Supplement to "Estimating Finite Mixtures of Ordinal Graphical Models"

We have organized the appendices in the following way. Appendix I describes the details about the M step of the proposed generalized EM algorithm. Appendix II evaluates the numerical performance of the proposed method when $K = 3$. Appendix III evaluates the numerical performance of the proposed method when $L = 7$. Appendix IV compares the numerical performance of the proposed method with the probit graphical model. Appendix V presents the details about the 33 survey questions for the sports fan application we present. Appendix VI summarizes the average responses to the survey questions. Appendix VII presents the aggregate level network generated by the probit graphical model.

Appendix I: Details about the M Step of the Proposed Generalized EM Algorithm

In the M step of the $(t+1)$-th iteration, we need to estimate the parameters $(\boldsymbol{\pi}, \boldsymbol{\Pi}, \boldsymbol{\Omega})$ that maximize the conditional expectation of the "complete" ℓ_1-penalized log-likelihood function $\mathcal{L}^{\text{cmp}} = \ell^{\text{cmp}} - \lambda \sum_{k=1}^{K} \|\boldsymbol{\Sigma}_k^{-1}\|_{1,\text{off}}$ given the latest membership probabilities $\boldsymbol{\Gamma}^{(t+1)}$ updated in the E step and current estimates $(\boldsymbol{\pi}^{(t)}, \boldsymbol{\Pi}^{(t)}, \boldsymbol{\Omega}^{(t)})$, which is also known as the Q function. Note that the maximization of the Q function is equivalent to the maximization of the conditional expectation of:

$$\sum_{i=1}^{N} \sum_{k=1}^{K} \tilde{\gamma}_{ik}^{(t+1)} \log \pi_k \mathbb{1}(\boldsymbol{y}_i \in C(\boldsymbol{x}_i, \hat{\boldsymbol{\Theta}})), \tag{18}$$

subject to $\sum_{k=1}^{K} \pi_k = 1$, and for $k = 1, \dots K$, the maximization of the conditional expectation of:

$$\sum_{i=1}^{N} \tilde{\gamma}_{ik}^{(t+1)} \left[\frac{1}{2} \log |\boldsymbol{\Sigma}_k^{-1}| - \frac{1}{2} (\boldsymbol{y}_i - \boldsymbol{\mu}_k)' \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{y}_i - \boldsymbol{\mu}_k) \right] \mathbb{1}(\boldsymbol{y}_i \in C(\boldsymbol{x}_i, \hat{\boldsymbol{\Theta}})) - \lambda_k \|\boldsymbol{\Sigma}_k^{-1}\|_{1,\text{off}}, \tag{19}$$

respectively.

The subproblem (18) can solved by using the method of Lagrange multipliers, which yields the closed-form update of π_k as:

$$\pi_k^{(t+1)} = \frac{1}{N} \sum_{i=1}^{N} \tilde{\gamma}_{ik}^{(t+1)}. \tag{20}$$

In the subproblem (19), $\boldsymbol{\mu}_k^{(t+1)}$ has the closed-form solution that:

$$\boldsymbol{\mu}_k^{(t+1)} = \frac{1}{N}\sum_{i=1}^{N} \tilde{\gamma}_{ik}^{(t+1)} \boldsymbol{y}_i, \tag{21}$$

which cannot be directly estimated but we use $\boldsymbol{x}_i$ to construct the surrogate and provide an approximate estimate. However, $\boldsymbol{\Sigma}_k^{-1}{}^{(t+1)}$ is more challenging to solve, and it is infeasible to generate random samples from the truncated multivariate Gaussian mixture distribution when $K > 1$. Hence, the empirical conditional second moment is unavailable. To address the computational challenge, we further reduce the subproblem (19) to a sequence of penalized estimation problems as follows:

