How social network influences human behavior: An integrated latent space approach

Ick Hoon Jin

Joint Work with Jina Park (Yonsei University) and Minjeong Jeon (UCLA)

May. 28. 2021

Introduction

- A frequently asked question in education and social science research is how contexts influence behavioral outcomes of individuals.
- The influence process can take various forms. For example, emotional or social support networks that students have in classrooms with their teachers, mentors, or friends are likely to play an important role in that process.
- The impact of individuals' social network on their behavior is often referred to as *social influence* in the literature.

Introduction

- In this study, we propose a new statistical modeling approach to identifying social influence effects.
- The key observation behind our idea is that in the social influence literature, behaviors are often measured at the scale level and used as covariates of the model for social network data.
- However, human mind and behaviors, such as their cognitive and non-cognitive abilities, psychology, and attitudes, are latent variables and measured based on multi-item scales. Here, we propose to model the behaviors of interest with the item-level data in the proposed approach.

Introduction

- Here, we propose to model the behaviors of interest with the item-level data in the proposed approach. By doing so, we essentially have two statistical models for individuals' item-level behavior data and social network data; then the task of examining social influences comes down to integrating the two statistical models.
- We will model the two types of data in a common modeling framework based on latent spaces: the latent space model (LSM) for social network data (Hoff et al., 2002) and the latent space item response model (LSIRM; Jeon et al. 2021) for item response behavior data.

Introduction

- The LSM measures the distances between the latent positions of individuals to model the probabilities of ties among them. The between-people distances indicate individuals' interactions in the network of interest, where shorter distances indicate stronger between-people connections.
- The LSIRM views binaray item response data as a bipartite network data that represent relationships between respondents and items. It measures the distances between respondents and items as a penalty term to model the respondents' probabilities of giving correct (or positive) responses, where the respondent-item distances represent their relations given the respective main effects.

Introduction

- A common feature of the LSM and the LSIRM is that both models locate individuals in a latent space based on their relationships with peers (for social network) and with behavior items (for item response network).
- After integrating the two latent spaces, we postulate that social influence can be studied by examining how the latent positions of the individuals determined by one type of data impact the other type of data.
- The configuration of the latent space for behavior response data reflects the individuals' social relations to their peers.

Introduction

- This way, social influence can be defined as the degree to which the positions individuals, determined based on their distances to peers in the social network of interest.
- Measuring the impact of the latent space configuration for item response data will tell us the effect of how individuals' social relationships influence their behavioral item responses

Illustration

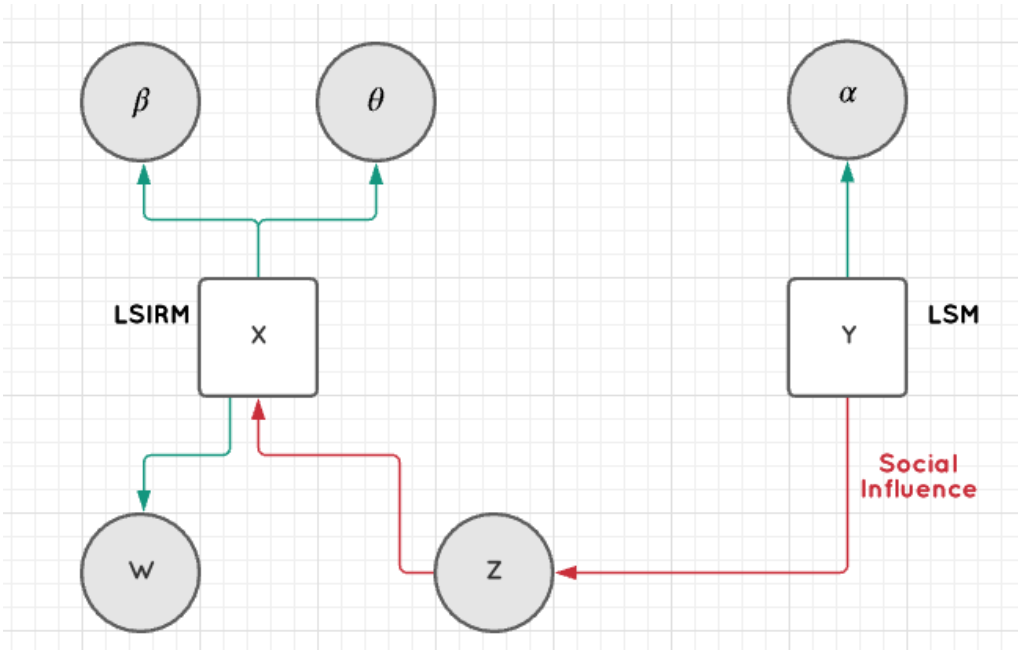


Figure: This figure illustrates our approach to social influence. The red line with an arrowhead shows the process of the social influence model. Under the social influence model, we estimate individuals' dependence structure from a peer network and then explain the item response dataset with the estimated dependent system.

