

Article

Latent Class Analysis with Arbitrary-Distribution Responses

Huan Qing ¹  and Xiaofei Xu ^{2,*} 
¹ School of Economics and Finance, Chongqing University of Technology, Chongqing 400054, China; qinghuan@cqut.edu.cn

² School of Mathematics and Statistics, Wuhan University, Wuhan 430070, China

* Correspondence: xiaofeix@whu.edu.cn

Abstract

The latent class model has been proposed as a powerful tool in understanding human behavior for various fields such as social, psychological, behavioral, and biological sciences. However, one important limitation of the latent class model is that it is primarily applied to data with binary responses or categorical responses, making it fail to model real-world data with continuous or negative responses. In many applications, ignoring the weights throws out a lot of potentially valuable information contained in the weights. To address this limitation, we propose a novel generative model, the arbitrary-distribution latent class model (adLCM). Our model enables the generation of data's response matrix from an arbitrary distribution with a latent class structure. When compared to the latent class model, our adLCM is both more realistic and general. To our knowledge, our adLCM is the first model for latent class analysis with any real-valued responses, including continuous, negative, and signed values, thereby extending the classical latent class model beyond its traditional limitation to binary or categorical outcomes. We investigate the identifiability of the model and propose an efficient algorithm for estimating the latent classes and other model parameters. We show that the proposed algorithm enjoys consistent estimation. The performance of our algorithm is evaluated using both computer-generated data and real-world personality test data.

Keywords: categorical data; latent class model; spectral method; SVD; arbitrary-distribution responses



Academic Editors: Éloi Bossé and Alessandro Giuliani

Received: 9 July 2025

Revised: 2 August 2025

Accepted: 12 August 2025

Published: 14 August 2025

Citation: Qing, H.; Xu, X. Latent Class Analysis with Arbitrary-Distribution Responses. *Entropy* **2025**, *27*, 866. <https://doi.org/10.3390/e27080866>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The latent class model (LCM) [1–3] is a powerful tool for categorical data, with many applications across various areas such as social, psychological, behavioral, and biological sciences. These applications include movie rating [4,5], psychiatric evaluation [6–9], educational assessments [10], political surveys [11–14], transport economics personal interviews [15], and disease etiology detection [16–18]. In categorical data, subjects (individuals) typically respond to several items (questions). LCM is a theoretical model that categorizes subjects into disjoint groups, known as latent classes, according to their response pattern to a collection of categorical items. Latent classes assist researchers in better understanding human behaviors. For example, in movie rating, latent classes may represent different groups of users with an affinity for certain movie themes; in psychological tests, latent classes may represent different types of personalities. In educational assessments, latent classes may indicate different levels of abilities. In political surveys, latent classes may represent distinct types of political ideologies. In transport economics personal interviews, each latent class stands for a partition of the population. In disease etiology detection, latent

classes may represent different disease categories. To infer latent classes for categorical data generated from LCM, various approaches have been developed in recent years, including maximum likelihood estimation techniques [19–23], nonnegative matrix factorization (NMF) [24], tensor-based methods [25,26], and spectral clustering approaches [27–29].

To mathematically describe categorical data, let R be the N -by- J observed response matrix such that $R(i, j)$ represents subject i 's response to item j , where N denotes the number of subjects and J denotes the number of items. For LCM, many researchers focus on binary choice data where elements of the observed response matrix R only take 0 or 1 [10,16,26,30–37]. LCM models the binary (or categorical) response matrix by generating its elements from a Bernoulli (or Binomial) distribution. Binary responses can be agree/disagree responses in psychiatric evaluation, correct/wrong responses in educational assessments, and presence/absence of symptoms in disease etiology detection. For real-world categorical data from various online personality tests in the link https://openpsychometrics.org/_rawdata/ (accessed on 11 August 2025), the ranges of most categorical responses are $\{0, 1, 2, \dots, m\}$, where m is an integer like 2, 5, and 10. However, categorical data is more than binary or categorical responses. Categorical data with negative or continuous responses is also commonly encountered in the real world, and ignoring such weighted data may lose potentially meaningful information [38]. For example, in the buyer–seller rating e-commerce data [39], elements of the observed response matrix take values of $\{-1, 0, 1\}$ (for convenience, we call such R as a signed response matrix in this paper) since sellers are rated by users by applying three levels of rating, “Positive”, “Neutral”, and “Negative”. In the users–jokes rating categorical data Jester 100 [40], with the data source link: <https://eigentaste.berkeley.edu/dataset/archive/> (accessed on 11 August 2025), all responses (i.e., ratings) are continuous numbers in the range $[-10, 10]$. The aforementioned data cannot be generated from a Bernoulli or Binomial distribution. On the other hand, a substantial body of work has been developed to address polytomous responses or continuous responses, such as profile analysis (LPA) and factor mixture models (FMMs) [41–47], while their restrictive Gaussian or factorial assumptions limit their applicability to more general response types with irregular scales, such as real-valued ratings or signed categorical scores as aforementioned. Therefore, it is desirable to develop a more flexible model for data with arbitrary-distribution responses. With this motivation, our key contributions to the literature of latent class analysis are summarized as follows.

