Supplementary Material 4: Additional Simulation Studies

1 Additional Simulation Studies

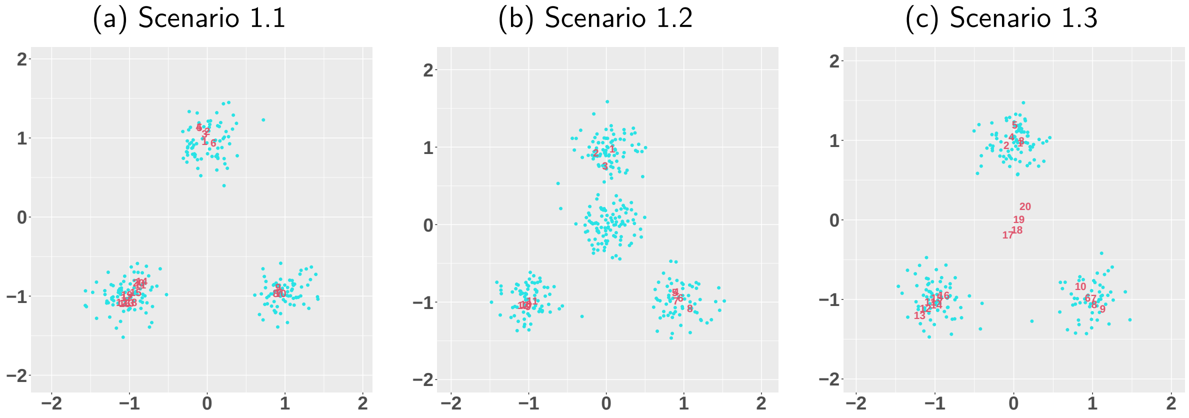
We conducted simulation studies to examine the performance of the proposed approach in various data-generating conditions. We focused on evaluating (1) the social influence parameter (δ) and (2) the overall model fit of the proposed modeling approach. Details of the study design, method, and results are provided in this section.

1.1 Study Design

We generated binary social network data of friendship $\mathbf{Y}_{n\times n}$ and binary item-level behavior data $\mathbf{X}_{n\times p}$ under our proposed modeling framework. We considered five different interaction map configurations, where respondents' interaction map configuration based on response behavior matched their social network space configuration in Scenarios 1.1 to 1.3, while the two interaction map configurations mismatch to some extent in Scenarios 2 and 3. We separately generated α, $\boldsymbol{\beta} = \{\beta_i\}_{i=1}^{p}$, and $\boldsymbol{\theta} = \{\theta_k\}_{k=1}^{n}$ from Uniform (-1, 1) for each dataset in all scenarios. We fix the weight parameter $\gamma = 1$ and set the social influence parameter $\delta = 1$.

Our simulation design was motivated by the main empirical data under investigation. For example, the number of students was between 150 and 300 in most schools. Therefore, we set the number of respondents to have a similar range in the data simulation. In addition, from the empirical data analysis, we observe that the school networks showed two to four main clusters. Following this result, we also set two to four respondent clusters in the data simulation. Additional details of each simulation scenario are described below.

Scenario 1.1 In Scenario 1.1, we considers three item and respondent clusters. As shown in Figure 1(a), one person cluster is associated with one item cluster such that: (1) respondents in Cluster 1 tend to give positive answers to item group 1 only; (2) respondents in Cluster 2 tend to give positive



(a) Scenario 1.1 (b) Scenario 1.2 (c) Scenario 1.3

Figure 1: interaction maps that illustrate Scenario 1.1 to 1.3. Red numbers and blue dots represent the latent positions of items and respondents, respectively.

responses to item group 2 only; and (3) respondents in Cluster 3 tend to give positive answers to item group 3 only. In this basic scenario, the person clusters from the social network space exactly match the person clusters from the item-response network. That is, people with only similar response behavior are assumed to be friends with each other in this scenario.

Under this setting, we generated the latent positions of individuals and behavior items as follows:

$$\mathbf{z}_k = \sum_{g=1}^{3} N\Big(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_z\Big) I\Big(g_k = g\Big) \quad \text{and} \quad \mathbf{w}_i = \sum_{g=1}^{3} N\Big(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_{w,g}\Big) I\Big(g_i = g\Big), \tag{1}$$

where $\mathbf{z}_k$ and $\mathbf{w}_i$ are the positions of individual k and item i, and g_k and g_i are the cluster/group membership for respondent k and item i, where $k = 1, 2, 3$ and $i = 1, 2, 3$. The group means and co-variances were set to as follows:

$$\boldsymbol{\mu}_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \boldsymbol{\mu}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad \boldsymbol{\mu}_3 = \begin{pmatrix} -1 \\ -1 \end{pmatrix}, \tag{2}$$

and

$$\begin{aligned} \boldsymbol{\Sigma}_z &= \begin{pmatrix} 0.2^2 & 0 \\ 0 & 0.2^2 \end{pmatrix}, \quad \boldsymbol{\Sigma}_{w,1} = \begin{pmatrix} 0.1^2 & 0 \\ 0 & 0.1^2 \end{pmatrix}, \\ \boldsymbol{\Sigma}_{w,2} &= \begin{pmatrix} 0.1^2 & -0.9 \cdot 0.1^2 \\ -0.9 \cdot 0.1^2 & 0.1^2 \end{pmatrix}, \quad \boldsymbol{\Sigma}_{w,3} = \begin{pmatrix} 0.1^2 & 0.9 \cdot 0.1^2 \\ 0.9 \cdot 0.1^2 & 0.1^2 \end{pmatrix}. \end{aligned} \tag{3}$$

Details of generating the simulation data for Scenario 1.1 are described in Algorithm 1 of the Supplementary Materials.