$$\boldsymbol{\Sigma}_k^{-1}{}^{(t+1)} = \underset{\boldsymbol{\Sigma}_k^{-1}}{\arg\max} \Big\{ \log|\boldsymbol{\Sigma}_k^{-1}| - \mathrm{tr}(\boldsymbol{\Sigma}_k^{-1}\boldsymbol{R}_k) - \lambda\|\boldsymbol{\Sigma}_k^{-1}\|_{1,\mathrm{off}} \Big\}, \tag{22}$$

where $\boldsymbol{R}_k$ is the latent correlation matrix in the k-th mixture given the current estimates, for $k = 1, \ldots, K$. The above penalized estimation problem is well studied and can be efficiently solved by Friedman et al. (2008).

It remains to effectively estimate $\boldsymbol{R}_k$ without requiring the realizations of latent variables $\boldsymbol{y}_i$'s. As a promising alternative, we may use a rank-based ensemble approach (Feng & Ning 2019), whose details are presented as follows. First, we obtain the binary form of the j-th ordinal variable with respect to level $l = 1, \ldots, L_j$ as $x_{ij}^{(l)} = \mathbb{1}(x_{ij} \geq l)$, for $j = 1, \ldots, p$. Second, we calculate the preliminary rank-based estimators based on Kendall's τ as follows:

$$\hat{\tau}_{jj'}^{(l,l')} = \frac{1}{\binom{N}{2}} \sum_{1 \leq i < i' \leq N} \mathrm{sign}\{(x_{ij}^{(l)} - x_{i'j}^{(l)})(x_{ij'}^{(l')} - x_{i'j'}^{(l')})\}, \tag{23}$$

with $\mathrm{sign}(0) = 0$. Third, we derive the estimate of latent correlation between variables Y_j and $Y_{j'}$ from $\hat{\tau}_{jj'}^{(l,l')}$ by solving:

$$\hat{R}_{jj'}^{(l,l')} = F^{-1}(\hat{\tau}_{jj'}^{(l,l')}; \hat{\Delta}_j^{(l)}, \hat{\Delta}_{j'}^{(l')}), \tag{24}$$

where

$$\hat{\Delta}_j^{(l)} = \Phi^{-1}(1 - \frac{1}{N}\sum_{i=1}^{N} x_{ij}^{(l)}), \tag{25}$$

$$\hat{\Delta}_{j'}^{(l')} = \Phi^{-1}(1 - \frac{1}{N}\sum_{i=1}^{N} x_{ij'}^{(l')}), \tag{26}$$

and

$$F(s;\Delta_j,\Delta_{j'})=2\{\Phi_2(\Delta_j,\Delta_{j'}|s)-\Phi(\Delta_j)\Phi(\Delta_{j'})\} \tag{27}$$

is an invertible function with respect to correlation s. Here, we have used the fact that:

$$\tau_{jj'}^{(l,l')}:=E(\hat{\tau}_{jj'}^{(l,l')})=F(\Sigma_{jj'};\Delta_j^{(l)},\Delta_{j'}^{(l')}) \tag{28}$$

and $\Phi_2(\cdot,\cdot|s)$ is the cumulative density function of the standard bivariate Gaussian density with correlation s as in Feng & Ning (2019). Finally, for (j,j'), the rank-based ensemble estimator of the (j,j')-th entry of $\boldsymbol{R}_k$ can be constructed as:

$$\hat{R}_{jj'}=\sum_{1\le l\le L_j,1\le l'\le L_{j'}}\frac{1}{L_jL_{j'}}\hat{R}_{jj'}^{(l,l')}. \tag{29}$$

Appendix II: Numerical Performance of the Proposed Method when $K = 3$

In this section, we examine the numerical performance of our proposed method when we assume the number of latent clusters is 3, i.e. $K = 3$.

