PSYCHOMETRIKA—VOL. 89, NO. 4, 1186–1202
DECEMBER 2024
https://doi.org/10.1007/s11336-024-09985-2

A NOTE ON ISING NETWORK ANALYSIS WITH MISSING DATA

SILIANG ZHANG
EAST CHINA NORMAL UNIVERSITY
YUNXIAO CHEN
LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

The Ising model has become a popular psychometric model for analyzing item response data. The statistical inference of the Ising model is typically carried out via a pseudo-likelihood, as the standard likelihood approach suffers from a high computational cost when there are many variables (i.e., items). Unfortunately, the presence of missing values can hinder the use of pseudo-likelihood, and a listwise deletion approach for missing data treatment may introduce a substantial bias into the estimation and sometimes yield misleading interpretations. This paper proposes a conditional Bayesian framework for Ising network analysis with missing data, which integrates a pseudo-likelihood approach with iterative data imputation. An asymptotic theory is established for the method. Furthermore, a computationally efficient Pólya–Gamma data augmentation procedure is proposed to streamline the sampling of model parameters. The method's performance is shown through simulations and a real-world application to data on major depressive and generalized anxiety disorders from the National Epidemiological Survey on Alcohol and Related Conditions (NESARC).

Key words: Ising model, iterative imputation, full conditional specification, network psychometrics, mental health disorders, major depressive disorder, generalized anxiety disorder.

Recent years have witnessed the emergence of network psychometrics (Borsboom, 2008; Marsman & Rhemtulla, 2022; van der Maas et al., 2006) , a family of statistical graphical models and related inference procedures, for analyzing and interpreting the dependence structure in psychometric data. These models embed psychometric items as nodes in an undirected or directed network (i.e., graph) and visualize their interrelationships through the network edges, which represent certain probabilistic conditional dependencies. Network psychometric methods concern the learning of the network structure. They have been developed under various settings, including undirected graphical models for cross-sectional data (Burger et al., 2022; Epskamp et al., 2018) , directed networks for longitudinal data (Borsboom et al., 2021; Gile & Handcock, 2017; Ryan et al., 2022) , and extended networks with latent variables for time series data or panel data (Epskamp, 2020) . These methods have received wide applications in education (Koponen et al., 2019; Siew, 2020; Simon de Blas et al., 2021; Sweet et al., 2013; Willcox & Huang, 2017) , psychology (Borsboom et al., 2021; Burgess & Hitch, 1999; Epskamp et al., 2018; Fried et al., 2017; Van Der Maas et al., 2017) , and health sciences (Brunson & Laubenbacher, 2018; Kohler et al., 2022; Luke & Harris, 2007; Mkhitaryan et al., 2019) .

Analyzing cross-sectional binary item response data with the Ising model (Ising, 1925) is common in network psychometric analysis. This analysis is typically performed based on a conditional likelihood (Besag, 1974) because the standard likelihood function is computationally infeasible when involving many variables. In this direction, Bayesian and frequentist methods have been developed, where sparsity-inducing priors or penalties are combined with the conditional likelihood for learning a sparse network structure (Epskamp & Fried, 2018; Li et al., 2019;

Correspondence should be made to Yunxiao Chen, Department of Statistics, London School of Economics and Political Science, Room 5.16 Columbia House, Houghton Street, London WC2A 2AE, UK. Email: y.chen186@lse.ac.uk

© 2024 The Author(s)

https://doi.org/10.1007/s11336-024-09985-2 Published online by Cambridge University Press

Marsman et al., 2022; Mazumder & Hastie, 2012; Van Borkulo et al., 2014; Yuan & Lin, 2007) . Besides, the Ising model is shown to be closely related to item response theory (IRT) models (Anderson & Yu, 2007; Holland, 1990) . The log-multiplicative association models (Anderson & Yu, 2007) , which are special cases of the Ising model, can be used as item response theory models and yield very similar results as IRT models. Furthermore, the Ising model and the conditional likelihood have been used for modeling the local dependence structure in locally dependent IRT models (Chen et al., 2018; Ip, 2002) .

Due to its construction, the conditional likelihood does not naturally handle data with missing values, despite the omnipresence of missing data in psychometric data. To deal with missing values in an Ising network analysis, listwise deletion (Fried et al., 2020; Haslbeck & Fried, 2017) and single imputation (e.g., Huisman, 2009, Armour et al., 2017, Lin et al., 2020) are typically performed, which arguably may not be the best practice. In particular, it is well known that listwise deletion is statistically inefficient and requires the missing completely at random (MCAR) assumption (Little & Rubin, 2019) to ensure consistent estimation. Moreover, a naïve imputation procedure, such as mode imputation, likely introduces bias into parameter estimation. A sophisticated imputation procedure must be developed to ensure statistical validity and computational efficiency.

In this note, we propose an iterative procedure for learning an Ising network. The proposed procedure combines iterative imputation via full conditional specification (FCS; Liu et al., 2014, van Buuren, 2018), and Bayesian estimation of the Ising network. We show that the FCS leads to estimation consistency when the conditional models are chosen to take logistic forms. In terms of computation, we propose a joint Pólya–Gamma augmentation procedure by extending the Pólya–Gamma augmentation procedure for logistic regression (Polson et al., 2013) . It allows us to efficiently sample parameters of the Ising model. Simulations are conducted to compare the proposed procedure with estimations based on the listwise deletion and single imputation. Finally, the proposed procedure and a complete-case analysis are applied to study the network of major depressive disorder (MDD) and generalized anxiety disorders (GAD) based on data from the National Epidemiological Survey on Alcohol and Related Conditions (NESARC; Grant et al., 2003). In this analysis, data missingness is mainly due to two screening items for GAD. That is, a respondent's responses to the rest of the MDD items are missing if they answered "no" to both screening items. This missing mechanism is Missing at Random (MAR; Little & Rubin, 2019). The complete-case analysis of missing data caused by screening items is known to be problematic in the literature of network psychometrics (Borsboom et al., 2017; McBride et al., 2023) . Our Bayesian estimate of the edge coefficient between the two screening items is negative based on the complete cases, which can be seen as a result of Berkson's paradox (De Ron et al., 2021) . In contrast, the proposed method makes use of all the observed data entries and obtains a positive estimate of this edge coefficient. An identifiability result about the Ising model under this special missing data setting in Appendix, the item content, and a simulation study mimicking this setting suggest that the result given by the proposed method is sensible. A GitHub repository for the proposed algorithm is available at https://github.com/slzhang-fd/IsingNetMissing-replication.

1. Proposed Method

1.1. Ising Model

Consider a respondent answering J binary items. Let $\mathbf{Y} = (Y_1, \ldots, Y_J)^\top \in \{0, 1\}^J$ be a binary random vector representing the respondent's responses. We say $\mathbf{Y}$ follows an Ising model if its probability mass function satisfies

https://doi.org/10.1007/s11336-024-09985-2 Published online by Cambridge University Press

$$P(\mathbf{Y}=\boldsymbol{y}\mid\mathbf{S})=\frac{1}{c(\mathbf{S})}\exp\left[\frac{1}{2}\boldsymbol{y}^{\top}\mathbf{S}\boldsymbol{y}\right]=\frac{1}{c(\mathbf{S})}\exp\left[\sum_{j=1}^{J}s_{jj}y_j/2+\sum_{j=1}^{J-1}\sum_{k=j+1}^{J}s_{jk}y_jy_k\right],\quad(1)$$

where $\mathbf{S}=(s_{ij})_{J\times J}$ is a J by J symmetric matrix that contains parameters of the Ising model and

$$c(\mathbf{S})=\sum_{\boldsymbol{y}\in\{0,1\}^J}\exp\left[\sum_{j=1}^{J}s_{jj}y_j/2+\sum_{j=1}^{J-1}\sum_{k=j+1}^{J}s_{jk}y_jy_k\right]$$

is a normalizing constant. The parameter matrix $\mathbf{S}$ encodes a network with the J items being the nodes. More specifically, an edge is present between nodes i and j if and only if the corresponding entry s_{ij} is nonzero. If an edge exists between nodes i and j, then Y_i and Y_j are conditionally dependent given the rest of the variables. Otherwise, the two variables are conditionally independent.

