Bayesian Edge Regression in Undirected Graphical Models to Characterize Interpatient Heterogeneity in Cancer

Zeya Wang\textsuperscript{a,b}, Veerabhadran Baladandayuthapani\textsuperscript{c}, Ahmed O. Kaseb\textsuperscript{d}, Hesham M. Amin\textsuperscript{e}, Manal M. Hassan\textsuperscript{f}, Wenyi Wang\textsuperscript{g}, and Jeffrey S. Morris\textsuperscript{h}

\textsuperscript{a}Department of Statistics, Rice University, Houston, TX; \textsuperscript{b}Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX; \textsuperscript{c}Department of Biostatistics, University of Michigan, Ann Arbor, MI; \textsuperscript{d}Department of Gastrointestinal Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX; \textsuperscript{e}Department of Hematopathology, The University of Texas MD Anderson Cancer Center, Houston, TX; \textsuperscript{f}Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX; \textsuperscript{g}Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, Philadelphia, PA; \textsuperscript{h}Department of Statistics, Rice University, Houston, TX; \textsuperscript{i}Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX; \textsuperscript{j}Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX; \textsuperscript{k}Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, Philadelphia, PA

ABSTRACT

It is well established that interpatient heterogeneity in cancer may significantly affect genomic data analyses and in particular, network topologies. Most existing graphical model methods estimate a single population-level graph for genomic or proteomic network. In many investigations, these networks depend on patient-specific indicators that characterize the heterogeneity of individual networks across subjects with respect to subject-level covariates. Examples include assessments of how the network varies with patient-specific prognostic scores or comparisons of tumor and normal graphs while accounting for tumor purity as a continuous predictor. In this article, we propose a novel edge regression model for undirected graphs, which estimates conditional dependencies as a function of subject-level covariates. We evaluate our model performance through simulation studies focused on comparing tumor and normal graphs while adjusting for tumor purity. In application to a dataset of proteomic measurements on plasma samples from patients with hepatocellular carcinoma (HCC), we ascertain how blood protein networks vary with disease severity, as measured by HepatoScore, a novel biomarker signature measuring disease severity. Our case study shows that the network connectivity increases with HepatoScore and a set of hub proteins as well as important protein connections are identified under different HepatoScore, which may provide important biological insights to the development of precision therapies for HCC.

1. Introduction

The proliferation of new technologies that can simultaneously measure genetic, transcriptomic, and proteomic markers have revolutionized biomedical research and contributed to the advent of precision therapy, whereby medical treatment strategies are tailored to individual patients on the basis of molecular characteristics of their disease. While certain individual genes have key biological roles in healthy and/or diseased cells, molecular processes relevant to the functional behavior of multicellular organisms or complex diseases are not determined by individual genetic factors, but rather complex interactions of various molecules at various molecular resolution levels. Network Biology is a nascent and burgeoning subfield of systems biology that involves the discovery and characterization of molecular interactions underlying complex diseases, including cancer. Graphical models, which characterize the conditional dependency structure among random variables, are widely used in genomic studies to build networks representing interactions among different biological units, including genes and proteins.

There has been a great deal of work on graphical models over the past decade. One model class shown to be useful for discovering biological networks is the undirected graphical model, for which nodes index random variables and edges connecting nodes represent the global conditional dependency structure among the variables (Lauritzen 1996). A popular tool in studying undirected graphs is the Gaussian graphical model, for which conditional independence, the absence of an edge, corresponds to a zero entry in the precision (or concentration) matrix of multivariate Gaussian distribution (Dempster 1972), which also attracts growing interest in the recent development of distributed statistical learning (Lee et al. 2017).

However, most of the graphical model work in existing literature involves estimation of a single network for a population, while inter-patient heterogeneity in many complex diseases, including cancer, suggests that these networks may vary across patients. Characterization of this heterogeneity has the potential to reveal insights into the differences in molecular processes across patients that can lead to the discovery of novel precision therapy strategies. One way of characterizing inter-
patient network heterogeneity is to assess how the networks vary across patient-level covariates. Two specific examples that have motivated this work include tumor purity and prognostic indices explaining inter-patient heterogeneity in cancer.

1.1. Accounting for Tumor Purity in Biological Networks

Tumor samples are inherently heterogeneous, with different types of cells present in a clinically derived sample, potentially confounding, to a large extent, the downstream analysis of gene expression or protein profiling of solid tumors (Farley 2015; Juntila and de Sauvage 2013). In practice, tumor samples invariably contain some contaminating normal tissue, and the proportion of a sample that is pure tumor, called tumor purity, varies from sample to sample. For this reason, many of the standard tumor versus normal comparisons are biased, typically attenuated, because they do not adjust for tumor purity and assume that the tumor samples are pure tumors. This principle also holds true in more advanced analyses including gene or protein networks, as any comparison of normal and tumor networks would be similarly biased by this factor. Deconvolution models such as DeMixT (Wang et al. 2018) can be fit to molecular data for tumor and normal samples in order to obtain an estimate of the tumor purity, $\pi_i$ for each sample $i = 1, \ldots, N$. Including this measurement as a continuous covariate in a graph regression model enables estimation of pure normal and pure tumor networks in a way that adjusts for this heterogeneous contamination.

1.2. Differential Biological Networks by Severity of Disease

The characterization of inter-patient heterogeneity within a given cancer can contribute to new precision therapy strategies. For example, important biological mechanisms can be revealed by assessments of how various gene-gene or protein-protein networks strengthen or weaken with advancing disease. In hepatocellular carcinoma (HCC), Morris et al. (2020) has developed a novel prognostic signature computed from a patient’s plasma protein profile called the HepatoScore. The HepatoScore for a given patient is a score $\pi_i \in [0, 1]$ that quantifies the degree of aberration in the patient’s blood protein profile relative to healthy subjects, with $\pi \approx 0$ indicating a protein profile essentially no different from a healthy subject, $\pi \approx 1$ indicating that patient’s profile is maximally aberrant, and $\pi$ in between (e.g., $\pi \approx 0.5$) indicating a moderate level of aberration relative to healthy subjects. Although determined without consideration of any patient-level clinical factors (i.e., unsupervised), HepatoScore has demonstrated remarkable prognostic separability (low/medium/high HepatoScore with median survival of 38.2/18.3/7.1 months) in a set of 767 HCC patients. This biological score contains more prognostic information than standard factors such as metastasis or nodal involvement, and provides a significant refinement of existing staging systems, for example, with metastatic HCC patients with low HepatoScore having substantially better prognosis than non-metastatic HCC patients with high HepatoScores. HepatoScore can be shown to be driven by a number of key proteins, including some in key pathways relevant to HCC such as growth hormone (GH), angiogenesis, and immune response. By modeling how protein networks vary across the continuous covariate HepatoScore, we can assess which protein-protein connections characterize advanced disease and provide molecular insights into this inter-patient heterogeneity.