- **Model.** We propose a novel, identifiable, and generative statistical model, the arbitrary-distribution latent class model (adLCM), for data with arbitrary-distribution responses, where the responses can be continuous or negative values. Our adLCM allows the elements of an observed response matrix R to be generated from any distribution provided that the population version of R under adLCM enjoys a latent class structure. For example, our adLCM allows R to be generated from Bernoulli, Normal, Poisson, Binomial, Uniform, and Exponential distributions, etc. By considering a specifically designed discrete distribution, our adLCM can also model signed response matrices. For details, please refer to Examples 1–7.
- **Algorithm.** We develop an easy-to-implement algorithm, spectral clustering with K-means (SCK), to infer latent classes for arbitrary-distribution response matrices generated from arbitrary distribution under the proposed model. Our algorithm is designed based on a combination of two popular techniques: the singular value decomposition (SVD) and the K-means algorithm.
- **Theoretical property.** We build a theoretical framework to show that SCK enjoys consistent estimation under adLCM. We also provide Examples 1–7 to show that the theoretical performance of the proposed algorithm can be different when the

observed response matrices R are generated from different distributions under the proposed model.

- Empirical validation. We conduct extensive simulations to validate our theoretical insights. Additionally, we apply our SCK approach to two real-world personality test datasets with meaningful interpretations.

The remainder of this paper is organized as follows. Section 2 offers a comprehensive review of related works. Section 3 describes the model. Section 4 details the algorithm. Section 5 establishes the consistency results and provides examples for further analysis. Section 6 contains numerical studies that verify our theoretical findings and examine the performance of the proposed method. Section 7 demonstrates the proposed method using two real-world datasets. Section 8 concludes the paper with a brief discussion of contributions and future work.

The following notations will be used throughout the paper. For any positive integer m , let $[m]$ and $I_{m \times m}$ be $[m] := \{1, 2, \dots, m\}$ and the $m \times m$ identity matrix, respectively. For any vector x and any $q > 0$, $\|x\|_q$ denotes x 's l_q -norm. For any matrix M , M' denotes its transpose, $\|M\|$ denotes its spectral norm, $\|M\|_F$ denotes its Frobenius norm, $\text{rank}(M)$ denotes its rank, $\sigma_i(M)$ denotes its i -th largest singular value, $\lambda_i(M)$ denotes its i -th largest eigenvalue ordered by magnitude, $M(i, :)$ denotes its i -th row, and $M(:, j)$ denotes its j -th column. Let $\mathbb{R}$ and $\mathbb{N}$ be the set of real numbers and nonnegative integers, respectively. For any random variable X , $\mathbb{E}(X)$ and $\mathbb{P}(X = a)$ are the expectation and the probability that X equals to a , respectively. Let $\mathbb{M}_{m,K}$ be the collection of all $m \times K$ matrices where each row has only one 1 and all others 0.