In Ising network analysis, the goal is to estimate the parameter matrix $\mathbf{S}$. The standard likelihood function is computationally intensive when J is large, as it requires computing a normalizing constant $c(\mathbf{S})$ which involves a summation of all the 2^J response patterns. To address this computational issue, Besag (1975) proposed a conditional likelihood which is obtained by aggregating the conditional distributions of Y_j given $\mathbf{Y}_{-j}=(Y_1,\ldots,Y_{j-1},Y_{j+1},\ldots,Y_J)^{\top}$, for $j=1,\ldots,J$, where the conditional distribution of Y_j given $\mathbf{Y}_{-j}$ takes a logistic regression form. More precisely, the conditional likelihood with one observation $\boldsymbol{y}$ is defined as

$$p^*(\boldsymbol{y}\mid\mathbf{S})=\prod_{j=1}^{J}p(y_j\mid\boldsymbol{y}_{-j},\mathbf{S})=\prod_{j=1}^{J}\frac{\exp\left[(s_{jj}/2+\sum_{k\neq j}s_{jk}y_k)y_j\right]}{1+\exp\left(s_{jj}/2+\sum_{k\neq j}s_{jk}y_k\right)}.\quad(2)$$

A disadvantage of the conditional likelihood is that it requires a fully observed dataset because missing values cannot be straightforwardly marginalized out from (2). In what follows, we discuss how missing data can be treated in the conditional likelihood.

1.2. Proposed Method

Consider a dataset with N observations. Let $\Omega_j\subset\{1,\ldots,N\}$ be the subset of observations whose data on item j are missing. For each observation i and item j, y_{ij} denotes the observed response if $i\notin\Omega_j$, and otherwise, y_{ij} is missing. Thus, the observed data include Ω_j and y_{ij}, for $i\in\{1,\ldots,N\}\backslash\Omega_j$ and $j=1,\ldots,J$.

The proposed procedure iterates between two steps—(1) imputing the missing values of y_{ij} for $i\in\Omega_j$, $j=1,\ldots,J$, achieved via a full conditional specification, and (2) sampling the posterior distribution of $\mathbf{S}$ given the most recently imputed data. Let t be the current iteration number. Further, let $\boldsymbol{y}_i^{(t-1)}=(y_{i1}^{(t-1)},\ldots,y_{iJ}^{(t-1)})^{\top}$, $i=1,\ldots,N$, be imputed data from the previous iteration, where $y_{ij}^{(t-1)}=y_{ij}$ for $i\notin\Omega_j$ and $y_{ij}^{(t-1)}$ is imputed in the $(t-1)$th iteration for $i\in\Omega_j$. For the tth iteration, the imputation and sampling steps are described as follows.

Imputation We initialize the imputation in the tth iteration with the previously imputed data set $(y_{i1}^{(t-1)},\ldots,y_{iJ}^{(t-1)})$, $i=1,\ldots,N$. Then, we run a loop over all the items, $j=1,\ldots,J$. In step j of the loop, we impute y_{ij} for all $i\in\Omega_j$, given the most recently imputed data

https://doi.org/10.1007/s11336-024-09985-2 Published online by Cambridge University Press

$(y_{i1}^{(t)}, \ldots, y_{i,j-1}^{(t)}, y_{ij}^{(t-1)}, \ldots, y_{iJ}^{(t-1)})$, $i = 1, \ldots, N$. We then obtain an updated data set $(y_{i1}^{(t)}, \ldots, y_{i,j}^{(t)}, y_{i,j+1}^{(t-1)}, \ldots, y_{iJ}^{(t-1)})$ by incorporating the newly imputed values for y_{ij}.

The imputation of each variable j is based on the conditional distribution of Y_j given $\mathbf{Y}_{-j}$. Under the Ising model, this conditional distribution takes a logistic regression form. For computational reasons to be discussed in the sequel, we introduce an auxiliary parameter vector $\boldsymbol{\beta}_j = (\beta_{j1}, \ldots, \beta_{jJ})^\top$ as coefficients in the logistic regression, instead of directly using $\mathbf{S}$ from the previous iteration to sample the missing y_{ij}s. Unlike the constraint of $s_{ij} = s_{ji}$ in the symmetric matrix $\mathbf{S}$, no constraints are imposed on $\boldsymbol{\beta}_j$, $j = 1, \ldots, J$, which makes the sampling computationally efficient; see discussions in Sect. 1.4. The imputation of variable j consists of the following two steps:

1. Sample auxiliary parameter vector $\boldsymbol{\beta}_j^{(t)}$ from the posterior distribution

$$p^{(t,j)}(\boldsymbol{\beta}_j) \propto \pi_j(\boldsymbol{\beta}_j) \prod_{i=1}^{N} \frac{\exp\left[(\beta_{jj}/2 + \sum_{k \neq j} \beta_{jk} y_{ik}^{(t,j-1)}) y_{ij}^{(t,j-1)}\right]}{1 + \exp(\beta_{jj}/2 + \sum_{k \neq j} \beta_{jk} y_{ik}^{(t,j-1)})}, \tag{3}$$

   where $\pi_j(\boldsymbol{\beta}_j)$ is the prior distribution for the auxiliary parameters $\boldsymbol{\beta}_j$.
2. Sample $y_{ij}^{(t)}$ for each $i \in \Omega_j$ from a Bernoulli distribution with success probability

$$\frac{\exp(\beta_{jj}^{(t)}/2 + \sum_{k \neq j} \beta_{jk}^{(t)} y_{ik}^{(t,j-1)})}{1 + \exp(\beta_{jj}^{(t)}/2 + \sum_{k \neq j} \beta_{jk}^{(t)} y_{ik}^{(t,j-1)})}. \tag{4}$$

After these two steps, we obtain $(y_{i1}^{(t)}, \ldots, y_{i,j}^{(t)}, y_{i,j+1}^{(t-1)}, \ldots, y_{iJ}^{(t-1)})$ by incorporating the newly imputed values for $y_{ij}, i \in \Omega_j$. We emphasize that only the missing values are updated. For $i \notin \Omega_j$, $y_{ij}^{(t)}$ is always the observed value of y_{ij}. After the loop over all the items, we have the imputed data set $(y_{i1}^{(t)}, \ldots, y_{iJ}^{(t)})$ as the output from this imputation step.

Sampling $\mathbf{S}$ Given the most recently imputed data $\boldsymbol{y}_i^{(t)}$, $i = 1, \ldots, N$, update $\mathbf{S}^{(t)}$ by sampling from the pseudo-posterior distribution

$$p(\mathbf{S} \mid \boldsymbol{y}_1^{(t)}, \ldots, \boldsymbol{y}_N^{(t)}) \propto \pi(\mathbf{S}) \prod_{i=1}^{N} p^*(\boldsymbol{y}_i^{(t)} \mid \mathbf{S}), \tag{5}$$

where $\pi(\mathbf{S})$ is the prior distribution for the Ising parameter matrix $\mathbf{S}$ and recall that $\prod_{i=1}^{N} p^*(\boldsymbol{y}_i^{(t)} \mid \mathbf{S})$ is the conditional likelihood.

Figure 1 visualizes the steps performed in the proposed method. Note that it is unnecessary to sample the parameter matrix $\mathbf{S}$ during the burn-in period and in every iteration after the burn-in period; thus, we employ a thinning step after the burn-in period. This is done to both decrease computational cost and reduce the auto-correlation in the imputed data. Moreover, we outline the proposed algorithm in Algorithm 1. The final estimate of $\mathbf{S}$ is obtained by averaging all the $\mathbf{S}^{(t)}$ obtained after the burn-in period. The computational details, including the sampling of auxiliary parameters and Ising parameter matrix and discussions of the computational complexity, are given in Sect. 1.4.

We remark that our method imputes the missing variables one by one for each observation. This method is chosen because simultaneously imputing all the missing variables is typically

https://doi.org/10.1007/s11336-024-09985-2 Published online by Cambridge University Press

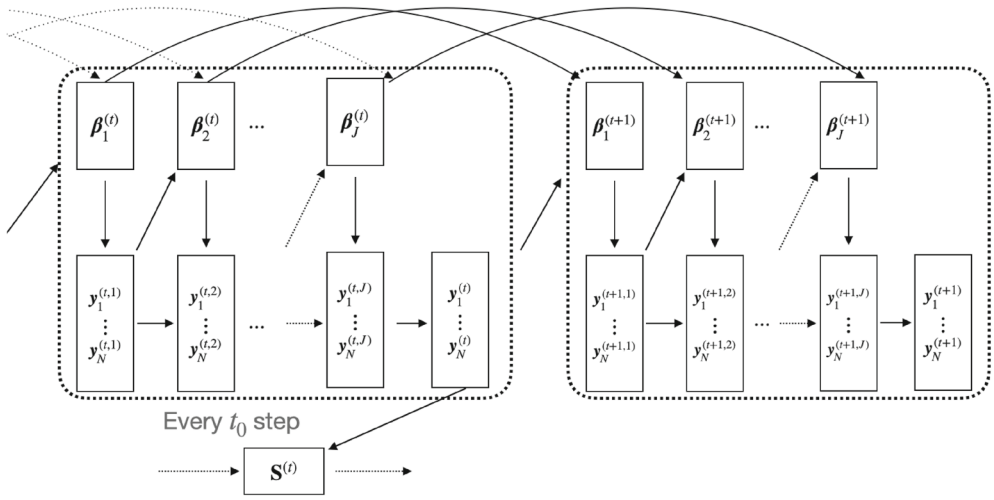


FIGURE 1.
Flowchart of the updating rule for the proposed method.

