Estimating Network Induced Large Correlation Matrix for High-dimensional Biomedical Data

Shuo Chen*

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

**email:* schen@mprc.umaryland.edu

and

Jian Kang

Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA

and

Yishi, Xing

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

and

Yunpeng, Zhao

Department. of Statistics, George Mason University, Fairfax, VA 22030, USA

and

Donald Milton

Maryland Institute for Applied Environmental Health, University of Maryland, College Park, MD 20742, USA

SUMMARY: The correlation matrix of high-dimensional biomedical data (e.g. genomics, proteomics, and neuroimaging data) often exhibits a complex and organized, yet latent graph topological structure. Estimating the correlation matrix is fundamental to understand inter-relationship between the massive variables. We propose a two step procedure that first detects the latent graph topological structure with parsimony from the sample correlation matrix and then regularizes the correlation matrix by leveraging the detected graph topological information. We show that the graph topological information guided regularization can reduce false positive and false negative rates simultaneously because it allows edges to borrow strengths from each other precisely. Several examples illustrate that the proposed latent graph topological structure widely exists across many biomedical platforms and identifying these network structures can effectively improve correlation matrix estimation and understanding the interactive relationships.

Biometrics 000, 000–000 DOI: 000

000 0000

Key words: Graph, Large correlation matrix, Network, Parsimony, Shrinkage, Thresholding, Topology.

1. Introduction

Recent advances in bio-technologies allow measuring multi-dimensional biological features simultaneously, for example, in genomics, proteomics, and neuroimaging research. The underlying biological machinery is often associated with the coordination between high-throughput features (Emilsson, 2008; Allen and Tibshirani, 2012). For a large biomedical data set $\mathbf{X}_{n\times p}$ with the sample size n and p variables, we rely on the estimation of the covariance matrix $\mathbf{\Sigma}$ or correlation matrix $\mathbf{R}$ to understand the inter-relationship between variables (Fan *et al*, 2015).

Regularization methods have been developed to estimate the high-dimensional covariance/correlation matrix. For instance, the ℓ_1 norm penalized maximum likelihood has been utilized to estimate the sparse precision matrix $\mathbf{\Theta} = \mathbf{\Sigma}^{-1}$ (Friedman *et al*, 2008; Banerjee *et al*, 2008; Yuan and Lin, 2007; Lam and Fan, 2009; Yuan, 2010; Cai and Liu, 2011; Shen *et al*, 2012) and the covariance matrix thresholding methods to directly regularize the sample covariance matrix (Bickel and Levina, 08; Rothman *et al*, 2009; Cai *et al*, 2011; Zhang, 2010; Fan *et al*, 2013; Liu *et al*, 2014). The thresholding regularization techniques have also been applied to correlation matrix $\mathbf{R}$ estimation (Qi and Sun, 2006; Liu *et al*, 2014; Cui *et al*, 2016). Mazumder and Hastie (2012) and Witten *et al* (2011) point out that the two sets of methods are naturally linked regarding vertex-partition of the whole graph and estimate of the graph edge skeleton.

Graph notations and definitions are used to describe the relationship between the p variables of $\mathbf{X}_{n\times p}$ (Yuan and Lin, 2007; Mazumder and Hastie, 2012). A finite undirected graph $G = \{V, E\}$ consists two sets, where the node/vertex set V represents variables $\mathbf{X} = (X_1, \cdots, X_p)$ with $|V| = p$ and the edge set E denotes relationships between the nodes. Let $e_{i,j}$ be the edge between nodes i and j. Then $e_{i,j}$ is an connected edge if nodes i and j are correlated in G. Under the sparsity assumption, the regularization algorithms

assign most edges as unconnected, and G may be decomposed to a set of maximal connected subgraphs (Witten *et al*, 2011; Mazumder and Hastie, 2012).

Many recent works estimate the covariance matrix by using the graph topological structural information. For example, Witten *et al* (2011), Hsieh *et al* (2012), Tan *et al* (2015) utilize the diagonal block structure and Bien *et al* (2016) use the *banding* structure to improve the estimation of the covariance matrix. When analyzing biomedical high-dimensional data sets, we note that the interactions between biological features often exhibit an interesting organized network graph topological pattern which consists a number of block/community subgraphs and a large random subgraph (see Figure 1). Therefore, we propose a new topological structure of G consisting of two components $G = G^1 \cup G^0$ where the first component $G^1 = \cup_{c=1}^{C_1} G_c$ is a stochastic block model structure and the second component $G^0 = \cup_{c=1}^{C_0} G_c^0$ (G_c^0 is a singleton) can be considered as an Erdös-Rényi random graph. We refer it as the $G^1 \cup G^0$ mixture structure. The $G^1 \cup G^0$ mixture structure is a special case of the stochastic block model, which contains many singletons and a number of communities (Bickel and Chen, 2009; Karrer and Newman, 2011; Zhao *et al*, 2011; Choi *et al*, 2012; Nadakuditi and Newman, 2012; Lei and Rinaldo, 2014). We observe the $G^1 \cup G^0$ mixture structure in many omics and imaging data though it is latent, and we demonstrate two examples among many in section 3.

However, in practice data sets are often noisy and thus conventional clustering algorithms may not easily identify the $G^1 \cup G^0$ mixture structure (Tan *et al*, 2015). We propose a new parsimonious algorithm to effectively recognize the latent $G^1 \cup G^0$ mixture structure from the sample matrix, which is robust to false positive noises (edges). Our new approach imposes a penalty term on the sizes of networks so that the $\{G_c\}$ of G^1 include as many highly correlated edges as possible while minimizing the sizes of edges in $\{G_c\}$. We assume that the (sample) false positive highly correlated edges are often distributed in a random pattern

rather in a community structure. Therefore, with the new penalty term, these false positive highly correlated edges are not likely to be included in community networks because they can great increase network sizes.

Next, we estimate the large correlation matrix based on the detected $G^1 \cup G^0$ mixture graph topological structure. Specifically, we perform thresholding for edges within and outside communities adaptively by using an Empirical Bayes approach, and the detected graph topology serves as prior knowledge. In this way, the decision of thresholding an edge is made upon both this edge's magnitude and the its *neighborhood* information via the detected graph topological structure. Our network based thresholding strategy allows edges to borrow strengths from each other (accounting for dependence) while avoiding the computationally difficult step to estimate *covariance between edges* (i.e. correlation of correlations). Different from the existing graph topology guided methods e.g. Hsieh *et al* (2012) and Tan *et al* (2015) which mainly focus on the edges within block components, we utilize information of edges from both inside and outside diagonal blocks.

We name the two step graph topology information guided regularization strategy as **N**etwork **I**nduced **C**orrelation matrix **E**stimation (NICE). The NICE method makes three contributions: i) we propose a new penalized objective function that is well-suited to estimate latent graph topological structures and robust to false positive noises ii) we fuse the graph topological information and threholding decision making procedure to simultaneously reduce false positive and false negative discovery rates; and iii) we develop computationally efficient algorithms. In addition, the detected graph topological structures may also help to reveal interesting underlying biological networks.

The paper is organized as follows. Section 2 describes the NICE algorithm. In Sections 3, we perform the simulation studies and model evaluation/comparisons and we apply our

method to a mass spectrometry proteomics data set and gene expression data. Concluding remarks are summarized in Section 4.

2. Methods

We consider the sample covariance $\mathbf{S}$ and sample correlation matrix $\widehat{\mathbf{R}}$=diag$(\mathbf{S})^{-1/2}$ $\mathbf{S}$ diag$(\mathbf{S})^{-1/2}$ as our input data (Qi and Sun, 2006; Liu *et al*, 2014; Fan *et al*, 2015). We can directly perform hard thresholding on the sample correlation matrix to estimate $\mathbf{R}$ by using $R^{\mathcal{T}}_{i,j} = \{\widehat{R}_{i,j} I(|\widehat{R}_{i,j}| > T)\}$ without exploring the underlying network structure, where T is a pre-specified or calculated threshold (Bickel and Levina, 08). However, applying the universal regularization rule (even when optimal T is provided) to each element (or column) may introduce numerous false positive and false negative findings due to various noises and measurement errors from the sample data. Therefore, we propose to leverage the information from the latent topological structure of the correlation matrix (i.e. graph G) to assist the decision making process adaptively.

