

Accounting for network noise in graph-guided Bayesian modeling of structured high-dimensional data

Wenrui Li^{1,*}, Changgee Chang², Suprateek Kundu³, Qi Long^{1,*}

¹Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, PA 19104, United States, ²Department of Biostatistics and Health Data Science, Indiana University School of Medicine, Indianapolis, IN 46202, United States, ³Department of Biostatistics, University of Texas MD Anderson Cancer Center, Houston, TX 77030, United States

*Corresponding authors: Wenrui Li, Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania, Philadelphia, PA 19104, USA (Wenrui.Li@Pennmedicine.upenn.edu); Qi Long, Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania, Philadelphia, PA 19104, USA (qlong@upenn.edu).

ABSTRACT

There is a growing body of literature on knowledge-guided statistical learning methods for analysis of structured high-dimensional data (such as genomic and transcriptomic data) that can incorporate knowledge of underlying networks derived from functional genomics and functional proteomics. These methods have been shown to improve variable selection and prediction accuracy and yield more interpretable results. However, these methods typically use graphs extracted from existing databases or rely on subject matter expertise, which are known to be incomplete and may contain false edges. To address this gap, we propose a graph-guided Bayesian modeling framework to account for network noise in regression models involving structured high-dimensional predictors. Specifically, we use 2 sources of network information, including the noisy graph extracted from existing databases and the estimated graph from observed predictors in the dataset at hand, to inform the model for the true underlying network via a latent scale modeling framework. This model is coupled with the Bayesian regression model with structured high-dimensional predictors involving an adaptive structured shrinkage prior. We develop an efficient Markov chain Monte Carlo algorithm for posterior sampling. We demonstrate the advantages of our method over existing methods in simulations, and through analyses of a genomics dataset and another proteomics dataset for Alzheimer's disease.

KEYWORDS: adaptive Bayesian shrinkage; latent scale network model; MCMC algorithm; noisy graph; structured high-dimensional prediction.

1 INTRODUCTION

Rapid advances in technologies have led to collection of various types of omics data, such as genomics and proteomics data, in many biomedical studies. Two notable examples are the Alzheimer's disease neuroimaging initiative (ADNI) study (Mueller et al., 2005) and Accelerating Medicines Partnership-Alzheimer's Disease (AMP-AD) study (Hodes and Buckholtz, 2016). These datasets can be used to build prediction models for AD risk or progression (eg, cell metabolism and cognitive score). The majority of existing predictive modeling methods do not incorporate biological knowledge such as functional genomics and functional proteomics. Extensive research has yielded much information on the association structure among variables (eg, genes and proteins) and underlying networks that can be represented by graphs. For example, in genomics studies, several biological databases include gene network information from previous studies (Kanehisa and Goto, 2000). There is a growing body of on knowledge-guided statistical learning methods for analysis of such data that can incorporate biological graphs (Zhao et al., 2019). These methods have been shown to have better performance in variable selection and prediction,

compared to their counterparts that do not use graph information.

However, the network information used in the graph-guided statistical learning methods is generally derived from existing databases such as the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto, 2000) and Search Tools for the Retrieval of Interacting Genes/Proteins (STRING) (Szklarczyk et al., 2021) and presumably available only up to some level of error or mis-specifications. Such noise may also be encountered due to limitations in the algorithms used to construct networks, which are often based on notions of association (eg, correlation, partial correlation, etc.) among experimental measurements of gene activity levels that are determined by some form of statistical inference (Tsimring, 2014). Importantly, mis-specifications may also arise due to the lack of information in the existing databases with regards to certain disease conditions that are of interest. To our best knowledge, existing work on graph-guided regression modeling has largely ignored the important issue of mis-specification in networks and routinely uses the given network directly in their modeling without accounting for noise. We note that while the recent work

by Zhao et al. (2016) addressed a related problem for handling missing edges in only part of the graph via a multiple-imputation approach, their method is limited in scope and cannot be directly applied to diverse and general settings involving noisy networks with varying degrees of mis-specification.

As remarked above, there appears to be little in the way of a formal and general treatment of the error propagation problem in graph-guided statistical learning in the context of regression modeling. However, there are several areas in which the probabilistic or statistical treatment of uncertainty enters prominently in network analysis. Model-based approaches include statistical methodology for predicting network topology with approaches that explicitly include a component for network noise (Jiang et al., 2011; Jiang and Kolaczyk, 2012), the “denoising” of noisy networks (Chatterjee, 2015), the adaptation of methods for vertex classification using networks observed with errors (Priebe et al., 2015), a regression model on network-linked data that is based on a flexible network effect assumption and is robust to errors in the network structure (Le and Li, 2020), and a general Bayesian framework for reconstructing networks from observational data (Young et al., 2020). The other common approach to network noise is based on a “signal plus noise” perspective. For example, Balachandran et al. (2017) introduced a simple model for noisy networks that, conditional on some true underlying network, assumes we observe a version of that network corrupted by an independent random noise that effectively flips the status of (non)edges. Later, Chang et al. (2022) developed method-of-moments estimators for the underlying rates of error when replicates of the observed network are available. The model-based approaches are problem-specific and cannot be directly used in graph-guided statistical learning. The approaches from the “signal plus noise” perspective assume constant error across all edges and thus cannot accommodate different levels of noise in network. In a somewhat different direction, uncertainty in network construction due to sampling has also been studied in some depth. See, for example, Kolaczyk (2009, Chapter 5) or Ahmed et al. (2014) for surveys of this area. However, in that setting, the uncertainty arises only from sampling—the subset of vertices and edges obtained through sampling is typically assumed to be observed without error.

Our contribution in this paper is to provide a graph-guided Bayesian modeling framework that accounts for network noise in the context of regression models with structured high-dimensional predictors. There is no existing work on graph-guided statistical learning that accounts for network noise. Our model treats the true graph as unknown or latent and has 2 key components. The regression model involves a fully Bayesian implementation of the adaptive structured shrinkage prior proposed by Chang et al. (2018) that incorporates the unknown true graph to facilitate variable selection. This regression model is coupled with a latent scale modeling framework (Ma et al., 2022) for the true underlying network that relies on 2 different network sources, that is, the graph extracted from existing databases such as KEGG and the estimated network derived from predictors in the observed dataset. The proposed approach reflects our understanding that the 2 sources of data may contain complimentary information on the true graph with some degree of mis-specification and thus can be used to inform the true

graph. We note that while there are methods for integrating multiple sources of graphs (Xie et al., 2021; Danaher et al., 2014), they cannot be directly used for our purpose. For brevity, our model is denoted as NoiseSH in short. We demonstrate the advantages of our approach via simulation studies involving varying degrees of network mis-specification, where the proposed approach demonstrates superior prediction performance. The practical importance of the proposed approach is further showcased via application to a genomics dataset and another proteomics dataset for AD.

