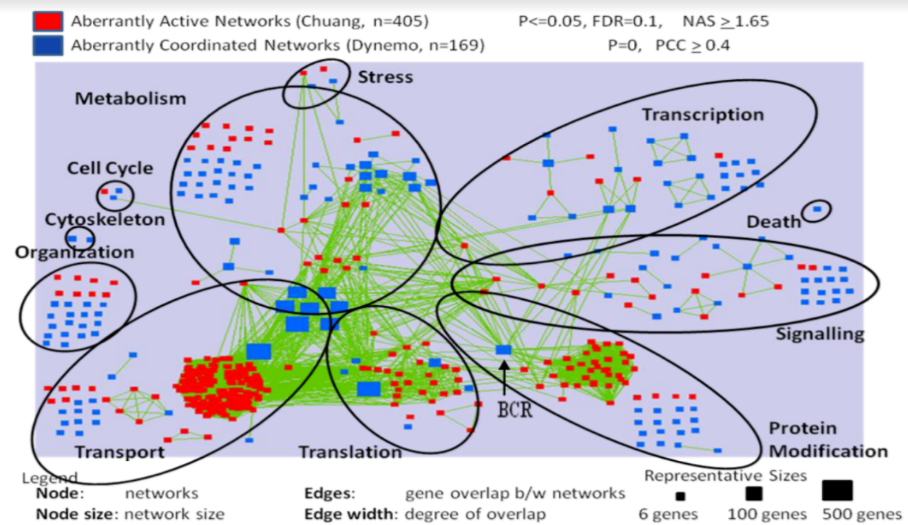
Regression analysis of networked data

Presenter: Soroor Hediyeh Zadeh & Lei Yu

Oct. 2017

Outlines

Motivation and Background: SNPs and Gene Regulation Networks



Motivation and Background: Electroencephalogram (EEG)

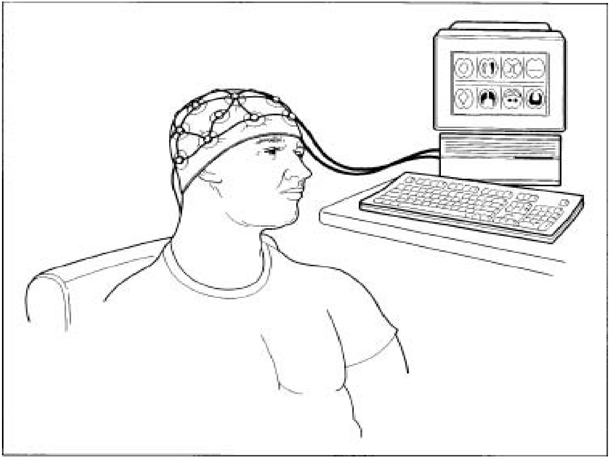


Figure: An electroencephalogram (EEG) is a test that measures and records the electrical activity of your brain.

Motivation and Background: Electroencephalogram (EEG)

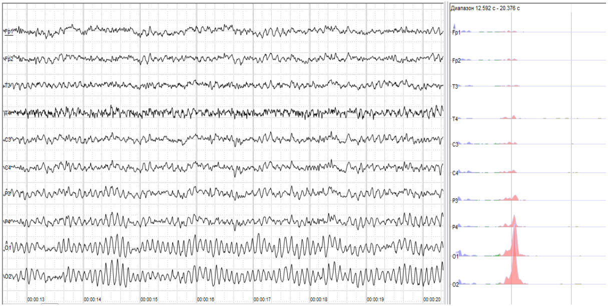


Figure: The computer records brain's electrical activity on the screen.

Motivation and Background: The main study

- In this paper, we consider the late slow wave outcome. (i.e. response variables)
- We discuss regression analysis of multi-dimensional response variables on covariates that are collected from networks.

Such measurements from the 64 electrodes are correlated in the EEG net, and the correlation is highly clustered according to subregions of memory functionality.