1.3. Literature on Heterogeneous Graphical Models

There are a number of papers in the existing literature on heterogeneous graphs, but none that precisely solves the problem underlying our motivating examples. There are a number of papers on “group graphs”, in which graphical models are jointly estimated for discretized groups of subjects in a way that estimates group-specific graphs, while also borrowing strength between groups on common edges. Two-sample inference can be used to test differential edges between groups. Xia et al. (2015) developed a multiple testing procedure to detect gene-by-gene interactions with binary traits, while Narayan et al. (2015) proposed a novel resampling, random penalization, and random effects method for testing to identify the functional brain connections between two groups from neuroimages. Many other works focus on more than two groups (Guo et al. 2011; Danaher et al. 2014; Cai et al. 2016; Liu et al. 2017; Saegusa and Shojaie 2016). Liu et al. (2017) extended the two-sample test to capture the structural similarities and differences among multiple Gaussian graphical models. Guo et al. (2011) jointly estimated multiple graphical models by incorporating a hierarchical penalty for common factors and group-specific factors. Danaher et al. (2014) developed a more general model by employing fused lasso or group lasso penalties to encourage shared edges across the estimated precision matrices. Saegusa and Shojaie (2016) applied a Laplacian shrinkage penalty to encourage similarity among estimates from related subpopulations, and further proposed a Laplacian penalty based on hierarchical clustering for unknown population structures. Most recently, Bayesian approaches have become popular in modeling group graphs to construct the differential biological networks (Peterson et al. 2015; Tan et al. 2017; Mitra et al. 2016; Lin et al. 2017). Peterson et al. (2015) utilized a Markov random field prior for encouraging the common structure between different groups. Tan et al. (2017) investigated metabolic associations with the effect of cadmium through inducing multiplicative priors on the graphical space. Lin et al. (2017) proposed a Bayesian neighborhood selection method that jointly estimates multiple Gaussian graphical models for data with both spatial and temporal structure through naturally incorporating this structure. Motivated by the progress made for the joint estimation of multiple graphs, there is a growing development to have heterogeneous graphical models with a more relaxed assumption for observations (Yang et al. 2014; Hao et al. 2017). Hao et al. (2017) proposed a method to learn a cluster structure of data while estimating multiple graphical models that does not need to specify the membership of observations. Yang et al. (2014) developed a class of mixed graphical models, in which each node-conditional distribution with a graphical model belongs to a possibly different univariate
exponential family, therefore allowing random variables to be from heterogeneous domain sets for complex data. While groups can be viewed as categorical covariates, these methods do not model graphical variation across the continuous covariates of primary interest as in our motivating examples and the present paper.

Other methods model heterogeneity across covariates with graphs or covariance matrices, but are not suitable for our setting. Hoff and Niu (2012) proposed a covariance regression model that regresses a covariance matrix on a set of explanatory variables using a factor analysis. Zou et al. (2017) studied different estimators to parameterize the covariance matrix as a function of predictors. Cai et al. (2012) proposed a covariate-adjusted Gaussian graphical model that regresses a $p$-dimensional vector of responses on a $q$-dimensional vector of covariates, but the precision matrix does not depend on predictors, so this method only evaluates how the nodes change with covariates, but not node-to-node dependencies. Zhou et al. (2010) and Kolar and Xing (2009) developed dynamic undirected graph models varying with time. Cheng et al. (2014) modeled multivariate binary data using an Ising model to study the change of dependency with covariates. While some of these designs model covariance heterogeneity, these methods either do not provide node-specific inference, deal with covariance rather than precision matrices, or cannot be applied to general regression settings with multiple covariates for Gaussian graphical models.

Liu et al. (2010) proposed Graph-optimized classification and regression trees to partition the covariate space and estimate the graph within each partitioned subspace. While quite flexible, as reported by Cheng et al. (2014), this model lacks interpretation of the graphical model and covariates, and it has the undesirable property that graphs constructed for covariate values close to each are not necessarily similar. A machine learning method proposed by Kolar et al. (2010) applied a penalized kernel smoothing approach and allowed the precision matrix to change with covariates. One weakness of this method is that it ignores the intrinsic symmetry of the precision matrix, which may result in contradictory, unclear results in neighborhood selection and subsequent interpretation. Similarly using a kernel regression-based approach, Lee and Xue (2018) proposed another covariate dependent graphical model that utilizes a nonparametric mixture of Gaussian graphical models with a single scalar covariate controlling mixture probability and distribution. The finite mixture model effectively clusters the subjects into discrete subgroups based on a partitioning of the covariate space, and then estimates separate graphs for each partition point. This method shares similar limitations as found with kernel-based methods: it lacks a clear interpretation of the change of graph structure with respect to the covariates. Plus, more fundamentally, it can only handle a single covariate, not multiple covariates as in a general graph regression modeling framework as we develop in this paper. Ni et al. (2018) constructed Bayesian graphical regression models for directed acyclic graphs (DAG), which enable directed graphs to vary with general covariates, but their approach does not work in the undirected graph setting, which poses additional challenging difficulties and is our primary interest here. To our knowledge, none of the existing literature has considered building a regression model for edges in undirected graphs allowing general linear model-based effects and multiple covariates, whose development is the primary goal of this manuscript.

1.4. Outline

In this article, we present a Bayesian method to perform edge regression for undirected graphical models. We define edge-specific conditional precision functions that allow the edge strengths of an undirected graphical model to vary with extra-aneous covariates. We estimate these elements of the precision matrix using a joint regression model while constraining the elements corresponding to a given node to be the same, and thus we guarantee symmetry in the corresponding edges of the precision matrix. We induce sparsity on both the edges and covariates through Bayesian global-local priors that introduce nonlinear shrinkage, after which posterior edge selection occurs to generate predicted graphs for given sets of covariates while accounting for multiple testing across edges and covariates using Bayesian false discovery rate (FDR) considerations. We demonstrate the performance of this method in a simulation study in the context of estimating gene networks that are specific for the tumor and stromal components by accounting for proportions of the two components in the observed mixed data, and by application to an HCC case study in which we assess heterogeneity of protein networks across the prognostic index HepatoScore.

The rest of the article is structured as follows. In Section 2, we provide a formal description of edge regression with several theoretical properties for undirected graphical models. Then we present our models with sampling scheme and posterior inference technique. We present our simulation studies in Section 3 and our HCC case study in Section 4. Section 5 contains a discussion and conclusions.

2. Methods

2.1. Edge Regression

A graphical model for a random $p$-vector $Y$ is defined by a tuple $\mathcal{G}_Y = (G, \mathcal{P}(Y))$, where $G$ is a graph and $\mathcal{P}(Y)$ denotes its associated distribution. $G = (V, E)$ represents a conditional independence structure among random variables by specifying a set of nodes $V = 1, 2, 3, \ldots, p$ and a set of edges $E \subseteq V \times V$. In this work, our intended focus for application is on moderated-sized graphs $G$ with nodes $V$ in the dozens to greater than 100 or so and thus edges $E$ from hundreds to thousands, which is useful in practice in studying genetic pathways, many of which are on that order of magnitude. Each node in graph $G$ corresponds to a random variable in $Y$. In an undirected graph, we have undirected edges $E$, where $(i, j) \in E$ if and only if $(j, i) \in E$. For example, a Gaussian graphical model is defined by assuming $\mathcal{P}(Y)$ is a Gaussian distribution with mean $\mu \in \mathbb{R}^p$ and covariance matrix $\Sigma \in \mathbb{R}^{p \times p}$. $Y_n \sim \mathcal{N}(\mu, \Sigma^{-1}), n = 1, \ldots, N$, where $Y_n$ is the observed data and $\Omega = \Sigma^{-1} \in \mathbb{R}^{p \times p}$ is the inverse covariance matrix (a.k.a., precision matrix or concentration matrix). In a Gaussian graphical model, $\Omega$ is a $p \times p$ symmetric positive definite matrix with elements $(\omega_{ij})$. If $\omega_{ii} = 0$, then the random variables $i$ and $j$ are conditionally
independent given all the other variables of \( Y \), which indicates that there is no edge in \( G \) between nodes \( i \) and \( j \). Therefore the conditional independence structure of graph \( G \) can be inferred from models for the precision matrix \( \Omega \), which is well-known as the covariance selection model.