The organization of this paper is as follows. In Section 2, we present the proposed methodology and the Markov chain Monte Carlo (MCMC) algorithm. Section 3 reports the practical performance of our method through simulation studies. In Section 4, we apply our method to a genomics dataset and another proteomics dataset for AD. Finally, we conclude in Section 5 with a discussion of future directions for this work. The details of the computation framework and cost are relegated to Web Appendices A and B.

2 METHODOLOGY

In this section, we describe our proposed model and associated MCMC algorithm.

2.1 Model specification

Suppose we have data of n observations $\{y_i, \mathbf{x}_i : i = 1, \dots, n\}$, where y_i is the outcome variable and \mathbf{x}_i is a vector of p predictors. Let $G = (V, E)$ be the true underlying undirected graph for the p predictors, with vertices $V = \{1, \dots, p\}$ and a set E of edges, where elements of E are unordered pairs $\{j, k\}$ of distinct vertices $j, k \in V$. Let $\tilde{G} = (V, \tilde{E})$ be the working graph extracted from the database, where we implicitly assume that the vertex set V is known. Denote the adjacency matrix of G by $\mathbf{A} = (A_{jk})_{p \times p}$ and that of \tilde{G} by $\tilde{\mathbf{A}} = (\tilde{A}_{jk})_{p \times p}$. Hence, $A_{jk} = 1$ if there is a true edge between the predictor j and the predictor k , and 0 otherwise, while $\tilde{A}_{jk} = 1$ if there is an edge in the working graph \tilde{G} between the j th vertex and the k th vertex, and 0 otherwise.

In addition to extracting the working graph \tilde{G} directly from the database, we can also obtain the information on the true underlying graph from the p predictors $\mathbf{x}_i, i = 1, \dots, n$. Let Ω^* denote the precision matrix for predictors calculated by an inverse covariance matrix estimation algorithm such as graphical lasso (Friedman et al., 2008) and BigQUIC (Hsieh et al., 2013). Let $\mathbf{A}^* = (A_{jk}^*)_{p \times p}$ be the corresponding adjacency matrix, that is, $A_{jk}^* = 1$ if $\Omega_{jk}^* \neq 0$, and 0 otherwise. The corresponding graph for \mathbf{A}^* is denoted by $G^* = (V, E^*)$.

We consider the linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_n), \quad (1)$$

where $\mathbf{y} = (y_1, \dots, y_n)^\top$, $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^\top$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$, $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^\top$, \mathbf{I}_n is the $n \times n$ identity matrix, and $\mathcal{N}(\cdot)$ denotes the Gaussian distribution. We adopt the Bayesian shrinkage approach with the structural information incorporated in Chang et al. (2018) for setting priors for $\boldsymbol{\beta}$ and σ^2 . Following Chang et al. (2018), we specify

$$\beta_j \sim \mathcal{L}(0, \lambda_j / \sigma), \quad \sigma^2 \sim \text{IG}(a_\sigma, b_\sigma), \quad j = 1, \dots, p, \quad (2)$$

where λ_j is the shrinkage parameter for β_j , $\mathcal{L}(\cdot)$, and $\mathcal{IG}(\cdot)$ denote the Laplace and inverse gamma distributions, respectively. The shrinkage parameters λ_j are treated as random variables that encode the graph knowledge G under an informative network-based prior. Specifically, we have

$$\boldsymbol{\theta} = (\log(\lambda_1), \dots, \log(\lambda_p))^T \sim \mathcal{N}(\boldsymbol{\mu}, \nu \boldsymbol{\Xi}^{-1}), \quad (3)$$

where $\boldsymbol{\Xi}(j, k) = \boldsymbol{\Xi}(k, j) = -\xi_{jk} = -\xi_{kj}$ for $1 \leq j < k \leq p$, $\boldsymbol{\Xi}(j, j) = 1 + \sum_{k \neq j} \xi_{jk}$ for $1 \leq j \leq p$, and we assign the following prior to $\boldsymbol{\xi} = \{\xi_{jk} : 1 \leq j < k \leq p\}$

$$\begin{aligned} \pi(\boldsymbol{\xi}) \propto |\boldsymbol{\Xi}|^{-1/2} \prod_{A_{jk}=1} \xi_{jk}^{a_\xi - 1} \\ \times \exp(-b_\xi \xi_{jk}) 1(\xi_{jk} > 0) \prod_{A_{jk}=0} \delta_0(\xi_{jk}), \quad (4) \end{aligned}$$

where δ_0 is the Dirac delta function (concentrated at 0 and $1(\cdot)$ is the indicator function). Note that the diagonally dominant structure of $\boldsymbol{\Xi}$ ensures positive definiteness. The partial correlations among the log-shrinkage parameters are positive if the corresponding variables are connected, and 0 otherwise. This feature encourages a similar degree of shrinkage for the regression coefficients if the corresponding predictors are connected, and otherwise an conditionally independent shrinkage is imposed. The mean vector $\boldsymbol{\mu}$ controls the overall sparsity of the model, and ν specifies the prior confidence on the choice of $\boldsymbol{\mu}$. Moreover, the shape parameter a_ξ and the rate parameter b_ξ influence the magnitudes and variability of the partial correlations.