2. Related Literature

Categorical data have been widely collected in various fields ranging from social and psychological research to political and transportation sciences [4–18]. The latent class model (LCM) has long been a cornerstone for modeling unobserved heterogeneity in categorical data that are not directly measurable [48,49]. Since [48] provided one of the earliest formal frameworks for latent structure models in sociology, LCM has been widely applied in psychology and medicine for identifying subgroups of individuals with similar symptom or trait profiles. For example, Meyer et al. [6] applied latent classes to represent different personality or clinical profiles in psychometric assessments. In political science, LCM is used to analyze voting patterns and public opinion. For instance, Poole [11] applied an unfolding approach to cluster legislators based on their yes/no voting patterns. Wu et al. [18] developed a nested latent class model that distinguished pathogen combinations causing childhood pneumonia in epidemiology. LCM is also employed in health outcomes research [16] or to cluster clinical profiles [17].

For the classical LCM, the responses of categorical data are often assumed to be binary or categorical with nonnegative integers, such as the binary responses of agree/disagree (or yes/no) or ordinal categories in certain surveys [26,30–37]. Compared to traditional clustering methods like K-means and K-modes, LCM provides a model-based clustering approach with the selection of the clusters based on rigorous statistical tests [28]. Various studies in the literature have extended and refined latent class techniques in diverse directions. For example, some studies focus on the fundamental concern of identifiability in LCM to evaluate the feasibility of recovering model parameters and latent classes [37,50]. Gyllenberg et al. [51] stated that the LCM with binary responses is not identifiable; this problem has received considerable attention in [34,37] for extended LCMs. Another extension is the grade-of-membership model (GoM), which allows individuals to have partial membership in multiple classes rather than a single discrete class [52]. Qing [28] detected mixed memberships in categorical data with polytomous responses based on the GoM model and derived theoretical guarantees

as well; see also [53–55]. Meanwhile, the estimation of LCM has also spurred extensive methodological innovation. For example, the maximum likelihood estimates are typically used via the expectation–maximization algorithm to iteratively refine class membership and parameter estimates [22,23,56]. Clinton et al. [12] considered Bayesian inference using MCMC techniques to analyze roll-call voting data, treating the class memberships as parameters to be sampled, see also [57–59] for more Bayesian reference. Recently, the tensor-based [25,26] and spectral-based algorithms [27,28] have also been popular to infer the latent memberships in the LCM framework.

As mentioned earlier, while classical latent class models are powerful for categorical data, they have a significant limitation as they typically assume binary or ordinal categories for the response. The literature has stated that the data could have arbitrary distribution with general responses such as negative or continuous values; ignoring this information may lead to a misunderstanding of the latent class structures [38]. Various methods have been proposed to address polytomous or continuous response data, such as latent profile analysis (LPA) and factor mixture models (FMM) [41–47]. However, for more general response types—such as the Jester joke data with ratings continuously varying within $[-10, 10]$ [40], e-commerce data with scores in the range $\{-1, 0, 1\}$ [39], or the Advogato trust data with relationship values of $\{0.6, 0.8, 1\}$ [60]—the classical LCM, LPA, and FMM are not applicable, due to LPA’s Gaussian assumption and FMM’s factorial assumption. To address this limitation, this paper proposes an arbitrary-distribution latent class model to enable the generation of data’s response matrix from an arbitrary distribution with a latent class structure. Hence, it allows any real-valued responses, including continuous, negative, and signed values, not merely limited to sampling weights, and thereby extends the classical LCM beyond its traditional response limitation.

In recent years, a noteworthy trend in latent structure learning is the use of spectral and other matrix/tensor-based methods to learn latent class structures as alternatives to traditional likelihood-based inference. Spectral algorithms offer an alternative by leveraging linear algebra (eigen-decomposition, singular value decomposition) or tensors of the data with theoretical soundness and ease of implementation [25]. In the context of network community detection—a problem analogous to finding latent classes of nodes—spectral clustering on graph Laplacians became a dominant approach for diverse types of networks [27,61–63]. Such algorithms have been extended to mixed-membership stochastic block models to allow overlapping communities [64,65]. In the latent class analysis realm, spectral algorithms have recently been developed to consistently estimate LCM parameters. For example, Chen and Gu [27] applied a spectral method for identifying GoM parameters from binary response data. Similarly, Anandkumar et al. [25] described tensor decomposition techniques to solve latent variable models via the factorization of third-order moments. These spectral methods, surveyed comprehensively by [66], achieve sound estimation and asymptotic consistency. Though recent advances extend spectral learning to polytomous categorical data, most existing applications still focus on binary settings. This paper focuses on the spectral method for arbitrary-distribution latent class modeling, and develops an easy-to-implement spectral clustering algorithm based on singular value decomposition (SVD) and the K-means algorithm.