Estimation of the LSM

The posterior distribution $\pi\left(\Theta^n \mid \mathbf{Y}\right)$ of the Bayesian LSM for social network data is written as follows:

$$\begin{aligned}\pi\left(\Theta^n \mid \mathbf{Y}\right) &\propto P\left(\mathbf{Y} \mid \Theta^n\right)\pi\left(\alpha\right)\pi\left(\gamma\right)\pi\left(\mathbf{Z}\right) \\ &= \prod_{k \neq l} \frac{\exp\left(\alpha - \gamma||\mathbf{z}_k - \mathbf{z}_l||\right)^{y_{kl}}}{1 + \exp\left(\alpha - \gamma||\mathbf{z}_k - \mathbf{z}_l||\right)} \pi\left(\alpha\right)\pi\left(\gamma\right)\pi\left(\mathbf{Z}\right), \end{aligned} \tag{1}$$

where $P\left(\mathbf{Y} \mid \Theta^n\right)$ is the likelihood of LSM.

Estimation of the LSIRM

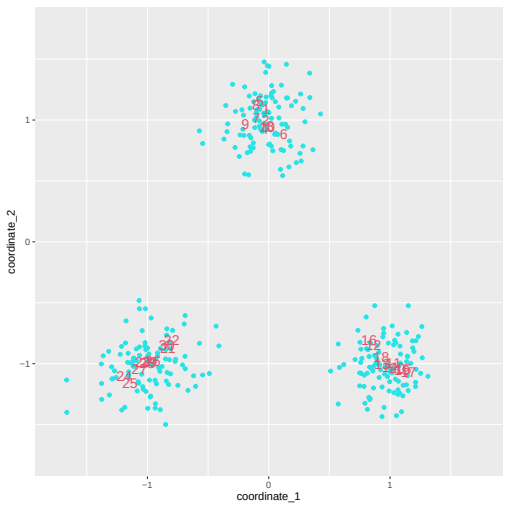
The posterior distribution $\pi\left(\Theta^r \mid \mathbf{X}, \widehat{\mathbf{Z}}\right)$ is then written as follows:

$$\begin{aligned}\pi\left(\Theta^r \mid \mathbf{X}, \widehat{\mathbf{Z}}\right) &\propto P\left(\mathbf{X} \mid \widehat{\mathbf{Z}}, \Theta^r\right)\pi\left(\boldsymbol{\beta}\right)\pi\left(\boldsymbol{\theta} \mid \sigma^2\right)\pi\left(\sigma^2\right)\pi\left(\mathbf{W}\right)\pi\left(\delta\right) \\ &= \prod_{k=1}^{n}\prod_{i=1}^{p}\frac{\exp\left(\beta_i + \theta_k - \delta||\hat{\mathbf{z}}_k - \mathbf{w}_i||\right)^{x_{ki}}}{1+\exp\left(\beta_i + \theta_k - \delta||\hat{\mathbf{z}}_k - \mathbf{w}_i||\right)} \\ &\times \pi\left(\boldsymbol{\beta}\right)\pi\left(\boldsymbol{\theta} \mid \sigma^2\right)\pi\left(\sigma^2\right)\pi\left(\mathbf{W}\right)\pi\left(\delta\right),\end{aligned} \tag{2}$$

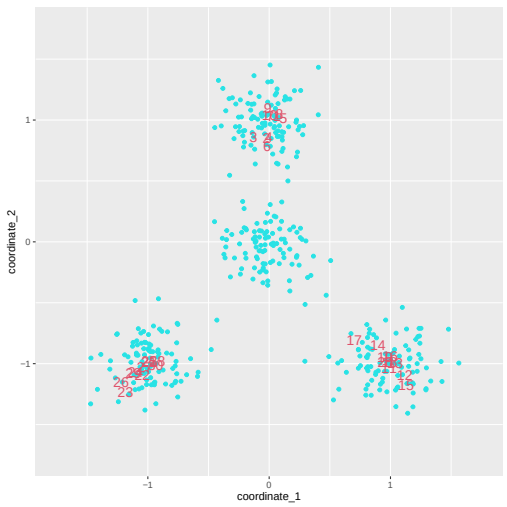
where $P\left(\mathbf{X} \mid \widehat{\mathbf{Z}}, \Theta^r\right)$ is the LSIRM for item response data $\mathbf{X}$, given the respondent latent positions obtained from (1).

