A Composite Likelihood Approach to Gaussian Network Differentiation with Application to Epigenetics

James Adefisoye, PhD[1], S Hasan Arshad, PhD[2,3] and Hongmei Zhang, PhD[1]

[1]Division of Epidemiology, Biostatistics, and Environmental Health Sciences, School of Public Health, University of Memphis, Memphis, TN, USA
[2]Allergy and Clinical Immunology Department, University of Southampton, Southampton, U.K.
[3]The David Hide Asthma and Allergy Research Centre, Isle of Wight, U.K.

ARTICLE HISTORY
Compiled October 10, 2025

ABSTRACT
For networks originated from dependent populations, methods to test network differentiation between the two populations are generally designed incorporating the nature of dependence. Doing so potentially complicates the inferencing process with heavy computing burden. Through simulations, we assess the value of using composite likelihood to carry out network comparisons under different statuses of population dependency. We apply the method to real-life epigenetic data and assess epigenetic network stability over time.

KEYWORDS
Undirected network, dependent network comparisons, network differentiation, MCMC, Bayesian methods, DNA methylation, composite likelihood.

1. Introduction

Epigenetics encapsulates environmental influences and life changes and regulates gene functions. One commonly studied epigenetic modification mechanism is DNA methylation. DNA methylation (DNA-m) is the process of chemical modification of the DNA base, usually the addition of a methyl (CH3) group at 5'-carbon of pyrimidine ring of cytosine nucleotide to the DNA to form 5-methylcytosine. Although most studies focus on features of individual CpG sites, joint activities among CpGs have been recognized, and the value of such joint activities on phenotypic characteristics has been suggested [1]. In this article, we focus on such activities, from the perspective of networks formed by CpG methylation sites.

Networks display the inter-connectivity of a set of entities, e.g., CpG sites, and are useful in understanding the relationship between these entities referred to as nodes. The connectivity of nodes (e.g., CpGs) in a network is indicated by the presence of an edge or lack thereof. Edges transmit details about the links between the nodes. Networks are generally classified as directed or undirected or a combination of the two. A directed network is one in which the edges indicating the connections between two nodes carry signals that drive the activities from one (parent) to the other (child). On the other

CONTACT Hongmei Zhang. Email: hzhang6@memphis.edu

hand, an undirected network is one that simply depicts an association or conditional dependence between two nodes without an implied flow or direction of relationship. Our focus in this work is on undirected Gaussian networks such that the connections between nodes of a network are determined by a covariance matrix or a precision matrix inferred based on observed variables (nodes) following a multivariate normal distribution.

Undirected Gaussian network structures between different conditions such as disease status (independent networks) or at different time points (dependent networks) can be differentiated. Methods to compare between independent networks have been proposed in multiple studies. Gill et al. [2] suggested a procedure based on genetic associations or interaction between genes to globally test differential undirected gene networks. Another approach to globally test network differentiation was proposed by Xia et al. [3], designed based on estimating difference in precision matrices between two differential undirected networks [4]. The work by Städler et al. [5] was under a framework similar to Zhao et al. [4]. A method to compare two graphs by comparing multivariate two-sample means on known graphs using Hotelling's T^2-tests was proposed by Jacob et al. [6]. More recently, He et al. [7], built upon inference of Gaussian graphic modeling and asymptotic normality of precision matrix components, proposed a test statistic to efficiently compare two precision matrices. Approaches comparing between covariance matrices can also be applied to compare agreement between two Gaussian undirected networks constructed using precision matrices, e.g., the works by Cai et al. [8] and Chang et al. [9]. To our knowledge, methods to compare dependent networks, on the other hand, are relatively limited except for the manifest-data likelihood (MDL) method proposed by Zhang et al. [10]. MDL has the ability to infer the underlying shared network between different time points as well as edges showing differentiation. However, the scalability of the approach when the number of nodes is large, limited its scope of application. This is potentially due to the complexity of the modeling to address dependence while comparing networks. To this end, we propose a simplified approach aiming to enhance computational efficiency. In particular, we adopt the concept of composite likelihood for this purpose.

Composite likelihood (CL) is an inference function obtained by summing individual component log-likelihood objects, regardless of their dependence status. The composite likelihood function is unbiased because each component in the function is a conditional density [11]. Irrespective of the nature of dependence among the individual components, each individual term in the summation is a valid log-likelihood, although a composite likelihood may be for a misspecified model. Its high efficiency, particularly, computation-wise and in modeling the joint distribution in high-dimensional response, as well as its robustness to model misspecification makes it appealing [11]. Composite likelihood-based methods have been proposed with varying applications, e.g., to infer ancestry probability in genetic studies [12], or to estimate genetic and environmental covariance in genome-wide association studies [13]. For more on composite likelihood-based methods, readers are referred to [11,14–18]. In this article, we simplify Zhang et al.'s MDL method by use of composite likelihood and examine the potential of this simplification in the comparison of undirected Gaussian networks, dependent or independent.

The remainder of this paper is outlined as follows. Section 2 introduces the proposed method. In section 3, simulations to demonstrate and evaluate the composite likelihood-based approach are discussed, and we also compare our approach with the MDL via the simulated data. In section 4, we apply the method to real-life epigenetic data and assess epigenetic network stability over time. Finally, in section 5, we discuss and summarized our work.

2. Methods

Let $\boldsymbol{X}_{1,n_1\times P}$ and $\boldsymbol{X}_{2,n_2\times P}$ represent two multivariate data sets with sample sizes n_1 and n_2, respectively, measured on P variables in two populations such that the data are complete and have no missing values. These two populations could be dependent or independent. Without loss of generality, we assume the data are centered. Let vector $\boldsymbol{X}_{1i}$ of length P denote data for subject i, $i = 1, \ldots, n_1$, from population 1 and assume $\boldsymbol{X}_{1i} \sim N(0, \Sigma_1)$, where Σ_1 is the covariance matrix of $\boldsymbol{X}_{1i}$. Similarly, $\boldsymbol{X}_{2i} \sim N(0, \Sigma_2), i = 1, \ldots, n_2$. We define $\Omega_k = \Sigma_k^{-1}$, a precision matrix under population k, $k = 1, 2$. Denote $\boldsymbol{X}_{1\cdot} = \{\boldsymbol{X}_{11}, \ldots, \boldsymbol{X}_{1n_1}\}$ and $\boldsymbol{X}_{2\cdot} = \{\boldsymbol{X}_{21}, \ldots, \boldsymbol{X}_{2n_2}\}$. When $\boldsymbol{X}_{1i}$ and $\boldsymbol{X}_{2i}$ are independent, the joint likelihood of Ω_k, is defined as,

$$\begin{aligned}
\mathcal{L}(\Omega_1, \Omega_2) &= \prod_{k=1}^{2} p(\boldsymbol{X}_{k\cdot}|\Omega_k) \qquad (1)\\
&= \prod_{k=1}^{2}\prod_{i=1}^{n_k} p(\boldsymbol{X}_{ki}|\Omega_k)\\
&= \prod_{k=1}^{2}(2\pi)^{-P(n_k/2)}\det(\Omega_k)^{n_k/2}\\
&\quad \exp\left(-\sum_{k=1}^{2}\left(\frac{n_k}{2}\right)\mathrm{tr}(\hat{\Sigma}_{X(k,k)}\Omega_k)\right),
\end{aligned}$$

where $\hat{\Sigma}_{X(1,1)}$ and $\hat{\Sigma}_{X(2,2)}$ are the sample covariance matrices for populations 1 and 2, respectively. Note that (1) becomes a composite or pseudo likelihood when the two populations are dependent, in which case the likelihood focuses on each population without accounting for the dependence between them. In the remainder of this article, we simply call (1) a composite likelihood.

The graph structures for the two populations can be concluded using Ω_1 and Ω_2 via binary adjacency matrices G_1 and G_2, respectively. An entry of 1 in $G_k, k = 1, 2$, denotes a connected edge and its corresponding entry of Ω_k is non-zero. An entry of 0 in G_k indicates a disconnected edge in a graph, and Ω_k is zero at that entry. It is assumed that a self-loop does not exist in any of the graphs, i.e., the diagonals of the adjacency matrices are all 0's.