The NICE method consists two steps: i) we first detect the latent topological structure of $G = G^1 \cup G^0$ mixture in G by applying the rule of parsimony; ii) we then apply empirical Bayes based thresholding to the sample correlation matrix guided by the detected graph topology.

2.1 *Parsimonious estimation of latent networks from sample correlation matrix*

We first define the weight matrix $\mathbf{W}$ based on the sample correlation matrix $\widehat{\mathbf{R}}$. An entry $w_{i,j}$ of $\mathbf{W}$ can be a transformed correlation coefficient between variables i and j that corresponds to the edge $e_{i,j}$ in G, for example, Fisher's Z transformation. $w_{i,j}$ is often a continuous metric. In the Supplementary Materials, we describe an empirical Bayes based procedure to calculate $w_{i,j}$ as a metric between 0 and 1. $\mathbf{W}$ is only used for the latent network detection rather than the regularization step.

We assume that G includes a set of community networks and many singletons as shown in Figure 1a, and edges within the networks are more likely to be connected than edges outside networks. However, in practice this topological structure is latent and the sample correlation matrix has no explicit graph topological pattern (1c). By implementing the objective function 1 we can recognize the latent graph topological structures (1d). We next perform permutation tests to evaluate the statistical significance for each G_c, and the statistically significant subgraphs $\{G_c\}$ are used to assist the estimation of the correlation matrix in the following step.

[Figure 1 about here.]

We aim to identify the latent $G^1 \cup G^0$ mixture structure from $\mathbf{W}$ by using penalized optimization. The heuristic is to identify a set of subgraphs $U = \cup_{c=1}^{C} G_c$ that maximizes the sum of weights of edges in the community networks while minimizing the community network edge sizes. The penalty term is used to avoid the disruption of false positive noises (edges). An interesting fact is that a singleton G_c contribute 0 edge and when the number of subgraphs equal to the number of nodes $C = |V|$ the sum of community network edge sizes $\cup_{c=1}^{C}|E_c| = 0$. On the contrary, $C = 1$ leads to $\cup_{c=1}^{C}|E_c| = |E|$. Therefore, C is considered to be related to the penalty term and a larger C increase the parsimony level.

Formally, we propose the objective function:

$$\underset{C,\{G_c\}_{c=1}^{C}}{\arg\max} \sum_{c=1}^{C} \exp\left[\log\left\{\sum(w_{i,j}|e_{i,j} \in G_c)\right\} - \lambda_0 \log(|E_c|)\right], \tag{1}$$

with following definitions and conditions:

1. G_c $(c = 1, \cdots, C)$ is a clique subgraph that $G_c = \{V_c, E_c\}$ and $|V_c| \geqslant 1$;
2. the size of the a subgraph G_c is determined by the number of edges $|G_c| = |E_c|$;
3. $\cup_{c=1}^{C} V_c = V$, $\cap_{c=1}^{C} V_c = \emptyset$ and $\cup_{c=1}^{C} E_c \subseteq E$.

The objective function is non-convex and difficult to be directly solved. We develop iterative algorithm to optimize C and $\{G_c\}$. In the Supplementary Materials, we provide the

detailed derivation and optimization algorithms, which links 1 to a spectral clustering related objective function (von Luxburg, 2007; Nadakuditi and Newman, 2012). By applying the penalty term, the objective function often selects a relatively large $\hat{C}$ and include many G_c as singletons to shrinkage the subgraph sizes. This objective function is well-suited to capture general graph topological structure from sample correlation matrix as it is less affected by the false positive noises by implementing the new penalty term (see more details in section 3.1). Optimizing (1) provides the estimates of underlying network topological structure within the large sample correlation matrix, which can be used to guide correlation matrix regularization.

2.2 *Graph topology oriented correlation matrix threshohlding*

To estimate the correlation matrix $\mathbf{R}$, we perform graph topology guided thresholding on the sample correlation matrix $\widehat{\mathbf{R}}$ by using Bayes factors (BF). Let $z_{i,j}$ be the Fisher's Z transformed sample correlation coefficient of $\widehat{R}_{i,j}$ and it follows a mixture distribution that $z_{i,j} \sim \pi_0 f_0(z_{i,j}) + \pi_1 f_1(z_{i,j})$ (Chen *et al*, 2016a). The distribution assumption is well supported by the example data sets in section 3.1.

Universal thresholding

Without considering prior information of the topology structure, the universal thresholding can be applied (Bickel and Levina, 08). For instance, an empirical Bayes framework implements a Bayes factor based via the (Efron, 2004, Shäfer and Strimmer, 2005). The hard-thresholding rule (Cai *et al*, 2011, Fan *et al*, 2015) is often employed for this purpose, which sets an edge to zero unless

$$\frac{\mathrm{P}(\delta_{i,j}=1|z_{i,j})}{\mathrm{P}(\delta_{i,j}=0|z_{i,j})} = \frac{f_1(z_{i,j})\pi_1}{f_0(z_{i,j})\pi_0} \geqslant T,$$

T is a constant that is linked to local fdr cutoff, and π_0 and π_1 are the proportions of null and non-null distributions correspondingly. For example, $T = 4$ is equivalent to the cutoff of local fdr of 0.2 (Efron, 2007). For instance, given $\pi_0 = 0.9$ and $\pi_1 = 0.1$, the universal

decision rule is that an edge is thresholded when Bayes factor is less than 36, i.e.,

$$BF_{i,j} = \frac{f_1(z_{i,j})}{f_0(z_{i,j})} \leqslant 36.$$

In practice, π_0 and π_1 are estimated based on the distribution of the statistics (e.g. $z_{i,j}$) and the Bayes factor cut-off is updated accordingly.

It has been well documented that the Bayes factor inferential models could adjust the multiplicity by adjusting the prior structure (Jeffreys, 1961; Kass and Raftery, 1995; Scott and Berger, 2006; Efron, 2007; Scott and Berger, 2010). The prior odds are tuned to control false positive rates, and a larger π_0 ($\pi_0 \to 1$) or a distribution of π_0 with larger mean leads to more stringent adjustment that may cause both low false positive discovery rates and high false negative discovery rates. Scott and Berger, 2006 suggest a prior distribution with median value around 0.9 and numerical methods have been developed to estimate π_0 (Wu *et al*, 2006; Efron, 2007).

However, there has been a long term challenge for all universal regularization methods (e.g. shrinkage or thresholding): the trade-off between false positive and false negative findings. In addition, edges may be dependent with each other and the mass univariate edge inference (universal regularization) ignoring the dependency structure may not estimate the large covariance and correlation matrix accurately. Yet, the direct estimation of the dependency structure between edges is challenging and sometimes not feasible. We propose one possible solution by leveraging latent graph topology to guide thresholding and account for the *dependency* between edges. The detected topological structure can seamlessly fuse into the empirical Bayes thresholding framework as prior knowledge and provides precise neighborhood information that allows edges to borrow strengths for each other.

