Foundations of Data Science
©American Institute of Mathematical Sciences
Volume **2**, Number **1**, March **2020**
doi:10.3934/fods.2020003
pp. **35–54**

# BAYESIAN INFERENCE FOR LATENT CHAIN GRAPHS

Deng Lu

Department of Statistics & Applied Probability, National University of Singapore
Singapore, 117546, SG

Maria De Iorio

Yale-NUS, Singapore, 138527, SG. &
Department of Statistical Science, University College London, UK

Ajay Jasra

Computer, Electrical and Mathematical Sciences and Engineering Division
King Abdullah University of Science and Technology
Thuwal, 23955, KSA

Gary L. Rosner

Oncology Biostatistics and Bioinformatics, Sidney Kimmel Comprehensive Cancer Center
Johns Hopkins University
Baltimore, MD 21205, USA

Abstract. In this article we consider Bayesian inference for partially observed Andersson-Madigan-Perlman (AMP) Gaussian chain graph (CG) models. Such models are of particular interest in applications such as biological networks and financial time series. The model itself features a variety of constraints which make both prior modeling and computational inference challenging. We develop a framework for the aforementioned challenges, using a sequential Monte Carlo (SMC) method for statistical inference. Our approach is illustrated on both simulated data as well as real case studies from university graduation rates and a pharmacokinetics study.

1. **Introduction.** Two common approaches to probabilistically describe conditional dependence structures are based on undirected graphs (Markov networks) and directed acyclic graphs (DAG) (Bayesian networks), e.g. [25]. The vertices of the graph represent variables, while edges represent a relation between each pair of vertices. In general, the presence (absence) of an edge between two vertices indicates conditional dependence (independence) between the two corresponding variables. Applications of Markov networks include biological networks, e.g., to study the dependence structure among genes from expression data [9, 11] and financial time series for forecasting and predictive portfolio analysis [8, 33]; while Bayesian networks occur in expert system research for providing rapid absorption and propagation of evidence [18, 24], path analysis for computing implied correlations [34], and psychometrics for causal pathways [6, 16].

Chain graphs (CG) [19, 20, 35] provide an elegant unifying framework which encompasses both Markov and Bayesian networks. These models allow for edges that are both directed and undirected and do not contain any semi-directed cycles.

2010 *Mathematics Subject Classification.* 62F15.
*Key words and phrases.* Chain graphs, Bayesian inference, sequential Monte Carlo.

Although chain graphs were first introduced in the late eighties, most research has focused on Bayesian networks and Gaussian Graphical Models. Recently they are receiving more attention as they have proved to be very useful in applications due to their ability to represent symmetric and non-symmetric relationships between the random variables of interest. However, there exist in the literature several different ways of interpreting chain graphs and what conditional independencies they encode, giving rise to different so-called *chain graph interpretations*. Whilst a directed edge works as in DAGs when it comes to representing independence, the undirected edge can be understood in different ways, giving rise to different interpretations. This implies that chain graphs can represent every graphical Markov model achievable by any DAG, whereas the opposite does not hold. The most common interpretations are the Lauritzen-Wermuth-Frydenberg (LWF) interpretation, the Andersson-Madigan-Perlman (AMP) interpretation and the multivariate regression (MVR) interpretation. Each interpretation has its own way of determining conditional independences in a CG and each interpretation subsumes another in terms of representable independence models; see [30]. Moreover, [17] discuss causal interpretation of chain graphs.

This work focuses on chain graphs of the AMP type [2, 22] and develops a statistical framework to infer the chain graph structure from data measured on a set of variables of interest with the goal of understanding their association pattern. Many inferential procedures have concentrated on the estimation of the parameters characterizing the graph, *given the chain graph topology*. For example, [12] consider likelihood-based parameter inference for chain graph models that are directly observed. Nevertheless, in the machine learning literature, methods based on optimization have been proposed to perform inference on the graph structure itself in the context of chain graphs, e.g. [23, 26]. We consider the case where not only are the parameters unknown, but the chain graph itself is unobserved. The main objective is to perform Bayesian inference in such a context.

There are several contributions that propose Bayesian methods in related contexts. In [28], the author proposes a method to perform full Bayesian inference for acyclic directed mixed graphs using Markov chain Monte Carlo (MCMC) methods. There are several similarities between the model there (based upon structural equation models (SEM) e.g. [7]) and the one developed here. However, the main difference is given by the structure of the latent graph and the associated constraints. In [28], the likelihood function is defined through linear mixed graph models usually referred to as structural equation models ([7]). Here we focus on AMP chain graphs described in detail later. In [32] a similar model to the one presented in [28] is developed, which uses a spike-and-slab prior to induce sparsity in the graph. The computational scheme is expected to outperform that in [28] for a large number of nodes. The prior structure in [32] could in principle be extended to accommodate chain graph structures at the cost of introducing constraints that could make computations infeasible.

The contributions of this article are as follows. Based upon the likelihood method in [12] we develop a new Bayesian model for latent AMP chain graphs. This model can also incorporate covariate information, if available. We introduce a sequential Monte Carlo (SMC) method as in [10] that improves upon MCMC-based methods in this context. Our approach is applied to real case studies from university graduation rates and a pharmacokinetics study. We find the performance of our algorithm to be stable and robust.

This article is structured as follows. In Section 2 we describe our model and prior specifications. In Section 2.3 we introduce the SMC algorithm. In Section 3 we present a simulation study and two real-life applications. In the appendix we detail further elements of our algorithm in Section 2.3.

## 2. Model.

### 2.1. Likelihood.

Let $Y = (Y_1, \dots, Y_p) \in \mathbb{R}^p$ be a random vector whose elements correspond to $p \in \mathbb{N}$ nodes of a graph. We assume we have $m \in \mathbb{N}$ observations on $Y$: $y_{1:m}$, $y_i \in \mathbb{R}^p$.

A graph $G = (V, E)$ is described by a set of nodes $V$ and edges $E$, with variables $Y_1, \dots, Y_p$ placed at the nodes. An edge $(u, v) \in E$ is called an undirected edge (denoted by $u - v$) if its opposite $(v, u) \in E$. An edge $(u, v) \in E$ is called a directed edge (denoted by $u \to v$) if its opposite $(v, u) \notin E$. The edges define the global conditional independence structure of the distribution. A sequence $(v_0, v_1, \dots, v_n)$ is called a path of length $n \geq 1$ from $v_0$ to $v_n$ if $(v_{i-1}, v_i) \in E$ for all $i = 1, \dots, n$. If $v_n = v_0$ and $n \geq 3$, the path is called an $n$-cycle. If $v_{i-1} \to v_i$ for (all)(at least one)(no) $i$, the path/cycle is called (directed)(semi-directed)(undirected). An AMP chain graph is a graph whose every edge is directed or undirected such that it does not contain any semi-directed cycles. Each graph can be identified by an adjacency matrix. Let $A$ be a $p \times p$ matrix with entries $(a_{ij})_{1 \leq i,j \leq p}$ where $a_{ij} \in \{0, \dots, r\}$ (let $r = 3$, which is a notation used later on) with $a_{ii} = 0$, for $i < j$

$$a_{ij} = \begin{cases} 0 & \text{no edge between } i \text{ and } j \\ 1 & \text{undirected edge between } i \text{ and } j \\ 2 & \text{directed edge from } i \text{ to } j \\ 3 & \text{directed edge from } j \text{ to } i \end{cases} \tag{1}$$

and, given the upper-triangular part of $A$, for $j < i$

$$a_{ji} = \begin{cases} a_{ij} & \text{if } a_{ij} \in \{0, 1\} \\ 2 & \text{if } a_{ij} = 3 \\ 3 & \text{if } a_{ij} = 2. \end{cases} \tag{2}$$

Given $p$, a labelling of the nodes and the adjacency matrix $A$ define a graph $G(A)$. Let $\mathcal{C}$ denote the set of possible chain graphs for a set of $p$ vertices. Given $A$, let $\Omega$ be a positive definite $p \times p$ (real-valued) matrix, such that if $a_{ij} \neq 1$ then $\omega_{ij} = 0$. In addition, given $A$, for an arbitrary real-valued $p \times p$ matrix $B$, then, if $a_{ij} \neq 2$ $b_{ij} = 0$. Finally set

$$\Sigma(B, \Omega) = (I - B)^{-1} \Omega^{-1} (I - B')^{-1} \tag{3}$$

where $I$ is the $p \times p$ identity matrix. To inherit the AMP chain graph property, $\Sigma(B, \Omega)$ should be a $p \times p$ positive definite matrix.