Scenario 1.2 In Scenario 1.2, we consider four-person clusters for the social network space. For the item-respondent network space, we consider four-person clusters and three-item clusters. As shown in Figure 1(b), Scenario 1.2 is identical to Scenario 1.1 except for the fourth person cluster in the middle of the space.

$\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, $\boldsymbol{\mu}_3$, $\boldsymbol{\Sigma}_z$ and item clusters that are identical to Scenario 1.1 were generated based on Equation (1). The group mean of the fourth person cluster was set as $\boldsymbol{\mu}_4 = \big(0, 0\big)^T$. Then, we generate the latent positions of individuals as follows:

$$\mathbf{z}_k \sim \sum_{g=1}^{4} N\Big(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_z\Big) I\Big(g_k = g\Big).$$

Scenario 1.3 In Scenario 1.3, we consider three-person clusters for the social network space as in Scenario 1.1. For the item-respondent network space, we consider three-person clusters and four-item clusters as displayed in Figure 1(c). Scenario 1.3 is identical to Scenario 1.1 except for the fourth item cluster in the middle of the space.

We generated the three person clusters that are identical to Scenario 1.1 based on Equation (1). $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, $\boldsymbol{\mu}_3$, $\boldsymbol{\Sigma}_{w,1}$, $\boldsymbol{\Sigma}_{w,2}$, and $\boldsymbol{\Sigma}_{w,3}$ are also identical to Equation (1) in Scenario 1.1. The group mean and co-variance for the fourth item cluster was set as

$$\boldsymbol{\mu}_4 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}_{w,4} = \begin{pmatrix} 0.1^2 & 0 \\ 0 & 0.1^2 \end{pmatrix}.$$

Moreover, we generated the latent positions of item $(p = 40)$ as follows:

$$\mathbf{w}_i \sim \sum_{g=1}^{4} N\Big(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_{w,g}\Big) I\Big(g_k = g\Big).$$

Scenario 2 Unlike Scenarios 1.1 to 1.3, we now assume that the social network space configuration mismatches the item-respondent network configuration. Specifically, for the social network, we consider two-person clusters, while three-person clusters and three-item clusters in the item-respondent network. We generated the item-respondent network data as in Scenario 1.1. For the social network data, we randomly generate the number of respondents between 75 and 150 per cluster. As displayed in Figures 2(a) and (b), the social network space appears quite different from the item-respondent network space, such that the person cluster in the bottom of the space does not match any of the person clusters in the item-respondent space.

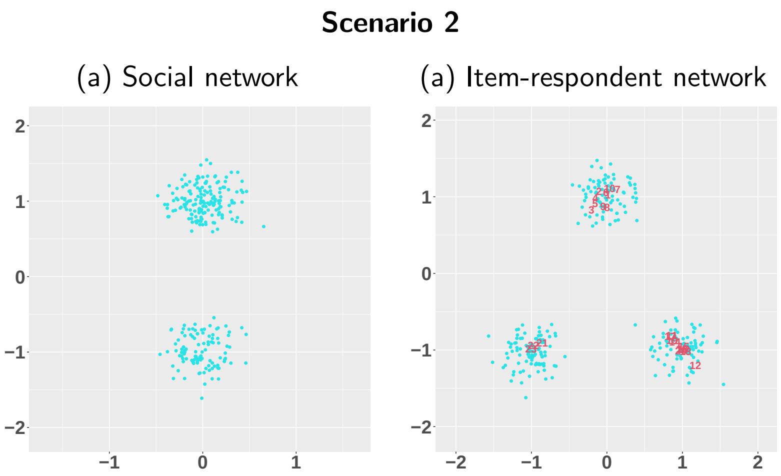


Figure 2: Social network interaction map (a) and item-respondent interaction map (b) considered in Scenario 2. Red numbers and blue dots represent latent positions for items and respondents, respectively.

The person and item positions of the item-respondent space were generated based on Eq (1). For the social network data, the latent position $\mathbf{z}'_k$ for respondent k were generated as follows:

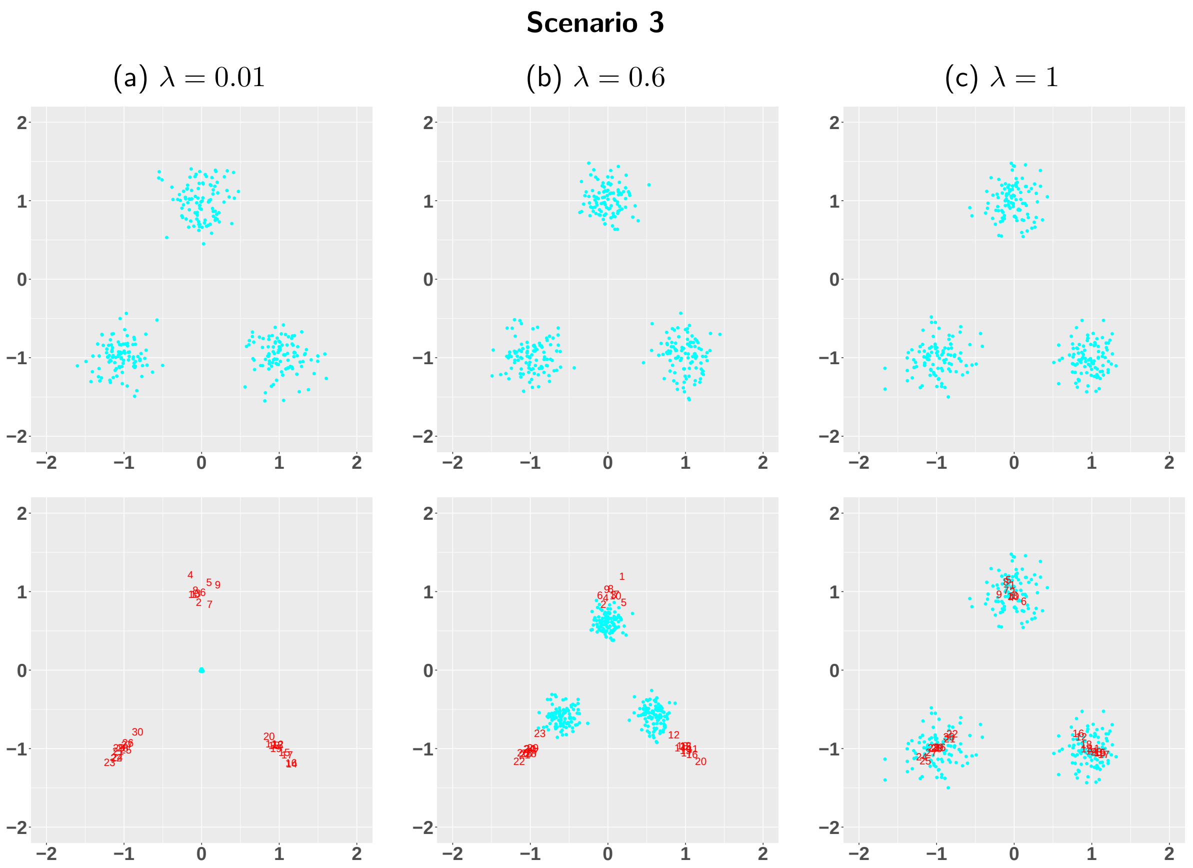
$$\mathbf{z}'_k \sim \sum_{g=1}^{2} N\Big(\boldsymbol{\mu}_{z'_g}, \boldsymbol{\Sigma}_{z'}\Big) I\Big(g_k = g\Big),$$

where

$$\boldsymbol{\mu}_{z',1} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \ \boldsymbol{\mu}_{z',2} = \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \text{ and } \boldsymbol{\Sigma}_{z'} = \begin{pmatrix} 0.2^2 & 0 \\ 0 & 0.2^2 \end{pmatrix}.$$

Details of how to generate the simulation data for Scenario 2 are described in Algorithm 2 of the Supplementary Materials.

Scenario 3 As in Scenario 2, we assume the interaction map configurations are different in the two network spaces. For the social network, we assume three-person clusters as in Scenario 1.1. For the item-respondent network, we consider three-person clusters and three-item clusters. However, this time, the person clustering structure varies across several sub-conditions, as illustrated in Figure 3. Specifically, the person positions are determined as $\mathbf{z}^*_k = \lambda \mathbf{z}_k$, where $\mathbf{z}_k$ is the person latent position from the network and λ determines the closeness of the clusters that can be interpreted as the effect of network to item response data. We consider $\lambda = (0.01,\ 0.05,\ 0.1,\ 0.2,\ 0.4,\ 0.6,\ 0.8,\ 1.0)$. When $\lambda = 1$, the person clusters are distinguished, and a dependent structure in a network apparently influences the respondent-item interactions. When $\lambda = 0.01$, the item clusters are nearly indistinguishable, and a network-dependent structure has the least effect on the respondent-item interactions. The probability of answering items



Scenario 3

(a) $\lambda = 0.01$ (b) $\lambda = 0.6$ (c) $\lambda = 1$

Figure 3: The top and bottom row are social network interaction map and item response interaction map for Scenario 3 of $\lambda = 0.01, 0.6$ and 1, respectively.

positively totally depends on the item difficulty and person characteristic parameters regardless of the dependent structures in a network. $\lambda = 1$ gives the condition equivalent to Scenario 1. Details of generating the simulation data for Scenario 3 are described in Algorithm 3 of the Supplementary Materials.

Summary Scenarios 1.1 to 1.3 show that the interaction map configuration and individuals' latent positions are the same in the spaces for item responses and social network data. Scenario 1.1 is the simplest setting where the number of person groups is the same as the number of items. This is no longer true in Scenarios 1.2 and 1.3. In Scenarios 2 and 3, the interaction map configuration is not the same in the social network and item-respondent networks. In Scenario 2, the number of person groups in the social network space is smaller than those in the item response space. In Scenario 3, the number of person groups is the same in the two spaces, but we control the influence of network dependence structure on respondent-item dependent structure. Note that there is no change in item-item dependent structures in Scenario 2 and 3.

1.2 Estimation

Per scenario, 200 datasets were generated. For each simulated dataset, we applied the proposed approach to estimate the social influence parameter with the MCMC algorithm described in Section **??**. We used 30,000 iterations, with the first 5,000 iterations being discarded as a burn-in process. From the remaining 25,000 iterations, 5,000 samples were collected at a time-space of 5 iterations. We adjusted the jumping rules for proposal distributions to achieve ideal acceptance rates (20% to 40%), identical to those in real data analysis. Details of the jumping rules are given in the Supplementary Material.

1.3 Result

Here we focused on evaluating the overall social influence parameter (δ) of the proposed approach. As a comparison, we additionally applied the network autocorrelation model and the linear-in-mean model as in the empirical data analysis.

