Selection and Estimation of Conditional Graphical Models

Stephen Salerno, MS and Yi Li, PhD
Department of Biostastics · University of Michigan

March 19, 2020

Outline

Background and Motivation

Model

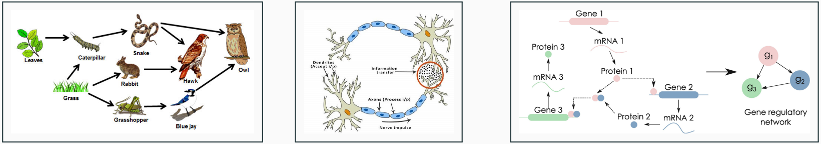
Algorithm

Conclusions

Background and Motivation

Biological Networks

Network-based structures appear everywhere in nature throughout various **biological systems**:



(a) Ecological **(b)** Neural **(c)** Gene Co-Regulation

Figure 1: Example biological networks.

(a) https://link.springer.com/protocol/10.1007/978-1-4939-8882-2_1;//

(b) https://medium.com/predict/artificial-neural-networks-mapping-the-human-brain-2e0bd4a93160;//

(c) https://ontrack-media.net/gateway/science7/g_s7m1l2s3.html

Biological Networks

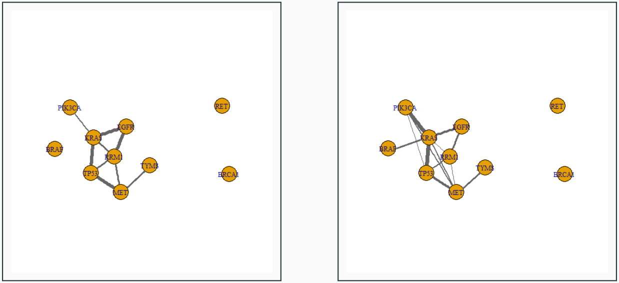
Modeling biological networks provides a **mathematical representation** of the **unit-to-unit connections** in these systems[1]

These network connections may **differ** by **important factors**

In particular, gene co-regulated networks may differ by **individual DNA profiles**[2–4] or characteristics such as **sex**[5–9]

Boston Lung Cancer Survival Cohort Example

Figure 2 shows that the network structure of lung cancer-related genes depends on the heterozygous or homozygous status of SNP chr1:1792215.



(a) Homozygous Subgroup

(b) Heterozygous Subgroup

Figure 2: Networks with eight lung cancer genes differing by SNP chr1:1792215 genotype among Boston Lung Cancer Survival Cohort (BLCSC) study patients.

Graphical Models

Graphical models provide a means of quantifying the relationship between nodes of a network through **conditional co-dependence** (edges)

- Let $\boldsymbol{X} = (X_1, \ldots, X_p)$ be a p-dimensional random vector
- The tuple $\mathcal{G}_{\boldsymbol{X}} = \{\mathcal{G},\ \mathcal{P}(\boldsymbol{X})\}$ defines a graphical model for $\boldsymbol{X}$ where $\mathcal{G}$ is a graph and $\mathcal{P}(\boldsymbol{X})$ is a given probability

An edge between two nodes is defined by a **non-zero partial correlation**

Conditional Gaussian Graphical Models

The **Gaussian distribution** is a natural choice for jointly modeling conditional independences, encoded by $\mathcal{G}$, for continuous outcomes

Conditional Gaussian graphical models (CGGMs) reparametrize the multivariate linear regression model to **explicitly exhibit**:[10]

- Partial correlations between predictors and responses
- Partial correlations among responses

Current CGGM Formulations

For observed $\{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i=1}^n$ where $\boldsymbol{x}_i$ is a p-vector of predictors and $\boldsymbol{y}_i$ is a q-vector of responses, CGGMs **currently** take the form:

$$\boldsymbol{y}_i = \boldsymbol{A}'\boldsymbol{x}_i + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}), \quad \forall i = 1, \ldots, n$$

where $\boldsymbol{A}$ is a $p \times q$ matrix of regression coefficients, $\boldsymbol{\Psi}$ is a $q \times q$ covariance matrix of Gaussian noise, and $\mathcal{N}(\cdot)$ denotes the Normal distribution