Motivated by the latent scale representation of networks in Ma et al. (2022), we specify the model for $\tilde{\mathbf{A}}$ as following: for $1 \leq j < k \leq p$,

$$\begin{aligned} \log \left\{ \frac{\mathbb{P}(\tilde{A}_{jk} = 1 | A_{jk} = 0)}{1 - \mathbb{P}(\tilde{A}_{jk} = 1 | A_{jk} = 0)} \right\} &= \tilde{s}_0 + \tilde{\mathbf{u}}_j \tilde{\mathbf{D}}_0 \tilde{\mathbf{u}}_{0k}^\top, \\ \log \left\{ \frac{\mathbb{P}(\tilde{A}_{jk} = 1 | A_{jk} = 1)}{1 - \mathbb{P}(\tilde{A}_{jk} = 1 | A_{jk} = 1)} \right\} &= \tilde{s}_1 + \tilde{\mathbf{u}}_j \tilde{\mathbf{D}}_1 \tilde{\mathbf{u}}_{1k}^\top, \quad (5) \end{aligned}$$

where \tilde{s}_0 and \tilde{s}_1 denote the intercepts common across edges, $\tilde{\mathbf{u}}_j = (\tilde{u}_{j,1}, \dots, \tilde{u}_{j,\tilde{r}})$ and $\tilde{\mathbf{u}}_j = (\tilde{u}_{j,1}, \dots, \tilde{u}_{j,\tilde{r}})$ are the vectors of latent scales for vertex j of length \tilde{r} , and $\tilde{\mathbf{D}}_0$ and $\tilde{\mathbf{D}}_1$ represent the $\tilde{r} \times \tilde{r}$ diagonal weight matrices having all positive elements, which control the contribution of the latent scales to the inner product. Since the latent scales are only identifiable up to a rotation in $\mathbb{R}^{\tilde{r}}$, we fix the first element of every latent scale (0.5 in our simulations and real data analysis) to preserve identifiability. Note that $\mathbb{P}(\tilde{A}_{jk} = 1 | A_{jk} = 0)$ can be interpreted as the probability of Type I error for a “test” of $H_0: \{j, k\} \notin E$ vs. $H_1: \{j, k\} \in E$, on vertex pair $\{j, k\} \notin E$. Similarly, $\mathbb{P}(\tilde{A}_{jk} = 1 | A_{jk} = 1)$ can be interpreted as the power for a test of the same hypothesis on vertex pair $\{j, k\} \in E$. Here, the intercept terms control the overall noise level of the working graph \tilde{G} , and the latent scales capture the variation in the edge noise.

Similarly, we specify a latent scale model for \mathbf{A}^* , that is, for $1 \leq j < k \leq p$,

$$\begin{aligned} \log \left\{ \frac{\mathbb{P}(A_{jk}^* = 1 | A_{jk} = 0)}{1 - \mathbb{P}(A_{jk}^* = 1 | A_{jk} = 0)} \right\} &= s_0^* + \mathbf{u}_{0j}^* \mathbf{D}_0^* \mathbf{u}_{0k}^{*\top}, \\ \log \left\{ \frac{\mathbb{P}(A_{jk}^* = 1 | A_{jk} = 1)}{1 - \mathbb{P}(A_{jk}^* = 1 | A_{jk} = 1)} \right\} &= s_1^* + \mathbf{u}_{1j}^* \mathbf{D}_1^* \mathbf{u}_{1k}^{*\top}, \quad (6) \end{aligned}$$

where s_0^* and s_1^* denote the intercepts common across edges, $\mathbf{u}_{0j}^* = (u_{0j,1}^*, \dots, u_{0j,r^*}^*)$ and $\mathbf{u}_{1j}^* = (u_{1j,1}^*, \dots, u_{1j,r^*}^*)$ are the vectors of latent scales for vertex j of length r^* , and \mathbf{D}_0^* and \mathbf{D}_1^* represent the $r^* \times r^*$ diagonal weight matrices having all positive elements. The distinct intercept terms and the latent scales in Equations 5 and 6 reflect our understanding that the 2 noisy graphs obtained from different sources are subject to different degrees of mis-specification.

We then assign the following priors to the intercept terms and the latent scales in the models for $\tilde{\mathbf{A}}$ and \mathbf{A}^* :

$$\begin{aligned} \tilde{s}_0 &\sim \mathcal{L}(\tilde{\mu}_{s,0}, 1/\tilde{\sigma}_{s,0}), \quad \tilde{s}_1 \sim \mathcal{L}(\tilde{\mu}_{s,1}, 1/\tilde{\sigma}_{s,1}), \\ s_0^* &\sim \mathcal{L}(\mu_{s,0}^*, 1/\sigma_{s,0}^*), \quad s_1^* \sim \mathcal{L}(\mu_{s,1}^*, 1/\sigma_{s,1}^*), \\ \begin{bmatrix} \tilde{\mathbf{u}}_{0j}^* \\ \tilde{\boldsymbol{\mu}}_0^* \end{bmatrix} &\sim \mathcal{N} \left(\begin{bmatrix} \tilde{\boldsymbol{\mu}}_0 \\ \boldsymbol{\mu}_0^* \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{\tilde{\mathbf{u}}_0 \tilde{\mathbf{u}}_0} & \boldsymbol{\Sigma}_{\tilde{\mathbf{u}}_0 \boldsymbol{\mu}_0^*} \\ \boldsymbol{\Sigma}_{\boldsymbol{\mu}_0^* \tilde{\mathbf{u}}_0} & \boldsymbol{\Sigma}_{\boldsymbol{\mu}_0^* \boldsymbol{\mu}_0^*} \end{bmatrix} \right), \quad j = 1, \dots, p, \\ \begin{bmatrix} \tilde{\mathbf{u}}_{1j}^* \\ \boldsymbol{\mu}_1^* \end{bmatrix} &\sim \mathcal{N} \left(\begin{bmatrix} \tilde{\boldsymbol{\mu}}_1 \\ \boldsymbol{\mu}_1^* \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{\tilde{\mathbf{u}}_1 \tilde{\mathbf{u}}_1} & \boldsymbol{\Sigma}_{\tilde{\mathbf{u}}_1 \boldsymbol{\mu}_1^*} \\ \boldsymbol{\Sigma}_{\boldsymbol{\mu}_1^* \tilde{\mathbf{u}}_1} & \boldsymbol{\Sigma}_{\boldsymbol{\mu}_1^* \boldsymbol{\mu}_1^*} \end{bmatrix} \right), \quad j = 1, \dots, p, \end{aligned}$$