In our proposed edge regression model, given another \( q \)-dimensional random vector \( X = (x_1, \ldots, x_q)^\top \), we consider \( \mathcal{G}_Y(X) = \{G(X), \mathcal{P}(Y|X)\} \), and the precision matrix for each observation \( Y_n \) given \( X = x_n \) is a function of \( X \), allowing the conditional independence structure to vary from observation to observation across different realizations of \( X \). In the following discussion, we use the term extraneous covariates to define \( X \). We denote the precision matrix dependent on \( X \) through \( \Omega(x) \) with elements \( \omega_{ij}(X) \). Here, we focus on linear assumptions in \( X \), leaving extensions to nonparametric representations to future work. Ni et al. (2018) shows a functional pairwise Markov property for directed acyclic graphs, which implies the pairwise Markov property still holds if given the covariates when modeling the graph \( \mathcal{G}_Y \) with \( \mathcal{P}(Y) \) as a function of the external covariates \( X \). Similarly, we have the following lemma for functional covariance selection that is used to represent edge regression for the covariance selection problem.

**Lemma 1.** (Functional Covariance Selection Rule) Assume \( Y \) has a multivariate Gaussian distribution given extraneous covariates \( X \) with a precision matrix \( \Omega(X) \). \( Y^i \perp \!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!

\[ \rho_{ij}(x) = \frac{\omega_{ij}(x)}{\sqrt{\omega_{ii}(x)\omega_{jj}(x)}} \]
As we previously mentioned, only $\omega^{ij}$ is assumed to vary across $X$ for focusing on learning the dynamic sparsity structure of off-diagonal elements and reducing the computational complexity, so $\theta^{ij}$ and $\epsilon^{ij}$ share the same scaling parameter $\omega^{ii}$. Since $\omega^{ij}()$ is the off-diagonal element of precision matrix corresponding to vertex $i$ and vertex $j$, we have $\omega^{ij}() = \omega^{ij}()$. Hence, we will constrain these two functions to be identical in the sampling scheme when we jointly perform these regressions. Consequently, we have $\beta_{s}^{ij} = \beta_{s}^{ij}$ for every $i \neq j$. We rewrite the full conditional probability of $Y_{s}$ as follows:

$$Y_{s} | Y^{-i}, \{\beta^{k}_{s} \}_{k=1}^{i-1}, \omega^{ij}, \{X_{s} \}_{i=1}^{q} \sim N \left( - \frac{\sum_{j \neq i}^{q} \sum_{i=1}^{q} \beta_{s}^{ij} X_{s} y_{j}}{\omega^{ij}}, \frac{1}{\omega^{ij}} \right) \quad (3)$$

### 2.4. Bayesian Adaptive Shrinkage

It has been widely observed that genomic and proteomic graphs tend to be sparse, and as previously discussed, the sparsity of a graph corresponds to sparsity in the estimated precision matrix given extraneous covariates. We will induce sparsity in the subject-specific precision matrix using a Bayesian approach involving shrinkage priors on the coefficients corresponding to the off-diagonal precision matrix elements.

The spike-slab prior (Mitchell and Beauchamp 1988), consisting of a mixture of a spike at 0 and a continuous slab, is a popular choice as a Bayesian sparsity prior. It provides true zero estimates for some variables in the model, yielding automatic edge selection in our graphical setting, plus it has some desirable theoretical properties (Johnstone et al. 2004; Scott and Berger 2010; Narisetty et al. 2014; Castillo et al. 2012). However, in high-dimensional settings involving a large number of variables or, as in our setting, a large number of potential graph edges, this prior can have computational problems in searching the enormously large underlying state space. Another alternative is to use global–local priors (Polson and Scott 2010) that involve scale mixtures of normals. These priors are absolutely continuous, making them computationally easy to work with even in high dimensional settings, and with a shape that induces a type of nonlinear shrinkage in which small magnitude coefficients shrink strongly toward zero, while large magnitude coefficients are left largely unaffected. As described below in Section 2.5, this nonlinear shrinkage effectively induces a type of sparsity in the graph edges, and posterior selection rules can be used to induce true zeros in the estimated graph structure for specific covariate levels (Polson and Scott 2010).

There are many potential global–local prior choices, including the Bayesian Lasso (Park and Casella 2008), Horseshoe (Carvalho et al. 2010), Dirichlet Laplace (Bhattacharya et al. 2015), Normal-Exponential-Gamma (Griffin and Brown 2011), and Normal-Gamma priors (Griffin and Griffin 2010). Here, we will use the Normal-Gamma prior, which has been shown to have outstanding sparsity properties (Griffin et al. 2010). This distribution is indexed by two parameters that together provide useful flexibility in capturing varying degrees of sparsity and heavy-tails in the distribution of coefficients. Furthermore, there are efficient Gibbs sampling schemes available for the Normal-Gamma prior (Griffin et al. 2010). Specifically, assuming $\omega^{ij} = \sum_{s=1}^{q} \beta_{s}^{ij} X_{s}$, we will assume the following Normal-Gamma prior for the coefficients $\beta_{s}^{ij}$:

$$\beta_{s}^{ij} \sim N(0, \psi_{s}^{ij}); \psi_{s}^{ij} \sim \text{Gamma}(\lambda_{s}, 1/(2\gamma^{2})). \quad (4)$$

The CPF $\theta^{ij}(x) = -\frac{\omega^{ij} x}{\omega^{ij}} - \sum_{i=1}^{q} \beta_{s}^{ij} X_{i} = \sum_{i=1}^{q} \frac{\beta_{s}^{ij}}{\omega^{ij}} X_{i}$ is still a linear function. For each $\beta_{s}^{ij}$ in edge regression, The latent scale parameter $\psi_{s}^{ij}$ serves as an adaptive shrinkage parameter across both edges and covariates. We allow the shape parameter $\lambda_{s}$ to vary across covariates, but borrow strength across edges, and set the scale parameter $\gamma$ to be common across covariates and edges. This hierarchical structure is constructed to have flexibility, yet borrow strength across edges within covariates, and then across covariates. For $\omega^{ii}$, which controls the variance parameter in the neighborhood selection model, we choose a vague prior such that $\omega^{ii} \propto 1$, as done by Griffin et al. (2010), for our following discussion. If $\omega^{ij}$ is given with a conjugate prior Gamma($a^{*}, b^{*}$), the full conditional distribution for $\omega^{ij}$ keeps the same form, so our sampling scheme can still be implemented by a Gibbs step. A graphical representation of this hierarchical formulation is shown in Figure 1.