Similar to the $K = 2$ simulation study, we would like to investigate the numerical performance of our proposed method in various simulation settings. Here, we consider three different factors: the number of respondents ($N = 200, 300$), the number of items/questions ($p = 30, 50$), and the mixing proportions $\left(\boldsymbol{\pi} = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right), \left(\frac{1}{6}, \frac{2}{6}, \frac{3}{6}\right)\right)$. Regarding the three different graph structures to represent three groups, we use two different types of neighbor chain graph structures, and a two-block diagonal graph structure. The first 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 second neighbor chain graph structure is constructed with 1's on the main diagonal, 0.5's on the sub-diagonal and super-diagonal, and 0.25's on the diagonal that lies directly below and to the left of the sub-diagonal and on the diagonal that lies directly above and to the right of the super-diagonal, and 0 at all other entries. The two-block diagonal graph structure is constructed with 1's on the main diagonal and 0.35 on the first and second block of off-diagonal entries with probability of 0.1.

Regarding the data generation procedure, it follows the same procedure as $K = 2$ except now we draw $\boldsymbol{z} \sim$ Multinomial(1; π_1,π_2,π_3) and if $\boldsymbol{z} = (1, 0, 0)$ draw $\boldsymbol{Y} \sim N(0, \boldsymbol{\Sigma}_1)$ else if $\boldsymbol{z} = (0, 1, 0)$ draw $\boldsymbol{Y} \sim N(0, \boldsymbol{\Sigma}_2)$ else draw $\boldsymbol{Y} \sim N(0, \boldsymbol{\Sigma}_3)$.

We check the overall performance of our proposed mixture of ordinal graphical models where we obtain the average values over all 2^3 different simulation settings. The results are summarized in Table 1 in this appendix. Our proposed method achieves much higher ATPR and lower AFPR than the naive finite mixture of Gaussian graphical models. We can also see that the graph structure recovery performance of our method is comparable to the oracle method. Regarding the graph estimation performance, our method performs better than the naive finite mixture of Gaussian graphical models.

We create many different simulation settings to check the performance of our proposed method more realistically. The results are summarized by factor in Table 2 and Table 3 in this

TABLE 1.

Overall summary of ATPR, AFPR, and AFL. The Gaussian method applies finite mixture of Gaussian graphical model to the ordinal data set. 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 and the corresponding standard deviations are written in the parentheses.

Gaussian	Oracle	Proposed
	ATPR	
0.69 (0.03)	0.84 (0.02)	0.74 (0.02)
	AFPR	
0.26 (0.02)	0.09 (0.02)	0.17 (0.03)
	AFL	
11 (1.32)	3.68 (0.61)	6.48 (1.42)

appendix. In all simulation settings, our proposed method shows better performance than the naive finite mixture of Gaussian graphical models and shows reasonable performance even compared to the oracle method.

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 tells us that overall our method shows convincing cluster recovering performance.

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. The Gaussian method applies finite mixture of Gaussian graphical model to the ordinal data set. 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.

		Gaussian	Oracle	Proposed
			ATPR	
Number of respondents (N)	200	0.69 (0.03)	0.83 (0.02)	0.73 (0.02)
	300	0.68 (0.03)	0.85 (0.02)	0.75 (0.02)
Number of items/questions (p)	30	0.69 (0.06)	0.83 (0.02)	0.74 (0.02)
	50	0.69 (0.02)	0.86 (0.02)	0.74 (0.02)
Mixing proportions ($\boldsymbol{\pi}$)	$\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$	0.703 (0.03)	0.85 (0.02)	0.75 (0.01)
	$\left(\frac{1}{6}, \frac{2}{6}, \frac{3}{6}\right)$	0.67 (0.02)	0.83 (0.03)	0.73 (0.02)
			AFPR	
Number of respondents (N)	200	0.27 (0.02)	0.1 (0.01)	0.18 (0.02)
	300	0.25 (0.03)	0.07 (0.01)	0.15 (0.03)
Number of items/questions (p)	30	0.26 (0.03)	0.08 (0.02)	0.18 (0.03)
	50	0.26 (0.06)	0.09 (0.02)	0.16 (0.02)
Mixing proportions ($\boldsymbol{\pi}$)	$\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$	0.26 (0.03)	0.09 (0.02)	0.19 (0.02)
	$\left(\frac{1}{6}, \frac{2}{6}, \frac{3}{6}\right)$	0.26 (0.03)	0.08 (0.02)	0.15 (0.02)

TABLE 3.