Study Design

(a) Scenario 1



(b) Scenario 2



(c) Scenario 3

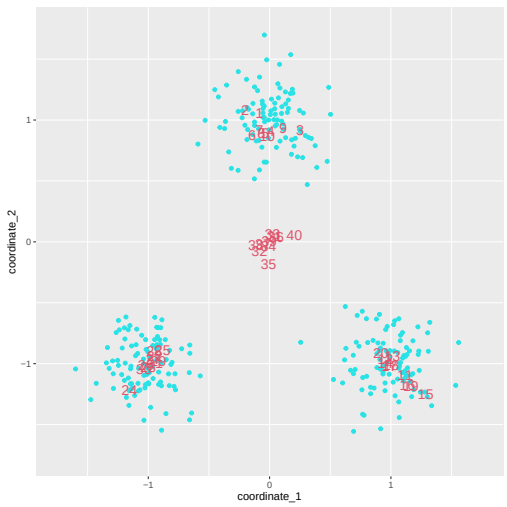
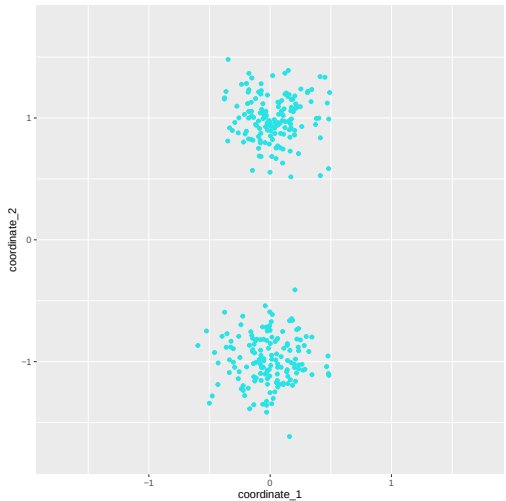


Figure: Latent spaces that illustrate Scenario 1 to 3. Red numbers and blue dots represent the latent positions of items and respondents, respectively.

Study Design

(a) Social network



(a) Item-respondent network

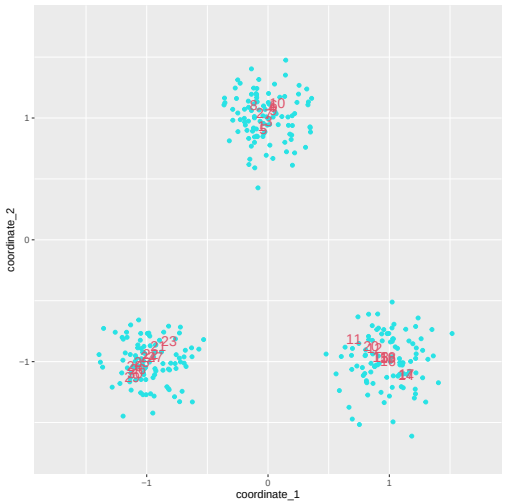


Figure: Social network latent space (a) and item-respondent latent space (b) considered in Scenario 4. Red numbers and blue dots represent latent positions for items and respondents, respectively.