	δ							ω	
	True	Min	25%	Median	Mean	75%	Max	Mean	# of ($\omega > 0.5$)
Scenario 1.1	1	0.86	0.95	0.98	0.98	1.01	1.13	0.98	200
Scenario 1.2	1	0.72	0.80	0.84	0.84	0.87	0.98	0.98	200
Scenario 1.3	1	0.73	0.87	0.91	0.91	0.95	1.03	0.98	200
Scenario 2	1	0.44	0.59	0.64	0.64	0.68	0.86	0.95	200

Table 1: The mean and five-number summaries of the posterior means for δ, and the mean of the posterior means ω across the 200 replicated data sets. # of ($\omega > 0.5$) indicates the number of data sets (of 200 all simulated data sets) with $\omega > 0.5$. True means data generating values in the four conditions.

Scenarios 1.1–1.3 and 2 Table 1 summarizes δ and ω estimates over 200 simulated data sets. In Scenarios 1.1 to 1.3, where the person positions perfectly match the social and item-respondent networks, the δ parameter was well recovered, ensuring the proposed estimation procedure operates appropriately. In Scenario 2 where the social network and item-respondent network were intentionally mismatched, δ was underestimated, which is not surprising given the non-trivial mismatch between the two network configurations in this condition. In all four scenarios, ω was satisfactorily recovered, providing evidence that the proposed approach can capture the presence (vs. absence) of social influence in the data being analyzed with sufficient precision, regardless of match or mismatch between the two network configurations.

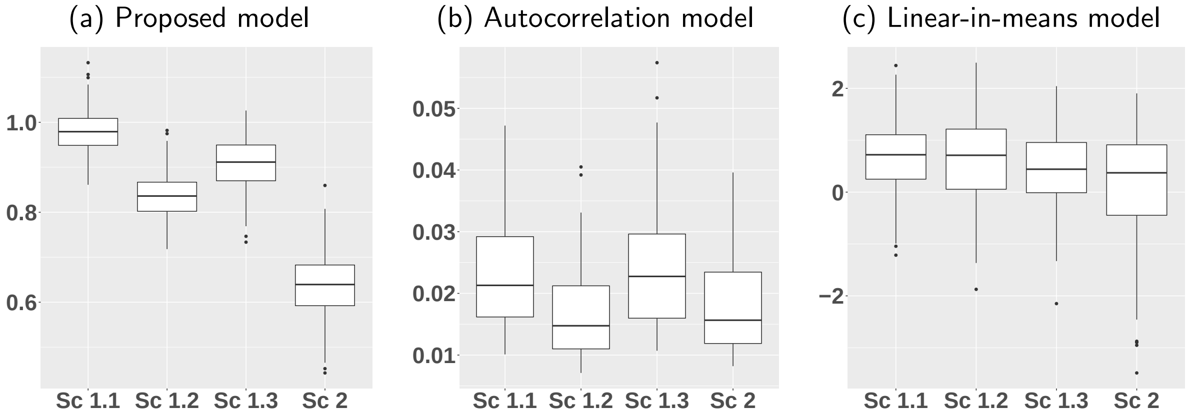


Figure 4: The boxplots of the posterior means of the social influence parameters from the three models in Scenarios 1.1 to 1.3 and 2.

Figure 4 displays the boxplots of the posterior means of the social influence parameters from the autocorrelation model and the linear-in-means model in comparison to the proposed model in Scenarios 1.1 to 1.3 and 2 over the 200 replicates.

The mean of the estimated autocorrelation parameters from the network autocorrelation model across the 200 simulated data sets were 0.0232, 0.0164, 0.0177, and 0.0240 for the four scenarios, respectively, where the credible intervals did not include 0 in all 200 data sets in all four scenarios. The patterns in the social influence effects among Scenarios 1.1, 1.2, and 1.3 appear to follow the patterns shown in the proposed approach, but the differences in the magnitude of the social influence were very small among the four scenarios. That is, even though the autocorrelation model could capture the presence of social influence to some degree, the autocorrelation model did not satisfactorily distinguish the differences in the respondent-item dependence structure (which defines social influence in the proposed framework). The mean of the estimated endogenous peer effect parameters from the linear-in-mean model across the 200 simulated data sets were 0.6513, 0.6219, 0.4426, and 0.1831 for the four scenarios, respectively. The proportion of the simulated data sets whose confidence intervals contains 0 were 51.5%, 49%, 44.5%, and 37% for the four scenarios, respectively. The medians of social influence estimates were indistinguishable among the four scenarios. The linear-in-mean model neither precisely captured the presence of social influence nor distinguished the differences in the social influence patterns defined in the proposed framework.

λ	Min	25%	Median	Mean	75%	Max
1.00	0.883	0.958	0.982	0.982	1.006	1.090
0.80	0.789	0.832	0.849	0.851	0.866	0.937
0.60	0.578	0.649	0.668	0.665	0.684	0.743
0.40	0.405	0.470	0.487	0.488	0.508	0.564
0.20	0.168	0.270	0.294	0.293	0.317	0.365
0.10	0.091	0.178	0.204	0.204	0.236	0.290
0.01	0.089	0.150	0.171	0.175	0.199	0.286

Table 2: The mean and five-number summary of the estimated posterior means for the social influence parameters from simulated datasets in Scenario 3.

Scenario 3 In Scenario 3, the social network and item-respondent network mismatch as in Scenario 2. Here we examine what happens when the degree of social influence decreased (or the degree of mismatch increased) as λ decreases from 1.0 to 0.01. Table 2 summarizes the δ parameter estimates across the 200 replicated data sets in the sub-conditions with varying λ.

