

RESEARCH ARTICLE

Empowering differential networks using Bayesian analysis

Jarod Smith¹, Mohammad Arashi^{1,2*}, Andriette Bekker¹**1** Department of Statistics, University of Pretoria, Pretoria, South Africa, **2** Department of Statistics, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, Mashhad, Iran

All these authors are contributed equally to this work.

* arashi@um.ac.ir



Abstract

Differential networks (DN) are important tools for modeling the changes in conditional dependencies between multiple samples. A Bayesian approach for estimating DNs, from the classical viewpoint, is introduced with a computationally efficient threshold selection for graphical model determination. The algorithm separately estimates the precision matrices of the DN using the Bayesian adaptive graphical lasso procedure. Synthetic experiments illustrate that the Bayesian DN performs exceptionally well in numerical accuracy and graphical structure determination in comparison to state of the art methods. The proposed method is applied to South African COVID-19 data to investigate the change in DN structure between various phases of the pandemic.

OPEN ACCESS

Citation: Smith J, Arashi M, Bekker A (2022) Empowering differential networks using Bayesian analysis. PLoS ONE 17(1): e0261193. <https://doi.org/10.1371/journal.pone.0261193>

Editor: Marton Karsai, Central European University, HUNGARY

Received: June 1, 2021

Accepted: November 24, 2021

Published: January 25, 2022

Copyright: © 2022 Smith et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data underlying the results presented in the study are available from <https://archive.ics.uci.edu/ml/datasets/spambase> for the spambase dataset. The corresponding COVID-19 data are available from <https://www.nicd.ac.za/diseases-a-z-index/disease-index-covid-19/surveillance-reports/> and <https://ourworldindata.org/coronavirus/country/south-africa>.

Funding: This work was based upon research supported in part by the National Research Foundation (NRF) of South Africa, SARChI Research Chair UID: 71199; Ref.:

Introduction

Probabilistic networks are becoming ever-present in a multitude of scientific disciplines. These networks aim to illustrate the relationships, if any, between the components of complex systems [1]. If the data is assumed to be Gaussian distributed with mean μ and covariance matrix Σ ; the precision matrix $\Theta := \{\theta_{ij}\}$, defined as the inverse of the covariance matrix $\Theta \equiv \Sigma^{-1}$, directly determines the conditional dependence relations and structure of the Gaussian undirected graphical model [2].

Differential network (DN) analysis is a statistical methodology that involves functions of at least two graphical models. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ define a graphical model with nodes $\mathcal{V} = \{1, 2, \dots, p\}$ and a set of edges $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. The graph visually depicts the conditional dependence structure between the nodes of the system. The adjacency matrix associated to a graphical model \mathcal{G} is the binary encoded $p \times p$ precision matrix where the entries of the matrix are equal to 1 if the corresponding precision matrix entry is nonzero and zero otherwise. Non-zero adjacency matrix entries indicate an edge between corresponding nodes of \mathcal{G} . For this work, the focus will be on the difference of two Gaussian graphical models (GGM), \mathcal{G}_1 and \mathcal{G}_2 that share the same set of nodes \mathcal{V} . In particular, the edge sets given here are equivalent to the adjacency matrices obtained from the GGM estimation. More specifically, assume that the observations, $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{n_1}$ and $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{n_2}$ are generated from a p variate Gaussian distribution, $N_p(\mu_1, \Sigma_1)$ and $N_p(\mu_2, \Sigma_2)$, respectively, where n_1 and n_2 indicate the respective sample

SRUG190308422768 grant No. 120839. The opinions expressed and conclusions arrived at are those of the authors and are not necessarily to be attributed to the NRF. The research of the corresponding author is supported by a grant from Ferdowsi University of Mashhad (N.2/55265).

Competing interests: The authors have declared that no competing interests exist.

Abbreviations: List of abbreviations: ADMM, Alternating direction method of multipliers; BAGLASSO, Bayesian adaptive graphical lasso; B-net, Bayesian differential network; CRAN, Comprehensive R archive network; DE, Double exponential; DN, Differential network; D-net, Iterative shrinkage-thresholding differential network; EL1, L1 eigenvalue loss; EL2, Frobenius eigenvalue loss; EXP, Exponential; FNR, False negative rate; FN, False negative; FP, False positive; F1, F1-score; GCL, Graphical cooperative lasso; GGM, Gaussian graphical model; IGL, Intertwined graphical lasso; L1, L1 loss; L2, Frobenius loss; MAP, Maximum a posteriori; MAXEL1, L1 loss on the largest eigenvalue; MINEL2, L1 loss on the smallest eigenvalue; MCC, Matthews correlation coefficient; SE, Sensitivity; SP, Specificity; TN, True negative; TP, True positive; **List of symbols:** $\tilde{\theta}_{ij}$, Element ij of the adaptive graphical lasso sample precision matrix; \mathcal{G} , Graphical model; \mathcal{E} , Set of edges of a graphical model \mathcal{G} ; \mathcal{V} , Set of nodes of a graphical model \mathcal{G} ; \mathbb{M}^+ , Space of positive definite matrices; Σ , Gaussian covariance matrix; n , Sample size; P , Partial correlation matrix; r , Gamma shape parameter; s , Gamma scale parameter; S , Sample covariance matrix; Δ , Differential network; η , Bayesian sparsity criterion threshold; Θ , Gaussian precision matrix; θ_{ij} , Element ij of the precision matrix Θ ; μ , Gaussian mean vector; ξ , Weights of the adaptive graphical lasso; ρ , Lasso penalty parameter; τ , Scale mixture of Gaussian scale parameter.

sizes that need not be equal. The interest here is estimating the DN ($\Delta = \Sigma_2^{-1} - \Sigma_1^{-1}$), that is the difference between two precision matrices. Numerous measures exist for comparing and evaluating the differences between graphical structures [1]. DN analysis is becoming increasingly popular and important, for example in biological systems where protein interaction networks can be utilised as informative biosignatures for prevalent diseases [3, 4]. The fundamental idea here is that, if two molecules interact with one another then a statistical dependency between them should be observed. Additionally, another application of DNs is multivariate statistical quadratic discriminant analysis [5, 6], under the Gaussian distribution assumption.

A key component of DN analysis is the estimation of covariance and precision matrix components. Numerous statistical matrix estimation, as well as graphical model determination methods exist within literature. In particular, from a frequentist approach [7], introduce a computationally efficient neighborhood selection procedure. The lasso is used for covariance estimation which enjoys consistency for sparse high-dimensional graphs. The approach is quite effective, in that the sparse precision matrix is estimated by fitting the lasso to each variable using the remaining as predictors. Finally, the estimated precision matrix entry (θ_{ij}) is non-zero if the estimated coefficient of i on j or vice versa is non-zero. Importantly, their algorithm can consistently estimate the set of non-zero entries in Θ , [8]. For a penalised likelihood methodology for sparse precision matrix estimation see [9, 10]. More so [11], estimate the undirected graphical model using both a block coordinate descent algorithm, as well as Nesterov's first order method [12]. Additionally [13], propose a ℓ_1 constraint estimation technique for both sparse and non-sparse high dimensional matrices with applicability on a wide range of sparsity patterns and class of matrices; precision estimation in GGMs for example. For a joint graphical model estimation approach see [14, 15].

Fully Bayesian treatments of GGM estimation are, also, well rooted in literature. In particular [16], introduce the Bayesian adaptive graphical lasso (BAGLASSO) which utilises a generalised Pareto distribution in the hierarchical formulation of the Bayesian graphical lasso. [17] provide a method for graphical model determination by invoking positive prior mass on the event that there is no conditional dependencies between variables. In terms of joint graphical model inference from a Bayesian perspective see [18]. Lastly [19], propose using Kullback-Leibler divergence and cross-validation for graphical model structure estimation.

Background

Recently, a plethora of statistical techniques have emerged for estimating DNs. These techniques can largely be classified into two main categories. The first estimating the individual precision matrices, Θ_1 and Θ_2 separately; where the estimated DN is the difference between the estimated precision matrices. For example, the methods and references for GGM estimation outlined in the introduction can be used to directly estimate Δ . The second methodology estimates both the precision matrices simultaneously. The approach here, typically penalises a joint loss function for both precision matrices. [20] provide a methodology for inference and estimation of functions of GGMs. In particular, the Intertwined Graphical Lasso (IGL) approach biases the estimation of the precision matrices towards a common value. More so, their Graphical Cooperative Lasso (GCL) utilises a group-penalty for solutions that favour a common sparsity pattern. [14, 21] estimate separate graphical models using a joint penalised loss function. [22] propose a method for estimating Δ directly which relaxes the need for both individual precision matrices to be sparse nor be estimated directly. Similarly [6, 23], utilise an alternating direction method of multipliers (ADMM) algorithm for estimating Δ from their joint ℓ_1 penalised convex loss function. More recently [24], introduce a computationally

efficient iterative shrinkage-thresholding algorithm for minimising the ℓ_1 loss function defined in [6], namely

$$L_1(\Delta) = \frac{1}{2} \text{trace}(\Delta^\top \mathbf{S}_1 \Delta \mathbf{S}_2) - \text{trace}(\Delta(\mathbf{S}_1 - \mathbf{S}_2)), \tag{1}$$

is convex and \mathbf{S}_1 and \mathbf{S}_2 are the sample covariance matrices. The DN estimate is obtained by minimising the penalised loss Eq (1). An analogous symmetric convex loss function and estimator is proposed by [23].