2.1. Graphs differentiation between two populations

Network structures under different conditions can be identical or differential. To facilitate a comparison between networks for different populations, we introduce an indicator variable η with $\eta = 1$ denoting two graphs being identical and $\eta = 0$ two graphs being differential. When the underlying two graphs are identical, $\Omega_1 = \Omega_2 = \Omega_c$, and consequently, $G_1 = G_2 = G_c$. To incorporate both situations, we define $\Omega^{(k)} = (1-\eta)\Omega_k + \eta\Omega_c$, and $G^{(k)} = (1-\eta)G_k + \eta G_c$, with both $\Omega^{(k)}$ and $G^{(k)}$ dependent on η. Consequently,

likelihood (1) is revised to

$$\prod_{k=1}^{2}\prod_{i=1}^{n_k} p(\boldsymbol{X}_{ki}|\Omega_k,\eta) = \prod_{k=1}^{2}(2\pi)^{-P(n_k/2)}\det\left(\Omega^{(k)}\right)^{n_k/2} \exp\left(-\sum_{k=1}^{2}\left(\frac{n_k}{2}\right)\mathrm{tr}(\hat{\Sigma}_{X(k,k)}\Omega^{(k)})\right),$$

and the focus is on inferring η, and corresponding graph parameters $\Omega^{(k)}$ and $G^{(k)}$.

2.2. Inference on graph differentiation status between two conditions

Let $\omega^{(k)}_{(m_1,m_2)}$ denote a unique entry of $\Omega^{(k)}$ at (m_1,m_2) such that $m_1 \leq m_2 = 1,\ldots,P$. Following the definition of $\Omega^{(k)}$, we have $\omega^{(k)}_{(m_1,m_2)} = (1-\eta)\omega_{k,(m_1,m_2)} + \eta\omega_{c,(m_1,m_2)}$. Similarly, let $g^{(k)}_{(m_1,m_2)}$ denote an entry of $G^{(k)}$ at (m_1,m_2), and we have $g^{(k)}_{(m_1,m_2)} = (1-\eta)g_{k,(m_1,m_2)} + \eta g_{c,(m_1,m_2)}$, $m_1 \leq m_2 = 1,\ldots,P$. At (m_1,m_2), ${g^{(k)}}_{(m_1,m_2)} = 1$ if nodes m_1 and m_2 are connected, i.e., the corresponding entry of $\Omega^{(k)}$ is non-zero, and ${g^{(k)}}_{(m_1,m_2)} = 0$ otherwise. Let Θ denote a collection of graphical parameters along with the network differentiation status parameter η, $\Theta = \{\boldsymbol{\omega}^{(k)}, \boldsymbol{g}^{(k)}, \eta, k = 1,2\}$, where $\boldsymbol{\omega}^{(k)}$ and $\boldsymbol{g}^{(k)}$ denote vectors composed of $\omega^{(k)}_{(m_1,m_2)}$ and $g^{(k)}_{(m_1,m_2)}$, respectively. We propose a fully Bayesian approach to draw inferences on these parameters.

Prior distributions. For network constructions, we utilize established prior distributions as in Zhang et al. [10]. We assign the spike-and-slab prior distribution [19–21] to each $\omega^{(k)}_{(m_1,m_2)}$ with $m_1 \neq m_2$,

$$p(\omega^{(k)}_{(m_1,m_2)}|g^{(k)}_{(m_1,m_2)}) = g^{(k)}_{(m_1,m_2)}N(0,\nu^2_{1,k}) + (1-g^{(k)}_{(m_1,m_2)})N(0,\nu^2_{0,k}),$$

conditional on $\Omega^{(k)} \in \mathcal{M}^+$ with $\mathcal{M}^+$ denoting a space of positive definite matrices. Hyper-parameters $\nu^2_{1,k}$ and $\nu^2_{0,k}$ are the variances in the two respective normal distributions. The spike component, represented by the Gaussian distribution $N(0,\nu^2_{0,k})$, is crucial for controlling sparsity, and the hyperparameter $\nu^2_{0,k}$, reflect the expected sparsity of a graph. Here, we use the same hyperparameters for the two populations, i.e. $\nu^2_{1,k} = \nu^2_1$ and $\nu^2_{0,k} = \nu^2_0$, assuming comparable sparsity in networks between the two populations. With a large ν^2_1, the corresponding normal prior distribution is a vague prior distribution, allowing a wide range of edge strength for each population. With ν^2_1 fixed and ν^2_0 selected based on sparsity level, we are able to control the overall sparsity level in the two populations while allowing the two populations to have different number of edges.

For parameters $\omega^{(k)}_{(m1,m2)}$ with $m_1 = m_2 = m$, its prior, $p(\omega^{(k)}_{(m,m)})$, is assumed to be an exponential distribution with parameter λ. We set $\lambda = 1$ as suggested by Wang et al. [22].

Bernoulli probability mass function with prior parameter π denoting our a priori belief of two nodes being connected was used for the components in $\boldsymbol{g}^{(k)}$, $p(g^{(k)}_{(m_1,m_2)}) = \pi^{g^{(k)}_{(m_1,m_2)}}(1-\pi)^{1-g^{(k)}_{(m_1,m_2)}}, m_1 \neq m_2$. We set $\pi = \mathcal{O}(2/(P-1))$ based on a common

assumption on the expected number of edges, $\mathcal{O}(P)$, for sparse graphs [22,23]. Finally, we assume a Bernoulli distribution with hyperparameter 0.5 for the prior mass function of η. We proposed this non-informative prior, as in most situations, we do not have prior knowledge regarding network similarity. If we believe in any particular status over the other, then an informative prior is preferred to reflect our prior belief and benefit posterior inference, especially when the sample size is small and signal in the data is weak.

The joint posterior distribution and its computation. Based on (1) and the defined prior distributions, we define the joint posterior distribution of Θ as

$$p(\Theta|\boldsymbol{X}) \propto \prod_{k=1}^{2} p(\boldsymbol{X}_{k.}|\Omega^{(k)}, \eta)p(\Omega^{(k)}|\eta, \boldsymbol{g}^{(k)})p(\boldsymbol{g}^{(k)}|\eta)p(\eta). \tag{2}$$

We draw posterior inferences of Θ through Markov Chain Monte Carlo (MCMC) simulations. In particular, we implement the Gibbs sampler to draw samples from the full conditional posterior distribution of each parameter derived from (2). In the following, we present these full conditional posterior distributions. We use $(\cdot)$ to represent a collection of other parameters upon which a parameter is conditioned.

We begin with the conditional posterior probability of η, the key parameter for network differentiation and a fundamental parameter for subsequent graph structure inference. From (2), we have

$$p(\eta = 1|(\cdot), \boldsymbol{X}) \propto \prod_{k=1}^{2} p(\boldsymbol{X}_{k.}|\Omega_c, \eta = 1, \boldsymbol{g}_c)p(\eta = 1). \tag{3}$$

Following a comparable argument as in Zhang et al. [10], under the composite likelihood framework, it can be shown straightforwardly that

$$\begin{aligned} p(\eta = 1|(\cdot), \boldsymbol{X}) &\approx [1 + \exp\{log(b) - log(a) + \lambda(n_1, n_2)\}]^{-1} \equiv p^{\lambda}(\eta = 1|(\cdot)), \\ \lambda(n_1, n_2) &= \frac{1}{2}(|E|\log(n) - |E_1|\log(n_1) - |E_2|\log(n_2)), \end{aligned} \tag{4}$$

where $n = n_1 + n_2$, $|E|$ represents the number of edges in an inferred network under the condition that $\eta = 1$ and $|E_k|, k = 1, 2$, is the number of edges in an inferred network under the condition of $\eta = 0$. We define a and b as,

$$
\begin{aligned}
a &= \prod_{k=1}^{2} p(\boldsymbol{X}_{k.}|\Omega_c, \eta = 1) \\
&= \prod_{k=1}^{2} (2\pi)^{-P(n_k/2)} \det(\Omega_c)^{n_k/2} \exp\left(-\sum_{k=1}^{2} \left(\frac{n_k}{2}\right) \mathrm{tr}(\hat{\Sigma}_{X(k,k)}\Omega_c) \right), \\
b &= \prod_{k=1}^{2} p(\boldsymbol{X}_{k.}|\Omega_k, \eta = 0) \\
&= \prod_{k=1}^{2} (2\pi)^{-P(n_k/2)} \det(\Omega_k)^{n_k/2} \exp\left(-\sum_{k=1}^{2} \left(\frac{n_k}{2}\right) \mathrm{tr}(\hat{\Sigma}_{X(k,k)}\Omega_k) \right).
\end{aligned} \tag{5}
$$