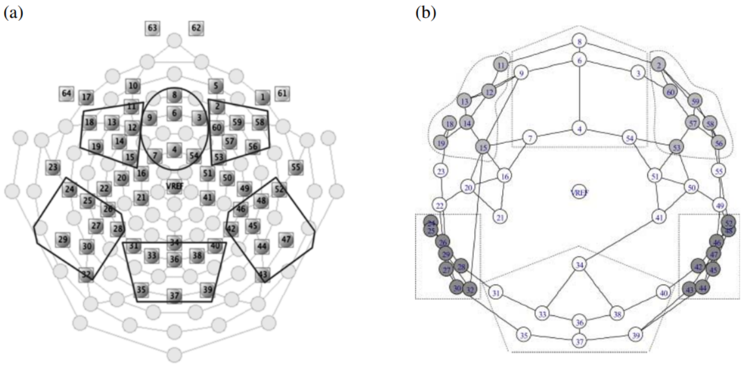


Fig. 1. (a) Layout of the 64-channel sensor net, where the six outlined clusters of nodes relate to auditory recognition memory and the remaining nodes belong to an additional cluster. (b) Sparse graphical representation of the learned network among electrodes based on the late slow wave data under voice stimulus from a stranger.

Motivation and Background: Challenges

• The existing methods only account for symmetric exchangeable correlations, whereas correlations between measurements in this data are not necessary symmetric

• Network dependence structure can not be easily incorporated to models such as marginal regression over nodes.

• proper estimation methods often involve estimating a large number of nuisance parameter, leading to loss of estimation efficiency.

Motivation and Background: The method in this research

We treat the EEG net as a network and develop a flexible dependence model that can better reflect the underlying relationships among the electrodes, for instance allowing for clustered and asymmetric dependence relationships.

Framework: Estimating functions

Variables:

• Suppose that the response variable y_{ij} and the associated p-dimensional covariate x_{ij} are measured at node, or vertex, j for subject (sample) $i(j = 1, ..., m; i = 1, ..., n)$.

• Let $y_i = (y_{i1}, ..., y_{im})^T$ and $x_i = (x_{i1}, ..., x_{im})^T$, which is an $m \times p$ matrix, and let $(y_i, x_i)(i = 1, ..., n)$ be be independent and identically distributed data from n subjects.

Population-average model framework:

$\mu_{ij} = E(y_{ij}|x_{ij}) = \mu(x_{ij}^T\beta)$, where $\mu(\cdot)$ is a known link function, β is a p-dimensional parameter vector of interesting, and $\mu_i = (\mu_{i1}, ..., \mu_{im})^T$

Framework: Estimating functions

•(Liang & Zeger, 1986) Generalized estiamting equations provide an estimate of $\hat{\beta}$ by solving the equation:

$$\sum_{i=1}^{n} \dot{\mu}^T V_i^{-1}(y_i - \mu_i) = 0 \tag{1}$$

• where $\dot{\mu}_i(\cdot)$ is the gradient vector of $\mu_i(\cdot)$ with respect to β.

•The second moment of y_i is specificed by $V_i = A_i^{\frac{1}{2}} R(\alpha) A_i^{\frac{1}{2}}$.

•$R(\alpha)$ is the correlation matrix.

•A_i is the diagonal matrix of matginal variances $var(y_{ij}|x_{ij})$.

(Because the number of nodes in a network is fixed, we write the variance V_i as simply V)

Framework: Estimating functions

•(Liang & Zeger, 1986) Generalized estiamting equations provide an estimate of $\hat{\beta}$ by solving the equation:

$$\sum_{i=1}^{n} \dot{\mu}^T V_i^{-1}(y_i - \mu_i) = 0 \tag{1}$$

• where $\dot{\mu}_i(\cdot)$ is the gradient vector of $\mu_i(\cdot)$ with respect to β.

•The second moment of y_i is specificed by $V_i = A_i^{\frac{1}{2}} R(\alpha) A_i^{\frac{1}{2}}$.