Study Design

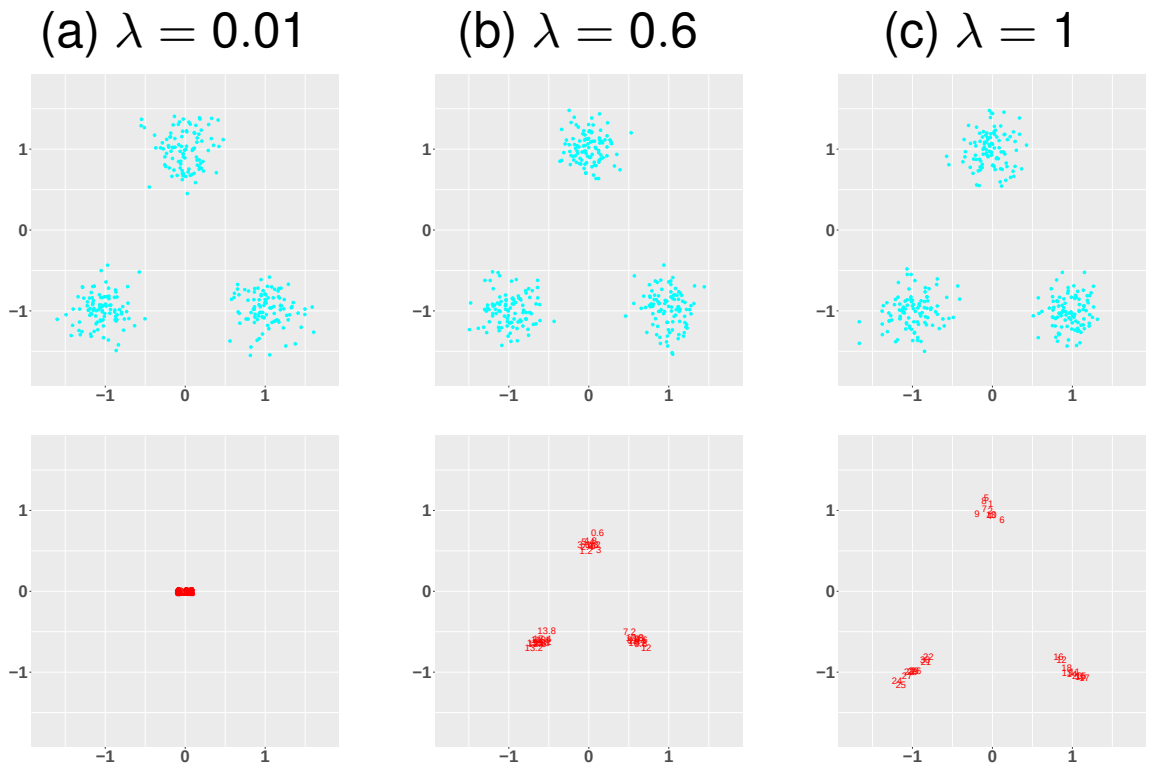


Figure: The top and bottom row are social network latent space and item response latent space for Scenario 5 of $\lambda = 0.01, 0.6$ and 1, respectively.