The absence of semi-directed cycles implies that the vertex set of a chain graph can be partitioned into so-called chain components such that edges within a chain component are undirected whereas the edges between two chain components are directed and point in the same direction. More precisely, the vertex set of a chain graph $G(A)$ can be partitioned into subsets $\mathscr{T}(G(A))$ such that all edges within each subset $\tau$ are undirected and edges between two different subsets $\tau \neq \tau'$ are directed. In the following, we assume that the partition $\tau \in \mathscr{T}(G(A))$ is maximal, that is, any two vertices in a subset $\tau$ are connected by an undirected path. Recall that the parents of a set of nodes $\mathcal{X}$ of $G$ is the set $\text{pa}(\mathcal{X}) = \{V_j \text{ s.t. } V_j \to V_i \in G, V_j \notin \mathcal{X} \quad \& \quad V_i \in \mathcal{X}\}$. Let $\text{pa}(\tau)$ be parents of $\tau$, $B_\tau = (B_{ij})_{i \in \tau, j \in \text{pa}(\tau)}$ and

$\Omega_\tau = (\omega_{ij})_{i,j\in\tau}$ sub-matrices of $B$ and $\Omega$, respectively. For a single observation $y \in \mathbb{R}^p$, $y_\tau \mid y_{\mathrm{pa}(\tau)}, B, \Omega \sim \mathcal{N}_{|\tau|}(B_\tau y_{\mathrm{pa}(\tau)}, \Omega_\tau^{-1})$, where, for $d \in \mathbb{N}$, $\mathcal{N}_d(\mu, \Sigma)$ denotes the $d-$dimensional Gaussian distribution with mean $\mu$ and covariance matrix $\Sigma$. Then the joint density of $y \in \mathbb{R}^p$ given $B, \Omega, A$ is

$$p(y \mid B, \Omega, A) = \prod_{\tau\in\mathscr{T}(G(A))} p(y_\tau \mid y_{\mathrm{pa}(\tau)}, B, \Omega). \tag{4}$$

The likelihood is then given by

$$p(y_{1:m} \mid B, \Omega, A) = \prod_{i=1}^{m} p(y_i \mid B, \Omega, A) \tag{5}$$

that is, the observations are i.i.d. given $B, \Omega, A$. These above assumptions on the precision matrix ensure that the corresponding graph satisfies the AMP Markov property as discussed in [2].

We remark that this structure is similar to the structural equation model proposed by [29], where also a Gaussian distribution is assumed for the variables corresponding to the nodes of the graph. The undirected edge $u - v$ is equivalent to the bidirected edge $u \leftrightarrow v$ in [29]. The difference is that the model in [29] requires the entries of the covariance matrix to be 0 if there is no undirected edge between two nodes, while we let the entries of the inverse of the covariance matrix to be 0 if there is no bidirected edge between two nodes. Algebraically, assume the structure equation in matrix form is $Y = BY + \varepsilon$. Let the covariance matrix of $\varepsilon$ be $\Sigma = (\sigma_{ij})$ and the concentration matrix be $\Omega = (\omega_{ij}) = \Sigma^{-1}$. The model in [29] satisfies the property that $\sigma_{ij} = 0$ if is no bidirected edge $Y_i \leftrightarrow Y_j$ in the graph. As a contrast, our model ensure that $\omega_{ij} = 0$ if is no undirected edge $Y_i - Y_j$ in the graph. The parameterization in [29] is for acyclic directed mixed graphs (directed mixed graphs that have no directed cycles), while ours is for chain graphs (do not contain semi-directed cycles), which is a wider class that contains acyclic directed mixed graphs.

2.2. **Prior distribution on the chain graph space.** We specify the prior on the set of possible chain graphs, by specifying a prior on the elements of the corresponding adjacency matrix. For $1 \le i < j \le p$, let

$$\mathbb{P}(a_{ij} = l \mid \pi_{ij}) = \pi_{ij}(l), \quad l = 0, \dots, r \tag{6}$$

where $\pi_{ij} = (\pi_{ij}(0), \dots, \pi_{ij}(r))$. We assume each vector $\pi_{ij}$ to follow a Dirichlet distribution with parameter $\alpha = (\alpha_0, \dots, \alpha_r)$ and the $\pi_{ij}$ to be conditionally independent given $\alpha$. Note that, if available, covariate information could be incorporated to model $\pi_{ij}$, as in [31] for example. Marginally, we have

$$\begin{aligned}\mathbb{P}(a_{ij} = l \mid \alpha) &= \frac{1}{B(\alpha)} \int \prod_{k=0}^{r} \pi_{ij}(k)^{\mathbb{I}(a_{ij}=l)} \prod_{k=0}^{r} \pi_{ij}(k)^{\alpha_s - 1} d\pi_{ij}(0) ... d\pi_{ij}(r) \\ &= \frac{\Gamma(1 + \alpha_{a_{ij}})}{\Gamma(\alpha_{a_{ij}})} \frac{\Gamma\left(\sum_{k=0}^{r} \alpha_k\right)}{\Gamma\left(1 + \sum_{k=0}^{r} \alpha_k\right)}, \qquad l = 0, \dots, r\end{aligned}$$

where

$$B(\alpha) = \frac{\prod_{k=0}^{r} \Gamma(\alpha_k)}{\Gamma\left(\sum_{k=0}^{r} \alpha_k\right)} \tag{7}$$

The prior for a graph $G(A)$ then becomes:

$$p\big((a_{ij})_{i<j} \mid \alpha\big) \propto \mathbb{I}_{\mathcal{C}}(G(A))\Big\{\prod_{i<j}\mathbb{P}(a_{ij}|\alpha)\Big\}. \tag{8}$$

Note that the prior does not integrate to 1, but has a finite mass.

We assume that given $G(A)$, $\Omega$ has a $G-$Wishart prior distribution ([21]) of parameters $\delta, D$, and given $A$ the non-zero entries of $B$ are i.i.d. univariate Gaussian random variables with mean $\xi$ and variance $\kappa$. Let $\mathcal{P}$ be the cone of $p\times p$ real-valued positive definite matrices. Then the joint prior for $\Omega$ and $B$ is given by:

$$p(\Omega, B|A) \propto \mathbb{I}_{\mathcal{P}}(\Sigma(B,\Omega))p(\Omega|A)p(B|A) \tag{9}$$

which does not integrate to 1, but has a finite mass.

2.3. **Posterior inference.** The posterior distribution $\pi$ of all the unknown of interest is given by:

$$\pi(B,\Omega,(a_{ij})_{i<j} \mid y_{1:m},\alpha) \propto p(y_{1:m}|B,\Omega,A)p(\Omega,B|A)p\big((a_{ij})_{i<j}|\alpha\big). \tag{10}$$

The structure of the model bears some resemblance to the models considered in [28, 32]. Besides the graph structure being different (namely, we do not allow bi-directional edges), we also impose the AMP constraint, mathematically represented by the term $\mathbb{I}_{\mathcal{P}}(\Sigma(B,\Omega))$ in $p(\Omega, B|A)$, which leads to a much more complex structure of the graph space than the aforementioned references. As a result, posterior exploration of graph space is more challenging and computationally demanding and requires carefully devised algorithms, in a sense more sophisticated, than the ones considered in [28, 32].

Our approach is to design an adaptive SMC sampler as in [15] (see [10] for the original algorithm and [5] for convergence results). SMC based algorithms offer the advantage of being easily parallelisable, often reducing computation times over serial methods in some scenarios. The strategy involves simulating $N$ samples (particles) to approximate the sequence of densities

$$\pi_t(B,\Omega,(a_{ij})_{i<j} \mid y_{1:m},\alpha) \propto \nu_t(B,\Omega,(a_{ij})_{i<j}|\alpha) \tag{11}$$

where

$$\nu_t(B,\Omega,(a_{ij})_{i<j} \mid \alpha) = \Big[p(y_{1:m} \mid B,\Omega,A)\Big]^{\phi_t} p(\Omega,B \mid A)p\big((a_{ij})_{i<j} \mid \alpha\big) \tag{12}$$

and $0=\phi_0<\cdots<\phi_T=1$. The motivation for this algorithm is well-documented (see the aforementioned references for details), and SMC algorithms have been successfully employed in may contexts to sample from high dimensional posterior distributions.

In the implementation of the algorithm, we require a Markov kernel (e.g., a MCMC kernel) $K_t$, $t\in\{1,2,\ldots,T\}$ that admits $\pi_t$ as an invariant distribution; this step is detailed in the appendix. The algorithm is summarised in Algorithm 1. For notational convenience, we set $u_t=(B_t,\Omega_t,(a_{t,ij})_{i<j})$. In the initialization stage, one simulates exactly from $\pi_0$ using rejection sampling. The sequence of $\{\phi_t\}$ is set as proposed in [37], and the choice of parameters for the MCMC kernel follows [15].

**Algorithm 1** SMC Sampler.

- **Initialize**. Set $t = 0$, for $i \in \{1, \ldots, N\}$ sample $u_0^{(i)}$ from $\pi_0$.
- **Iterate**: Set $t = t+1$. Determine $\phi_t$. If $\phi_t = 1$ set $t = T$, otherwise determine the parameters of the MCMC kernel $K_t$.
  - Resample $(\hat{u}_{t-1}^{(1)}, \ldots, \hat{u}_{t-1}^{(N)})$ using the weights $(w_t^{(1)}, \ldots, w_t^{(N)})$ if effective sample size (ESS) falls below a threshold, where, for $i \in \{1, \ldots, N\}$,

$$w_t^{(i)} = \Big(\frac{\nu_t(u_{t-1}^{(i)})}{\nu_{t-1}(u_{t-1}^{(i)})}\Big)\Big(\sum_{j=1}^{N} \frac{\nu_t(u_{t-1}^{(j)})}{\nu_{t-1}(u_{t-1}^{(j)})}\Big)^{-1}$$

  - Sample $u_t^{(i)} \mid \hat{u}_{t-1}^{(i)}$ from $K_t$ for $i \in \{1, \ldots, N\}$.

3. **Simulation and real data study.** In the following numerical experiments, we set $\delta = 3$ and $D = I_p$ for the G-Wishart prior of $\Omega$, and $\xi = 0, \kappa = 1$ for the distribution of the non-zero elements of $B$. The number of particles for the SMC is $N = 500$. The MCMC steps are Metropolis-Hastings kernels.