Based on (2), the full conditional posterior distributions of parameters in $\Omega^{(k)}$ can be derived,

$$
\begin{aligned}
p(\omega^{(k)}_{(m_1,m_2)}|(\cdot), \boldsymbol{X}) &\propto \prod_{k=1}^{2} p(\boldsymbol{X}_{k.}|\Omega^{(k)}, \eta) p(\omega^{(k)}_{(m_1,m_2)}|\eta, g^{(k)}_{(m_1,m_2)}), \\
p(\omega^{(k)}_{(m,m)}|(\cdot), \boldsymbol{X}) &\propto \prod_{k=1}^{2} p(\boldsymbol{X}_{k.}|\Omega^{(k)}, \eta) p(\omega^{(k)}_{(m,m)}|\lambda),
\end{aligned}
$$

where $m = 1, \ldots, P$, $m_1 < m_2$, $m_1, m_2 = 1, \ldots, P$.

Similarly, the conditional posterior distributions of the parameters in graph structures $G^{(k)}$ can be derived,

$$
p(g^{(k)}_{(m_1,m_2)}|(\cdot), \boldsymbol{X}) \propto \prod_{k=1}^{2} p(\boldsymbol{X}_{k.}|\Omega^{(k)}, \eta) p(\omega^{(k)}_{(m_1,m_2)}|\eta, g^{(k)}_{(m_1,m_2)}) p(g^{(k)}_{(m_1,m_2)}).
$$

Hyper-parameters ν_0 and ν_1, $\nu_0 < \nu_1$, define the sparsity of a graph a priori. At a fixed value of ν_1, the inclusion or exclusion of an edge can be controlled with varying values of ν_0 where larger values of ν_0 leads to the inclusion of a smaller number of edges, and on the other hand, smaller values of ν_0 leads to the inclusion of a larger number of edges. The selection of ν_0 is crucial to the estimate of underlying graphs. In this article, we follow the suggestion of Zhang et al. [10] and choose the value of ν_0 based on a pre-specified level of sparsity. It was suggested that the starting value of ν_0 needs to be small. Here, we set $\nu_0 = 0.02$.

3. Simulations

Through simulations, our goal is to evaluate the robustness of the composite-likelihood-based approach in network comparisons with respect to different statuses of dependence between two populations.

3.1. Network structures and settings

To assess the method, we consider two situations of relationships between two populations: independence between two populations (S1), and dependence between two populations (S2). For each situation, the networks for the two populations are either truly identical or truly differential.

S1. Chain graphs, G_1 and G_2, are generated for each population independently, following the idea of Fan et al. [24]. We create a covariance matrix with a covariance structure of auto-regressive process of order one for each network. In a covariance matrix, the diagonal entries are 1's and an off-diagonal entry $\{m_1, m_2 = 1, \ldots, P, m_1 \neq m_2\}$ is defined as $\exp\{-a|s_{m_1} - s_{m_2}|\}$ with $s_{m_1} - s_{m_1-1} \overset{i.i.d.}{\sim}$ Unif(0.5,1), $m_1 = 2, \ldots, P$, and $s_1 < s_2 < \cdots < s_P$, where a is a parameter that determines the magnitude of link between two nodes and was set at $a = 0.1$. Based on this method, the partial correlations resulting from the precision matrices vary between -0.4 and 0.9. For networks that are truly identical, the same precision matrix is applied to generate data for each population using multivariate normal distribution with mean zero.

To generate differential networks, a chain graph for population 1 is defined as above with corresponding precision matrix Ω_1. For the second population, the precision matrix Ω_2 is formulated by forcing $\frac{P}{2} - 1$ consecutive entries of Ω_1 to be zero, we then generate data from each precision matrix for each population using multivariate normal with mean zero.

S2. When generating two dependent networks, for identical networks, we first use the method in S1 to generate two identical but independent chain networks, G_1 and G_2. To introduce dependence, we also generate a network, G_0, shared between the two populations. The shared network is a nearest-neighbor network and generated according to Li and Gui [25]. The generation of the shared network uses pairwise distances between nodes, with each node generated from a Uniform distribution on a $[0, 1] \times [0, 1]$ space. We identified t nearest neighbors using the pairwise distances, where we set $t = 2$ in our simulations. Then, entries in the precision matrix for each pair of connected nodes are generated based on the algorithm by Li and Gui [25]. The diagonal entries are 1 in the resulting precision matrix while the off-diagonal entries vary between -0.2 and 0.2 denoting low partial correlation between the nodes. The data from the shared network were then added to the data corresponding to G_1 and G_2 to introduce dependence between the two populations.

For differential networks, the network for one population is a chain network (G_1) plus the shared network (G_0), and for the other population, it is only the shared network. The shared graph is a nearest neighbor network as described above under this dependent situation (S2).

For each of the above four settings, the number of nodes is set at $P = 10$ and $P = 30$. To demonstrate finite sample properties, different sample sizes, $n_1 = n_2 = n = 50, 100, 200$ and 500 are considered. For each case, we generate 100 Monte Carlo (MC) replicates to address sampling error.

To evaluate the convergence of posterior sampling, we ran four chains with diffused starting values on multiple randomly selected MC replicates. After noticing fast convergence (Figure B.1 in the Appendix as an illustration), for each of the 100 MC replicates, we run 1,000 iterations, of which 700 iterations were used as burn-in, and the last 300 used to draw the inferences.

Model performances are assessed using four statistics, 1) empirical power (EP) of

correct detection with respect to network comparison (identical or differential), 2) median proportion of true positives for edge connection (TP; sensitivity) in a network, 3) median proportion of true negatives (TN; specificity) of a network, and 4) median proportion of correct connections (CC) of edges. Correct connections combines information from both sensitivity and specificity, and identifies edges with observed edge connection statuses (connected or not connected) that is consistent with the true underlying network structure. We also provide 95% empirical credible interval for each of the model comparison statistics except for power.

3.2. Performance of the method

3.2.1. Findings when two populations are independent

We first discuss findings when the underlying two populations are independent. With 10 nodes (Table 1), when two networks are either truly identical or truly differential, the powers to detect the underlying truth are high, at 100% for all sample sizes. Regarding the quality of the inferred graphs, the proportions of true positives, true negatives, and consequently, the overall correct connections are all high, regardless of the sample size or the underlying network differentiation status. In Table 1 and throughout the findings in this section, there is a pattern of slightly decreased proportions of true negatives as sample sizes increase, e.g., 1.0 at $n = 50$ vs. 0.875 at $n = 500$ for identical networks, which means more edges are falsely identified. This phenomenon is due to finite sample properties analogous to what is commonly observed in hypothesis testing; as sample size becomes large, there is a higher chance to reject the null hypothesis, even if the null is true.

When increasing the number of nodes to $P = 30$, similar patterns as described above are observed (Table 2). That is, the powers to detect the underlying truth are 100% for all sample sizes, irrespective of the underlying truth. The proportions of TP, TN, and CC are also high and comparable to those in Table 1.

Overall, when the two populations are independent, for all the settings considered, the patterns observed on the assessment of network differentiation and network inferences are as expected, since population independence fits the underlying assumption in the construction of the composite likelihood. The next subsection discusses results when underlying two populations are truly dependent.

3.2.2. Findings when two populations are dependent

In this subsection, we use figures to describe the patterns when comparing findings across different sample sizes and different number of nodes. Corresponding tables are in the Appendix (Table A.1 and Table A.2).