**Sampling scheme:** We adapt the scheme of Griffin et al. (2010) to sample $\lambda_{s}$ and $\gamma$ simultaneously by specifying exponential and inverse-gamma hyperpriors. Enabled by this hierarchical specification of the Normal-Gamma prior, we implement a Metropolis-within-Gibbs sampling scheme to update each parameter sequentially. The Gibbs steps involve a multivariate Gaussian for $\beta_{s}$, generalized inverse Gaussians for $\omega^{ii}$ and $\psi_{s}^{ij}$, a Gamma distribution for the scale parameter $\gamma^{-2}$, and a Metropolis-Hastings step is used to update the shape parameter $\lambda_{s}$. After sampling the parameters in the edge regression model, we subsequently obtain posterior samples for the subject-specific precision matrices for the subjects of interest in the dataset, and we could also produce posterior precision matrices for any other hypothetical subjects with specific levels of the covariates $x$. The steps of the sampler are summarized in Algorithm 1 (Supplementary Section B) and the corresponding computational details are given in the Supplementary Section B and C. Recall that the CPF $\theta^{ij}(x)$ is used to define the edge strength $\rho^{ij}(x)$ through $X$. When the edge strength of the subject-level graph is varying across $x$, the adaptive shrinkage priors imposed on each $\beta_{s}^{ij}$ induce different degrees of shrinkage on $\theta^{ij}(x)$ across $x ((1))$. Note that the Normal-Gamma prior induces a ridge-type prior $N(0, \psi_{s}^{ij})$ on each $\beta_{s}^{ij}$, so with a linear CPF, the prior induced to $\theta^{ij}(X = x)$ will still be a ridge prior, of which the variance item is controlled by $x$. The intercept term is also given sparsity priors, inducing sparsity across the edges overall. This implies that, a priori, we expect most edges do not vary strongly with a given $x$, but only a subset of edges. With the shrinkage induced onto the edges across subjects, a Bayesian FDR control procedure will be proposed to select edges for each subject-level graph, which finally induces sparsity at each subject-level graph.

### 2.5. Posterior Inference and Thresholding

We perform edge selection to estimate covariate-specific graphs by thresholding posterior probabilities of edge inclusion (PPI) for each edge based on the MCMC samples. For a given set of
In this section, we present simulation studies to investigate the performance of our Bayesian edge regression, designed to mimic the setting of tumor heterogeneity discussed in the introduction. As previously stated, most researchers interested in contrasting normal and tumor networks would not account for tumor purity, but instead would estimate normal and tumor graphs from the respective samples either using independent or group graphical models. Thus, we will compare our Bayesian edge regression method with three commonly used approaches for estimating multiple graphical models, the fused graphical lasso; the group graphical lasso (Danaher et al. 2014); as well as the Laplacian shrinkage for inverse covariance matrices from heterogeneous populations (LASICH) (Saeegus and Shojaie 2016), and an approach for solving covariate-dependent graphical model, the nonparametric finite mixture of Gaussian graphical model (NFMGGM) (Lee and Xue 2018). Additionally, we include a comparison with a method of Bayesian inference of multiple Gaussian graphical models (BIMGGM) (Peterson et al. 2015), which is a Bayesian approach to inference on group graphs. We further run the proposed Bayesian edge regression model with binary covariates mimicking the group definition used when applying the other group models, which is denoted as Bayesian edge regression (group case) in the following discussion. For each simulation, we run 20,000 MCMC iterations, in which the first 10,000 iterations are discarded as a “burn-in” period, and thin out the chain using every 10th sample.

3.1. Data Generation for Simulation

In order to test the ability of our Bayesian edge regression method to account for tumor purity in network estimation, we simulate data in a way to mimic the real-life setting of normal contamination in tumor samples. In our simulation study, we use a similar setting to construct precision matrices from Peterson et al. (2015) and include 20 nodes to represent 20 genes, which produces a proper degree of sparsity with around 10 ~ 20% of possible edges included for the generated precision matrix. From here on, we use the more general term “normal” to represent the stroma component. Let $\mathbf{Y}_n$ be the log2-transformed expressions from the clinically derived tumor sample $n$. According to Ahn et al. (2013) and Wang et al. (2018),

![Figure 1](image-url) A graphical representation of edge regression with normal-gamma prior. Single arrows are probabilistic edges; double arrows are deterministic edges; squares are observed data; circles are random variables. The total number of instances of each variable that is enclosed in the same plate is given by the constant in the corner of that plate. $\rho^{ij}$ is the partial correlation for edge $(i,j)$. 

Figure 1. A graphical representation of edge regression with normal-gamma prior.
the observed expressions before log2-transformation of gene expression data are well-modeled as a linear mixture of the expressions from the normal and the tumor components. It follows that

\[ 2^{Y_n} = (1 - \pi_n)2^{N_n} + \pi_n2^{T_n}, \quad (5) \]

where with a multivariate extension for gene interactions the log2-transformed expressions from the normal component \( N_n \sim \mathcal{N}(\mu_N, \Omega_N^{-1}) \) and those from the tumor component \( T_n \sim \mathcal{N}(\mu_T, \Omega_T^{-1}) \). \( \pi_n \in [0, 1] \) is the proportion of the tumor component before log2-transformation, that is, the measured tumor purity for sample \( n \). Following (5), we generated \( Y_n \) for each sample \( n \) from the simulated expressions \( N_n \) and \( T_n \). For simplicity, we set \( \mu_N = 0 \) and \( \mu_T = 0 \) in our simulation. We also generate \( N_n^{\star} \sim \mathcal{N}(0, \Omega_N^{-1}) \) as a reference group for normal component, where \( \pi_n = 0 \). We provide two simulations with different set-up of precision matrix.

### 3.2. Simulation 1: Without Overlap in Tumor and Normal Graphs

\( \Omega_T \), where off-diagonal elements \( \omega_T^{ij} = \omega_T^{ji} \) uniformly sampled from \([-0.5, -0.3] \cup [0.3, 0.5] \) for \( i = 1, \ldots, 18 \). \( \Omega_N \), where off-diagonal elements \( \omega_N^{ij} = \omega_N^{ji} \) uniformly sampled from \([-0.5, -0.3] \cup [0.3, 0.5] \) for \( i = 1, \ldots, 19 \). For both \( \Omega_T \) and \( \Omega_N \), all the diagonal elements are one and all the other elements are left with zero. \( \Omega_T \) and \( \Omega_N \) are truly sparse with just 18 and 19 edges. They do not have any overlapping edges by construction. We simulate reference normal samples of size \( N_N = 50 \) and mixed tumor samples of size \( N_T = 150 \) with \( \{\pi_n\}_{n=1}^{150} \) generated from an arithmetic sequence from 0.01 to 0.99. \( \pi_n \) will be considered as the extraneous covariate to our edge regression model and also taken as a fixed value in our following experiments. We randomly generated 100 datasets for this simulation.