where $\begin{bmatrix} \tilde{\boldsymbol{\mu}}_0 \\ \boldsymbol{\mu}_0^* \end{bmatrix}$ and $\begin{bmatrix} \tilde{\boldsymbol{\mu}}_1 \\ \boldsymbol{\mu}_1^* \end{bmatrix}$ are with sizes $\begin{bmatrix} \tilde{r} \times 1 \\ r^* \times 1 \end{bmatrix}$, and $\begin{bmatrix} \boldsymbol{\Sigma}_{\tilde{\mathbf{u}}_0 \tilde{\mathbf{u}}_0} & \boldsymbol{\Sigma}_{\tilde{\mathbf{u}}_0 \boldsymbol{\mu}_0^*} \\ \boldsymbol{\Sigma}_{\boldsymbol{\mu}_0^* \tilde{\mathbf{u}}_0} & \boldsymbol{\Sigma}_{\boldsymbol{\mu}_0^* \boldsymbol{\mu}_0^*} \end{bmatrix}$ and $\begin{bmatrix} \boldsymbol{\Sigma}_{\tilde{\mathbf{u}}_1 \tilde{\mathbf{u}}_1} & \boldsymbol{\Sigma}_{\tilde{\mathbf{u}}_1 \boldsymbol{\mu}_1^*} \\ \boldsymbol{\Sigma}_{\boldsymbol{\mu}_1^* \tilde{\mathbf{u}}_1} & \boldsymbol{\Sigma}_{\boldsymbol{\mu}_1^* \boldsymbol{\mu}_1^*} \end{bmatrix}$ are with sizes $\begin{bmatrix} \tilde{r} \times \tilde{r} & \tilde{r} \times r^* \\ r^* \times \tilde{r} & r^* \times r^* \end{bmatrix}$. The latent scales embedded in Equations 5 and 6 have some prior correlations (ie, $\boldsymbol{\Sigma}_{\tilde{\mathbf{u}}_0 \boldsymbol{\mu}_0^*}$ and $\boldsymbol{\Sigma}_{\tilde{\mathbf{u}}_1 \boldsymbol{\mu}_1^*}$), so the latent scales in the 2 models can be learned jointly.

By convention, we assume the true (non)edge A_{jk} follows a Bernoulli distribution with probability q and assign the beta distributed prior to q :

$$\begin{aligned} A_{jk} &\sim \text{Bern}(q), \quad 1 \leq j < k \leq p, \\ q &\sim \mathcal{B}(a_q, b_q), \end{aligned}$$

where $\text{Bern}(\cdot)$ denotes the Bernoulli distribution, and $\mathcal{B}(\cdot)$ denotes the Beta distribution.

2.2 MCMC algorithm

We developed an efficient MCMC algorithm for sampling the parameters in our model. To facilitate the sampling, we use 2 data augmentation techniques: replacing the Laplace prior with a scale mixture of normals and augmenting Pólya-Gamma latent variables for Bernoulli distributed variables A_{jk} and A_{jk}^* , $1 \leq j < k \leq p$. The log-shrinkage parameter $\boldsymbol{\theta}$ is sampled by the Metropolis–Hastings (MH) algorithm, all other parameters are sampled using the Gibbs sampler. The details of the computation framework can be found in Web Appendix A.

3 SIMULATION STUDIES

In this section, we conducted some simulations to evaluate the prediction performance of our model in comparison with several existing methods. We considered different graph scenarios, including correctly specified working graphs and noise-contaminated working graphs.

3.1 Simulation set-up

The simulated data were generated from the following model:

$$y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i, \quad 1 \leq i \leq n, \quad (7)$$

where $\mathbf{x}_i \sim \mathcal{N}(0, \Sigma_X)$ and $\epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon^2)$. We set $p = 150$. The number of nonzero coefficients in $\boldsymbol{\beta}$ is 15, and the true nonzero effect sizes were randomly generated from a uniform distribution with boundaries 0.5 and 2. The sample sizes for the training set, the validation set and test set are 70, 70, and 200, respectively. The residual variance was fixed at $\sigma_\epsilon^2 = 2$.

We employed the mechanism described in Chang et al. (2018) to generate \mathbf{A} and Σ_X , which resembles practical gene networks encountered in applications where a subset of important and unimportant genes lie in the same pathway and are connected. We used an estimated graph \mathbf{A}^* obtained by graphical lasso and a working graph $\tilde{\mathbf{A}}$ given to fit the model. The tuning parameter in graphical lasso was chosen by cross-validation. The working graph may be correctly specified or mis-specified, as follows:

- (1) $\tilde{\mathbf{A}} = \mathbf{A}$, where \mathbf{A} allows edges between important and unimportant variables.
- (2) $\tilde{\mathbf{A}} = \mathbf{A}$, where \mathbf{A} does not allow edges between important and unimportant variables.
- (3) \mathbf{A} is the same as in scenario (1), but $\tilde{\mathbf{A}}$ includes only a subset of edges in \mathbf{A} for which the corresponding partial correlations rank in the top 20%.
- (4) \mathbf{A} is the same as in scenario (2), but $\tilde{\mathbf{A}}$ includes only a subset of edges in \mathbf{A} for which the corresponding partial correlations rank in the top 20%.
- (5) \mathbf{A} is the same as in scenario (1), but $\tilde{\mathbf{A}}$ includes only a subset of edges in \mathbf{A} for which the corresponding partial correlations rank in the top 20%, and randomly generated edges such that the number of edges in $\tilde{\mathbf{A}}$ is the same as that in \mathbf{A} .
- (6) \mathbf{A} is the same as in scenario (2), but $\tilde{\mathbf{A}}$ includes only a subset of edges in \mathbf{A} for which the corresponding partial correlations rank in the top 20%, and randomly generated edges such that the number of edges in $\tilde{\mathbf{A}}$ is the same as that in \mathbf{A} .
- (7) \mathbf{A} is the same as in scenario (1), but $\tilde{\mathbf{A}}$ is randomly generated with the same number of edges as \mathbf{A} .
- (8) \mathbf{A} is the same as in scenario (2), but $\tilde{\mathbf{A}}$ is randomly generated with the same number of edges as \mathbf{A} .

Scenarios (1) and (2) are the cases where the true graph is completely known. Scenario (2) does not allow edges between important and unimportant variables and thus it is the ideal setting for approaches that encourage connected variables to be jointly included or excluded in the model. Scenarios (3) and (4) mimic the situation where only strong signals from the true

graph are known. Scenarios (5) and (6) are the cases where strong signals from the true graph are known and false edges exist. Scenarios (7) and (8) are considered as the worst cases, where the edges in the working graphs are completely random. Although each of the 8 scenarios consists of a distinct edge structure for the working graph, scenarios (1), (3), (5), and (7) have the same pathway membership information, as do scenarios (2), (4), (6), and (8).

