PSYCHOMETRIKA—VOL. 87, NO. 1, 83–106
MARCH 2022
https://doi.org/10.1007/s11336-021-09781-2

ESTIMATING FINITE MIXTURES OF ORDINAL GRAPHICAL MODELS

KEVIN H. LEE

WESTERN MICHIGAN UNIVERSITY

QIAN CHEN

UNIVERSITY OF NEBRASKA–LINCOLN

WAYNE S. DESARBO

PENNSYLVANIA STATE UNIVERSITY

LINGZHOU XUE

PENNSYLVANIA STATE UNIVERSITY

Graphical models have received an increasing amount of attention in network psychometrics as a promising probabilistic approach to study the conditional relations among variables using graph theory. Despite recent advances, existing methods on graphical models usually assume a homogeneous population and focus on binary or continuous variables. However, ordinal variables are very popular in many areas of psychological science, and the population often consists of several different groups based on the heterogeneity in ordinal data. Driven by these needs, we introduce the finite mixture of ordinal graphical models to effectively study the heterogeneous conditional dependence relationships of ordinal data. We develop a penalized likelihood approach for model estimation, and design a generalized expectation-maximization (EM) algorithm to solve the significant computational challenges. We examine the performance of the proposed method and algorithm in simulation studies. Moreover, we demonstrate the potential usefulness of the proposed method in psychological science through a real application concerning the interests and attitudes related to fan avidity for students in a large public university in the United States.

Key words: Gaussian mixture model, Gaussian graphical model, ordinal data, latent variables, network psychometrics, EM algorithm.

1. Introduction

Graphical models provide a probabilistic approach for modelling the conditional dependence structure of complex systems using graph theory (Lauritzen 1996). In many psychological studies, it is of interest to study the relations among psychological constructs (e.g., attitudes, cognitions, emotions, intelligence), psychopathological symptoms, behaviors, or other psychometric indicators (Borsboom & Molenaar 2015). Recently, graphical models have been introduced to the

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11336-021-09781-2.

The authors wish to recognize and thank the Co-Editors, the Associate Editor, and three anonymous referees for their insightful and constructive comments.
The work of Lingzhou Xue was supported in part by the National Science Foundation (NSF) Grants DMS-1811552, DMS-1953189, and CCF-2007823.
Kevin H. Lee and Qian Chen are co-first authors who contributed equally to this paper.

Correspondence should be made to Lingzhou Xue, Department of Statistics, Pennsylvania State University, 318 Thomas Building, University Park, PA 16802, USA. Email: lzxue@psu.edu

discipline of psychometrics as a promising alternative to classic approaches using latent variables (e.g., Borsboom 2008; Cramer et al. 2010; Schmittmann et al. 2013; Borsboom & Cramer 2013; Epskamp et al. 2017, 2018; Marsman et al 2018). Network psychometrics has received an increasing amount of attention in clinical psychology, psychiatry, social psychology, and other domains (e.g., Fried et al. 2015; Isvoranu et al. 2016; Dalege et al. 2016). Compared to traditional psychometric models, graphical models use an alternative graph-based representation to study the conditional relations among the variables of interest. More specifically, the undirected graphical models represent random variables as nodes and use edges to express conditional dependence relationships. In other words, the nonexistence of an edge between two nodes corresponds to the conditional independence of these two variables giving all other variables. There exists a certain level of equivalence between a particular class of graphical models called Ising models and traditional models typically employed in psychometrics such as logistic regression models, log-linear models, and multi-dimensional item response theory models (Marsman et al. 2018; Epskamp et al. 2018). However, graphical models hypothesize that the network is formed by mutually reinforcing variables, thus providing a new conceptualization of why variables cluster (Epskamp et al. 2018).

In the past two decades, substantial progress has been made in developing new methods, algorithms, and applications of graphical models. The main focus of the current literature in statistics and machine learning is on the graphical modeling of the complex conditional dependence structure among *binary* variables via Ising models, or among *continuous* variables via Gaussian graphical models. The estimation of sparse graphical models is increasingly important to model complex interactions in a large-scale system. Ising models were initially introduced by Ising (1925) in statistical physics for studying magnetic interactions. The joint distribution of Ising models is also known as the quadratic exponential binary distribution (Cox & Wermuth 1994) in the statistics literature. Gaussian graphical models were first studied by Dempster (1972) as the covariance selection problem, and the covariance structure of Gaussian graphical models can be simplified by estimating sparse off-diagonal elements of the inverse of the covariance matrix (also known as the precision matrix). Penalized likelihood estimation (Tibshirani 1996; Fan & Li 2001; Fan et al. 2014) has become a standard procedure to achieve the sparse estimation of such graphical models and provide interpretable results. Penalized estimation of sparse Ising models was developed by Höing & Tibshirani (2009), Ravikumar et al. (2010) and Xue et al. (2012) among others, and the sparse estimation of Gaussian graphical models and their variants were studied by Meinshausen & Bühlmann (2006), Yuan & Lin (2007), Friedman et al. (2008), Liu et al. (2009), Cai et al. (2011), Xue & Zou (2012), Ma et al. (2021), and many others. Recently, graphical models for mixed continuous and discrete variables were also considered by Chen et al. (2015), Lee & Hastie (2015), Haslbeck & Waldorp (2016), Fan et al. (2017) and Cheng et al. (2017), among others.

Graphical models for *ordinal* data has received much less attention. The multivariate probit model has been used to model the joint distribution of ordinal data through a latent multivariate Gaussian distribution. Motivated by the success of multivariate probit analysis (Amemiya 1974; Albert & Chib 1993; Bock & Gibbons 1996; Chib & Greenberg 1998) and polychotomous Rasch model (von Davier & Carstensen 2007), Guo et al. (2015) introduced probit graphical models to study the conditional dependence structure of ordinal variables. The sparse estimation of probit graphical models for ordinal data is very challenging in the presence of latent variables in the resulting log-likelihood function. Guo et al. (2015) designed an approximate EM-like algorithm to estimate the parameters, Suggala et al. (2017) introduced a two-stage procedure based on the bivariate marginal log likelihood function, and Feng & Ning (2019) proposed a rank-based ensemble estimation approach.

Despite these advances, existing ordinal graphical models usually assume a homogeneous population. However, the population is often heterogeneous and is comprised of several sub-

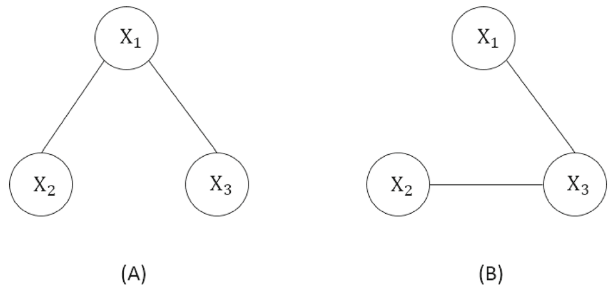


FIGURE 1.
The illustration of mixtures of graphical models with two mixture components (A) and (B).

groups. For example, in psychopathological research, the data randomly sampled from several sub-populations may have different symptomatic structures. As pointed out by Brusco et al. (2019), ignoring the heterogeneity of the data leads to an estimation of network that may not represent any of the underlying population. In the current literature, a number of papers account for heterogeneity in Gaussian graphical models and Ising models. On the one hand, Ruan et al. (2011) and Lee & Xue (2018) studied the finite mixtures of Gaussian graphical models to deal with heterogeneity. On the other hand, Brusco et al. (2019) introduced a two-step procedure, using the clustering for the first step and fitting a graphical model to each cluster, to address the heterogeneity problem in Ising models, and Marsman (2019) recently pointed out the bridge between idiographic and cross-sectional approaches in Ising models.

However, none of these existing papers consider the simultaneous estimation of the heterogeneous conditional relations for ordinal data in the context of graphical models. In the statistics and psychometrics literature, mixture models provide a powerful approach to make use of latent variables and account for the heterogeneous sub-population of individuals. Mixture models incorporate latent group memberships into generalized linear models and probit analysis for ordinal data to model different parameters for heterogeneous sub-populations, including Lwin & Martin (1989), Wedel & DeSarbo (1995), Greene & Hensher (2003), Grün & Leisch (2008), Breen & Luijkx (2010), and many others. We would like to extend the finite mixture method into ordinal graphical modeling to account for heterogeneity.

In this paper, we aim to develop finite mixtures of ordinal graphical models to describe the heterogeneous conditional dependence relationships in ordinal data. Our proposed method simultaneously estimates latent groups of the studied population and creates ordinal graphical models for the identified groups. We use a toy example to illustrate the graphical modeling of heterogeneous conditional dependencies. As shown in Fig. 1, two mixture components (A) and (B) specify two different conditional relations among ordinal variables X_1, X_2, and X_3, and the proposed model is a mixture of (A) and (B) weighted by their corresponding mixing proportions. The mixtures of ordinal graphical models divide the studied population into different groups based on conditional dependencies.

The proposed model shares the similar philosophy with the mixture of probits that generalize probit analysis (Lwin & Martin 1989). More specifically, we introduce the finite mixture of ordinal graphical models as the discretization of a latent multivariate Gaussian mixture distribution, which combines the strengths of multivariate probit models, mixture models, and Gaussian graphical models. The heterogeneous relationships between ordinal variables are characterized by the latent mixture of Gaussian graphical models, and inferred by the corresponding precision matrix.

We propose the penalized likelihood approach for the sparse estimation of the mixtures of ordinal graphical models. Specifically, we solve the maximum penalized likelihood estimation of both mixing proportions and heterogeneous precision matrices for ordinal data. It is worth noting that the model estimation of the proposed model is non-trivial and needs to overcome more significant computational challenges than traditional approaches for estimating mixtures of ordinal data such as Greene & Hensher (2003), Grün & Leisch (2008), and Breen & Luijkx (2010). On the one hand, the likelihood function consists of two different sets of latent variables: one set of latent variables in probit models to generate the ordinal data, and the other set of latent variables in Gaussian mixture models to represent the underlying group memberships. On the other hand, unlike traditional approaches, the sparse estimation imposes the non-smooth penalty function to regularize the likelihood function, which leads to solving a challenging non-convex and non-smooth optimization problem. Traditionally, mixture models employ expectation-maximization (EM) algorithms (Dempster et al. 1977) to handle these latent variables by iteratively solving the Expectation step (E step) and the Maximization step (M step). Unfortunately, with our proposed model, the M step would require the realizations of latent truncated Gaussian mixture random variables that can not be estimated or approximated when there are multiple latent groups. Hence, we need new insights to solve this significant computational challenge. To this end, we develop a generalized EM algorithm to effectively estimate the probits of latent Gaussian mixture models. In particular, without requiring the generation of random samples from the latent Gaussian mixture distribution, we use a rank-based estimation method in the M step to estimate the latent correlation matrix for each mixture.