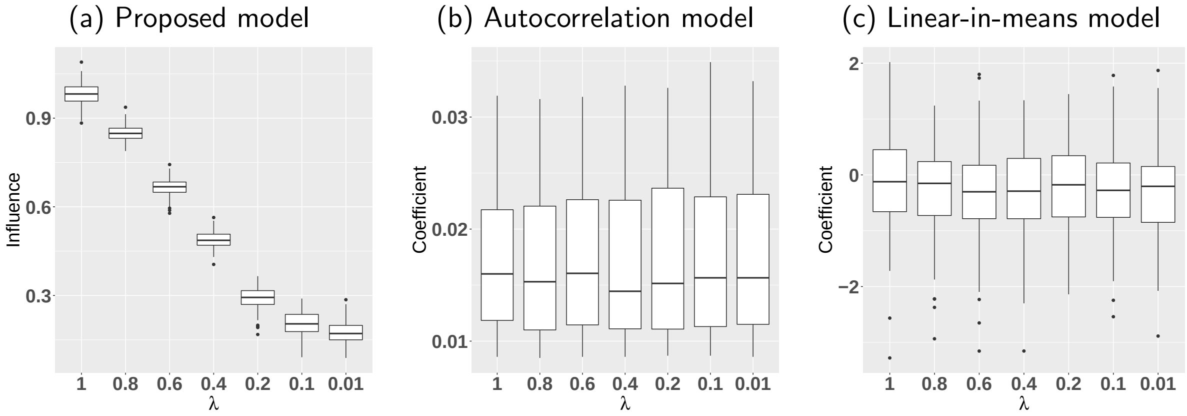


Figure 5: Boxplots of the social influence parameter estimates from (a) the proposed model, (b) the network autocorrelation model, and (c) the linear-in-means model over 200 replicates by λ values in Scenario 3.

Figure 5 shows the boxplots of the posterior means of the social influence parameters over the 200 replicates from (a) the proposed approach, (b) the network autocorrelation model, and (c) linear-in-means model. Overall, the recovery of the δ parameter of the proposed approach was satisfactory across the seven λ conditions. Although the precision appeared to decrease as λ became close to zero, the 95% HPD of δ never included zero in all 200 replicates in all λ conditions.

As we observed with Scenarios 1.1 to 1.3 and 2, the social influence parameter estimates from the autocorrelation model and the linear-in-means model did not vary much across the seven λ conditions, adding another evidence that these two alternative models did not seem to satisfactorily distinguish the differences in the respondent-item dependence structure, which defines social influence in the proposed framework.

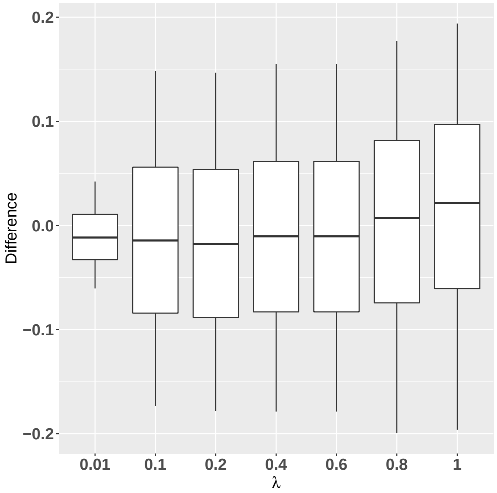


Figure 6: The boxplot of $p_{ki} - \hat{p}_{ki}$ for the LSIRM for social influence model. Outliers are suppressed.

λ	Min	25%	Median	Mean	75%	Max
0.01	-0.39	-0.05	-0.01	-0.01	0.03	0.37
0.10	-0.75	-0.14	-0.01	-0.01	0.11	0.68
0.20	-0.81	-0.14	-0.02	-0.02	0.11	0.70
0.40	-0.83	-0.14	-0.01	-0.01	0.12	0.71
0.60	-0.83	-0.14	-0.01	-0.01	0.12	0.71
0.80	-0.87	-0.15	0.01	-0.01	0.14	0.71
1.00	-0.87	-0.14	0.02	-0.00	0.16	0.72

Table 3: Mean and five-number summaries of $p_{ki} - \hat{p}_{ki}$ for the proposed model across the 200 replicates.

Model Fit Evaluation To evaluate the goodness-of-fit of the proposed approach with the simulated data, we focused on Scenario 3, which includes conditions with match ($\lambda = 1$) and mismatch ($\lambda < 1$) between the social network and the item-response network. As λ becomes closer to 1.0 (further away from 1.0), the mismatch increases. We assessed differences between the true and estimated probabilities of the positive responses, $p_{ki} - \hat{p}_{ki}$, as in the empirical study, across the 200 replicates in all λ conditions. Figure 6 shows the boxplots and Table 3 lists the five number summaries of $p_{ki} - \hat{p}_{ki}$ over 200 replicates.

The results suggest that the discrepancies between p_{ki} and $\hat{p}_{ki}$ are minimal in all λ conditions, indicating that the goodness-of-fit of the proposed approach was satisfactory even when there was a mismatch, small or large, between the social network and the item-respondent network $(\lambda < 1)$.