Network based thresholding

In a network induced correlation matrix, an edge with sample correlation value $z_{i,j}$ is more likely to be truly connected within than outside a network community because the

within community 'neighbor' edges are more connected. Thus, we incorporate the topological location information of an edge into the regularization procedure. We first calculate the prior odds for edges (to be connected)within and outside community networks separately by:

$$\theta_{in} = \frac{\mathrm{P}(\delta_{i,j} = 0 | e_{i,j} \in G_c, \forall c)}{\mathrm{P}(\delta_{i,j} = 1 | e_{i,j} \in G_c, \forall c)} = \frac{\pi_0^{in}}{\pi_1^{in}},$$
$$\theta_{out} = \frac{\mathrm{P}(\delta_{i,j} = 0 | e_{i,j} \notin G_c, \forall c)}{\mathrm{P}(\delta_{i,j} = 1 | e_{i,j} \notin G_c, \forall c)} = \frac{\pi_0^{out}}{\pi_1^{out}},$$

Clearly, the within community edges are more connected by and thus $\pi_1^{in} > \pi_1 > \pi_1^{out}$ and $\pi_0^{out} > \pi_0 > \pi_0^{in}$, and $\theta_{out} \geqslant \theta_{all} \geqslant \theta_{in}$.

Let edges inside and outside of the detected communities follow distributions:

$$z_{i,j} | e_{i,j} \in G_c \sim \pi_0^{in} f_0(z_{i,j}) + \pi_1^{in} f_1(z_{i,j}) \quad \text{and}$$
$$z_{i,j} | e_{i,j} \notin G_c \sim \pi_0^{out} f_0(z_{i,j}) + \pi_1^{out} f_1(z_{i,j}),$$

respectively.

The proportions are different for the inside and outside networks, as well as overall edges, yet we assume that the null $f_0(z_{i,j})$ and non-null $f_1(z_{i,j})$ distributions are identical (the distribution assumption is well supported by examples in Section 3.1). By using the identified the latent community networks where edges are more correlated in step one, we propose the network based thresholding rule. Denote by $\widehat{R}_{i,j}^{\mathcal{T}}$ the thresholded correlation estimates for nodes i and j. Let $\widehat{BF}_{i,j} = \widehat{f}_1(z_{i,j})/\widehat{f}_0(z_{i,j})$ be an estimate of $BF_{i,j}$. Let $\widehat{\theta}_{in}$ and $\widehat{\theta}_{out}$ be the estimates of θ_{in} and θ_{out}, respectively.

If $e_{i,j} \in G_c$,

$$\widehat{R}_{i,j}^{\mathcal{T}} = \begin{cases} \widehat{R}_{i,j} & \text{if } \widehat{BF}_{i,j} \geqslant T \cdot \widehat{\theta}_{in}; \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

else if $e_{i,j} \notin G_c$,

$$\widehat{R}_{i,j}^{\mathcal{T}} = \begin{cases} \widehat{R}_{i,j} & \text{if } \widehat{BF}_{i,j} \geqslant T \cdot \widehat{\theta}_{out}; \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Equivalently, the we provide estimate of the edge set $\widehat{E}$ by using:

$$\begin{aligned} \widehat{\delta}_{i,j}^{in} &= I\left(\widehat{BF}_{i,j} \geqslant T \cdot \widehat{\theta}_{in}\right) \\ \widehat{\delta}_{i,j}^{out} &= I\left(\widehat{BF}_{i,j} \geqslant T \cdot \widehat{\theta}_{out}\right), \end{aligned} \tag{4}$$

where $\delta_{i,j}$ is an indicator variable that $\delta_{i,j} = 1$ when variables i and j are correlated with each other, otherwise $\delta_{i,j} = 0$.

We obtain $\widehat{f}_1$, $\widehat{f}_0$, $\widehat{\theta}_{in}$ and $\widehat{\theta}_{out}$ by the following steps. First, assume the edge-specific Fisher's Z transformed sample correlation coefficients in $\widehat{\mathbf{R}}$ follow a mixture distribution: $f(z_{i,j}) = \pi_0^{all} f_0(z_{i,j}) + \pi_1^{all} f_1(z_{i,j})$. $\widehat{\pi}_0^{all}, \widehat{\pi}_1^{all}$, $\widehat{f}_0$ and $\widehat{f}_1$ can be estimated by using algorithms for local *fdr* (Efron, 2007). Next, using $\widehat{f}_0$ and $\widehat{f}_1$ in the previous step we estimate $\widehat{\pi}_0^{in}$ for in-network edges $e_{i,j} \in G_c$ as the only parameter in $f^{in}(z_{i,j}) = \pi_0^{in} f_0(z_{i,j}) + \pi_1^{in} f_1(z_{i,j})$ via maximum likelihood estimation. In results, we calculate $\widehat{\theta}_{in} = \widehat{\pi}_0^{in}/\widehat{\pi}_1^{in}$ ($\widehat{\pi}_1^{in} = 1 - \widehat{\pi}_0^{in}$). For edges outside of networks ($z_{i,j}$ that $e_{i,j} \notin G_c$), we estimate $\widehat{\pi}_0^{out}$ in $f^{out}(z_{i,j}) = \pi_0^{out} f_0(z_{i,j}) + \pi_1^{out} f_1(z_{i,j})$ by following steps above, and calculate $\widehat{\theta}_{out} = \widehat{\pi}_0^{out}/\widehat{\pi}_1^{out}$ by following the same procedure.

In practice, our graph topological structure detection algorithm produces a very small odds ratio $\widehat{\theta}^{in}/\widehat{\theta}^{out}$ when the informative edges are distributed in an organized pattern. Thus, the choice of T has little impact on thresholding.

The detected graph topology provides the prior knowledge of the 'neighborhood' and 'location' for each edge.

A community network defines a neighborhood (spatial closeness) of *edges* with explicit boundaries and edges within the same neighborhood can borrow power from each other. Many statistical models are developed to account for dependency based on the neighborhood definition, for example, the Ising prior and conditional autoregressive (CAR) model (Besag

and Kooperberg, 1995). Nevertheless, unlike data in spatial or imaging statistics the sample correlation/covariance matrix of large biomedical data often include no available information about the exact spatial location or closeness on edges. Alternatively, the detected graph topological structure provides a graph topological 'closeness' of edges and thus accounts for the dependency between edges.

2.3 *Reduced false positive and negative discovery rates by using NICE thresholding*

We show that under a mild condition, the NICE method can simultaneously reduce false positive and false negative finding rates.

CONDITION 1: Let $\omega = (\sum_{c=1}^{C} |V_c| \times (|V_c| - 1)/2)/(|V| \times (|V| - 1)/2)$ the proportion of edges inside community networks and $\int_{z_0}^{\infty} f(z_{i,j}) = F(z_0)$. z_0 is the universal threshold cut-off value , $z_{0,in}$ is the within networks threshold cut-off value, $z_{0,out}$ is the within networks threshold cut-off value.

$$\frac{F_0(z_0) - F_0(z_{0,out})}{F_0(z_{0,in}) - F_0(z_0)} > \frac{\omega \pi_0^{in}}{(1-\omega)\pi_0^{out}}$$

$$\frac{F_1(z_0) - F_1(z_{0,out})}{F_1(z_{0,in}) - F_1(z_0)} < \frac{\omega \pi_1^{in}}{(1-\omega)\pi_1^{out}}.$$

This condition is generally valid for network induced correlation matrix because by implementing the parsimonious estimation of network topological structure f^{in} is distinct from f^{out} (see Figures 2 and 3). Thus, we have $\pi_0^{in} \ll \pi_0^{out}$ and $\pi_1^{in} \gg \pi_1^{out}$, and condition holds.

THEOREM 1: *Suppose Condition 1 holds, we have both 1)* $E(\sum_{i<j} I(\widehat{\delta}_{ij}^{NICE} = 1 | \delta_{ij} = 0)) \leqslant E(\sum_{i<j} I(\widehat{\delta}_{ij}^{Univ} = 1 | \delta_{ij} = 0))$ *the expected false positively thresholded edges by using the graph topology oriented thresholding (NICE) method are less than the universal thresholding method; 2)* $E(\sum_{i<j} I(\widehat{\delta}_{ij}^{NICE} = 0 | \delta_{ij} = 1)) \leqslant E(\sum_{i<j} I(\widehat{\delta}_{ij}^{Univ} = 0 | \delta_{ij} = 1))$*, the expected false negatively thresholded edges by using the graph topology oriented thresholding (NICE) method are less than the universal thresholding method.*

The proof is in the Supplementary Materials.