•$R(\alpha)$ is the correlation matrix.

•A_i is the diagonal matrix of matginal variances $var(y_{ij}|x_{ij})$.

(Because the number of nodes in a network is fixed, we write the variance V_i as simply V)

• This method have low efficiency if the working correlation $R(\alpha)$ does not represent the true correlation structure.

Framework: Estimating functions

Therefore, we use a popular approach procedure of Qu et al (2000) improving the efficiency of generalized estimating equations estimators.

$$R^{-1}(\alpha) = \sum_{k=0}^{K} a_k M_k, \tag{2}$$

• Where M_0 is the identity matrix
• $M_k (k = 1, ..., K)$ are known symmetric basis matrices with elements equal to either 0 or 1, and the a_k are unknown coefficients that may depend on the parameter α.

Framework: Estimating functions

Graphical interpretation of basis matrices

The basis matrices is determined by the structure of Network:

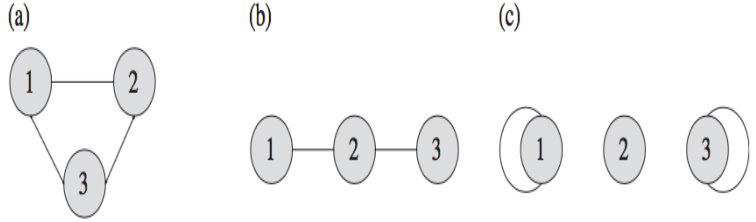


Fig. 2. Graphical display of basis matrices (a) M_{comp} (M_1), (b) M_{chain} (M_1^*), and (c) M_2^* for a three-node network.

Framework: Estimating functions

Then, the generalized estimating equations may be written as a linear combination of estimating functions given by the extended score vector

$$\bar{q}_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} q_i(\beta) = \frac{1}{n}\sum_{i=1}^{n} f(x) = \begin{pmatrix} \dot{\mu}_i^T A_i^{-1}(y_i - \mu_i) \\ \hat{\mu}_i^T A^{-1/2} M_1 A_i^{-1/2}(y_i - \mu_i) \\ \vdots \\ \hat{\mu}_i^T A^{-1/2} M_K A_i^{-1/2}(y_i - \mu_i) \end{pmatrix}, \tag{3}$$

where the dimension of $\bar{q}_n(\beta)$ is $p(K+1)$.

• The quadratic inference function does not require estimation of the nuisance parameter α.

Framework: Estimating functions

• Similar to generalized method of moments, the quadratic inference function method minimizes a quadratic objective function of the form

$$n\bar{q}_n^T(\beta)\Gamma^{-1}(\beta)\bar{q}_n(\beta),$$

where the optimal weighting matrix is $\Gamma(\beta) = var\{q_i(\beta)\}$, which may be consistently estimated by the sample covariance matrix $\bar{\Gamma}_n = n^{-1}\sum_{i=1}^{n} q_i(\beta)q_i^T(\beta)$

Framework: Estimating functions

• Similar to generalized method of moments, the quadratic inference function method minimizes a quadratic objective function of the form

$$n\bar{q}_n^T(\beta)\Gamma^{-1}(\beta)\bar{q}_n(\beta),$$

where the optimal weighting matrix is $\Gamma(\beta) = var\{q_i(\beta)\}$, which may be consistently estimated by the sample covariance matrix
$\bar{\Gamma}_n = n^{-1}\sum_{i=1}^{n} q_i(\beta)q_i^T(\beta)$
• In practice, $M_1, ..., M_k$ is not easy to be determined since the graphical structure always very complex.