When two networks are from two dependent populations, the composite likelihood-based method is promising with respect to the power of detecting underlying network differentiation status (Figures 1A and 2A), irrespective of the number of nodes. In particular, when two networks are truly identical, the powers are high at or close to 100% for all sample sizes and different numbers of nodes, $P = 10$ and $P = 30$ (Figure 1A). When the underlying two networks are differential, the method shows high power, at 100% for all sample sizes when $P = 30$. The power statistics are also 100% for larger sample sizes, when $P = 10$, although the power is average for smaller sample sizes (46% and 48% for $n = 50$ and $n = 100$, respectively) (Figure 2 A).

Another robust finding is the high proportions of true negatives even when the sam-

Table 1. Simulation results for independent networks with 10 nodes.

n	EP (%)	TP (95% EI)	TN (95% EI)	CC (95% EI)
Underlying truth: identical networks (P=10 nodes, $\|E_{X_1}\| = \|E_{X_2}\| = 9$ edges)				
50	100	1.0 (0.889, 1.0)	1.0 (0.958, 1.0)	0.989 (0.950, 1.0)
100	100	1.0 (0.915, 1.0)	0.986 (0.937, 1.0)	0.989 (0.944, 1.0)
200	100	1.0 (0.915, 1.0)	0.958 (0.882, 1.0)	0.967 (0.844, 1.0)
500	100	1.0 (0.915, 1.0)	0.875 (0.791, 0.944)	0.900 (0.822, 0.956)
Underlying truth: differential networks (P=10 nodes, $\|E_{X_1}\| = 9$, $\|E_{X_2}\| = 7$ edges)				
50	100	X_1: 1.0 (0.889, 1.0)	1.0 (0.958, 1.0)	1.0 (0.944, 1.0)
		X_2: 1.0 (0.60, 1.0)	1.0 (0.963, 1.0)	0.989 (0.933, 1.0)
100	100	X_1: 1.0 (0.889, 1.0)	1.0 (0.944, 1.0)	0.989 (0.956, 1.0)
		X_2: 1.0 (0.9, 1.0)	1.0 (0.968, 1.0)	0.989 (0.972, 1.0)
200	100	X_1: 1.0 (0.889, 1.0)	0.972 (0.909, 1.0)	0.967 (0.928, 1.0)
		X_2: 1.0 (0.9, 1.0)	0.988 (0.950, 1.0)	0.989 (0.950, 1.0)
500	100	X_1: 1.0 (0.889, 1.0)	0.861 (0.792, 0.944)	0.889 (0.828, 0.944)
		X_2: 1.0 (0.9, 1.0)	0.975 (0.913, 1.0)	0.978 (0.922, 1.0)

Table 2. Simulation results for independent networks with 30 nodes.

n	EP (%)	TP (95% EI)	TN (95% EI)	CC (95% EI)
Underlying truth: identical networks (P = 30 nodes, $\|E_{X_1}\| = \|E_{X_2}\| = 29$ edges)				
50	100	0.983 (0.931, 1.0)	0.998 (0.991, 1.0)	0.997 (0.988, 1.0)
100	100	0.983 (0.948, 1.0)	0.994 (0.991, 1.0)	0.995 (0.991, 0.999)
200	100	0.983 (0.948, 1.0)	0.992 (0.983, 0.998)	0.991 (0.983, 0.997)
500	100	0.983 (0.948, 1.0)	0.968 (0.956, 0.980)	0.969 (0.957, 0.980)
Underlying truth: differential networks (P = 30 nodes, $\|E_{X_1}\| = 29$, $\|E_{X_2}\| = 27$ edges)				
50	100	X_1: 0.983 (0.931, 1.0)	0.998 (0.993, 1.0)	0.997 (0.99, 1.0)
		X_2: 0.967 (0.867, 1.0)	0.998 (0.995, 1.0)	0.997 (0.992, 1.0)
100	100	X_1: 0.983 (0.948, 1.0)	0.998 (0.991, 1.0)	0.997 (0.99, 0.999)
		X_2: 1.0 (0.933, 1.0)	0.998 (0.993, 1.0)	0.998 (0.993, 1.0)
200	100	X_1: 0.983 (0.939, 1.0)	0.993 (0.985, 0.999)	0.992 (0.984, 0.998)
		X_2: 1.0 (0.933, 1.0)	0.996 (0.991, 1.0)	0.996 (0.99, 0.999)
500	100	X_1: 0.983 (0.939, 1.0)	0.969 (0.956, 0.981)	0.971 (0.958, 0.982)
		X_2: 1.0 (0.933, 1.0)	0.989 (0.981, 0.996)	0.989 (0.98, 0.996)

ple size is small (Figures 1C and 2C). Together with the relatively lower proportions of true positives (Figures 1B and 2B), the findings indicate the conservativeness of the method when claiming an edge. For the proportions of true positives, although they are lower than those observed when underlying populations are independent, they increase as the sample sizes increase. The high proportions of true negatives and relatively reasonable proportions of true positives result in overall acceptable proportions of correct connections (Figures 1D and 2D).

Here, we would like to delve a little deeper into the relatively low proportions of true positives when two populations are dependent. We use the setting when two graphs are truly differential as an example to lay out our discussion. Recall that the networks represented by X_1 are a combination of G_1 (unique network) and G_0 (shared network) while the networks by X_2 are G_0 only. We observe that the proportions of true positives for networks corresponding to X_2 are higher than those for graphs corresponding to X_1. This phenomenon is likely due to the scenarios implemented to simulate dependent networks since the assumed underlying truth G_1+G_0 for X_1, and thus lower proportions

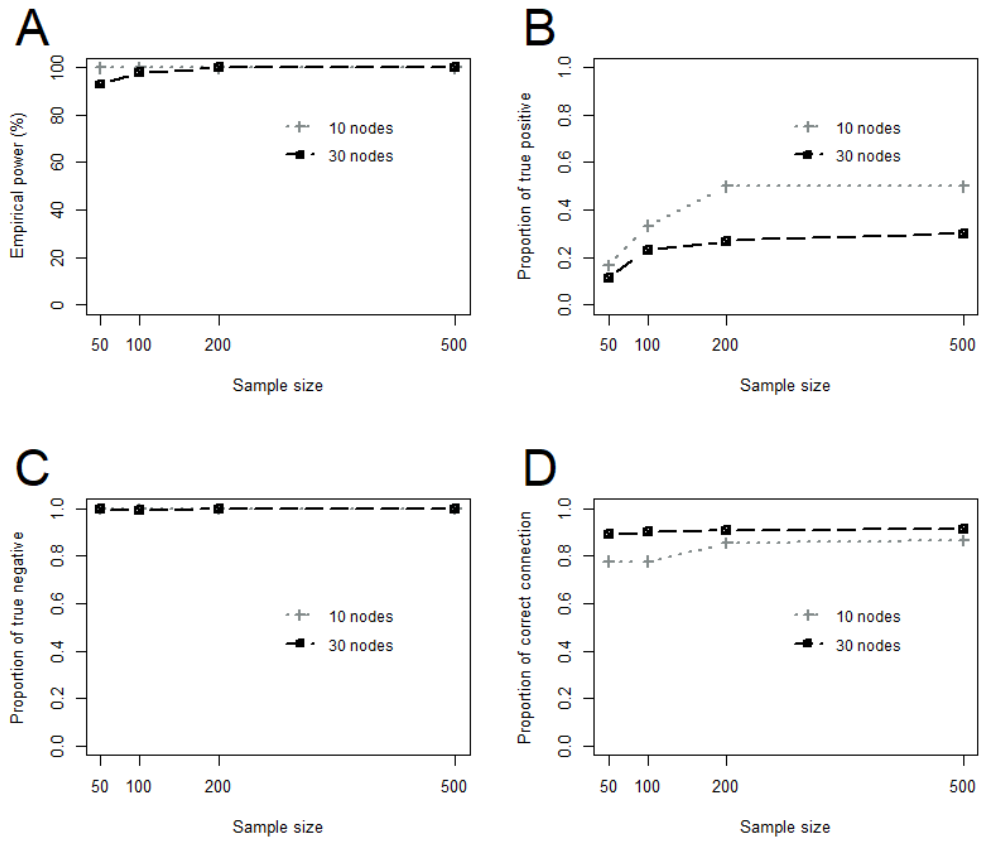

Figure 1. Model assessment statistics when two dependent networks are identical. When $P = 10$, the number of edges are G_0=9 and G_1=G_2=9. When $P = 30$, the number of edges of G_0=26 and G_1=G_2=29.

of true positives are concluded for networks corresponding to X_1. However, given the increased proportions of true positives for X_2 with G_0 as the underlying truth, we expect the proposed approach has a strong potential to correctly infer the underlying true graph determined by $\Omega_1+\Omega_0$, which is likely not a simple addition of G_1 and G_0. In addition, this observation leads to an interesting direction of future work while enjoying the simplicity and robustness of the composite likelihood framework. That is, when two populations are potentially dependent, to what extent we are able to learn the shared network.