3.1. **Simulated example.** In the simulated example, we assume that the $p$ random variables corresponding to the nodes of the graph are independent, i.e., $a_{ij} = 0$ for all $1 \leq i, j \leq p$. This is a benchmark example to evaluate the ability of our algorithm to recover the structure as this is a very simple graph structure. We consider $p = 10$ vertices and $m = 100$ observations. To generate the data, we set the precision matrix $\Omega$ equal to identity matrix $I_p$, and the entries of matrix $B$ are set to be 0. For the Dirichlet prior of $\pi_{ij}$ (the probability of occurrence of each type of edges), we consider $\alpha = (3, 1, 1, 1)$. This prior implies a higher probability of no connection and assumes the same prior probability for any type of edge between two nodes. The reason we prefer this choice of hyper-parameters to $\alpha = (1, 1, 1, 1)$ (which corresponds to a uniform prior) is due to the computational time required by the initialization step to generate the chain graph.

The simulation results are summarized in Figure 1. The effective sample size (ESS) drops very fast as the algorithm begins and goes into a steady state after several resampling procedures. The acceptance rate of $(A, \Omega, B)$ is empirically acceptable as it does not fall below a very small value (is around 0.1 after 100 SMC steps). Using the weighted samples $\{W_T^{(n)}, A_T^{(n)}\}_{n=1}^N$ obtained at the last step of SMC, we estimate the posterior probability $\mathbb{P}(a_{ij} = 0 \mid y_{1:m}, \alpha), 1 \leq i < j \leq p$ by

$$\widehat{\mathbb{P}}(a_{ij} = 0 \mid y_{1:m}, \alpha) = \sum_{p=1}^{N} W_T^{(p)} \mathbb{I}_{\{0\}}(a_{T,ij}^{(p)}), \tag{13}$$

and summarize the results in Table 1. From the table, we can see that all the estimated posterior probabilities are greater than 0.7, and most of them are above 0.9. This suggests that our algorithm is able to recover the structure of the graph (independence) used to generate the data.

3.2. **University graduation rates.** We investigate the performance of our algorithm when the latent graph has a more complex structure. We consider data first presented in [14] that stem from a study for college ranking carried out in 1993. After initial analysis, Druzdzel and Glymour focus on $p = 8$ variables:

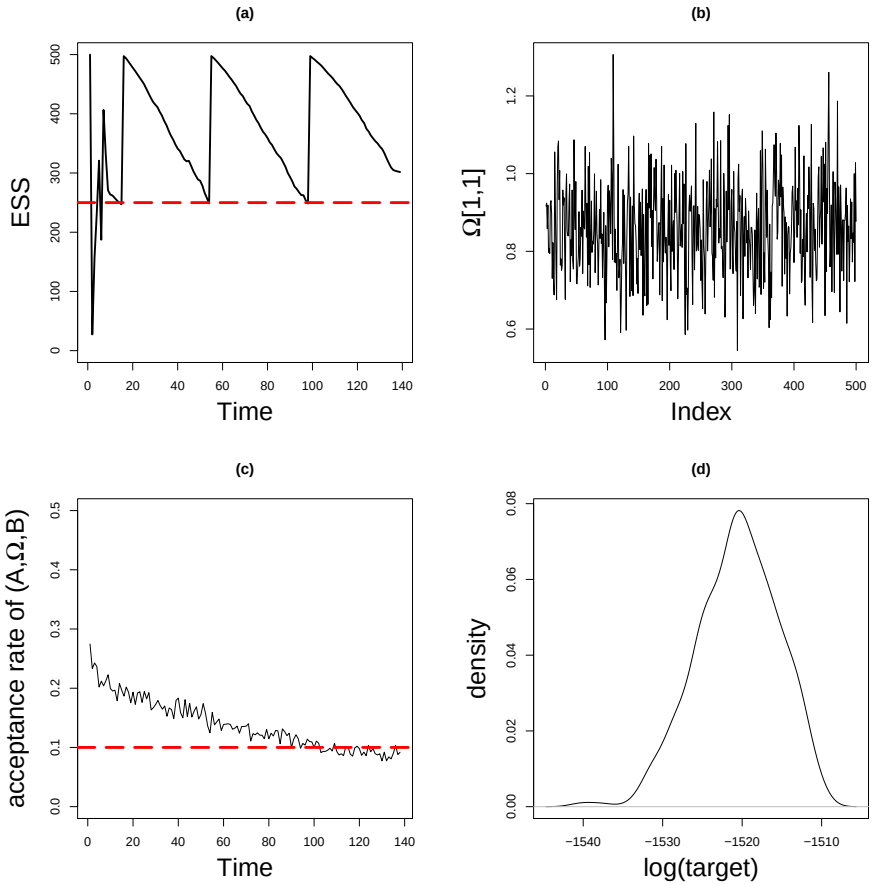



FIGURE 1. Simulation results for the independent case: (a) ESS in each SMC step; (b) plot of $\Omega[1,1]$ across on particles; (c) acceptance rates in each SMC step; (d) distribution of the log(target) (i.e., log of $\pi(B, \Omega, (a_{ij})_{i<j} \mid y_{1:m}, \alpha)$) at the end of the algorithm.

TABLE 1. Posterior probability $\mathbb{P}(a_{ij} = 0 | y_{1:m}, \alpha), 1 \leq i < j \leq p$.

| Nodes | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.882 | 0.892 | 0.920 | 0.932 | 0.920 | 0.870 | 0.874 | 0.908 | 0.934 |
| 2 | | 0.902 | 0.906 | 0.828 | 0.898 | 0.934 | 0.784 | 0.804 | 0.906 |
| 3 | | | 0.914 | 0.890 | 0.880 | 0.890 | 0.908 | 0.900 | 0.890 |
| 4 | | | | 0.900 | 0.866 | 0.882 | 0.770 | 0.918 | 0.900 |
| 5 | | | | | 0.788 | 0.952 | 0.912 | 0.830 | 0.918 |
| 6 | | | | | | 0.806 | 0.932 | 0.906 | 0.908 |
| 7 | | | | | | | 0.904 | 0.738 | 0.916 |
| 8 | | | | | | | | 0.918 | 0.740 |
| 9 | | | | | | | | | 0.952 |

| | |
|---|---|
| *spend* | average spending per student, |
| *strat* | student-teacher ratio, |
| *salar* | faculty salary, |
| *rejr* | rejection rate, |
| *pacc* | percentage of admitted students who accept university's offer, |
| *tstsc* | average test scores of incoming students, |
| *top*10 | class standing of incoming freshmen, and |
| *apgra* | average percentage of graduation. |

Based on $m = 159$ universities, the correlation matrix of these eight variables is estimated in [14]. Conditional on this correlation matrix, [13] obtain a chain graph through an approach that called SIN model selection (a procedure that partition the p-values into three disjoints, a significant set S, an indeterminate set I, and a non-significant set N) with significance level 0.15. To get a chain graph with different AMP and LWF Markov properties, [12] further deleted the undirected edge between *top*10 and *rejr* and the undirected edge between *salar* and *top*10, and introduced an undirected edge between *top*10 and *rejr*. The resulting graph is shown in Figure 2, and the corresponding adjacency matrix is shown in Table 2. This graph has three chain components $\tau_1 = \{spend, strat, salar\}, \tau_2 = \{top10, tstsc, rejr, pacc\}$ and $\tau_3 = \{apgra\}$. The original article [14] provide some insight about the causal relationship between some of the variables. The average spending per student (*spend*), student-teacher ratio (*strat*) and faculty salary (*salar*) are determined based on budget considerations and are not influenced by any of the five remaining variables. Rejection rate (*rejr*) and percentage of students who accept the university's offer from among those who are offered admission (*pacc*) precede the average test scores (*tstsc*) and class standing (*top*10) of incoming freshmen. The latter two are measures taken over matriculating students. Finally, graduation rate (*apgra*) does not influence any of the other variables.