2 Algorithm for Generating Simulation Study 2 Data

Algorithm 1: How to Generate a Simulated Dataset for Scenario 1.1.

```
Generate $\alpha \sim$ Uniform$(-1, 1)$;
for i in 1:p do
    Generate $\beta_i \sim$ Uniform$(-1, 1)$;
    Generate $\mathbf{w}_i \sim \sum_{g=1}^{3}$ Normal$\Big(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_{w,g}\Big) I\Big(g_i = g\Big)$;
end
for k in 1:n do
    Generate $\theta_k \sim$ Uniform$(-1, 1)$;
    Generate $\mathbf{z}_k \sim \sum_{g=1}^{3}$ Normal$\Big(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_z\Big) I\Big(g_k = g\Big)$;
end
for k in 1:n do
    if $k \neq 1$ then
        for l in 1:(k-1) do
            Generate $Y_{kl} = Y_{lk} \sim$ Binomial$\Big(1, \frac{\exp(\pi_{y,kl})}{1+\exp(\pi_{y,kl})}\Big)$ where $\pi_{y,kl} = \alpha - \gamma||\mathbf{z}_k - \mathbf{z}_l||$;
        end
    end
    for i in 1:p do
        Generate $X_{ki} \sim$ Binomial$\Big(1, \frac{\exp(\pi_{x,ki})}{1+\exp(\pi_{x,ki})}\Big)$ where $\pi_{x,ki} = \beta_i + \theta_k - \delta||\mathbf{z}_k - \mathbf{w}_i||$;
    end
end
```

Algorithm 2: How to Generate a Simulated Dataset for Scenario 2.

```
Generate $\alpha \sim$ Uniform$(-1, 1)$;
for $i$ in 1:p do
    Generate $\beta_i \sim$ Uniform$(-1, 1)$;
    Generate $\mathbf{w}_i \sim \sum_{g=1}^{3}$ Normal$\Big(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_{w,g}\Big) I\Big(g_i = g\Big)$;
end
for $k$ in 1:n do
    Generate $\theta_k \sim$ Uniform$(-1, 1)$;
    Generate $\mathbf{z}_k \sim \sum_{g=1}^{3}$ Normal$\Big(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_z\Big) I\Big(g_k = g\Big)$;
    Generate $\mathbf{z}'_k \sim \sum_{g=1}^{2}$ Normal$\Big(\boldsymbol{\mu}_{g'}, \boldsymbol{\Sigma}_{z'}\Big) I\Big(g_k = g\Big)$;
end
for $k$ in 1:n do
    if $k \neq 1$ then
        for $l$ in 1:(k-1) do
            Generate $Y_{kl} = Y_{lk} \sim$ Binomial$\Big(1, \frac{\exp(\pi_{y,kl})}{1+\exp(\pi_{y,kl})}\Big)$ where $\pi_{y,kl} = \alpha - \gamma||\mathbf{z}'_k - \mathbf{z}'_l||$;
        end
    end
    for $i$ in 1:p do
        Generate $X_{ki} \sim$ Binomial$\Big(1, \frac{\exp(\pi_{x,ki})}{1+\exp(\pi_{x,ki})}\Big)$ where $\pi_{x,ki} = \beta_i + \theta_k - \delta||\mathbf{z}_k - \mathbf{w}_i||$;
    end
end
```

Algorithm 3: How to Generate a Simulated Dataset for Scenario 3.

```
Generate $\alpha \sim$ Uniform$(-1, 1)$;
for i in 1:p do
    Generate $\beta_i \sim$ Uniform$(-1, 1)$;
    Generate $\mathbf{w}_i \sim \sum_{g=1}^{3}$ Normal$\Big(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_{w,g}\Big) I\Big(g_i = g\Big)$;
end
for k in 1:n do
    Generate $\theta_k \sim$ Uniform$(-1, 1)$;
    Generate $\mathbf{z}_k \sim \sum_{g=1}^{3}$ Normal$\Big(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_z\Big) I\Big(g_k = g\Big)$;
end
for k in 1:n do
    if $k \neq 1$ then
        for l in 1:(k-1) do
            Generate $Y_{kl} = Y_{lk} \sim$ Binomial$\Big(1, \frac{\exp(\pi_{y,kl})}{1+\exp(\pi_{y,kl})}\Big)$ where $\pi_{y,kl} = \alpha - \gamma||\mathbf{z}_k - \mathbf{z}_l||$;
        end
    end
    for i in 1:p do
        Generate $X_{ki} \sim$ Binomial$\Big(1, \frac{\exp(\pi_{x,ki})}{1+\exp(\pi_{x,ki})}\Big)$ where $\pi_{x,ki} = \beta_i + \theta_k - \delta||\lambda\mathbf{z}_k - \mathbf{w}_i||$;
    end
end
```