### 3.3. Simulation 2: With Overlap in Tumor and Normal Graphs

\( \Omega_T \), where off-diagonal elements \( \omega_T^{ij} = \omega_T^{ji} \) uniformly sampled from \([-0.5, -0.3] \cup [0.3, 0.5] \) for \( i = 1, \ldots, 18 \). All the diagonal elements are one and all the other elements are left with zero. \( \Omega_N \), where we remove 30 edges randomly from \( \Omega_T \) by substituting these 30 nonzero elements with zero and randomly add 30 edges to \( \Omega_T \) by substituting these 30 zero elements with values uniformly sampled from \([-0.6, -0.4] \cup [0.4, 0.6] \). To ensure \( \Omega_N \) is positive definite, following Danaher et al. (2014), we divide each off-diagonal element by 1.5 times the sum of the absolute value of all the off-diagonal elements in its row. Then we average the transformed matrix with its transpose to guarantee it is symmetric. Although this procedure is able to guarantee the generated matrix will be positive definite, it can bring weak signals to \( \Omega_N \), which makes the estimation even more difficult. We allow \( \Omega_T \) and \( \Omega_N \) to have seven overlapping edges, and \( \Omega_N \) has relatively weak edge strengths. We simulate reference normal samples of size \( N_N = 100 \) and mixed tumor samples of size \( N_T = 200 \) and generate \( \pi_n \) as in Simulation 1. We randomly generated 100 datasets for this simulation. We also provide two more simulations that simulate \( \Omega_N \) and \( \Omega_T \) with scale-free networks and nearest-neighbor networks in Supplementary Section D (Guo et al. 2011). The graph structures for all the simulation settings are shown in Supplementary Figure 1.

We compare the results of our edge regression with application of the fused and group graphical lassos1, LASICH2, NFMGGM3, and BIMGGM4, respectively, to the tumor and normal measurements, \( Y \) and \( N \), respectively. The application of these group graph methods corresponds to what might be the usual practice of estimating tumor graphs from tumor samples without adjusting for tumor purity and normal contamination, and estimation of the normal graph from normal controls, so has practical scientific relevance. For running our method and NFMGGM, all the genes are normalized to have a mean of zero and a standard deviation of one with all the samples. For running all the group graphical models, the data are normalized to have a mean of zero and a standard deviation of one, respectively, within the tumor and normal group. More details of the implementations for all the methods are in the Supplementary Section D.

In our Bayesian edge regression model, following (1) we parameterize the dependence of \( \omega^{ij}(.) \) on \( X \):

\[ \omega^{ij}(\pi) = \beta^{ij}(1 - \pi) + \alpha^{ij}(\pi). \quad (6) \]

Under this parameterization, \( \alpha^{ij} \) represents the precision element for pure tumor samples, and \( \beta^{ij} \) the precision element for a pure normal sample, with the sample specific edges given by a linear combination as determined by their tumor purity \( \pi \). This model allows us to reweight samples based on tumor purity to estimate the pure normal and pure tumor graphs. For an additional comparison of our approach using discrete predictors, we also ran our Bayesian edge regression model with binary covariates mimicking the group definition used when applying the other group models (i.e., Bayesian edge regression (group case)). In this application, we use two binary covariates to encode the membership of tumor and normal groups and add an additional covariate to capture the interaction effects between tumor and normal groups (see more details in the supplementary material, Section D).

We implement these methods across 100 simulated datasets for both the first and second simulations. We compare the methods in terms of accuracy in estimating the graph structure using the area under the ROC curve (AUC) and true positive rate (TPR) and false positive rate (FPR). We have two regularization parameters for both our graph edge regression (\( k \) and \( \delta \)), NFMGGM (\( k \) and \( h \)), LASICH, and the two graphical lasso methods (\( \lambda_1 \) and \( \lambda_2 \)). Thus, for all these methods we compute a bivariate AUC (McGuffey et al. 2018) by varying both two parameters at the same time, and then computing the expected AUC (baUC) by binning results on a grid of 1 - specificity, and computing the average sensitivity within those values. BIMGGM is reported with a univariate AUC given it only requires one tuning parameter. For better observing how each tuning parameter affects the model performance for these methods with two

1Available in the R package JGL.
2Available in the R package LASICH.
3Implementation requested from the authors.
4https://odin.mdacc.tmc.edu/~cbpeterson/software.html.
hyperparameters, we further report the best univariate AUC over one hyperparameter given the other is fixed. To compare TPR and FPR for a single choice of regularization parameters, we use $\kappa = 0.1$ and $\delta = 0.1$ for the Bayesian edge regression methods that are respectively built with continuous covariates and discrete covariates, and we choose $\lambda_1$ and $\lambda_2$ or $\lambda$ and $h$ for the compared methods following the previously mentioned guidelines (Supplementary Section D). We report the results for BIMGGM with the same $\delta = 0.1$ for posterior thresholding with Bayesian local FDR. In addition to the TPR and FPR reported with the selected model, we also report a TPR corresponding to $FPR \approx 0.1$ for each method, providing a fair comparison of methods using common criteria. The ROC curves for these two simulations are given in Figure 2, and tables showing the AUC, TPR, and FPR for normal, tumor, and overall are given in Supplementary Section D.

In Simulation 1, we see that all methods yield relatively high bAUC for the normal graphs, but the Bayesian edge regression with continuous covariates has much better bAUC (0.915) for the tumor graph than the other group graph methods (0.706, 0.793, 0.808, and 0.813) (see supplementary material, Table 2). This is related to the fact that, unlike the group graph methods, our edge regression can adjust for continuous variables, such as the tumor purity, and thus reduce biases in parameter estimations for the tumor graph. The proposed method outperforms the kernel regression-based method NFMGGM (0.810) for the tumor graph as well, which suggests a parametric edge regression leads to better performance in this case. Also, note that the FPRs for the Bayesian edge regression method are all below 0.100, while the NFMGGM and graphical lasso methods have high FPR for the choices of their regularization parameters. We can further find that the proposed method is reported with the highest overall TPR (0.903 versus 0.660, 0.697, 0.735, 0.772, 0.781, 0.740) among all the methods when the model is chosen such that $FPR \approx 0.1$ (see Supplementary Table 3).

In Simulation 2, a more challenging setting with weaker signal, we also observe that our Bayesian edge regression method with continuous covariates produces higher bAUC for both the normal (0.801 versus 0.748, 0.770, 0.751, 0.738, 0.770) and tumor (0.911 versus 0.758, 0.813, 0.852, 0.799, 0.868) graphs (see Supplementary Table 4). Once again, the graph lasso methods with the chosen regularization parameters resulted in much higher FPRs ($>0.500$), while the Bayesian edge regression was much lower (0.047 for normal and 0.269 for tumor; see the results in Supplementary Section D). Similarly, the proposed method leads to the highest overall TPR (0.700 versus 0.423, 0.417, 0.443, 0.397, 0.638, 0.592) when $FPR \approx 0.1$ (see Supplementary Table 5). From the two simulations, we find that implementing a group graphical model on our Bayesian edge regression framework (i.e., Bayesian edge regression (group))
obtains a performance as similar as other group graph methods, but worse than the regression model with continuous covariates (overall AUC: 0.874 versus 0.931 in Simulation 1; 0.819 versus 0.856 in Simulation 2) (Supplementary material, Tables 2 and 4), which again highlights the importance of adjusting for tumor purity as a continuous covariate when estimating tumor and normal networks.

4. Proteomic Networks in Hepatocellular Carcinoma

4.1. Markers in HCC

Hepatocellular carcinoma (HCC) is the fifth most common cancer worldwide and the third-leading cause of cancer-related death, and it has an increasing incidence in developing countries. In the USA, it is the fastest growing cause of cancer-related mortality in men, and with the alarming increase of hepatitis C, it is expected to continue to grow in incidence in the coming years. Over 80% of patients present with advanced disease and underlying cirrhosis (Fattovich et al. 2004; Sanyal et al. 2010), which prevents curative treatment options. There are only a few approved systemic therapies for HCC, for example, sorafenib, and various targeted therapies are being assessed in combination, including therapies targeting angiogenesis pathways. New targeted therapies and precision therapy strategies are clearly needed.