3.2.3. Comparison of the performance of CL and MDL

Via simulations, to further assess the performance of the composite likelihood (CL)-based approach, we compare the findings with those from the MDL approach in Zhang et al. [10] The simulated data are the 100 Monte Carlo replicates based on the settings outlined in Section 3.1 at $n = 50$ and the same statistics (power, proportions of TP, TN, and CC) are used to compare the two approaches.

Results for the CL-based approach (or CL for simplicity) are in Table 1 ($P = 10$) and Table 2 ($P = 30$) with $n = 50$ and the results for MDL-based methods (or MDL for simplicity) are in Table 3 (independent populations) and Table 4 (dependent popu-

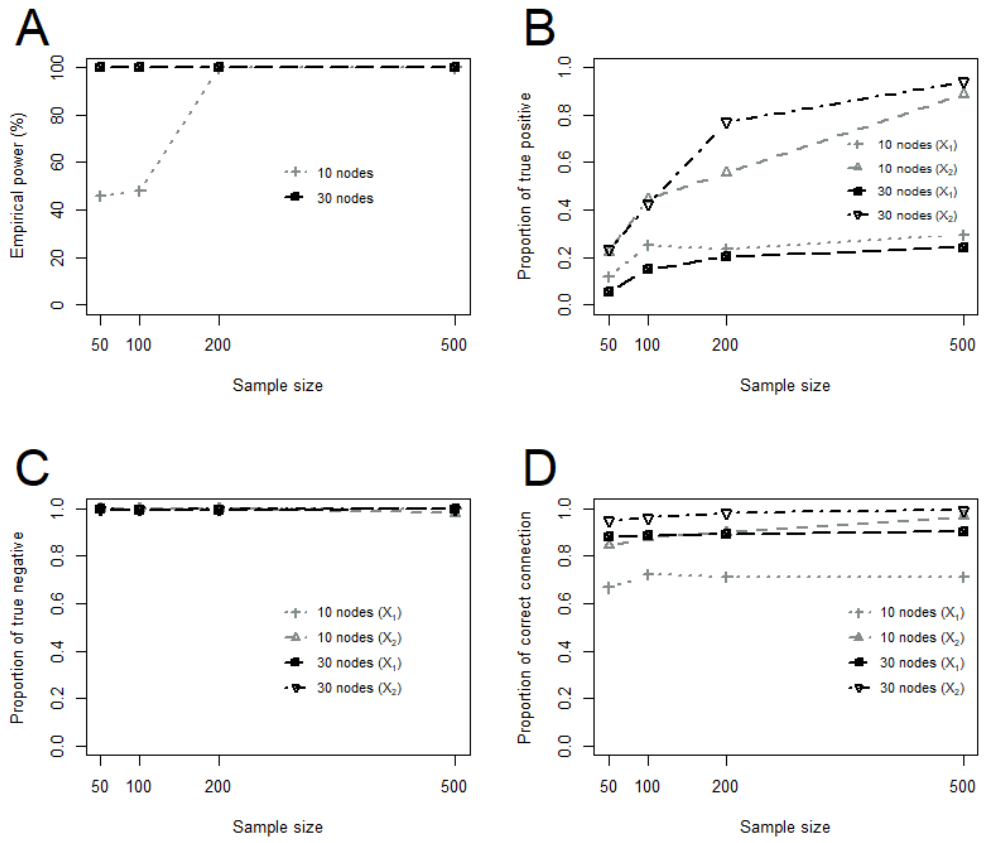

Figure 2. Model assessment statistics when two dependent networks are differential. When $P = 10$, the number of edges of $G_1 = 9$, and that of $G_0 = 9$. When $P = 30$, the number of edges of $G_1 = 29$, and that of $G_0 = 26$.

lations).

When two networks are independent, regardless of the number of nodes and network differentiation status (identical or differential), the CL-based method outperforms the MDL-based approach (Tables 1 and 2). The power to detect the underlying differentiation status under CL is always at least as high as that from the MDL-based approach (Tables 1 and 2). With respect to network construction based on CL, most statistics (the median proportions of TP, TN and CC) are higher than 0.95 and all statistics are higher than 0.85; overall, these statistics are much better than those from the MDL-based approach.

Under the situation of two dependent populations, with $P = 10$, when the two networks are truly identical, CL shows a much higher power than MDL to detect the underlying truth of network differentiation, 100% from CL (Figure 1A and Appendix Table A.1) vs 20% from MDL (the first row of Table 4)). When the two networks are truly differential, MDL outperforms CL, 100% for MDL (second row of Table 4)) and 46% for CL (2A). However, as seen in Figure 2A, the power of CL to detect the underlying truth quickly increases as sample sizes increase. When $P = 30$, regardless of the network differentiation status (identical or differential), the power of CL is much higher than that from MDL when the two networks are identical and comparable to that from MDL for differential networks.

As for graph inference under the situation of dependent populations, the proportions of TP and CC from CL are relatively lower than those from MDL, leading to relatively higher proportions of TN under CL. The underlying cause of such phenomena is likely to be the same as that noted for the situation when two populations are independent, i.e., MDL has the capacity to address shared networks. This point is especially clearer when we look at the graph inference statistics for X_2. The underlying true graph for X_2 is G_0, the shared graph. As shown in Figure 2, even at n = 50, the inferences of the graph for X_2 from CL are better than those for X_1. However, this is not the case for MDL as seen in Table 4 when two networks are differential. The much lower proportions of TP and higher proportions of TN from MDL for X_2 compared to those for X_1 are actually as expected, simply because the true graph on top of the shared graph G0 is an empty graph. On the other hand, the graph inference statistics for X_1 from MDL overall outperform those from CL at $n = 50$, which indicates a better fit of MDL for graph inferences when two populations are potentially dependent.

Table 3. Select comparative results for independent networks using the MDL method

n	EP (%)	TP (95% EI)	TN (95% EI)	CC (95% EI)
Underlying truth: identical networks ($P = 10$ nodes, $\|E_{X_1}\| = \|E_{X_2}\| = 9$ edges)				
50	96	1.0 (0.778, 1.0)	0.833 (0.012, 0.988)	0.867 (0.209, 0.991)
Underlying truth: differential networks ($P = 10$ nodes, $\|E_{X_1}\| = 9$, $\|E_{X_2}\| = 7$ edges)				
50	85	X_1: 1.0 (0.828, 1.0)	0.833 (0.056, 0.972)	0.867 (0.244, 0.978)
		X_2: 1.0 (0.8, 1.0)	0.9 (0.05, 1.0)	0.911 (0.156, 1.0)
Underlying truth: identical networks ($P = 30$ nodes, $\|E_{X_1}\| = \|E_{X_2}\| = 27$ edges)				
50	95	0.966 (0.897, 1.0)	0.443 (0.007, 0.581)	0.474 (0.073, 0.605)
Underlying truth: differential networks ($P = 30$ nodes, $\|E_{X_1}\| = 29$, $\|E_{X_2}\| = 27$ edges)				
50	99	X_1: 0.966 (0.897, 1.0)	0.438 (0.008, 0.605)	0.476 (0.074, 0.627)
		X_2: 1.0 (0.867, 1.0)	0.602 (0.012, 0.728)	0.616 (0.046, 0.737)