The shrinkage-thresholding algorithm proposed by [24], based on the fast-iterative shrinkage-thresholding algorithm in [25], aims to minimise Eq (1). The objective function is given by

$$\arg \min_{\Delta \in \mathbb{R}^{p \times p}} L_1(\Delta) + \rho \|\Delta\|_1,$$

where $\|\Delta\|_1 = \sum_{i < j} \sum_{i=1}^p |\hat{\Theta}_{2:ij} - \hat{\Theta}_{1:ij}|$. The lasso tuning parameter, ρ , controls the strength of the penalty term and resultantly the amount of shrinkage (precision matrix entries shrunk towards zero) too. The optimisation objective converges to the solution sequentially using a quadratic approximation and a gradient descent algorithm. The efficiency of the procedure is attested to this approach, resulting in superior computational complexity in contrast to the ADMM approaches by [6, 23]. To conclude this section it is worth noting that the iterative shrinkage-thresholding method will be used for experimental comparison later.

The main contributions of this study are as follows.

1. A framework for Bayesian DN estimation is developed. That is, the DN is estimated by separately estimating each Gaussian graphical model, referred to as the components.
2. The graphical lasso is applied as the thresholding method in the Bayesian precision matrix estimation in order to efficiently capture sparse patterns in the DN, hence developing the BAGLASSO. A threshold selection strategy, based on a conjugate Wishart prior, that accommodates both dense and sparse graphical structures determination is explored. The aforementioned strategy, applied to each component of the DN, ensures an accurately sparse DN estimate.
3. The proposed Bayesian DN efficiently improves the existing classical DN estimation for a number of known network structures.
4. An R package for the BAGLASSO block Gibbs sampler has been developed for the interested practitioner and is available on The Comprehensive R Archive Network (CRAN) as [abglasso](#).

The Bayesian DN

A fully Bayesian treatment of DNs remains unexplored and the novel methodology here aims to develop a simple yet highly accurate Bayesian DN estimation procedure. The novel contribution utilises the BAGLASSO as a launching point to separately estimate the components of the DN. The subsections that follow develop the framework for individual component estimation from a Bayesian viewpoint. Moreover, the framework has been develop for low $p = 10$ to moderate, $p = 50 - 100$, dimensions where $n \geq p$.

The Bayesian graphical lasso prior

Recall that the graphical lasso objective is maximising the penalized log-likelihood

$$\arg \min_{\Theta \in \mathbb{M}^+} l(\Theta),$$

where

$$l(\Theta) = \log(\det\Theta) - \text{trace}\left(\frac{\mathbf{S}}{n}\Theta\right) - \rho\|\Theta\|_1 \tag{2}$$

and M^+ is the space of positive definite matrices, \mathbf{S} is the sample covariance matrix and n the sample size, respectively. More over, $\rho \geq 0$ is the shrinkage parameter and $\Theta = (\theta_{ij})$ is the precision matrix. The Bayesian connection to the graphical lasso problem is the maximum a posteriori (MAP) estimate, assuming a random sample from $N_p(\boldsymbol{\mu}, \Theta^{-1})$, of the following

$$p(\Theta | \lambda) = C^{-1} \prod_{i < j} \{\text{DE}(\theta_{ij} | \lambda)\} \prod_{i=1}^p \{\text{EXP}(\theta_{ii} | \lambda)\} \quad (\Theta \in \mathbb{M}^+). \tag{3}$$

The prior distribution is given by the product of a double exponential (DE) with form $p(y) = \lambda/2 \exp(-\lambda|y|)$ for the off diagonal elements and an exponential (EXP) with form $p(y) = \lambda \exp(-\lambda y)1_{y > 0}$, otherwise. The value of Θ which maximizes the posterior density is the graphical lasso estimate in Eq (2) when $\rho = \lambda/n$. Within the Bayesian context λ is treated as the shrinkage parameter. The formulation and interpretations of the graphical lasso prior in Eq (3) have been studied in [26]. The aim therein is the development of varying regularization to infer block structures within the graphical models and efficiently estimating the maximum a posteriori of the corresponding posterior distribution. [16] make use of this prior formulation for the convenience (scale mixture of Gaussian formulation of the double exponential) in the development of their efficient block Gibbs sampler, in addition to allowing for the use of a gamma hyperprior on the shrinkage parameter λ for improved precision matrix estimation.

Hierarchical representation

The Gibbs sampler for sampling the precision matrix Θ from the posterior distribution, defined below in Eq (5), associated with the prior in Eq (3), is constructed using a hierarchical representation of Eq (3). This particular hierarchical representation of the prior in Eq (3) is presented by [16], whom follow the same approach as in the development of the Gibbs sampler for the Bayesian lasso in [27]. The Gibbs sampler in [27] utilises the structure of the double exponential distribution as a scale mixture of Gaussians, assuming independence of the conditional double exponential priors [28, 29], in their hierarchical representation to simulate regression parameters from the desired posterior distribution. The positive definite constraint on Θ in Eq (3) implies that the Gaussian components for θ_{ij} (DE parameters) in the scale mixture formulation are no longer independent given the scale parameters. To address this issue, the hierarchical representation of the graphical lasso prior in Eq (3) is given by

$$p(\boldsymbol{\theta} | \boldsymbol{\tau}, \lambda) = C_{\boldsymbol{\tau}}^{-1} \prod_{i < j} \left\{ \frac{1}{\sqrt{2\pi\tau_{ij}}} \exp\left(-\frac{\theta_{ij}^2}{2\tau_{ij}}\right) \right\} \prod_{i=1}^p \left\{ \frac{\lambda}{2} \exp\left(-\frac{\lambda}{2}\theta_{ii}\right) \right\} \quad (\Theta \in \mathbb{M}^+), \tag{4}$$

where $\boldsymbol{\theta} := \{\theta_{ij}\}_{i < j}$ is a vector of the upper triangular matrix entries of Θ and $\boldsymbol{\tau} = \{\tau_{ij}\}_{i < j}$ the scale parameters. The normalising constant, $C_{\boldsymbol{\tau}}$, has no closed-form solution. Obtaining the marginal distribution Eq (3), [16] propose a mixing density proportional to an exponential density with rate parameter $\lambda^2/2$ and simple substitution circumvents the intractable normalising constant. Finally, the hierarchical representation in Eq (4) is used in the development of

the block Gibbs sampler, available in the [S1 File](#), with a target posterior distribution given by

$$p(\Theta, \tau | \mathbf{Y}, \lambda) \propto \det \Theta^{\frac{n}{2}} \exp\left\{-\text{trace}\left(\frac{1}{2} \mathbf{S} \Theta\right)\right\} \prod_{i < j} \left\{ \tau_{ij}^{-\frac{1}{2}} \exp\left(-\frac{\theta_{ij}^2}{2\tau_{ij}}\right) \exp\left(-\frac{\lambda^2}{2} \tau_{ij}\right) \right\} \times \prod_{i=1}^p \left\{ \exp\left(-\frac{\lambda}{2} \theta_{ii}\right) \right\} (\Theta \in \mathbb{M}^+). \tag{5}$$

BAGLASSO

It is well known that the double exponential prior in [Eq \(3\)](#) may over-shrink (under-shrink) large (small) coefficients in Θ . The limitations within a regression context have been studied in [\[30–32\]](#) with alternative proposals. The BAGLASSO, Bayesian analog to the adaptive graphical lasso [\[33\]](#), exploits the framework and flexibility of the hierarchical representation in [Eq \(4\)](#) to address the aforementioned limitation. This extension serves to improve the accuracy of the precision matrix estimates obtained from the posterior in [Eq \(5\)](#) by allowing for different shrinkage parameters λ_{ij} for each corresponding off-diagonal precision matrix entry θ_{ij} . Recall that the adaptive graphical lasso is given by

$$\arg \min_{\Theta \in \mathbb{M}^+} l(\Theta),$$

where

$$l(\Theta) = \log(\det \Theta) - \text{trace}\left(\frac{\mathbf{S}}{n} \Theta\right) - \lambda \sum_{1 \leq i \leq p} \sum_{1 \leq j \leq p} \xi_{ij} |\theta_{ij}|, \tag{6}$$

and $\xi_{ij} = 1/|\tilde{\theta}_{ij}|^\alpha$ for $\alpha > 0$ are the adaptive weights and the weight matrix $(\tilde{\theta}_{ij})$ is the sample precision matrix.

The form of the Bayesian graphical lasso in [Eq \(3\)](#) enables the selection of an appropriate hyperprior on the shrinkage parameter λ , recall that $\rho = \lambda/n$ in the Bayesian formulation of [Eq \(2\)](#). Adhering to the positive definite constraint on Θ , the prior normalising constant in [Eq \(3\)](#) when a single λ is applied to all elements in Θ can be obtained by applying the substitution $\tilde{\Theta} = \lambda \Theta$. Thereafter, a gamma prior $\lambda \sim \text{GA}(r, s)$ and corresponding conditional posterior $\lambda \sim \text{GA}(r + p(p + 1), s + \|\Theta\|_1/2)$ can be obtained and sampled from. When allowing for individual λ_{ij} 's for different off-diagonal θ_{ij} 's, the normalising constant C will inevitably depend on λ_{ij} . To address this a hierarchical formulation can be used to construct a set of prior distributions, serving as the the extension of the graphical lasso prior in [Eq \(3\)](#), for various λ_{ij} that mitigate the complications associated with posterior simulation due to the intractable normalising constant. This extension is the BAGLASSO and, assuming a random sample from $N_p(\mu, \Theta^{-1})$, is given by

$$p(\Theta | \{\lambda_{ij}\}_{i < j}) = C_{\{\lambda_{ij}\}_{i < j}}^{-1} \prod_{i < j} \{ \text{DE}(\theta_{ij} | \lambda_{ij}) \} \prod_{i=1}^p \left\{ \text{EXP}(\theta_{ii} | \frac{\lambda_{ii}}{2}) \right\} (\Theta \in \mathbb{M}^+),$$

$$p(\{\lambda_{ij}\}_{i < j} | \{\lambda_{ii}\}_{i=1}^p) \propto C_{\{\lambda_{ij}\}_{i < j}} \prod_{i < j} \text{GA}(r, s). \tag{7}$$

The normalising constant $C_{\{\lambda_{ij}\}_{i < j}}$ is intractable, as mentioned above, and the set $\{\lambda_{ii}\}_{i=1}^p$ are hyperparameters for the diagonal elements of Θ . Simple substitution yields that computation of λ_{ij} is simplified by circumventing the intractable normalising constant.