TABLE 2. The adjacency matrix corresponding to chain graph in Figure 2.

| | strat | spend | salar | top10 | tstsc | rejr | pacc | apgra |
|---|---|---|---|---|---|---|---|---|
| strat | 0 | 1 | 1 | 2 | 0 | 0 | 0 | 0 |
| spend | 1 | 0 | 1 | 2 | 2 | 2 | 0 | 0 |
| salar | 1 | 1 | 0 | 0 | 2 | 2 | 2 | 2 |
| top10 | 3 | 3 | 0 | 0 | 1 | 0 | 1 | 0 |
| tstsc | 0 | 3 | 3 | 1 | 0 | 1 | 0 | 2 |
| rejr | 0 | 3 | 3 | 0 | 1 | 0 | 1 | 0 |
| pacc | 0 | 0 | 3 | 1 | 0 | 1 | 0 | 2 |
| apgra | 0 | 0 | 3 | 0 | 3 | 0 | 3 | 0 |

To test the performance of the proposed Bayesian chain graph model, we build an empirical baseline graph through the following procedure. Starting with a graph with only one node denoted by $G$, each time we add a new node to the existing graph. For every node in the subgraph not containing the new node, we consider the candidate graphs that are generated by adding one of four types of correlation (no edge, undirected edge and directed edges in either direction) between the new node and the node in the subgraph. We choose the graph with smallest BIC value

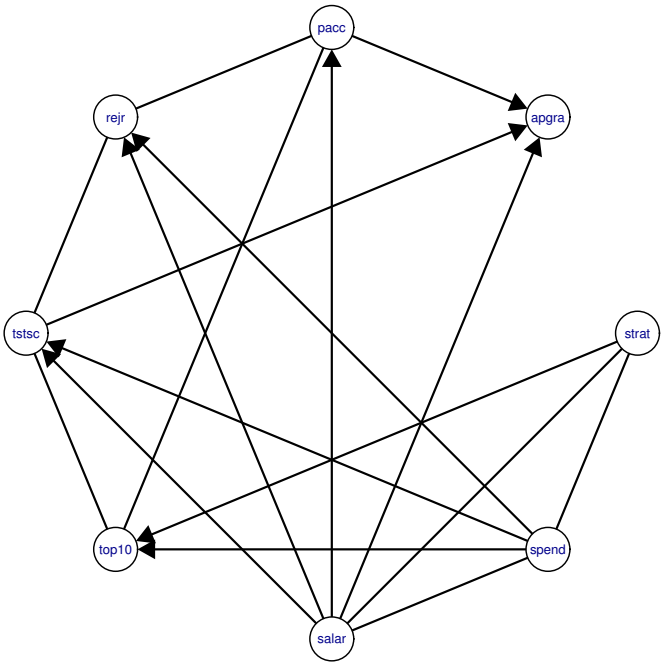



FIGURE 2. Chain Graph Estimate presented in [12].

obtained by fitting a structural equation model (SEM). SEM is a multivariate statistical analysis technique that is commonly used to analyse structural relationships. In general, the formulation of the SEM is given by the basic equation $v = Av + u$, where $v$ and $u$ are vectors of random variables. The parameter matrix $A$ contains regression coefficients and the matrix $P = \mathbb{E}(uu')$ gives covariances among the elements of $u$. The chain graph model described in this work can be written as $Y = BY + U$, where $U$ is a multivariate Gaussian random vector with covariance matrix $\Omega^{-1}$. This is consistent with the structural equation model formulation. Figure 3 shows the derived graph and Table 3 shows the corresponding adjacency matrix.

TABLE 3. The adjacency matrix corresponding to the chain graph in Figure 3.

| | strat | spend | salar | top10 | tstsc | rejr | pacc | apgra |
|---|---|---|---|---|---|---|---|---|
| strat | 0 | 1 | 2 | 2 | 2 | 2 | 2 | 2 |
| spend | 1 | 0 | 1 | 2 | 2 | 2 | 2 | 2 |
| salar | 3 | 1 | 0 | 1 | 2 | 2 | 3 | 2 |
| top10 | 3 | 3 | 1 | 0 | 2 | 1 | 0 | 2 |
| tstsc | 3 | 3 | 3 | 3 | 0 | 1 | 0 | 2 |
| rejr | 3 | 3 | 3 | 1 | 1 | 0 | 3 | 0 |
| pacc | 3 | 3 | 2 | 0 | 0 | 2 | 0 | 2 |
| apgra | 3 | 3 | 3 | 3 | 3 | 0 | 3 | 0 |

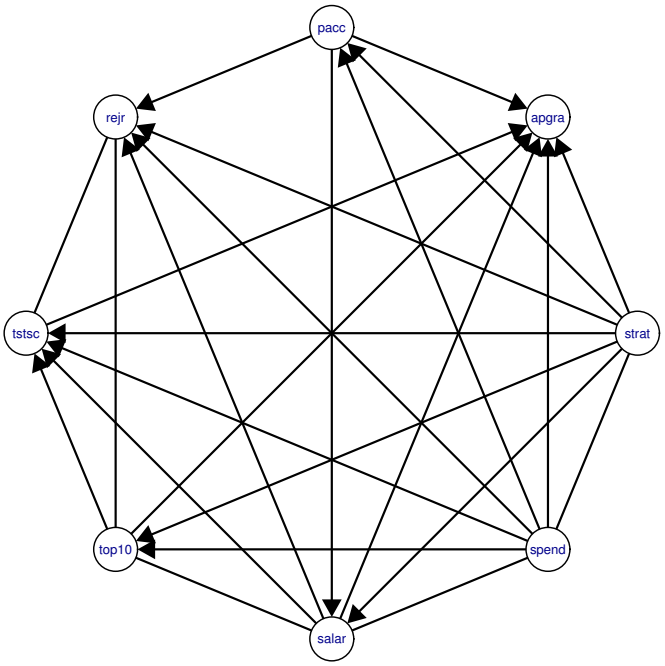



FIGURE 3. Empirical Graph.

We compare posterior inference from our model with the chain graphs in Figure 2 and Figure 3. For the SMC sampler, we set the number of samples $N = 5000$. For the Dirichlet prior, we first consider a prior based on the analysis of [12], i.e., choosing $\alpha = (0.39, 0.25, 0.36, 0.05)$ by matching the probabilities of each type of edges in Figure 2. More precisely, the number of no-edges, undirected edges and directed edges from $i$ to $j$ are 11, 7 and 10 according to the graphs in Figure 2. So the corresponding proportions of these three types of edge are the first three components of $\alpha$. The last component of $\alpha$ is chosen to be 0.05 to ensure the occurrence of directed edge from $j$ to $i$ and the probability of this type of edge is not large compared to the other three types. We also perform posterior inference with $\alpha = (1, 1, 1, 1)$, which is a uniform prior, and with $\alpha = (1, 3, 3, 3)$, which favours more connections. The remaining parameters are specified as in the previous section.

Based on the weighted samples $\{W_T^{(n)}, A_T^{(n)}\}_{n=1}^N$, we first estimate the posterior probability of occurrence of each edge $\widehat{\mathbb{P}}(a_{ij} = k \mid y_{1:m}, \alpha)$, $1 \le i < j \le p$ by

$$\widehat{\mathbb{P}}(a_{ij} = k \mid y_{1:m}, \alpha) = \sum_{p=1}^{N} W_T^{(p)} \mathbb{I}_{\{k\}}(a_{T,ij}^{(p)}), \quad k = 0, 1, 2, 3. \tag{14}$$

Then the entries of the estimated adjacency matrix $A$ are given by

$$a_{ij} = \underset{k=0,1,2,3}{\arg\max}\, \widehat{\mathbb{P}}(a_{ij} = k \mid y_{1:m}, \alpha). \tag{15}$$

Figure 4-6 show the estimated chain graphs obtained under each prior. The structure of the chain graph obtained from setting $\alpha = (0.39, 0.25, 0.36, 0.05)$ is obviously more similar to Figure 2 than that obtained using $\alpha = (1, 1, 1, 1)$. This is not surprising since the prior with $\alpha = (0.39, 0.25, 0.36, 0.05)$ is very informative and

forces the posterior distribution of the edges to be similar to Figure 2. The estimated chain graph obtained setting the hyper-parameters equal to $\alpha = (1, 3, 3, 3)$ favours more connections compared with other two priors. These figures also show that the type of edges is affected by $\alpha$. For example, if we look at the edges connecting tstsc and other nodes, we can see that there are more directed edges from tstsc to nodes strat, spend, salar and top10 in Figure 6 compared with Figure 4. The reason is that the probability of directed edges from tstsc to these 4 nodes (strat, spend, salar and top10) is $3/(1 + 3 + 3 + 3) = 0.3$ for prior with $\alpha = (1, 3, 3, 3)$, and is $0.05/(0.39, 0.25, 0.36, 0.05) = 0.047$ for prior with $\alpha = (0.39, 0.25, 0.36, 0.05)$. If we compare our results with the conclusions in [14], the estimated chain graph obtained under the informative prior indeed shows that *spend*, *strat* and *salar* are *parents* of other variables, and *apgra* is a *child* of the others. However, it does not show that *rejr* and *pacc* precede *tstsc* and *top*10 (the graph just shows the opposite relation). Similarly, the graph obtained under the prior with $\alpha = (1, 3, 3, 3)$ shows that *tstsc* is a parent of *salar*, *strat* and *spend*, which is a contrast to the available prior information. This can be improved by modifying the ordering of nodes and choosing a small value of $\alpha_3$.

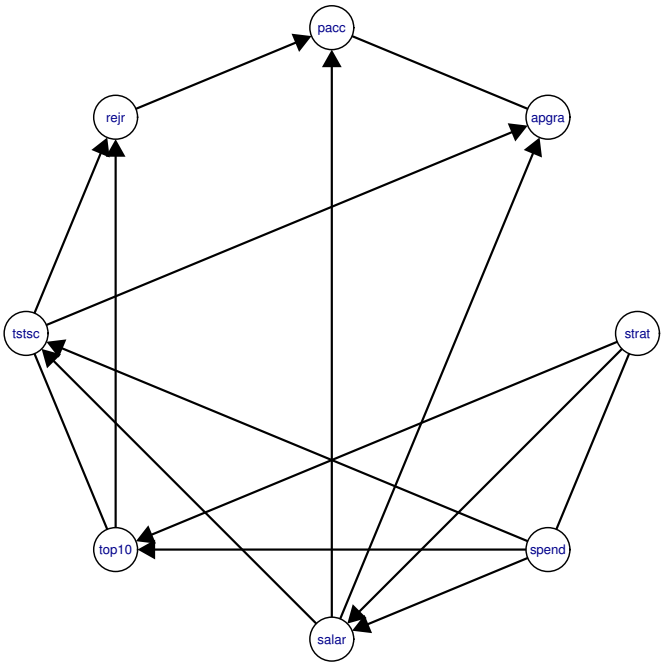



FIGURE 4. posterior estimated chain graph using a Dirichlet prior with $\alpha = (0.39, 0.25, 0.36, 0.05)$.