To the best of our knowledge, we introduce the first method for estimating finite mixtures of ordinal graphical models and bridge the gap between network modeling and practice for the analysis of ordinal data in network psychometrics. We examine the performance of our proposed method in extensive simulation studies and compare its performance with that of the existing models. Moreover, we demonstrate the performance of the proposed method in a real-world sports marketing application for a large public university in the nation. In this study, 307 students responded to a survey concerning the university's Division 1 NCAA football program. Of interest to psychometricians, participants assessed 33 statements concerning their interests and attitudes related to fan avidity using a 7-point Likert scale. Participants also answered several demographics questions (e.g., age, gender, fraternity, GPA). A better understanding of the heterogeneous relations among avid fans' interests and attitudes provides the insights for practitioners to devise marketing strategy and increase revenue. However, there has been little research that explores the heterogeneous associations among respondents' interests and attitudes based on their responses to Likert scale survey questions. The proposed method addresses this important research question.

We organize the rest of the paper as follows. We introduce the proposed model and present the model estimation and computational details in Sect. 2. We demonstrate the performance of the proposed method and algorithm via simulation studies in Sect. 3, and provide an empirical application in Sect. 4. In Sect. 5, we conclude with a discussion of our contributions, limitations, and directions for future research. The additional technical details and numerical results are presented in Appendices I–VII of the supplementary material.

2. The Proposed Methodology

This section presents the model specification for finite mixtures of ordinal graphical models in Sect. 2.1, a penalized likelihood approach for model estimation in Sect. 2.2, and a generalized EM algorithm in Sect. 2.3.

2.1. Model Specification

Suppose there are p survey items/questions based on psychological constructs, behaviors, or symptoms that are assessed on an ordinal scale. Let $\boldsymbol{X} = (X_1, \ldots, X_p)'$ be the ordinal categorical variables associated with these p survey items/questions. For any $j = 1, \ldots, p$, the number of ordered categories of X_j is denoted by L_j. For ease of presentation, the ordered categories of X_j are coded as $1, \ldots, L_j$ respectively, or equivalently, $X_j \in \{1, \ldots, L_j\}$.

We assume there are N respondents independently answering these survey questions. We define their responses as the ordinal data $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$. Suppose there are K unknown heterogeneous groups among $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$. The conditional dependence relationship is invariant within the same group, but different across groups. In what follows, we introduce the finite mixture of ordinal graphical models to account for the heterogeneous conditional dependence relationships. The proposed model employs the discretization of a latent multivariate Gaussian mixture distribution and shares the similar philosophy with the mixture of probits (Lwin & Martin 1989).

To begin with, we assume that each X_j is discretized from the latent continuous counterpart Y_j for $j = 1, \ldots, p$. Given the ordered thresholds $-\infty = \theta_j^0 < \theta_j^1 < \ldots < \theta_j^{L_j} = \infty$, we have:

$$X_j = \sum_{l=0}^{L_j - 1} \mathbb{1}(Y_j \geq \theta_j^l), \quad \text{for } j = 1, \ldots, p, \tag{1}$$

where $\mathbb{1}(\cdot)$ is the indicator function. The ordinal categories are represented by the corresponding intervals between the ordered thresholds. In other words, $X_j = l$ if and only if Y_j falls in the interval $[\theta_j^{l-1}, \theta_j^l)$ for $j = 1, \ldots, p$.

The conditional dependence relationship of the observable ordinal random vector $\boldsymbol{X}$ is characterized by the joint distribution of the latent continuous random vector $\boldsymbol{Y} = (Y_1, \ldots, Y_p)'$. We assume that $\boldsymbol{Y}$ comes from the following Gaussian mixture distribution:

$$\pi_1 N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) + \cdots + \pi_K N_p(\boldsymbol{\mu}_K, \boldsymbol{\Sigma}_K), \tag{2}$$

where K is the number of mixtures, $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_K)'$ consists of the non-negative mixing proportions such that $\sum_{k=1}^K \pi_k = 1$ with $0 \leq \pi_k \leq 1$, and $N_p(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ denotes the p-dimensional Gaussian distribution with mean vector $\boldsymbol{\mu}_k$ and covariance matrix $\boldsymbol{\Sigma}_k$ for $k = 1, \ldots, K$. The probit graphical model for ordinal data (Guo et al. 2015; Suggala et al. 2017; Feng & Ning 2019) is a special example of the proposed model with $K = 1$. To ensure the identifiability of number of mixtures, we assume K be the smallest integer such that $\pi_k > 0$ for $1 \leq k \leq K$, and $(\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m) \neq (\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n)$ for $1 \leq m \neq n \leq K$ (Huang et al. 2017). Regarding the identifiability of model parameters, Allman et al. (2009) studied the generic identifiability of parameters in latent structure models where the set of all uniquely identifiable parameters has a complement of Lebesgue measure zero in the full parameter space, and in general, the generic identifiability of model parameters is sufficient for data analysis purpose. As will be shown in Sect. 3, our proposed method achieved promising estimation performance in simulation studies. We will study the generic identifiability of model parameters estimated by our proposed methods in the future. Without loss of generality, we also assume that Y_j's have unit variances, namely, the diagonal elements of $\boldsymbol{\Sigma}_k$ are 1's for $k = 1, \ldots, K$. The mixing proportions π_k's can be modeled by the function of concomitant variables (such as demographics and psychographics) to provide a more informative characterization of mixtures (Dayton & Macready 1988, Wedel 2002; DeSarbo et al. 2017).

Let $\boldsymbol{\Omega} = \{\boldsymbol{\Sigma}_1^{-1}, \ldots, \boldsymbol{\Sigma}_K^{-1}\}$ be the set of precision matrices (namely, the inverse of covariance matrices) for the latent Gaussian mixture distribution (2). In fact, $\boldsymbol{\Omega}$ can be directly translated into a finite mixture of Gaussian graphical models for latent variables $\boldsymbol{Y}$, which was studied by Ruan et al. (2011) and Lee & Xue (2018). The sparsity pattern of $\boldsymbol{\Omega}$ encodes the heterogeneous conditional dependence relationships among latent variables $\boldsymbol{Y}$. Further, given the fact that ordinal variables $\boldsymbol{X}$ are generated from the discretization of $\boldsymbol{Y}$, the sparsity pattern of $\boldsymbol{\Omega}$ also implies the heterogeneous conditional relations among ordinal variables $\boldsymbol{X}$. Therefore, we refer to (1)—(2) as the mixture of ordinal graphical models.

2.2. Model Estimation

We propose the penalized estimation procedure for estimating the finite mixtures of sparse ordinal graphical models. Recall that $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_K)'$ and $\boldsymbol{\Omega} = \{\boldsymbol{\Sigma}_1^{-1}, \ldots, \boldsymbol{\Sigma}_K^{-1}\}$. Here, we define $\boldsymbol{\Pi} = \{\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K\}$ and $\boldsymbol{\Theta} = \{\theta_j^l : j = 1, \ldots, p; l = 1, \ldots, L_j - 1\}$ as the set of ordered thresholds for the discretization of $\boldsymbol{Y}$ to generate the ordinal variables $\boldsymbol{X}$. Namely, we propose the penalized estimation procedure to estimate the model parameters $(\boldsymbol{\pi}, \boldsymbol{\Pi}, \boldsymbol{\Omega}, \boldsymbol{\Theta})$.

We first consider the joint probability density function of the observed ordinal variables $\boldsymbol{X}$ and latent variables $\boldsymbol{Y}$, denoted by $f(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\pi}, \boldsymbol{\Pi}, \boldsymbol{\Omega}, \boldsymbol{\Theta})$. After deriving $f(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\pi}, \boldsymbol{\Pi}, \boldsymbol{\Omega}, \boldsymbol{\Theta})$, we obtain the marginal probability density function for the observed ordinal variables $\boldsymbol{X}$ as:

$$f(\boldsymbol{x}; \boldsymbol{\pi}, \boldsymbol{\Pi}, \boldsymbol{\Omega}, \boldsymbol{\Theta}) = \int_{\boldsymbol{y} \in \mathbb{R}^p} f(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\pi}, \boldsymbol{\Pi}, \boldsymbol{\Omega}, \boldsymbol{\Theta}) d\boldsymbol{y}.$$

The log-likelihood function for the observed ordinal data $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ can be written as follows:

$$\ell(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N; \boldsymbol{\pi}, \boldsymbol{\Pi}, \boldsymbol{\Omega}, \boldsymbol{\Theta}) = \sum_{i=1}^{N} \log f(\boldsymbol{x}_i; \boldsymbol{\pi}, \boldsymbol{\Pi}, \boldsymbol{\Omega}, \boldsymbol{\Theta}). \tag{3}$$

We use penalized likelihood estimation (Tibshirani 1996; Fan & Li 2001; Fan et al. 2014) for estimating the mixture of sparse ordinal graphical models as follows:

$$\max_{(\boldsymbol{\pi}, \boldsymbol{\Pi}, \boldsymbol{\Omega}, \boldsymbol{\Theta})} \sum_{i=1}^{N} \log f(\boldsymbol{x}_i; \boldsymbol{\pi}, \boldsymbol{\Pi}, \boldsymbol{\Omega}, \boldsymbol{\Theta}) - \lambda \sum_{k=1}^{K} \|\boldsymbol{\Sigma}_k^{-1}\|_{1,\text{off}}, \tag{4}$$

where $\lambda \sum_{k=1}^{K} \|\boldsymbol{\Sigma}_k^{-1}\|_{1,\text{off}}$ is the penalty function that is defined as sum of absolute values of the off-diagonal entries of $\boldsymbol{\Sigma}_k^{-1}$. Here, the entry-wise ℓ_1 norm $\|\boldsymbol{\Sigma}_k^{-1}\|_{1,\text{off}}$ encourages the sparse estimation of precision matrices, and the penalization parameter λ controls the level of sparsity in the precision matrices. We choose λ using the cross validation based on the conditional likelihood given the estimates of latent variables.

The penalized estimation procedure posts a significant computational challenge to solve the non-convex and non-smooth problem (4) effectively and efficiently. In Sect. 2.3, we will propose a new generalized EM algorithm to address the computational challenge of (4).

What remains is the explicit derivation of the joint probability density function. Note that

$$f(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\pi}, \boldsymbol{\Pi}, \boldsymbol{\Omega}, \boldsymbol{\Theta}) = f(\boldsymbol{y}; \boldsymbol{\pi}, \boldsymbol{\Pi}, \boldsymbol{\Omega}) f(\boldsymbol{x}|\boldsymbol{y}; \boldsymbol{\Theta}) = f(\boldsymbol{y}; \boldsymbol{\pi}, \boldsymbol{\Pi}, \boldsymbol{\Omega}) \prod_{j=1}^{p} f(x_j|y_j; \boldsymbol{\Theta}), \tag{5}$$

where we have used the fact that each coordinate of $\boldsymbol{y}$ is discretized independently. In what follows, we derive the explicit forms of $f(\boldsymbol{y}; \boldsymbol{\pi}, \boldsymbol{\Pi}, \boldsymbol{\Omega})$ and $f(\boldsymbol{x}|\boldsymbol{y}; \boldsymbol{\Theta})$ respectively.