3. Arbitrary-Distribution Latent Class Model

Unlike most researchers that focus on binary responses, in our arbitrary-distribution response setting in this paper, all elements of the observed response matrix R are allowed to be any real value, i.e., $R \in \mathbb{R}^{N \times J}$.

Consider categorical data with N subjects and J items, where the N subjects belong to K disjoint extreme latent profiles (also known as latent classes). Throughout this paper,

the number of classes K is assumed to be a known integer. To describe the membership of each subject, we let Z be a $N \times K$ matrix such that $Z(i, k)$ is 1 if subject i belongs to the k -th extreme latent profile and $Z(i, k)$ is 0 otherwise. Call Z the classification matrix in this paper. For each subject $i \in [N]$, it is assumed to belong to a single extreme latent profile. For convenience, define ℓ as an N -by-1 vector whose i -th entry $\ell(i)$ is k if the i -th subject belongs to the k -th extreme latent profile for $i \in [N]$. Thus, for subject $i \in [N]$, we have $Z(i, \ell(i)) = 1$ and the other $(K - 1)$ entries of the $K \times 1$ classification vector $Z(i, :)$ are 0.

Introduce the $J \times K$ item parameter matrix $\Theta \in \mathbb{R}^{J \times K}$. For $k \in [K]$, our arbitrary-distribution latent class model (adLCM) assumes that $\Theta(j, k)$ collects the conditional-response expectation for the response of the i -th subject to the j -th item under arbitrary distribution $\mathcal{F}$ provided that subject i belongs to the k -th extreme latent profile. Specifically, for $i \in [N]$, $j \in [J]$, given the classification vector $Z(i, :)$ of subject i and the item parameter matrix Θ , our adLCM assumes that for arbitrary distribution $\mathcal{F}$, the conditional response expectation of the i -th subject to the j -th item is

$$\mathbb{E}(R(i, j) | Z(i, :), \Theta) = \sum_{k=1}^K Z(i, k) \Theta(j, k) \equiv \Theta(j, \ell(i)). \quad (1)$$

Based on Equation (1), our adLCM can be simplified as follows.

Definition 1. Let $R \in \mathbb{R}^{N \times J}$ denote the observed response matrix. Let $Z \in \mathbb{M}_{N, K}$ be the classification matrix and $\Theta \in \mathbb{R}^{J \times K}$ be the item parameter matrix. For $i \in [N]$, $j \in [J]$, our arbitrary-distribution latent class model (adLCM) assumes that for an arbitrary distribution $\mathcal{F}$, $R(i, j)$ are independent random variables generated from the distribution $\mathcal{F}$ and the expectation of $R(i, j)$ under the distribution $\mathcal{F}$ should satisfy the following formula:

$$\mathbb{E}(R(i, j)) = R_0(i, j), \text{ where } R_0 := Z\Theta'. \quad (2)$$

Definition 1 says that adLCM is determined by the classification matrix Z , the item parameter matrix Θ , and the distribution $\mathcal{F}$. For brevity, we denote adLCM by $adLCM(Z, \Theta, \mathcal{F})$. Under adLCM, $\mathcal{F}$ is allowed to be any distribution as long as Equation (2) is satisfied under $\mathcal{F}$, i.e., adLCM only requires the expectation (i.e., population) response matrix R_0 of the observed response matrix R to be $Z\Theta'$ under any distribution $\mathcal{F}$.

Remark 1. For the case that $\mathcal{F}$ is a Bernoulli distribution, all elements of Θ have a range in $[0, 1]$, R only contains binary responses (i.e., $R(i, j) \in \{0, 1\}$ for $i \in [N]$, $j \in [J]$ when $\mathcal{F}$ is a Bernoulli distribution), and Equation (1) becomes $\mathbb{P}(R(i, j) = 1 | Z(i, :), \Theta) = \Theta(j, \ell(i))$. For this case, adLCM reduces to the LCM model for data with binary responses.