Estimation and Evaluation

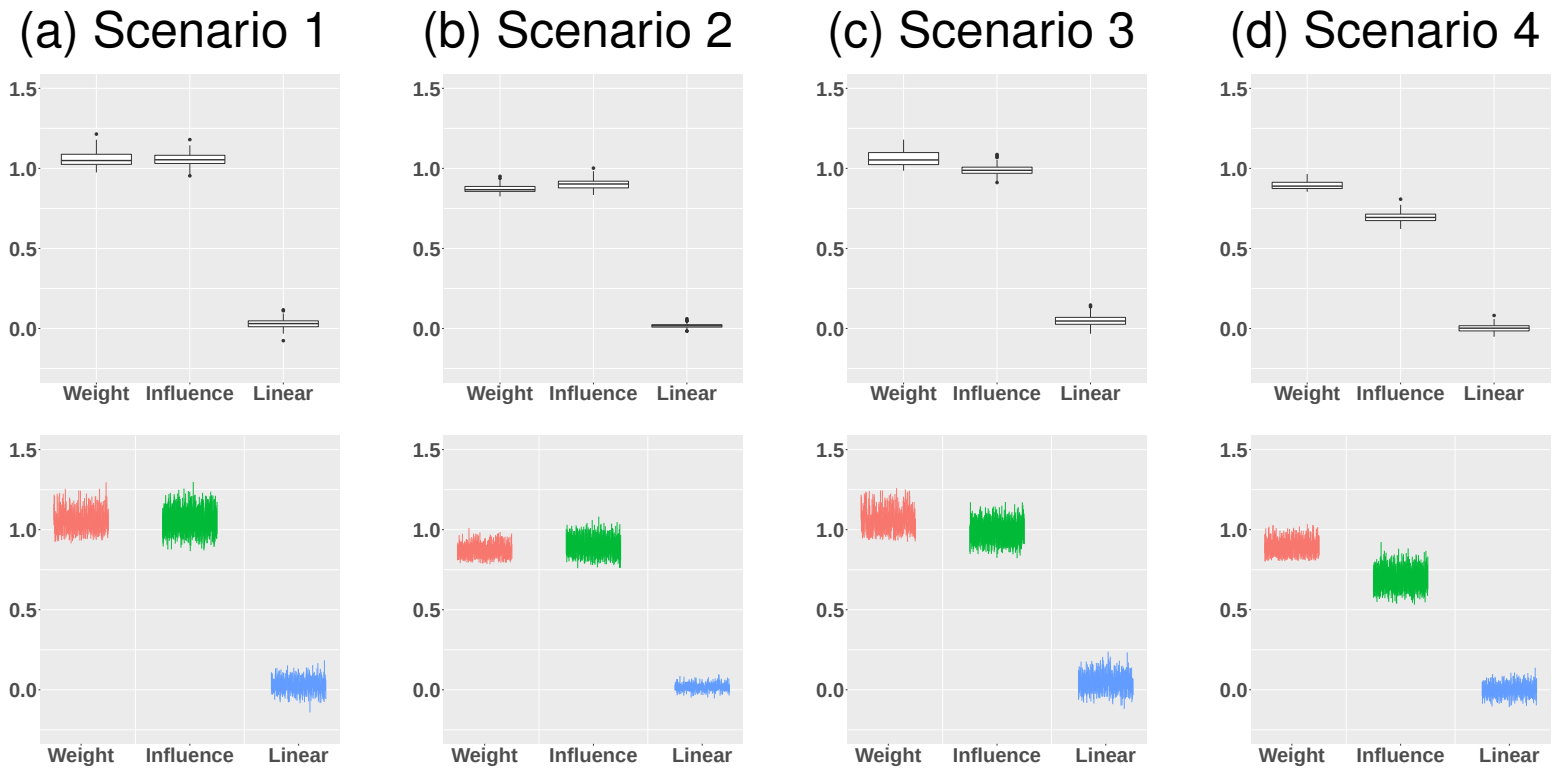


Figure: Graphical summaries of weight, influence parameters, and regression coefficients in Scenarios 1 – 4. The first row displays the boxplots of the posterior means of the three parameters. The second row presents the line plots that represent the 95% HPD intervals of weight and influence parameters and confidence intervals for regression coefficients for all simulated datasets in each scenario.

Estimation and Evaluation

	δ **from Social Influence Model**					
	min	25%	median	75%	max	mean
Scenario 1	0.954	1.031	1.054	1.082	1.180	1.056
Scenario 2	0.834	0.878	0.903	0.920	1.002	0.902
Scenario 3	0.912	0.969	0.988	1.008	1.087	0.989
Scenario 4	0.622	0.674	0.694	0.714	0.808	0.695

Table: Mean and five-number summaries of the estimated posterior means for the influence parameters from 200 simulated datasets for Scenario 1-4

Estimation and Evaluation

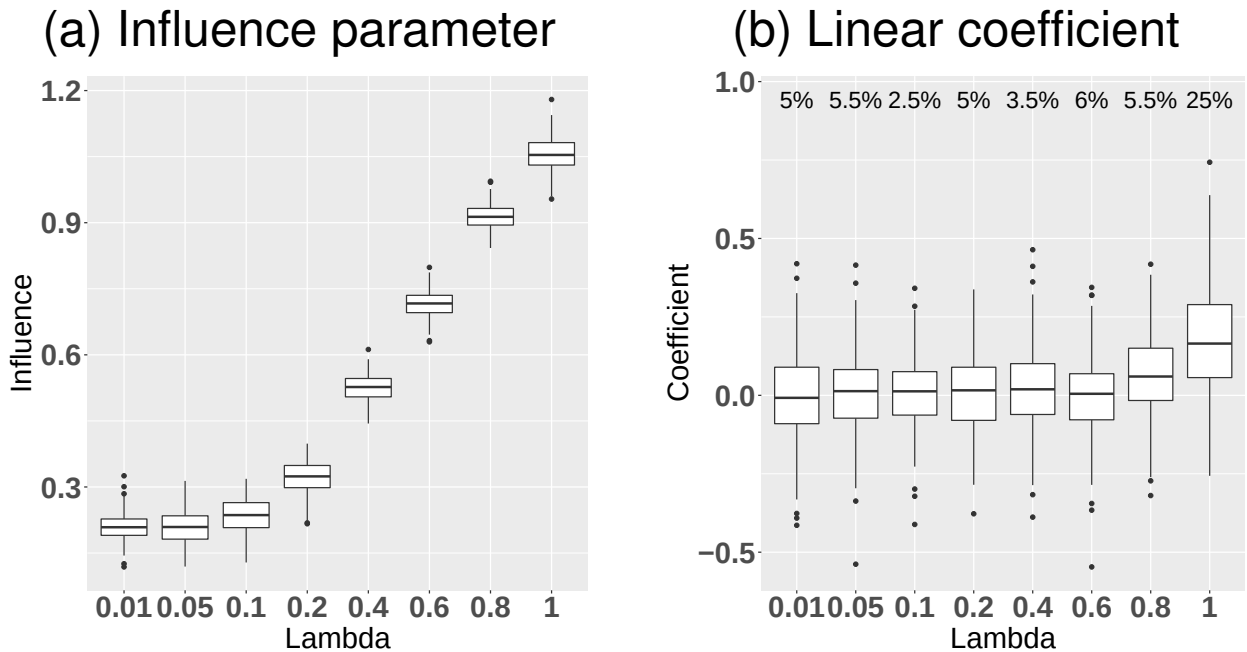


Figure: Graphical summaries of changes in the influence parameters and regression coefficients by λ for Scenario 5. Each boxplot represents the distribution of posterior means for influence parameters and regression coefficients. The boxplots on the left and right sides are for social influence and selection model, respectively. Boxplots in the top and bottom rows represent the change of influence parameters and regression coefficients as λ increases, respectively. The numbers above boxplots show the percentages not including zero in either 95% HPD or confidence intervals. For influence parameters, none of HPD intervals include zero even $\lambda = 0.01$.