The BAGLASSO selects the amount of shrinkage λ_{ij} proportionally to the current value of θ_{ij} . To see this [\[16\]](#), demonstrate that the conditional posterior, $\lambda_{ij} | \Theta \sim \text{GA}(r + 1, |\theta_{ij}| + s)$, has an expected value that is inversely related to magnitude of θ_{ij} . The data augmented block

Gibbs sampler for the hierarchical representation in Eq (7) is the fundamental building block upon which the novel Bayesian DN is devised.

Technicalities on conditional dependencies

Recall that the precision matrix directly determines the conditional dependence relations and structure of the undirected graphical model. Therefore, correctly estimating the precision matrices with sparse structures is essential to adequately gauge the conditional dependency relations between variables. The task to estimate the precision matrix for both $n < p$ and $p \leq n$ remains challenging and regularization is often required [34–36]. A popular choice of prior for Bayesian posterior inference regarding network structure is the conjugate Wishart [37]. An alternative thresholding strategy is presented which is an adaption of the recommendation by [32]. In particular the conjugate Wishart $W(3, \epsilon \mathbf{I}_p)$ prior is used. The corresponding posterior is $W(3 + n, (\mathbf{S} + \epsilon \mathbf{I}_p)^{-1})$, where $\epsilon = 0.001$ and \mathbf{I}_p a p dimensional identity matrix. The posterior samples are used to compute the posterior distribution of the $p \times p$ partial correlation matrix $P := \{\rho_{ij}\}$. The recommended strategy here suggests $\theta_{ij} \neq 0$ for $i \neq j$ if

$$|E_h(\rho_{ij} | \mathbf{Y})| > \eta, \quad (8)$$

where η may vary depending on the underlying graph structure. The Bayesian posterior thresholding recommendation by [16] claim that $\theta_{ij} \neq 0$ for $i \neq j$ if and only if

$$\frac{\tilde{\rho}_{ij}}{E_g(\rho_{ij} | \mathbf{Y})} > \eta. \quad (9)$$

Noting that $\tilde{\rho}_{ij}$ is the posterior sample mean estimate of the partial correlation under graphical lasso priors in Eq (3); g is the standard conjugate Wishart $W(3, \mathbf{I}_p)$ and h the standard conjugate Wishart $W(3, \epsilon \mathbf{I}_p)$. Moreover, $\eta \in [0, 1]$ with the lower and upper bounds resulting in a completely dense or sparse estimate, respectively.

The original recommendation for η in Eq (9) is 0.5. The forthcoming synthetic data analysis section describes the simulation procedure, as well as, illustrates the performance of the Bayesian DN with regards to different graph structures, namely an AR(1), AR(2), sparse random, scale-free, band, cluster, star and circle. The goal here is to suggest a suitable sparsity threshold region under the varying graph structures for the recommended sparsity criterion in Eq (8). The Bayesian DN is applied across all graph structures with thresholds, η , in the range of 0.2 and 0.6 in increments of 0.02. The absolute sparsity error is computed for each graph structure for each Bayesian sparsity criterion in Eqs (8) and (9), respectively. The results are based on the median of 40 replications and the Matthews Correlation Coefficient (MCC), see [38], is used to determine the best performing threshold. Fig (1a)–(1i) display the optimal threshold, based on the top performing MCC, for each graph structure and Bayesian sparsity criterion for $p = 10$. The optimal threshold plots for $p = 30$ and $p = 100$ are available in the S1 File. The optimal threshold based on Eq (8), η^* , for the Bayesian DN is, in most cases, in the neighborhood of the minimum absolute sparsity error and in the region of $\eta^* \in \{0.2 - 0.4\}$. Both Bayesian sparsity criterion candidates perform comparably well noting, however, that Eq (8) requires less computation.

Synthetic data analysis

The synthetic experiment is designed to test the parameter estimation and graphical structure determination of the DN estimation for both the novel Bayesian approach (referred to as ‘B-net’) and the iterative shrinkage-thresholding estimator (referred to as ‘D-net’) from [24]. The

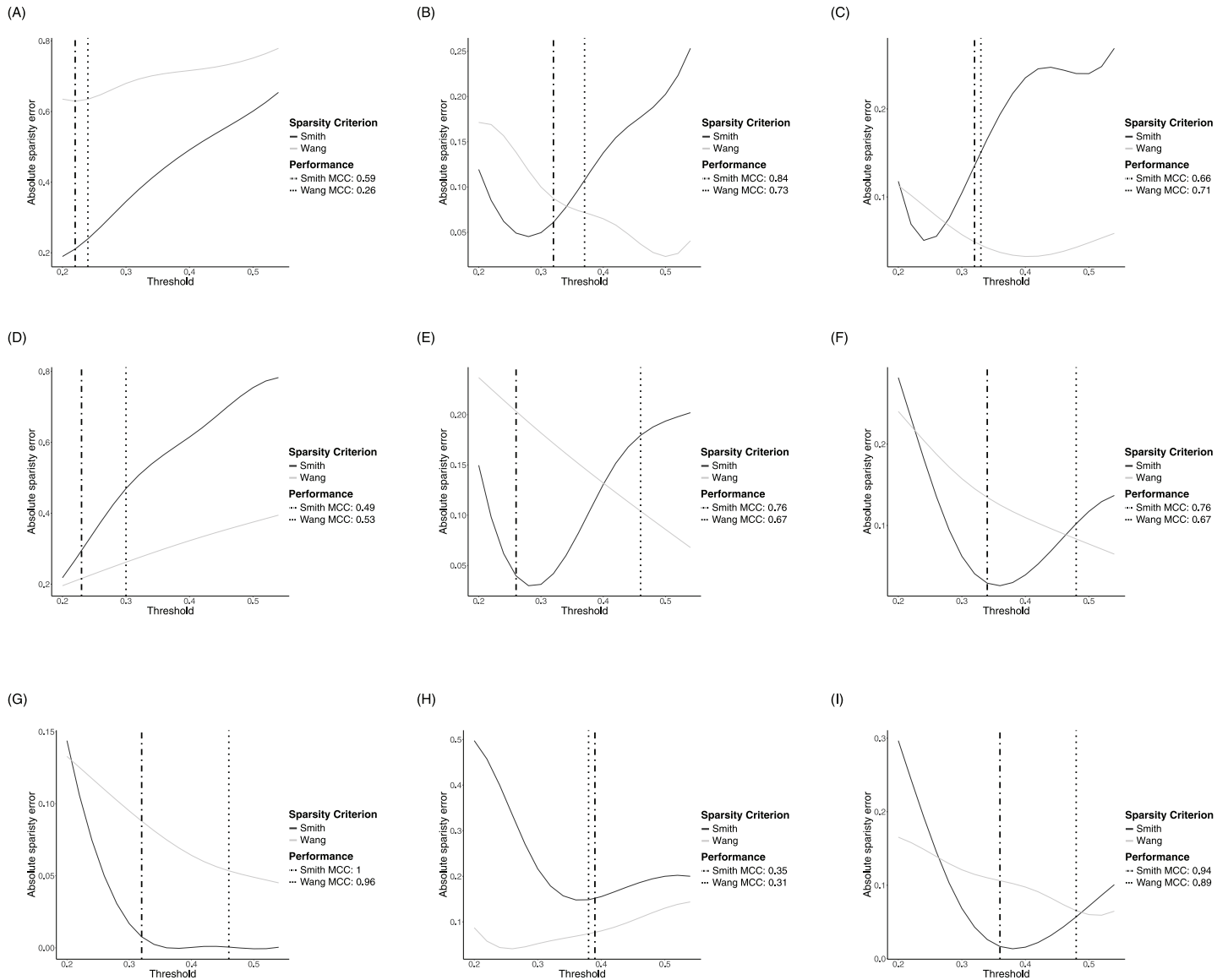


Fig 1. Optimal Bayesian sparsity threshold selection. The median of the absolute sparsity error and best performing MCC for various graph structures under varying thresholds for each Bayesian sparsity criterion in Eq (9) (dotted) and Eq (8) (dot-dash) for dimension $p = 10$. The best performing threshold is indicated by a vertical line with the accompanying MCC value displayed in the legend. (a) Model 1: AR(1). (b) Model 2: AR(2). (c) Model 3: at most 80% sparse. (d) Model 4: at most 40% sparse. (e) Model 5: scale-free. (f) Model 6: band. (g) Model 7: cluster. (h) Model 8: star. (i) Model 9: circle.