Existing CGGM Literature

Existing CGGMs typically only allow the mean structure, not the network structures, to vary with predictors[11–13]

- Precision matrix, $\boldsymbol{\Psi}^{-1}$, specifies a 'covariate-adjusted' Gaussian graph
- The conditional dependence structure is estimated after taking into account confounding effects on the mean structure
- Regression coefficients $\boldsymbol{A}$ relate to associations in the mean (e.g. gene expression level) structure

A Baysian Gaussian Graphical Model

A related work in the context of Bayesian **directed acyclic graphs** allows **network structures**, not means, to vary with predictors[14]

- Models the conditional independence function as the product of a smooth function and a thresholding function
- Accounts for functional nonlinearity in edge-covariate relationships
- Allows the structure of the graph to vary with multiple covariates

Our Proposal

We propose a **new class** of conditional (undirected) graphical models, where **both** the mean and network structures depend on covariates

- Jointly model the mean and covariance functions given covariates
- Parsimonious representation of these sources of variation
- Accomodates low- and high-dimensional settings

Model

Notation and Assumptions

Let $\boldsymbol{x} = (x_1, \ldots, x_p)^T$ be an observed p-vector of covariates and $\boldsymbol{y} = (y_1, \ldots, y_q)^T$ be an observed q-vector of outcomes. We assume:

$$\boldsymbol{y}|\boldsymbol{x} \sim \mathcal{N}_q\left(\boldsymbol{\mu}(\boldsymbol{x}), \boldsymbol{\Theta}^{-1}(\boldsymbol{x})\right) \tag{1}$$

where $\boldsymbol{\mu}(\boldsymbol{x})$ is a q-dimensional mean vector and $\boldsymbol{\Theta}(\boldsymbol{x})$ is a $q \times q$ positive-definite precision matrix, both of which depend on $\boldsymbol{x}$, and $\mathcal{N}_q(\cdot)$ denotes the q-dimensional multivariate Normal distribution

Parameterizing the Mean and Covariance

We seek to achieve a **parsimonious**, interpretable representation of the mean and covariance structures:

- We parameterize the mean vector by $\boldsymbol{\mu}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$, where $\boldsymbol{A}$ is a $q \times p$ regression coefficient matrix
- We further parameterize $\boldsymbol{\Theta}^{-1}(\boldsymbol{x})$ as $\boldsymbol{\Theta}^{-1}(\boldsymbol{x}) = \boldsymbol{\Psi} + \boldsymbol{B}\boldsymbol{x}\boldsymbol{x}'\boldsymbol{B}'$ where $\boldsymbol{\Psi}$ is a $q \times q$ positive-definite matrix and $\boldsymbol{B}$ is a $q \times p$ matrix
- Since $\boldsymbol{\Psi}$ is positive-definite and $\boldsymbol{B}\boldsymbol{x}\boldsymbol{x}'\boldsymbol{B}'$ is a rank-1 matrix that depends on $\boldsymbol{x}$, then $\boldsymbol{\Theta}^{-1}(\boldsymbol{x})$ is also positive-definite

Dimensionality and Sparsity

Given the context of the scientific question, the **dimensionality** of the data may vary greatly:

- This approach accomodates both the **low-** $(p, q << n)$ and **high-dimensional** $(p, q >> n)$ settings
- When $p, q >> n$, we impose **sparsity conditions** on the selection and estimation of $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{\Psi}$ through **regularization**
- In the **special case** when $p, q << n$ and $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{\Psi}$ are dense, the problem reduces to Hoff and Niu's covariance regression[15]

Random Effects Model Representation

We can conveniently express this formulation as a **random effects** model:

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \gamma \cdot \boldsymbol{B}\boldsymbol{x} + \boldsymbol{\varepsilon} \tag{2}$$

where $\gamma \sim N(0,1)$, $\boldsymbol{\varepsilon} \sim \mathcal{N}_q\left(\boldsymbol{0}, \boldsymbol{\Psi}\right)$, and $\gamma \perp \boldsymbol{\varepsilon}$. Thus:

- $\mathrm{E}\left[\boldsymbol{y}\right] = \boldsymbol{A}\boldsymbol{x} = \boldsymbol{\mu}(\boldsymbol{x})$
- $\mathrm{E}\left[\left(\boldsymbol{y} - \boldsymbol{\mu}\left(\boldsymbol{x}\right)\right)\left(\boldsymbol{y} - \boldsymbol{\mu}\left(\boldsymbol{x}\right)\right)'\right] = \boldsymbol{B}\boldsymbol{x}\boldsymbol{x}'\boldsymbol{B}' + \boldsymbol{\Psi} = \boldsymbol{\Theta}^{-1}\left(\boldsymbol{x}\right)$

Note: We have $\boldsymbol{\mu}(\boldsymbol{x})$ and $\boldsymbol{\Theta}^{-1}(\boldsymbol{x})$ where $\boldsymbol{x}$ is a common set of predictors. We can consider $\boldsymbol{\mu}(\boldsymbol{x})$ and $\boldsymbol{\Theta}^{-1}(\boldsymbol{x}^*)$ where $\boldsymbol{x}^* \subseteq \boldsymbol{x}$ or $\boldsymbol{x}^* \not\subset \boldsymbol{x}$.

Log-Likelihood Function

Given n i.i.d. samples, $\{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i=1}^{n}$, and with the γ_i are known, we consider the **complete data** log-likelihood function:

$$
\begin{aligned}
\ell\left(\boldsymbol{A}, \boldsymbol{\Psi}, \boldsymbol{B}, \gamma\right) &= \log \left[\prod_{i=1}^{n} (2\pi)^{-\frac{q}{2}} \, |\boldsymbol{\Psi}|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \left[\boldsymbol{y}_i - \left(\boldsymbol{A} + \gamma_i \cdot \boldsymbol{B} \right) \boldsymbol{x}_i \right]' \boldsymbol{\Psi}^{-1} \left[\boldsymbol{y}_i - \left(\boldsymbol{A} + \gamma_i \cdot \boldsymbol{B} \right) \boldsymbol{x}_i \right] \right\} \right] \\
&\propto \frac{1}{2} \sum_{i=1}^{n} \log |\boldsymbol{\Psi}|^{-1} - \left[\boldsymbol{y}_i - \left(\boldsymbol{A} + \gamma_i \cdot \boldsymbol{B} \right) \boldsymbol{x}_i \right]' \boldsymbol{\Psi}^{-1} \left[\boldsymbol{y}_i - \left(\boldsymbol{A} + \gamma_i \cdot \boldsymbol{B} \right) \boldsymbol{x}_i \right]
\end{aligned} \tag{3}
$$

In reality, the γ_i are **unknown** random effects, thus there is added computational difficulty, as we cannot observe $\boldsymbol{B}$ and $\boldsymbol{\Psi}$, only $\boldsymbol{\Theta}^{-1}$

Algorithm

Our Problem

Our Problem

Maximize: $\ell(\boldsymbol{A}, \boldsymbol{\Psi}, \boldsymbol{B})$ subject to: $\|[\boldsymbol{A}|\boldsymbol{B}]\|_1 \leq \lambda_1$ and $\|\boldsymbol{\Psi}\|_1 \leq \lambda_2$

- λ_1 and λ_2 are **tuning parameters** and $\|\cdot\|_1$ is the ℓ_1 norm
- Constraints **control sparsity** in the mean/covariance coefficients and our 'baseline' heterogeneity, respectively
- Exploiting the random-effects representation in (2), we establish a penalized expectation-maximization **(EM) algorithm**

Estimation via Penalized EM-Algorithm

E-Step

As γ_i are **unobserved**, we replace them in the likelihood with their:

- Conditional Expectation: $\mathrm{E}[\gamma_i | \boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{\Psi}, \boldsymbol{B}]$
- Conditional Variance: $\mathrm{Var}[\gamma_i | \boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{\Psi}, \boldsymbol{B}]$

M-Step

We formulate two ℓ_1-constrained optimization problems to iteratively obtain estimates for our parameters $\{\boldsymbol{A}, \boldsymbol{B}\}$ and $\boldsymbol{\Psi}$

- These problems can be expressed in their **primal-dual** form
- And solved with straight-forward **linear programming** approaches