We generated 50 data sets, each of which contains the training set, the validation set, and the test set. The prediction performance was assessed in terms of mean squared prediction error (MSPE) for the test data. We set $\hat{r} = r^* = 5$ in the latent scale models. We chose priors $\sigma^2 \sim \mathcal{IG}(1, 1)$, $(a_\xi, b_\xi) = (2, 1)$ for $\xi_{jk}, q \sim \mathcal{B}(1, 1)$, and $\tilde{\boldsymbol{\mu}}_0 = \tilde{\boldsymbol{\mu}}_1 = \boldsymbol{\mu}_0^* = \boldsymbol{\mu}_1^* = (0, 0, 0, 0, 0)^\top$, which are fairly uninformative. The remaining parameters were determined by choosing those with the smallest MSPE in the validation set. We ran MCMC chains of 10 000 samples. The number of burn-in samples is 1000. And we used the trace and autocorrelation plots to evaluate the samples.

3.2 Methods to compare

Our methods were compared to the lasso (Lasso), the adaptive Lasso (ALasso), the group Lasso (GLasso) (Yuan and Lin, 2006; Zeng and Breheny, 2016), the network-constrained regularization approach (Net) (Li and Li, 2008), the expectation-maximization approach for Bayesian variable selection (denoted as EMVS) proposed by Ročková and George (2014) and its extension to incorporate structural information (denoted as EMVSS), the expectation-maximization estimator for Bayesian shrinkage approach (denoted as EMSH) proposed by Chang et al. (2018) and its extension to incorporate structural information (denoted as EMSHS). Among these approaches, GLasso treats pathways as overlapping groups, while Net, EMSHS, and EMVSS incorporate the graph information. For Lasso and ALasso, we used the glmnet R package where the initial consistent estimator for ALasso was given by the ridge regression. We used the R packages gmpregOverlap (Zeng and Breheny, 2016) and glmgraph (Chen et al., 2015) for GLasso and Net, respectively. The unpublished R codes for EMVS and EMVSS were provided to us by the authors. For EMSH and EMSHS, we used the R package EMSHS (Chang et al., 2018). The tuning parameters in these methods were chosen by minimizing the MSPE in the validation set.

3.3 Results

The MSPEs for the test data are presented in Table 1. When the working graph $\tilde{\mathbf{A}}$ is correctly specified or misses some true edges (scenarios (1)-(4)), the structured variable selection methods EMVSS, EMSHS, and NoiseSH have superior prediction performance. When the working graph $\tilde{\mathbf{A}}$ contains false edges (scenarios (5)-(8)), NoiseSH outperforms all other methods. Therefore, NoiseSH accounts for the network noise when a non-trivial error is present. The MSPEs of NoiseSH in scenarios (3)-(8) (the working graph $\tilde{\mathbf{A}}$ is misspecified) are close, which implies that the prediction performance of NoiseSH is robust to the noise level of the working graph. Taken together, the above results suggest that combining 2 sources of network information to inform the underlying true network results in considerable pre-

TABLE 1 Mean of mean squared prediction errors for the test data.

Scenario	Lasso	ALasso	GLasso	Net	EMVS	EMVSS	EMSH	EMSHS	NoiseSH
(1)	3.977 (1.127)	3.975 (1.747)	5.556 (2.291)	5.095 (1.534)	3.583 (1.023)	3.012 (0.648)	3.541 (1.117)	2.898 (0.699)	2.841 (0.507)
(2)	3.939 (0.764)	3.654 (1.650)	5.476 (1.764)	4.844 (1.332)	3.544 (0.921)	2.990 (0.559)	3.537 (0.967)	2.870 (0.695)	2.878 (0.494)
(3)	3.977 (1.127)	3.975 (1.747)	5.556 (2.291)	5.301 (1.657)	3.583 (1.023)	3.491 (0.955)	3.541 (1.117)	3.448 (1.148)	3.425 (0.728)
(4)	3.939 (0.764)	3.654 (1.650)	5.476 (1.764)	5.213 (1.381)	3.544 (0.921)	3.428 (0.831)	3.537 (0.967)	3.432 (1.059)	3.420 (0.662)
(5)	3.977 (1.127)	3.975 (1.747)	5.556 (2.291)	5.562 (1.757)	3.583 (1.023)	3.642 (1.018)	3.541 (1.117)	3.618 (1.147)	3.434 (0.713)
(6)	3.939 (0.764)	3.654 (1.650)	5.476 (1.764)	5.327 (1.441)	3.544 (0.921)	3.616 (0.916)	3.537 (0.967)	3.607 (0.878)	3.432 (0.668)
(7)	3.977 (1.127)	3.975 (1.747)	5.556 (2.291)	5.894 (1.807)	3.583 (1.023)	3.674 (1.070)	3.541 (1.117)	3.751 (1.209)	3.449 (0.709)
(8)	3.939 (0.764)	3.654 (1.650)	5.476 (1.764)	5.545 (1.465)	3.544 (0.921)	3.670 (0.958)	3.537 (0.967)	3.686 (0.936)	3.446 (0.719)

In the parentheses are the standard deviations of the mean squared prediction errors.

TABLE 2 The average false negative rate (FNR), the average false positive rate (FPR), the average accuracy (accuracy) for edges in the graphs estimated by NoiseSH (\hat{A}), working graphs (\tilde{A}), and graphs obtained by graphical lasso (A^*).

Scenario	Graph	FNR	FPR	Accuracy
(1)	\hat{A}	0.052	0.000	1.000
	\tilde{A}	0.000	0.000	1.000
	A^*	0.431	0.073	0.925
(2)	\hat{A}	0.032	0.000	1.000
	\tilde{A}	0.000	0.000	1.000
	A^*	0.386	0.071	0.927
(3)	\hat{A}	0.805	0.000	0.994
	\tilde{A}	0.800	0.000	0.994
	A^*	0.431	0.073	0.925
(4)	\hat{A}	0.806	0.000	0.994
	\tilde{A}	0.801	0.000	0.994
	A^*	0.386	0.071	0.927
(5)	\hat{A}	0.936	0.000	0.993
	\tilde{A}	0.794	0.006	0.988
	A^*	0.431	0.073	0.925
(6)	\hat{A}	0.936	0.001	0.992
	\tilde{A}	0.797	0.006	0.989
	A^*	0.386	0.071	0.927
(7)	\hat{A}	0.999	0.001	0.992
	\tilde{A}	0.993	0.007	0.985
	A^*	0.431	0.073	0.925
(8)	\hat{A}	0.994	0.001	0.992
	\tilde{A}	0.994	0.007	0.986
	A^*	0.386	0.071	0.927

diction gains when the working graph is mis-specified, while performing comparably or better compared to existing structured regression approaches without noise correction in the presence of no or very limited mis-specification.