Algorithm 1: Ising Network Analysis with Iterative Imputation

Data: Given observed data, initial values for the Ising model parameters $\mathbf{S}^{(0)}$, randomly imputed missing data $\boldsymbol{y}_1^{(0)}, \ldots, \boldsymbol{y}_N^{(0)}$, MCMC length T, burn-in size T_0, thinning steps size t_0. Let auxiliary parameters $\boldsymbol{\beta}_j^{(0)} = \boldsymbol{s}_j^{(0)}$, $j = 1, \ldots, J$.

for each iteration $t = 1$ to T **do**

  for each $j = 1$ to J **do**

    Sample auxiliary parameter vector $\boldsymbol{\beta}_j^{(t)}$ from $p^{(t,j)}(\boldsymbol{\beta}_j)$.

    Sample $y_{ij}^{(t)}$ for each $i \in \Omega_j$ from the Bernoulli distribution given in (4).

  end

  if $t > T_0$ and t is a multiple of t_0 **then**

    Sample $\mathbf{S}^{(t)}$ from $p(\mathbf{S} \mid \boldsymbol{y}_1^{(t)}, \ldots, \boldsymbol{y}_N^{(t)})$ given in (5).

  end

end

Output: $\hat{\mathbf{S}} = \frac{1}{M - M_0} \sum_{t \in \{(M_0+1)t_0, \ldots, Mt_0\}} \mathbf{S}^{(t)}$, where $M = \lfloor T/t_0 \rfloor$ and $M_0 = \lfloor T_0/t_0 \rfloor$.

computationally infeasible, especially when some observation units have many missing values. Simultaneous imputation requires evaluating the joint distribution of the missing variables given the observed ones, whose computational complexity grows exponentially with the number of missing values. In contrast, the proposed method is based on unidimensional conditional distributions, which is computationally more feasible. We also note that the proposed method has several variants that should also work well. These variants are discussed in Sect. 3.

1.3. Statistical Consistency

As our method is not a standard Bayesian inference procedure, we provide an asymptotic theory under the frequentist setting to justify its validity. In particular, we show that the $\mathbf{S}$ parameter

https://doi.org/10.1007/s11336-024-09985-2 Published online by Cambridge University Press

sampled from the pseudo-posterior distribution converges to the true parameter $\mathbf{S}_0$, under the assumptions that the Ising model is correctly specified and the data are MAR.

Consider one observation with a complete data vector $\mathbf{Y} = (Y_1, \ldots, Y_J)^\top$. Further, let $\mathbf{Z} = (Z_1, \ldots, Z_J)^\top$ be a vector of missing indicators, where $Z_{ij} = 1$ if Y_{ij} is observed and $Z_{ij} = 0$ otherwise. We further let $\mathbf{Y}_{obs} = \{Y_j : Z_j = 1, j = 1, \ldots, J\}$ and $\mathbf{Y}_{mis} = \{Y_j : Z_j = 0, j = 1, \ldots, J\}$ be the observed and missing entries of $\mathbf{Y}$, respectively. Consider the joint distribution of observable data $(\mathbf{Y}_{obs}, \mathbf{Z})$, taking the form

$$P(\mathbf{Y}_{obs} = \boldsymbol{y}_{obs}, \mathbf{Z} = \mathbf{z} \mid \mathbf{S}, \boldsymbol{\phi}) = \sum_{y_j : z_j = 0} \left(\exp\left(\boldsymbol{y}^\top \mathbf{S} \boldsymbol{y} / 2 \right) / c(\mathbf{S}) \right) q(\mathbf{z} \mid \boldsymbol{y}, \boldsymbol{\phi}), \tag{6}$$

where $\exp\left(\boldsymbol{y}^\top \mathbf{S} \boldsymbol{y}/2\right) / c(\mathbf{S})$ is the distribution of $\mathbf{Y} = \boldsymbol{y}$ under the Ising model, $q(\mathbf{z} \mid \boldsymbol{y}, \boldsymbol{\phi})$ denotes the conditional probability of $\mathbf{Z} = \mathbf{z}$ given $\mathbf{Y} = \boldsymbol{y}$, and $\boldsymbol{\phi}$ denotes the unknown parameters of this distribution. The MAR assumption, also known as the ignorable missingness assumption, means that the conditional distribution $q(\mathbf{z} \mid \boldsymbol{y}, \boldsymbol{\phi})$ depends on $\boldsymbol{y}$ only through the observed entries, i.e., $q(\mathbf{z} \mid \boldsymbol{y}, \boldsymbol{\phi}) = q(\mathbf{z} \mid \boldsymbol{y}_{obs}, \boldsymbol{\phi})$. In that case, (6) can be factorized as

$$P(\mathbf{Y}_{obs} = \boldsymbol{y}_{obs}, \mathbf{Z} = \mathbf{z} \mid \mathbf{S}, \boldsymbol{\phi}) = q(\mathbf{z} \mid \boldsymbol{y}_{obs}, \boldsymbol{\phi}) \times \left(\sum_{y_j : z_j = 0} \exp\left(\boldsymbol{y}^\top \mathbf{S} \boldsymbol{y} / 2 \right) / c(\mathbf{S}) \right). \tag{7}$$

Consequently, the inference of $\mathbf{S}$ does not depend on the unknown distribution $q(\mathbf{z} \mid \boldsymbol{y}, \boldsymbol{\phi})$.

As shown in Liu et al. (2014), the MAR assumption, together with additional regularity conditions, ensures that the iterative imputation of the missing responses converges to the imputation distribution under a standard Bayesian procedure as the number of iterations and the sample size N go to infinity. A key to this convergence result is the compatibility of the conditional models in the imputation step—the logistic regression models are compatible with the Ising model as a joint distribution. The validity of the imputed samples further ensures the consistency of the estimated Ising parameter matrix. We summarize this result in Theorem 1.

Theorem 1. *Assume the following assumptions hold:* (1) *The Markov chain for missing data, generated by the iterative imputation algorithm (Algorithm 1), is positive Harris recurrent and thus admits a unique stationary distribution;* (2) *the missing data process is ignorable; and* (3) *A regularity condition holds for prior distributions of Ising model parameters and auxiliary parameters, as detailed in the supplementary material. Let* $\pi_N^*(\mathbf{S})$ *be the posterior density of* $\mathbf{S}$ *implied by the stationary distribution of the proposed method. Given the true parameters* $\mathbf{S}_0$ *for the Ising model and any* $\varepsilon > 0$, *we have* $\pi_N^*(\mathbf{S})$ *concentrates at* $\mathbf{S}_0$,

$$\int_{B_\varepsilon(\mathbf{S}_0)} \pi_N^*(\mathbf{S}) d\mathbf{S} \to 1, \tag{8}$$

in probability as $N \to \infty$. $B_\varepsilon(\mathbf{S}_0) = \{\mathbf{S} : \|\mathbf{S} - \mathbf{S}_0\| < \varepsilon\}$ *is the open ball of radius* ε *at* $\mathbf{S}_0$.

We provide intuitions about this consistency result. Suppose that the data are generated by an Ising model. The iterative imputation method ensures that the parameters of the logistic regressions are close to those implied by the true Ising model, and thus, the conditional distributions we use to impute the missing values are close to those under the true model. This further guarantees that the joint distribution of the imputed data given the observed ones is close to that under the true Ising model, and consequently, the Ising model parameters we learn from the imputed data are close to those of the true model.

https://doi.org/10.1007/s11336-024-09985-2 Published online by Cambridge University Press

1.4. Computational Details

In what follows, we discuss the specification of the prior distributions and the sampling of auxiliary parameters $\boldsymbol{\beta}_j$ and Ising model parameters $\mathbf{S}$.

Sampling $\boldsymbol{\beta}_j$ We set independent mean-zero normal priors for entries of $\boldsymbol{\beta}_j$. For the intercept parameter β_{jj}, we use a weakly informative prior by setting the variance to 100. For the slope parameters $\beta_{jk}, k \neq j$, we set a more informative prior by setting the variance to be 1, given that these parameters correspond to the off-diagonal entries of $\mathbf{S}$, which are sparse and typically do not take extreme values. The sampling of the auxiliary parameters $\boldsymbol{\beta}_j$, following (3), is essentially a standard Bayesian logistic regression problem. We achieve it by a Markov chain Monte Carlo (MCMC) sampler called the Pólya–Gamma sampler (Polson et al., 2013) .