Remark 2. It should be noted that Equation (2) does not hold for all distributions. For instance, we cannot set $\mathcal{F}$ as a t -distribution because the expectation of a t -distribution is always 0, which cannot capture the latent structure required by adLCM; $\mathcal{F}$ cannot be a Cauchy distribution whose expectation does not even exist; $\mathcal{F}$ cannot be a Chi-square distribution because the expectation of a Chi-square distribution is its degrees of freedom, which is a fixed positive integer and cannot capture the latent structure required by adLCM. We will provide some examples to demonstrate that Equation (2) can be satisfied for different distributions of $\mathcal{F}$. For details, please refer to Examples 1–7.

Remark 3. It should also be noted that the ranges of the observed response matrix R and the item parameter matrix Θ depend on distribution $\mathcal{F}$. For example, when $\mathcal{F}$ is a Bernoulli distribution, $R \in \{0, 1\}^{N \times J}$ and $\Theta \in [0, 1]^{J \times K}$; when $\mathcal{F}$ is a Poisson distribution, $R \in \mathbb{N}^{N \times J}$ and $\Theta \in [0, +\infty)^{J \times K}$. If we let $\mathcal{F}$ be a Normal distribution, $R \in \mathbb{R}^{N \times J}$ and $\Theta \in (-\infty, +\infty)^{J \times K}$. For details, please refer to Examples 1–7.

The following proposition shows that adLCM is identifiable as long as there exists at least one subject for every extreme latent profile.

Proposition 1 (Identifiability). *Consider the adLCM as in Equation (2): when each extreme latent profile has at least one subject, the model is identifiable: for any other valid parameter set $(\tilde{Z}, \tilde{\Theta})$, if $\tilde{Z}\tilde{\Theta}' = Z\Theta'$, then (Z, Θ) and $(\tilde{Z}, \tilde{\Theta})$ are identical up to a permutation of the K extreme latent profiles.*

All proofs of theoretical results developed in this paper are given in the Appendix A. The condition that each extreme latent profile must contain at least one subject means that each extreme latent profile cannot be an empty set and we have $\text{rank}(Z) = K$.

Remark 4. *Note that Z and $\tilde{Z}$ are the same up to a permutation of the K latent classes in Proposition 1. A permutation is acceptable since the equivalence of Z and $\tilde{Z}$ should not rely on how we label each of the K extreme latent profiles. A similar argument holds for the identity of Θ and $\tilde{\Theta}$.*

The observed response matrix R along with the ground-truth classification matrix Z and the item parameter matrix Θ can be generated using our adLCM as follows: let $R(i, j)$ be a random variable generated by distribution $\mathcal{F}$ with expected value $R_0(i, j)$ for $i \in [N], j \in [J]$, where $R_0 = Z\Theta'$ satisfies the latent structure required by adLCM. In latent class analysis, given the observed response matrix R generated from $\text{adLCM}(Z, \Theta, \mathcal{F})$, our goal is to infer the classification matrix Z and the item parameter matrix Θ . Proposition 1 ensures that the model parameters Z and Θ can be reliably inferred from the observed response matrix R . In the following two sections, we will develop a spectral algorithm to fit adLCM and show that this algorithm yields consistent estimation.

4. A Spectral Method for Parameter Estimation

In addition to providing a more general model for latent class analysis, we are also interested in estimating the model parameters. In this section, we focus on the parameter estimation problem within the adLCM framework by developing an efficient and easy-to-implement spectral method.

To provide insight into developing an algorithm for the adLCM, we first consider an oracle case where we observe the expectation response matrix R_0 given in Equation (2). We would like to estimate Z and Θ from R_0 . Recall that the item parameter matrix Θ is a J -by- K matrix; here we let $\text{rank}(\Theta) = K_0$, where K_0 is a positive integer and it is no larger than K . As $R_0 = Z\Theta'$, $\text{rank}(Z) = K$, and $\text{rank}(\Theta) = K_0 \leq K$, we see that R_0 is a rank- K_0 matrix. As the number of extreme latent profiles K is usually far smaller than the number of subjects N and the number of items J , the N -by- J population response matrix R_0 enjoys a low-dimensional structure. Next, we will demonstrate that we can greatly benefit from the low-dimensional structure of R_0 when we aim to develop a method to infer model parameters under adLCM.