<https://doi.org/10.1371/journal.pone.0261193.g001>

iterative shrinkage-thresholding estimator uses the lasso penalty and Bayesian Information Criterion (BIC) for model estimation and selection, respectively. For all simulations, the assumption is that the observations, $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{n_1}$ and $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{n_2}$ are generated from a Gaussian $N_p(0, \Sigma_1)$ and $N_p(0, \Sigma_2)$ respectively. The true DN is

$$\Delta = \Sigma_2^{-1} - \Sigma_1^{-1},$$

where the true precision matrices are $\Theta_1 = \Sigma_1^{-1}$ and $\Theta_2 = \Sigma_2^{-1}$. The Bayesian DN applies the BAGLASSO Eq (7) to each sample, i.e. separately estimates the precision matrices. Furthermore, for excellent performance set $r = 10^{-2}$ and $s = 10^{-6}$, see S1 File for more details, for the

hyperparameters of the prior distributions of λ_{ij} for $i < j$ and $\lambda_{ii} = 1$ for $i = 1, \dots, p$. The iterative shrinkage-thresholding approach jointly estimates the precision matrices for Eq (1). The following 9 graphical structure variations are considered—where the structure of each is applied to each component in the DN's composition to achieve the desired structure in the DN itself—in the simulation:

- *structure 1*: An AR(1) model.
 - *Component 1*: $\theta_{ij} = 0.7^{|i-j|}$.
 - *Component 2*: $\theta_{ij} = 0.75^{|i-j|}$.
- *structure 2*: An AR(2) model.
 - *Component 1*: $\theta_{ii} = 0.1$, $\theta_{i,i-1} = \theta_{i-1,i} = 0.05$ and $\theta_{i,i-2} = \theta_{i-2,i} = 0.025$.
 - *Component 2*: $\theta_{ii} = 1$, $\theta_{i,i-1} = \theta_{i-1,i} = 0.5$ and $\theta_{i,i-2} = \theta_{i-2,i} = 0.25$.
- *structure 3*: A sparse random model where both components have approximately up to 80% off-diagonal elements set to zero.
- *structure 4*: A moderately sparse random model where both components have approximately up to 40% off-diagonal elements set to zero.
- *structure 5*: A scale-free model where the second component is a scalar multiple of the first.
- *structure 6*: A band or diagonal model.
 - *Component 1*: $\theta_{ii} = 1$, $\theta_{ij} = 0.2$ for $1 \leq i \neq j \leq p/2$, $\theta_{ij} = 0.5$ for $p/2 + 1 \leq i \neq j \leq p$ and $\theta_{ij} = 0$ otherwise.
 - *Component 2*: $\theta_{ii} = 1$, $\theta_{ij} = 0.7$ for $1 \leq i \neq j \leq p/2$, $\theta_{ij} = 0.9$ for $p/2 + 1 \leq i \neq j \leq p$ and $\theta_{ij} = 0$ otherwise.
- *structure 7*: A cluster model containing two disjoint groups.
 - *Component 1*: $\theta_{ii} = 1$, $\theta_{ij} = 0.5$ for $1 \leq i \neq j \leq p/2$, $\theta_{ij} = 0.5$ for $p/2 + 1 \leq i \neq j \leq p$ and $\theta_{ij} = 0$ otherwise.
 - *Component 2*: $\theta_{ii} = 1$, $\theta_{ij} = 0.9$ for $1 \leq i \neq j \leq p/2$, $\theta_{ij} = 0.9$ for $p/2 + 1 \leq i \neq j \leq p$ and $\theta_{ij} = 0$ otherwise.
- *structure 8*: A star model with every node connected to the first node.
 - *Component 1*: $\theta_{ii} = 1$, $\theta_{1,i} = \theta_{i,1} = 0.1$ and $\theta_{ij} = 0$ otherwise.
 - *Component 2*: $\theta_{ii} = 1$, $\theta_{1,i} = \theta_{i,1} = 2.1$ and $\theta_{ij} = 0$ otherwise.
- *structure 9*: A circular model.
 - *Component 1*: $\theta_{ii} = 2$, $\theta_{i,i-1} = \theta_{i-1,i} = 1$ and $\theta_{1,p} = \theta_{p,1} = 0.45$.
 - *Component 2*: $\theta_{ii} = 4$, $\theta_{i,i-1} = \theta_{i-1,i} = 2$ and $\theta_{1,p} = \theta_{p,1} = 0.95$.

The sample sizes and dimensions for each model are $n_1 = n_2 \in \{50, 100, 200\}$ and $p_1 = p_2 \in \{10, 30, 100\}$, respectively. The Bayesian estimates are based on 10000 Monte Carlo iterations after 5000 burn-in iterations. To assess the performance of DN matrix estimation, six loss functions are considered and defined in Table 1, where p denotes the dimension and γ_i the i^{th} eigenvalue, respectively. Notice that some loss functions utilise the true DN matrix and its estimates, while others utilise the eigenvalues and their respective estimates. Table 2 reports the

Table 1. Loss functions used in the synthetic data analysis to assess the numerical accuracy of the B-net and D-net estimates.

Measure	Loss function	Abbreviation
Matrix L_1 -norm	$\ \hat{\Delta}-\Delta\ _1 = \max_{1 \leq j \leq p} \sum_{i=1}^p \hat{\Delta}_{ij} - \Delta_{ij} $	L1
Frobenius loss	$\ \hat{\Delta}-\Delta\ _F$, where $\ A\ _F^2 = \text{trace}(AA^T)$	L2
L_1 eigenvalue loss	$\sum_{i=1}^p \hat{\gamma}_i - \gamma_i /p$	EL1
L_2 eigenvalue loss	$\sum_{i=1}^p (\hat{\gamma}_i - \gamma_i)^2/p$	EL2
L_1 loss on the largest eigenvalue	$ \hat{\gamma}_{\max} - \gamma_{\max} $	MAXEL1
L_1 loss on the smallest eigenvalue	$ \hat{\gamma}_{\min} - \gamma_{\min} $	MINEL1

<https://doi.org/10.1371/journal.pone.0261193.t001>

median of L1, L2, EL1, EL2, MAXEL1 and MINEL1 for $p = 10, 30, 100$ in structures 1–9 based on 40 replications. For each scenario, the best performing measure is boldfaced.

The eigenvalue based loss functions are designed to investigate the extremes of the eigenvalue spectrum. In particular, the MAXEL1 loss function highlights which estimator is favourable in a principal component setting, [39]. A couple of observations are worth noting from Tables 2 and 3. First, the D-net estimator performs better with the AR(1) structure. Second, the B-net estimator performs exceptionally well in remaining structures. Third, the standard errors for both DN estimation techniques remain relatively consistent throughout the dimension spectrum considered, noting that the D-net estimator yields, in general, better results.

Table 2. Synthetic study median loss results.

	AR(1)		AR(2)		S80		S40		SF		Band		Cluster		Star		Circle	
	B-net	D-net	B-net	D-net	B-net	D-net	B-net	D-net	B-net	D-net	B-net	D-net	B-net	D-net	B-net	D-net	B-net	D-net
p = 10																		
L1	1.04	0.64	1.13	1.49	2.26	3.41	6.03	7.56	2.24	2.79	1	1.21	0.85	1.68	18.47	18	2.73	2
L2	0.91	0.64	1.41	2.09	2.21	3.8	6.09	7.45	2.14	2.22	1.39	1.91	1.05	2.48	8.73	8.49	3.87	4.3
EL1	0.14	0.1	0.24	0.53	0.23	0.61	0.63	1.12	0.11	0.48	0.22	0.43	0.25	0.58	1.19	1.2	0.42	1.23
EL2	0.03	0.03	0.09	0.35	0.1	0.47	0.47	1.73	0.02	0.49	0.07	0.21	0.1	0.52	6.02	7.2	0.27	1.83
MAXEL1	0.17	0.51	0.49	1.07	0.2	0.72	0.7	1.57	0.15	1.37	0.44	0.55	0.64	1.42	5.45	6	0.58	1.91
MINEL1	0.28	0.06	0.09	0.45	0.16	0.67	0.9	1.09	0.15	1.37	0.44	0.49	0.36	0.22	5.52	6	0.6	1.91
p = 30																		
L1	1.86	1.27	1.04	1.52	10.48	13.66	25.16	30.8	5.98	5.21	0.97	1	5.89	5.65	58.07	58	2.72	2.06
L2	1.98	1.35	1.93	3.82	13.11	17.03	26.02	32	3.77	2.87	2.21	3.67	8.39	8.2	15.26	15.23	4.96	7.65
EL1	0.16	0.12	0.17	0.58	0.77	2.4	1.75	4.67	0.22	0.31	0.22	0.59	0.72	0.74	0.71	0.72	0.15	1.25
EL2	0.04	0.06	0.04	0.43	0.75	7.36	3.93	28.49	0.1	0.27	0.07	0.42	2.1	2.21	7.29	7.73	0.04	1.92
MAXEL1	0.56	1.07	0.22	1.2	1.04	3.41	2.84	8.34	1.05	1.62	0.49	0.81	5.41	5.55	10.45	10.77	0.64	1.93
MINEL1	0.38	0.14	0.16	0.53	1.14	3.4	3.03	8.23	0.85	1.62	0.49	0.81	0.19	0.35	10.45	10.77	0.64	1.93
p = 100																		
L1	2.11	1.33	1.02	1.35	44.18	46.52	91.89	96.26	7.04	7.03	0.97	1	19.88	19.61	198.16	198.32	2.81	2.03
L2	3.36	2.62	3.1	7.07	58.3	62.13	103.75	108.41	5.56	3.88	3.87	7.04	28.2	28	28.17	28.19	9.17	14.09
EL1	0.14	0.13	0.15	0.63	3.37	5.17	5.04	9.08	0.11	0.23	0.27	0.63	0.78	0.78	0.4	0.4	0.18	1.27
EL2	0.03	0.07	0.03	0.5	14.36	36.49	33.86	114.22	0.03	0.15	0.1	0.49	7.69	7.83	7.78	7.83	0.06	1.98
MAXEL1	0.46	1.29	0.07	1.35	6.22	9.97	10.41	18.07	0.1	1.77	0.53	1	19.4	19.58	19.72	19.79	0.83	1.98
MINEL1	0.48	0.17	0.31	0.67	6.18	10.04	9.86	17.81	0.1	1.77	0.53	1	0.2	0.38	19.72	19.79	0.83	1.98

Summary of L1, L2, EL1, EL2, MAXEL1 and MINEL1 for an AR(1), AR(2), sparse random, scale-free, band, cluster, star and circle graphical model. The median loss values reported here are based on 40 replications for both the B-net and D-net estimators. The best performing values are boldfaced.

<https://doi.org/10.1371/journal.pone.0261193.t002>

Table 3. Synthetic study standard error loss results.