The average Frobenius norm loss (AFL) by the Gaussian method, oracle method, and the proposed method. The Gaussian method applies finite mixture of Gaussian graphical model to the ordinal data set. 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.

		Gaussian	Oracle	Proposed
Number of respondents (N)	200	10.6 (0.63)	3.70 (0.67)	6.36 (1.46)
	300	11.5 (1.74)	3.66 (0.65)	6.60 (1.60)
Number of items/questions (p)	30	10.4 (1.97)	3.11 (0.03)	5.24 (0.53)
	50	11.1 (2.48)	4.25 (0.04)	7.73 (0.55)
Mixing proportions ($\boldsymbol{\pi}$)	$\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$	10.8 (1.72)	3.70 (0.66)	6.93 (1.44)
	$\left(\frac{1}{6}, \frac{2}{6}, \frac{3}{6}\right)$	11.2 (0.99)	3.66 (0.66)	6.03 (1.45)

TABLE 4.

The Rand Index (RI) and root average squared error (RASE) of mixing proportions by the Gaussian method and the proposed method. The Gaussian method applies finite mixture of Gaussian graphical model to the ordinal data set. 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.

		Gaussian	Proposed
		RI	
Number of respondents (N)	200	0.49 (0.03)	0.67 (0.07)
	300	0.51 (0.03)	0.72 (0.05)
Number of items/questions (p)	30	0.49 (0.04)	0.67 (0.05)
	50	0.51 (0.02)	0.72 (0.06)
Mixing proportions ($\boldsymbol{\pi}$)	$\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$	0.52 (0.03)	0.73 (0.02)
	$\left(\frac{1}{6}, \frac{2}{6}, \frac{3}{6}\right)$	0.48 (0.02)	0.66 (0.06)
Overall		0.5 (0.03)	0.7 (0.06)
		$\text{RASE}_{\boldsymbol{\pi}}$	
Number of respondents (N)	200	0.36 (0.08)	0.1 (0.02)
	300	0.38 (0.03)	0.11 (0.01)
Number of items/questions (p)	30	0.39 (0.06)	0.1 (0.02)
	50	0.35 (0.05)	0.11 (0.02)
Mixing proportions ($\boldsymbol{\pi}$)	$\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$	0.33 (0.04)	0.1 (0.01)
	$\left(\frac{1}{6}, \frac{2}{6}, \frac{3}{6}\right)$	0.4 (0.04)	0.12 (0.01)
Overall		0.37 (0.05)	0.11 (0.02)

Appendix III: Numerical Performance of the Proposed Method when $L = 7$

In this section, we examine the numerical performance of our proposed method when the ordinal levels for survey items/questions are increased to 7 from 5.

Here, we consider a finite mixture of neighbor chain graph structure and random graph structure, and assume the number of latent clusters is known as 2, i.e., $K = 2$. 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. Next 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 and this ensures about 60 randomly chosen edges in the graph. The mixing proportions are set to be $\pi_1 = \frac{1}{2}$ and $\pi_2 = \frac{1}{2}$. The number of respondents and the number of survey items/questions are chosen to be $N = 100$ and $p = 50$ respectively. We consider two different ordinal levels for survey items/questions: $L = 5$ and 7.

We first examine the performance of graph structure recovery using ATPR and AFPR and the estimation of precision matrices using AFL. The results are summarized in Tables 5 and 6. Our proposed method achieves stable ATPR and AFPR even when the ordinal levels are increased to 7. Moreover, by comparing our method with the oracle method, we can see that the graph structure recovery performance of our method is reasonably good.