Let $\phi(\boldsymbol{y}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ be the probability density function of the p-dimensional Gaussian distribution with mean vector $\boldsymbol{\mu}_k$, covariance matrix $\boldsymbol{\Sigma}_k$, and precision matrix $\boldsymbol{\Sigma}_k^{-1}$ for $k = 1, \ldots, K$. Note that $\boldsymbol{y}$ follows the multivariate Gaussian mixture distribution, namely,

$$f(\boldsymbol{y}; \boldsymbol{\pi}, \boldsymbol{\Pi}, \boldsymbol{\Omega}) = \sum_{k=1}^{K} \pi_k \phi(\boldsymbol{y}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k). \tag{6}$$

Next, we will study the conditional distribution function $f(\boldsymbol{x}|\boldsymbol{y}; \boldsymbol{\Theta})$. For $j = 1, \ldots, p$, since $x_j = \sum_{l=0}^{L_j-1} \mathbb{1}(y_j \geq \theta_j^l)$, we know that:

$$f(x_j|y_j; \boldsymbol{\Theta}) = \mathbb{1}(y_j \in [\theta_j^{x_j-1}, \theta_j^{x_j})). \tag{7}$$

Combining (5) and (7), we immediately obtain that:

$$f(\boldsymbol{x}|\boldsymbol{y}; \boldsymbol{\Theta}) = \prod_{j=1}^{p} \mathbb{1}(y_j \in [\theta_j^{x_j-1}, \theta_j^{x_j})) = \mathbb{1}(\boldsymbol{y} \in C(\boldsymbol{x}, \boldsymbol{\Theta})), \tag{8}$$

where $C(\boldsymbol{x}, \boldsymbol{\Theta})$ is defined as the hyper-cube $[\theta_1^{x_1-1}, \theta_1^{x_1}) \times \ldots \times [\theta_p^{x_p-1}, \theta_p^{x_p})$ associated with $\boldsymbol{x}$.

Given (5), (6) and (8), the joint probability density function can be written as follows:

$$f(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\pi}, \boldsymbol{\Pi}, \boldsymbol{\Omega}, \boldsymbol{\Theta}) = \sum_{k=1}^{K} \pi_k \phi(\boldsymbol{y}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \cdot \mathbb{1}(\boldsymbol{y} \in C(\boldsymbol{x}, \boldsymbol{\Theta})). \tag{9}$$

Therefore, we maximize the following explicit penalized likelihood function to obtain model parameters $(\boldsymbol{\pi}, \boldsymbol{\Pi}, \boldsymbol{\Omega}, \boldsymbol{\Theta})$:

$$\max_{(\boldsymbol{\pi}, \boldsymbol{\Pi}, \boldsymbol{\Omega}, \boldsymbol{\Theta})} \sum_{i=1}^{N} \log \int_{\boldsymbol{y}_i \in \mathbb{R}^p} \left[\sum_{k=1}^{K} \pi_k \phi(\boldsymbol{y}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \cdot \mathbb{1}(\boldsymbol{y}_i \in C(\boldsymbol{x}_i, \boldsymbol{\Theta})) \right] d\boldsymbol{y}_i - \lambda \sum_{k=1}^{K} \|\boldsymbol{\Sigma}_k^{-1}\|_{1,\text{off}}. \tag{10}$$

In the next subsection, we will present the computational details to efficiently maximize (10).

2.3. *Computation*

We introduce the latent random variables $\boldsymbol{z}_i = (z_{i1}, \ldots, z_{iK})$, $i = 1, \ldots, N$, satisfying that

$$z_{ik} = \begin{cases} 1 & \text{if } \boldsymbol{y}_i \text{ belongs to the } k\text{-th group}, \\ 0 & \text{otherwise}. \end{cases} \tag{11}$$

Here, the latent variables $\boldsymbol{z}_i$ identify the respondent i's unobserved group membership.

The EM algorithm (Dempster et al. 1977) provides a powerful tool to deal with latent variables in mixture models. The EM algorithm not only provides the parameter estimates for different groups but also estimates posterior probabilities which can be used to find the membership of each observation. The EM algorithm can leverage a parallel implementation to decrease the computing time it takes to estimate parameters and potentially handle the larger problems. Moreover, Dwivedi et al. (2018) recently provided theoretical guarantees for EM algorithms when applied to misspecified Gaussian mixture models (e.g. with an under-specified number of components).

Following the spirit of the EM algorithm, we view the collected ordinal data $\boldsymbol{x}_i, i = 1, \ldots, N$ to be incomplete, and treat the latent variables as "missing data". Different from the traditional mixture models, our penalized estimation problem (10) includes two different sets of latent variables: one set of latent variables $\boldsymbol{y}_i$ in probit models to generate the ordinal data, and the other set of latent variables z_i in the latent Gaussian mixture models to specify the corresponding group memberships. It is possible to generate random samples from the truncated multivariate normal distribution in the M step, but, as pointed out in Guo et al. (2015), the computational cost of the Gibbs sampler is extremely high even when p is of moderate size. Under the mixture model setting, the M step depends on the estimated conditional expectations in the E step. We need to run the Gibbs sampler at every iteration, and the total computational costs would be significantly higher. Thus, it is not practical to use the MCMC approach to construct the empirical conditional second moment in our case. Moreover, unlike traditional approaches, the sparse estimation imposes the non-smooth penalty function to regularize the likelihood function, which leads to solving a challenging non-convex and non-smooth optimization problem.

To address the computational challenges for solving (10), we propose the generalized EM algorithm. Note that (10) depends on the parameters $(\boldsymbol{\pi}, \boldsymbol{\Pi}, \boldsymbol{\Omega}, \boldsymbol{\Theta})$ and two sets of latent variables. Among them, $\boldsymbol{\Theta}$ has a closed-form estimator $\hat{\boldsymbol{\Theta}} = \{\hat{\theta}_j^l : j = 1, \ldots, p; l = 1, \ldots, L_j - 1\}$, which can be explicitly derived as follows:

$$\hat{\theta}_j^l = \begin{cases} -\infty & \text{if } l = 0, \\ \Phi^{-1}\left(\frac{1}{N}\sum_{i=1}^N \mathbb{1}(x_{ij} < l + 1)\right) & \text{if } l = 1, 2, \ldots, L_j - 1, \\ +\infty & \text{if } l = L_j, \end{cases} \tag{12}$$

for $j = 1, \ldots, p$, where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution and $\Phi^{-1}(\cdot)$ is the inverse of $\Phi(\cdot)$. Given the closed-form estimate of $\boldsymbol{\Theta}$, we will iteratively solve the E step and the M step to obtain the estimates of the remaining parameters $(\boldsymbol{\pi}, \boldsymbol{\Pi}, \boldsymbol{\Omega})$. For iteration $t = 0, 1, \ldots$, given the current estimates of parameters $(\boldsymbol{\pi}^{(t)}, \boldsymbol{\Pi}^{(t)}, \boldsymbol{\Omega}^{(t)})$, the E step gives an approximate estimate of the conditional expectations of z_{ik}, and the M step obtains all the updated estimates of parameters $(\boldsymbol{\pi}^{(t+1)}, \boldsymbol{\Pi}^{(t+1)}, \boldsymbol{\Omega}^{(t+1)})$.

In what follows, we present the details about the E step and the M step of the proposed generalized EM algorithm. Given the "complete" data $\{(\boldsymbol{x}_i, \boldsymbol{y}_i, z_i), i = 1, \ldots, N\}$, the "complete" log-likelihood function is written as:

$$\ell^{\text{cmp}} = \sum_{i=1}^N \left[\sum_{k=1}^K z_{ik}(\log \pi_k + \log \phi(\boldsymbol{y}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\right] \mathbb{1}(\boldsymbol{y}_i \in C(\boldsymbol{x}_i, \hat{\boldsymbol{\Theta}})), \tag{13}$$

and the "complete" ℓ_1-penalized log-likelihood function becomes:

$$\mathcal{L}^{\text{cmp}} = \ell^{\text{cmp}} - \lambda \sum_{k=1}^K \|\boldsymbol{\Sigma}_k^{-1}\|_{1,\text{off}}. \tag{14}$$

__E step__: Let $\pi_k^{(t)}$, $\boldsymbol{\mu}_k^{(t)}$, and $\boldsymbol{\Sigma}_k^{-1^{(t)}}$ be the estimate of π_k, $\boldsymbol{\mu}_k$, and $\boldsymbol{\Sigma}_k^{-1}$ for k at the tth iteration. In the E step of the $(t+1)$th iteration, we compute the conditional expectation of z_{ik} given current estimates $\pi_k^{(t)}$, $\boldsymbol{\mu}_k^{(t)}$, and $\boldsymbol{\Sigma}_k^{-1(t)}$ for $k = 1, \ldots, K$. From Bayes' rule, the conditional expectation of z_{ik} is of the following form:

$$\gamma_{ik}^{(t+1)} = \frac{\pi_k^{(t)} \phi(\boldsymbol{y}_i | \boldsymbol{\mu}_k^{(t)}, \boldsymbol{\Sigma}_k^{(t)})}{\sum_{k'=1}^{K} \pi_{k'}^{(t)} \phi(\boldsymbol{y}_i | \boldsymbol{\mu}_{k'}^{(t)}, \boldsymbol{\Sigma}_{k'}^{(t)})}. \tag{15}$$

Although $\gamma_{ik}^{(t+1)}$ cannot be directly estimated, we use $\boldsymbol{x}_i$ to construct the surrogate as suggested by Guo et al. (2015), and provide an approximate estimate $\tilde{\gamma}_{ik}^{(t+1)}$. The approximate estimate of the conditional expectation of z_{ik} works well in numerical studies. For ease of notation, we define $\boldsymbol{\Gamma}^{(t+1)} = \{\tilde{\gamma}_{ik}^{(t+1)} : i = 1, \ldots, N; k = 1, \ldots, K\}$. These estimates are also called the membership probabilities as the output of the E step.