Finally, we compare our results using SEM. We use package `sem` in R to fit a SEM to these data. Note that when fitting the SEM model the graph topology is fixed. Table 4 presents a brief summary of three different chain graphs using the *sem* function. From the table, we can see that the chain graph selected by the Bayesian chain graph model with prior $\alpha = (0.39, 0.25, 0.36, 0.05)$ has smaller AIC and BIC values compared with the graph selected by the SIN model selection

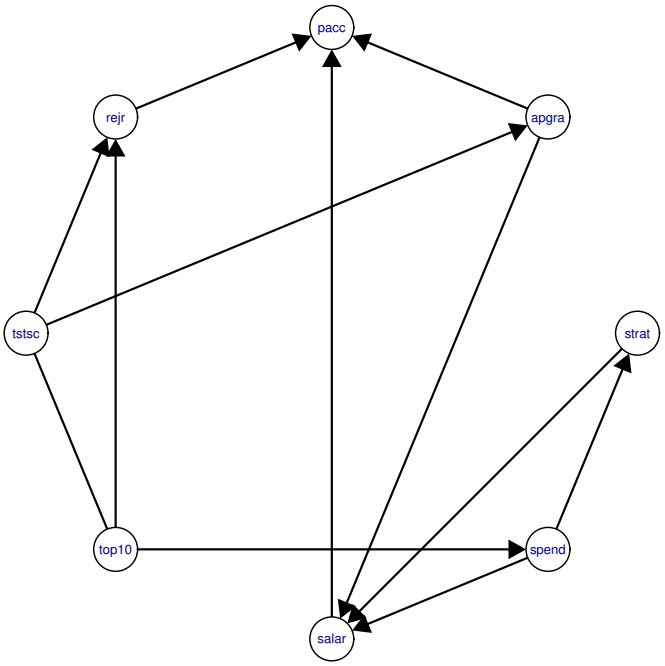



FIGURE 5. posterior estimated chain graph using a Dirichlet prior with $\alpha = (1, 1, 1, 1)$.

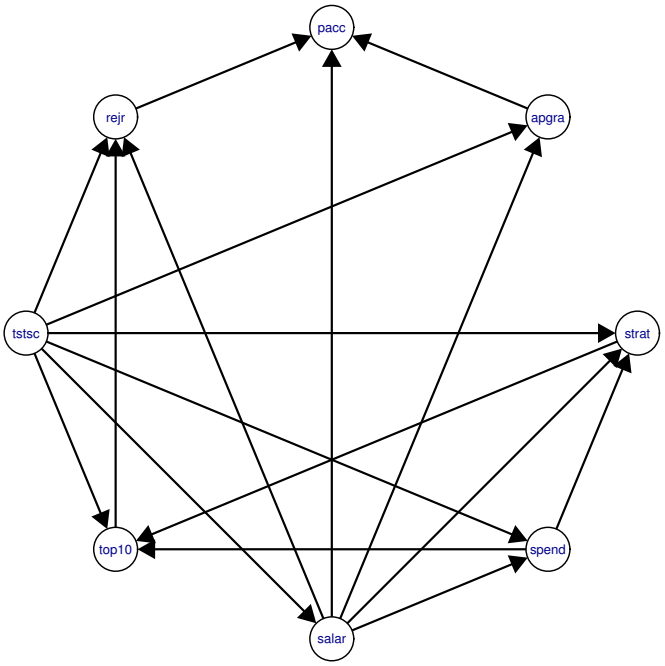



FIGURE 6. posterior estimated chain graph using a Dirichlet prior with $\alpha = (1, 3, 3, 3)$.

procedure and the empirical graph, which suggests that our algorithm can take advantage of available prior information and perform better.

TABLE 4. Summaries of different chain graphs using package SEM.

| Base chain graph | | | | Chain graph selected by SIN | |
|---|---|---|---|---|---|
| Edge | p-value | Edge | p-value | Edge | p-value |
| strat — spend | 1.630e-14 | pacc → salar | 1.137e-06 | strat — spend | 1.630e-14 |
| strat → salar | 1.382e-06 | pacc → rejr | 1.470e-03 | strat — salar | 1.082e-05 |
| spend — salar | 2.629e-11 | strat → apgra | 8.237e-02 | strat → top10 | 1.935e-09 |
| strat → top10 | 4.743e-07 | spend → apgra | 6.067e-02 | spend — salar | 7.156e-13 |
| spend → top10 | 2.822e-28 | salar → apgra | 4.794e-03 | spend → top10 | 5.979e-34 |
| top10 — salar | 8.931e-03 | top10 → apgra | 4.253e-01 | spend → tstsc | 3.995e-12 |
| strat → tstsc | 3.140e-03 | tstsc → apgra | 1.096e-10 | spend → rejr | 2.909e-03 |
| spend → tstsc | 6.634e-01 | pacc → apgra | 2.711e-03 | salar → tstsc | 2.350e-05 |
| salar → tstsc | 2.008e-04 | | | salar → rejr | 1.323e-03 |
| top10 → tstsc | 6.831e-19 | | | salar → pacc | 1.827e-14 |
| strat → rejr | 1.954e-01 | | | salar → apgra | 1.570e-02 |
| spend → rejr | 3.621e-03 | | | top10 — tstsc | 1.256e-09 |
| salar → rejr | 2.575e-04 | | | top10 — pacc | 5.020e-01 |
| top10 — rejr | 1.816e-04 | | | tstsc — rejr | 8.297e-03 |
| tstsc — rejr | 1.003e-02 | | | tstsc → apgra | 8.352e-19 |
| strat → pacc | 2.585e-02 | | | rejr — pacc | 5.617e-03 |
| spend → pacc | 4.109e-07 | | | pacc → apgra | 5.481e-03 |
| AIC | | BIC | | AIC | BIC |
| 67.887 | | -13.319 | | 80.838 | -24.919 |
| Chain graph selected by algorithm ($\alpha = (0.39, 0.25, 0.36, 0.05)$) | | Chain graph selected by algorithm ($\alpha = (1, 1, 1, 1)$) | | Chain graph selected by algorithm ($\alpha = (1, 3, 3, 3)$) | |
| Edge | p-value | Edge | p-value | Edge | p-value |
| strat — spend | 1.630e-14 | spend → strat | 2.150e-52 | spend → strat | 1.536e-42 |
| strat → salar | 1.727e-06 | strat → salar | 9.127e-07 | salar → strat | 4.952e-06 |
| strat → top10 | 3.597e-09 | spend → salar | 2.484e-24 | strat → top10 | 1.636e-07 |
| spend → salar | 4.956e-33 | top10 → spend | 2.068e-25 | tstsc → strat | 8.237e-01 |
| spend → top10 | 1.086e-33 | salar → pacc | 7.304e-10 | salar → spend | 9.659e-11 |
| spend → tstsc | 4.059e-12 | apgra → salar | 3.751e-11 | spend → top10 | 3.881e-10 |
| salar → tstsc | 5.524e-05 | top10 — tstsc | 2.163e-14 | tstsc → spend | 2.886e-08 |
| salar → pacc | 1.012e-18 | top10 → rejr | 2.504e-03 | tstsc → salar | 3.654e-27 |
| salar → apgra | 3.201e-05 | tstsc → rejr | 2.847e-04 | salar → rejr | 8.844e-02 |
| top10 — tstsc | 2.295e-10 | tstsc → apgra | 1.428e-43 | salar → pacc | 1.062e-08 |
| top10 → rejr | 1.951e-03 | rejr → pacc | 1.480e-04 | salar → apgra | 1.036e-04 |
| tstsc → rejr | 2.348e-04 | apgra → pacc | 3.587e-03 | tstsc → top10 | 1.033e-20 |
| tstsc → apgra | 1.367e-18 | | | top10 → rejr | 9.160e-03 |
| rejr → pacc | 1.032e-03 | | | tstsc → rejr | 5.443e-03 |
| pacc — apgra | 5.315e-03 | | | tstsc → apgra | 2.644e-17 |
| | | | | rejr → pacc | 3.048e-04 |
| | | | | apgra → pacc | 5.759e-03 |
| AIC | BIC | AIC | BIC | AIC | BIC |
| 61.372 | -50.522 | 102.93 | -18.169 | 58.401 | -47.356 |

3.3. **Tenofovir study.** In this section, we illustrate our Bayesian model for AMP chain graphs on a real data application from the RMP-02/MTN-006 study ([3]). Tenofovir (TFV) is a medication used to treat chronic hepatitis B and to prevent and treat HIV. TFV 1% gel demonstrated 39% protective efficacy in women using the gel within 12 hours before and after sexual activity in the Centre for the AIDs Programme of Research in South Africa 004 study [1]. Daily dosing of tenofovir disoproxil fumarate (TDF)/emtricitabine provides 62 to 73% protection against HIV transmission in serodiscordant men and women enrolled in the Partners PrEP study [4]. The RMP-02 study was designed to evaluate the systemic safety and biologic effects of oral TFV combined with a gel formulation of the drug for application rectally and vaginally. The study enrolled 18 patients, all of whom received a single oral dose of TFV, and randomized each patient to receive either the gel formulation or a placebo several weeks later. Details about the phase 1 study are given in [3]. [27] present analyses of the ancillary studies into the treatment's biologic effects.