TABLE 5.

Graph structure recovery of the oracle method and the proposed method. 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 100 repetitions and the corresponding standard deviations are written in the parentheses.

Number of Levels	Oracle	Proposed
	ATPR	
5	0.94 (0.02)	0.83 (0.05)
7	0.94 (0.02)	0.83 (0.06)
	AFPR	
5	0.14 (0.03)	0.2 (0.02)
7	0.14 (0.03)	0.2 (0.03)

TABLE 6.

The average Frobenius norm loss by the oracle method and the proposed method. 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 100 repetitions and the corresponding standard deviations are written in the parentheses.

Number of Levels	Oracle	Proposed
	AFL	
5	3.59 (0.05)	4.30 (0.12)
7	3.58 (0.05)	4.46 (0.12)

Appendix IV: Numerical Performance Comparison of the Proposed Method with the Probit Graphical Model

In this section, we compare the numerical performance of our proposed method with the probit graphical model.

Here, we consider a finite mixture of neighbor chain graph structure and two-block diagonal graph structure. 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 two-block diagonal graph structure is constructed with 1's on the main diagonal and 0.25 on the first and second block of off-diagonal entries with probability of 0.1 and this ensures about 30 randomly chosen edges in each block and overall about 60 edges in the graph. The mixing proportions are set to be $\pi_1 = \frac{1}{2}$ and $\pi_2 = \frac{1}{2}$. The number of respondents and the number of survey items/questions are chosen to be $N = 100$ and $p = 50$ respectively. The number of ordinal levels for survey items/questions is set to be $L = 5$.

Tables 7 and 8 summarize the performance of graph structure recovery and estimation. Our proposed method achieves much higher ATPR than the probit graphical model at the similar level of AFPR. Our proposed method also shows convincing graph structure recovery performance when it is compared to the oracle method. The estimation performance in terms of the Frobenius norm loss shows that our method performed better than the probit graphical model.

In this simulation, we only examine the performance of the probit graphical model under the balanced setting with $K = 2$, i.e., $\boldsymbol{\pi} = (\frac{1}{2}, \frac{1}{2})$, and our method outperforms the probit graphical model. The estimation performance of the probit graphical model is much worse under the unbalanced setting or when there is a more significant difference between the underlying true graph structures.

TABLE 7.

Graph structure recovery of the probit graphical model, oracle method, and the proposed method. 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 100 repetitions and the corresponding standard deviations are written in the parentheses.

PGM	Oracle	Proposed
	ATPR	
0.73 (0.03)	0.88 (0.07)	0.83 (0.05)
	AFPR	
0.19 (0.02)	0.08 (0.03)	0.2 (0.02)

TABLE 8.

The average Frobenius norm loss by the probit graphical model, oracle method, and the proposed method. 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 100 repetitions and the corresponding standard deviations are written in the parentheses.

PGM	Oracle	Proposed
	AFL	
4.49 (0.20)	3.95 (0.06)	4.21 (0.12)

Appendix V: Details about the 33 Likert scale survey questions.

ID	Statement of the survey questions
1	I played varsity sports in high school
2	One of my parents or siblings is a X alumnus
3	I try to keep physically fit while attending X
4	I am an avid X football sports fan
5	People who know me consider me to be a knowledgeable football fan
6	I usually get very frustrated and angry when the X football team does not win
7	When I view X football, I tend to admire the beauty of the sport
8	I tend to be a very socially active person
9	I enjoy talking about sports to my friends
10	I often imagine I am one of the football players on the field
11	People who know me would consider me to be very studious
11	People who know me would consider me to be very studious
12	When I view X football games, I get so into the action that I lose touch with what is happening around me
13	I tend to admire the X football players that compete
14	I consider myself an expert on X football
15	I usually pay attention mostly to athletes with the most notoriety
16	I frequently visit web sites related to X football
17	I play intramural sports at X
18	Most of my friends are interested in X football
19	I frequently attend fraternity/sorority parties
20	I am an avid fan of Big Ten football
21	I am an avid NFL fan
22	I have paid higher than face value for tickets to watch X football
23	I attend games for the social experience more than to watch the game action
24	I frequently watch college football games of different teams on TV
25	Compared to the average X student, I tend to date less often
26	I emphasize my studies above the social aspects at X
27	I would love a career somewhere in the sports industry
28	I am an avid fan of collegiate football
29	I frequently watch College Football Game Day on ESPN
30	I am an avid fan of other X sports
31	I frequently watch NFL football games on TV
32	I enjoy sports related movies
33	Part of the reason I chose to attend X was because of their football team