M step: In the M step of the $(t+1)$th iteration, we estimate the parameters $(\boldsymbol{\pi}, \boldsymbol{\Pi}, \boldsymbol{\Omega})$ that maximize the conditional expectation of $\mathcal{L}^{\text{cmp}} = \ell^{\text{cmp}} - \lambda \sum_{k=1}^{K} \|\boldsymbol{\Sigma}_k^{-1}\|_{1,\text{off}}$, with respect to the current conditional distribution of $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N$ given the updated membership probabilities $\boldsymbol{\Gamma}^{(t+1)}$ and current estimates $(\boldsymbol{\pi}^{(t)}, \boldsymbol{\Pi}^{(t)}, \boldsymbol{\Omega}^{(t)})$, namely:

$$Q(\boldsymbol{\pi}, \boldsymbol{\Pi}, \boldsymbol{\Omega}) = E_{\boldsymbol{y}_1,\ldots,\boldsymbol{y}_N \mid \boldsymbol{\Gamma}^{(t+1)}, (\boldsymbol{\pi}^{(t)}, \boldsymbol{\Pi}^{(t)}, \boldsymbol{\Omega}^{(t)})} \left\{ \ell^{\text{cmp}} - \lambda \sum_{k=1}^{K} \|\boldsymbol{\Sigma}_k^{-1}\|_{1,\text{off}} \right\}, \tag{16}$$

subject to the constraints that $0 \le \pi_k \le 1$ and $\sum_{k=1}^{K} \pi_k = 1$. The conditional expectation of $\mathcal{L}^{\text{cmp}}$ is also known as the Q function in the EM algorithm. More specifically, the above Q function corresponds to setting up the conditional expectation of:

$$\sum_{k=1}^{K} \left[\sum_{i=1}^{N} \tilde{\gamma}_{ik}^{(t+1)} \left(\log \pi_k + \log \phi(\boldsymbol{y}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right) \mathbb{1}(\boldsymbol{y}_i \in C(\boldsymbol{x}_i, \hat{\boldsymbol{\Theta}})) - \lambda \|\boldsymbol{\Sigma}_k^{-1}\|_{1,\text{off}} \right]. \tag{17}$$

Note that the closed-form solutions of $\boldsymbol{\pi}^{(t+1)}$ and $\boldsymbol{\Pi}^{(t+1)}$ can be solved from the maximization of the above Q function (16), but $\boldsymbol{\Omega}^{(t+1)}$ is more challenging to solve. When $K = 1$, Guo et al. (2015) proposed a MCMC approach to construct the empirical conditional second moment to estimate $\boldsymbol{\Omega}^{(t+1)}$. However, as pointed out by Guo et al. (2015), the computational costs of the MCMC approach are extremely high even when p is of moderate size. Moreover, different from Guo et al. (2015), under the mixture model setting, the M step of our EM algorithm depends on the estimated conditional expectations in the E step. We need to run the Gibbs sampler at every iteration of our EM algorithm, and the total computational costs would be significantly higher. Thus, it is not practical to use the MCMC approach to construct the empirical conditional second moment in our case. Instead, we use a rank-based ensemble method (Feng & Ning 2019) to estimate the latent correlation matrix for each mixture, without requiring the generation of random samples from the latent Gaussian mixture distribution. The main idea of rank-based ensemble method is first binarize the ordinal variable at each level and construct a set of preliminary rank-based estimators and then combine the preliminary estimators into a single estimator. Theoretical property and empirical advantage over a single preliminary estimator is shown in Feng & Ning (2019). We describe the technical details about this M step in Appendix I.

We alternate between the E step and the M step until the estimates of parameters converge. The proposed algorithm is summarized in Algorithm 1 below. The theoretical guarantees such as the local convergence for the EM algorithm with high-dimensional data were recently studied in the literature, for instance, Städler et al. (2010), Balakrishnan et al. (2017) and Lee & Xue (2018) among others. Although the objective function is non-concave and non-differentiable, the proposed EM algorithm is expected to achieve the local convergence that every cluster point in the sequence $\{(\boldsymbol{\pi}^{(t)}, \boldsymbol{\Pi}^{(t)}, \boldsymbol{\Omega}^{(t)}) : t = 0, 1, 2, \ldots\}$ is a stationary point of the objective function. Please see Section 4.1 and Definition 4.1 of Lee & Xue (2018) for more details about the local convergence and definition of cluster point.

Algorithm 1
The proposed generalized EM algorithm

---

1. Initialize $(\boldsymbol{\Gamma}^{(0)}, \boldsymbol{\pi}^{(0)}, \boldsymbol{\Pi}^{(0)}, \boldsymbol{\Omega}^{(0)})$.
2. **Repeat the following for iteration** $t = 0, 1, 2, \ldots$
3. E step: update membership probabilities $\boldsymbol{\Gamma}^{(t+1)}$ given current estimates $(\boldsymbol{\pi}^{(t)}, \boldsymbol{\Pi}^{(t)}, \boldsymbol{\Omega}^{(t)})$.
4. M step: solve $(\boldsymbol{\pi}^{(t+1)}, \boldsymbol{\Pi}^{(t+1)}, \boldsymbol{\Omega}^{(t+1)})$ from:

$$\max_{(\boldsymbol{\pi}, \boldsymbol{\Pi}, \boldsymbol{\Omega})} Q(\boldsymbol{\pi}, \boldsymbol{\Pi}, \boldsymbol{\Omega}) = \max_{(\boldsymbol{\pi}, \boldsymbol{\Pi}, \boldsymbol{\Omega})} E_{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N \mid \boldsymbol{\Gamma}^{(t+1)}, (\boldsymbol{\pi}^{(t)}, \boldsymbol{\Pi}^{(t)}, \boldsymbol{\Omega}^{(t)})} \left\{ \ell^{\text{cmp}} - \lambda \sum_{k=1}^{K} \|\boldsymbol{\Sigma}_k^{-1}\|_{1,\text{off}} \right\}$$

   subject to the constraint that $\sum_{k=1}^{K} \pi_k = 1$.
5. **Until convergence.**

We randomly sample the initial values of $\gamma_{ik}^{(0)}$ from the uniform distribution on [0, 1], and each vector $\boldsymbol{\gamma}_i^{(0)}$ is normalized to achieve the sum-to-one constraint that $\sum_{k=1}^{K} \gamma_{ik}^{(0)} = 1$ for any i. Then, we use the M step to obtain the initial values $\boldsymbol{\pi}^{(0)}$, $\boldsymbol{\Pi}^{(0)}$, and $\boldsymbol{\Omega}^{(0)}$.

After obtaining the parameter estimates, we assign the respondent i to the derived group according to corresponding estimated posterior probabilities. In practice, we need to choose the number of groups (i.e., K). The traditional Bayesian information criterion does not work well for the selection of K, and we discuss the details in Sect. 5. We will explore a data-driven approach to properly choose K in the future.

3. Simulation Studies

In this section, we present simulation studies to examine the numerical performance for estimating mixtures of ordinal graphical models with respect to differing factors across various settings. We consider four different factors: the number of respondents ($N = 100, 200$), the number of items/questions ($p = 30, 50$), the mixing proportions $\left(\boldsymbol{\pi} = \left(\frac{1}{2}, \frac{1}{2}\right), \left(\frac{1}{3}, \frac{2}{3}\right)\right)$, and the graph structures (Neighbor Chain Graph/Random Graph, Neighbor Chain Graph/Stochastic Block Graph). We choose these factors with two levels to show the numerical performance of our proposed method in different simulation settings and to represent potential various real-world applications. Here, we focus on the case when the number of latent clusters is 2, i.e., $K = 2$. We present the case when the number of latent clusters is 3, i.e. $K = 3$ in the Appendix II. In addition, we concentrate on the scenario where the number of ordinal levels for survey items/questions is 5, i.e., $L = 5$ in the main article, but we also examined the performance of the proposed method when the number of ordinal levels for survey items/questions is 7, i.e. $L = 7$ and it is presented in the Appendix III.

To investigate the performance of the proposed mixture of ordinal graphical models, we compare our proposed method with (i) the naive finite mixture of Gaussian graphical model

(Ruan et al. 2011) and (ii) the oracle method as in Guo et al. (2015). The naive finite mixture of Gaussian graphical model represents the case when we naively apply Gaussian graphical models to ordinal synthetic data sets. The oracle method uses the oracle information about the true group assignment and the latent continuous data, and it directly applies the graphical Lasso algorithm to the latent continuous data in each group. Although the oracle method never happens with real data, it serves as an ideal benchmark for comparison purposes in simulation studies. We would expect that our proposed model outperforms the naive finite mixture of Gaussian graphical model and shows comparable performance with the oracle method. It is expected that the oracle method would perform the best since it uses true group assignment and latent continuous data. But with good clustering results, we expect our model shows comparable performance with the oracle method. In addition, we have also compared our method with the probit graphical model (PGM) (Guo et al. 2015; Feng & Ning 2019) and the results are presented in the Appendix IV. The probit graphical model represents the case when we ignore the underlying mixture in synthetic data sets. We expect that our method shows better performance than the probit graphical model because the probit graphical model is not designed to handle the heterogeneity in the population.

To evaluate the parameter estimation and recovery of graphical structures, we define

- The average Frobenius norm loss (AFL) for the estimation of precision matrices:

$$\begin{aligned}\text{AFL} &= \frac{1}{K}\sum_{k=1}^{K}\|\hat{\boldsymbol{\Sigma}}_k^{-1} - \boldsymbol{\Sigma}_k^{-1}\|_{\text{F}}\\ &= \frac{1}{K}\sum_{k=1}^{K}\sqrt{\sum_{i,j}\left(\hat{\boldsymbol{\Sigma}}_k^{-1}(i,j) - \boldsymbol{\Sigma}_k^{-1}(i,j)\right)^2},\end{aligned}$$

- The root average squared error (RASE) for the estimation of mixing proportions:

$$\text{RASE}_\pi = \sqrt{\frac{1}{K}\sum_{k=1}^{K}(\hat{\pi}_k - \pi_k)^2},$$

- The average true positive rate (ATPR):

$$\text{ATPR} = \frac{1}{K}\sum_{k=1}^{K}\text{TPR}_k = \frac{1}{K}\sum_{k=1}^{K}\frac{\text{TP}_k}{\text{TP}_k + \text{FN}_k},$$

- The average false positive rate (AFPR):

$$\text{AFPR} = \frac{1}{K}\sum_{k=1}^{K}\text{FPR}_k = \frac{1}{K}\sum_{k=1}^{K}\frac{\text{FP}_k}{\text{TN}_k + \text{FP}_k},$$

where $\hat{\boldsymbol{\Sigma}}_k^{-1}$'s and $\hat{\pi}_k$'s denote the estimators of $\boldsymbol{\Sigma}_k^{-1}$'s and π_k's respectively. Here, TP_k is the number of true positives which counts true non-zero edges that are estimated as non-zero, TN_k is the number of true negatives which counts true zero edges that are estimated as zero, FP_k is the number of false positives which counts true zero edges that are estimated as non-zero, and FN_k is the number of false negatives which counts true non-zero edges that are estimated as zero. The TPR_k and FPR_k ranges from 0 to 1 and TPR_k a value of 1 indicates perfect recovery of the non-zero edges and FPR_k a value of 0 indicates a perfect recovery of the zero edges. Note that the Frobenius norm loss (AFL) measures the overall element wise error in the estimated precision matrix compared to the true precision matrix, and the root average squared error (RASE) measures the error in the estimated mixing proportions compared to the true mixing proportion used to simulate data set. Hence, the smaller values of AFL or RASE indicate the more accurate estimation of precision matrices or mixing proportions. We follow the idea of the distance-based labelling method (Yao 2015) to avoid a potential label-switching issue when we calculate the values of AFL or RASE in simulation studies.

