero eigenvalues of L equals the number of disconnected subgraphs (faster than NBS breadth first searching).
- Step 2: Perform spectral clustering algorithm the cut solution does not depend on the initial point, with the optimal selection of K by using the 'quantity and quality' criteria.
- Step 3: Draw inferences on the detected network.







Step 1 Compute the degree matrix D_s, where $D_s(i,i) = \sum_{j=1}^{N} W_s(i,j)$.

Step 2 Find the eigen-solutions $[V_s, L_s]$ of $D_s^{-\frac{1}{2}} W_s D_s^{-\frac{1}{2}}$, i.e., solving $D_s^{-\frac{1}{2}} W_s D_s^{-\frac{1}{2}} V_s = V_s L_s$ and $V_s V_s = I_N$. Then compute $Z_s = D_s^{-\frac{1}{2}} V_s$.

Step 3 Normalize Z_s by setting $X_s = \mathrm{diag}^{-1}[\mathrm{diag}(Z_s Z_s^T)] Z_s$, where operation $\mathrm{diag}(A)$ extracts the diagonal elements of matrix A as a vector; and $\mathrm{diag}^{-1}(a)$ creates a matrix with diagonal elements equal to a and off-diagonal being zeros.

Step 4 Set the convergence criterion parameter $\rho^* = 0$, and initialize a $K \times K$ matrix R_s by the following steps: denote by R_s^k the kth column of R_s for $k = 1, \ldots, K$. Set $R_s^1 = [X_s(i,1), \ldots, X_s(i,K)]^{\mathrm{T}}$, where i is randomly selected from $\{1, \ldots, N\}$. We denote the first column of R_s as R_s^1 and the k_{th} column as R_s^k. Then update the rest of the columns by following.

For $k = 2, \ldots, K$, iteratively update $R_s^k = [X_s(i_k,1), \ldots, X_s(i_k,K)]^{\mathrm{T}}$ where

$$i_k = \underset{i \in \{1,\ldots,N\}}{\arg\min}\ c_{k-1}(i), \text{ and } c_{k-1} = \sum_{l=1}^{k-1} |X_s R_s^l|.$$

Step 5 Minimize the objective function: $\sum_s ||Y - X_s R_s||^2 = ||Y - \widehat{XR}||^2$.

where $||\cdot||$ stands for Frobenius norm; and $\widehat{XR} = \sum_s t_s X_s R_s$ with

$$t_s = \frac{1/||X_s R_s - XcRc||^2}{\sum_s 1/||X_s R_s - XcRc||^2}.$$

The term $XcRc$ is the centroid of $X_s R_s$ which minimizes $\sum_{s' \neq s} ||X_s R_s - X_{s'} R_{s'}||^2$ with respect to $X_{s'} R_{s'}$.

Then $Y(i,l) = 1$, where $l = \arg\max_{k \in \{1,\ldots,K\}} \widehat{XR}(i,k)$, $i \in \{1, \ldots, N\}$ *and* $l \in \{1, \ldots, K\}$.

Step 6 Conduct singular value decomposition on the matrix $Y^T X_s$

$Y^T X_s = U_s \Omega_s V_s{}^T$

$\rho = \sum_s tr(\Omega_s)$

If $|\rho - \rho^*| <$ pre-assigned error limit then output Y,

else, update $R_s = V_s U_s^T$.

Step 7 Go to Step 5.

Remarks

- Our first step is is equal to NBS but in a computationally efficient way.
- In our network detection, K plays as a shrinkage parameter; and a larger K will lead to fewer networks.
- Our model provides a new algorithm to automatically detect a mixture of stochastic block model and random graph model.
- The resulting discrete solutions are nearly global-optimal and the estimate is consistent (Lei and Rinaldo 2014).
- It is RARE to observe a topological pattern.

Significance of a network

- Multiple test adjustment for the object oriented network biomarker, because we detect many networks.
- Zalesky et. al. 2010 developed a multiple testing control procedure in 'weak sense' (regarding individual edge) by permutation tests by using the size of the network (number of suprathreshold edges).
- The number of edges should be proportional to the number of the nodes.