Appendix VI: The average responses to the 33 Likert scale survey questions.

Question ID	Group 1	Group 2
1	5.77	5.36
2	2.39	2.99
3	6.01	5.88
4	5.96*	5.35*
5	5.10	4.83
6	5.67*	4.84*
7	5.02	4.74
8	6.34*	5.92*
9	5.66	5.77
10	3.60*	2.95*
11	5.68	5.42
12	4.38*	3.91*
13	5.31*	4.60*
14	3.69*	3.13*
15	4.41	4.20
16	3.93	3.94
17	4.51	4.26
18	6.51*	6.21*
19	3.82	3.73
20	4.96	4.78
21	5.07	5.18
22	3.98	3.73
23	3.58	3.71
24	4.39*	5.14*
25	3.40	3.59
26	4.20	4.44
27	4.64	4.71
28	5.01	4.94
29	4.67	5.09
30	4.47	4.43
31	5.17	5.40
32	5.70	5.88
33	4.54	4.16

Note: '*'denotes the significant difference between two average responses with $\alpha = 0.05$.

Appendix VII: Empirical Results of the Probit Graphical Model

We have applied the probit graphical model (PGM) to the real data set. Figure 1 presents the network generated by the PGM. Because the PGM ignores the underlying heterogeneity of the studied population, its generated network is not able to capture the mixtures in the conditional relations among sub-populations. For example, as shown in Figure 1, Node 1 and 17 are not connected with any other nodes, and Node 17 is only connected with Node 24. However, with our proposed model, as shown in Figure 3 of the main paper, Node 1 and 17 are hubs for Group 1, and Node 24 is a hub for Group 2.

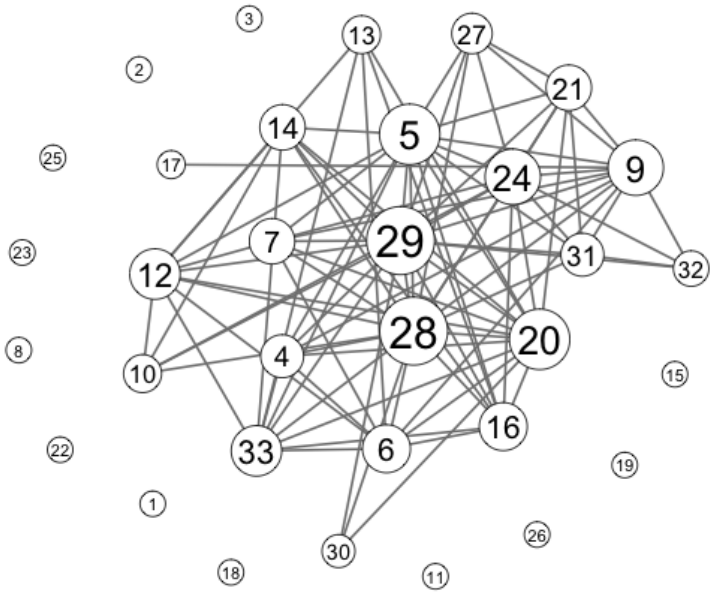

Figure 1.
The estimated conditional relations by the probit graphical model (PGM).