Framework: Data-driven network topology

One solution given by Qu & Lindsay (2003) is the so-called adaptive procedure, which requires only estimation of the covariance matrix. It follows from the Cayley–Hamilton theorem that the inverse of an $m \times m$ positive-definite matrix may be written as

$$V^{-1} = \frac{(-1)^{m-1}}{|V|}(c_1 I + c_2 V + \ldots + c_{m-1} V^{m-2} + V^{m-1}),$$

where c_j $(j = 1, \ldots, m-1)$ are certain suitable coefficients. Consequently, the optimal weight matrix $V^{-1}\dot{\mu}$ for a basic estimating function $s = y - \mu(\beta)$ lies in the space spanned by the columns of $\dot{\mu}$, $V\dot{\mu}, \ldots, V^{m-1}\dot{\mu}$.

Framework: Data-driven network topology

For sake of parsimony, Qu & Lindsay (2003) suggested the gradient direction generated by the first two columns, suggested including only the gradient direction generated by the first two columns, $\dot{\mu}$ and $V\dot{\mu}$. This gives the extended score vector

$$\bar{h}_n(\beta) = \begin{pmatrix} \bar{h}_n^{(1)} \\ \bar{h}_n^{(2)} \end{pmatrix} = \frac{1}{n}\sum_{i=1}^{n} \begin{pmatrix} \dot{\mu}_i^T(y_i - \mu_i) \\ \dot{\mu}_i^T V(y_i - \mu_i) \end{pmatrix}, \tag{4}$$

where V is consistently estimated by $\hat{V} = n^{-1}\sum_{i=1}^{n} s_i s_i^T$ with $s_i = y_i - \mu_i(\beta)$.

Framework: Data-driven network topology

For sake of parsimony, Qu & Lindsay (2003) suggested the gradient direction generated by the first two columns, suggested including only the gradient direction generated by the first two columns, $\dot{\mu}$ and $V\dot{\mu}$. This gives the extended score vector

$$\bar{h}_n(\beta) = \begin{pmatrix} \bar{h}_n^{(1)} \\ \bar{h}_n^{(2)} \end{pmatrix} = \frac{1}{n}\sum_{i=1}^{n} \begin{pmatrix} \dot{\mu}_i^T(y_i - \mu_i) \\ \dot{\mu}_i^T V(y_i - \mu_i) \end{pmatrix}, \quad (4)$$

where V is consistently estimated by $\hat{V} = n^{-1}\sum_{i=1}^{n} s_i s_i^T$ with $s_i = y_i - \mu_i(\beta)$.

•The number of parameters to be estimated in V is large, especially in the case of complex networks, and thus overfitting may occur in determining the network dependence structure.

Proposed Method: Hybrid quadratic inference function

We propose to construct the extended score

$$\bar{g}_n(\beta|\gamma) = \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} \dot{\mu}_i^T A^{-1}(y_i - \mu_i) \\ \dot{\mu}_i^T\{\gamma A_i^{-\frac{1}{2}}\Pi A_i^{-\frac{1}{2}} + (1-\gamma)V\}(y_i - \mu_i) \end{pmatrix}, \qquad (5)$$

where Π is the prior target structure, a given adjacency matrix.
$\gamma \in [0,1]$ denotes the shrinkage intensity coefficient.
The right-most expression in (5) is intended to provide an improvement in estimation efficiency.

Proposed Method: Hybrid quadratic inference function

Let $U_i(\gamma) = \gamma A_i^{-\frac{1}{2}} \Pi A_i^{-\frac{1}{2}} + (1-\gamma)V$, a linear shrinkage estimator of V(Ledoit& Wolf, 2004). For $\gamma = 1$ the shrinkage estimator fully favours the prior target Π, whereas for $\gamma = 0$ it reduces to the unrestricted covariance V.