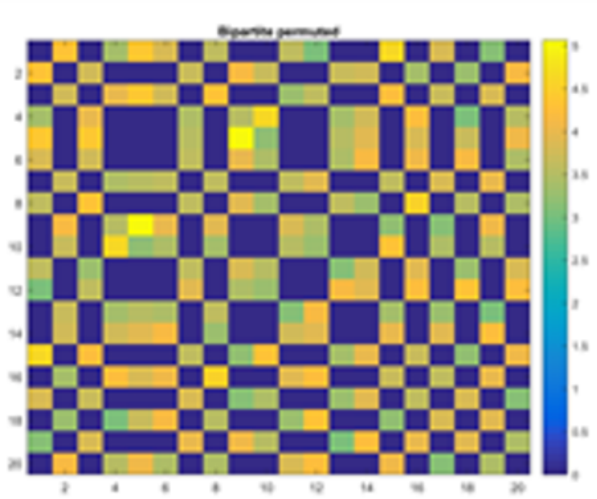
Significance of a network

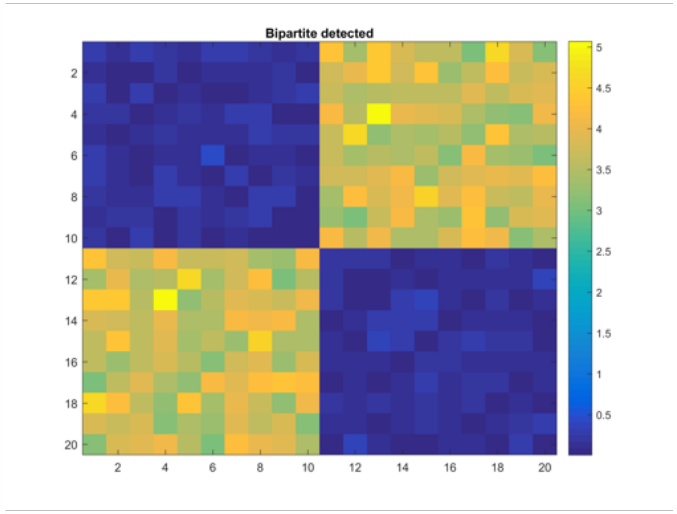
We are testing an strong non-Euclidean object oriented statistic $S_k = \{V_k, E_k, \mathcal{T}_k\}$.
H_0: the edges in a detected network are as informative as the overall graph G
H_1: the edges in a detected network are more informative as the overall graph G

- Permutation test (shuffling nodes' label 20,000 times), and use $\chi^2_{2E_k} = -2\sum_{E_k} ln(p_{ij})$ and corresponding p-values as test statistics, and single edge can be tested as well.
- Exact test: binarizing edges first (e.g. $p < 0.1$), calculate the exact test p value by hypergeometric distribution and Bonferroni correction for the number of networks.
- Bayes Factor based test BF$=\frac{\prod_{e_{ij}\in G_k} f_1(t_{ij})}{\prod_{e_{ij}\in G_k} f_0(t_{ij})}\frac{\pi_1}{\pi_0}$, π_0 prior reflects how rare to observe an informative network.

Where shall we proceed next?

- A community based network seems good, but what if the brain is more complicated?
- We make a further a tempt to investigate more detailed graph topology to improve the power and lower the false positive discovery rates.
- For illustration, we use a example K-partite graph to further investigate the within community topology.
- The algorithm is similar to Pard, however in an 'opposite' way.





K-partite graph detection

- Step 1: Look into the adjacency matrix for a detected network by using Pard,

$$W_{ij} = \begin{cases} -\log(p_{ij}) & \text{if } p_{ij} \leq p_0; \\ 0 & \text{if } p_{ij} > p_0. \end{cases}$$

- Step 2: within a detected network, we hope to identify the k-partite communities:

$$\arg\max_{\{\widetilde{A}_k\}_{k=1}^{K_q}} \sum_{k=1}^{K_q} \frac{\sum_{i\in\widetilde{A}_k, j\notin\widetilde{A}_k} -\log(p_{ij})}{|\widetilde{A}_k|} \tag{3}$$

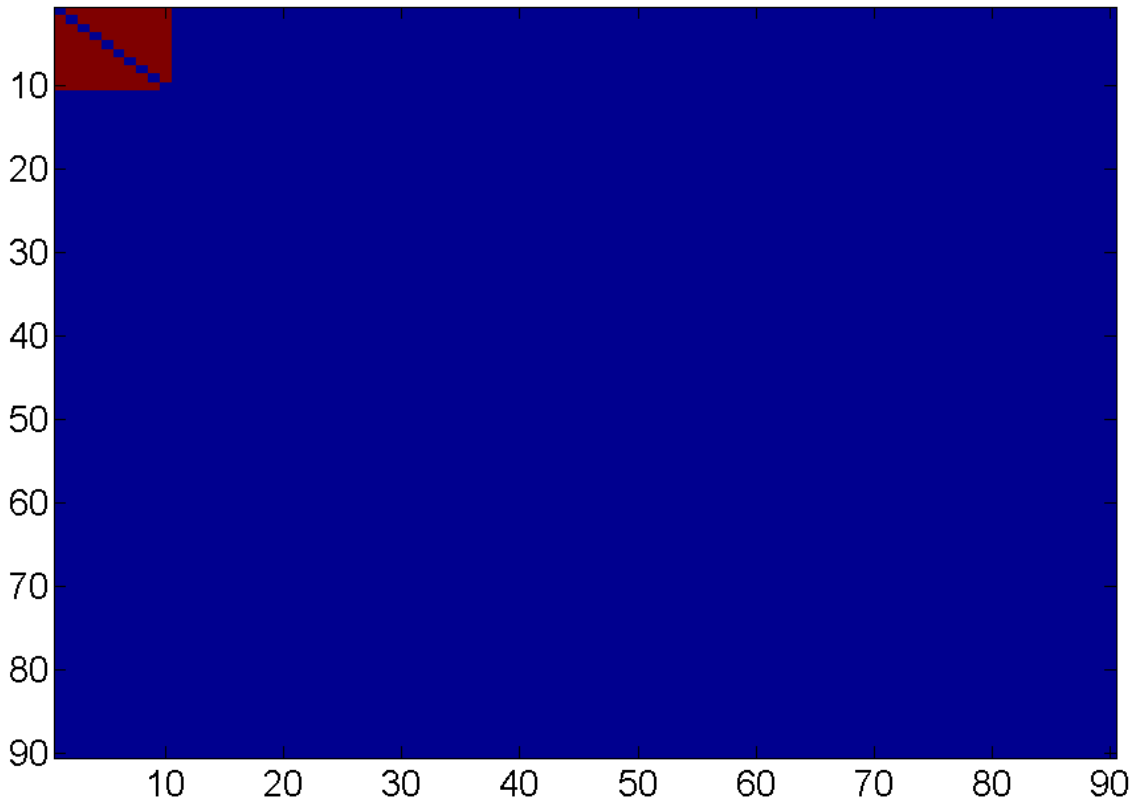
- Step 3: applying the 'quantity and quality' standard to ensure all informative edges are moved to the off-diagonal.