Estimation and Evaluation

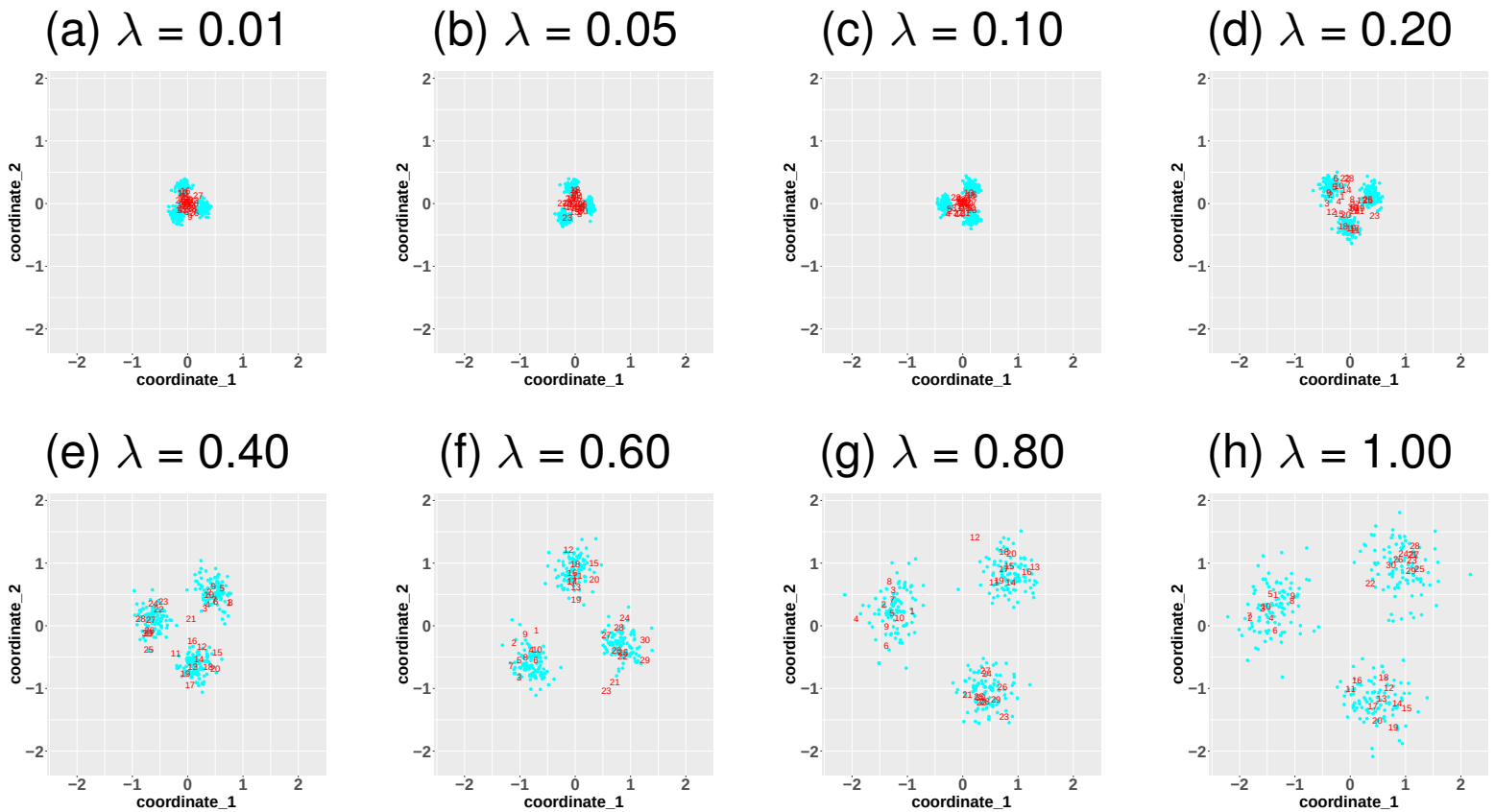


Figure: Estimated interaction maps for Scenario 5. For the social influence model, we selected the dataset whose influence parameter value is the closest to the median and visualized its interaction map estimated in LSIRM.