To obtain $\boldsymbol{\beta}_j^{(t)}$, this sampler starts with $\boldsymbol{\beta}_j^{(t-1)}$ from the previous step. It constructs an MCMC transition kernel by a data argumentation trick. More precisely, the following two steps are performed.

1. Given $\boldsymbol{\beta}^{(t-1)}$, independently sample N augmentation variables, each from a Pólya–Gamma distribution (Barndorff-Nielsen et al., 1982) .
2. Given the N augmentation variables, sample $\boldsymbol{\beta}^{(t)}$ from a J-variate normal distribution.

The details of these two steps are given in the supplementary material, including the forms of the Pólya–Gamma distributions and the mean and covariance matrix of the J-variate normal distribution. We choose the Pólya–Gamma sampler because it is very easy to construct and computationally efficient. It is much easier to implement than Metropolis–Hastings samplers which often need tuning to achieve good performance.

We comment on the computational complexity of the sampling of $\boldsymbol{\beta}_j$. The sampling from the Pólya–Gamma distribution has a complexity $O(NJ)$, and the sampling from the J-variate normal distribution has a complexity of $O(NJ^2)+O(J^3)$. Consequently, a loop of all the $\boldsymbol{\beta}_j$, $j = 1, \ldots, J$, has a complexity of $O((N+J)J^3)$.

Sampling $\mathbf{S}$ Since $\mathbf{S}$ is a symmetric matrix, we reparametrize it by vectorizing its off-diagonal entries (including the diagonal entries). Specifically, the reparameterization is done by half-vectorization, denoted by $\boldsymbol{\alpha} = \text{vech}(\mathbf{S}) = (s_{11}, \ldots, s_{J1}, s_{22}, \ldots, s_{J2}, \ldots, s_{JJ})^\top \in \mathbb{R}^{J(J+1)/2}$. It is easy to see that $\text{vech}(\cdot)$ is a one-to-one mapping between $\mathbb{R}^{J(J+1)/2}$ and $J \times J$ symmetric matrices. Therefore, we impose a prior distribution on $\boldsymbol{\alpha}$ and sample $\boldsymbol{\alpha}^{(t)}$ in the tth iteration when $\mathbf{S}$ is sampled. Then we let $\mathbf{S}^{(t)} = \text{vech}^{-1}(\boldsymbol{\alpha}^{(t)})$.

Recall that a thinning step is performed, and t_0 is the gap between two samples of $\mathbf{S}$. Let t be a multiple of t_0 and $\boldsymbol{\alpha}^{(t-t_0)} = \text{vech}(\mathbf{S}^{(t-t_0)})$ be previous value of $\boldsymbol{\alpha}$. The sampling of $\boldsymbol{\alpha}^{(t)}$ is also achieved by a Pólya–Gamma sampler, which involves the following two steps similar to the sampling of $\boldsymbol{\beta}_j$.

1. Given $\boldsymbol{\alpha}^{(t-t_0)}$, independently sample NJ augmentation variables, each from a Pólya–Gamma distribution.
2. Given the NJ augmentation variables, sample $\boldsymbol{\alpha}^{(t)}$ from a $J(J+1)/2$-variate normal distribution.

The details of these two steps are given in the supplementary material. We note that the computational complexity of sampling the NJ augmentation variables is $O(NJ^2)$, and that of sampling $\boldsymbol{\alpha}^{(t)}$ is $O(NJ^5) + O(J^6)$, resulting in an overall complexity $O((N+J)J^5)$. Comparing the complexities of the imputation and sampling $\mathbf{S}$ steps, we notice that the latter is computationally much more intensive. This is the reason why we choose to impute data by introducing auxiliary parameters $\boldsymbol{\beta}_j$s rather than using Ising network parameters $\mathbf{S}$ so that the iterative imputation mixes much faster in terms of the computation time. In addition, we only sample $\mathbf{S}$ every t_0 iterations

https://doi.org/10.1007/s11336-024-09985-2 Published online by Cambridge University Press

for a reasonably large t_0 to avoid a high computational cost and also reduce the auto-correlation between the imputed data.

We remark that Marsman et al. (2022) considered a similar Ising network analysis problem based on fully observed data, in which they proposed a Bayesian inference approach based on a spike-and-slab prior to learning $\mathbf{S}$. Their Bayesian inference is also based on a Pólya–Gamma sampler. However, they combined Gibbs sampling with a Pólya–Gamma sampler, updating one parameter in $\mathbf{S}$ at a time. This Gibbs scheme often mixes slower than the joint update of $\mathbf{S}$ as in the proposed method and, thus, is computationally less efficient. The proposed Pólya–Gamma sampler may be integrated into the framework of Marsman et al. (2022) to improve their computational efficiency.

2. Numerical Experiments

We illustrate the proposed method and show its power via simulation studies and a real-world data application. In Sect. 2.1, we conduct two simulation studies, evaluating the proposed method under two MAR scenarios, one of which involves missingness due to screening items. A further simulation study is carried out, applying our method to a 15-node Ising model governed by the MCAR mechanism. Detailed exposition of this study can be found in the supplementary materials.

2.1. Simulation

Study I We generate data from an Ising model with $J = 6$ variables. Missing values are generated under an MAR setting that is not MCAR. The proposed method is then compared with Bayesian inference based on (1) listwise deletion and (2) a single imputation, where the single imputation is based on the imputed data from the Tth iteration of Algorithm 1, recalling that T_0 is the burn-in size.

We configure the true parameter matrix $\mathbf{S}_0$ as follows. Since $\mathbf{S}_0$ is a symmetric matrix, we only need to specify its upper triangular matrix and then the diagonal entries. For the upper triangular entries (i.e., s_{jl}, $j < l$), we randomly assign 50% of them to zero to introduce a moderately sparse setting. In addition, the nonzero parameters are then generated by sampling from a uniform distribution over the set $[-1, -0.4]\cup[0.4, 1]$. The intercept parameters s_{jj}, $j = 1, \ldots, J$ are set to zero. The true parameter values are given in the supplementary material. Missing data are simulated by masking particular elements under an MAR mechanism. In particular, we have $z_{i6} = 1$, so that the sixth variable is always observed. We further allow the missingness probabilities of the first five variables (i.e., $z_{ij} = 0$, $j = 1, \ldots, 5$) to depend on the values of y_{i6}. The specific settings on $p(z_{ij} = 0 \mid y_{i6})$, $j = 1, \ldots, 5$ are detailed in the supplementary material. Data are generated following the aforementioned Ising model and MAR mechanism for four different sample sizes, $N = 1000$, 2000, 4000, and 8000, respectively. For each sample size, 50 independent replications are created.

Three methods are compared—the proposed method, Bayesian inference with a single imputation, and Bayesian inference based on complete cases from listwise deletion. The Bayesian inference for complete data is performed by sampling parameters from the posterior implied by the pseudo-likelihood and a normal prior, which is a special case of the proposed method without iterative imputation steps. All these methods shared the same initial values $s_{jl}^{(0)} \sim U(-0.1, 0.1)$, $1 \leq j \leq l \leq J$. For our proposed method, we set the length of the Markov chain Monte Carlo (MCMC) iterations to $T = 5000$ and a burn-in size of $T_0 = 1000$, with a thinning parameter $k_0 = 10$. This setup leads to an effective total of 400 MCMC samples for the Ising parameter matrix $\mathbf{S}$. Notably, identical MCMC length and burn-in configuration are applied during parameters inference in the single imputation and complete-case analyses.

https://doi.org/10.1007/s11336-024-09985-2 Published online by Cambridge University Press

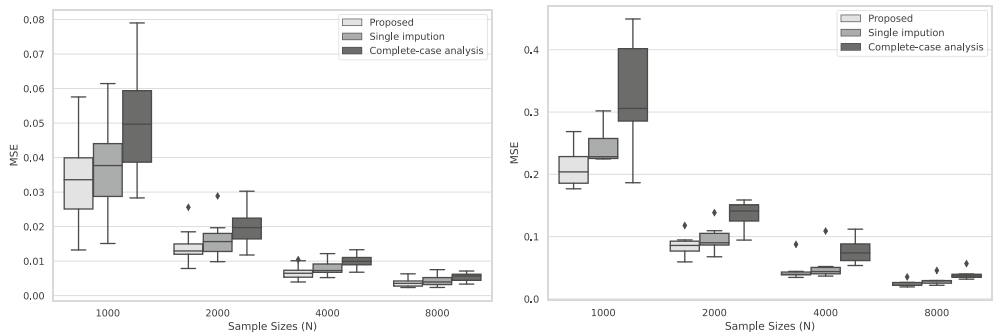


FIGURE 2.
a Boxplots of MSEs for edge parameters s_{jl}. **b** Boxplots of MSEs for intercept parameters s_{jj}.