	AR(1)		AR(2)		S80		S40		SF		Band		Cluster		Star		Circle	
	B-net	D-net	B-net	D-net	B-net	D-net	B-net	D-net	B-net	D-net	B-net	D-net	B-net	D-net	B-net	D-net	B-net	D-net
p = 10																		
L1	0.19	0.05	0.14	0.09	0.72	1.16	1.37	2.42	0.16	0.01	0.29	0.19	0.33	0.12	0.79	0.12	0.69	0.12
L2	0.15	0.03	0.15	0.04	0.78	1.02	1.07	2.11	0.09	0.01	0.37	0.15	0.37	0.05	0.42	0.06	0.89	0.01
EL1	0.02	0.01	0.07	0.04	0.12	0.24	0.19	0.77	0.02	0.01	0.09	0.1	0.08	0.03	0.06	0.02	0.14	0.03
EL2	0.01	0.01	0.05	0.05	0.11	0.41	0.39	3.51	0.01	0.01	0.06	0.1	0.09	0.04	0.8	0.29	0.17	0.1
MAXEL1	0.11	0.01	0.25	0.08	0.27	0.5	0.57	1.28	0.08	0.01	0.17	0.18	0.3	0.08	0.4	0.12	0.48	0.1
MINEL1	0.13	0.04	0.1	0.08	0.26	0.57	0.64	1.46	0.08	0.01	0.17	0.22	0.12	0.09	0.38	0.12	0.5	0.1
p = 30																		
L1	0.22	0.01	0.14	0.11	1.45	1.25	1.63	2.28	0.11	0.01	0.16	0.04	0.08	0.04	0.22	0.03	0.5	0.05
L2	0.2	0.01	0.14	0.01	1.03	1.23	1.07	1.23	0.05	0.01	0.27	0.2	0.04	0.01	0.06	0.01	1.07	0.01
EL1	0.02	0.01	0.03	0.02	0.2	0.25	0.18	0.3	0.01	0.01	0.04	0.04	0.01	0.01	0.01	0.01	0.05	0.01
EL2	0.01	0.01	0.01	0.03	0.39	1.58	0.86	4.6	0.01	0.01	0.03	0.07	0.03	0.02	0.15	0.05	0.05	0.03
MAXEL1	0.15	0.02	0.12	0.06	0.61	1.31	0.76	1.58	0.06	0.01	0.1	0.11	0.04	0.03	0.11	0.04	0.28	0.04
MINEL1	0.16	0.02	0.07	0.05	0.61	1.37	0.83	1.6	0.03	0.01	0.11	0.12	0.05	0.03	0.11	0.04	0.28	0.04
p = 100																		
L1	0.24	0.01	0.12	0.02	2.26	2.85	2.49	4.02	0.1	0.01	0.01	0.01	0.03	0.01	0.07	0.36	0.19	0.01
L2	0.2	0.01	0.05	0.01	1.17	1.1	1.85	1.71	0.01	0.01	0.02	0.01	0.02	0.01	0.01	0.05	1.31	0.01
EL1	0.01	0.01	0.01	0.01	0.17	0.11	0.22	0.16	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.04	0.01
EL2	0.01	0.01	0.01	0.01	1.45	1.56	2.98	3.84	0.01	0.01	0.01	0.01	0.01	0.01	0.06	0.08	0.01	0.01
MAXEL1	0.19	0.01	0.07	0.01	0.75	0.89	1.08	1.32	0.03	0.01	0.03	0.03	0.01	0.01	0.07	0.1	0.14	0.01
MINEL1	0.12	0.01	0.05	0.01	0.6	0.89	0.89	1.52	0.03	0.01	0.03	0.03	0.01	0.01	0.07	0.1	0.14	0.01

Summary of L1, L2, EL1, EL2, MAXEL1 and MINEL1 for an AR(1), AR(2), sparse random, scale-free, band, cluster, star and circle graphical model. The median standard errors reported here are based on 40 replications for both the B-net and D-net estimators. The best performing values are boldfaced.

<https://doi.org/10.1371/journal.pone.0261193.t003>

This may be due to the fact that the best performing tuning parameter in the solution path leads to highly sparse estimates. The B-net estimation procedure inherits the utilisation of multiple penalty parameters in the precision matrix estimation, leading to robust estimation of the precision matrices.

To assess the performance on graphical structure determination, the specificity, sensitivity, false negative rate, f1 score and the MCCs are computed and defined in Table 4. Noting that, TP, TN, FP and FN denote the number of true positives, true negatives, false positives and false negatives, respectively. Values of specificity, sensitivity, f1-score and MCC closer to one, imply better classification performance. The closer the values of false negative rate are to zero the better. Further insights on the performance metrics are discussed in [40]. The sparsity for

Table 4. Performance measures used to assess classification accuracy of the B-net and D-net graphical models estimates.

Measure	Performance function	Abbreviation
Specificity	$\frac{TN}{TN+FP}$	SP
Sensitivity	$\frac{TP}{TP+FN}$	SE
False negative rate	$\frac{FP}{FP+TN}$	FNR
F1-score	$\frac{TP}{TP+\frac{1}{2}(FP+FN)}$	F1
Matthews Correlation Coefficient	$\frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$	MCC

<https://doi.org/10.1371/journal.pone.0261193.t004>

Table 5. Synthetic study median performance results.

	AR(1)		AR(2)		S80		S40		SF		Band		Cluster		Star		Circle	
	B-net	D-net	B-net	D-net	B-net	D-net	B-net	D-net	B-net	D-net	B-net	D-net	B-net	D-net	B-net	D-net	B-net	D-net
p = 10																		
SE	0.72	0.11	0.82	0.21	0.6	0.53	0.65	0.36	0.89	NA	0.89	0.44	1	0.25	0.22	0.11	0.9	0
SP	1	1	0.97	0.7	0.89	0.82	0.87	0.94	0.9	NA	0.95	0.93	1	0.87	0.8	1	0.95	0.98
PR	0.87	0.5	0.3	0.12	0.22	0.22	0.72	0.56	0.18	NA	0.16	0.11	0.4	0.16	0.06	0.02	0.2	0
MC	0.45	0.11	0.76	-0.07	0.54	0.39	0.43	0.24	0.71	NA	0.83	0.41	1	0.14	0.11	0.3	0.88	-0.07
F1	0.84	0.2	0.84	0.24	0.67	0.56	0.77	0.51	0.76	NA	0.86	0.4	1	0.34	0.25	0.2	0.9	0
FNR	0.28	0.89	0.18	0.79	0.4	0.47	0.35	0.64	0.11	NA	0.11	0.56	0	0.75	0.78	0.89	0.1	1
p = 30																		
SE	0.31	0.01	0.75	0.04	0.23	0.08	0.37	0.03	0.79	NA	0.9	0.24	0.89	0.02	0.03	0.02	0.97	0
SP	0.97	1	0.99	0.93	0.98	0.99	0.82	1	0.92	NA	1	1	0.99	0.98	1	1	1	0.99
MC	0.17	0.02	0.8	-0.04	0.35	0.16	0.16	0.05	0.52	NA	0.92	0.46	0.89	-0.02	0.06	0.08	0.95	-0.03
F1	0.48	0.01	0.82	0.04	0.36	0.14	0.53	0.06	0.53	NA	0.93	0.38	0.93	0.04	0.06	0.03	0.95	0
FNR	0.69	0.99	0.25	0.96	0.77	0.92	0.63	0.97	0.21	NA	0.1	0.76	0.11	0.98	0.97	0.98	0.03	1
p = 100																		
SE	0.21	0	0.73	0	0.04	0.01	0.15	0	0.49	NA	0.95	0.05	0.79	0	0.02	0.12	0.99	0
SP	0.99	1	1	1	1	1	0.92	1	0.98	NA	1	1	1	1	1	1	1	1
MC	0.33	0.01	0.82	0	0.14	0.05	0.07	0.01	0.39	NA	0.95	0.22	0.81	0	0.08	0.35	0.98	-0.01
F1	0.34	0	0.82	0	0.08	0.01	0.25	0	0.4	NA	0.95	0.1	0.88	0	0.04	0.22	0.98	0
FNR	0.79	1	0.27	1	0.96	0.99	0.85	1	0.51	NA	0.05	0.95	0.21	1	0.98	0.88	0.01	1

Summary of SE, SP, F1, MC and for an AR(1), AR(2), sparse random, scale-free, band, cluster, star and circle graphical model. The median performance scores reported here are based on 40 replications for both the B-net and D-net estimators. The best performing values are bold-faced and scores that were not attainable due to single class classification are encoded as NA.

<https://doi.org/10.1371/journal.pone.0261193.t005>

the B-net estimator is determined by the thresholding rule in Eq (8) and the thresholds, η , associated with the MCC values in Fig (1a)–(1i). Similarly, the best performing tuning parameter in the solution path of the D-net algorithm determines the sparsity of the estimator. The median performance scores, based on 40 repetitions, for each graphical structure is presented in Table 5. The main diagonals of the adjacency matrices were not included in the scoring.

The B-net estimator generally outperforms the D-net estimator across all models and all dimensions according to the MCC, f1-score, sensitivity and false negative rate, with the exception of the star case for $p = 100$. Fig (3a)–(3i) display the true and inferred undirected DN graphs for both the B-net and D-net estimators for $p = 10$; higher dimensions are available in the S1 File. Lastly, Fig (2a)–(2i) display the true and inferred adjacency matrices for $p = 10$. Both Figs 2 and 3 visually demonstrate the superiority of the B-net estimator.

Real data analysis

This section focuses on applying the novel Bayesian DN estimator, B-net, as well as the iterative shrinkage-thresholding estimator, D-net, to the spambase dataset, available at <https://archive.ics.uci.edu/ml/datasets/spambase> to investigate changes in DN structure between spam and non-spam data. In addition, the B-net estimator is applied to South African COVID-19 data, obtained from <https://www.nicd.ac.za/diseases-a-z-index/disease-index-covid-19/surveillance-reports/>, <https://ourworldindata.org/coronavirus/country/south-africa> and <https://mediahack.co.za/datastories/coronavirus> to investigate the change in DN structure between various phases of the pandemic.