Uncovering the graph topological structure is important to understand the interactive relationships between multivariate variables (nodes) and the dependency between edges. We show that the detected topological structure can also provide prior knowledge to assist large covariance/correlation matrix regularization and estimation. The network based regularization approach utilizes the additional yet latent graph structure information and reduces false positive and negative discovery rates simultaneously. We summarize the overall NICE algorithm of both steps in details in the Supplementary Materials (Algorithm **??**).

3. Data Analysis

3.1 *Data examples*

We apply the NICE method to two publicly available high-dimensional biomedical data sets. By using these two examples, we show that the latent $G^1 \cup G^0$ mixture structure widely exists in data across platforms (e.g. proteomics, genomics, and imaging data, yet due to space limitation we only demonstrate two data types).

3.1.1 *Proteomics data.* The first example is matrix-assisted laser desorption ionization time of flight mass spectrometry (MALDI-TOF MS) proteomics data from human 288 subjects (Yildiz *et al*, 2007). The data assess the relative abundance of peptides/proteins in human serum. Each raw mass spectrum consists roughly 70,000 data points. After preprocessing steps including registration, wavelets denoising, alignment, peak detection, quantification, and normalization (Chen *et al*, 2009), 184 features are extracted to represent the most abundant protein and peptide features in the serum. Each feature is located at a distinct m/z value that could be linked to a specific peptide or protein with some ion charges (feature id label). The original paper utilizes the proteomics data to enhance understanding of lung

cancer pathology at the molecular level. In this paper, we focus on estimating the correlation matrix to investigate interactive relationships between these features.

We apply NICE to detect correlated peptide/protein networks and estimate correlation matrix based on the Fisher's Z transformed sample correlation matrix (Figure 2a). First, the penalized objective function 1 is implemented to capture the latent $G^1 \cup G^0$ mixture structure. The estimation results are $\hat{C} = 77$, and that seven significant community networks (G_1) are detected and the rest are singletons (G_0) (see Figure 2b). Figure 2b reorders features Figure 2a by the detected topological structure. Generally, features within networks are more correlated than features outside networks

We show that the distributions of edges inside and outside networks in Figure 2c. Clearly, f^{in} and f^{out} show distinct distributions, and f^{out} is close to the null distribution for which all edges are not connected. We estimate $\hat{\pi}_0^{all} = 0.78$, $\hat{\pi}_0^{out} = 0.83$, and $\hat{\pi}_0^{in} = 0.001$. Then, we apply the network based thresholding to estimate $\hat{E}$ and the correlation matrix. The estimate $\hat{E}$ and thresholding rule $\{\hat{\delta}_{i,j}\}$ are shown in Figure 2d. The network detection results provide informative inferences of the interactive relationship between these proteomics features. In this data example, each network represents a group of related protein and peptides that can be confirmed by proteomics mass spectrometry literature. For example, the most correlated network three consists a list of proteins of normal and variant hemoglobins with one and two charges (Lee *et al*, 2011) including normal hemoglobins α and β with one charge and two charges (at m/z 15127, 15868, 7564, and 7934). The highly correlated networks of biomedical features may provide guidance to identify a set of biomarkers for future research that allow to borrow power between each other.

[Figure 2 about here.]

3.1.2 *Gene expression data.* The second data example is gene expression profiling data based on Affy Human Genome U133A 2.0 array. The data is publicly available at Gene

Expression Omnibus (GEO) with accession code: GSE17156, GSE30550, GSE52428 and used by Dream Challenge (see more details at https://www.synapse.org/#!Synapse:syn5647810/wiki/399110). Blood samples were collected for 110 healthy controls at baseline. We focus on 1924 gene expression features that are commonly observed in human blood, and normalized data is used for analysis. The input data for our model is a 1924×1924 sample correlation matrix (Figure 3a). The sample correlation matrix show no explicit organized topological structures. By applying the penalized objective function in 1, we identify the latent $G^1 \cup G^0$ mixture structure (Figure 3b). Note that Figure 3b is a isomorphic graph to Figure 3a with reordered nodes. With $\hat{C} = 613$, four large networks and a long list of singletons and small networks (with 2 or 3 nodes) are detected because of the penalty term.

Figure 3c shows that edges inside and outside community networks follow distinct distributions. We estimate $\hat{\pi}_0^{all} = 0.84$, $\hat{\pi}_0^{out} = 0.99$, and $\hat{\pi}_0^{in} = 0.05$. The distribution of edges outside of community networks is also close to the null distribution of non-connected edges, whereas the distribution of edges inside networks again centers around 0.5. By applying the network guided thresholding, we obtain the estimated correlation matrix and $\hat{E}$ as shown in figure 3d.

Interestingly, the latent $G^1 \cup G^0$ mixture structure shows in both data examples, which can also be identified in large data from many other platforms including neuroimaging activation and connectivity data, DNA methylation data, and etc. (Chen *et al*, 2016a). In addition, for most of these data sets the inside and outside network edge distributions tend to be distinct with f^{out} close to the null distribution and f^{in} centers around 0.5. This further verifies that our assumptions of mixture distribution and condition 1 are generally valid.

In comparison, when we apply existing methods (e.g. *glasso*), the latent $G^1 \cup G^0$ mixture structure can not identified based on the estimated covariance or inverse covariance matrix.

Many inside network edges are (false negatively) regularized to zero (see Supplementary Materials).

[Figure 3 about here.]

3.2 *Simulation Studies*

We conduct numerical studies to evaluate the performance of our approach, and compare it with several existing methods.

3.2.1 *Synthetic data sets.* We simulate each data set with $p = 100$ variables, and thus $|V| = 100$ and $|E| = \binom{100}{2} = 4950$. We assume that the correlation matrix includes two community networks, and the first include 15 nodes and the second 10 nodes. The induced networks are complete subgraphs (cliques) that all edges are connected within these two networks and no other edges are connected outside the two networks (Figure 1a). Next, we permute the order of the nodes to mimic the practical data where the topological structure is latent. Figure 1b represents the connected edges in the matrix. Let vector $\mathbf{x}^k_{p\times 1}$ follow a multivariate normal distribution, with zero mean and covariance matrix $\mathbf{\Sigma}_{p\times p}$, and the sample size is n. $\sigma_{i,j}$ is an entry at the ith row and jth column of $\mathbf{\Sigma}$, $\sigma_{i,j} = 1$ if $i = j$ (then $\mathbf{\Sigma} = \mathbf{R}$), and $\sigma_{i,j} = \rho$ if $e_{i,j} \in G_c$ (inside network edges) and $\sigma_{i,j} = 0$ when $e_{i,j} \notin G_c$ (outside network edges). We simulate 100 data sets at three different settings with different levels of signal to noise ratio (SNR) by using various sample sizes n and values of ρ. A larger sample size reduces the asymptotic variance of $\widehat{\sigma}_{i,j}$ and thus the noise level is lower; and a higher absolute value of ρ represents higher signal level. A higher SNR leads to more distinct empirical distributions of $\widehat{\sigma}_{i,j}$ between inside network edges $e_{i,j} \in G_c$ and outside network edges $e_{i,j} \notin G_c$. Figure 1c demonstrates a calculated correlation matrix based on a simulated data set.