Penalized EM-Algorithm: E-Step

In the **low-dimensional** setting, $\gamma_i | \boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{\Psi}, \boldsymbol{B} \sim \mathcal{N}(m_i, v_i)$ where:

$$v_i = \mathrm{Var}[\gamma_i | \boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\Psi}, \boldsymbol{B}] = (1 + \boldsymbol{x}_i' \boldsymbol{B}' \boldsymbol{\Psi}^{-1} \boldsymbol{B} \boldsymbol{x}_i)^{-1}$$

$$m_i = \mathrm{E}[\gamma_i | \boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\Psi}, \boldsymbol{B}] = v_i (\boldsymbol{y}_i - \boldsymbol{A} \boldsymbol{x}_i)' \boldsymbol{\Psi}^{-1} \boldsymbol{B} \boldsymbol{x}_i$$

In the **high-dimensional** setting, these integrals are intractable. Conditional means/variances are approximated with **Laplace's method**

E-Step: Complete Data Log Likelihood

We utilize the expressions derived for the conditional means and variances of γ_i in the **expected complete-data log likelihood** as follows:

$$
\begin{aligned}
&Q(\boldsymbol{A}, \boldsymbol{\Psi}, \boldsymbol{B} | \hat{\boldsymbol{A}}, \hat{\boldsymbol{\Psi}}, \hat{\boldsymbol{B}}) = -2 \cdot \mathrm{E}[\ell(\boldsymbol{A}, \boldsymbol{\Psi}, \boldsymbol{B}) | \hat{\boldsymbol{A}}, \hat{\boldsymbol{\Psi}}, \hat{\boldsymbol{B}}] \\
&\qquad \propto n \log |\boldsymbol{\Psi}| + \sum\nolimits_{i=1}^{n} \mathrm{E}\left[\left(\boldsymbol{y}_i - \hat{\boldsymbol{A}}\boldsymbol{x}_i - \gamma_i \cdot \boldsymbol{B}\boldsymbol{x}_i \right)' \boldsymbol{\Psi}^{-1} \left(\boldsymbol{y}_i - \hat{\boldsymbol{A}}\boldsymbol{x}_i - \gamma_i \cdot \boldsymbol{B}\boldsymbol{x}_i \right) | \hat{\boldsymbol{A}}, \hat{\boldsymbol{\Psi}}, \hat{\boldsymbol{B}} \right] \\
&\qquad = n \log |\boldsymbol{\Psi}| + \sum\nolimits_{i=1}^{n} \left\{ \left(\boldsymbol{y}_i - \hat{\boldsymbol{A}}\boldsymbol{x}_i - m_i \boldsymbol{B}\boldsymbol{x}_i \right)' \boldsymbol{\Psi}^{-1} \left(\boldsymbol{y}_i - \hat{\boldsymbol{A}}\boldsymbol{x}_i - m_i \boldsymbol{B}\boldsymbol{x}_i \right) + s_i \boldsymbol{x}_i' \boldsymbol{B}' \boldsymbol{\Psi}^{-1} \boldsymbol{B}\boldsymbol{x}_i s_i \right\}
\end{aligned}
$$

where $s_i = \sqrt{v_i}$

E-Step: Alternative Notation

We then construct the following **augmented matrices**:

$$\boldsymbol{X}^* = \begin{bmatrix} \boldsymbol{x}_1' & \cdots & \boldsymbol{x}_n' & \boldsymbol{0}_1' & \cdots & \boldsymbol{0}_p' \\ m_1\boldsymbol{x}_1' & \cdots & m_n\boldsymbol{x}_n' & s_1\boldsymbol{x}_1' & \cdots & s_n\boldsymbol{x}_n' \end{bmatrix}_{2n\times 2p}'$$

$$\boldsymbol{Y}^* = \begin{bmatrix} \boldsymbol{Y}_{n\times q}' \\ \boldsymbol{0}_{n\times q}' \end{bmatrix}_{2n\times q} \qquad \boldsymbol{C}^* = \begin{bmatrix} \boldsymbol{A}_{p\times q} \\ \boldsymbol{B}_{p\times q} \end{bmatrix}_{p\times 2q}'$$