Table 2 displays the average false negative rate (FNR), the average false positive rate (FPR), the average accuracy (accuracy) for edges in the graphs estimated by NoiseSH (\hat{A}), working graphs (\tilde{A}), and graphs obtained by graphical lasso (A^*). By FNR (FPR), we mean the proportion of true edges (non-edges) being

identified as non-edges (edges). Accuracy is the fraction of correctly identified edges and non-edges. Combining the results in Tables 1 and 2 reveals some interesting observations. The graph \hat{A} estimated by our method NoiseSH yields the smallest FPR that is or close to 0, and highest accuracy in all settings. When the working graph \tilde{A} contains false edges (scenarios (5)-(8)), \tilde{A} may have higher FNR than \hat{A} and A^* , but our NoiseSH method still outperforms the competing methods in prediction as shown in Table 1, including the methods that use \tilde{A} with higher FPR and lower FNR. This suggests that including false edges (ie, higher FPR) has a greater deleterious impact on prediction than excluding true edges (ie, higher FNR) in structured regression models. Another notable implication is that our method only needs to learn a subset of true edges in \tilde{A} to yield better performance than the existing methods. This is an appealing property of NoiseSH, as graph estimation by itself is a very challenging problem that our method is not designed to address.

4 DATA APPLICATION

We applied the proposed method to analyses of 1 AD genomics dataset from the ADNI study (Mueller et al., 2005) and 1 AD proteomics dataset from the AMP-AD study (Hodes and Buckholtz, 2016).

4.1 Application to an ADNI study

The ADNI is a large scale multisite longitudinal study where researchers at 63 sites track the progression of AD in the human brain through the process of normal aging, early mild cognitive impairment, and late mild cognitive impairment to dementia or AD (Mueller et al., 2005). The goal of the study is to validate diagnostic and prognostic biomarkers that can predict the progress of AD. In our application, we investigated the association of patients' gene expression levels with their fluorodeoxyglucose-positron emission tomography (FDG-PET). The FDG-PET averaged over the regions of interest measures cell metabolism. The cells affected by AD tend to show reduced metabolism.

TABLE 3 Mean of mean squared prediction errors for the ADNI data set.

Method	MSPE
Lasso	1.028 (0.094)
ALasso	1.016 (0.090)
GLasso	1.028 (0.094)
Net	1.091 (0.107)
EMVS	1.028 (0.094)
EMVSS	1.039 (0.097)
EMSH	1.024 (0.099)
EMSHS	1.030 (0.086)
NoiseSH	0.953 (0.095)

In the parentheses are the standard deviations of the mean squared prediction errors. Abbreviations: ADNI, Alzheimer's disease neuroimaging initiative; MSPEs, mean squared prediction errors.

As a screening step, we first conduct univariate regression of each gene in relation to the FDG-PET and select genes with a false discovery rate of <0.05 for subsequent analyses. There are 374 selected genes. The gene network (\tilde{G}) was retrieved from the KEGG database (Kanehisa and Goto, 2000), including 41 edges. There are 675 subjects in the dataset. We randomly divided them into a training set, a validation set, and a testing set, with equal size. We used graphical lasso to estimate the precision matrix for predictors and obtained the corresponding graph A^* . The tuning parameter in graphical lasso was chosen by cross-validation. Then we fitted our method and the competing methods used in the simulation study and recorded the MSPE for the testing samples. The tuning parameters in the competing methods were determined by choosing those with the smallest MSPE in the validation set. For our proposed approach, we set $\tilde{r} = r^* = 5$ in the latent scale models, and chose priors $\sigma^2 \sim \mathcal{IG}(1, 1)$, $(a_\xi, b_\xi) = (2, 1)$ for ξ_{jk} , $q \sim \mathcal{B}(1, 1)$, and $\tilde{\mu}_0 = \tilde{\mu}_1 = \mu_0^* = \mu_1^* = (0, 0, 0, 0, 0)^\top$, which are fairly uninformative. The remaining parameters were chosen by minimizing the MSPE in the validation set. We repeated this procedure for 50 times.

Table 3 shows the MSPEs for the test set. NoiseSH yields the best prediction performance, which demonstrates the advantage of our method. In addition, we used the 95% credible intervals to select genes and conducted the pathway enrichment analysis via the ToppGene Suite (Chen et al., 2009). The selected genes, such as HDAC2 and ABCA7, and the enriched pathways, including ALK signaling, have been shown to be associated with AD (De Roeck et al., 2019; Mahady et al., 2019; Majumder et al., 2021). The list of selected genes and the details of the pathway enrichment analysis can be found in [Web Appendices C and D](#), respectively. Our data analyses demonstrate that NoiseSH yields biologically meaningful results.

4.2 Application to an AMP-AD study

The AMP-AD program is focused on discovering novel therapeutic targets for AD, understanding how these novel targets operate in gene, protein, and metabolic networks, and assessing the druggability of these novel targets (Hodes and Buckholtz, 2016). One of the studies included in the AMP-AD project is the Religious Orders Study (ROS), which is a longitudinal clinical-pathologic cohort study of aging and AD conducted by Rush

TABLE 4 Mean of mean squared prediction errors for the AMP-AD data set.

Method	MSPE
Lasso	1.001 (0.076)
ALasso	0.943 (0.120)
GLasso	1.001 (0.076)
Net	0.808 (0.076)
EMVS	1.001 (0.076)
EMVSS	0.751 (0.061)
EMSH	0.703 (0.065)
EMSHS	0.709 (0.073)
NoiseSH	0.687 (0.066)

In the parentheses are the standard deviations of the mean squared prediction errors. Abbreviations: AMP-AD, accelerating medicines partnership-Alzheimer's disease; MSPEs, mean squared prediction errors.