To assess the clustering performance, we calculate the average value of the Rand Index (RI) (Rand 1971). The measure $\text{RI}(\boldsymbol{z}, \hat{\boldsymbol{z}})$ calculates the proportion of pairs whose estimated labels correspond to the true labels in terms of being assigned to the same or different groups:

$$\text{RI}(\boldsymbol{z}, \hat{\boldsymbol{z}}) = \binom{N}{2}^{-1} \sum_{i<j}^{N} (I\{z_i = z_j\} I\{\hat{z}_i = \hat{z}_j\} + I\{z_i \neq z_j\} I\{\hat{z}_i \neq \hat{z}_j\}).$$

The Rand Index measure ranges from 0 to 1 and it basically measures the similarity between two clustering results. In our simulation study, the Rand Index will be equal to 1 when estimated labels are identical to the true labels. The Adjusted Rand Index also gives similar results in our simulation studies since we have fixed number of clusters.

Regarding the data generation procedure, we first generate the latent continuous data in the following way:

1. Draw $\boldsymbol{z} \sim \text{Multinomial}(1; \pi_1, \pi_2)$
2. If $z = 1$ draw $\boldsymbol{Y} \sim N(0, \boldsymbol{\Sigma}_1)$ else draw $\boldsymbol{Y} \sim N(0, \boldsymbol{\Sigma}_2)$

Regarding the mixing proportion, we first consider when both $\pi_1 = \pi_2 = \frac{1}{2}$ and next we consider $\pi_1 = \frac{1}{3}$ and $\pi_2 = \frac{2}{3}$. Regarding the graph structure, $\boldsymbol{\Sigma}^{-1}$, the neighbor chain graph structure is constructed with 1's on the main diagonal, 0.5's on the sub-diagonal and super-diagonal, and 0 at all other entries. The random graph structure is constructed with 1's on the main diagonal and 0.25 on the off-diagonal entries with probability of 0.05. Both the neighbor chain graph and random graph are constructed similarly as in Lafit et al. (2019). For the stochastic block graph structure, the edge set is generated with the block matrix, $\begin{pmatrix} 0.150 & 0.005 \\ 0.005 & 0.075 \end{pmatrix}$ and the number of vertices in two groups are set equally. For the edge values for block graph, 0.35 is assigned when $p = 30$ and 0.25 is assigned when $p = 50$. After the graph structures are obtained, we take the inverse and re-scale them to create the covariance matrices, where the diagonal entries are 1's.

Next we obtain observed ordinal data following the similar procedure in Feng & Ning (2019). We simulate $\boldsymbol{X} = (X_1, \ldots, X_p)'$, where $X_j = \sum_{l=1}^{L_j-1} \mathbb{1}(Y_j \geq \theta_j^l)$ for $j = 1, \ldots, p$. Here, the sequence of thresholds θ_j^l are drawn uniformly on $[\Phi^{-1}((l-0.5)\frac{1}{L_j}), \Phi^{-1}((l+0.5)\frac{1}{L_j})]$ for $l = 1, \ldots, L_j - 1$. This procedure guarantees the randomness of the simulated thresholds and ensures that the difference of the number of samples at each level is not too large.

We first check the overall performance of our proposed mixture of ordinal graphical models where we obtain the average values over all 2^4 different simulation settings. We generate 50 different synthetic data sets for each simulation setting and compute the average of metrics. The results are summarized in Table 1. Our proposed method achieves much higher ATPR and lower AFPR than the naive finite mixture of Gaussian graphical models. Moreover, by comparing our method with the oracle method, we can see that the graph structure recovery performance of our method is reasonably good. Regarding the graph estimation performance, our method achieves smaller AFL compared to the naive finite mixture of Gaussian graphical models. This result tells

TABLE 1.
Overall summary of ATPR, AFPR, and AFL.

Gaussian	Oracle	Proposed
ATPR		
0.68 (0.05)	0.84 (0.04)	0.82 (0.06)
AFPR		
0.27 (0.03)	0.09 (0.03)	0.20 (0.03)
AFL		
10.9 (1.90)	3.52 (0.50)	6.63 (1.21)

The Gaussian method applies the finite mixture of the Gaussian graphical model to the ordinal data. The oracle method applies the graphical Lasso algorithm to the latent continuous data in each group. The oracle method is an ideal benchmark but not feasible in practice. The results are averaged over different scenarios of fixed level for each factor and the corresponding standard deviations are written in the parentheses.

us that our proposed method estimates the strength of conditional dependencies better than the naive finite mixture of Gaussian graphical models.

To illustrate the graph estimation performance more clearly, we choose one simulation setting and pick the case which shows the median performance among 50 repeats. We use the heatmap to provide a graphical representation of the precision matrix. Specifically, the heatmap is used to visualize the sparsity pattern of the precision matrix, where the white color denotes no conditional dependence between two ordinal variables, and the grayscale represents different strength of conditional dependence between two ordinal variables. We use the `ggplot2` package in R to plot the heatmaps of both estimated precision matrices and true precision matrices in Fig. 2. As shown in Fig. 2, we appropriately recover the random sparse graph structure in Group 1 and the neighbor chain graph structure in Group 2.

We check the performance of our proposed method in different simulation settings thoroughly and summarize the results by factor in Tables 2 and 3. In all simulation settings, our proposed method shows better performance than the naive finite mixture of Gaussian graphical models and shows satisfying results compared to the oracle method. More specifically, in the settings with larger numbers of respondents (N) and smaller numbers of items/questions (p), the proposed method shows better performance than the naive finite mixture of Gaussian graphical models and similar performance with that of the oracle model, as expected. In contrast, alternative conditions with smaller numbers of respondents (N) and larger numbers of items/questions (p) are more challenging, but our proposed method still shows better results than the naive finite mixture of Gaussian graphical models. In addition, our proposed model shows convincing performance in two different mixing proportion settings and in two different graph structure settings. These results show that our proposed method performs well in different mixing proportions settings and can recover different types of graph structures.

Next, we examine the estimation performance of mixing proportions and the clustering performance of our method in various simulation settings using $\text{RASE}_{\boldsymbol{\pi}}$ and RI respectively. The results reported in Table 4 tell us that overall our method shows convincing clustering performance and particularly performs well when the number of respondents (N) is large.

We also examine the performance of our proposed method when we assume the number of latent clusters is 3, i.e. $K = 3$. Similar to before we create diverse simulation settings and summarize both overall result and the results by factor. More details about simulation settings can be found in Appendix II. Here, our method achieves better performance of both graph structure recovery and graph estimation than the naive finite mixture of graphical models and comparable

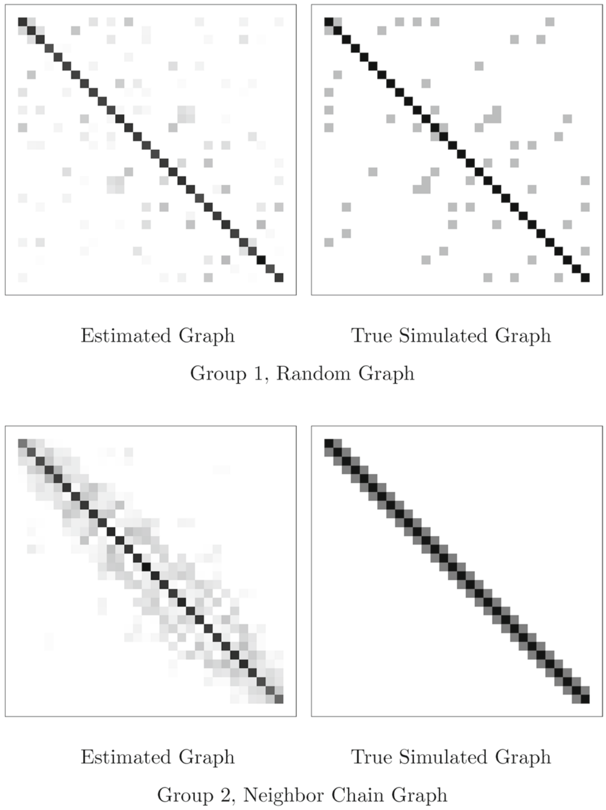


FIGURE 2.
Heatmaps of the estimated graph (left) and the heatmaps of the true simulated graph (right) when $p = 30$. The white color denotes no conditional dependence between two ordinal variables, and the grayscale represents different strength of conditional dependence between two ordinal variables.

performance to the oracle method. In addition, we obtain the reasonable clustering performance. The tables that summarize the results can be found in Appendix II.

Next, we examine the performance of the proposed method when we increase the number of ordinal levels for survey items/questions from 5 to 7. Here, we focus on one simulation setting and consider two different ordinal levels for survey items/questions: $L = 5$ and $L = 7$. Our proposed method performs well in both graph structure recovery and graph estimation even when the ordinal levels are increased to $L = 7$. Results are summarized in the tables in Appendix III.

Lastly, we also compare the numerical performance or our method with the probit graphical model and the results are presented in the Appendix IV. Here, we again focus on one simulation setting and compared numerical performance among the probit graphical model, oracle method, and the proposed method. As we expected our method shows better performance in graph structure recovery and graph estimation than the probit graphical model as the probit graphical model is not designed to handle the heterogeneity in the population. Results can be found in the Appendix IV.

TABLE 2.
The average true positive rate (ATPR) and the average false positive rate (AFPR) by the Gaussian method, oracle method, and the proposed method.

		Gaussian	Oracle	Proposed
			ATPR	
Number of respondents (N)	100	0.67 (0.02)	0.81 (0.03)	0.79 (0.05)
	200	0.7 (0.06)	0.87 (0.02)	0.84 (0.07)
Number of items/questions (p)	30	0.7 (0.05)	0.84 (0.04)	0.86 (0.06)
	50	0.66 (0.03)	0.84 (0.04)	0.78 (0.03)
Mixing proportions ($\boldsymbol{\pi}$)	$\left(\frac{1}{2}, \frac{1}{2}\right)$	0.71 (0.05)	0.85 (0.03)	0.81 (0.06)
	$\left(\frac{1}{3}, \frac{2}{3}\right)$	0.66 (0.03)	0.83 (0.04)	0.82 (0.07)
Graph structures	Neighbor chain/Random	0.68 (0.05)	0.82 (0.04)	0.8 (0.05)
	Neighbor chain/Block	0.68 (0.05)	0.85 (0.03)	0.83 (0.07)
			AFPR	
Number of respondents (N)	100	0.28 (0.02)	0.11 (0.02)	0.23 (0.03)
	200	0.27 (0.04)	0.07 (0.01)	0.18 (0.02)
Number of items/questions (p)	30	0.29 (0.03)	0.09 (0.03)	0.22 (0.03)
	50	0.26 (0.03)	0.09 (0.02)	0.18 (0.02)
Mixing proportions ($\boldsymbol{\pi}$)	$\left(\frac{1}{2}, \frac{1}{2}\right)$	0.28 (0.04)	0.1 (0.03)	0.21 (0.04)
	$\left(\frac{1}{3}, \frac{2}{3}\right)$	0.27 (0.03)	0.08 (0.02)	0.19 (0.03)
Graph structures	Neighbor chain/Random	0.28 (0.03)	0.09 (0.03)	0.2 (0.04)
	Neighbor chain/Block	0.27 (0.03)	0.09 (0.03)	0.2 (0.03)

TABLE 3.
The average Frobenius norm loss (AFL) by the Gaussian method, oracle method, and the proposed method.