In our simulated data sets, 150 edges are connected and 4800 edges are unconnected,

which together represent the graph edge skeleton E. For both precision matrix shrinkage and covariance matrix thresholding methods we treat a non-zero entry $\widehat{\delta}_{i,j} = 1$ (Mazumder and Hastie, 2012) as a connected edge. We summarize the false positive (FP) edges $\widehat{\delta}_{i,j} = 1$ when the edge is not connected and $e_{i,j} \notin G_c$ and false negative (FN) edges $\widehat{\delta}_{i,j} = 0$ when the edge is connected and $e_{i,j} \in G_c$. We compare FN and FP counts of each method by contrasting the estimated $\widehat{E}$ with the truth E. We compare our method with universal thresholding (Thresh), *glasso*, el_1 minimization for inverse matrix estimation (CLIME), and adaptive thresholding (AThres) by comparing the FP and FN edges of estimating the graph edge skeleton E (Bickel and Levina, 08; Friedman *et al*, 2008; Cai *et al*, 2011; Cai and Liu, 2011).

[Table 1 about here.]

3.2.2 *Simulation Study Results.* The simulation results are summarized in Table 1. Rather than selecting a single tuning parameter λ for glasso and other methods by cross-validation, we explore all possible choices within a reasonable range and use the one with best performance for comparison. Cross the 100 simulation data sets, we summarize the 25%, 50%, and 75% quantiles of the number of FP and FN edges to assess the performance of each method. The results show that the NICE algorithm outperforms the competing methods even when optimal tuning parameters are used (after comparing with the truth) for these methods. One possible reason could be the NICE algorithm thresholds the correlation matrix based on the topological structure rather than the a universal shrinkage or thresholding strategy. More importantly, our approach is the only method can automatically detect the underlying $G^1 \cup G^0$ mixture topological structure. When the graph topological structure does not exist, the performance of all methods are similar across all settings. The matrix norm loss is not compared, because the community networks are small in size and norm comparison are likely determined by the false positive edges outside network communities. We note that the

methods with sparsity assumption (e.g. *glasso* and CLIME) may miss many connected edges (false negative discovery rates are higher) even when small tuning parameter is used (false positive rates are high). Therefore, when a latent topological structure exists the sparsity assumption may not be valid because a cluster of features within a network are all correlated with each other and many of them can be conditionally independent.

In summary, the numerical results demonstrate that our new method not only provides more accurate estimation of the correlation matrix and the edge set E than the competing method, but also automatically detects the community networks where highly correlated edges distribute in an organized fashion.

4. Discussion and Conclusion

We develop the NICE approach to bridge the correlation matrix estimation and graph topological structure detection via a flexible empirical Bayesian framework. Recognizing the latent network topological structure can not only reveal underlying biological pathways, but also guide the decision making procedure of regularization.

The latent $G^1 \cup G^0$ mixture graph structure exist widely in high-throughput biomedical data across platforms, however, the conventional network detection and clustering algorithms may not detect it due to the impact of false positive noises. For instance, a few false positive edges may lead to detecting a large networks with low proportion of highly connected edges. The proposed penalized network estimation objective function can identify the mixture structure because it is less affected by false positive noises. Interestingly, we find that the number of networks is related to the penalty term because a larger C generate many singletons. Efficient optimization algorithms of the penalized objective function is developed. The computational cost of NICE algorithm is low (for our simulation example the algorithm takes about 40 seconds by using a i7 CPU and 24G memory desktop), and thus it is ready to scale up for larger data sets.

More complex subgraph topological structure of G may exist implicitly within the sample correlation matrix (e.g. multipartite graph in Chen *et al*, 2016b). The detected organized subnetworks (with more complex graph topological structures) can increase the objective function 1 as the quality term increases and quantity term is almost unchanged. Therefore, the refined graph topological structure detection algorithms, for instance, K-partite, rich-club, and overlapped subgraphs could further assist to optimize the objective function, which are compatible with the objective function 1. In future works, under the same framework more graph topological structure automatic detection tools will be developed, tested, applied when needed.

The new Bayes factor based thresholding approach naturally incorporates detected network topological structure from step one as prior knowledge. The updated thresholding values are determined by each edge's 'location' on the detected graph topological 'map'. Therefore, edges can borrow strengths with each other with higher precision based on detected topological structure, which also provides a flexible pathway to account for the dependency between edges. With additional information from the detected topological structure and appropriate modeling strategy, our new thresholding approach reduces false positive and false negative rates simultaneously when topological structures exist. Clearly, the performance of graph topological structure detection influences the accuracy of correlation matrix thresholding because it determines the empirical distributions of $z_{i,j}^{in}$ and $z_{i,j}^{out}$ and thus $\widehat{\theta}_{in}$ and $\widehat{\theta}_{out}$. Therefore, the two steps of the NICE algorithm are seamlessly connected as the parsimonious property of the network detection ensures the efficiency and accuracy of the following regularization step. Edges outside networks are subject to more stringent thresholds whereas edges inside networks are less likely to be thresholded. This decision rule is data-oriented and determined by the latent spatial distributions of edges in the sample correlation matrix.

In our application, only positive (correlation) edges are distributed in an organized graph

topology and the negative (correlation) edges are randomly distributed. Based on the network based thresholding, negative (correlation) edges are thresholded. Our methods are ready to be extended to the scenario that negatively correlated edges show a organized topological structure. The numerical studies and example data application have demonstrated excellent performance of the NICE algorithm regarding false positive/negative findings and latent network detection.

In addition, the NICE algorithm does not require that the large biomedical data follows a multivariate Gaussian distribution. It is straightforward to extend the sample correlation matrix to other sample metrics, for example maximal information coefficients (Kinney and Atwal, 2014) for continuous data and polychoric correlation coefficients for categorical data (Bonett and Price, 2005) because graph topology oriented thresholding are based on the empirical distribution of the coefficients.

Acknowledgements

The research is based upon work supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via DJF-15-1200-K-0001725.

References

Allen, G. I., & Tibshirani, R. (2012). Inference with transposable data: modelling the effects of row and column correlations. Journal of the Royal Statistical Society: Series B, 74(4), 721-743.

Banerjee, O., El Ghaoui, L. and d'Aspremont, A. (2008). Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. *The Journal of Machine Learning Research* **9**, 485-516.

Besag, J., Kooperberg, C. (1995). On conditional and intrinsic autoregressions. *Biometrika*, **82**(4), 733-746.

Bickel, P.J., Levina, E. (2008). Covariance regularization by thresholding. *Ann. Statist.* **36**, no. 6, 2577–2604.

Bickel, P. J., Chen, A. (2009). A nonparametric view of network models and NewmanGirvan and other modularities. *Proceedings of the National Academy of Sciences*, **106**(50), 21068-21073.

Bien, J., Bunea, F., Xiao, L. (2016). Convex banding of the covariance matrix. Journal of the American Statistical Association, 111(514), 834-845.

Bonett, D. G., Price R. M. (2005). Inferential Methods for the Tetrachoric Correlation Coefficient. *Journal of Educational and Behavioral Statistics*, **30**, 213.

Cai, T., Liu, W. and Luo, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, **106**, 594–607.

Cai, T., Liu, W. (2011). Adaptive thresholding for sparse covariance matrix estimation. *J. Amer. Statist. Assoc.* **106** (494), 672–684.

Cai, T. T., Ren, Z., Zhou, H. H. (2014). Estimating structured high-dimensional covariance and precision matrices: Optimal rates and adaptive estimation. *The Annals of Statistics*, **38**, 2118-2144.

Chen, S., Li, M., Hong, D., Billheimer, D., Li, H., Xu, B. J., Shyr, Y. (2009). A novel comprehensive wave-form MS data processing method. *Bioinformatics*, **25**(6), 808-814.

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.

Chen, S., Bowman, F. D., & Mayberg, H. S. (2016). A Bayesian hierarchical framework for

modeling brain connectivity for neuroimaging data. *Biometrics*, 72(2), 596-605.