University that enrolled individuals from religious communities for longitudinal clinical analysis and brain donation (Bennett et al., 2012). We used the proteomics data set from ROS to investigate the association of individuals' protein levels with their mini-mental state examination (MMSE) scores. The MMSE is a 30-point questionnaire that is used extensively in clinical and research settings to measure cognitive impairment (Pangman et al., 2000).

There are 61 proteins included in our analysis. The protein-protein interaction network (\tilde{G}) and pathway information were obtained from the STRING database, including 105 edges and 24 pathways. We analyzed 539 subjects with no cognitive impairment, mild cognitive impairment, or AD. The subjects were randomly divided into a training set, a validation set, and a testing set, with equal size. The graphical lasso was used to estimate the precision matrix for predictors and obtain the corresponding graph A^* . The tuning parameter in graphical lasso was chosen by cross-validation. Then we fitted our model and the competing methods used in the simulation study. The tuning parameters in the competing methods were determined by choosing those with the smallest MSPE in the validation set. For our proposed approach, we set $\tilde{r} = r^* = 5$ in the latent scale models, and chose priors $\sigma^2 \sim \mathcal{IG}(1, 1)$, $(a_\xi, b_\xi) = (2, 1)$ for ξ_{jk} , $q \sim \mathcal{B}(1, 1)$, and $\tilde{\mu}_0 = \tilde{\mu}_1 = \mu_0^* = \mu_1^* = (0, 0, 0, 0, 0)^\top$, which are fairly uninformative. The remaining parameters were chosen by minimizing the MSPE in the validation set. We recorded the MSPE for the testing samples. This procedure was repeated for 50 times.

Table 4 shows the MSPEs for the test set. NoiseSH leads to the smallest MSPE, which shows the prediction advantage of our approach. In addition, we used the 95% credible intervals to select proteins. The selected proteins are encoded by genes including CD47 and STX1A, which have been shown to be associated with AD (Gheibihayat et al., 2021; Williams et al., 2021). Using the related genes, a pathway enrichment analysis identified a number of enriched pathways such as serotonin neurotransmitter release cycle and dopamine neurotransmitter release cycle, which have been shown to be associated with AD (Meltzer et al., 1998; Pan et al., 2019). Our data analyses show that NoiseSH yields biologically meaningful results.

5 DISCUSSION

We have developed a Bayesian framework for predicting an outcome in structured high-dimensional settings. Our approach uses 2 sources of network information, namely, the noisy graph extracted from database and the graph estimated for predictors in the data at hand to inform the latent true graph. Our simulation experiments and analyses of 2 real datasets demonstrate the superiority of our proposed method over several existing methods. To relieve the computation burden, we recommend users to select covariates that are most correlated with the outcome before applying our method.

We have pursued a Bayesian approach to the problem of accounting for network noise in graph-guided statistical learning. If network replicates are available, a frequentist approach offers a natural alternative. A starting point could be to employ method-of-moments techniques to estimate network noise under a “signal plus noise” model (Chang et al., 2022) and constructing an unbiased estimator for the degree of each node (Li et al., 2021). The unbiased estimators of degrees can then be used in the network-constrained regularization criterion, for example, the network-constrained regularization for linear model in Li and Li (2008).

Our work here sets the stage for extensions to generalized linear models for general types of outcomes, such as binary or categorical. We could also extend our approach to models beyond regression, including support vector machines and linear discriminant analysis. Another future direction is to extend our adaptive structured shrinkage prior to more general classes of priors on the shrinkage parameters, which would translate to more diverse penalties.

ACKNOWLEDGMENTS

The complete ADNI Acknowledgment is available at http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

The ROS proteomics data were provided through the Accelerating Medicine Partnership for AD (U01AG046161 and U01AG061357) based on samples provided by the Rush Alzheimer’s Disease Center, Rush University Medical Center, Chicago. Data collection was supported through funding by National Institute on Aging grants P30AG10161, R01AG15819, R01AG17917, R01AG30146, R01AG36836, U01AG32984, and U01AG46152, the Illinois Department of Public Health, and the Translational Genomics Research Institute (Higginbotham et al., 2020).

SUPPLEMENTARY MATERIALS

Supplementary material is available at *Biometrics* online.

Web Appendices referenced in Sections 1, 2, and 4, and the code for implementing the proposed method are available with this paper at the Biometrics website on Oxford Academic. The code is also available at <https://github.com/wenrui-li/NoiseSH>.

FUNDING

This work was supported in part by National Institutes of Health grants (RF1 AG063481 and R01 AG071174). This work was also supported by the National Institute of Mental Health award R01 MH120299. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

CONFLICT OF INTEREST

None declared.

DATA AVAILABILITY

The data that support the findings in this paper are available in the Alzheimer’s disease neuroimaging initiative database at <https://adni.loni.usc.edu> and the Accelerating Medicines Partnership Alzheimer’s Disease Knowledge Portal hosted on Synapse at <https://www.synapse.org> with Data ID syn3219045.

REFERENCES

- Ahmed, N. K., Neville, J. and Kompella, R. (2014). Network sampling: from static to streaming graphs. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 8, 7.
- Balachandran, P., Kolaczyk, E. D. and Viles, W. D. (2017). On the propagation of low-rate measurement error to subgraph counts in large networks. *The Journal of Machine Learning Research*, 18, 2025–2057.
- Bennett, D. A., Schneider, J. A., Arvanitakis, Z. and Wilson, R. S. (2012). Overview and findings from the religious orders study. *Current Alzheimer Research*, 9, 628–645.
- Chang, C., Kundu, S. and Long, Q. (2018). Scalable Bayesian variable selection for structured high-dimensional data. *Biometrics*, 74, 1372–1382.
- Chang, J., Kolaczyk, E. D. and Yao, Q. (2022). Estimation of subgraph densities in noisy networks. *Journal of the American Statistical Association*, 117, 361–374.
- Chatterjee, S. (2015). Matrix estimation by universal singular value thresholding. *The Annals of Statistics*, 43, 177–214.
- Chen, J., Bardes, E. E., Aronow, B. J. and Jegga, A. G. (2009). Toppgene suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Research*, 37, W305–W311.
- Chen, L., Liu, H., Kocher, J.-P. A., Li, H. and Chen, J. (2015). glmgraph: an R package for variable selection and predictive modeling of structured genomic data. *Bioinformatics*, 31, 3991–3993.
- Danaher, P., Wang, P. and Witten, D. M. (2014). The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 76, 373–397.
- De Roeck, A., Van Broeckhoven, C. and Sleegers, K. (2019). The role of ABCA7 in Alzheimer’s disease: evidence from genomics, transcriptomics and methylomics. *Acta Neuropathologica*, 138, 201–220.
- Friedman, J., Hastie, T. and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9, 432–441.
- Gheibihayat, S. M., Cabezas, R., Nikiforov, N. G., Jamialahmadi, T., Johnston, T. P. and Sahebkar, A. (2021). CD47 in the brain and neurodegeneration: an update on the role in neuroinflammatory pathways. *Molecules*, 26, 3943.
- Higginbotham, L., Ping, L., Dammer, E. B., Duong, D. M., Zhou, M., Gearing, M. et al. (2020). Integrated proteomics reveals brain-based cerebrospinal fluid biomarkers in asymptomatic and symptomatic Alzheimer’s disease. *Science Advances*, 6, eaaz9360.
- Hodes, R. J. and Buckholtz, N. (2016). Accelerating medicines partnership: Alzheimer’s disease (AMP-AD) knowledge portal aids