		Gaussian	Oracle	Proposed
Number of respondents (N)	100	11.1 (0.93)	3.55 (0.53)	6.51 (1.25)
	200	10.7 (2.57)	3.48 (0.51)	6.75 (1.24)
Number of items/questions (p)	30	9.81 (1.60)	3.03 (0.04)	5.55 (0.39)
	50	11.9 (1.54)	4.01 (0.07)	7.71 (0.53)
Mixing proportions ($\boldsymbol{\pi}$)	$\left(\frac{1}{2}, \frac{1}{2}\right)$	10.3 (2.00)	3.54 (0.52)	7.02 (1.27)
	$\left(\frac{1}{3}, \frac{2}{3}\right)$	11.4 (1.69)	3.50 (0.52)	6.24 (1.08)
Graph structures	Neighbor chain/Random	10.7 (1.86)	3.53 (0.54)	6.56 (1.34)
	Neighbor chain/Block	11.1 (2.00)	3.5 (0.50)	6.7 (1.15)

4. An Empirical Study

In this section, we apply the proposed model to a real-world sports marketing data set. The data set was collected and analyzed by DeSarbo (2010) and DeSarbo et al. (2017), where 307 university students responded to an online questionnaire about the university's Division 1 NCAA football program. To be consistent with DeSarbo (2010) and DeSarbo et al. (2017), we refer to the university as University X or simply X. The questionnaire contains 33 questions concerning the respondents' interests and attitudes about the football program, college and professional football,

TABLE 4.
The Rand Index (RI) and root average squared error (RASE) of mixing proportions by the Gaussian method and the proposed method.

		Gaussian	Proposed
		RI	
Number of respondents (N)	100	0.53 (0.02)	0.69 (0.05)
	200	0.57 (0.02)	0.9 (0.02)
Number of items/questions (p)	30	0.55 (0.04)	0.81 (0.11)
	50	0.54 (0.02)	0.79 (0.12)
Mixing proportions ($\boldsymbol{\pi}$)	$\left(\frac{1}{2}, \frac{1}{2}\right)$	0.55 (0.05)	0.81 (0.09)
	$\left(\frac{1}{3}, \frac{2}{3}\right)$	0.55 (0.01)	0.78 (0.14)
Graph structures	Neighbor chain/Random	0.55 (0.03)	0.79 (0.12)
	Neighbor chain/Block	0.55 (0.03)	0.8 (0.12)
Overall		0.55 (0.03)	0.8 (0.11)
		$\text{RASE}_{\boldsymbol{\pi}}$	
Number of respondents (N)	100	0.44 (0.05)	0.11 (0.05)
	200	0.44 (0.07)	0.04 (0.01)
Number of items/questions (p)	30	0.44 (0.07)	0.07 (0.04)
	50	0.44 (0.05)	0.08 (0.06)
Mixing proportions ($\boldsymbol{\pi}$)	$\left(\frac{1}{2}, \frac{1}{2}\right)$	0.41 (0.04)	0.05 (0.01)
	$\left(\frac{1}{3}, \frac{2}{3}\right)$	0.47 (0.06)	0.09 (0.06)
Graph structures	Neighbor chain/Random	0.45 (0.07)	0.08 (0.05)
	Neighbor chain/Block	0.43 (0.05)	0.07 (0.05)
Overall		0.44 (0.06)	0.07 (0.05)

and other sports as well as their demographics (e.g., age, gender, fraternity, GPA). Respondents were asked to report how much they agree or disagree with each statement using a 7-point Likert scale. A higher score suggests a higher level of agreement with the statement. The selection of these statements was based on an extensive literature review, in-depth interviews, and pretesting. See Appendix V for the details about these 33 statements.

It is worth pointing out that the 33 Likert scale survey questions includes the self measurements of fan avidity, which refers to the level of interest, involvement, passion, enthusiasm, and loyalty a fan exhibits to an entity (e.g., a sport team) (DeSarbo 2010). In psychometrics and marketing research, it is important to understand the heterogeneity of respondents' interests and attitudes related to fan avidity. Recognizing such heterogeneity is a fruitful area of research with potential major implications on marketing strategy, operations, and revenue.

However, there are very few research that explores the heterogeneous associations among respondents' interests and attitudes based on their responses to Likert scale survey questions. The proposed method addresses this research question. Let $\boldsymbol{x}_i$ be the i-th respondent's responses to the 33 Likert scale survey questions for $i = 1, \ldots, 307$. DeSarbo et al. (2017) has analyzed this data, which proposed a constrained segmentation methodology to examine the relationship between fan avidity and its various behavioral manifestations. They identified two distinct subgroups, thus, we expect there are two distinct subgroups and use $K = 2$. This choice is also supported by computing the measures of fit (e.g., mean absolute error, squared Spearman's correlation, etc.) between the observed ordinal data and the predicted values, where $K = 2$ obtains the best fit measures.

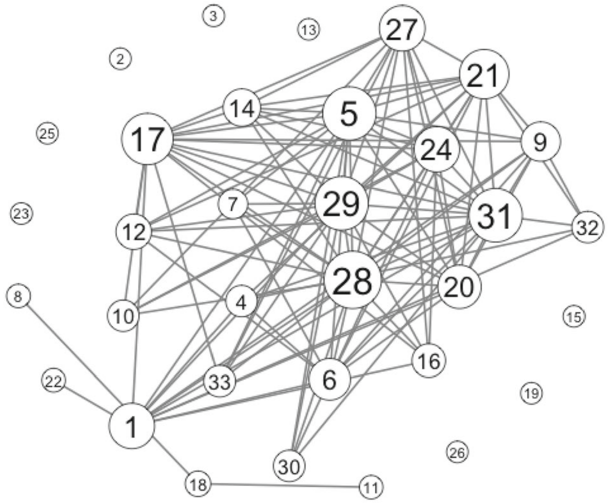
TABLE 5.
Summary of network statistics and demographics for two estimated groups.

Group ID	1	2
N	96	211
Maximum degree	18	14
Total degree	108	82
Degree centrality	6.55	4.97
Betweenness centrality	8.48	8.91
Fan avidity of X football	5.97	5.36
Female (%)	40.62%	35.07%
Fraternity (%)	27.08%	21.33%
GPA	3.52	3.49
Age	20.23	20.31

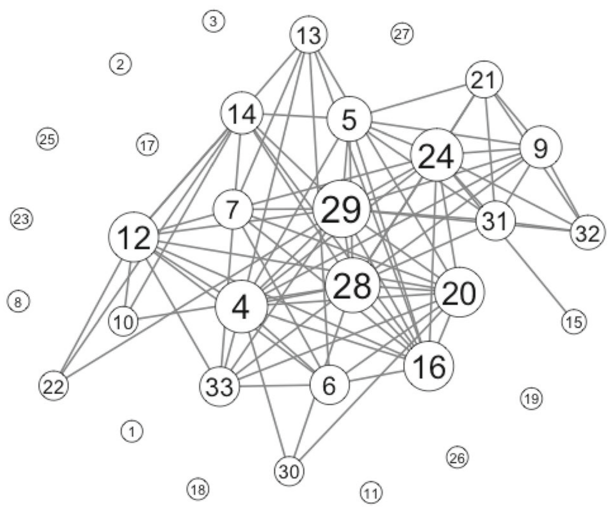
With the observed ordinal data ($\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{307}$), the proposed method identifies two heterogeneous groups with 96 and 211 respondents respectively, denoted by Group 1 and Group 2. The mixing proportions (i.e., the sizes of the derived groups, π_k) are 0.31 for Group 1 and 0.69 for Group 2. We summarize the descriptive network statistics and demographics for two estimated groups in Table 5. The proposed method constructs a graphical model for each group, which delineates the estimated group-level conditional dependencies among the 33 ordinal variables. We use the stability selection (Meinshausen & Bühlmann 2010; Xue et al. 2012) based on the resampling technique to select the stable edges that appear 70% out of 100 graphs estimated from the resampled data. The estimated stable edges of both graphical models are shown in Fig. 3, where the node size is proportional to its degree (i.e., the number of edges connected to a node).

In what follows, we first compare the characteristics of Group 1 and Group 2 in terms of their estimated graph structures, psychographic profiles, and demographics, and then focus on the heterogeneous conditional relations between the 4th survey question on the fan avidity about X football and other interests/attitudes. Overall, we find two heterogeneous conditional dependence structures in Group 1 and Group 2.

As shown in Fig. 3 and Table 5, two estimated graphs have different network structures. The total degree (i.e., the total number of edges) of Group 1's network is 108, the maximum degree (i.e., the maximum number of edges of a node) is 18, the average degree centrality is 6.55, and the average betweenness centrality is 8.48; for Group 2's network, the total degree is 82 and the maximum degree is 14, and the average degree centrality is 4.97, and the average betweenness centrality is 8.91. In addition, two estimated graphs have different hub nodes whose number of edges greatly exceeds the average. For Group 1, the hub nodes are related to professional and college football as well as sports in general, such as "avid fan of collegiate football" (Node 28), "knowledgeable football fan" (Node 5), "watch NFL games on TV" (Node 31), "watch College Football Game Day on ESPN" (Node 29), "play intramural sports" (Node 17), "avid NFL fan" (Node 21), "avid fan of big ten football" (Node 20), "play varsity sports in high school" (Node 1), "I would love a career in the sports industry" (Node 27), and "watch college football games of different teams on TV" (Node 24). For Group 2, in addition to the common hub nodes related to college and big ten football (i.e., Node 28, Node 29, Node 20, Node 24) shared with Group 1, the other hub nodes are specifically related to X football program, such as "avid X football sports fan" (Node 4), "visit websites related to X football" (Node 16), and "lose touch with what's happening around me when viewing X football game" (Node 12). Figure 4 also confirms the difference in the network structure between Group 1 and 2. Figure 4 plots the degrees of nodes in Group 1



(A) The estimated graphical model for Group 1



(B) The estimated graphical model for Group 2

FIGURE 3.
The estimated ordinal graphical models of 33 Likert scale survey questions. Nodes are scaled to degree.

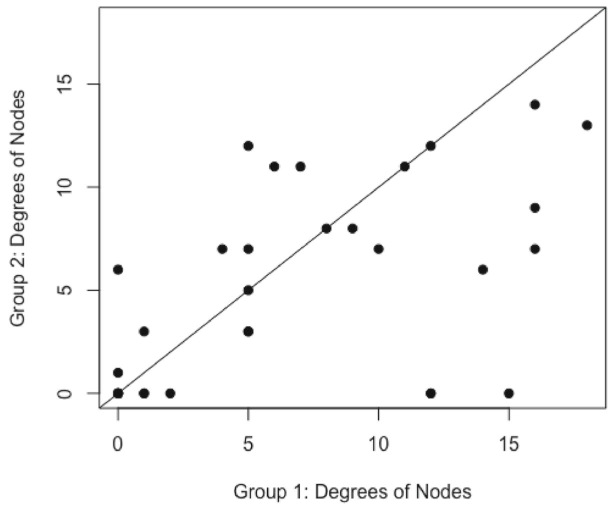


FIGURE 4.
The scatter plot of node degrees of Group 1 vs Group 2.

on the *x* axis and that of the same nodes in Group 2 on the *y* axis. The majority of dots fall far from the diagonal indicate the network structure are quite different between these two groups. We also present the centrality plots for these two groups in Fig. 5, which provide another evidence of the non-negligible difference in the network structure between the two groups. In these centrality plots, the *y* axis represents the node index, and the *x* axis represents the centrality measures (i.e., degree centrality, betweenness centrality).