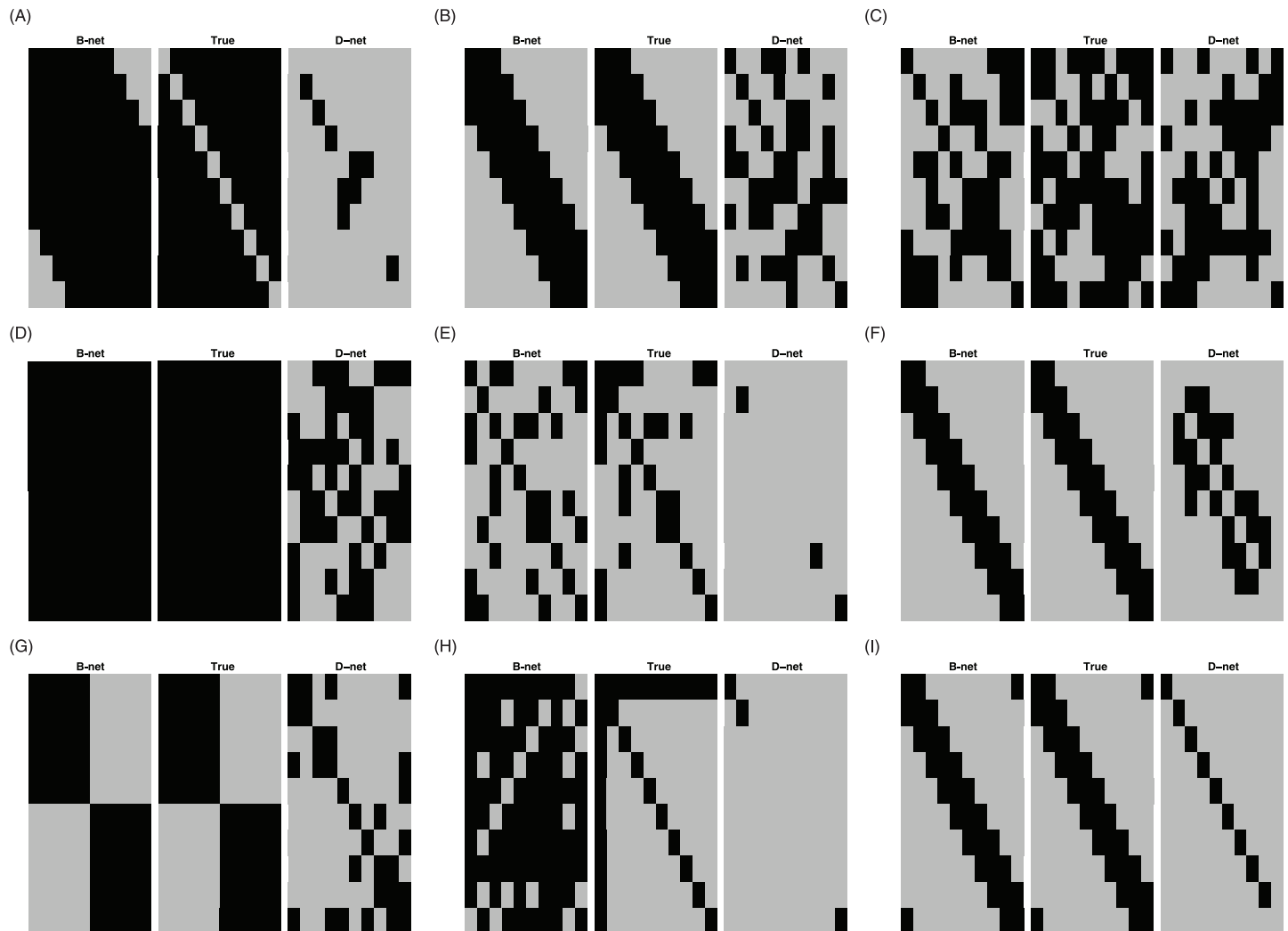


Fig 2. DN adjacency matrix heatmaps. Comparison of the true DN, B-net and D-net adjacency matrices for an AR(1), AR(2), sparse random, scale-free, band, cluster, star and circle graphical model and $p = 10$. (a) Model 1: AR(1). (b) Model 2: AR(2). (c) Model 3: at most 80% sparse. (d) Model 4: at most 40% sparse. (e) Model 5: scale-free. (f) Model 6: band. (g) Model 7: cluster. (h) Model 8: star. (i) Model 9: circle.

<https://doi.org/10.1371/journal.pone.0261193.g002>

Spam data

The objective here is to compare the B-net and D-net graphical model estimates of the spam and non-spam emails. The dataset consists of 1813 spam emails and 2788 non-spam emails. The attributes include, amongst others, the average length of uninterrupted sequences of capital letters; total number of capital letters in the e-mail; an indicator denoting whether the e-mail was considered spam or not, in this study.

Following the approach of [24], the data is standardised using a non-paranormal transformation in order to satisfy the Gaussian assumption. The B-net estimates are based on 10000 iterations of the Monte Carlo sampler after 5000 burn-in iterations. Fig 4 illustrates the difference between the B-net and D-net estimates. Both estimators indicate the presence of several common hub features namely, 'edu', 'original', 'direct', 'lab', 'telnet' and 'addresses'. It is clear from both panes that the state of the covariance matrix structure between the spam and non-spam emails may very well be different. Furthermore, given that Hewlett-Packard Labs

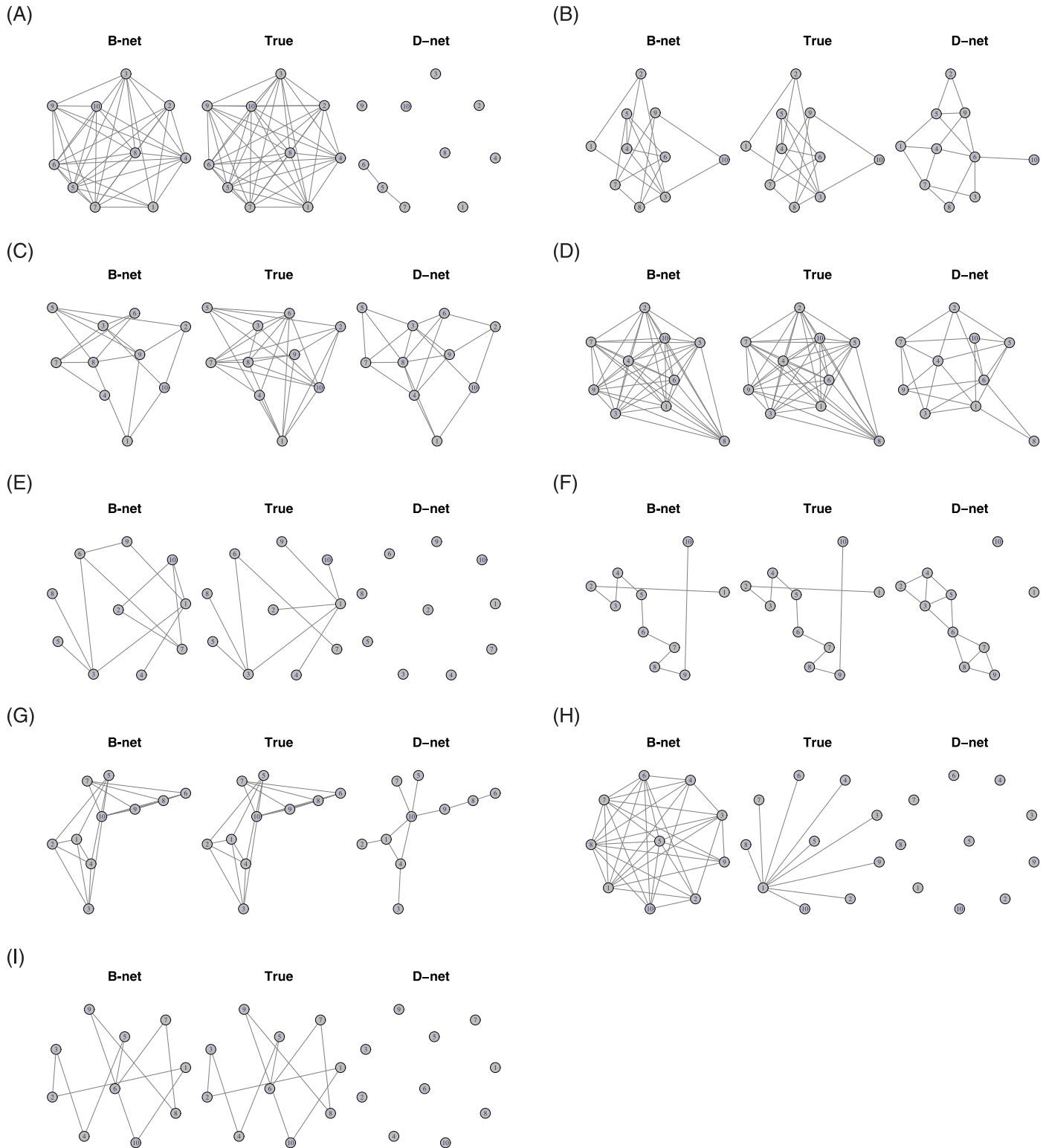


Fig 3. DN adjacency matrix graphical models. Comparison of the true DN, B-net and D-net graphical structure estimates for an AR(1), AR(2), sparse random, scale-free, band, cluster, star and circle graphical model and $p = 10$. (a) Model 1: AR(1). (b) Model 2: AR(2). (c) Model 3: at most 80% sparse. (d) Model 4: at most 40% sparse. (e) Model 5: scale-free. (f) Model 6: band. (g) Model 7: cluster. (h) Model 8: star. (i) Model 9: circle.