and write the expected value of the complete data log-likelihood as:

$$\begin{aligned} Q(\boldsymbol{A}, \boldsymbol{\Psi}, \boldsymbol{B}|\hat{\boldsymbol{A}}, \hat{\boldsymbol{\Psi}}, \hat{\boldsymbol{B}}) &= -2 \cdot \mathrm{E}[\ell(\boldsymbol{A}, \boldsymbol{\Psi}, \boldsymbol{B})|\hat{\boldsymbol{A}}, \hat{\boldsymbol{\Psi}}, \hat{\boldsymbol{B}}] \\ &\propto n \log|\boldsymbol{\Psi}| + [\boldsymbol{Y}^* - \boldsymbol{X}^*(\boldsymbol{C}^*)']' \, \boldsymbol{\Psi}^{-1} \, [\boldsymbol{Y}^* - \boldsymbol{X}^*(\boldsymbol{C}^*)'] \end{aligned}$$

Penalized EM-Algorithm: M-Step

In the **low-dimensional setting**, we have convenient, closed-form updates for both $\hat{\boldsymbol{C}}^*$ and $\hat{\boldsymbol{\Psi}}$:[15]

- $\hat{\boldsymbol{C}}\boldsymbol{*} = (\boldsymbol{Y}^*)'\boldsymbol{X}^*\left[(\boldsymbol{X}^*)'\boldsymbol{X}^*\right]^{-1}$
- $\hat{\boldsymbol{\Psi}} = \frac{1}{n}\left[\boldsymbol{Y}^* - \boldsymbol{X}^*(\boldsymbol{C}^*)'\right]'\left[\boldsymbol{Y}^* - \boldsymbol{X}^*(\boldsymbol{C}^*)'\right]$

In the **high-dimensional setting**, we iteratively update $\hat{\boldsymbol{C}}^*$ and $\hat{\boldsymbol{\Psi}}$ using linear programming approaches for constrained ℓ_1 minimization[13,16,17]

M-Step: Notation

Let:

- $\bar{y^*} = (2n)^{-1} \sum_{i=1}^{2n} y_i^*$
- $\bar{x^*} = (2n)^{-1} \sum_{i=1}^{2n} x_i^*$

and define:

- $S_{x^* y^*} = (2n)^{-1} \sum_{i=1}^{2n} \left(y_i^* - \bar{y^*} \right) \left(x_i^* - \bar{x^*} \right)'$
- $S_{x^* x^*} = (2n)^{-1} \sum_{i=1}^{n} \left(x_i^* - \bar{x^*} \right) \left(x_i^* - \bar{x^*} \right)'$
- $S_{y^* y^*} = (2n)^{-1} \sum_{i=1}^{2n} \left(y_i^* - \hat{\boldsymbol{C}}^* x_i^* \right) \left(y_i^* - \hat{\boldsymbol{C}}^* x_i^* \right)'$

M-Step: Updating C^*

We estimate $\boldsymbol{C}^*$ by solving the constrained optimization problem:

$$\hat{\boldsymbol{C}}^* \in \underset{\boldsymbol{C}^* \in R^{p \times 2q}}{\arg\min} \left\{ |\boldsymbol{C}^*|_1 : |S_{x^*y^*} - \boldsymbol{C}^* S_{x^*x^*}|_\infty \leqslant \lambda_1 \right\}$$

where λ_1 is the tuning parameter

- Exploiting the **separability** of the penalty function, this this is equivalently carried out as p separate optimization problems
- Expressing this minimization problem in its **primal-dual** form, we utilize a multivariate variation on the **Dantzig selector**[13,16]

M-Step: Updating $\boldsymbol{\Psi}$

Given $\boldsymbol{C}^*$, we estimate $\boldsymbol{\Psi}$ by solving the constrained optimization problem:

$$\hat{\Psi} \in \underset{\Psi \in R^{q\times q}}{\arg\min} \left\{ |\Psi|_1 : \left| I_{q\times q} - S_{y^* y^*} \Psi \right|_\infty \leqslant \lambda_2 \right\}$$

where λ_2 is the tuning parameter:

- We again exploit the **separability** of the penalty function and solve q separate optimization problems
- Expressing this minimization problem in its **primal-dual** form, we utilize a variation on the **CLIME algorithm**[17]

Symmetry Conditions on $\hat{\Psi}$

We impose a **symmetry condition** on $\hat{\Psi}$ as in Cai et al. (2013):[13]

$$\hat{\boldsymbol{\Psi}} = \left(\hat{\psi}_{ij}\right)$$

$$\hat{\psi}_{ij} = \hat{\psi}_{ji} = \hat{\psi}^1_{ij} \cdot \mathbb{I}\left(\left|\hat{\psi}^1_{ij}\right| \leqslant \left|\hat{\psi}^1_{ji}\right|\right) + \hat{\psi}^1_{ji} \cdot \mathbb{I}\left(\left|\hat{\psi}^1_{ij}\right| > \left|\hat{\psi}^1_{ji}\right|\right)$$

where $\mathbb{I}(\cdot)$ is the indicator function

Tuning λ_1 and λ_2

- We run the EM-Algorithm over a **grid** of candidate λ_1 and λ_2 values
- Tuning of λ_1 and λ_2 is carried out via *k*-fold **cross-validation**
- Optimal λ_1 and λ_2 are evaluated **jointly** using the Bayesian Information Criterion

Conclusions

Future Work

- Work through several computational considerations, including parallelization and converting `R` code to `C++`
- Run simulations comparing the selection and estimation accuracy and performance time of our method to existing methods
- Develop an `R` package and submit to the Comprehensive `R` Archive Network (CRAN)
- Analyze data from the Boston Lung Cancer Study Cohort

(brief) Simulated Case

We simulate $\boldsymbol{A}$ and $\boldsymbol{B}$ to form complex, though *not* biologically plausible coefficient matrices for the mean and covariance functions:

- $p = 70,\ q = 100,\ n = 50$
- $\boldsymbol{x}_i = (x_1, \ldots, x_p)' \sim \mathrm{Bin}(p, 1/q);\ i = 1, \ldots, n$

(brief) Simulated Case

Example results are given in Figure (3) below:

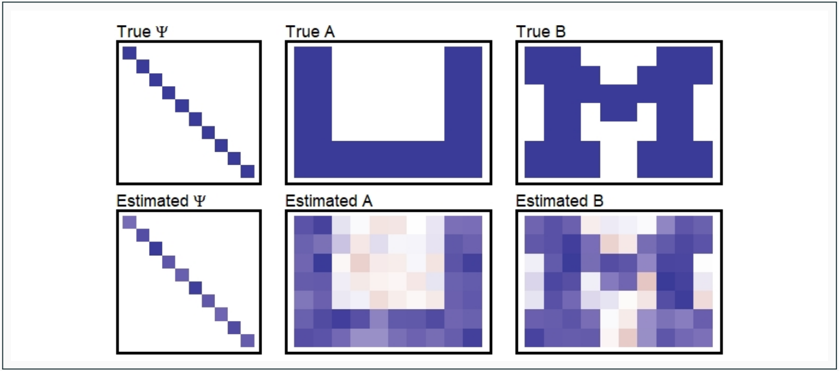


Figure 3: Recovery of complex mean and covariance coefficient structures

Conclusions

We have proposed a **novel method** for the selection and estimation of conditional Gaussian graphical models:

- We jointly model the mean and covariance structure of our Gaussian graph conditional on low- or high-dimensional covariates
- We offer a parsimonious random-effects model representation with computationally efficient and straightforward estimation techniques
- Parameters of $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{\Psi}$ have a direct interpretation in terms of how heteroscedasticity co-occurs in $\boldsymbol{y}$

References i

[1] S. R. Proulx, D. E. Promislow, and P. C. Phillips, "Network thinking in ecology and evolution," *Trends in ecology & evolution*, vol. 20, no. 6, pp. 345–353, 2005.

[2] S. Bandyopadhyay, M. Mehta, D. Kuo, M. K. Sung, R. Chuang, E. J. Jaehnig, B. Bodenmiller, K. Licon, W. Copeland, M. Shales, D. Fiedler, J. Dutkowski, A. G. abd H. van Attikum, K. M. Shokat, R. D. Kolodner, W. K. Huh, R. Aebersold, M. C. Keogh, N. J. Krogan, and T. Ideker, "Rewiring of genetic networks in response to DNA damage," *Science*, vol. 330, no. 6009, pp. 1385–1389, 2010.