Chen, S., Bowman, F. D., Xing, Y. (2016). Differentially Expressed Functional Connectivity Networks with K-partite Graph Topology. arXiv preprint arXiv:1603.07211.

Choi, D. S., Wolfe, P. J., Airoldi, E. M. (2012). Stochastic blockmodels with a growing number of classes. *Biometrika*, **99**, 273-284.

Chung, F. R. (1997). Spectral Graph Theory (CBMS Regional Conference Series in Mathematics, No. 92), American Mathematical Society.

Cui, Y., Leng, C., Sun, D. (2016). Sparse estimation of high-dimensional correlation matrices. *Computational Statistics & Data Analysis*, **93**, 390-403.

Efron, B. (2004). Large-Scale Simultaneous Hypothesis Testing: The Choice of a Null Hypothesis. *Journal of the American Statistical Association*, **99**, 96-104.

Efron, B. (2007). Size, power and false discovery rates. *The Annals of Statistics*, **35**(4), 1351-1377.

Efron, B., Turnbull, B., Narasimhan, B. (2008). locfdr: Computes local false discovery rates. *R package*, 195.

El Karoui, N. (2010). High-dimensionality effects in the markowitz problem and other quadratic programs with linear constraints: risk underestimation. *The Annals of Statistics*, **38**, 3487–3566.

Emilsson, V., Thorleifsson, G., Zhang, B., Leonardson, A. S., Zink, F., Zhu, J., ... & Mouy, M. (2008). Genetics of gene expression and its effect on disease. Nature, **452**(7186), 423.

Fan, J., Liao, Y., Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. With 33 discussions by 57 authors and a reply by Fan, Liao and Mincheva. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75**, no. 4, 603–680.

Fan, J., Liao, Y., Liu, H. (2015). Estimating Large Covariance and Precision Matrices. arXiv preprint arXiv:1504.02995.

Friedman, J., Hastie, T., Tibshirani, R. Sparse inverse covariance estimation with the graphical lasso. (2008).*Biostat.* **9**(3), 432–441.

Friedman, J., Hastie, T., Tibshirani, R. (2010). Applications of the lasso and grouped lasso to the estimation of sparse graphical models (pp. 1-22). Technical report, Stanford University.

Hsieh, C.J., Banerjee, A., Dhillon, I.S., Ravikumar, P.K.(2012) A divide-and-conquer method for sparse inverse covariance estimation *Advances in Neural Information Processing Systems* , 2330-2338

Lam, C. and Fan, J. (2009). Sparsistency and rates of convergence in large covariance matrix estimation. *Annals of statistics* **37** 42-54.

Lee, B. S., Jayathilaka, G. L. P., Huang, J. S., Vida, L. N., Honig, G. R., Gupta, S. (2011). Analyses of in vitro nonenzymatic glycation of normal and variant hemoglobins by MALDI-TOF mass spectrometry. *Journal of biomolecular techniques: JBT*, **22**(3), 90.

Lei, J., Rinaldo, A. (2014). Consistency of spectral clustering in stochastic block models. *The Annals of Statistics*, **43**(1), 215-237.

Liu, H., Wang, L. and Zhao, T. (2014). Sparse covariance matrix estimation with eigenvalue constraints. *Journal of Computational and Graphical Statistics*, **23**, 439-459.

Jeffreys, H. (1961). Theory of Probability, 3rd ed. Clarendon Press, Oxford.

Kass, R. E., Raftery, A. E. (1995). Bayes factors. *Journal of the american statistical association*, **90**(430), 773-795.

Karrer, B., Newman, M. E. (2011). Stochastic block models and community structure in networks. *Physical Review E*, **83**(1), 016107.

Kinney, J. B., Atwal, G. S. (2014). Equitability, mutual information, and the maximal information coefficient. *Proceedings of the National Academy of Sciences*, **111**(9), 3354-3359.

Mazumder, R. and Hastie, T. (2012). Exact covariance thresholding into connected components for large-scale graphical lasso. *The Journal of Machine Learning Research*, **13**(1), 781-794.

Nadakuditi, R. R., Newman, M. E. (2012). Graph spectra and the detectability of community structure in networks. *Physical review letters*, **108**(18), 188701.

Rothman, A. J., Levina, E. and Zhu, J. (2009). Generalized thresholding of large covariance matrices. *Journal of the American Statistical Association* **104**, 177-186.

Qi, H. and Sun, D. (2006). A quadratically convergent newton method for computing the nearest correlation matrix. S*IAM journal on matrix analysis and applications* **28** 360-385.

Scott, J. G., Berger, J. O. (2006). An exploration of aspects of Bayesian multiple testing. *Journal of Statistical Planning and Inference*, **136**(7), 2144-2162.

Scott, J. G., Berger, J. O. (2010). Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem. *The Annals of Statistics*, **38**(5), 2587-2619.

Schäfer, J., Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical applications in genetics and molecular biology*, **4**(1).

Shen, X., Pan, W. and Zhu, Y. (2012). Likelihood-based selection and sharp parameter estimation. *Journal of the American Statistical Association* **1**07 223232.

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

von Luxburg, U. A tutorial on spectral clustering. (2007), *Stat. Comput.* **17** (4), 395–416.

Tan, K. M., Witten, D., Shojaie, A. (2015). The cluster graphical lasso for improved estimation of Gaussian graphical models. *Computational statistics & data analysis*, 85, 23-36.

Witten, D. M., Friedman, J. H., Simon, N. (2011). New insights and faster computations for the graphical lasso. *Journal of Computational and Graphical Statistics*, **20**(4), 892-900.

Wu, B., Guan, Z., Zhao, H. (2006). Parametric and nonparametric FDR estimation revisited. *Biometrics*, **62**(3), 735-744.

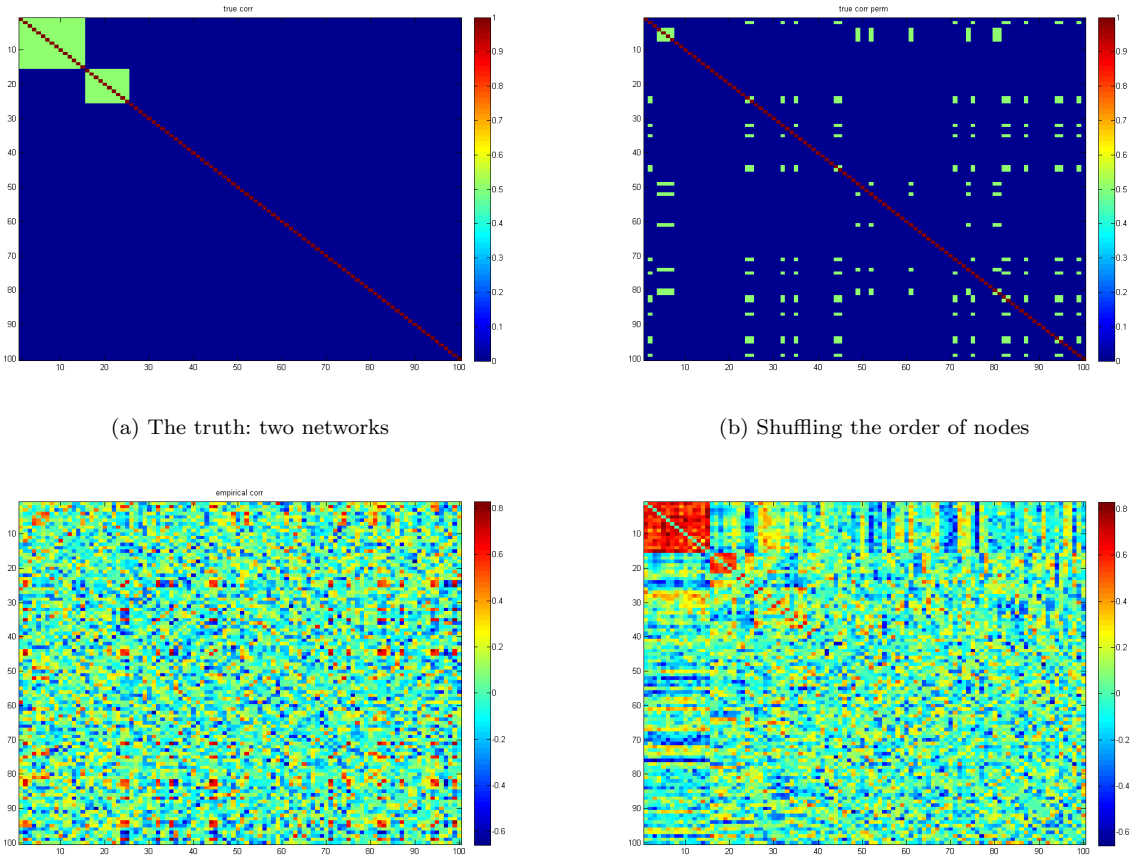
Yildiz, P. B., Shyr, Y., Rahman, J. S., Wardwell, N. R., Zimmerman, L. J., Shakhtour, B., ... Massion, P. P. (2007). Diagnostic accuracy of MALDI mass spectrometric analysis of unfractionated serum in lung cancer. *Journal of thoracic oncology*, **2**(10), 893-915.