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

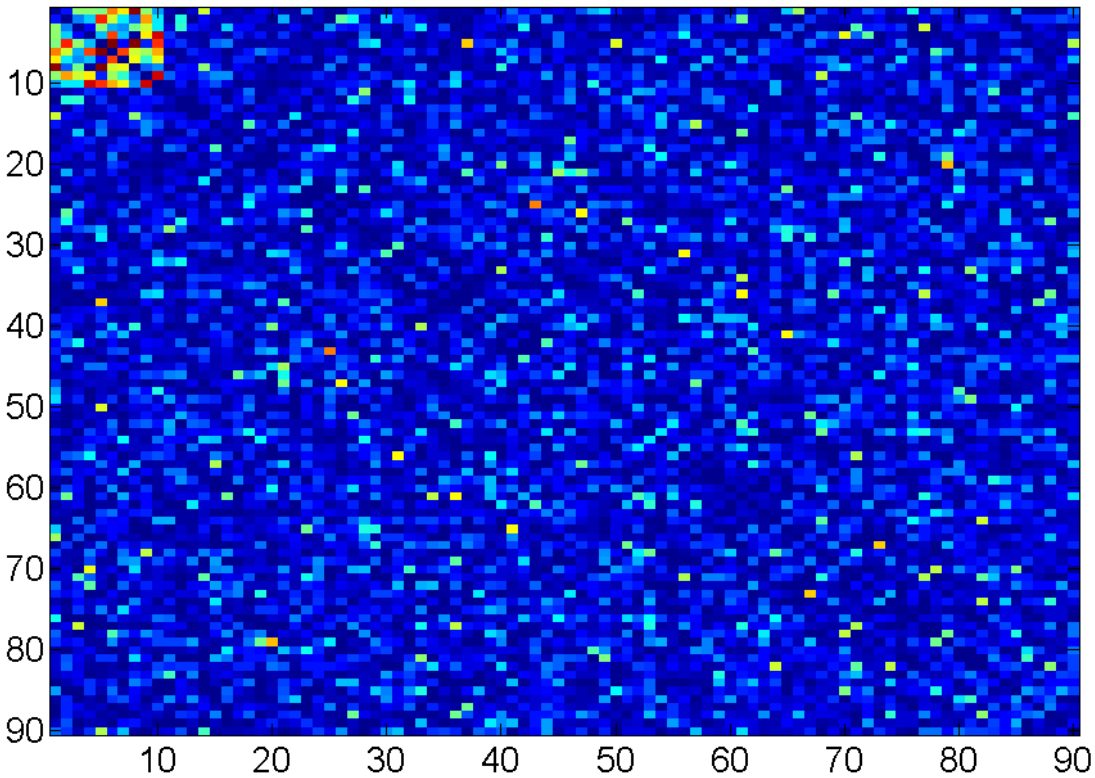
Results

Results

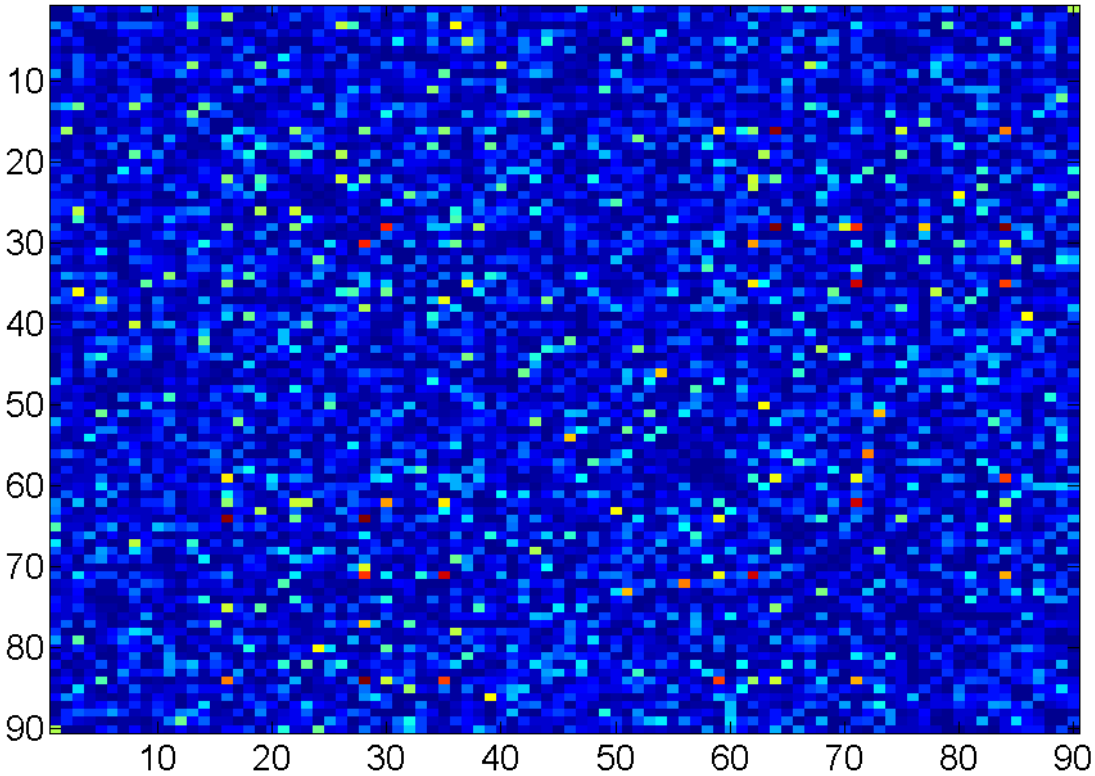
Simulation



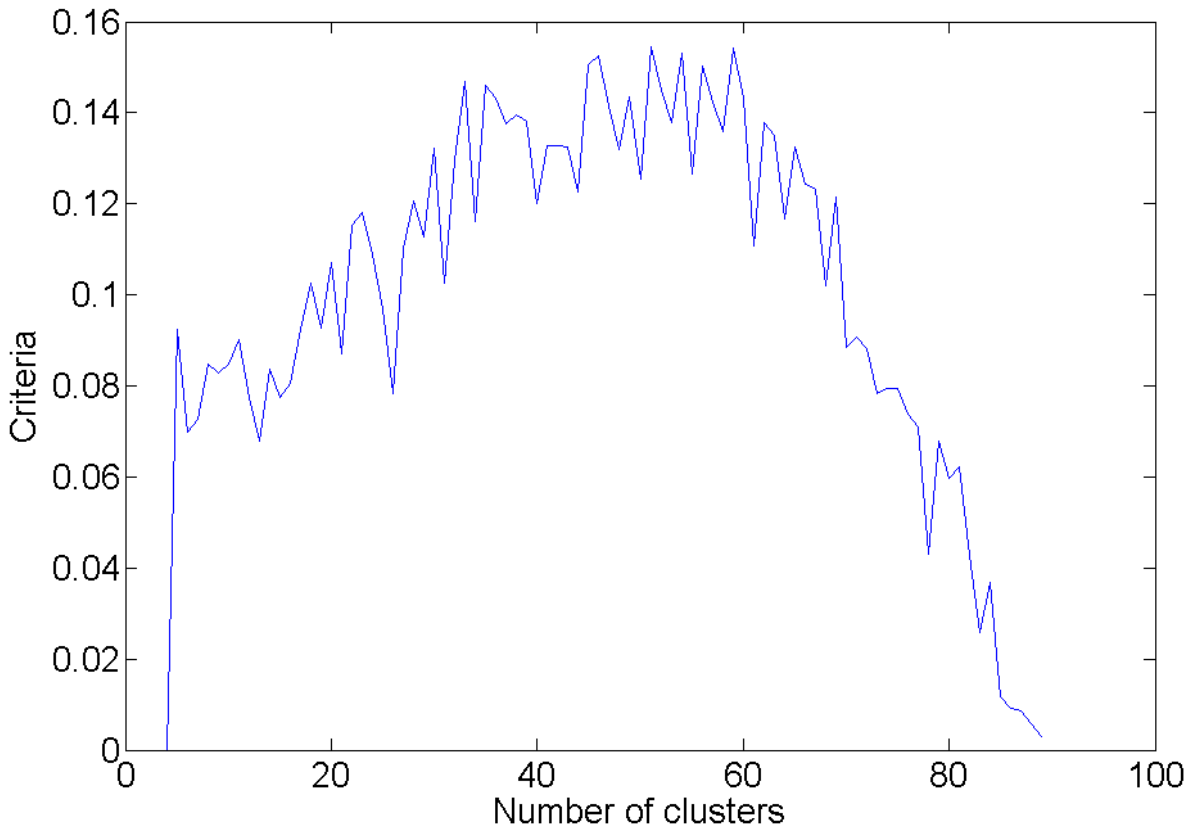
Simulation



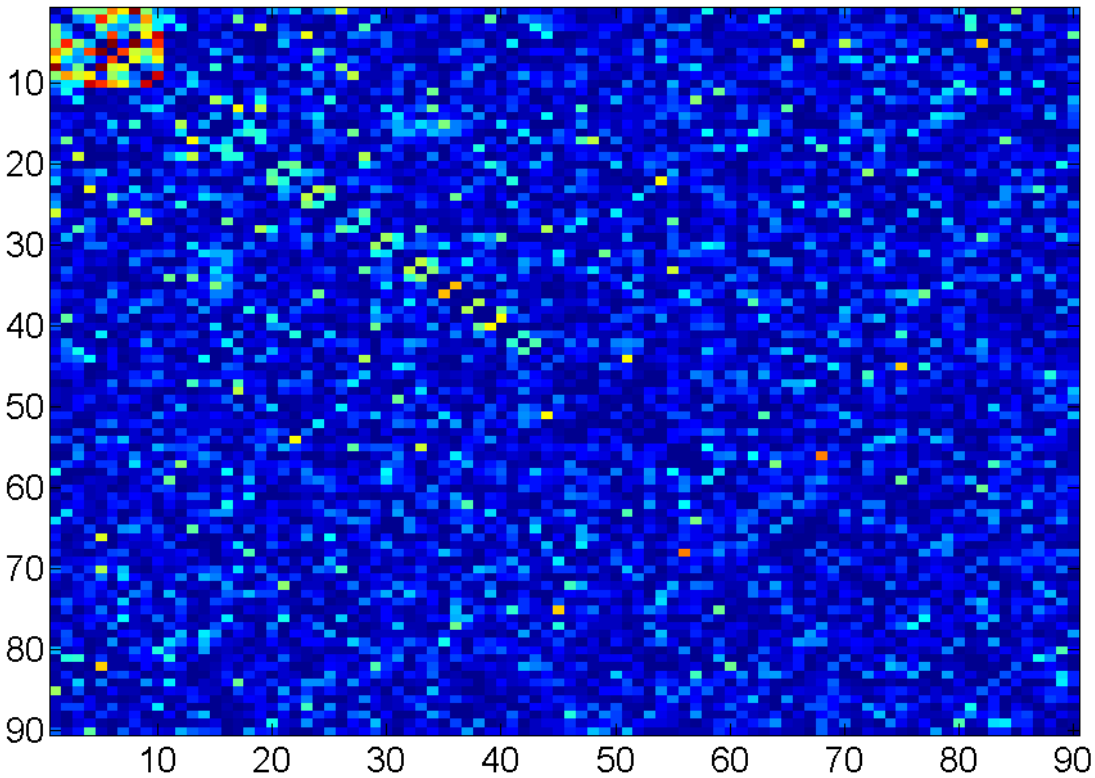
Simulation



Simulation Results



Simulation Results



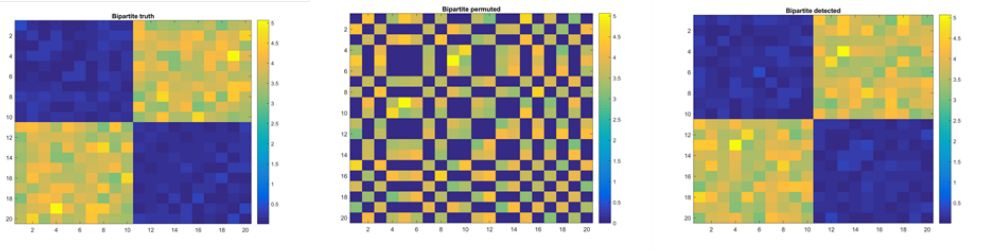
Simulation Results of Pard

TABLE I. Simulation results under different settings