Cytokines are blood proteins secreted by various types of cells in the immune system that have an effect on other cells. There is significant evidence that numerous cytokines mediate processes involved in the liver, including inflammation, necrosis, cholestasis, fibrosis, and regeneration, and are a key factor in many liver diseases, including HCC (Tilg 2001). Other biological pathways instrumental for HCC include inflammation, metabolic pathways, immune response, growth factor, and angiogenesis (Dhanasekaran et al. 2016; Aravalli et al. 2008). A deeper characterization of the molecular basis of interpatient heterogeneity of HCC, including behavior within these pathways, has the potential to contribute to new, targeted precision therapy strategies for HCC.

A recently developed novel prognostic measure characterizing inter-patient heterogeneity in HCC from blood protein profiles is the HepatoScore (Morris et al. 2020). This biological prognostic score has been shown to dramatically refine HCC staging systems, e.g. accurately delineating a subset of metastatic patients with low HepatoScore who have substantially better prognosis than non-metastatic patients with high HepatoScore. Although it is a biological score based only on blood protein levels including no clinical factors, HepatoScore by itself outperforms all existing staging systems and prognostic factors (Morris et al. 2020). While a global score computed from the entire panel of circulating proteins, HepatoScore is primarily driven by a subset of key circulating proteins within various molecular pathways relevant to HCC, most notably the immune response, GH, and angiogenesis pathways. It is thought that these pathways play a major role in characterizing the patient’s cancer and prognosis, and deeper characterization of the interrelationships across these proteins may yield important biological insights.

The dataset analyzed in this article involves measurements of proteins that are obtained using CytokineMAP (Myriad RBM, Austin, TX) on plasma samples from 767 HCC patients (Morris et al. 2020). The proteins considered here include 71 proteins from immune system, GH, and angiogenesis pathways plus alpha-fetoprotein (AFP), an important protein for HCC used for early detection and prognosis. After scaling all proteins to have a mean of zero and variance of one, we ran our graph edge regression model as outlined above. Given HepatoScore \( \pi \in [0, 1] \), our model for the conditional precision edge \((i, j)\) is given by \( \omega^{ij}(\pi) = \beta^{ij}(1 - \pi) + \alpha^{ij}(\pi) \).

Under this parameterization \( \beta^{ij} \) represents the edge strength for low HepatoScore \((\pi = 0)\), \( \alpha^{ij} \) the edge strength for high HepatoScore \((\pi = 1)\), and a linear combination assumed for a moderate HepatoScore, for example, with the edge strength for \( \pi = 0.5 \) given by \( 0.5\beta^{ij} + 0.5\alpha^{ij} \). The graph edge strengths for any continuous HepatoScore \( \pi \in [0, 1] \) can be computed from this model. We used the same prior setting as in Section 3, and chose a tuning parameter \( \sigma_\delta \) that provides an acceptance rate of the Metropolis step at around 20%–30%, determined after the burn-in period. We ran the MCMC sampler for 10,000 iterations after a burn-in of 10,000, then thinning to keep every 10th sample.

To assess convergence, we observed trace plots and ran a Geweke convergence diagnostic for all parameters. The histogram of the Geweke \( p \)-values suggests that the chain converged satisfactorily (Supplementary Figure 4). Within the sampler, we also obtained posterior samples for the predicted precision matrices corresponding to a low \((\pi = 0)\), medium \((\pi = 0.5)\), and high \((\pi = 1)\) HepatoScore as described above, and applied our posterior edge selection approach based on \( \delta = 0.1 \) and \( \kappa = 0.15 \), also considering \( \kappa = 0.1 \) and \( \kappa = 0.2 \) for sensitivity. We also ran NFMGGM for comparison, with results in Supplementary Section E.

While our method produces predicted graphs for any \( \pi \in [0, 1] \), for interpretation we focus on three levels of \( \pi \in \{0, 0.5, 1\} \). Figure 3 contains the estimated graphs for a low HepatoScore \((\pi = 0)\), medium HepatoScore \((\pi = 0.5)\), and high HepatoScore \((\pi = 1)\), with edge direction indicated by color (green = positive, red = negative), edge strength by line width, and node size indicating the number of connecting edges. Blue lines indicate edges shared across all levels of \( \pi \), and their direction is given by panel (d). It is clear that the number of graph edges increases with HepatoScore values, indicating that the protein network connectivity increases for more invasive forms of HCC. Figure 4 summarizes the number of connections within each pathway and between each pair of pathways, as a function of HepatoScore \( \pi \), and Supplementary material, Table 15 shows the number of edges within and between different pathways in the respective graphs for \( \pi = 0, 0.5 \) and 1. We see that the number of intra-pathway connections within each of the three pathways strongly increases with HepatoScore, especially for a high HepatoScore \((\pi > 0.8)\), with more than twice the number of edges than a low HepatoScore. The number of inter-pathway edges increases with the HepatoScore even more strongly, with a three to fourfold increase, much notably between the angiogenesis and immune pathways with 40 edges for \( \pi = 1 \) and only 15 edges for \( \pi = 0 \). The increased connectivity could correspond to increased activity within these important pathways, and increased cross talk between them. This could have important implications for the underlying molecular biology, and it needs to be followed up to validate and assess the biological implications of these associations.
Supplementary Section E contains the results using $\kappa = 0.1$ and 0.2, which demonstrate the same substantive effects, although of course with a greater and fewer number of total edges, respectively, in the graphs.

Hub proteins are proteins with many connections in the graph, and they may be involved in multiple regulatory activities. Different hub genes are identified for these three graphs (Supplementary Table 16). IGFBP-3 has been identified as a hub gene in the high HepatoScore graph and with a moderate degree of connectivity in the low HepatoScore graphs, where the connected nodes are different. IGFBP-3 has been considered as an effective predictor for HCC patients with chronic HCV infections, and it is a transcription factor encoding proteins to suppress HCC cell proliferation, so the reduction of IGFBP-3 is significantly associated with the development of HCC (Aleem et al. 2012; Ma et al. 2016). There are many edges that vary over the HepatoScore. We highlight a few notable ones here and present the rest in Supplementary material, Section E. Figure 5 contains a plot for three edges, which presents the edge strength as a function of HepatoScore $\pi$ along with 95% credible intervals and the corresponding PPI $\delta = 0.1, \kappa = 0.15$.

6Ckine is strongly associated with MIP-3, $\beta$ for medium and high HepatoScores, whereas this edge is not apparent in the graphs for a low HepatoScore. The regulation between 6Ckine and MIP-3, $\beta$ has been previously reported to play a determinant role in accumulating antigen-loaded mature dendritic cells (Caux et al. 2000). AFP/MIP-3, $\alpha$ is another pair that shows no correlation in the graph for low HepatoScore, but a positive correlation for high HepatoScore. AFP ($\alpha$-fetoprotein) is a tumor marker for liver cancer. The levels of AFP have
been reported to relate with MIP-3, α levels in HCC, where the serum levels of MIP-3, α are increased (Yamauchi et al. 2003). CA-15-3 is well known to detect breast cancer and distinguish from non-cancerous lesions, and its level has been shown to be increased for end-stage liver disease patients (Pissaia et al. 2009; Szekanecz et al. 2008). MCP-1 is a protein secreted by the HCC microenvironment that can promote progression, angiogenesis, and metastasis in cancer through recruiting and modifying mesenchymal stromal cells (MSCs). CA-15-3/MCP-1 shows negative correlation for π = 0 and positive correlation for π = 1, which corresponds to these previous empirical findings of elevated levels of CA-15-3 and MCP-1 in liver disease.