Table 4. Select comparative results for dependent networks using the MDL method

n	EP (%)	TP (95% EI)	TN (95% EI)	CC (95% EI)
Underlying truth: identical networks ($P = 10$ nodes, $\|E_{X_1}\| = \|E_{X_2}\| = 29$ edges)				
50	20	0.5 (0.167, 0.75)	0.879 (0.758, 0.97)	0.778 (0.644, 0.889)
Underlying truth: differential networks ($P = 10$ nodes, $\|E_{X_1}\| = 9$, $\|E_{X_2}\| = 7$ edges)				
50	100	X_1: 0.412 (0.294, 0.578)	0.857(0.0, 0.964)	0.688 (0.578, 0.8)
		X_2: 0.111 (0.0, 0.222)	0.972 (0.0, 1.0)	0.8 (0.733, 0.822)
Underlying truth: identical networks ($P = 30$ nodes, $\|E_{X_1}\| = 29$, $\|E_{X_2}\| = 27$ edges)				
50	16	0.404 (0.308, 0.962)	0.950 (0.110, 0.971)	0.880 (0.212, 0.908)
Underlying truth: differential networks ($P = 30$ nodes, $\|E_{X_1}\| = 29$, $\|E_{X_2}\| = 27$ edges)				
50	99	X_1: 0.389 (0.315, 0.919)	0.953 (0.136, 0.971)	0.880 (0.265, 0.909)
		X_2: 0.038 (0.0, 1.0)	0.972 (0.101, 0.995)	0.915 (0.179, 0.938)

3.3. Comparison of computing times between CL and MDL

To examine the efficiency of the proposed CL approach in network comparison, data simulated under scenario S2 with $n = 50, 100, 200, 500$, and $P = 10$ are used. The time spent in one iteration is recorded and compared to that of MDL. Regardless of the

sample size, the CL approach only needs less than one-third of the computing time, compared to if dependency is incorporated into the calculations as in the MDL method (Figure 3A and 3B).

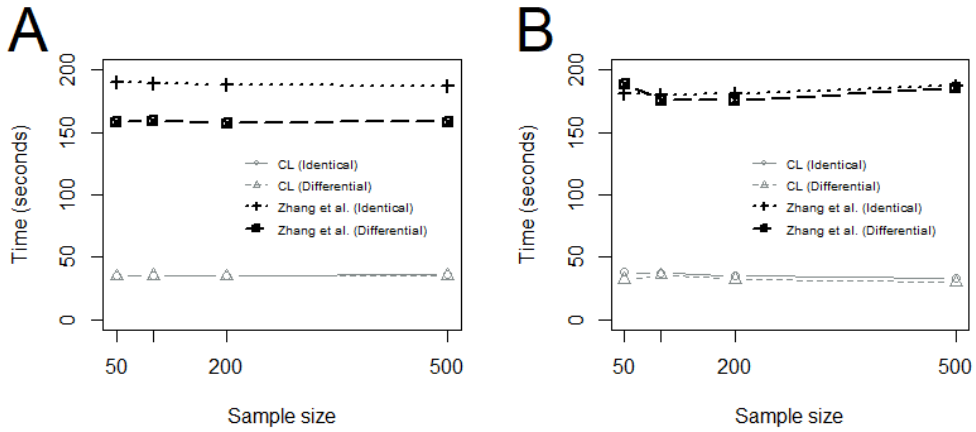

Figure 3. Computing times of CL vs. MDL. (A) shows the times for independent graphs with 10 nodes. (B) shows the times for dependent graphs with 10 nodes.

4. Application to real-life data

4.1. Background of the data

The study was motivated by the potentially joint activities among CpG sites and changes of such activities over time, as noted in the introduction section. By applying the proposed method (CL) to DNA-m at certain CpG sites, we demonstrate the performance of CL and compare the results with those from MDL [10].

DNA-m data measured in a birth cohort established in 1989-1990 on the Isle of Wight, United Kingdom, are analyzed [26]. In particular, DNA-m at 22 CpG sites at ages 10, 18, and 26 years examined in Zhang et al. [10] are included in this study and we compare network differentiation between ages 10 and 26 (n=171) using the CL-based approach. During adolescence, children experience substantial changes mentally and physically, and we expect to observe network differentiation between these two ages.

We use the methods proposed in Section 2.2 to draw inferences on the parameters of the model, including network comparisons. To tune ν_0 based on sparsity level, following our assumption, the lower bound on the number of edges is set at $\mathcal{O}(P)$=22, and the upper bound is relaxed to the number of possible edges from 22 nodes, i.e., 231 as the upper bound. To draw posterior inferences, in total, 1,000 iterations are run in the Gibbs sampler with 500 as burn-in iterations. Running longer chains does not change our conclusion, indicating potential convergence.

4.2. Results

With the CL-based approach, the estimated probability that the networks are differential between ages 10 and 26 is 0.994, strongly supporting network changes from pre- to post-adolescence. Most of the 16 edges identified at age 10 years were kept at age 26 years with new edges detected at age 26 years within (e.g., gene *GFI1*) and between

Table 5. The 22 CpG sites considered from IOW Cohort

Chr	Gene	CpG	CpG Index
1	*GFI1*	cg06338710	1
		cg09662411	2
		cg09935388	3
		cg10399789	4
		cg12876356	5
		cg14179389	6
		cg18146737	7
		cg18316974	8
5	*AHRR*	cg05575921	9
6	*HLA-DPB2*	cg11715943	10
7	*CNTNAP2*	cg25949550	11
	ENSG00000225718	cg04598670	12
	MYO1G	cg04180046	13
		cg12803068	14
		cg19089201	15
8	*EXT1*	cg03346806	16
14	*TTC7B*	cg18655025	17
15	*CYP1A1*	cg05549655	18
		cg11924019	19
		cg18092474	20
		cg22549041	21
21	*RUNX1*	cg12477880	22

Note: "Chr" denotes the chromosome location of CpG

genes (Figure 4), indicating potential impact of adolescence transition on epigenetic activities.

On the other hand, the long duration between these two ages may also contribute to the detected network differentiation. That is, with more than 15 years apart from pre-adolescence to post-adolescence, epigenetic changes following adolescence transition are likely to be more pronounced, thus making it easier for the CL-based approach to detect such changes. To evaluate this postulation, we assess network differentiation between ages 10 and 18 years ($n = 325$). The estimated probability of network differentiation is 0.102, suggesting that it is more likely that the two networks are identical rather than differential. This finding differs from that in Zhang et al. but is consistent with our postulation and the results of our simulations. The CL-based approach overall has lower power compared to MDL when two dependent networks are truly differential. At age 18 years, changes in epigenetic activities with respect to age 10 are potentially not as prominent and established as that at age 26 years, and thus the CL- but not ML-based approach has difficulty to detect such differentiation. However, when two populations are dependent, the CL is superior to the MDL approach as demonstrated by our extensive simulations.

5. Discussion and conclusion

This work proposes a composite likelihood (CL)-based approach to compare networks between two populations that could be dependent or independent. The CL-based approach is designed based on the method proposed by Zhang et al. [10], which is particularly for dependent networks. Simulations and real-data applications are applied to demonstrate and evaluate the CL-based method.