By the idea of shrinkage estimation (Stein, 1956), our regularization procedure involves shrinking the estimation of the covariance V towards a known piror structure. We propose to construct the extend score:

$$\bar{g}_n(\beta|\gamma) = \frac{\gamma}{n} \sum_{i=1}^{n} \begin{pmatrix} \dot{\mu}_i^T A^{-1}(y_i - \mu_i) \\ \dot{\mu}_i^T A_i^{-\frac{1}{2}} \Pi A_i^{-\frac{1}{2}} (y_i - \mu_i) \end{pmatrix} + \frac{1-\gamma}{n} \sum_{i=1}^{n} \begin{pmatrix} \dot{\mu}_i^T A^{-1}(y_i - \mu_i) \\ \dot{\mu}_i^T V (y_i - \mu_i) \end{pmatrix}, \tag{6}$$

Proposed Method: Hybrid quadratic inference function

We call (6) the hybrid extended score vector; Consequently, given a shrinkage coefficient γ, we can estimate β by minimizing

$$Q_n(\beta|\gamma) = n\bar{g}_n^T(\beta|\gamma)\Gamma^{-1}(\beta|\gamma)\bar{g}_n(\beta|\gamma), \tag{7}$$

where Γ is consistently estimated by

$$\bar{\Gamma}_n = n^{-1}\sum_{i=1}^{n} g_i(\beta|\gamma)g_i^T(\beta|\gamma)$$

Since the estimatir of β depends on the choice of shrinkage coefficient γ, it is denoted by $\hat{\beta}(\gamma)$.

Proposed Method: Asymptotic properties

The estimator β is not only consistent but also asymptotically normally distributed. With a known target structure Π and a fixed shrinkage coefficient γ, these large-sample properties remain valid for the proposed estimator in (7). In other words, $\hat{\beta}(\gamma) \to \beta_0$ in probability as $n \to \infty$, and

$$\sqrt{n}\{\hat{\beta}(\gamma) - \beta_0\} \to N\{0, J^{-1}(\beta_0|\gamma)\} \tag{8}$$

in distribution as $n \to \infty$, where $J(\beta_0|\gamma) = G^T(\beta_0|\gamma)\Gamma^{-1}(\beta_0|\gamma)G(\beta_0|\gamma)$ is the Godambe inforamtion of $g_i(\beta_0|\gamma)$, provided that $\bar{\Gamma}_n(\hat{\beta}|\gamma) \to \Gamma(\beta_0|\gamma)$ and $\dot{\bar{g}}_n(\hat{\beta}|\gamma) \to G(\beta_0|\gamma)$ in probability.

Proposed Method: Chioce of the shrinkage coefficient

We propose to select γ by minimizing the trace of the inverse of the Godambe information matrix $J(\beta_0|\gamma)$, in order to maximize estimation efficiency over $\gamma \in [0, 1]$:

$$\tilde{\gamma} = \mathit{argmin}\ \mathit{tr}\{J^{-1}(\beta_0|\gamma)\} \tag{9}$$

After we determine the suitable γ, we put $\tilde{\gamma}$ into function (7) and get $\hat{\beta}$