Figure 2 gives the plots for the mean squared errors (MSE) of the estimated edge parameters and intercept parameters under different sample sizes and for different methods. The MSE for each parameter s_{jl} is defined as

$$\frac{1}{50}\sum_{k=1}^{50}(\hat{s}_{k,jl} - s_{0,jl})^2. \tag{9}$$

Here, $\hat{s}_{k,jl}$ denotes the estimated value from the kth replication while $s_{0,jl}$ refers to the true value. Each box in panel (a) corresponds to the 15 edge parameters, and each box in panel (b) corresponds to the 6 intercept parameters. We notice that the listwise deletion procedure introduces biases into the edge and intercept estimation, resulting in the MSEs for certain parameters not decaying toward zero as the sample size grows. Additionally, both the proposed method and the single imputation method offer accurate parameter estimation, with MSEs decaying toward zero as the sample size increases. Notably, the proposed method is substantially more accurate than the single imputation method, suggesting that aggregating over multiple imputed datasets improves the estimation accuracy. Furthermore, for smaller sample sizes, the complete-case analysis seems to yield slightly more accurate estimates of the edge parameters than the single imputation method. Across four sample sizes, the median computational times for obtaining the results of the proposed method were 33, 50, 88, and 185 s, respectively.[1]

Study II: Missing due to screening items Missingness due to screening items is commonly encountered in practice, posing challenges to the network analysis (Borsboom et al., 2017; McBride et al., 2023) . This occurs, for example, in surveys where initial screening questions determine respondents' eligibility or relevance for subsequent questions. Suppose respondents indicate a lack of relevant experience (i.e., their answers to the screening items are all negative). In that case, they are not prompted to answer certain follow-up questions, making the missingness of these responses depend on their answers to the screening questions and, thus, MAR. Our real data example in Sect. 2.2 involves two screening items, which results in a large proportion of missing data.

We consider a simulation setting involving two screening items to evaluate the proposed method's performance under this setting. Similar to Study I, we consider a setting with six items,

[1] All simulations were performed on Intel-based systems with the following configuration: Ubuntu 22.04.3 LTS operating system; Intel(R) Xeon(R) Platinum 8375C CPU @ 2.90 GHz; Python version 3.11.2; Numpy version 1.23.5.

https://doi.org/10.1007/s11336-024-09985-2 Published online by Cambridge University Press

the first two of which are the screening items. The full data are generated under an Ising model, whose parameters are given in the supplementary material, where the corresponding network has six positive edges including one between the two screening items. The responses to the screening items are always set as observed for any observation. When an observation's responses to the screening items are both zero, their responses to the rest of the four items are regarded as missing.

We consider a single sample size $N = 8000$ and generate 50 independent datasets. We apply the proposed method, the single imputation method, and the complete-case analysis. For each estimation procedure, we set the MCMC iterations $T = 5000$, the burn-in size $T_0 = 1000$, and the thinning parameter $k_0 = 10$. These methods are compared in terms of MSEs and biases for parameter estimation.

Table 1 presents the result. For all the edge parameters except for s_{12}, the three estimation methods work well, though the single imputation method is slightly less accurate, as indicated by its slightly larger MSEs. However, the complete-case estimate is substantially negatively biased for s_{12}, the edge between two screening items. At the same time, the imputation-based methods are still accurate, with the proposed method having a smaller MSE than that of the single imputation method. This result confirms that running a complete-case analysis on data involving screening items is problematic while performing the imputation-based methods, especially the proposed method, yields valid results.

We provide discussions on this result. The negative bias for s_{12} in the complete-case analysis is due to a selection bias, typically referred to as Berkson's paradox (De Ron et al., 2021). The complete-case analysis excludes all the response vectors with negative responses to both screening items. Consequently, a positive response on one screening item strongly suggests a negative response on the other, regardless of the responses to the rest of the items. This results in a falsely negative conditional association between the two screening items. In fact, one can theoretically show that the frequentist estimate of s_{12} based on the maximum pseudo-likelihood is negative infinity. The finite parameter estimate in Table 1 for s_{12} is due to the shrinkage effect

TABLE 1.
MSEs and biases for edge parameters.

Edge	Proposed	Single imputation	Complete-case analysis
	MSE | Bias	MSE | Bias	MSE | Bias
s_{12}	0.007 | 0.029	0.007 | 0.032	57.060 | −7.524
s_{13}	0.012 | 0.030	0.019 | 0.020	0.012 | 0.030
s_{14}	0.010 | 0.008	0.011 | −0.002	0.010 | 0.011
s_{15}	0.014 | −0.007	0.020 | −0.002	0.014 | −0.005
s_{16}	0.011 | −0.020	0.017 | −0.027	0.011 | −0.017
s_{23}	0.003 | 0.012	0.004 | 0.009	0.004 | 0.012
s_{24}	0.005 | 0.004	0.005 | 0.001	0.005 | 0.004
s_{25}	0.007 | 0.011	0.009 | 0.013	0.007 | 0.012
s_{26}	0.005 | −0.005	0.006 | −0.008	0.005 | −0.003
s_{34}	0.004 | −0.004	0.004 | −0.003	0.004 | −0.003
s_{35}	0.006 | 0.014	0.007 | 0.017	0.006 | 0.014
s_{36}	0.008 | −0.009	0.007 | −0.017	0.008 | −0.009
s_{45}	0.006 | −0.004	0.006 | −0.003	0.006 | −0.004
s_{46}	0.006 | −0.002	0.006 | −0.003	0.006 | −0.003
s_{56}	0.007 | 0.001	0.008 | −0.002	0.007 | 0.002

https://doi.org/10.1007/s11336-024-09985-2 Published online by Cambridge University Press

of the prior distribution that we impose. On the other hand, the proposed method makes use of the observed frequency of the (0, 0) response pattern for the two screening items, in addition to the frequencies of the fully observed response vectors. As shown by the identifiability result in Appendix, these frequencies are sufficient for identifying all the parameters of the Ising model.

2.2. A Real Data Application

We analyze the dataset for the 2001–2002 National Epidemiological Survey of Alcohol and Related Conditions (NESARC), which offers valuable insights into alcohol consumption and associated issues in the US population (Grant et al., 2003) . The dataset consists of 43,093 civilian non-institutionalized individuals aged 18 and older. In this analysis, we focus on two specific sections of the survey that concern two highly prevalent mental health disorders—major depressive disorder (MDD) and generalized anxiety disorder (GAD). Because MDD and GAD have high symptom overlap (Hettema, 2008) and often co-occur (Hasin et al., 2005) , it is important to perform a joint analysis of the symptoms of the two mental health disorders and study their separation. In particular, Blanco et al. (2014) performed factor analysis based on the same data and found that the two mental health disorders have distinct latent structures. We reanalyze the data, hoping to gain some insights from the network perspective of the two mental health disorders.

Following Blanco et al. (2014), we consider data with nine items measuring MDD and six items measuring GAD. These items are designed according to the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) (American Psychiatric Association, 2000) . These items ask the participants if they have recently experienced certain symptoms; see Table 2 for their short descriptions. After eliminating samples with entirely absent values across the 15 items, a total of 42,230 cases remain in the dataset. Note that any "Unknown" responses in the original data are converted into missing values. The dataset exhibits a significant degree of missingness, with only 2,412 complete cases for the 15 items, representing approximately 6% of the total cases. Specifically, the missing rate for each item is given in Table 2. Importantly, items D1 and D2 function as screening items and, thus, have a very low missing rate. The respondents did not need to answer items D3–D9 if the responses to D1 and D2 were "No" or "Unknown," resulting in high missing rates for these items. This pattern suggests that the missing data in this study is not MCAR. The GAD items A1–A6 also have a screening item, which results in the high

TABLE 2.
Descriptions of MDD and GAD items and their missing rates.

MDD item description	
D1 (0.1%): Depressed mood	D5 (68.5%): Psychomotor agitation/retardation
D2 (0.2%): Diminished interest	D6 (68.0%): Fatigue/loss of energy
D3 (68.5%): Weight loss or gain	D7 (67.9%): Feelings of guilt
D4 (67.9%): Insomnia or hypersomnia	D8 (67.9%): Diminished concentration
	D9 (67.7%): Recurrent thoughts of death
GAD Item Description	
A1 (91.8%): Restlessness	A4 (91.8%): Irritability
A2 (91.9%): Easily fatigued	A5 (91.9%): Muscle tension
A3 (91.8%): Difficulty concentrating	A6 (91.8%): Sleep disturbance

https://doi.org/10.1007/s11336-024-09985-2 Published online by Cambridge University Press

missing rates in A1 through A6. Following the treatment in Blanco et al. (2014), these screening items are not included in the current analysis.