Let $R_0 = U\Sigma V'$ be the compact singular value decomposition (SVD) of R_0 such that Σ is a $K_0 \times K_0$ diagonal matrix collecting the K_0 nonzero singular values of R_0 . Write $\Sigma = \text{diag}(\sigma_1(R_0), \sigma_2(R_0), \dots, \sigma_{K_0}(R_0))$. The $N \times K_0$ matrix U collects the corresponding left singular vectors and it satisfies $U'U = I_{K_0 \times K_0}$. Similarly, the $J \times K_0$ matrix V collects the corresponding right singular vectors and it satisfies $V'V = I_{K_0 \times K_0}$. For $k \in [K]$, let N_k be the number of subjects that belong to the k -th extreme latent profile, i.e., $N_k = \sum_{i=1}^N Z(i, k)$. The ensuing lemma constitutes the foundation of our estimation method.

Lemma 1. *Under $\text{adLCM}(Z, \Theta, \mathcal{F})$, let $R_0 = U\Sigma V'$ be the compact SVD of R_0 . The following statements are true.*

- The left singular vectors matrix U can be written as

$$U = ZX, \quad (3)$$

where X is a $K \times K_0$ matrix.

- U has K distinct rows such that for any two distinct subjects i and $\bar{i}$ that belong to the same extreme latent profile (i.e., $\ell(i) = \ell(\bar{i})$), we have $U(i, :) = U(\bar{i}, :)$.
- Θ can be written as

$$\Theta = V\Sigma U'Z(Z'Z)^{-1}. \quad (4)$$

- Furthermore, when $K_0 = K$, for all $k \in [K], l \in [K]$, and $k \neq l$, we have

$$\|X(k, :) - X(l, :)\|_F = (N_k^{-1} + N_l^{-1})^{1/2}. \quad (5)$$

From now on, for the simplicity of our further analysis, we let $K_0 \equiv K$. Hence, the last statement of Lemma 1 always holds.

The second statement of Lemma 1 indicates that the rows of U corresponding to subjects assigned to the same extreme latent profile are identical. This circumstance implies that the application of a clustering algorithm to the rows of U can yield an exact reconstruction of the classification matrix Z after a permutation of the K extreme latent profiles.

In this paper, we adopt the K-means clustering algorithm, an unsupervised learning technique that groups similar data points into K clusters. This clustering technique is detailed as follows,

$$(\bar{Z}, \bar{X}) = \arg \min_{\bar{Z} \in \mathbb{M}_{N,K}, \bar{X} \in \mathbb{R}^{K \times K}} \|\bar{Z}\bar{X} - \bar{U}\|_F^2, \quad (6)$$

where $\bar{U}$ is any $N \times K$ matrix. For convenience, call Equation (6) a “Run K-means algorithm on all rows of $\bar{U}$ with K clusters to obtain $\bar{Z}$ ” because we are interested in the classification matrix $\bar{Z}$. Let $\bar{U}$ in Equation (6) be U ; the second statement of Lemma 1 guarantees that $\bar{Z} = Z\mathcal{P}$, $\bar{X} = \mathcal{P}'X$, where $\mathcal{P}$ is a $K \times K$ permutation matrix, i.e., the running K-means algorithm on all rows of U exactly recovers Z up to a permutation of the K extreme latent profiles.

After obtaining Z from U , Θ can be recovered subsequently by Equation (4). The above analysis suggests the following Algorithm 1, Ideal SCK, where SCK stands for Spectral Clustering with K-means. Ideal SCK returns a permutation of (Z, Θ) , which also supports the identifiability of the proposed model as stated in Proposition 1.

Algorithm 1 Ideal SCK

Require: The expectation response matrix R_0 and the number of extreme latent profiles K .

Ensure: A permutation of Z and Θ .