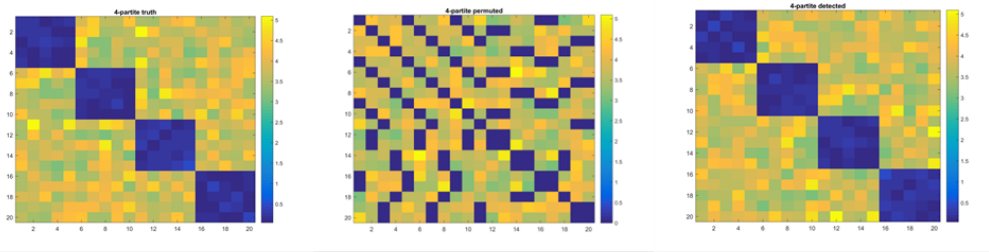
	Pard			FDR			NBS		
	FP	FN	Network	FP	FN	Network	FP	FN	Network
Size = 10; $\sigma^2 = 1$	0.3 ± 0.17	0	Yes	0.29 ± 0.06	40.42 ± 0.35	No	0	45	No
Size = 5; $\sigma^2 = 1$	4.519 ± 0.47	0	Yes	0.19 ± 0.05	9.45 ± 0.10	No	0	10	No
Size = 20; $\sigma^2 = 1$	1.33 ± 0.51	0	Yes	3.21 ± 0.18	134.43 ± 1.12	No	0	190	No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