The biologic effects we examine here concern the pharmacokinetics (PK) of TFV and its active metabolite tenofovir diphosphate (TFVdp), as well as the pharmacodynamics of the drug. The PK studies evaluate subjects' TFV and TFVdp concentrations in multiple physiologic compartments (i.e., tissues and cells) across multiple time points during the study. Table 5 lists the compartments.

TABLE 5. Tissues and cell types examined in the PK studies

| Compound | Compartment | Notation |
|---|---|---|
| TFV | Blood plasma | $\text{TFV}_{plasma}$ |
| TFV | Rectal biopsy tissue | $\text{TFV}_{tissue}$ |
| TFV | Rectal fluid | $\text{TFV}_{rectal}$ |
| TFVdp | Rectal biopsy tissue | $\text{TFVdp}_{tissue}$ |
| TFVdp | Total mononuclear cells in rectal tissue | $\text{Total}_{\text{MMC}}$ |
| TFVdp | $CD4^+$ lymphocytes from MMC | $\text{CD4}^+_{\text{MMC}}$ |
| TFVdp | $CD4^-$ lymphocytes from MMC | $\text{CD4}^-_{\text{MMC}}$ |

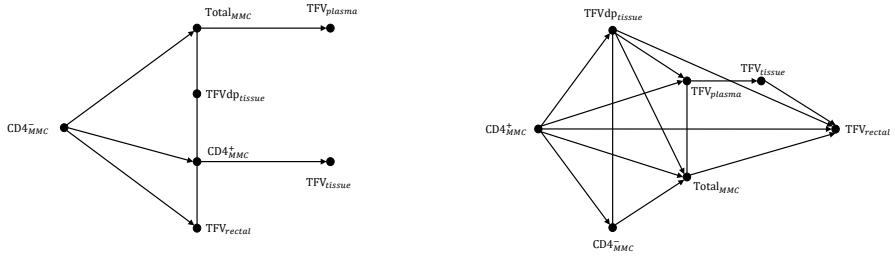
[27] demonstrate that tissue HIV infectibility (cumulative p24) is correlated with *in vivo* concentrations of both TFV and TFVdp. Statistically significant, nonlinear dose-response relationships with reduced tissue infectibility are found for one TFV compartment and four TFVdp compartments; the dose-response relationships are highly significant for TFVdp in whole rectal tissue, $\text{CD4}^+_{\text{MMC}}$, $\text{CD4}^-_{\text{MMC}}$ and $\text{Total}_{\text{MMC}}$ compartments. Furthermore, [36] conduct a comprehensive pharmacokinetic study of rectally administered TFV gel that describes the distribution of TFV and TFVdp into various tissue compartments relevant to HIV infection. They argue that TFV rectal fluid concentrations may be reasonable bio-indicators of plasma and rectal tissue concentrations, making it easier to estimate adherence and TFV concentrations in the target tissue. Therefore, the correlations between the TFV and TFVdp in compartments can be helpful in studying HIV suppression procedure and providing a measure of drug efficacy, enabling more advanced population pharmacokinetic modelling methods. In the following analysis, we investigate the correlation structure of p = 7 concentration levels in Table 5 collected at visit 12. From clinical knowledge, we would expect the following associations:

- $\text{TFV}_{plasma}$ is associated with $\text{TFV}_{tissue}$ (blood levels and tissue levels),
- $\text{TFV}_{tissue}$ is associated with $\text{TFV}_{rectal}$ (rectal tissue and rectal fluid),
- $\text{TFV}_{tissue}$ is associated with $\text{Total}_{\text{MMC}}$ (rectal tissue and mononuclear cells in rectal tissue),
- $\text{Total}_{\text{MMC}}$ is associated with $\text{CD4}^+_{\text{MMC}}$ and $\text{CD4}^-_{\text{MMC}}$ (total and constituents).

We normalise the observations of each variable to have zero mean and standard deviation of one. The number of observations is $m = 11$, and the number of particles is set to $N = 5000$. We fit the model using as hyper-parameters in the Dirichlet prior both $\alpha = (1, 1, 1, 1)$, which corresponds to the uniform prior, and $\alpha = (1, 3, 3, 3)$, which favours the presence of connections. This latter prior choice is more suitable for small sample sizes. The remaining parameters are set as in the previous section.

Based on weighted samples $\{W_T^{(n)}, (A, B, \Omega)_T^{(n)}\}_{n=1}^N$, we calculate the posterior probabilities $\mathbb{P}\big((A, B, \Omega)_T^{(n)} \mid y_{1:n}, \alpha\big)$. As posterior estimate of the resulting chain graph, we report the one obtained from the adjacency matrix $A_T^{(n^*)}$ where

$n^* = \arg\max_n \mathbb{P}\big((A, B, \Omega)_T^{(n)} \mid y_{1:n}, \alpha\big)$, and it is shown in Figure 7 (for both prior settings). Obviously, the chain graph obtained under the prior with $\alpha = (1, 3, 3, 3)$ has more connections. Moreover, this graph also shows that the $\text{TFV}_{tissue}$ is related to $\text{TFV}_{plasma}$ and $\text{TFV}_{rectal}$, the $\text{TFV}_{tissue}$ is associated with $\text{Total}_{\text{MMC}}$ through $\text{TFV}_{plasma}$, and $\text{CD4}^+_{\text{MMC}}$ causes $\text{Total}_{\text{MMC}}$ and $\text{CD4}^-_{\text{MMC}}$. This is consistent with the clinical knowledge mentioned before. However, some of these associations are missing in the chain graph obtained under the uniform prior, e.g. the edge between $\text{TFV}_{plasma}$ and $\text{TFV}_{tissue}$ (they are independent given $\text{Total}_{\text{MMC}}$ and $\text{CD4}^+_{\text{MMC}}$ under the AMP chain graph property).




(a) $\alpha = (1, 1, 1, 1)$ (b) $\alpha = (1, 3, 3, 3)$

FIGURE 7. Chain graph with highest posterior probability.

We investigate some of our results by considering a SEM model using the *sem* function in the R package `sem`. We summarise the results in Table 6. The AIC and BIC values of the model corresponding to the chain graph obtained under the uniform prior are missing, which may due to the singular Hessian matrix when estimating the covariance matrix of the parameter. From the table, it is evident that the p-values of some edges are large. For example, the edge $\text{CD4}^+_{\text{MMC}} \to \text{TFV}_{tissue}$ in the chain graph obtained under uniform prior has a p-value 0.704. These results suggest that these relationships are not significant in a SEM model.

When inadequate fit of a structural equation model is observed, model modification is often conducted followed by retesting of the modified model. The most popular statistic is the modification index, which is a chi-square score test statistic with degree of freedom one. The modification index provides an estimated value in which the model's chi-square test statistic would decrease if the corresponding parameter is added to the model and respecified as a free parameter. We perform model modification using the *modIndices* function in `sem` package, which calculates modification indices and estimates parameter changes for the fixed and constrained parameters in a structural equation model. Table 7 shows the five largest modification indices for both the $A$ matrix and $P$ matrix for the model corresponding to the chain graph obtained under the Dirichlet prior with $\alpha = (1, 3, 3, 3)$. The modification indices suggest that a better fit to the data would be achieved by adding association between $\text{TFV}_{rectal}$ and $\text{TFV}_{plasma}$ to the model. The small sample size makes it obviously challenging to estimate associations. Furthermore, the level in rectal tissue may depend on whether or not the patient received the TFV gel or the placebo, which may make the correlation of $\text{TFV}_{tissue}$ with other variables unclear.

TABLE 6. Summaries of two different chain graphs using package SEM.

| Chain graph selected by algorithm ($\alpha = (1,1,1,1)$) | | Chain graph selected by algorithm ($\alpha = (1,3,3,3)$) | |
|---|---|---|---|
| Edge | p-value | Edge | p-value |
| $CD4^-_{MMC} \rightarrow Total_{MMC}$ | 9.881e-19 | $CD4^+_{MMC} \rightarrow TFVdp_{tissue}$ | 5.687e-01 |
| $CD4^-_{MMC} \rightarrow CD4^+_{MMC}$ | 4.091e-81 | $CD4^+_{MMC} \rightarrow CD4^-_{MMC}$ | 2.941e-46 |
| $CD4^-_{MMC} \rightarrow TFV_{rectal}$ | 1.496e-01 | $CD4^+_{MMC} \rightarrow TFV_{plasma}$ | 1.210e-02 |
| $Total_{MMC}$ — $TFVdp_{tissue}$ | 1.589e-02 | $CD4^+_{MMC} \rightarrow Total_{MMC}$ | 7.028e-10 |
| $TFVdp_{tissue}$ — $CD4^+_{MMC}$ | 1.812e-02 | $CD4^+_{MMC} \rightarrow TFV_{rectal}$ | 1.874e-03 |
| $CD4^+_{MMC}$ — $TFV_{rectal}$ | 8.583e-03 | $TFVdp_{tissue}$ — $CD4^-_{MMC}$ | 8.477e-02 |
| $Total_{MMC} \rightarrow TFV_{plasma}$ | 1.815e-03 | $TFVdp_{tissue} \rightarrow TFV_{rectal}$ | 2.991e-01 |
| $CD4^+_{MMC} \rightarrow TFV_{tissue}$ | 7.043e-01 | $TFVdp_{tissue} \rightarrow TFV_{plasma}$ | 2.584e-01 |
| | | $TFVdp_{tissue} \rightarrow Total_{MMC}$ | 1.162e-13 |
| | | $CD4^-_{MMC} \rightarrow Total_{MMC}$ | 2.352e-22 |
| | | $TFV_{plasma} \rightarrow TFV_{tissue}$ | 4.259e-01 |
| | | $TFV_{plasma}$ — $Total_{MMC}$ | 5.719e-02 |
| | | $TFV_{tissue} \rightarrow TFV_{rectal}$ | 7.550e-01 |
| | | $Total_{MMC} \rightarrow TFV_{rectal}$ | 2.337e-02 |
| AIC | BIC | AIC | BIC |
| Inf | Inf | 46.875 | -11.909 |