3 Selected Interaction Maps from Additional Simulation Studies

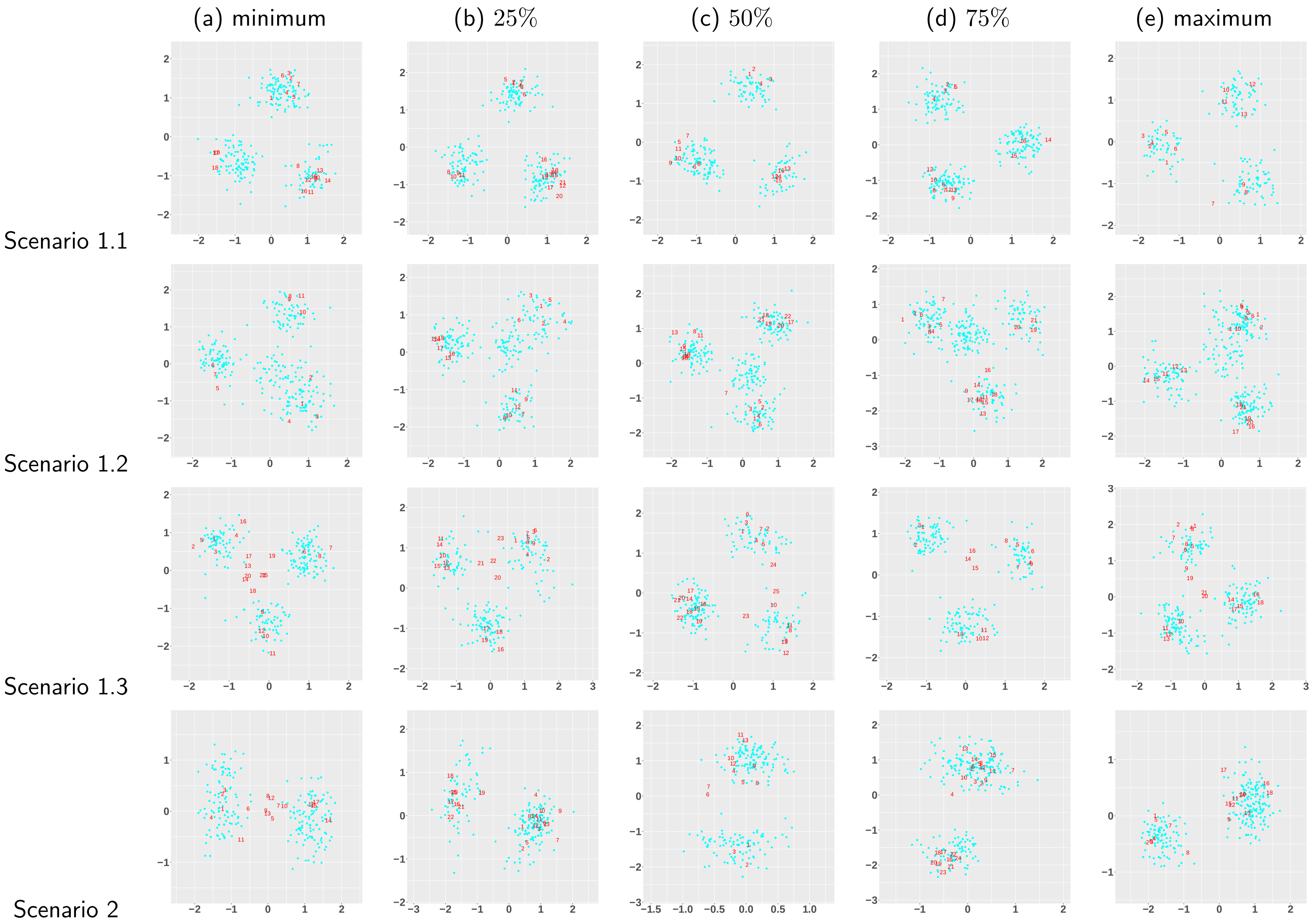


Figure 7: Interaction maps estimated from the latent space item response model using the social influence Model: Scenario 1.1, 1.2, 1.3, and 2