Simulation Results of K-partite graph detection

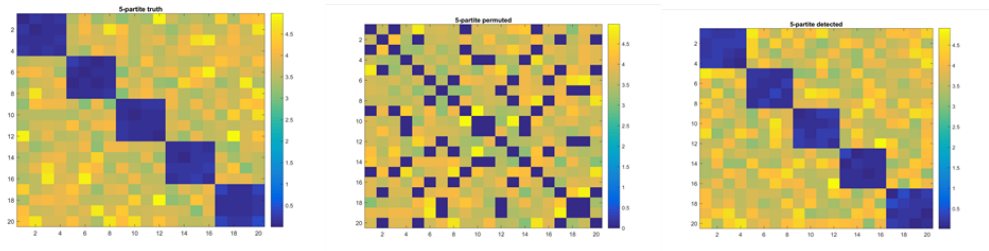
Bipartite

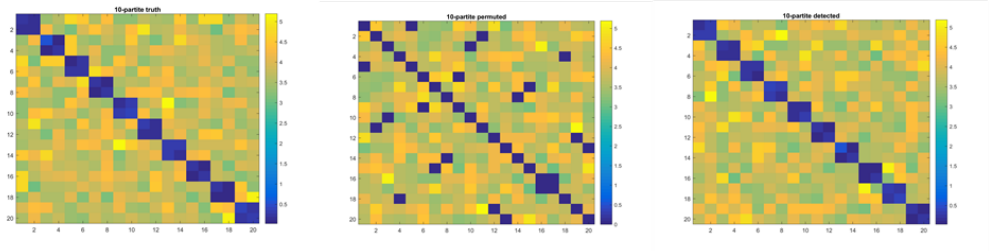


4-partite



5-partite





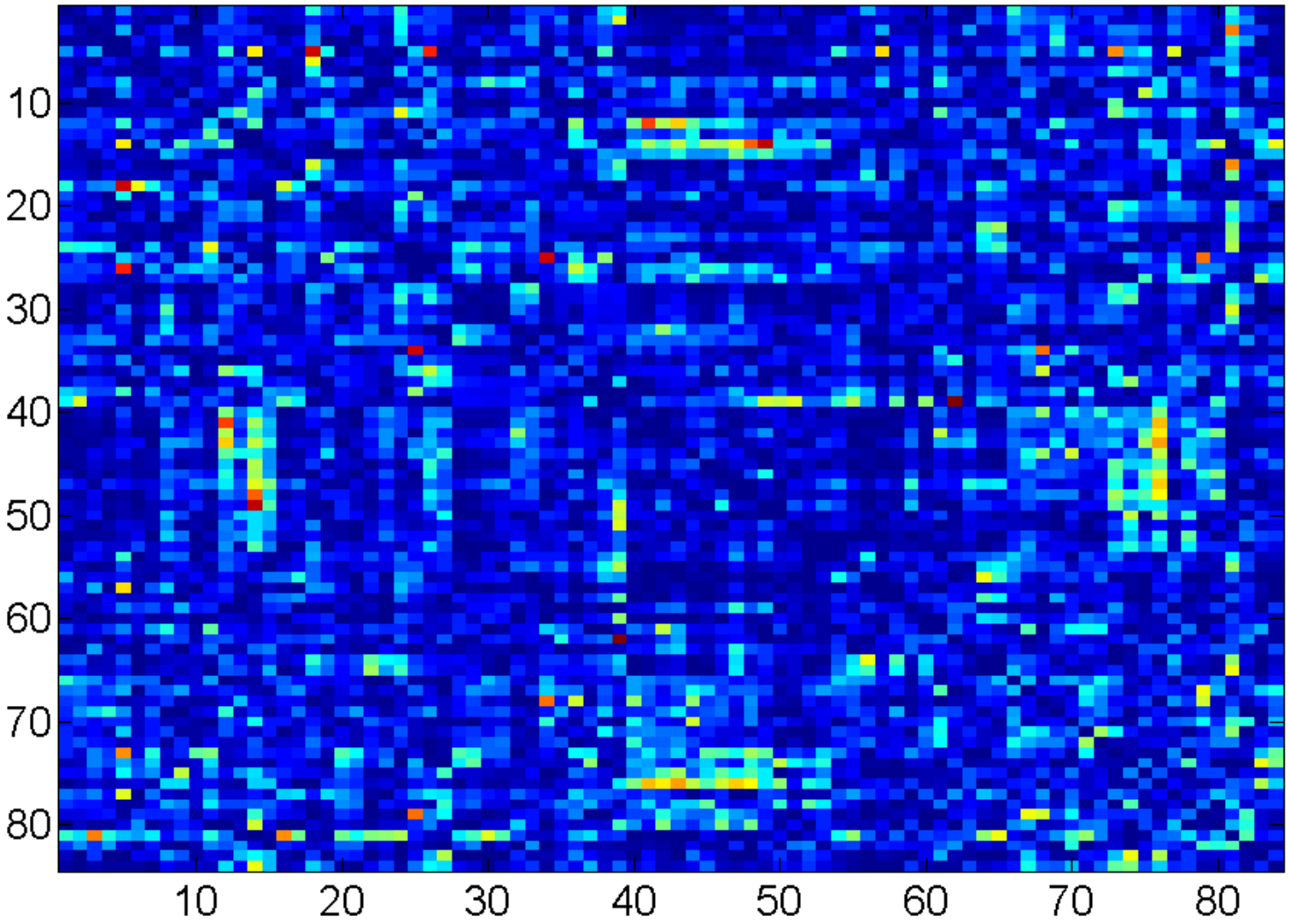
Simulation Results of K-partite graph detection

Sig Difference *	FP(k=2)	FN	FP(k=4)	FN	FP(k=5)	FN	FP(k=10)	FN
3	0.02(0.14)	0.78(1.52)	0.47(2.01)	0.82(1.16)	0.41(1.7)	0.93(1.15)	1.03(1.82)	1.36(1.1)
2.53	0.08(0.94)	0.75(1.59)	0.29(1.52)	0.92(1.2)	0.6(1.97)	1.03(1.28)	1.21(1.85)	1.5(1.24)
3.5	0.08(1.05)	0.81(1.69)	0.42(1.89)	0.94(1.25)	0.5(1.85)	0.92(1.05)	0.96(1.7)	1.36(1.11)
4	0.06(0.71)	0.58(1.32)	0.22(1.41)	0.88(1.11)	0.46(1.92)	0.9(1.15)	1.07(1.87)	1.32(1.21)
4.5	0.09(0.1)	0.84(1.66)	0.41(1.85)	0.77(1.11)	0.37(1.63)	0.86(1.17)	1.1(1.89)	1.46(1.21)
5	0.16(1.5)	0.91(1.68)	0.45(1.91)	0.86(1.2)	0.32(1.54)	0.88(1.12)	0.89(1.84)	1.46(1.13)