TABLE 7. Summaries of modification indices for the model corresponding to the chain graph obtained under prior with $\alpha = (1,3,3,3)$.

| 5 largest modification indices, $A$ matrix (regression coefficients) | | 5 largest modification indices, $P$ matrix (variances/covariances) | |
|---|---|---|---|
| $TFV_{rectal} \rightarrow Total_{MMC}$ | 3.281 | $TFV_{rectal}$ — $Total_{MMC}$ | 3.394 |
| $TFV_{rectal} \rightarrow TFV_{plasma}$ | 2.219 | $TFV_{rectal}$ — $TFV_{plasma}$ | 2.881 |
| $CD4^-_{MMC} \rightarrow TFV_{rectal}$ | 0.709 | $TFV_{rectal}$ — $CD4^-_{MMC}$ | 0.709 |
| $TFV_{rectal} \rightarrow TFVdp_{tissue}$ | 0.654 | $TFV_{rectal}$ — $TFVdp_{tissue}$ | 0.709 |
| $TFV_{rectal} \rightarrow CD4^-_{MMC}$ | 0.555 | $TFV_{tissue}$ — $TFVdp_{tissue}$ | 0.389 |

4. **Conclusions.** In this article we propose a novel Bayesian model for latent AMP chain graphs, for which observations are available only on the nodes of the graph. Posterior inference is performed through a specially devised SMC algorithm. We investigate the ability of the model to recover a range of structures, also when prior knowledge is available. The performance of the SMC sampler is stable and consistent in our numerical study. However, the sampler is not suitable when the number of nodes $p$ is large, as the initializing step is difficult. Moreover, the proposed algorithm does not scale well with respect to $p$. The computational cost of computing the probabilities in the adjacency matrix is $\mathcal{O}(p^2)$, and the computation of the normalizing constant in the $G$-Wishart distribution is quite expensive when $p$ is large (approximately $\mathcal{O}(p^3)$).

Several extensions of this work are possible. First, the algorithm can be extended to large $p$ by choosing a more efficient initial proposal $q$ (for example, a Dirichlet prior that endorses sparse structure of the adjacency matrix, like $\alpha = (\alpha_0, \ldots, \alpha_r)$ with $\alpha_0 > \alpha_1 + \cdots + \alpha_r$) and by actually exploiting parallel computing techniques.

Second, the model can be extended to accommodate multiple groups of observations, allowing borrowing information across groups or time periods. Third, we could avoid sampling $\Omega$ and $B$ by using a Laplace approximation when calculating the posterior probability.

**Acknowledgments.** We acknowledge the contribution of the study team and participants of the RMP-02/MTN-006 study and thank them for sharing the study data. Partial support for this research came from grant numbers U19 AI060614 and UM1 AI106707 from the U.S. National Institutes of Health. GLR was partially supported by grant P30CA006973 from the U.S. National Cancer Institute.

**Appendix** A. **MCMC kernel.** The updates are taken in the order $A, \Omega, B$ together, then $\Omega$ and $B$. We use Metropolis-Hastings steps. The updates for $\Omega$ and $B$ are preformed element-wise. For each element of $\Omega$ and $B$ a Gaussian random walk with variance $\sigma_1^2$ is used as proposal. The acceptance probability is calculated as usual and note that if any of the constraints on $\Omega$ or $B$ are violated, then the acceptance probability is zero.

The update for $A, \Omega, B$ together is much more complicated. We denote with $(A^{(n)}, \Omega^{(n)}, B^{(n)})$ the current values of $A, \Omega, B$ and the proposed states with $(A_c^{(n)}, \Omega_c^{(n)}, B_c^{(n)})$. The move proceeds by picking an $(a_{ij}^{(n)})_{i<j}$ uniformly at random and then proposing one of the other possible 3 values with uniform probability. Note that at this stage, we will check if the proposed graph is a chain graph and if not, the move is rejected instantly. Based upon this proposal, we propose new values of $\Omega$ and $B$ as detailed below. Suppose the selected edge is $(i, j), i < j$, consider the following scheme:

(i) If the transformation of the selected edge is "$0 \to 1$", we take $\Omega_c^{(n)} = \Omega^{(n)}$ except for elements $\Omega_c^{(n)}[i, j]$ and $\Omega_c^{(n)}[j, i]$. For these elements, we set $\Omega_c^{(n)}[i, j] \sim \mathcal{N}(0, \sigma_2^2)$ and $\Omega_c^{(n)}[j, i] = \Omega_c^{(n)}[i, j]$. Finally, we set $B_c^{(n)} = B^{(n)}$.

(ii) If the transformation of the selected edge is "$0 \to 2$", we take $B_c^{(n)} = B^{(n)}$ except for element $B_c^{(n)}[j, i]$. For this element, we set $B_c^{(n)}[j, i] \sim \mathcal{N}(\xi, \kappa)$. Finally, we set $\Omega_c^{(n)} = \Omega^{(n)}$.

(iii) If the transformation of the selected edge is "$0 \to 3$", we take $B_c^{(n)} = B^{(n)}$ except for element $B_c^{(n)}[i, j]$. For this element, we set $B_c^{(n)}[i, j] \sim \mathcal{N}(\xi, \kappa)$. Finally, we set $\Omega_c^{(n)} = \Omega^{(n)}$.

(iv) If the transformation of the selected edge is "$1 \to 0$", we take $\Omega_c^{(n)} = \Omega^{(n)}$ except for elements $\Omega_c^{(n)}[i, j]$ and $\Omega_c^{(n)}[j, i]$. For these elements, we set $\Omega_c^{(n)}[i, j] = \Omega_c^{(n)}[j, i] = 0$. Finally, we set $B_c^{(n)} = B^{(n)}$.

(v) If the transformation of the selected edge is "$1 \to 2$", we take $\Omega_c^{(n)} = \Omega^{(n)}$ except for elements $\Omega_c^{(n)}[i, j]$ and $\Omega_c^{(n)}[j, i]$. For these elements, we set $\Omega_c^{(n)}[i, j] = \Omega_c^{(n)}[j, i] = 0$. Finally, we take $B_c^{(n)} = B^{(n)}$ except for element $B_c^{(n)}[j, i]$. For this element, we set $B_c^{(n)}[j, i] \sim \mathcal{N}(\xi, \kappa)$.

(vi) If the transformation of the selected edge is "$1 \to 3$", we take $\Omega_c^{(n)} = \Omega^{(n)}$ except for elements $\Omega_c^{(n)}[i, j]$ and $\Omega_c^{(n)}[j, i]$. For these elements, we set $\Omega_c^{(n)}[i, j] = \Omega_c^{(n)}[j, i] = 0$. Finally, we take $B_c^{(n)} = B^{(n)}$ except for element $B_c^{(n)}[i, j]$. For this element, we set $B_c^{(n)}[i, j] \sim \mathcal{N}(\xi, \kappa)$.

(vii) If the transformation of the selected edge is "$2 \to 0$", we take $B_c^{(n)} = B^{(n)}$ except for element $B_c^{(n)}[j,i]$. For this element, we set $B_c^{(n)}[j,i] = 0$. Finally, we set $\Omega_c^{(n)} = \Omega^{(n)}$.

(viii) If the transformation of the selected edge is "$2 \to 1$", we take $B_c^{(n)} = B^{(n)}$ except for element $B_c^{(n)}[j,i]$. For this element, we set $B_c^{(n)}[j,i] = 0$. Finally, we take $\Omega_c^{(n)} = \Omega^{(n)}$ except for elements $\Omega_c^{(n)}[i,j]$ and $\Omega_c^{(n)}[j,i]$. For these elements, we set $\Omega_c^{(n)}[i,j] \sim \mathcal{N}(0,\sigma_2^2)$ and $\Omega_c^{(n)}[j,i] = \Omega_c^{(n)}[i,j]$.

(ix) If the transformation of the selected edge is "$2 \to 3$", we take $B_c^{(n)} = B^{(n)}$ except for elements $B_c^{(n)}[j,i]$ and $B_c^{(n)}[i,j]$. For these elements, we set $B_c^{(n)}[j,i] = 0$, and $B_c^{(n)}[i,j] \sim \mathcal{N}(\xi,\kappa)$. Finally, we set $\Omega_c^{(n)} = \Omega^{(n)}$.

(x) If the transformation of the selected edge is "$3 \to 0$", we take $B_c^{(n)} = B^{(n)}$ except for element $B_c^{(n)}[i,j]$. For this element, we set $B_c^{(n)}[i,j] = 0$. Finally, we set $\Omega_c^{(n)} = \Omega^{(n)}$.