We apply the proposed method and the complete-case analysis to the data. For each method, 10 MCMC chains with random starting values are used, each having 10,000 MCMC iterations and a burn-in size 5000. The Gelman–Rubin statistics are always below 1.018, confirming the satisfactory convergence of all 120 parameters for both methods. The estimated network structures for MDD and GAD items are presented in Fig. 3, where an edge is shown between two variables when the absolute value of the estimated parameter is greater than 0.5. We emphasize that this threshold is applied only for visualization purposes, rather than for edge selection. Consequently, the edges in Fig. 3 should only be interpreted as edges with large estimated parameters, rather than truly nonzero edges. The nine MDD items are shown as blue nodes at the bottom, and the six GAD items are shown as orange nodes at the top. The edges are colored blue and orange, which represent positive and negative parameter estimates, respectively. In addition, the line thickness of the edges indicates their magnitude. A clear difference between the two methods is the edge between D1 "depressed mood most of the day, nearly every day," and D2 "markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day," which are two screening questions in the survey that all the participants responded to. The estimated parameter for this edge has a large absolute value under each of the two methods, but the estimated parameter is negative in the complete-case analysis, while it is positive according to the proposed method. As revealed by the result of Study II in Sect. 2.1, the negative edge estimate of the edge between the screening items given by the complete-case analysis is spurious. Considering the content of these items, we believe that the estimate from the proposed method is more sensible. Other than this edge, the remaining structure of the two networks tends to be similar, but with some differences. In particular, we see that the complete-case analysis yields more edges than the proposed method; for example, the edges of A4–A5, A1–D5, D1–D6, D1–D7, D1–D8, and D8–D9 appear in the estimated network from the complete-case analysis but not in that of the proposed method, while

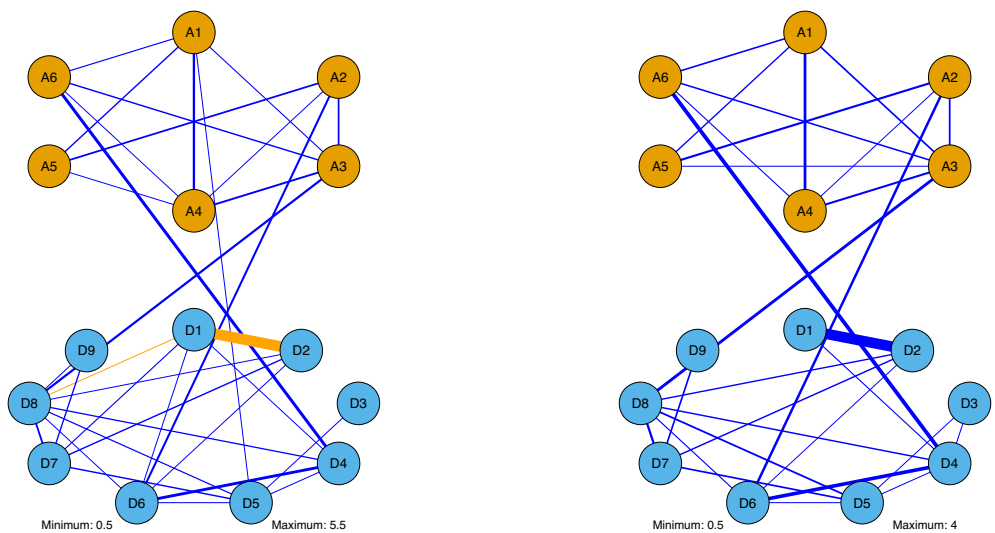


FIGURE 3.
Estimated network structure for MDD and GAD. **a** Complete-case analysis. **b** Proposed method.

https://doi.org/10.1007/s11336-024-09985-2 Published online by Cambridge University Press

only two edges, A3–A5 and D3–D4, are present in the network estimated by the proposed method but absent in the network from the complete-case analysis. We believe this is due to the higher estimation variance of the complete-case analysis caused by its relatively small sample size.

Finally, our analysis shows that the symptoms of each mental health disorder tend to densely connect with each other in the Ising network, while the symptoms are only loosely but positively connected between the two mental health disorders. The edges between the two mental health disorders identify the overlapping symptoms, including "D4: Insomnia or hypersomnia" and "A6: Sleep disturbance," "A2: Easily fatigued" and "D6: Fatigue/loss of energy," and "A3: Difficulty concentrating" and "D8: Diminished concentration." These results suggest that MDD and GAD are two well-separated mental health disorders, despite their high symptom overlap and frequent co-occurrence. This result confirms the conclusion of Blanco et al. (2014) that GAD and MDD are closely related but different nosological entities.

3. Concluding remarks

In this paper, we propose a new method for Ising network analysis in the presence of missing data. The proposed method integrates iterative imputation into a Bayesian inference procedure based on conditional likelihood. An asymptotic theory is established that guarantees the consistency of the proposed estimator. Furthermore, a Pólya–Gamma machinery is proposed for the sampling of Ising model parameters, which yields efficient computation. The power of the proposed method is further shown via simulations and a real data application. An R package has been developed that will be made publicly available upon the acceptance of the paper.

The current work has several limitations that require future theoretical and methodological developments. First, this manuscript concentrates on parameter estimation for the Ising model in the presence of missing data. However, the problem of edge selection (Borsboom, 2022; Marsman et al., 2022; Noghrehchi et al., 2021; Roçkovóa, 2018) requires future investigation. There are several possible directions. One direction is to view it as a multiple testing problem and develop procedures that control certain familywise error rates or the false discovery rate for the selection of edges. To do so, one needs to develop a way to quantify the uncertainty for the proposed estimator. It is non-trivial, as the proposed method is not a standard Bayesian procedure, and we still lack a theoretical understanding of the asymptotic distribution of the proposed procedure. In particular, it is unclear whether the Bernstein–von Mises theorem that connects Bayesian and frequentist estimation holds under the current setting. Another direction is to view it as a model selection problem. In this direction, we can use sparsity-inducing priors to better explore the Ising network structure when it is sparse. We believe that the proposed method, including the iterative imputation and the Pólya–Gamma machinery, can be adapted when we replace the normal prior with the spike-and-slab prior considered in Marsman et al. (2022). This adaptation can be done by adding some Gibbs sampling steps. In addition, it is of interest to develop an information criterion that is computationally efficient while statistically consistent. This may be achieved by computing an information criterion, such as the Bayesian information criterion, for each imputed dataset and then aggregating them across multiple imputations. Finally, the proposed method has several variants that may be useful for problems of different scales. For problems of a relatively small scale (i.e., when J is small), we may perform data imputation using the sampled $\mathbf{S}$ instead of using auxiliary parameters $\boldsymbol{\beta}_j$s. This choice will make the algorithm computationally more intensive, as the sampling of $\mathbf{S}$ has a high computational complexity. On the other hand, it may make the estimator statistically more efficient as it avoids estimating the auxiliary parameters $\boldsymbol{\beta}_j$s, whose dimension is higher than $\mathbf{S}$. For very large-scale problems, one may estimate the Ising model parameters based only on the auxiliary parameters $\boldsymbol{\beta}_j$s. For example, we may estimate s_{ij} by averaging the value of $(\beta_{ij}+\beta_{ji})/2$ over the iterations. This estimator is computationally more

https://doi.org/10.1007/s11336-024-09985-2 Published online by Cambridge University Press

efficient than the proposed one, as it avoids sampling $\mathbf{S}$ given the imputed datasets. This estimator should still be consistent but may be statistically slightly less efficient than the proposed one.

Acknowledgments

We thank the editor, associate editor, and three anonymous referees for their careful review and valuable comments. Siliang Zhang's research is partially supported by Shanghai Science and Technology Committee Rising-Star Program (22YF1411100) and National Natural Science Foundation of China (12301373).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Appendix: Identifiability of Ising Model with Two Screening Items

We investigate the identifiability of the Ising model parameters when there are two screening items.