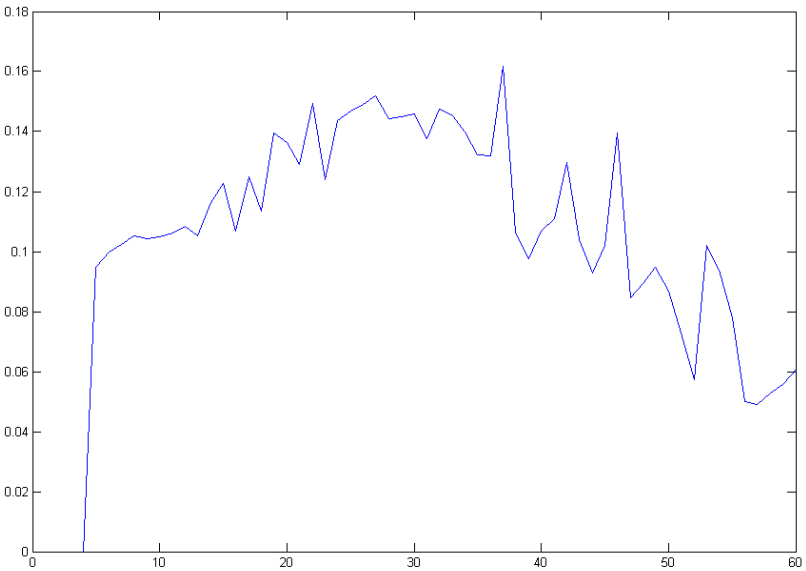
rs-fMRI Study

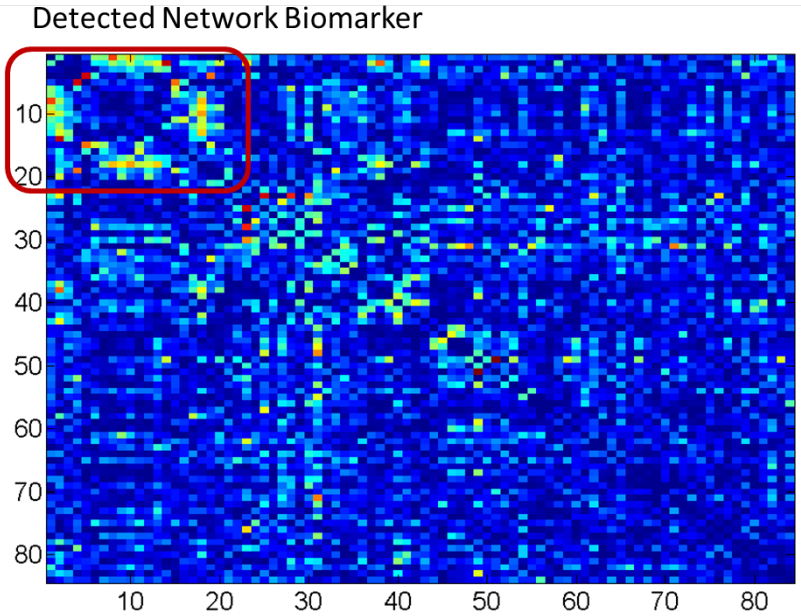
- 28 participants in the control group and 14 in the case group.
- 90 nodes lead to 4005 edges;
- Network is a subset of 4005 edges and 90 nodes;
- The connectivity biomarkers require multiple testing correction and network detection.