The assessments of these statements are also different between Group 1 and Group 2, as summarized in Appendix VI. Out of the 33 statements, there are statistically significant differences in the average scores of 8 statements. Specifically, the members of Group 1 seem to be more avid about X football, get very frustrated and angry when the X football team does not win, lose touch with what is happening around when viewing X football game, be experts on X football, be socially active, have most friends interested in X football, imagine to be the football players on the field, and admire football players. There are no statistically significant differences in demographics (e.g., age, gender, or GPA) due to the somewhat homogeneous student population.

Next, we focus on the different conditional relations between fans' avidity about X football (i.e., Node 4, one of the 33 ordinal variables) and other interests and attitudes. The estimated conditional relations in Group 1 is shown in Fig. 6a, and the estimated conditional relations in Group 2 is shown in Fig. 6b. Because Fig. 6 only presents the conditional dependencies involving fan avidity, the relations presented in Fig. 6 is a subset of those presented in Fig. 3. For both groups, emotional identification is related to fan avidity: being an avid fan of X football is conditionally associated with getting frustrated and angry when X does not win (Node 6). Except for emotional identification, we do find more different conditional relations between the two groups. For Group 1, being an avid fan of X football is conditionally associated with being a fan of professional football (i.e., "avid NFL fan"—Node 21, "watch NFL games on TV"—Node 31) and enjoying sports more broadly (i.e., "play varsity sports in high school"—Node 1, "enjoy sports-related movies"—Node 32). For Group 2, being an avid fan of X football is conditionally associated with being an avid fan of college football (i.e., "avid fan of collegiate football"—Node 28, "avid fan of big ten football"—Node 20, "watch college football games of different teams on TV"—Node 24, "watch college football game day on ESPN"—Node 29), showing strong interests and attachment in X football and other X sports (i.e., "part of the reason of attending X was

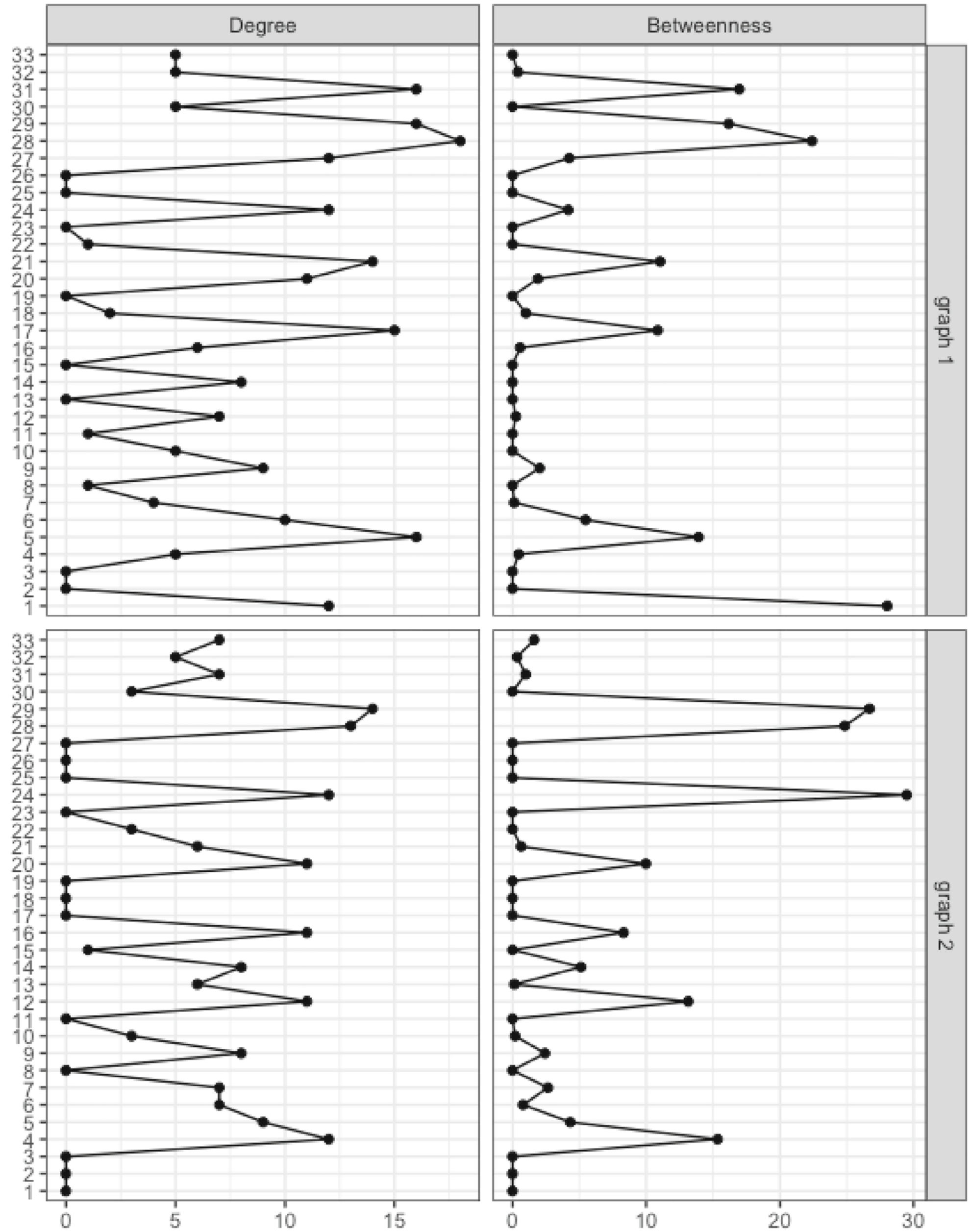
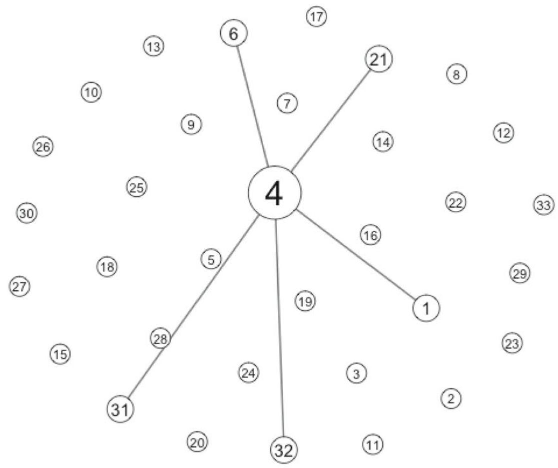
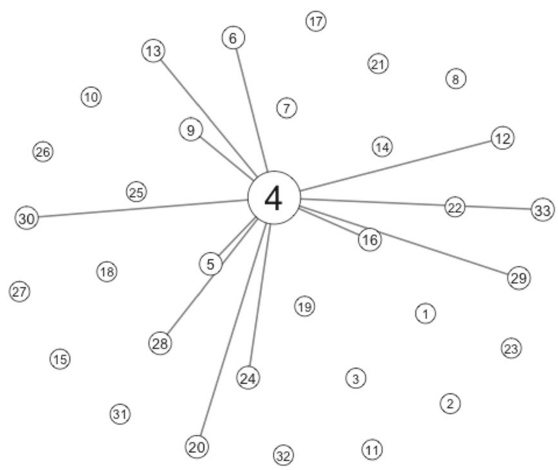


FIGURE 5.
The centrality plots of the two groups identified by the proposed model.

because of their football team"—Node 33, "lose touch with what's happening around me when viewing X football game"—Node 12, "visit websites related to X football"—Node 16,"avid fan of other X sports"—Node 30), acquiring football knowledge and admiring professionals (i.e., "knowledgeable football fan"—Node 5, "admire football players"—Node 13) and social aspect (i.e., "enjoy talking about sports with friends"—Node 9).



(A) The estimated conditional relations in Group 1.



(B) The estimated conditional relations in Group 2.

FIGURE 6.
The estimated conditional relations between X football fan avidity and others.

For comparison purpose, we have also applied the probit graphical model (PGM) (Guo et al. 2015; Feng & Ning 2019) to this dataset. We have presented the resulting graph in Appendix VII. Because the probit graphical model ignores underlying heterogeneity, the generated graph fails to capture the differences between latent groups.

5. Conclusion and Future Work

Although ordinal variables are commonly used in many areas of psychological science, the estimation of heterogeneous associations among ordinal variables has not been explored in network psychometrics. The proposed method and algorithm in this paper aim to fill this important gap in the current literature. Methodologically, we introduce the finite mixture of ordinal graphical models and propose a penalized likelihood approach to effectively estimate the heterogeneous conditional dependence relationships within ordinal data. Computationally, the proposed generalized EM algorithm effectively estimates the parameters despite the intractable likelihood function. After solving these modeling and computational challenges, we examine the performance of our proposed method and algorithm in extensive simulation studies, and demonstrate the potential usefulness in psychological science through a real application to study the interests and attitudes related to sport fan avidity. To the best of our knowledge, the proposed methodology is the first network psychometric framework to explore the heterogeneity in ordinal data.

In what follows, we discuss several limitations and research topics to extend the methodology and applicability of the proposed model. Firstly, we have assumed that the number of mixtures is known in this paper. Like most latent class modeling approaches, it is important to assess the number of mixtures for estimating the mixtures of ordinal graphical models. However, the likelihood-based model selection procedures such as the BIC can not be directly used since the likelihood function cannot be explicitly computed. We should point out that, although the parameters can be estimated by using the proposed generalized EM algorithm, the log-likelihood function (3) still requires solving the intractable high-dimensional integral over the latent variables. We will explore an effective data-driven approach to choose the appropriate number of mixtures in the future.

Secondly, as suggested in Marsman et al. (2019), it is a fundamental research topic to study a formal test of homogeneity against heterogeneity in network psychometrics. Similar to Brusco et al. (2019), we do not perform such a test when estimating the heterogeneous conditional dependence relationships of ordinal data. Under the proposed framework, it is possible to explore the hypothesis testing problem that $H_0 : K = 1$ versus $H_1 : K > 1$, which points out another important future work.

Thirdly, it is interesting to study the joint estimation of multiple ordinal graphical models. In the current literature, when observations belong to different known classes, Guo et al. (2011) and Danaher et al. (2014) studied the joint estimation of Gaussian graphical models that share certain characteristics. However, the Gaussian assumption is essential for the joint estimation procedure Guo et al. (2011), Danaher et al. (2014), which can not be used for estimating multiple ordinal graphical models. Future work may consider studying the joint estimation procedure without the presence of this Gaussian assumption.