Simulations demonstrate that the method is powerful in detecting the underlying truth when two networks are under independent conditions, irrespective of the graphs'

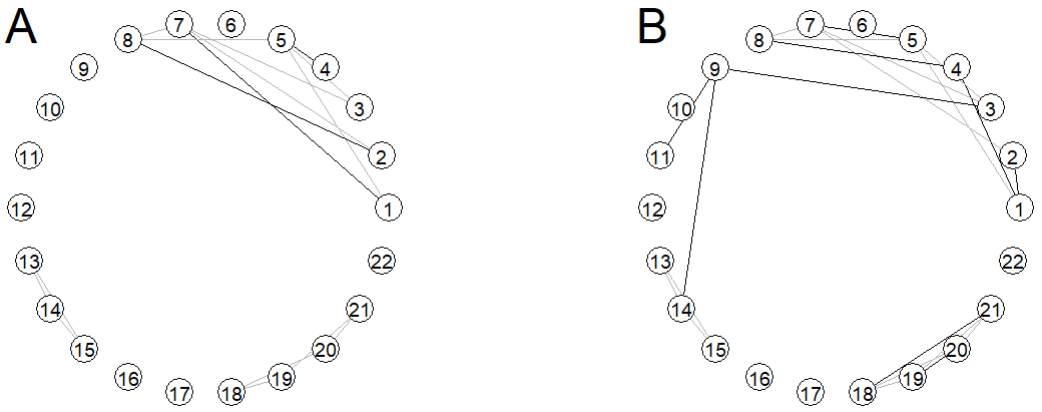

Figure 4. Networks estimated at ages 10 and 26 years, respectively. (A) Age 10 years (16 edges). (B) Age 26 years (22 edges). The thick, darker lines indicate links that are different between the two networks.

differential status. This is as expected and fits the underlying assumption of the proposed approach. When two networks are from two dependent populations, the potential of the composite likelihood approach to detect underlying network differentiation is high and comparable overall to the situation when two populations are independent, except for dependent graphs with a relatively small number of nodes and relatively small sample sizes.

When comparing the performances of the CL- and MDL-based methods, particularly when networks are from two independent populations, the CL-based method outperforms MDL in terms of both the power of detecting network differentiation and the quality of graph constructions. Under the situation of two dependent populations, when the two networks are identical, compared to the MDL-based method, the CL-based approach has a much higher power to detect the underlying truth of network differentiation even when the sample size is small. When the two networks are truly differential, sparsity level of a graph seems to affect the performance of the CL-based approach. Sparsity level is defined as the number of edges divided by the number of possible edges; the smaller the number, the higher the sparsity level is. The CL-based approach has a lower power compared to the MDL-based method when sample size is small if networks are less sparse, e.g., when the number of nodes is 10, the sparsity levels of the two networks are 9/45=0.2 and 18/45=0.4, respectively. However, when the sparsity level is relatively high (e.g., at $P = 30$, the sparsity levels are 0.06 and 0.13 for the two networks), the power of detecting underlying differentiation is extremely high and consistent with that from the MDL-based method. This is understandable, since the CL-based approach does not consider shared background as in the MDL and more noise is introduced into the comparison when network sparsity level is low (less sparse), consequently causing difficulty to differentiate between two networks if sample sizes are small. Notably, such inferiority is substantially reduced when the sample size is increased.

Our real data applications support the findings in simulations. That is, when two dependent networks are truly differential, the MDL-based approach is more powerful to detect the underlying truth compared to the CL-based method, in that CL was able to detect network changes between ages 10 and 26 but not between 10 and 18 years. These real data applications, accompanied by findings in the simulations, indicate that

the CL-based method has its value in the detection of underlying network differentiation. Since CL requires stronger "signal" when underlying dependent networks are truly differential, if CL concludes a high probability of network differentiation, then it is very likely that the finding reveals the underlying truth. As such, when two populations are independent, we would still suggest the CL method over the MDL approach due to its computing efficiency. When two populations are dependent, the CL method can be used as a "screening" technique; if CL shows two networks are differential, then the MDL can be utilized to obtain more detailed information including assessment of network differentiation, estimate of shared network, along with the different network structures unique to each population. An effort to improve the power of the CL-based method when two populations are dependent, while maintaining its computing efficiency, is certainly along the line of future work.

In our method, we focus on a non-informative prior for graph comparison in order to be as objective as possible. However, when the sample size is small and/or the signal in the data is likely to be weak, non-informative priors may cause difficulty to differentiate between two graphs even if the two networks are truly differential. In such situations, prior knowledge or belief will be highly appreciated and benefit posterior inferences. This consideration of prior distribution assignment is also applicable to graph constructions.

Furthermore, as noted in our simulations, the proposed CL-based approach has difficulty to handle underlying shared graphs between two dependent populations. Since the focus of our work is on detecting network differentiation, addressing the role of shared graphs is not within the scope of the present study. However, it can be a promising future research direction under the framework of composite likelihood. If successful, we do not have to rely on complex modeling for dependence between populations and will achieve both computing efficiency and graph inference with sound quality. One way is to incorporate a modeling explicitly for the shared graph between two populations using composite likelihood, but a careful design is needed to model the networks in the two populations on top of the shared graphs.

Acknowledgment

The authors are thankful to the High Performance Computing at the University of Memphis.

Funding

This work is supported by NIAID/NIH R21AI175891 and R01AI121226.

References

[1] Hotta K, Kitamoto A, Kitamoto T, et al. Identification of differentially methylated region (dmr) networks associated with progression of nonalcoholic fatty liver disease. Scientific reports. 2018;8(1):1–11.

[2] Gill R, Datta S, Datta S. A statistical framework for differential network analysis from microarray data. BMC bioinformatics. 2010;11(1):1.

[3] Xia Y, Cai T, Cai TT. Testing differential networks with applications to the detection of gene-gene interactions. Biometrika. 2015;102(2):247–266.

[4] Zhao SD, Cai TT, Li H. Direct estimation of differential networks. Biometrika. 2014; 101(2):253–268.

[5] Städler N, Dondelinger F, Hill SM, et al. Molecular heterogeneity at the network level: high-dimensional testing, clustering and a tcga case study. Bioinformatics. 2017; 33(18):2890–2896.

[6] Jacob L, Neuvial P, Dudoit S. More power via graph-structured tests for differential expression of gene networks. The Annals of Applied Statistics. 2012;:561–600.

[7] He H, Cao S, Zhang Jg, et al. A statistical test for differential network analysis based on inference of gaussian graphical model. Scientific reports. 2019;9(1):1–8.

[8] Cai T, Liu W, Xia Y. Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. Journal of the American Statistical Association. 2013; 108(501):265–277.

[9] Chang J, Zhou W, Zhou WX, et al. Comparing large covariance matrices under weak conditions on the dependence structure and its application to gene clustering. Biometrics. 2017;73(1):31–41.

[10] Zhang H, Huang X, Arshad H. Comparing dependent undirected gaussian networks. Bayesian Analysis. 2022;1(1):1–26.

[11] Varin C, Reid N, Firth D. An overview of composite likelihood methods. Statistica Sinica. 2011;:5–42.

[12] Zhang H, Stern H. Assessment of ancestry probabilities in the presence of genotyping errors. Theoretical and Applied Genetics. 2006;112:472–482.

[13] Gao B, Yang C, Liu J, et al. Accurate genetic and environmental covariance estimation with composite likelihood in genome-wide association studies. PLoS genetics. 2021; 17(1):e1009293.

[14] Varin C. On composite marginal likelihoods. Asta advances in statistical analysis. 2008; 92(1):1–28.

[15] Larribe F, Fearnhead P. On composite likelihoods in statistical genetics. Statistica Sinica. 2011;:43–69.

[16] Lindsay BG. Composite likelihood methods. Comtemporary Mathematics. 1988; 80(1):221–239.

[17] Besag J. Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society: Series B (Methodological). 1974;36(2):192–225.

[18] Besag J. Statistical analysis of non-lattice data. Journal of the Royal Statistical Society Series D: The Statistician. 1975;24(3):179–195.

[19] George EI, McCulloch RE. Variable selection via Gibbs sampling. Journal of the American Statistical Association. 1993;88:881–889.

[20] Ishwaran H, Rao JS. Spike and slab gene selection for multigroup microarray data. Journal of the American Statistical Association. 2005;100(471):764–780.

[21] Ishwaran H, Rao JS. Spike and slab variable selection: Frequentist and Bayesian strategies. The Annals of Statistics. 2005;33:730–773.

[22] Wang H, et al. Scaling it up: Stochastic search structure learning in graphical models. Bayesian Analysis. 2015;10(2):351–377.

[23] Jones B, Carvalho C, Dobra A, et al. Experiments in stochastic computation for high-dimensional graphical models. Statistical Science. 2005;:388–400.