[3] J. Greshock, K. Nathanson, A. Medina, M. R. Ward, M. Herlyn, B. L. Weber, and T. Z. Zaks, "Distinct patterns of DNA copy number alterations associate with *BRAF* mutations in melanomas and melanoma-derived cell lines," *Genes Chromosomes Cancer*, vol. 48, pp. 419–428, 2009.

[4] V. Lázár, S. Ecsedi, L. Vízkeleti, Z. Rákosy, G. Boross, B. Szappanos, B. A., G. Emri, R. Adány, and M. Balázs, "Marked genetic differences between *BRAF* and *NRAS* mutated primary melanomas as revealed by array comparative genomic hybridization," *Melanoma Research*, vol. 22, pp. 202–214, 2012.

[5] D. Ronen and N. Benvenisty, "Sex-dependent gene expression in human pluripotent stem cells," *Cell Reports*, vol. 8, no. 4, pp. 923–932, 2014.

[6] D. Iacobas, S. Iacobas, N. Thomas, and D. C. Spray, "Sex-dependent gene regulatory networks of the heart rhythm," *Functional & Integrative Genomics*, vol. 10, no. 1, pp. 73–86, 2010.

[7] J. M. Ranz, C. I. Castillo-Davis, C. D. Meiklejohn, and D. L. Hartl, "Sex-dependent gene expression and evolution of the Drosophila transcriptome," *Science*, vol. 300, no. 5626, pp. 1742–1745, 2003.

[8] K. H. Clodfelter, M. G. Holloway, P. Hodor, S.-H. Park, W. J. Ray, and D. J. Waxman, "Sex-dependent liver gene expression is extensive and largely dependent upon STAT5b: STAT5b-dependent activation of male genes and repression of female genes revealed by microarray analysis," *Mol Endocrinol*, vol. 20, pp. 1333–1351, 2006.

References ii

[9] F. Conforti, L. Pala, V. Bagnardi, G. Viale, T. De Pas, E. Pagan, E. Pennacchioli, E. Cocorocchio, P. F. Ferrucci, and F. De Marinis, "Sex-based heterogeneity in response to lung cancer immunotherapy: a systematic review and meta-analysis," *JNCI: Journal of the National Cancer Institute*, vol. 111, no. 8, pp. 772–781, 2019.

[10] J. Chiquet, T. Mary-Huard, and S. Robin, "Structured regularization for conditional gaussian graphical models," *Statistics and Computing*, vol. 27, no. 3, pp. 789–804, 2017.

[11] J. Yin and H. Li, "A sparse conditional Gaussian graphical model for analysis of genetical genomic data," *The Annals of Applied Statistics*, vol. 5, no. 4, pp. 2630–2650, 2011.

[12] B. Li, H. Chuns, and H. Zhao, "Sparse estimation of conditional graphical models with application to gene networks," *Journal of the American Statistical Association*, vol. 107, no. 497, pp. 152–167, 2012.

[13] T. Cai, H. Li, W. Liu, and J. Xie, "Covariate-adjusted precision matrix estimation with an application in genetical genomics," *Biometrika*, vol. 100, pp. 139–156, 2013.

[14] Y. Ni, F. C. Stingo, and V. Baladandayuthapani, "Bayesian graphical regression," *Journal of the American Statistical Association*, vol. 114, no. 525, pp. 184–197, 2019.

[15] P. D. Hoff and X. Niu, "A covariance regression model," *Statistica Sinica*, pp. 729–753, 2012.

[16] E. Candes, T. Tao *et al.*, "The dantzig selector: Statistical estimation when p is much larger than n," *The annals of Statistics*, vol. 35, no. 6, pp. 2313–2351, 2007.

[17] T. Cai, W. Liu, and X. Luo, "A constrained l1 minimization approach to sparse precision matrix estimation," *Journal of the American Statistical Association*, vol. 106, no. 494, pp. 594–607, 2011.