Proposition 1. *Consider an Ising model with $J \geq 3$, true parameters $\mathbf{S}_0 = (s^0_{ij})_{J\times J}$, and the first two items being the screening items. We define $p_0(\mathbf{y}) := P(\mathbf{Y} = \mathbf{y}|\mathbf{S}_0)$, for any $\mathbf{y} \in \mathcal{A} = \{(x_1, \ldots, x_J)^\top \in \{0,1\}^J : x_1 = 1 \text{ or } x_2 = 1\}$, and $p_0(0,0) := P(Y_1 = 0, Y_2 = 0|\mathbf{S}_0)$ under the Ising model. Then there does not exist an Ising parameter matrix $\mathbf{S} \neq \mathbf{S}_0$ such that $p_0(\mathbf{y}) = P(\mathbf{Y} = \mathbf{y}|\mathbf{S})$, for any $\mathbf{y} \in \mathcal{A}$ and $p_0(0,0) = P(Y_1 = 0, Y_2 = 0|\mathbf{S})$ under the Ising model.*

Proof. We first prove the statement for $J \geq 4$. Suppose that $p_0(\mathbf{y}) = P(\mathbf{Y} = \mathbf{y}|\mathbf{S})$, for any $\mathbf{y} \in \mathcal{A}$ and $p_0(0,0) = P(Y_1 = 0, Y_2 = 0|\mathbf{S})$. We will prove that $\mathbf{S} = \mathbf{S}_0$.
We start by considering items 1, 2, 3, 4. We define the set $\mathcal{A}_{3,4} = \{(x_1, \ldots, x_J)^\top \in \mathcal{A} : x_5 = \ldots = x_J = 0\}$. We note that $\mathcal{A}_{3,4}$ has 12 elements. Using $\boldsymbol{y}_d = (1,0,0,0,0,\ldots,0)^\top \in \mathcal{A}_{3,4}$ as the baseline pattern, for any $\boldsymbol{y} \in \mathcal{A}_{3,4}$ such that $\boldsymbol{y} \neq \boldsymbol{y}_d$ we have

$$\log\left[\frac{\exp(\frac{1}{2}\boldsymbol{y}^\top\mathbf{S}\boldsymbol{y})}{c(\mathbf{S})} \Big/ \frac{\exp(\frac{1}{2}\boldsymbol{y}_d^\top\mathbf{S}\boldsymbol{y}_d)}{c(\mathbf{S})}\right] = \frac{1}{2}\boldsymbol{y}^\top\mathbf{S}\boldsymbol{y} - \frac{1}{2}\boldsymbol{y}_d^\top\mathbf{S}\boldsymbol{y}_d = \frac{1}{2}\boldsymbol{y}^\top\mathbf{S}_0\boldsymbol{y} - \frac{1}{2}\boldsymbol{y}_d^\top\mathbf{S}_0\boldsymbol{y}_d.$$

That gives us 11 linear equations for 10 parameters, $s_{ij}, i, j \leq 4$. By simplifying these equations, we have (1) two linear equations for (s_{11}, s_{12}, s_{22})

$$\begin{aligned} s_{11} - s_{22} &= s^0_{11} - s^0_{22} \\ 2s_{12} + s_{22} &= 2s^0_{12} + s^0_{22} \end{aligned} \tag{10}$$

https://doi.org/10.1007/s11336-024-09985-2 Published online by Cambridge University Press

and (2) $s_{1i} = s^0_{1i}$, $s_{2i} = s^0_{2i}$, $s_{ij} = s^0_{ij}$ for all $i, j = 3, 4$.
We repeat the above calculation for any four-item set involving items 1 and 2. By choosing any item pair $i, j > 2$, $i \neq j$ and repeating the above calculation with patterns in the set $\mathcal{A}_{i,j} = \{(x_1, \ldots, x_J)^\top \in \mathcal{A} : x_l = 0, l \neq 1, 2, i, j\}$, we have $s_{1i} = s^0_{1i}$, $s_{2i} = s^0_{2i}$, $s_{ij} = s^0_{ij}$ for all $i, j > 2, i \neq j$. With the above argument and (10), we only need to show $s_{11} = s^0_{11}$ to prove $\mathbf{S} = \mathbf{S}_0$. To prove $s_{11} = s^0_{11}$, we use $p_0(0,0)/p_0(\boldsymbol{y}_d) = P(Y_1 = 0, Y_2 = 0|\mathbf{S})/P(\mathbf{Y} = \boldsymbol{y}_d|\mathbf{S})$ and that $s_{ij} = s^0_{ij}$ for all $i, j > 2, i \neq j$ which we have proved. We have

$$\sum_{\boldsymbol{y} \in \mathcal{A}_0} \exp\left(-\frac{1}{2}s^0_{11} + \frac{1}{2}\boldsymbol{y}^\top \mathbf{S}_0 \boldsymbol{y}\right) = \sum_{\boldsymbol{y} \in \mathcal{A}_0} \exp\left(-\frac{1}{2}s_{11} + \frac{1}{2}\boldsymbol{y}^\top \mathbf{S}_0 \boldsymbol{y}\right),$$

where $\mathcal{A}_0 = \{(x_1, \ldots, x_J)^\top \in \{0,1\}^J : x_1 = x_2 = 0\}$. As the right-hand side of the above equation is a strictly monotone decreasing function of s_{11}, we know that $s_{11} = s^0_{11}$ is the only solution to the above equation. This proves the $J \geq 4$ case.
We now move on to the case when $J = 3$. We consider $\mathcal{A} = \{(x_1, x_2, x_3)^\top \in \{0,1\}^3 : x_1 = 1 \text{ or } x_2 = 1\}$ and $\boldsymbol{y}_d = (1,0,0)^\top$. Using $\boldsymbol{y}_d$ as the baseline, for any $\boldsymbol{y} \in \mathcal{A}$, $\boldsymbol{y} \neq \boldsymbol{y}_d$, we construct five linear equations given by $\log(p_0(\boldsymbol{y})/p_0(\boldsymbol{y}_d)) = \log(P(\mathbf{Y} = \boldsymbol{y}|\mathbf{S})/P(\mathbf{Y} = \boldsymbol{y}_d|\mathbf{S}))$. From these equations, we obtain: (1) two linear equations for (s_{11}, s_{12}, s_{22})

$$\begin{aligned} s_{11} - s_{22} &= s^0_{11} - s^0_{22} \\ 2s_{12} + s_{22} &= 2s^0_{12} + s^0_{22} \end{aligned} \tag{11}$$

and (2) $s_{13} = s^0_{13}$, $s_{23} = s^0_{23}$, $s_{33} = s^0_{33}$. Again, with the above equations, $\mathbf{S} = \mathbf{S}^0$ if $s_{11} = s^0_{11}$. To show $s_{11} = s^0_{11}$, we use $p_0(0,0)/p_0(\boldsymbol{y}_d) = P(Y_1 = 0, Y_2 = 0|\mathbf{S})/P(\mathbf{Y} = \boldsymbol{y}_d|\mathbf{S})$ and $s_{33} = s^0_{33}$. We have

$$\exp\left(-\frac{1}{2}s^0_{11}\right) + \exp\left(-\frac{1}{2}s^0_{11} + \frac{1}{2}s^0_{33}\right) = \exp\left(-\frac{1}{2}s_{11}\right) + \exp\left(-\frac{1}{2}s_{11} + \frac{1}{2}s^0_{33}\right).$$

As the right-hand side is a strictly monotone decreasing function of s_{11}, we know that $s_{11} = s^0_{11}$ is the only solution to the above equation. This proves the $J = 3$ case, which completes the proof. □

Following the same proof strategy as above, it can be further shown that s_{11}, s_{12}, and s_{22} are not identified in the complete-case analysis, while the rest of the parameters are. This is consistent with the result of Simulation Study II, where the other model parameters can be accurately estimated while the estimates of s_{11}, s_{12}, and s_{22} are substantially different from the corresponding true parameters.

References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders: DSM-IV-TR (4th*. Text (Revision): American Psychiatric Association.

Anderson, C. J., & Yu, H.-T. (2007). Log-multiplicative association models as item response models. *Psychometrika, 72*(1), 5–23.

Armour, C., Fried, E. I., Deserno, M. K., Tsai, J., & Pietrzak, R. H. (2017). A network analysis of DSM-5 posttraumatic stress disorder symptoms and correlates in U.S. military veterans. *Journal of Anxiety Disorders, 45*, 49–59.

Barndorff-Nielsen, O., Kent, J., & Sørensen, M. (1982). Normal variance-mean mixtures and z distributions. *International Statistical Review/Revue Internationale de Statistique, 50*(2), 145–159.