- Alzheimer's drug discovery through open data sharing. *Expert Opinion on Therapeutic Targets*, 20, 389–391.
- Hsieh, C.-J., Sustik, M. A., Dhillon, I. S., Ravikumar, P. K. and Poldrack, R. (2013). Big & quic: sparse inverse covariance estimation for a million variables. *Advances in Neural Information Processing Systems*, 26, 3165–3173.
- Jiang, X., Gold, D. and Kolaczyk, E. D. (2011). Network-based auto-probit modeling for protein function prediction. *Biometrics*, 67, 958–966.
- Jiang, X. and Kolaczyk, E. D. (2012). A latent eigenprobit model with link uncertainty for prediction of protein–protein interactions. *Statistics in Biosciences*, 4, 84–104.
- Kanehisa, M. and Goto, S. (2000). Kegg: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 28, 27–30.
- Kolaczyk, E. D. (2009). *Statistical Analysis of Network Data*, New York, NY: Springer.
- Le, C. M. and Li, T. (2020). *Linear regression and its inference on noisy network-linked data*, arXiv, arXiv:2007.00803, preprint.
- Li, C. and Li, H. (2008). Network-constrained regularization and variable selection for analysis of genomic data. *Bioinformatics*, 24, 1175–1182.
- Li, W., Sussman, D. L. and Kolaczyk, E. D. (2021). *Causal inference under network interference with noise*, arXiv, arXiv:2105.04518, preprint.
- Ma, X., Kundu, S. and Stevens, J. (2022). Semi-parametric bayes regression with network-valued covariates. *Machine Learning*, 111, 3733–3767.
- Mahady, L., Nadeem, M., Malek-Ahmadi, M., Chen, K., Perez, S. E. and Mufson, E. J. (2019). HDAC 2 dysregulation in the nucleus basalis of Meynert during the progression of Alzheimer's disease. *Neuropathology and Applied Neurobiology*, 45, 380–397.
- Majumder, P., Chanda, K., Das, D., Singh, B. K., Chakrabarti, P., Jana, N. R. et al. (2021). A nexus of mir-1271, PAX4 and ALK/RYK influences the cytoskeletal architectures in Alzheimer's disease and type 2 diabetes. *Biochemical Journal*, 478, 3297–3317.
- Meltzer, C. C., Smith, G., DeKosky, S. T., Pollock, B. G., Mathis, C. A., Moore, R. Y. et al. (1998). Serotonin in aging, late-life depression, and Alzheimer's disease: the emerging role of functional imaging. *Neuropsychopharmacology*, 18, 407–430.
- Mueller, S. G., Weiner, M. W., Thal, L. J., Petersen, R. C., Jack, C., Jagust, W. et al. (2005). The Alzheimer's disease neuroimaging initiative. *Neuroimaging Clinics*, 15, 869–877.
- Pan, X., Kamanga, A. C., Wen, S. W., Wu, X., Acheampong, K. and Liu, A. (2019). Dopamine and dopamine receptors in Alzheimer's disease: a systematic review and network meta-analysis. *Frontiers in Aging Neuroscience*, 11, 175.
- Pangman, V. C., Sloan, J. and Guse, L. (2000). An examination of psychometric properties of the mini-mental state examination and the standardized mini-mental state examination: implications for clinical practice. *Applied Nursing Research*, 13, 209–213.
- Priebe, C. E., Sussman, D. L., Tang, M. and Vogelstein, J. T. (2015). Statistical inference on errorfully observed graphs. *Journal of Computational and Graphical Statistics*, 24, 930–953.
- Ročková, V. and George, E. I. (2014). EMVS: the em approach to Bayesian variable selection. *Journal of the American Statistical Association*, 109, 828–846.
- Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S. et al. (2021). The string database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Research*, 49, D605–D612.
- Tsimring, L. S. (2014). Noise in biology. *Reports on Progress in Physics*, 77, 026601.
- Williams, J. B., Cao, Q. and Yan, Z. (2021). Transcriptomic analysis of human brains with Alzheimer's disease reveals the altered expression of synaptic genes linked to cognitive deficits. *Brain Communications*, 3, fcab123.
- Xie, S., Zeng, D. and Wang, Y. (2021). Integrative network learning for multi-modality biomarker data. *The Annals of Applied Statistics*, 15, 64–87.
- Young, J.-G., Cantwell, G. T. and Newman, M. (2020). *Robust Bayesian inference of network structure from unreliable data*, arXiv, arXiv:2008.03334, preprint.
- Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68, 49–67.
- Zeng, Y. and Breheny, P. (2016). Overlapping group logistic regression with applications to genetic pathway selection. *Cancer Informatics*, 15, CIN–S40043.
- Zhao, Y., Chang, C. and Long, Q. (2019). Knowledge-guided statistical learning methods for analysis of high-dimensional-omics data in precision oncology. *JCO Precision Oncology*, 3, 1–9.
- Zhao, Y., Chung, M., Johnson, B. A., Moreno, C. S. and Long, Q. (2016). Hierarchical feature selection incorporating known and novel biological information: identifying genomic features related to prostate cancer recurrence. *Journal of the American Statistical Association*, 111, 1427–1439.