In Supplementary Figure 8, we show several additional connections we consider biologically meaningful, as we discuss in Supplementary Section E. We further summarize how edge connectedness varies with the HepatoScore in Figure 4 (see Supplementary Figures 9 and 10 for κ = 0.1 and 0.2). These results suggest that the protein networks in these important pathways differ in HCC patients with more and less advanced stages of the disease, with connectivity increasing with HepatoScore, with a dramatically greater number of connections for the HCC patients with higher HepatoScore and the most poor prognoses. Biological studies investigating these differences have the potential to reveal insights into the molecular heterogeneity distinguishing these patients from those with a much better prognosis, and this knowledge can contribute toward our efforts to identify sorely needed new precision therapy strategies for HCC. These discoveries were made possible by the novel modeling framework we have introduced in this article.

5. Discussion

In this article, we introduce a Bayesian edge regression model for construction of non-static undirected graphs with edge strengths varying with extraneous covariates. To deal with potential high dimensionality, we use global-local priors to effectively induce sparsity into the underlying graphs, and we use posterior probabilities to infer important graph edges for a given set of covariates. Based on node-wise regressions, that have been shown with good performance for graph reconstruction (Leday et al. 2015; Meinshausen and Bühlmann 2006; Ha et al. 2021), our method is primarily focused and recommended for applied settings in which the focus is on edge detection rather than estimation of the full precision or covariance matrix. This modeling framework allows researchers to study how clinical and biological factors lead to heterogeneous genomic or proteomic networks varying across patients. We demonstrate how this method could be used to incorporate tumor purity in estimating structural differences between tumor and normal graphs, and to assess how graphs vary across a continuous prognostic factor HepatoScore to explain interpatient heterogeneity in HCC. Our construction is general and can be used in any setting with multivariate data and covariates for which assessment of how conditional dependencies across the variables vary across continuous or discrete covariates are of interest. While motivated by the setting of continuous covariates, the method is based on a general regression framework in which any number of continuous or discrete covariates or interactions can be included. We also provide freely available code for fitting our models.

Our sampling scheme employs a Gibbs sampler, which yields posterior samples for the model and predicted graphs for any set of covariates that can be used for posterior inference. Our hyperpriors for the normal-gamma shrinkage prior are set to accommodate the borrowing of information for regularization parameters of covariate coefficient across different edges. Our sampling procedure is also easy to implement and requires only minimal tuning of shrinkage hyperparameters. We show the parameterization of our conditional precision function and its practicality through simulation study, and we demonstrate that our method is able to provide a reasonable sensitivity and specificity in edge selection. The parameterization is flexible and shown to be able to borrow strength in a group-specific setting by introducing interactions.

Our fully Bayesian method is designed for application to moderate-sized graphs from dozens to over 100 nodes, the scale of our motivating examples, and our method scales well to these sizes. It is not intended for enormous graphs with 1000s or 10,000s of nodes, a setting that would require enormous sample sizes for estimability anyway. As is commonly done in genomic settings (Peterson et al. 2015; Telesca et al. 2012; Chun et al.
Supplementary Materials

The supplementary materials additionally detail the steps of the MCMC algorithm used for edge regression as well as the derivation of the algorithm. The supplementary materials also present more simulation settings and results as well as more implementation details for the methods in simulations. Additional result analyses for our HCC case study are also provided in the supplementary materials.

Acknowledgments

This work was supported by NSF grants 1550088 and 1463233, plus NIH grants P30-CA-016672, CA-160736, CA-158113, CA-178744, CA-183793, CA220299, CA-221707, CA-239342, CA-244845, P30-CA46592, and U11-TR001878, and start-up funds from the U-M Rogel Cancer Center and University of Michigan School of Public Health. We thank the associate editor and referees whose constructive comments greatly improved this paper.

References

Ahn, J., Yuan, Y., Parmigiani, G., Suaroakar, M. B., Diao, L., Wistuba, I. I., and Wang, W. (2013), “Demix: Deconvolution for Mixed Cancer Transcriptsomes Using Raw Measured Data,” Bioinformatics, 29, 1865–1871. [538]

Aleem, E., Elshayeb, A., Elhabache, N., Mansour, A. R., Gowily, A., and Hela, A. (2012), “Serum igfbp-3 is a More Effective Predictor Than igf-1 and igf-2 for the Development of Hepatocellular Carcinoma in Patients With Chronic hcv Infection,” Oncology Letters, 3, 704–712. [542]

Aravalli, R. N., Steer, C. I., Cressman, E. N. (2008), “Molecular Mechanisms of Hepatocellular Carcinoma,” Hepatology, 48, 2047–2063. [541]

Bhattacharya, A., Pati, D., Pillai, N. S., Dunson, D. B. (2015), “Dirichlet–Laplace Priors for Optimal Shrinking,” Journal of the American Statistical Association, 110, 1479–1490. [537]

Cai, T. T., Polson, N. G., and Scott, J. G. (2012), “Covariate-Adjusted Precision Matrix Estimation With an Application in Genetical Genomics,” Biometrika, 100, 139–156. [535]

Cai, T. T., Li, H., Liu, W., Xie, J. (2016), “Joint Estimation of Multiple High-Dimensional Precision Matrices,” Statistica Sinica, 26, 445–464. [534]

Carvalho, C. M., Polson, N. G., and Scott, J. G. (2010), “The Horseshoe Estimator for Sparse Signals,” Biometrika, 97, 465–480. [537]

Castillo, I., and van der Vaart, A. (2012), “Needles and Straw in a Haystack: Posterior Concentration for Possibly Sparse Sequences,” The Annals of Statistics, 40, 2069–2101. [537]

Caux, C., Ait-Yahia, S., Chemin, K., de Roulleter, O., Dieu-Noisjean, M. C., Homey, B., Massacrier, C., Vanbervliet, B., Zlotnik, A., Vicari, A. (2000), “Dendritic Cell Biology and Regulation of Dendritic Cell Trafficking by Chemokines,” in Springer Seminars in Immunopathology, Vol. 22, Springer, pp. 345–369. [542]

Cheng, J., Levina, E., Wang, P., and Zhu, J. (2014), “A Sparse Ising Model With Covariates,” Biometrics, 70, 943–953. [535]

Chun, H., Zhang, X., and Zhao, H. (2015), “Gene Regulation Network Inference With Joint Sparse Gaussian Graphical Models,” Journal of Computational and Graphical Statistics, 24, 954–974. [544]

Danaheer, 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, 76, 373–397. [534,538,539]

Dempster, A. P. (1972), “Covariance Selection,” Biometrics, 28, 157–175. [533]

Dhanasekaran, R., Bandoh, S., and Roberts, L. R. (2016), “Molecular Pathogenesis of Hepatocellular Carcinoma and Impact of Therapeutic Advances,” F1000Research, 5. [541]

Farley, P. (2015), “‘Purity’ of Tumor Samples May Significantly Bias Genomic Analyses.” Available at: http://medicalxpress.com/news/2015-12-purity-tumor-samples-significantly-bias.html [534]