References

Albert, J. H., & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, *88*(422), 669–679.

Allman, E. S., Matias, C., & Rhodes, J. A. (2009). Identifiability of parameters in latent structure models with many observed variables. *The Annals of Statistics*, *37*(6A), 3099–3132.

Amemiya, T. (1974). Bivariate probit analysis: Minimum chi-square methods. *Journal of the American Statistical Association*, *69*(348), 940–944.

Balakrishnan, S., Wainwright, M. J., & Yu, B. (2017). Statistical guarantees for the EM algorithm: From population to sample-based analysis. *The Annals of Statistics*, *45*(1), 77–120.

Bock, R. D., & Gibbons, R. D. (1996). High-dimensional multivariate probit analysis. *Biometrics*, *52*(4), 1183–1194.

Borsboom, D. (2008). Psychometric perspectives on diagnostic systems. *Journal of Clinical Psychology*, *64*(9), 1089–1108.

Borsboom, D., & Cramer, A. O. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, *9*, 91–121.

Borsboom, D., & Molenaar, D. (2015). Psychometrics. *International Encyclopedia of the Social & Behavioral Sciences*, *19*(2), 418–422.

Breen, R., & Luijkx, R. (2010). Mixture models for ordinal data. *Sociological Methods & Research*, *39*(1), 3–24.

Brusco, M. J., Steinley, D., Hoffman, M., Davis-Stober, C., & Wasserman, S. (2019). On Ising models and algorithms for the construction of symptom networks in psychopathological research. *Psychological Methods*, *24*(6), 735–753.

Cai, T., Liu, W., & Luo, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, *106*(494), 594–607.

Chen, S., Witten, D. M., & Shojaie, A. (2015). Selection and estimation for mixed graphical models. *Biometrika*, *102*(1), 47–64.

Cheng, J., Li, T., Levina, E., & Zhu, J. (2017). High-dimensional mixed graphical models. *Journal of Computational and Graphical Statistics*, *26*(2), 367–378.

Chib, S., & Greenberg, E. (1998). Analysis of multivariate probit models. *Biometrika*, *85*(2), 347–361.

Cox, D. R., & Wermuth, N. (1994). A note on the quadratic exponential binary distribution. *Biometrika*, *81*(2), 403–408.

Cramer, A. O., Waldorp, L. J., Van Der Maas, H. L., & Borsboom, D. (2010). Comorbidity: A network perspective. *Behavioral and Brain Sciences*, *33*(2–3), 137–150.

Dalege, J., Borsboom, D., van Harreveld, F., van den Berg, H., Conner, M., & van der Maas, H. L. (2016). Toward a formalized account of attitudes: The causal attitude network (can) model. *Psychological Review*, *123*(1), 2.

Danaher, P., Wang, P., & Witten, D. M. (2014). The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society: Series B*, *76*(2), 373–397.

Dayton, C. M., & Macready, G. B. (1988). Concomitant-variable latent-class models. *Journal of the American Statistical Association*, *83*(401), 173–178.

Dempster, A. P. (1972). Covariance selection. *Biometrics*, *28*, 157–75.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, *39*(1), 1–22.

DeSarbo, W. S. (2010). A spatial multidimensional unfolding choice model for examining the heterogeneous expressions of sports fan avidity. *Journal of Quantitative Analysis in Sports*, *6*(2), 1–24.

DeSarbo, W. S., Chen, Q., & Blank, A. S. (2017). A parametric constrained segmentation methodology for application in sport marketing. *Customer Needs and Solutions*, *4*(4), 37–55.

Dwivedi, R., Ho, N., Khamaru, K., Wainwright, M. J. & Jordan, M. I. (2018), Theoretical guarantees for the EM algorithm when applied to mis-specified gaussian mixture models. In *Proceedings of the 32nd international conference on neural information processing systems* (pp. 9704–9712).

Epskamp, S., Maris, G., Waldorp, L. J. & Borsboom, D. (2018). Network psychometrics. *The Wiley handbook of psychometric testing: A multi-disciplinary reference on survey, scale and test development* (pp. 953–986).

Epskamp, S., Rhemtulla, M., & Borsboom, D. (2017). Generalized network psychometrics: Combining network and latent variable models. *Psychometrika*, *82*(4), 904–927.

Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, *96*(456), 1348–1360.

Fan, J., Liu, H., Ning, Y., & Zou, H. (2017). High dimensional semiparametric latent graphical model for mixed data. *Journal of the Royal Statistical Society: Series B*, *79*(2), 405–421.

Fan, J., Xue, L., & Zou, H. (2014). Strong oracle optimality of folded concave penalized estimation. *The Annals of Statistics*, *42*(3), 819.

Feng, H., & Ning, Y. (2019). High-dimensional mixed graphical model with ordinal data: Parameter estimation and statistical inference. In *The 22nd international conference on artificial intelligence and statistics* (pp. 654–663).

Fried, E. I., Bockting, C., Arjadi, R., Borsboom, D., Amshoff, M., Cramer, A. O., et al. (2015). From loss to loneliness: The relationship between bereavement and depressive symptoms. *Journal of Abnormal Psychology*, *124*(2), 256.

Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, *9*(3), 432–441.

Greene, W. H., & Hensher, D. A. (2003). A latent class model for discrete choice analysis: Contrasts with mixed logit. *Transportation Research Part B: Statistical Methodology*, *37*(8), 681–698.

Grün, B., & Leisch, F. (2008). Finite mixtures of generalized linear regression models. In *Recent advances in linear models and related areas* (pp. 205–230). Springer.

Guo, J., Levina, E., Michailidis, G., & Zhu, J. (2011). Joint estimation of multiple graphical models. *Biometrika*, *98*(1), 1–15.

Guo, J., Levina, E., Michailidis, G., & Zhu, J. (2015). Graphical models for ordinal data. *Journal of Computational and Graphical Statistics*, *24*(1), 183–204.

Haslbeck, J. M., & Waldorp, L. J. (2016). mgm: Structure estimation for time-varying mixed graphical models in high-dimensional data **30**, 39–81. arXiv:1510.06871

Höfling, H., & Tibshirani, R. (2009). Estimation of sparse binary pairwise Markov networks using pseudo-likelihoods. *Journal of Machine Learning Research*, *10*, 883–906.

Huang, T., Peng, H., & Kun, Z. (2017). Model selection for Gaussian mixture models. *Statistica Sinica*, *27*(1), 147–169.

Ising, E. (1925). Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik*, *31*(1), 253–258.

Isvoranu, A.-M., van Borkulo, C. D., Boyette, L.-L., Wigman, J. T., Vinkers, C. H., Borsboom, D., et al. (2016). A network approach to psychosis: Pathways between childhood trauma and psychotic symptoms. *Schizophrenia Bulletin*, *43*(1), 187–196.

Lafit, G., Tuerlinckx, F., Myin-Germeys, I., & Ceulemans, E. (2019). A partial correlation screening approach for controlling the false positive rate in sparse Gaussian graphical models. *Scientific Reports*, *9*(1), 1–24.

Lauritzen, S. L. (1996). *Graphical models*. Oxford: Clarendon Press.

Lee, J. D., & Hastie, T. J. (2015). Learning the structure of mixed graphical models. *Journal of Computational and Graphical Statistics*, *24*(1), 230–253.

Lee, K. H., & Xue, L. (2018). Nonparametric finite mixture of Gaussian graphical models. *Technometrics*, *60*(4), 511–521.

Liu, H., Lafferty, J., & Wasserman, L. (2009). The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *Journal of Machine Learning Research*, *10*, 2295–2328.

Lwin, T., & Martin, P. (1989). Probits of mixtures. *Biometrics*, *45*(3), 721–732.

Ma, S., Xue, L., & Zou, H. (2021). Alternating direction methods for latent variable Gaussian graphical model selection. *Neural Computation*, *25*(8), 2172–2198.

Marsman, M. (2019). The idiographic ising model. PsyArXiv Preprints https://psyarxiv.com/h3ka5.

Marsman, M., Borsboom, D., Kruis, J., Epskamp, S., van Bork, R., Waldorp, L., et al. (2018). An introduction to network psychometrics: Relating Ising network models to item response theory models. *Multivariate Behavioral Research*, *53*(1), 15–35.

Marsman, M., Waldorp, L., & Borsboom, D. (2019). Towards an encompassing theory of network models. PsyArXiv Preprints https://psyarxiv.com/n98qt.

Meinshausen, N., & Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, *34*(3), 1436–1462.

Meinshausen, N., & Bühlmann, P. (2010). Stability selection. *Journal of the Royal Statistical Society: Series B*, *72*(4), 417–473.

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, *66*(336), 846–850.

Ravikumar, P., Wainwright, M. J., & Lafferty, J. D. (2010). High-dimensional Ising model selection using ℓ_1-regularized logistic regression. *The Annals of Statistics*, *38*(3), 1287–1319.

Ruan, L., Yuan, M., & Zou, H. (2011). Regularized parameter estimation in high-dimensional Gaussian mixture models. *Neural Computation*, *23*(6), 1605–1622.

Schmittmann, V. D., Cramer, A. O., Waldorp, L. J., Epskamp, S., Kievit, R. A., & Borsboom, D. (2013). Deconstructing the construct: A network perspective on psychological phenomena. *New Ideas in Psychology*, *31*(1), 43–53.

Städler, N., Bühlmann, P., & Van De Geer, S. (2010). ℓ_1-penalization for mixture regression models. *Test*, *19*(2), 209–256.

Suggala, A. S., Yang, E., & Ravikumar, P. (2017). Ordinal graphical models: A tale of two approaches. In *International conference on machine learning* (pp. 3260–3269).

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B*, *58*(1), 267–288.

von Davier, M., & Carstensen, C. H. (2007). *Multivariate and mixture distribution Rasch models: Extensions and applications*. Berlin: Springer.

Wedel, M. (2002). Concomitant variables in finite mixture models. *Statistica Neerlandica*, *56*(3), 362–375.

Wedel, M., & DeSarbo, W. S. (1995). A mixture likelihood approach for generalized linear models. *Journal of Classification*, *12*(1), 21–55.

Xue, L., & Zou, H. (2012). Regularized rank-based estimation of high-dimensional nonparanormal graphical models. *The Annals of Statistics*, *40*(5), 2541–2571.

Xue, L., Zou, H., & Cai, T. (2012). Nonconcave penalized composite conditional likelihood estimation of sparse Ising models. *The Annals of Statistics*, *40*(3), 1403–1429.

Yao, W. (2015). Label switching and its solutions for frequentist mixture models. *Journal of Statistical Computation and Simulation*, *85*(5), 1000–1012.

Yuan, M., & Lin, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika*, *94*(1), 19–35.

Manuscript Received: 15 MAR 2020
Final Version Received: 25 MAY 2021
Accepted: 28 MAY 2021
Published Online Date: 30 JUN 2021