<https://doi.org/10.1371/journal.pone.0261193.g003>

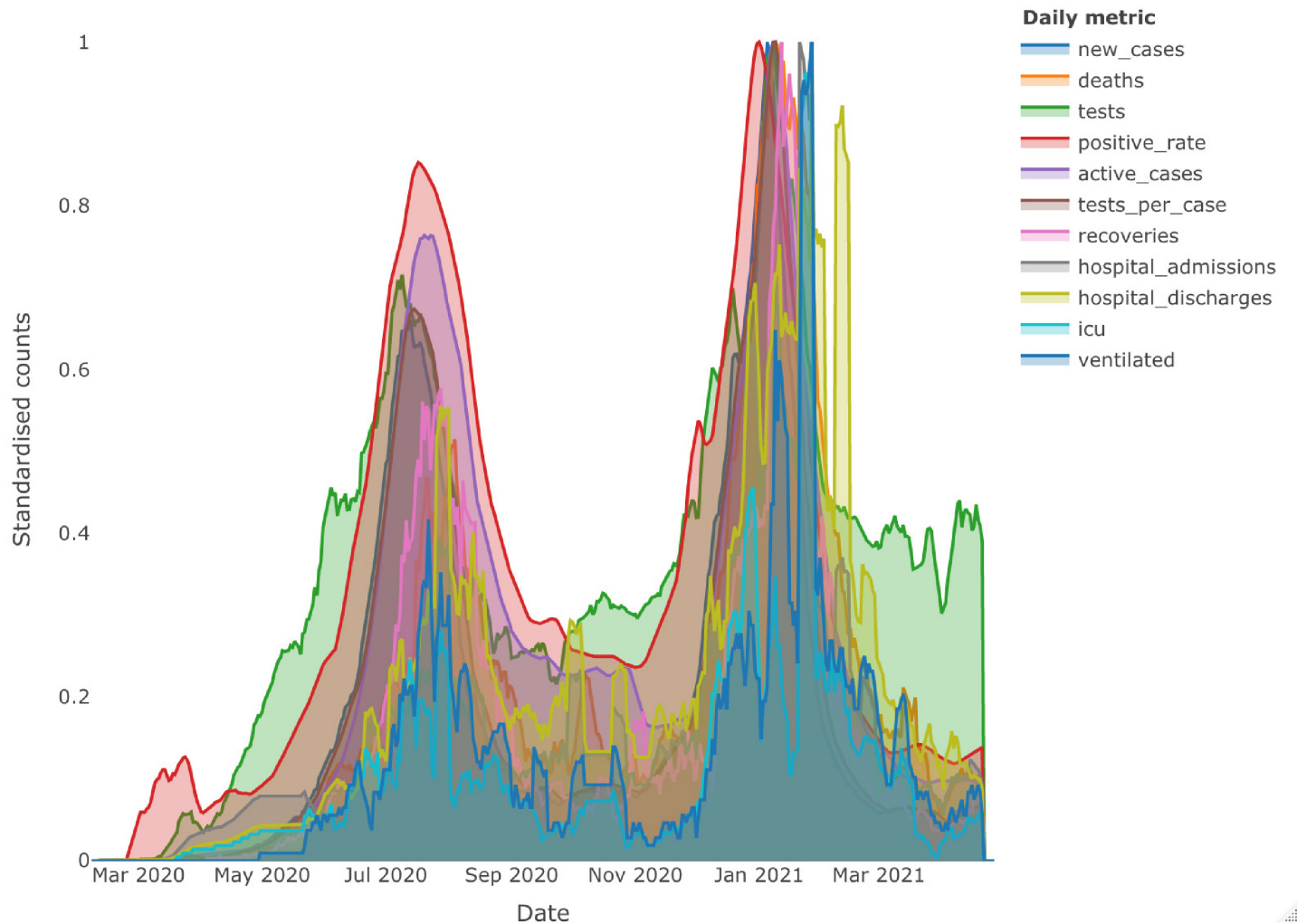


Fig 5. South African COVID-19 daily metrics over time. 7-day moving average filled area line plots with standardised counts for daily new cases; deaths; tests; positive test rate; active cases; tests per active case; recoveries; hospital admissions; hospital discharges; ICU admissions and ventilated patients.

<https://doi.org/10.1371/journal.pone.0261193.g005>

metrics simultaneously, the crisp group-wise multi modality diminishes rather rapidly. The figures in Fig 6 illustrate the higher proportions of hub features present in the DNs. Interestingly, the Bayesian DN provides insight to the change in interaction between daily metrics between perceived homogeneous pandemic phases, that is comparisons between the two peaks and two post-peak plateaus. This change in behaviour could be as a result of the change in population adherence to public sanitation awareness; weather conditions; virus mutations or complacency of over time.

Discussion

The Bayesian differential network estimator is the first of its kind which utilises the excellent graphical structure determination and matrix estimation of the Bayesian graphical lasso [16]. In comparison with the state of the art iterative shrinkage-thresholding approach, the Bayesian differential network offers MCMC outputs that allow the user to gain deeper insight and inference in the estimation procedure. The numerical accuracy of the Bayesian differential network

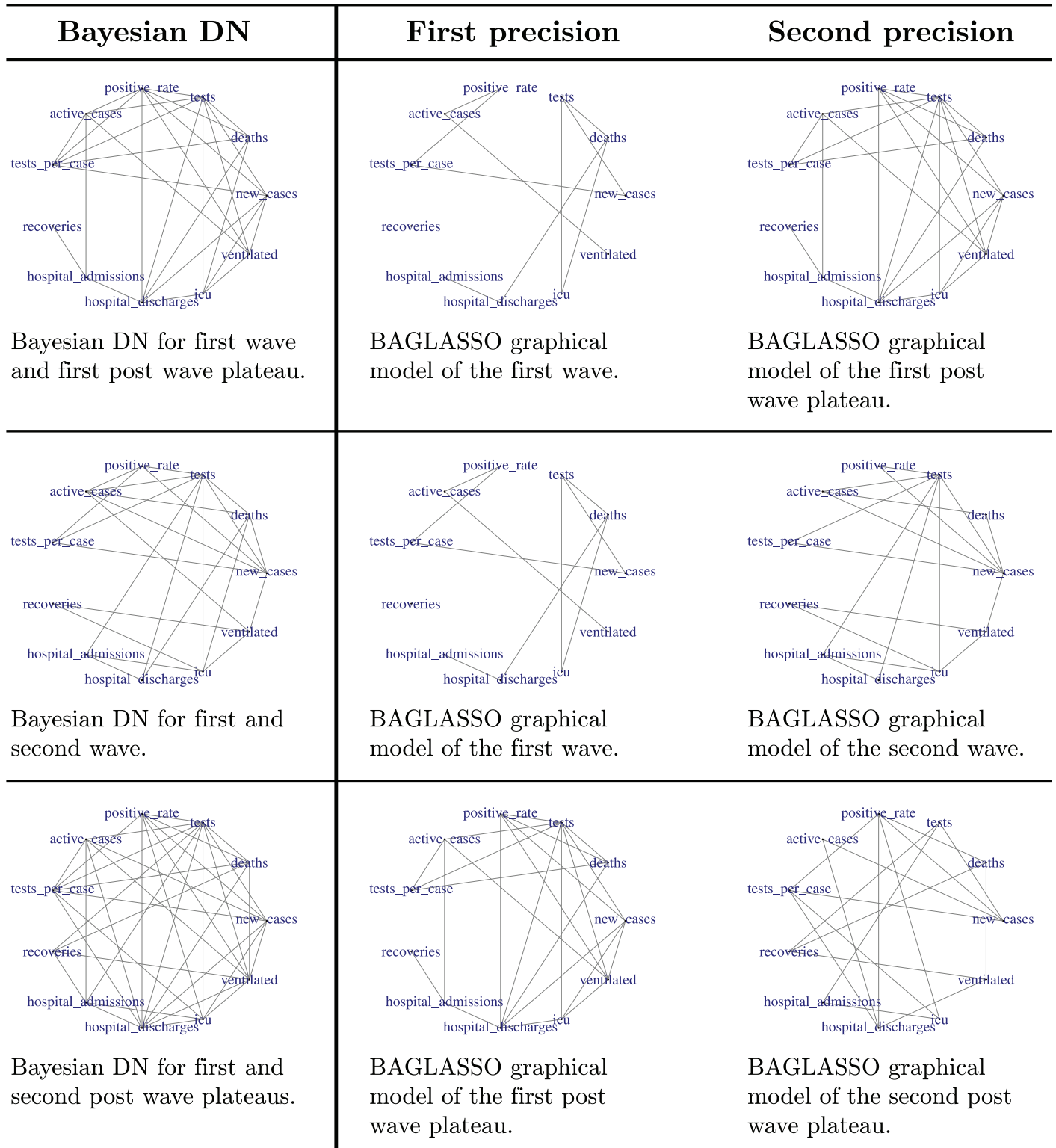


Fig 6. Bayesian DN estimates of South African COVID-19 data. The Bayesian DN and corresponding BAGLASSO graphical models between the first two waves; the first wave and the following plateau and finally the difference between the first and second post wave plateaus. The *p*-values from the Box's *M*-test for homogeneity of covariance matrices between the contributing precision matrices were all less than 0.001 [43].

<https://doi.org/10.1371/journal.pone.0261193.g006>

is, in general, superior to that of the iterative shrinkage-thresholding estimator. Moreover, the Bayesian proposal captures both sparse and dense precision matrix patterns in some well-known graphical structures more accurately. The latter being a result of the Wishart prior's ability to accommodate the variability and adjustment to the data. Furthermore, the thresholding technique for sparse estimation is designed such that it accounts for the effect of prior allocation through the posterior expectation.

The graphical structure learning is a crucial component of the Bayesian differential network estimator. The ad hoc approach provided in Eq (8) suggests a suitable sparsity threshold under varying graph structures. The Bayesian differential network also provides key insights to changes in the interactive behaviour of real data metrics ranging from filtering spam emails to COVID-19 life cycles. For high-dimensional data, the block Gibbs sampler may be adjusted to incorporate the singular normal distribution presented in [44] in the hierarchical representation Eq (7). Furthermore, research on simultaneous Bayesian estimation and optimisation of both Σ_1^{-1} and Σ_2^{-1} in the construction of the differential network is underway.