[24] Fan J, Feng Y, Wu Y. Network exploration via the adaptive lasso and scad penalties. The

508 annals of applied statistics. 2009;3(2):521.
509 [25] Li H, Gui J. Gradient directed regularization for sparse gaussian concentration graphs,
510 with applications to inference of genetic networks. Biostatistics. 2006;7(2):302–317.
511 [26] Arshad SH, Holloway JW, Karmaus W, et al. Cohort profile: The isle of wight whole
512 population birth cohort (iowbc). International journal of epidemiology. 2018;47:1043–
513 1044.

Appendix

A. Model assessment statistics for dependent networks

Table A.1. Simulation results for dependent networks with 10 nodes

n	EP (%)	TP (95% EI)	TN (95% EI)	CC (95% EI)
Underlying truth: identical networks (P=10 nodes, $\|E_{X_1}\| = \|E_{X_2}\| = 9$ edges)				
50	100	0.167 (0.167, 0.333)	1.0 (0.970, 1.0)	0.778 (0.756, 0.817)
100	100	0.333 (0.167, 0.417)	1.0 (0.970, 1.0)	0.778 (0.756, 0.817)
200	100	0.5 (0.333, 0.5)	1.0 (0.970, 1.0)	0.856 (0.816, 0.867)
500	100	0.5 (0.458, 0.583)	1.0 (0.970, 1.0)	0.867 (0.844, 0.889)
Underlying truth: differential networks (P=10 nodes, $\|E_{X_1}\| = 9$, $\|E_{X_2}\| = 9$ edges)				
50	46	X_1: 0.117 (0.059, 0.235)	1.0 (0.964, 1.0)	0.667 (0.633, 0.7)
		X_2: 0.222 (0.111, 0.444)	1.0 (0.972, 1.0)	0.844 (0.805, 0.889)
100	48	X_1: 0.25 (0.14, 0.25)	1.0 (0.966, 1.0)	0.722 (0.689, 0.733)
		X_2: 0.444 (0.222, 0.667)	1.0 (0.972, 1.0)	0.878 (0.839, 0.922)
200	100	X_1: 0.235 (0.132, 0.294)	1.0 (0.973, 1.0)	0.711 (0.667, 0.733)
		X_2: 0.556 (0.333, 0.778)	1.0 (0.972, 1.0)	0.9 (0.867, 0.933)
500	100	X_1: 0.294 (0.235, 0.353)	0.982 (0.947, 1.0)	0.711 (0.683, 0.750)
		X_2: 0.889 (0.5, 1.0)	1.0 (0.972, 1.0)	0.967 (0.933, 1.0)

Table A.2. Simulation results for dependent networks with 30 nodes

n	EP (%)	TP (95% EI)	TN (95% EI)	CC (95% EI)
Underlying truth: identical networks (P=30 nodes, $\|E_{X_1}\| = \|E_{X_2}\| = 29$ edges)				
50	93	0.115 (0.077, 0.154)	0.997 (0.993, 1.0)	0.892 (0.885, 0.897)
100	98	0.231 (0.173, 0.269)	0.995 (0.989, 0.999)	0.902 (0.895, 0.91)
200	100	0.269 (0.24, 0.327)	0.997 (0.991, 1.0)	0.911 (0.904, 0.917)
500	100	0.298 (0.269, 0.327)	0.999 (0.994, 1.0)	0.915 (0.91, 0.92)
Underlying truth: differential networks (P=30 nodes, $\|E_{X_1}\| = 29$, $\|E_{X_2}\| = 29$ edges)				
50	100	X_1: 0.056 (0.019, 0.093)	0.997 (0.992, 1.0)	0.88 (0.874, 0.887)
		X_2: 0.231 (0.115, 0.346)	0.998 (0.994, 1.0)	0.951 (0.944, 0.957)
100	100	X_1: 0.148 (0.093, 0.185)	0.993 (0.988, 0.999)	0.889 (0.88, 0.897)
		X_2: 0.423 (0.308, 0.577)	0.995 (0.991, 0.999)	0.962 (0.953, 0.973)
200	100	X_1: 0.204 (0.153, 0.269)	0.993 (0.988, 0.998)	0.894 (0.886, 0.905)
		X_2: 0.769 (0.615, 0.885)	0.999 (0.995, 1.0)	0.983 (0.976, 0.993)
500	100	X_1: 0.241 (0.204, 0.269)	0.997 (0.991, 1.0)	0.903 (0.896, 0.913)
		X_2: 0.942 (0.826, 1.0)	0.999 (0.995, 1.0)	0.994 (0.987, 1.0)

B. Trace plots assessing convergence

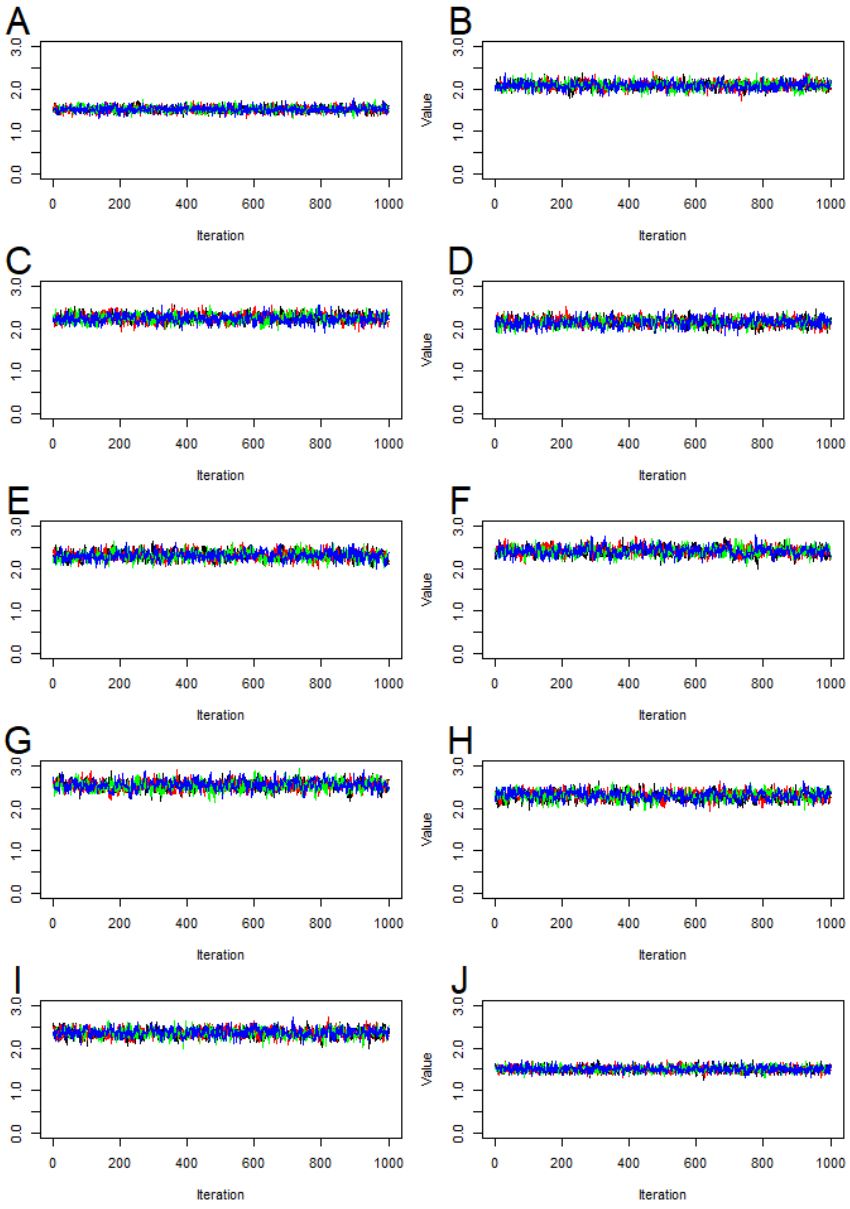

Figure B.1. Trace plots of posterior samples for the diagonal elements of a precision matrix (Ω_1) with 10 nodes. (A) corresponds to the first diagonal element. (B) corresponds to the second diagonal element, and so on.