Estimation and Evaluation

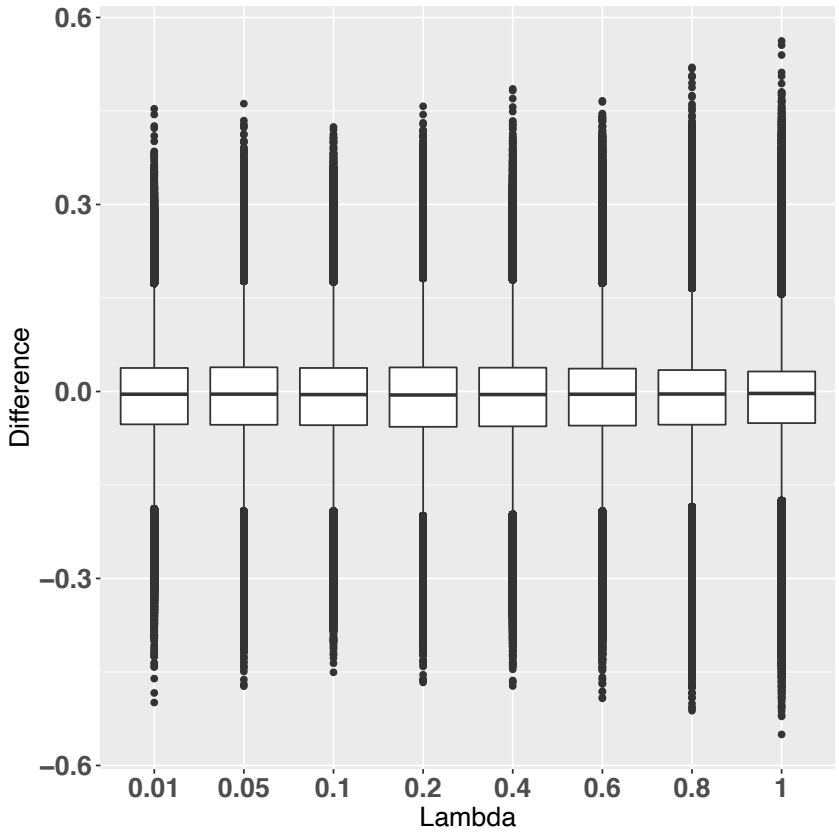


Figure: boxplot of $p_{r,ki} - \hat{p}_{r,ki}$ for the social influence model. As the differences get close to 0, the performance of the model is better.

Summary Plots

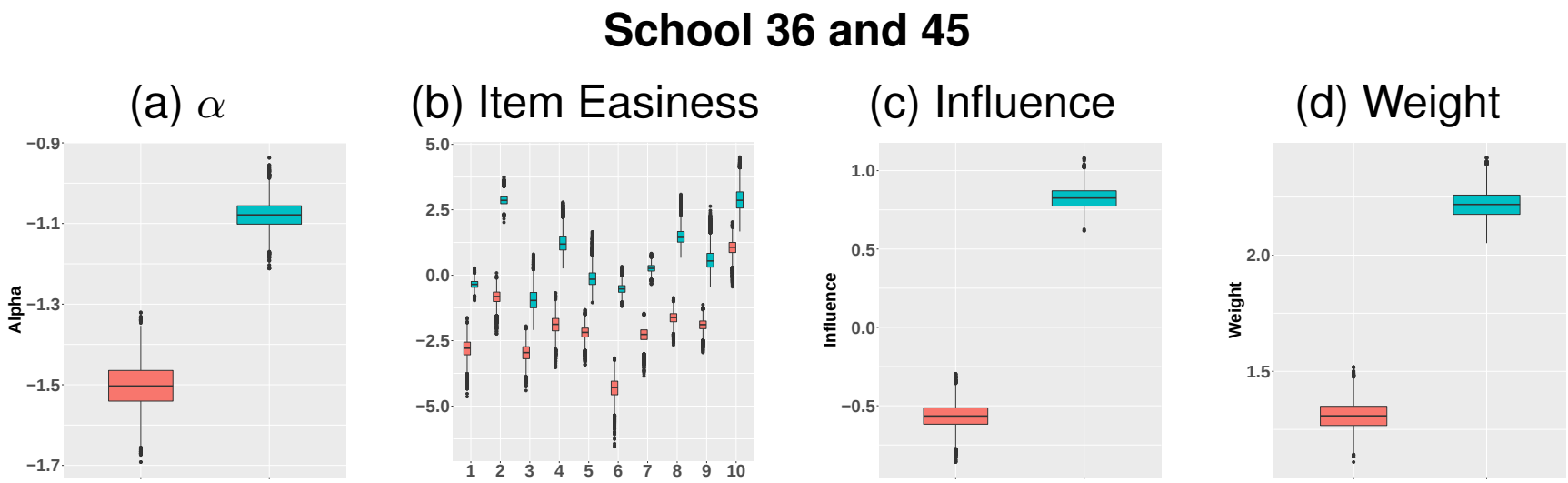


Figure: Summary plots of parameter estimates for School 36 and 45. School 45 has the highest percentage of black students and is shown in red color, and School 36 has the highest percentage of white students and is shown in blue color.