Supporting information

S1 File. Supplementary material. Contains a block Gibbs sampler, as well as, additional optimal threshold; adjacency heatmaps and graphical network figures for dimensions $p = 30$ and $p = 100$.
(PDF)

Acknowledgments

We would like to sincerely thank both anonymous reviewers for their generous comments on the manuscript. The astute feedback was most welcomed, insightful and above all greatly improved the presentation, scientific justification and readability of the paper.

Author Contributions

Conceptualization: Jarod Smith, Mohammad Arashi, Andriëtte Bekker.

Data curation: Jarod Smith.

Formal analysis: Jarod Smith, Mohammad Arashi, Andriëtte Bekker.

Funding acquisition: Mohammad Arashi.

Investigation: Jarod Smith, Mohammad Arashi, Andriëtte Bekker.

Methodology: Jarod Smith, Mohammad Arashi, Andriëtte Bekker.

Project administration: Mohammad Arashi, Andriëtte Bekker.

Resources: Mohammad Arashi, Andriëtte Bekker.

Software: Jarod Smith.

Supervision: Mohammad Arashi, Andriëtte Bekker.

Validation: Jarod Smith, Mohammad Arashi, Andriëtte Bekker.

Visualization: Jarod Smith.

Writing – original draft: Jarod Smith.

Writing – review & editing: Jarod Smith, Mohammad Arashi, Andriëtte Bekker.

References

1. Shojaie A. Differential network analysis: A statistical perspective. *Wiley Interdisciplinary Reviews: Computational Statistics*. 2020; p. e1508.
2. Koller D, Friedman N. *Probabilistic Graphical Models: Principles and Techniques*. MIT press; 2009.
3. Chuang H, Lee E, Liu Y, Lee D, Ideker T. Network-based classification of breast cancer metastasis. *Molecular Systems Biology*. 2007; 3(1):140. <https://doi.org/10.1038/msb4100180> PMID: 17940530
4. Taylor I, Linding R, Warder-Farley D, Liu Y, Pesquita C, Faria D, et al. Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nature Biotechnology*. 2009; 27(2):199–204. <https://doi.org/10.1038/nbt.1522> PMID: 19182785
5. Li Q, Shao J. Sparse quadratic discriminant analysis for high dimensional data. *Statistica Sinica*. 2015; 25:457–473.
6. Jiang B, Wang X, Leng C. A direct approach for sparse quadratic discriminant analysis. *The Journal of Machine Learning Research*. 2018; 19(1):1098–1134.
7. Meinshausen N, Bühlmann P. High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*. 2006; 34(3):1436–1462. <https://doi.org/10.1214/009053606000000281>
8. Lauritzen S. *Graphical models*. Oxford: Clarendon Press; 1996.
9. Yuan M, Lin Y. Model selection and estimation in the Gaussian graphical model. *Biometrika*. 2007; 94(1):19–35. <https://doi.org/10.1093/biomet/asm018>
10. Friedman J, Hastie T, Tibshirani R. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*. 2008; 9(3):432–441. <https://doi.org/10.1093/biostatistics/kxm045> PMID: 18079126
11. Banerjee O, Ghaoui LE, d'Aspremont A. Model Selection Through Sparse Maximum Likelihood Estimation for Multivariate Gaussian or Binary Data. *The Journal of Machine Learning Research*. 2008; 9:485–516.
12. Nesterov Y. Smooth minimization of non-smooth functions. *Mathematical Programming*. 2005; 103(1):127–152. <https://doi.org/10.1007/s10107-004-0552-5>
13. Cai T, Liu W, Luo X. A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*. 2011; 106(494):594–607. <https://doi.org/10.1198/jasa.2011.tm10155>
14. Guo J, Levina E, Michailidis G, Zhu J. Joint estimation of multiple graphical models. *Biometrika*. 2011; 98(1):1–15. <https://doi.org/10.1093/biomet/asq060> PMID: 23049124
15. Danaher P, Wang P, Witten D. The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society: Series B (Methodological)*. 2014; 76(2):373–397. <https://doi.org/10.1111/rssb.12033> PMID: 24817823
16. Wang H. Bayesian graphical lasso models and efficient posterior computation. *Bayesian Analysis*. 2012; 7(4):867–886. <https://doi.org/10.1214/12-BA729>
17. Banerjee S, Ghosal S. Bayesian structure learning in graphical models. *Journal of Multivariate Analysis*. 2015; 136:147–162. <https://doi.org/10.1016/j.jmva.2015.01.015>
18. Peterson C, Stingo F, Vannucci M. Bayesian inference of multiple Gaussian graphical models. *Journal of the American Statistical Association*. 2015; 110(509):159–174. <https://doi.org/10.1080/01621459.2014.896806> PMID: 26078481
19. Williams D, Piironen J, Vehtari A, Rast P. Bayesian estimation of Gaussian graphical models with predictive covariance selection. *arXiv preprint arXiv:180105725*. 2018;.
20. Chiquet J, Grandvalet Y, Ambroise C. Inferring multiple graphical structures. *Statistics and Computing*. 2011; 21(4):537–553. <https://doi.org/10.1007/s11222-010-9191-2>
21. Zhu Y, Li L. Multiple matrix gaussian graphs estimation. *Journal of the Royal Statistical Society: Series B (Methodological)*. 2018; 80(5):927–950. <https://doi.org/10.1111/rssb.12278> PMID: 30505211
22. Zhao S, Cai T, Li H. Direct estimation of differential networks. *Biometrika*. 2014; 101(2):253–268. <https://doi.org/10.1093/biomet/asu009> PMID: 26023240
23. Yuan H, Xi R, Chen C, Deng M. Differential network analysis via lasso penalized D-trace loss. *Biometrika*. 2017; 104(4):755–770. <https://doi.org/10.1093/biomet/asx049>
24. Tang Z, Yu Z, Wang C. A fast iterative algorithm for high-dimensional differential network. *Computational Statistics*. 2020; 35(1):95–109. <https://doi.org/10.1007/s00180-019-00915-w>
25. Beck A, Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*. 2009; 2(1):183–202. <https://doi.org/10.1137/080716542>
26. Marlin B, Schmidt M, Murphy K. Group sparse priors for covariance estimation. In: *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*; 2009.

27. Park T, Casella G. The bayesian lasso. *Journal of the American Statistical Association*. 2008; 103(482):681–686. <https://doi.org/10.1198/016214508000000337>
28. Andrews DF, Mallows CL. Scale mixtures of normal distributions. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1974; 36(1):99–102.
29. West M. On scale mixtures of normal distributions. *Biometrika*. 1987; 74(3):646–648. <https://doi.org/10.1093/biomet/74.3.646>
30. Li Q, Lin N. The Bayesian elastic net. *Bayesian Analysis*. 2010; 5(1):151–170. <https://doi.org/10.1214/10-BA506>
31. Griffin JE, Philip J. Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis*. 2010; 5(1):171–188.
32. Carvalho CM, Polson NG, Scott JG. The horseshoe estimator for sparse signals. *Biometrika*. 2010; 97(2):465–480. <https://doi.org/10.1093/biomet/asq017>
33. Fan J, Feng Y, Wu Y. Wishart distributions for decomposable graphs. *The Annals of Applied Statistics*. 2009; 3(2):521–541. <https://doi.org/10.1214/08-AOAS215SUPP> PMID: 21643444
34. Dempster A. Covariance selection. *Biometrics*. 1972; p. 157–175.
35. Mazumder R, Hastie T. The graphical lasso: New insights and alternatives. *Electronic Journal of Statistics*. 2012; 6:2125–2149. <https://doi.org/10.1214/12-EJS740> PMID: 25558297
36. Wang J, Jiang B. An efficient ADMM algorithm for high dimensional precision matrix estimation via penalized quadratic loss. *Computational Statistics & Data Analysis*. 2020; 142(2):106812. <https://doi.org/10.1016/j.csda.2019.106812>
37. Letac G, Massam H. Wishart distributions for decomposable graphs. *The Annals of Statistics*. 2007; 35(3):1278–1323. <https://doi.org/10.1214/009053606000001235>
38. Matthews BW. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA)-Protein Structure*. 1975; 405(2):442–451. [https://doi.org/10.1016/0005-2795\(75\)90109-9](https://doi.org/10.1016/0005-2795(75)90109-9) PMID: 1180967
39. Banerjee S, Monni S, Wells M. A regularized profile likelihood approach to covariance matrix estimation. *Journal of Statistical Planning and Inference*. 2013; 179:36–59. <https://doi.org/10.1016/j.jspi.2016.06.004>
40. Iwendi C, Khan S, Anajemba JH, Mittal M, Alenezi M, Alazab M. The use of ensemble models for multiple class and binary class classification for improving intrusion detection systems. *Sensors*. 2020; 20(9):2559. <https://doi.org/10.3390/s20092559> PMID: 32365937
41. Iwendi C, Bashir AK, Peshkar A, Sujatha R, Chatterjee JM, Pasupuleti S, et al. COVID-19 patient health prediction using boosted random forest algorithm. *Frontiers in Public Health*. 2020; 8(357). <https://doi.org/10.3389/fpubh.2020.00357> PMID: 32719767
42. Iwendi C, Moqurrab SA, Anjum A, Khan S, Mohan S, Srivastava G. N-sanitization: A semantic privacy-preserving framework for unstructured medical datasets. *Computer Communications*. 2020; 161:160–171. <https://doi.org/10.1016/j.comcom.2020.07.032>
43. Box GE. A general distribution theory for a class of likelihood criteria. *Biometrika*. 1949; 36(3/4):317–346. <https://doi.org/10.2307/2332671> PMID: 15402070
44. Bland RP, Owen DB. A note on singular normal distributions. *Annals of the Institute of Statistical Mathematics*. 1966; 18(1):113–116. <https://doi.org/10.1007/BF02869521>