Input data



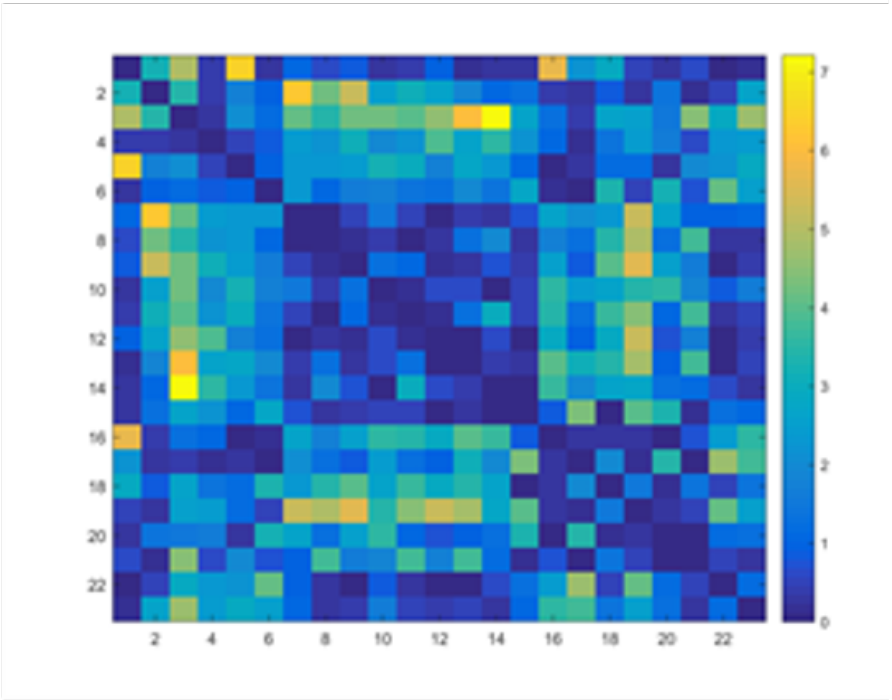
Data analysis results

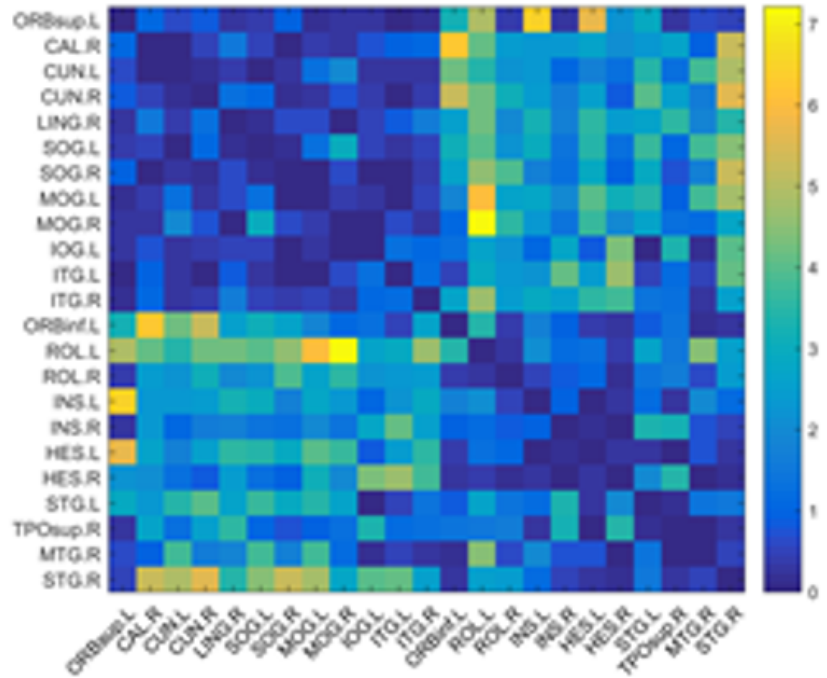




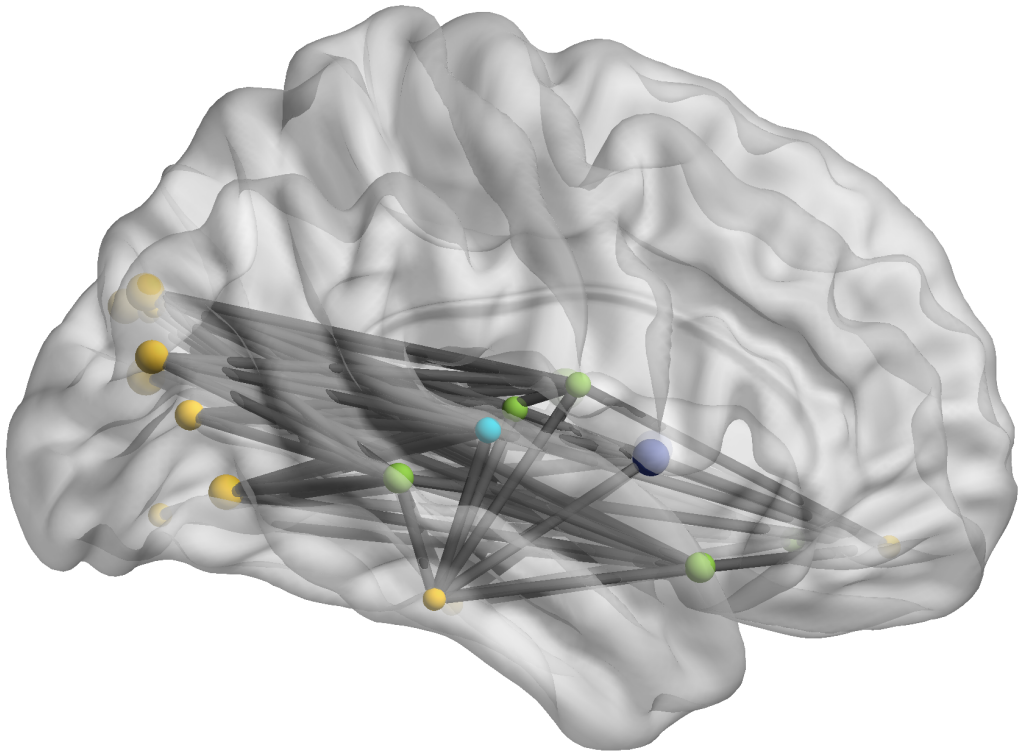
Three networks are found significant on permutation tests.

K-partite topology

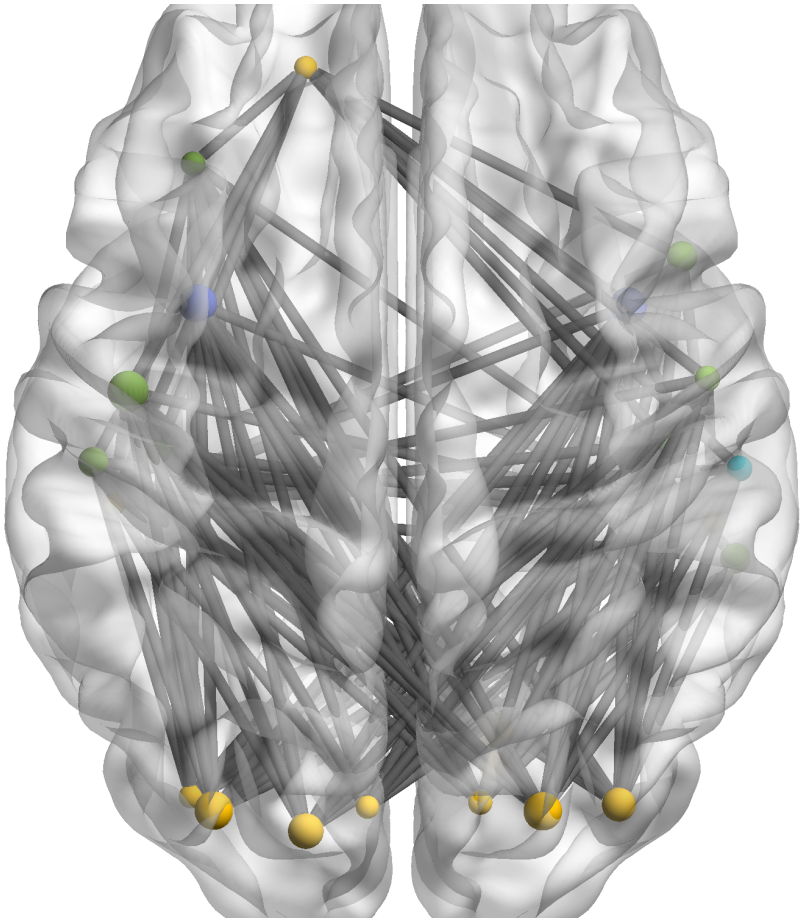




Bi-partite Topology Results (1st network)



Bi-partite Topology Results (1st network)



Conclusions and Future Work

- We developed novel strategies to detect graph topology object oriented biomarkers and inferences.
- Brain is organized and complex, we shall build statistical models considering by leveraging these properties wisely.
- The inferences fully account for the topological structure (constrained by brain anatomy), and thus can account for the edge and edge dependence structure (with lower computational load).
- Still need prove the consistency K-partite graph?
- Deep learning neural networks based on the network biomarkers.
- Applications to other high-throughput data.
- Software development.

Acknowledgements

- Yishi Xing, Ph.D student EECE, UMD
- Yang Jiao, Ph.D student EECE, UMD
- Guoqing Wang, Johns Hopkins, former UMD student
- Jian Kang, University of Michigan
- DuBois Bowman, Columbia University

Thank you!