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

https://doi.org/10.1007/s11336-024-09985-2 Published online by Cambridge University Press

Besag, J. (1975). Statistical analysis of non-lattice data. *Journal of the Royal Statistical Society: Series D (The Statistician), 24*(3), 179–195.
Blanco, C., Rubio, J. M., Wall, M., Secades-Villa, R., Beesdo-Baum, K., & Wang, S. (2014). The latent structure and comorbidity patterns of generalized anxiety disorder and major depressive disorder: A national study. *Depression and Anxiety, 31*(3), 214–222.
Borsboom, D. (2008). Psychometric perspectives on diagnostic systems. *Journal of Clinical Psychology, 64*(9), 1089–1108.
Borsboom, D. (2022). Possible futures for network psychometrics. *Psychometrika, 87*(1), 253–265.
Borsboom, D., Deserno, M. K., Rhemtulla, M., Epskamp, S., Fried, E. I., McNally, R. J., & Waldorp, L. J. (2021). Network analysis of multivariate data in psychological science. *Nature Reviews Methods Primers, 1*(1), 58.
Borsboom, D., Fried, E. I., Epskamp, S., Waldorp, L. J., van Borkulo, C. D., van der Maas, H. L., & Cramer, A. O. (2017). False alarm? a comprehensive reanalysis of "evidence that psychopathology symptom networks have limited replicability" by Forbes, Wright, Markon, and Krueger (2017). *Journal of Abnormal Psychology, 126*(7), 989–999.
Brunson, J. C., & Laubenbacher, R. C. (2018). Applications of network analysis to routinely collected health care data: A systematic review. *Journal of the American Medical Informatics Association, 25*(2), 210–221.
Burger, J., Isvoranu, A. M., Lunansky, G., Haslbeck, J. M. B., Epskamp, S., Hoekstra, R. H. A., & Blanken, T. F. (2022). Reporting standards for psychological network analyses in cross-sectional data. *Psychological Methods, 28*(4), 806–824.
Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review, 106*(3), 551–581.
Chen, Y., Li, X., Liu, J., & Ying, Z. (2018). Robust measurement via a fused latent and graphical item response theory model. *Psychometrika, 83*(3), 538–562.
De Ron, J., Fried, E. I., & Epskamp, S. (2021). Psychological networks in clinical populations: Investigating the consequences of Berkson's bias. *Psychological Medicine, 51*(1), 168–176.
Epskamp, S. (2020). Psychometric network models from time-series and panel data. *Psychometrika, 85*(1), 206–231.
Epskamp, S., Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods, 50*(1), 195–212.
Epskamp, S., & Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychological Methods, 23*(4), 617–634.
Epskamp, S., Waldorp, L. J., Mõttus, R., & Borsboom, D. (2018). The Gaussian graphical model in cross-sectional and time-series data. *Multivariate Behavioral Research, 53*(4), 453–480.
Fried, E. I., van Borkulo, C. D., Cramer, A. O., Boschloo, L., Schoevers, R. A., & Borsboom, D. (2017). Mental disorders as networks of problems: A review of recent insights. *Social Psychiatry and Psychiatric Epidemiology, 52*(1), 1–10.
Fried, E. I., Von Stockert, S., Haslbeck, J., Lamers, F., Schoevers, R., & Penninx, B. (2020). Using network analysis to examine links between individual depressive symptoms, inflammatory markers, and covariates. *Psychological Medicine, 50*(16), 2682–2690.
Gile, K. J., & Handcock, M. S. (2017). Analysis of networks with missing data with application to the national longitudinal study of adolescent health. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 66*(3), 501–519.
Grant, B. F., Moore, T., Shepard, J., & Kaplan, K. (2003). *Source and accuracy statement: Wave 1 national epidemiologic survey on alcohol and related conditions (NESARC)* (p. 52). National Institute on Alcohol Abuse and Alcoholism.
Hasin, D. S., Goodwin, R. D., Stinson, F. S., & Grant, B. F. (2005). Epidemiology of major depressive disorder: Results from the national epidemiologic survey on alcoholism and related conditions. *Archives of General Psychiatry, 62*(10), 1097–1106.
Haslbeck, J. M., & Fried, E. I. (2017). How predictable are symptoms in psychopathological networks? A reanalysis of 18 published datasets. *Psychological Medicine, 47*(16), 2767–2776.
Hettema, J. M. (2008). The nosologic relationship between generalized anxiety disorder and major depression. *Depression and Anxiety, 25*(4), 300–316.
Holland, P. W. (1990). The Dutch identity A new tool for the study of item response models. *Psychometrika, 55*, 5–18.
Huisman, M. (2009). Imputation of missing network data: Some simple procedures. *Journal of Social Structure, 10*(1), 1–29.
Ip, E. H. (2002). Locally dependent latent trait model and the Dutch identity revisited. *Psychometrika, 67*, 367–386.
Ising, E. (1925). Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik, 31*(1), 253–258.
Kohler, K., Jankowski, M. D., Bashford, T., Goyal, D. G., Habermann, E. B., & Walker, L. E. (2022). Using network analysis to model the effects of the SARS Cov2 pandemic on acute patient care within a healthcare system. *Scientific Reports, 12*(1), 10050.
Koponen, M., Asikainen, M. A., Viholainen, A., & Hirvonen, P. E. (2019). Using network analysis methods to investigate how future teachers conceptualize the links between the domains of teacher knowledge. *Teaching and Teacher Education, 79*, 137–152.
Li, Y., Craig, B. A., & Bhadra, A. (2019). The graphical horseshoe estimator for inverse covariance matrices. *Journal of Computational and Graphical Statistics, 28*(3), 747–757.
Lin, S.-Y., Fried, E. I., & Eaton, N. R. (2020). The association of life stress with substance use symptoms: A network analysis and replication. *Journal of Abnormal Psychology, 129*(2), 204–214.
Little, R. J., & Rubin, D. B. (2019). *Statistical analysis with missing data* (Vol. 793). John Wiley & Sons.
Liu, J., Gelman, A., Hill, J., Su, Y.-S., & Kropko, J. (2014). On the stationary distribution of iterative imputations. *Biometrika, 101*(1), 155–173.

https://doi.org/10.1007/s11336-024-09985-2 Published online by Cambridge University Press

Luke, D. A., & Harris, J. K. (2007). Network analysis in public health: History, methods, and applications. *Annual Review of Public Health, 28*, 69–93.
Marsman, M., Huth, K., Waldorp, L., & Ntzoufras, I. (2022). Objective Bayesian edge screening and structure selection for Ising networks. *Psychometrika, 87*(1), 47–82.
Marsman, M., & Rhemtulla, M. (2022). Guest editors' introduction to the special issue "network psychometrics in action": Methodological innovations inspired by empirical problems. *Psychometrika, 87*(1), 1–11.
Mazumder, R., & Hastie, T. (2012). The graphical Lasso: New insights and alternatives. *Electronic Journal of Statistics, 6*, 2125–2149.
McBride, O., van Bezooijen, J., Aggen, S. H., Kendler, K. S., & Fried, E. I. (2023). Quantifying skip-out information loss when assessing major depression symptoms. *Journal of Psychopathology and Clinical Science, 132*(4), 396–408.
Mkhitaryan, S., Crutzen, R., Steenaart, E., & de Vries, N. K. (2019). Network approach in health behavior research: How can we explore new questions? *Health Psychology and Behavioral Medicine, 7*(1), 362–384.
Noghrehchi, F., Stoklosa, J., Penev, S., & Warton, D. I. (2021). Selecting the model for multiple imputation of missing data: Just use an IC! *Statistics in Medicine, 40*(10), 2467–2497.
Polson, N. G., Scott, J. G., & Windle, J. (2013). Bayesian inference for logistic models using Póolya-Gamma latent variables. *Journal of the American Statistical Association, 108*(504), 1339–1349.
Roçkovóa, V. (2018). Bayesian estimation of sparse signals with a continuous spike-and-slab prior. *The Annals of Statistics, 46*(1), 401–437.
Ryan, O., Bringmann, L. F., & Schuurman, N. K. (2022). The challenge of generating causal hypotheses using network models. *Structural Equation Modeling: A Multidisciplinary Journal, 29*(6), 953–970.
Siew, C. S. (2020). Applications of network science to education research: Quantifying knowledge and the development of expertise through network analysis. *Education Sciences, 10*(4), 1–16.
Simon de Blas, C., Gomez Gonzalez, D., & Criado Herrero, R. (2021). Network analysis: An indispensable tool for curricula design. A real case-study of the degree on mathematics at the URJC in Spain. *PLOS ONE, 16*(3), 1–21.
Sweet, T. M., Thomas, A. C., & Junker, B. W. (2013). Hierarchical network models for education research: Hierarchical latent space models. *Journal of Educational and Behavioral Statistics, 38*(3), 295–318.
Van Borkulo, C. D., Borsboom, D., Epskamp, S., Blanken, T. F., Boschloo, L., Schoevers, R. A., & Waldorp, L. J. (2014). A new method for constructing networks from binary data. *Scientific Reports, 4*(1), 1–10.
van Buuren, S. (2018). *Flexible imputation of missing data* (2nd ed.). Chapman and Hall/CRC.
van der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review, 113*(4), 842–861.
Van Der Maas, H. L., Kan, K.-J., Marsman, M., & Stevenson, C. E. (2017). Network models for cognitive development and intelligence. *Journal of Intelligence, 5*(2), 16.
Willcox, K. E., & Huang, L. (2017). Network models for mapping educational data. *Design Science, 3*, e18.
Yuan, M., & Lin, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika, 94*(1), 19–35.

Manuscript Received: 20 JUN 2023
Accepted: 5 JUN 2024
Published Online Date: 6 JUL 2024

https://doi.org/10.1007/s11336-024-09985-2 Published online by Cambridge University Press