Yuan, M. and Lin, Y. (2007). Model selection and estimation in the gaussian graphical model. *Biometrika*, **94**, 19–35.

Yuan, M. (2010). High dimensional inverse covariance matrix estimation via linear programming. *Journal of Machine Learning Research*, **11**, 2261-2286.

Zhang, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, 894942.

Zhao, Y., Levina, E., Zhu, J. (2011). Community extraction for social networks. *Proceedings of the National Academy of Sciences*, **108**(18), 7321-7326.



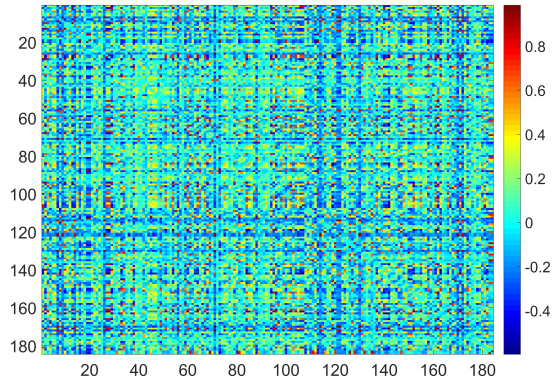
(a) The truth: two networks

(b) Shuffling the order of nodes

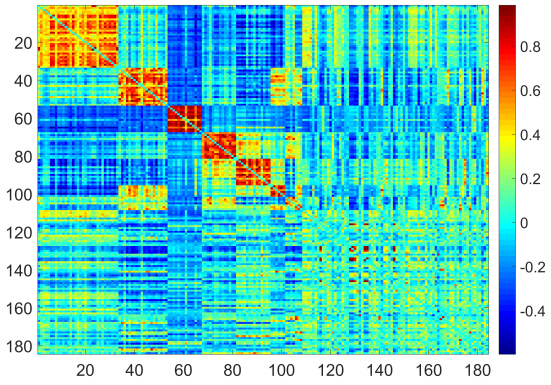
(c) The input data for NICE

(d) Network detection results

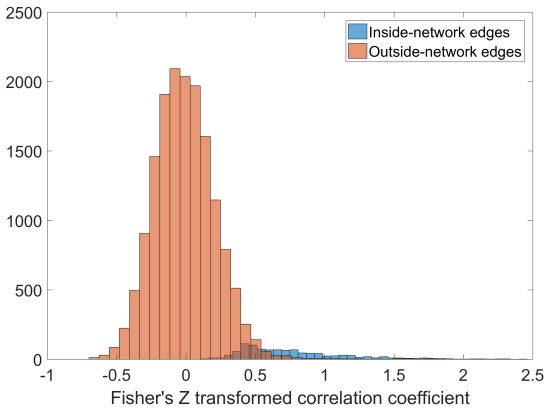
Figure 1: An example of a network induced correlation matrix: $|V|$=100 nodes and $|E|$=4950 edges, there are two networks (a) and in practice they are implicit (b); it may be difficult to recognize the the latent $G^1 \cup G^0$ mixture structure when looking at the sample correlation matrix (c); the proposed objective function is robust to false positive noise and identify the latent $G^1 \cup G^0$ mixture structure from the sample correlation matrix.



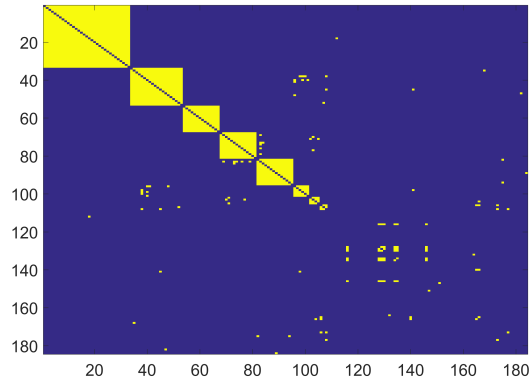
(a) Sample correlation



(b) Detecting latent $G^1 \cup G^0$ mixture structure

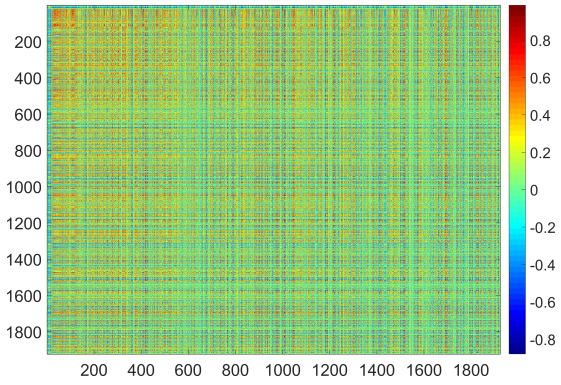


(c) Edges inside and outside networks

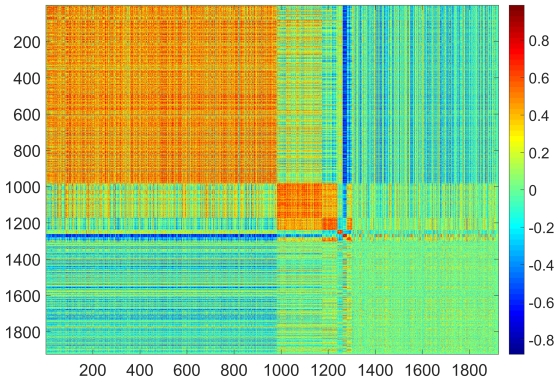


(d) Estimated edge set $\hat{\mathbf{E}}$

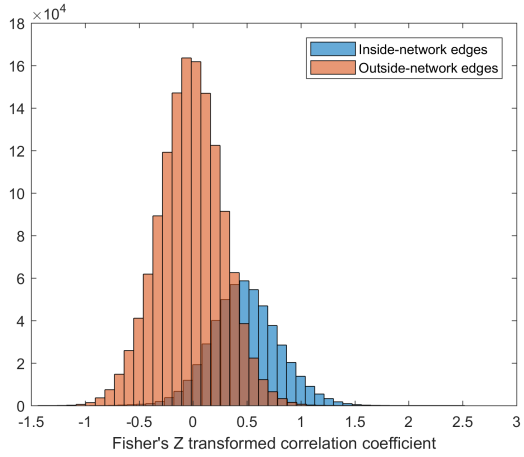
Figure 2: Application of the NICE to the example data set one. (a) is the heatmap of sample correlation matrix; (b) demonstrates the latent $G^1 \cup G^0$ mixture structure by reordering the variables in the heatmap; (c) shows the distributions of edges inside and outside the networks; (d) is the estimated $\hat{E}$ based on the NICE thresholding.