(xi) If the transformation of the selected edge is "$3 \to 1$", we take $B_c^{(n)} = B^{(n)}$ except for element $B_c^{(n)}[i,j]$. For this element, we set $B_c^{(n)}[i,j] = 0$. Finally, we take $\Omega_c^{(n)} = \Omega^{(n)}$ except for elements $\Omega_c^{(n)}[i,j]$ and $\Omega_c^{(n)}[j,i]$. For these elements, we set $\Omega_c^{(n)}[i,j] \sim \mathcal{N}(0,\sigma_2^2)$ and $\Omega_c^{(n)}[j,i] = \Omega_c^{(n)}[i,j]$.

(xii) If the transformation of the selected edge is "$3 \to 2$", we take $B_c^{(n)} = B^{(n)}$ except for elements $B_c^{(n)}[i,j]$ and $B_c^{(n)}[j,i]$. For these elements, we set $B_c^{(n)}[i,j] = 0$, and $B_c^{(n)}[j,i] \sim \mathcal{N}(\xi,\kappa)$. Finally, we set $\Omega_c^{(n)} = \Omega^{(n)}$.

The acceptance probabiity for this move is easily calculated.

## REFERENCES


[1] K. Q. Abdool, K. S. S. Abdool and J. A. Frohlich, Effectiveness and safety of tenofovir gel, an antiretroviral microbicide, for the prevention of HIV infection in women, *Science*, **329** (2010), 1168–1174.

[2] S. A. Andersson, D. Madigan and M. D. Perlman, Alternative Markov properties for chain graphs, *Scand. J. Statist.*, **28** (2001), 33–85.

[3] P. A. Anton, R. D. Cranston, A. Kashuba, C. W. Hendrix, N. N. Bumpus, N. R. Harman, J. Elliott, L. Janocko, E. Khanukhova, R. Dennis, W. G. Cumberland, C. Ju, A. C. Dieguez, C. Mauck and I. McGowan, RMP-02/MTN-006: A phase rectal safety, acceptability, pharmacokinetic, and pharmacodynamic study of tenofovir 1% gel compared with oral tenofovir disoproxil fumarate, *AIDS Res Hum Retroviruses*, **28** (2012), 1412–1421.

[4] J. M. Baeten, D. Donnell and P. Ndase, *et al*, Antiretroviral prophylaxis for HIV prevention in heterosexual men and women, *N Engl J Med*, **367** (2012), 399–410.

[5] A. Beskos, A. Jasra, N. Kantas and A. Thiery, On the convergence of adaptive sequential Monte Carlo, *Ann. Appl. Probab.*, **26** (2016), 1111–1146.

[6] B. C. Boerebach, K. M. Lombarts, C. Keijzer, M. J. Heineman and O. A. Arah, The teacher, the physician and the person: How faculty's teaching performance influences their role modeling, *PLoS One*, **7** (2012), e32089.

[7] K. Bollen, *Structural Equation Models with Latent Variables*, Wiley: New York, 1989.

[8] C. M. Carvalho and M. West, Dynamic matrix-variate graphical modelso, *Bayesian Anal.*, **2** (2007), 69–97.

[9] H. Chun, X. Zhang and H. Zhao, Gene regulation network inference with joint sparse Gaussian graphical models, *J. Comp. Graph. Statist.*, **24** (2015), 954–974.

[10] P. Del Moral, A. Doucet and A. Jasra, Sequential Monte Carlo samplers, *J. Roy. Statist. Soc. Ser. B*, **68** (2006), 411–436.

[11] A. Dobra, C. Hans, B. Jones, J. R. Nevins, G. Yao and M. West, Sparse graphical models for exploring gene expression data, *J. Mult. Anal.*, **90** (2004), 196–212.

[12] M. Drton and M. Eichler, Maximum Likelihood Estimation in Gaussian Chain Graph Models under the Alternative Markov Property, *Scand. J. Statist.*, **33** (2006), 247–257.

[13] M. Drton and M. D. Perlman, A SINful approach to Gaussian graphical model selection, *Journal of Statistical Planning and Inference*, **138** (2008), 1179–1200.

[14] M. J. Druzdel and C. Glymour, Causal inferences from databases: Why universities lose students, in *Computation, Causation, and Discovery* (eds C. Glymour and G. F. Cooper), AAAI Press, Menlo Park, CA., (1999), 521–539.

[15] A. Jasra, D. A. Stephens, A. Doucet and T. Tsagaris, Inference for Lévy driven stochastic volatility models via adaptive sequential Monte Carlo, *Scand. J. Statist.*, **38** (2011), 1–22.

[16] G. Kanayama, H. G. Pope and J. I. Hudson, Associations of anabolic-androgenic steroid use with other behavioral disorders: an analysis using directed acyclic graphs, *Psychol Med*, **48** (2018), 2601–2608.

[17] S. L. Lauritzen and T. S. Richardson, Chain graph models and their causal interpretations, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **64** (2002), 321–348.

[18] S. L. Lauritzen and D. J. Spiegelhalter, Local computations with probabilities on graphical structures and their applications to expert systems (with discussion), *J. R. Statist. Soc. B*, **50** (1988), 157–224.

[19] S. L. Lauritzen and N. Wermuth, *Mixed Interaction Models*, Institut for Elektroniske Systemer, Aalborg Universitetscenter, 1984.

[20] S. L. Lauritzen and N. Wermuth Graphical models for association between variables, some of which are qualitative and some quantitative, *Ann. Statist*, **17** (1989), 31–57.

[21] A. Lenkoski and A. Dobra, Computational aspects related to inference in Gaussian graphical models with the G-Wishart prior, *Journal of Computational and Graphical Statistics*, **20** (2011), 140–157.

[22] M. Levitz, M. D. Perlman and D. Madigan, Separation and completeness properties for AMP chain graph Markov models, *Annals of statistics*, **29** (2001), 1751–1784.

[23] C. McCarter and S. Kim, On sparse Gaussian chain graph models, *Advances in Neural Information Processing Systems (NIPS)*, **2** (2014), 3212–3220.

[24] J. Pearl A constraint propagation approach to probabilistic reasoning, in *Uncertainty in Artificial Intelligence* (eds. L. M. Kanal and J. Lemmer), North-Holland, Amsterdam, (1986), 357–370.

[25] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, The Morgan Kaufmann Series in Representation and Reasoning. Morgan Kaufmann, San Mateo, CA, 1988.

[26] J. M. Pena, Learning marginal AMP chain graphs under faithfulness, in *European Workshop on Probabilistic Graphical Models* (eds. Linda C. van der Gaag and Ad J. Feelders), Springer, (2014), 382–395.

[27] N. Richardson-Harman, C. W. Hendrix, N. N. Bumpus, C. Mauck, R. D. Cranston, K. Yang, J. Elliott, K. Tanner and I. McGowan, Correlation between compartmental tenofovir concentrations and an ex vivo rectal biopsy model of tissue infectibility in the RMP-02/MTN-006 phase 1 study, *PLoS One*, **9** (2014), e111507.

[28] R. Silva, A MCMC approach for learning the structure of gaussian acyclic directed mixed graphs, in *Statistical Models for Data Analysis* (eds. P. Giudici, S. Ingrassia and M. Vichi), Springer: New York, (2013), 343–351.

[29] R. Silva and Z. Ghahramani, The Hidden Life of Latent Variables: Bayesian learning with mixed graph models, *J. Mach. Learn. Res.*, **10** (2009), 1187–1238.

[30] D. Sonntag and J. M. Pena, On expressiveness of the chain graph interpretations, *International Journal of Approximate Reasoning*, **68** (2016), 91–107.

[31] L. Tan, A. Jasra, M., De Iorio and T. Ebbels, Bayesian Inference for multiple Gaussian graphical models, *Ann. Appl. Stat.*, **11** (2017), 2222–2251.

[32] H. Wang, Scaling It Up: Stochastic search structure learning in graphical models, *Bayes. Anal*, **10** (2015), 351–377.

[33] H. Wang, C. Reesony and C. M. Carvalho, Dynamic financial index models: Modeling conditional dependencies via graphs, *Bayesian Anal.*, **6** (2011), 639–663.

[34] N. Wermuth, Linear recursive equations, covariance selection and path analysis, *J. Am. Statist. Assoc*, **75** (1980), 963–972.

[35] N. Wermuth, and S. L. Lauritzen, On substantive research hypotheses, conditional independence graphs and graphical chain models (with discussion), *J. Roy. Statist. Soc. Ser. B*, **52** (1990), 21–72.

[36] K. H. Yang, H. Hendrix, N. Bumpus and J. Elliott, *et. al*, A multi-compartment single and multiple dose pharmacokinetic comparison of rectally applied tenofovir 1% gel and oral tenofovir disoproxil fumarate, *PLOS One*, **9** (2014), e106196.
[37] Y. Zhou, A. M. Johansen and J. A. Aston, Towards Automatic Model Comparison: An Adaptive Sequence Monte Carlo Approach, *J. Comp. Graph. Statist.*, **25** (2016), 701–726.

*E-mail address*: denglu@u.nus.edu
*E-mail address*: maria@yale-nus.edu.sg
*E-mail address*: ajay.jasra@kaust.edu.sa
*E-mail address*: grosner1@jhmi.edu