Summary Plots

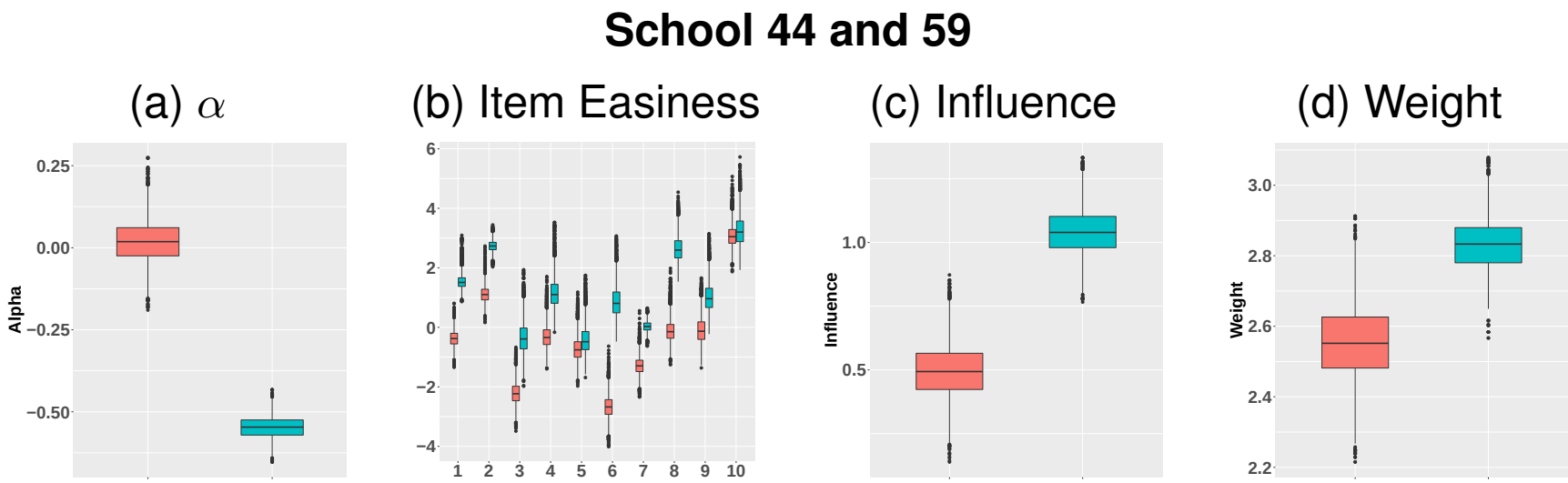


Figure: Summary plots of parameter estimates for School 44 and 59. School 44 and School 59 have the highest and smallest percentage of providing free or reduced lunch to students, respectively. This implies School 44 is located in the poorest area, and School 59 is located in the wealthiest area among sampled schools in New Jersey.

Influence Parameter

	Influence Parameter			Vanilla model coefficient	
	Posterior Mean	95% HPD Interval		Coefficient Estimate	P-Value
School 36	0.8259	(0.6801,	0.9624)	0.0391	0.399
School 45	-0.5664	(-0.7258,	-0.4117)	0.0594	0.257
School 44	0.4957	(0.2793,	0.7015)	0.0368	0.569
School 59	1.0398	(0.8646,	1.2116)	0.0682	0.107

Table: Posterior mean and 95% HPD intervals of the influence parameters for School 36, 45, 44, and 59.

Conclusion

- We have introduced a new novel approach for identifying social influence and social selection effects from individuals' social network and behavioral data.
- In our proposed model, evaluating social influence and social selection effects becomes the task of integrating two latent space models, one for network, and the other for item-level behavior data. Specifically, our social influence model estimates interactions in item response data given the respondent's latent positions from a network.
- The magnitude of social influence and selection are estimated from the weight parameters of distance terms in latent space models.

Hoff, P., A. Raftery, and M. S. Handcock (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association 97*, 1090–1098.

Jeon, M., I. H. Jin, M. Schweinberger, and S. Baugh (2021). Mapping unobserved item-respondent interactions: A latent space item response model with interaction map. *Psychometrika In Press*, 1–38.