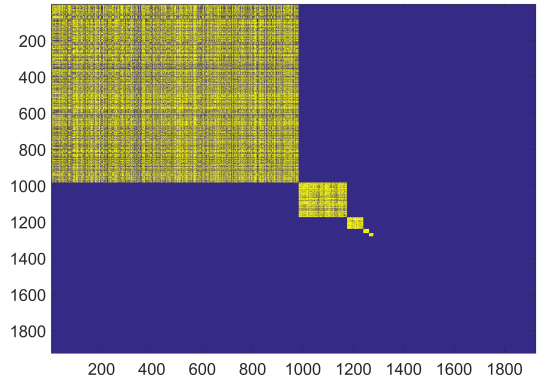
(a) Sample correlation



(b) Weight matrix $\mathbf{W}$



(c) Edges inside and outside networks



(d) Estimated edge set $\hat{\mathbf{E}}$

Figure 3: Application of the NICE to the example data set two. (a) is the sample correlation matrix; (b) demonstrates the latent $G^1 \cup G^0$ mixture structure by reordering the variables in the heatmap; (c) shows the distributions of edges inside and outside the networks; (d) is the estimated $\hat{E}$ based on the NICE thresholding.

Table 1: Median along with 25% and 75% quantiles of FP and FN

Method	Tuning Par.	$\sigma = 0.5$, $n = 25$ FP Med.	FP Quantiles	FN Med.	FN Quantiles	$\sigma = 0.5$, $n = 50$ FP Med.	FP Quantiles	FN Med.	FN Quantiles	$\sigma = 0.7$, $n = 25$ FP Med.	FP Quantiles	FN Med.	FN Quantiles
glasso	0.1	1673	(1648, 1702)	59	(55, 64)	1621	(1591.5, 1640)	44	(40, 46)	1581.5	(1557, 1606)	45.5	(42, 48)
	0.2	1008.5	(989, 1025)	59	(53.5, 64.5)	630	(610, 644)	38	(33.5, 43)	932.5	(920, 955.5)	36	(32, 40)
	0.3	546	(529.5, 560)	56	(48, 63.5)	151	(141, 162.5)	38	(30.5, 43)	500.5	(490, 516)	28	(23.5, 33)
	0.4	211.5	(200.5, 222.5)	60	(50.5, 72)	19	(16, 21)	48.5	(38, 58)	194	(186, 204.5)	24.5	(20, 29)
	0.5	51	(46, 59)	80.5	(66, 96)	1	(0, 2)	82.5	(67, 96.5)	47	(41.5, 54)	28	(22.5, 35)
	0.6	7	(5, 10)	112.5	(97, 125.5)	0	(0, 0)	130	(118.5, 137)	6	(5, 8.5)	41	(31, 51)
	0.7	0	(0, 1)	140	(131.5, 146)	0	(0, 0)	149	(147, 150)	0	(0, 1)	75	(61.5, 89)
	0.8	0	(0, 0)	149	(148, 150)	0	(0, 0)	150	(150, 150)	0	(0, 0)	127	(119.5, 135)
	0.9	0	(0, 0)	150	(150, 150)	0	(0, 0)	150	(150, 150)	0	(0, 0)	149	(149, 150)
	1.0	0	(0, 0)	150	(150, 150)	0	(0, 0)	150	(150, 150)	0	(0, 0)	150	(150, 150)
CLIME	0.1	1082.5	(1047.5, 1108)	56	(48, 64.5)	993.5	(981, 1024)	39	(32, 45.5)	1054	(1021, 1079)	48.5	(40, 56)
	0.2	353	(339.5, 367.5)	79.5	(69, 87.5)	241.5	(231.5, 251.5)	61	(54, 67.5)	345	(328, 359)	70	(59, 78)
	0.3	63	(57, 69)	110	(98.5, 115)	25	(22, 29)	92	(84.5, 100)	64	(59, 68)	98	(87, 103)
	0.4	0	(0, 1)	140	(135, 144)	0	(0, 0)	130	(124, 135)	0	(0, 1)	134	(129, 139)
	0.5	0	(0, 0)	150	(150, 150)	0	(0, 0)	150	(150, 150)	0	(0, 0)	150	(150, 150)
Thres	0.1	2017.5	(1963.5, 2067.5)	0	(0, 2)	1978.5	(1944.5, 2021.5)	0	(0, 0)	2021.5	(1968.5, 2061)	0	(0, 1)
	0.3	1292.50	(1252, 1331)	2	(0, 5)	1249.5	(1220.5, 1288.5)	0	(0, 0)	1293.5	(1251, 1341.5)	1	(0, 3)
	0.5	721.5	(699, 752)	5	(1, 12)	689	(673.5, 721)	0	(0, 1)	722	(693, 756)	3	(1, 10.5)
	0.7	344.5	(325, 360)	14	(7, 26.5)	328.5	(311.5, 349.5)	1	(0, 2)	342.5	(324, 363)	10	(3, 21.5)
	0.9	132	(121, 143.5)	30	(18, 45)	129.5	(121, 142)	3	(1, 7)	133	(123.5, 146)	24	(12, 39.5)
	1.1	41.5	(35, 46)	55.5	(40, 78.5)	40.5	(36.5, 47.5)	10	(4.5, 17)	40	(35.5, 46.5)	49.5	(28, 63)
	1.3	9	(6, 10)	92	(74, 112)	10	(8, 12)	25	(13, 37)	9	(6, 11)	78	(54.5, 89)
	1.5	1	(0, 2)	126	(112.5, 137)	2	(1, 3)	50.5	(32.5, 68)	1	(0, 2)	106	(92.5, 114)
	1.7	0	(0, 0)	145	(138.5, 148)	0	(0, 0)	85.5	(67, 102.5)	0	(0, 0)	132.5	(120.5, 138.5)
	1.9	0	(0, 0)	150	(149, 150)	0	(0, 0)	120.5	(105, 130)	0	(0, 0)	147	(144, 149)
AThres	0.3	2593	(2566.5, 2627.5)	2	(0, 5)	2538.5	(2509.5, 2571)	0	(0, 0)	2594	(2563, 2619)	1	(0, 3)
	0.5	1460	(1421.5, 1486)	5	(1, 12)	1412.5	(1379.5, 1440)	0	(0, 1)	1453	(1419.5, 1491)	3	(1, 10.5)
	0.7	691.5	(667, 717)	14	(7, 26.5)	668.5	(646, 697)	1	(0, 2)	695.5	(665.5, 720)	10	(3, 21.5)
	0.9	271.5	(258, 291.5)	30	(18, 45)	265	(252, 283.5)	3	(1, 7)	270.5	(255.5, 288)	24	(12, 39.5)
	1.1	83	(75, 95)	55.5	(40, 78.5)	85	(75.5, 95.5)	10	(4.5, 17)	82	(74, 89.5)	49.5	(28, 63)
	1.3	18	(15, 21)	92	(74, 112)	22	(18.5, 25.5)	25	(13, 37)	18	(14.5, 22)	78	(54.5, 89)
	1.5	2	(1, 4)	126	(112.5, 137)	4	(3, 6)	50.5	(32.5, 68)	3	(1, 3)	106	(92.5, 114)
	1.7	0	(0, 0)	145	(138.5, 148)	0	(0, 1)	85.5	(67, 102.5)	0	(0, 0)	132.5	(120.5, 138.5)
	1.9	0	(0, 1)	150	(149, 150)	0	(0, 0)	120.5	(105, 130)	0	(0, 0)	147	(144, 149)
NICE	None	44	(15, 98)	3	(0, 27)	11	(1, 30)	0	(0, 4)	32.5	(13.5, 71)	14	(4, 38.5)