Simulation

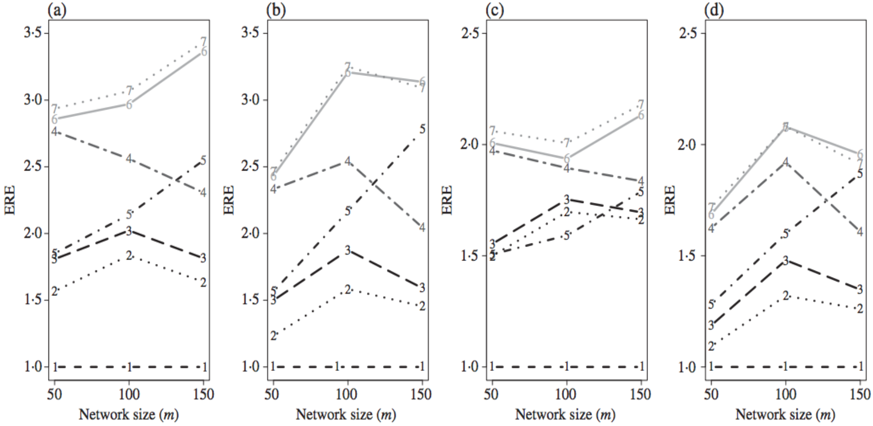


Fig. 4. Comparison of empirical relative efficiency (ERE) under the five-subregion network N3 with sample size n, number of nodes $m = 50, 100, 150$ and $\Pi^* = \Pi_{\text{CL}}$: (a) R^a_{CL} and $n = 100$; (b) R^a_{CL} and $n = 500$; (c) R^b_{CL} and $n = 100$; (d) R^b_{CL} and $n = 500$. In each panel the labelled lines indicate: 1, generalized estimating equations oracle where the reference equals 1; 2, $\hat{\beta}(\Pi = \Pi^*, \gamma = \hat{\gamma}^*)$; 3, $\hat{\beta}(\gamma = 0)$; 4, $\hat{\beta}(\Pi = \Pi^*, \gamma = 1)$; 5, $\hat{\beta}(\Pi = M_{\text{chain}}, \gamma = 1)$; 6, generalized estimating equations independence; 7, $\hat{\beta}(\Pi = M_{\text{comp}}, \gamma = 1)$.

Table of regression coefficients

Table 3. *Estimated regression coefficients $\hat{\beta}$ for the infant memory data with respect to mother's voice stimulus* ($*$: p-value $< 0{\cdot}05$), *with estimated standard errors in parentheses. The first two columns of values are for $\hat{\beta}(\Pi = \Pi^*, \gamma = \hat{\gamma}^*)$ under two types of network structure suggested by our collaborators with different optimal shrinkage coefficients; the third and fourth columns of values are for $\hat{\beta}(\gamma = 0)$ and the spatial analysis-of-variance mixed-effects model; the final row lists the estimated sums of variances for $\hat{\beta}$*

Parameter	$\Pi^* = \Pi_{7\text{comp}}$ $\hat{\gamma}^* = 0{\cdot}875$	$\Pi^* = \Pi_{\text{stranger}}$ $\hat{\gamma}^* = 0{\cdot}583$	$\gamma = 0$	Spatial ANOVA mixed-effects model
age	−0·003 (0·002)	−0·003 (0·001)	−0·003 (0·001)*	−0·001 (0·002)
lead	−0·006 (0·003)	−0·005 (0·003)	−0·006 (0·003)*	0·000 (0·004)
group	0·158 (0·174)	0·158 (0·174)	0·176 (0·173)	0·587 (0·271)*
left fc	−0·803 (0·220)*	−0·854 (0·218)*	−0·811 (0·220)*	−0·824 (0·335)*
middle fc	−0·360 (0·189)	−0·338 (0·183)	−0·363 (0·186)	−0·580 (0·275)*
right fc	−1·375 (0·218)*	−1·327 (0·215)*	−1·373 (0·218)*	−1·045 (0·343)*
left po	−0·167 (0·259)	−0·359 (0·251)	−0·200 (0·251)	0·466 (0·370)
middle po	−0·056 (0·281)	−0·110 (0·280)	−0·071 (0·282)	1·566 (0·367)*
right po	0·573 (0·240)*	0·603 (0·230)*	0·559 (0·229)*	1·065 (0·392)*
group × left fc	0·714 (0·344)*	0·571 (0·380)	0·482 (0·374)	−1·143 (0·762)
group × middle fc	0·120 (0·312)	0·092 (0·339)	−0·081 (0·336)	−1·167 (0·703)
group × right fc	−0·462 (0·379)	−0·458 (0·388)	−0·612 (0·390)	−1·101 (0·746)
group × left po	0·056 (0·392)	0·167 (0·413)	0·246 (0·410)	0·207 (0·593)
group × middle po	−0·112 (0·484)	−0·080 (0·480)	−0·006 (0·491)	−0·277 (0·689)
group × right po	−1·427 (0·374)*	−1·417 (0·366)*	−1·337 (0·367)*	−0·959 (0·689)
$\text{tr}\{\widehat{\text{var}}(\hat{\beta})\}$	1·263	1·306	1·314	3·763

fc, frontal-central; po, parietal-occipital; ANOVA, analysis of variance.