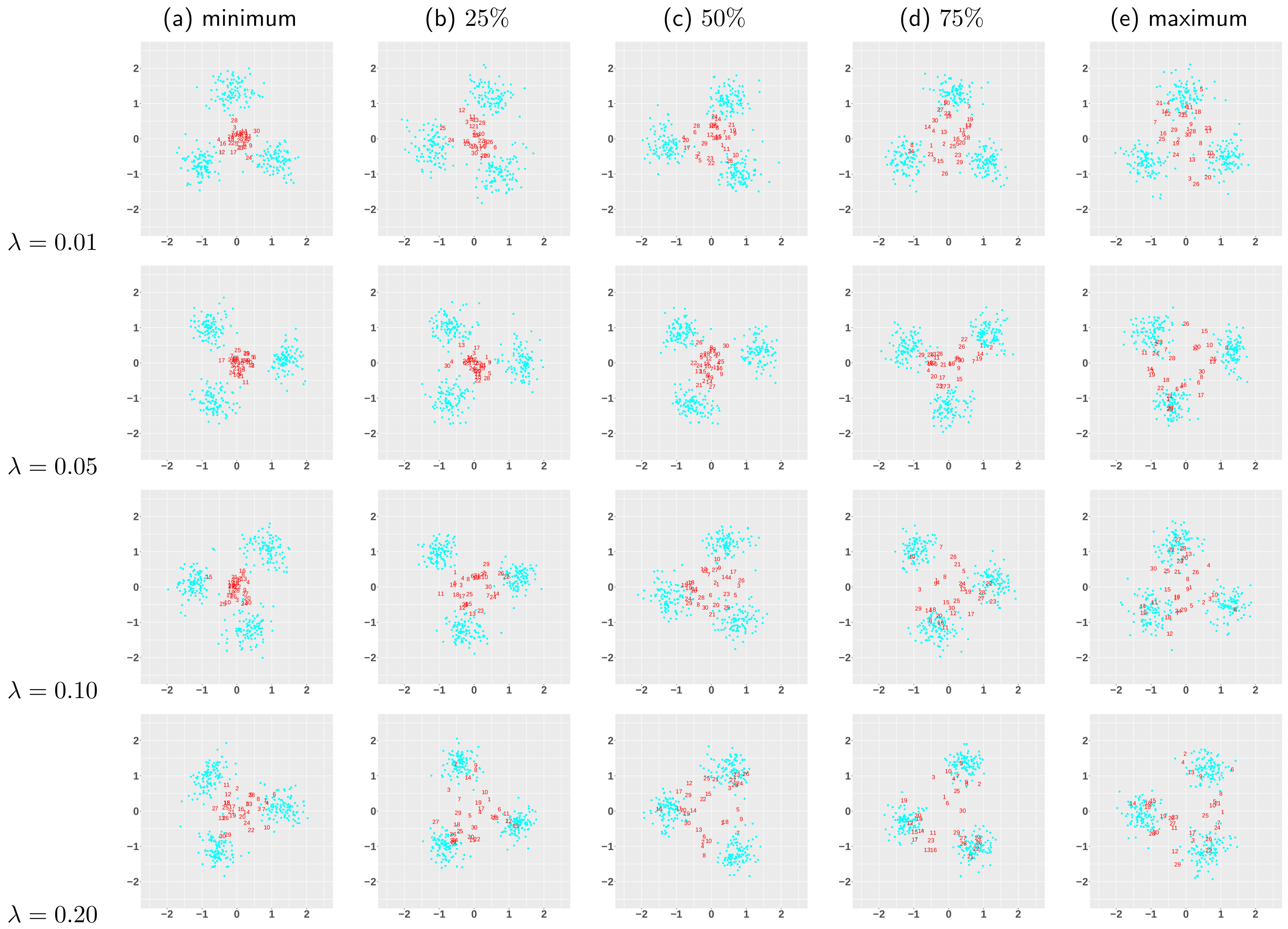


Figure 8: Interaction maps estimated from the latent space item response model using the social influence Model: Scenario 3 Part 1

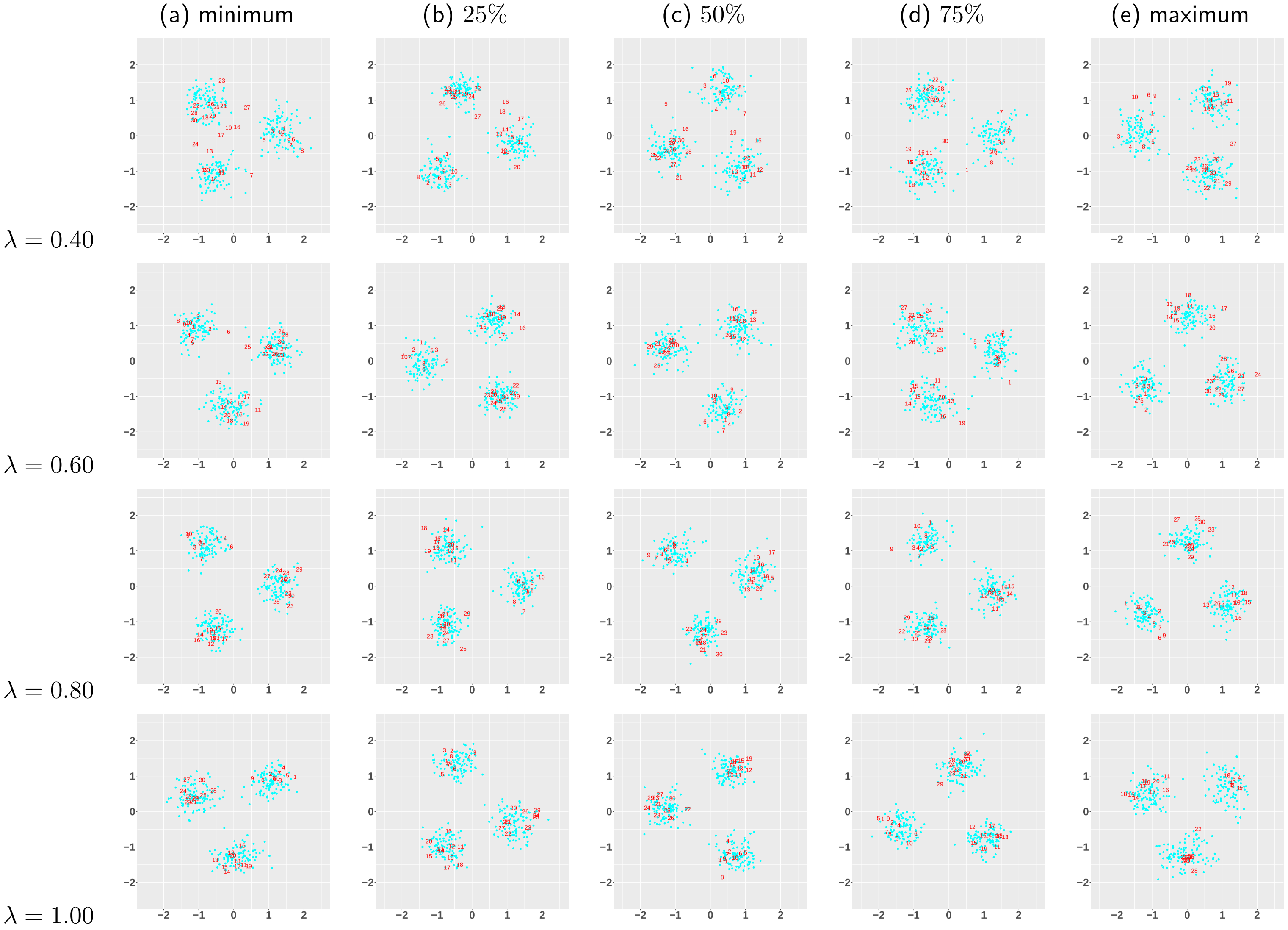


Figure 9: Interaction maps estimated from the latent space item response model using the social influence Model: Scenario 3 part 2