Fattovich, G., Stroffolini, T., Zagni, I., Donato, F. (2004), “Hepatocellular Carcinoma in Cirrhosis: Incidence and Risk Factors,” Gastroenterology 127, S35–S50. [541]
Griffin, J. E., and Brown, P. J. (2011), “Bayesian Hyper-Lasso With Non-Convex Penalization,” *Australian and New Zealand Journal of Statistics*, 53, 423–442. [537]

Griffin, J. E., and Griffin, J. E. (2010), “Inference With Normal-Gamma Prior Distributions in Regression Problems,” *Bayesian Analysis*, 5, 171–188. [537]

Hastie, T., and Tibshirani, R. (1993), “Varying Coefficient Models,” *Journal of the Royal Statistical Society*, Series B, 55, 757–796. [536]

Hof, P. D., and Niu, X. (2012), “A Covariance Regression Model,” *Statistica Sinica*, 729–753. [535]

Horst, F., and Sillanpää, M. (2006), “Bayesian Mapping of Genotype

Liu, W. (2017), “Structural Similarity and Difference Testing on Multiple Graphical Models,” *Journal of the American Statistical Association*, 110, 1372–1384. [537]

Meinshausen, N., and Bühlmann, P. (2006), “High-Dimensional Graphs and Variable Selection With the Lasso,” *The Annals of Statistics* pp. 1436–1462. [534,544]

Mitchell, T. J., and Beauchamp, J. J. (1988), “Bayesian Variable Selection in Linear Regression,” *Journal of the American Statistical Association*, 83, 1023–1032. [537]

Mitra, R., Müller, P., Ji, Y. (2016), “Bayesian Graphical Models for Differential Pathways,” *Bayesian Analysis*, 11, 99–124. [534]

Morris, J. S., Brown, P. J., Herrick, R. C., Bagherly, K. A., and Coombes, K. R. (2008), “Bayesian Analysis of Mass Spectrometry Proteomic Data Using Wavelet-Based Functional Mixed Models,” *Biometrics*, 64, 479–489. [536]

Morris, J. S., Hassan, M. M., Zohner, Y. E, Wang, Z., Xiao, L., Rashid, A., Haque, A., Abdel-Wahab, R., Mohamed, Y. I., Ballard, K. L., Wolf, R. A., George, B., Li, L., Allen, G., Weylandt, M., Li, D., Wang, Raghav, K., Yao, J., Amin, H. M., Kaseb, A. O. (2020), “Hepatocyte-14: Measures of Biological Heterogeneity Significantly Improve Prediction of Hepatocellular Carcinoma Risk,” *Hepatology*, 73, 2278–2292. [534,541]

Narayan, M., Allen, G. I., Tomson, S. (2015), “Two Sample Inference for Populations of Graphical Models With Applications to Functional Connectivity,” arxiv:1502.03853. [534]

Narisetty, N. N. et al. (2014), “Bayesian Variable Selection With Shrinking and Diffusing Priors,” *The Annals of Statistics*, 42, 789–817. [537]

Ni, Y. et al. (2018), “Bayesian Graphical Regression,” *Journal of the American Statistical Association*, 1–14. [535,536,544]

Park, T., and Casella, G. (2008), “The Bayesian Lasso,” *Journal of the American Statistical Association*, 103, 681–686. [537]

Peng, J. et al. (2012), “Partial Correlation Estimation by Joint Sparse Regression Models,” *Journal of the American Statistical Association*. [536,544]

Petersen, C. et al. (2015), “Bayesian Inference of Multiple Gaussian Graphical Models,” *Journal of the American Statistical Association*, 110, 159–174. [534,538,543]

Pissaia, A. et al. (2009), “Significance of Serum Tumor Markers Carcinoembryonic Antigen, Ca 19-9, Ca 125, and Ca 15-3 in Pre-Orthotopic Liver Transplantation Evaluation,” in *Transplantation Proceedings*, pp. 682–684. [543]

Polson, N. G., and Scott, J. G. (2010), “Shrink Globally, Act Locally: Sparse Bayesian Regularization and Prediction,” *Bayesian Statistics*, 9, 501–538. [537]

Saegusa, T. and Shojaita, A. (2016), “Joint Estimation of Precision Matrices in Heterogeneous Populations,” *Electronic Journal of Statistics*, 10, 1341. [534,538]

Sanyal, A. J. et al. (2010), “The Etiology of Hepatocellular Carcinoma and Consequences for Treatment,” *The Oncologist*, 15, 14–22. [541]

Scott, J. G., and Berger, J. O. (2010), “Bayes and Empirical-Bayes Multiplicity Adjustment in the Variable-Selection Problem,” *The Annals of Statistics*, 38, 2587–2619. [537]

Szekanecz, É., Szucs, G., Szekanecz, Z., Tarr, T., Antal-Szalmás, P., Szamosi, S., Szántó, I., Kiss, E. (2008), “Tumor-Associated Antigens in Systemic Sclerosis and Systemic Lupus Erythematosus: Associations With Organ Manifestations, Immunolaboratory Markers and Disease Activity Indices,” *Journal of Autoimmunity*, 31, 372–376. [543]

Tan, L. S., Jara, A., De Iorio, M., Eibels, T. M. D. (2017), “Bayesian Inference for Multiple Gaussian Graphical Models With Application to Metabolic Association Networks,” *The Annals of Applied Statistics*, 11, 2222–2251. [534]

Telesca, D., Müller, P., Kornblau, S. M., Suchard, M. A., and Ji, Y. (2012), “Modeling Protein Expression and Protein Signaling Pathways,” *Journal of the American Statistical Association*, 107, 1372–1384. [543]

Tilg, H. (2001), “Cytokines and Liver Disease,” *Canadian Journal of Gastroenterology and Hepatology*, 15, 661–668. [541]

Wang, Z., Cao, S., Morris, J. S., Ahn, J., Liu, R., Tsykucheva, S., Gao, F., Li, B., Lu, W., Tang, X., Wistuba, I. I., Bowden, M., Musci, L., Loda, M., Parmigiani, G., Holmes, C. C., Wang, W. (2018), “Transcriptome Deconvolution of Heterogeneous Tumor Samples With Immune Infiltration,” *iScience*, 9, 451–460. [534,538]

Xia, Y., Cai, T. T., and Cai, T. (2015), “Testing Differential Networks With Applications to the Detection of Gene-Gene Interactions,” *Biometrika*, 102, 247–266. [534]

Yamauchi, K., Akbar, S. M., Horikke, N., Michitaka, K., and Onji, M. (2003), “Increased Serum Levels of Macrophage Inflammatory Protein-3α in Hepatocellular Carcinoma: Relationship With Clinical Factors
and Prognostic Importance During Therapy,” *International Journal of Molecular Medicine*, 11, 601–605. [543]

Yang, E., Baker, Y., Ravikumar, P., Allen, G., Liu, Z. (2014), “Mixed Graphical Models Via Exponential Families,” in *Artificial Intelligence and Statistics*, 33, 1042–1050. [534]

Zhou, S., Lafferty, J., and Wasserman, L. (2010), “Time Varying Undirected Graphs,” *Machine Learning*, 80, 295–319. [535]

Zou, T., Lan, W., Wang, H., and Tsai, C. L. (2017), “Covariance Regression Analysis,” *Journal of the American Statistical Association*, 112, 266–281. [535]