- 1: Obtain $U\Sigma V'$, the top K SVD of R_0 .
 - 2: Run K-means algorithm on all rows of U with K clusters to obtain $Z\mathcal{P}$, a permutation of Z .
 - 3: Equation (4) gives $V\Sigma U'Z\mathcal{P}((Z\mathcal{P})'Z\mathcal{P})^{-1} = \Theta\mathcal{P}$, a permutation of Θ .
-

For the real case, the response matrix R is observed rather than the expectation response matrix R_0 . We now move from the ideal scenario to the real scenario, intending to estimate Z and Θ when the observed response matrix R is a random matrix generated from an unknown distribution $\mathcal{F}$ satisfying Equation (2) with K extreme latent profiles under adLCM. The expectation of R is R_0 according to Equation (2) under adLCM, so intuitively, the singular values and singular vectors of R will be close to those of R_0 . Set $\hat{R} = \hat{U}\hat{\Sigma}\hat{V}'$ as

the top K SVD of R , where $\hat{\Sigma}$ is a $K \times K$ diagonal matrix collecting the top K singular values of R . Write $\hat{\Sigma} = \text{diag}(\sigma_1(R), \sigma_2(R), \dots, \sigma_K(R))$. As $\mathbb{E}(R) = R_0$ and the $N \times J$ matrix R_0 has K nonzero singular values, while the other $(\min(N, J) - K)$ singular values are zeros; we see that $\hat{R}$ should be a good approximation of R_0 . Matrices $\hat{U} \in \mathbb{R}^{N \times K}$, $\hat{V} \in \mathbb{R}^{J \times K}$ collect the corresponding left and right singular vectors and satisfy $\hat{U}'\hat{U} = \hat{V}'\hat{V} = I_{K \times K}$. The above analysis implies that $\hat{U}$ should have roughly K distinct rows because $\hat{U}$ is a slightly perturbed version of U . Therefore, to obtain a good estimation of the classification matrix Z , we should apply the K-means algorithm on all rows of $\hat{U}$ with K clusters. Let $\hat{Z}$ be the estimated classification matrix returned by applying the K-means method on all rows of $\hat{U}$ with K clusters. Then we are able to obtain a good estimation of Θ according to Equation (4) by setting $\hat{\Theta} = \hat{V}\hat{\Sigma}\hat{U}'\hat{Z}(\hat{Z}'\hat{Z})^{-1}$. Algorithm 2, referred to as SCK, is a natural extension of the Ideal SCK from the oracle case to the real case. Note that in our SCK algorithm there are only two inputs: the observed response matrix R and the number of latent classes K , i.e., SCK does not require any tuning parameters.

Algorithm 2 Spectral Clustering with K-means (SCK for short)

Require: The observed response matrix $R \in \mathbb{R}^{N \times J}$ and the number of extreme latent profiles K .

Ensure: $\hat{Z}$ and $\hat{\Theta}$.

- 1: Obtain $\hat{R} = \hat{U}\hat{\Sigma}\hat{V}'$, the top K SVD of R .
 - 2: Run K-means algorithm on all rows of $\hat{U}$ with K clusters to obtain $\hat{Z}$.
 - 3: Obtain an estimate of Θ by setting $\hat{\Theta} = \hat{R}'\hat{Z}(\hat{Z}'\hat{Z})^{-1}$.
-

Here, we evaluate the computational cost of our SCK algorithm. The computational cost of the SVD step involved in the SCK approach is $O(\max(N^2, J^2)K)$. For the K-means algorithm, its complexity is $O(NlK^2)$ with l being the number of K-means iterations. In all experimental studies considered in this paper, l is set as 100 for the K-means algorithm. The complexity of the last step in SCK is $O(JNK)$. Since $K \ll \min(N, J)$ in this paper, as a consequence, the total time complexity of our SCK algorithm is $O(\max(N^2, J^2)K)$.

5. Theoretical Properties

In this section, we present comprehensive theoretical properties of the SCK algorithm when the observed response matrix R is generated from the proposed model. Our objective is to demonstrate that the estimated classification matrix $\hat{Z}$ and the estimated item parameter matrix $\hat{\Theta}$ both concentrate around the true classification matrix Z and the true item parameter matrix Θ , respectively.

Let $\mathcal{T} = \{\mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_K\}$ be the collection of true partitions for all subjects, where $\mathcal{T}_k = \{i : Z(i, k) = 1 \text{ for } i \in [N]\}$ for $k \in [K]$, i.e., $\mathcal{T}_k$ is the set of the true partition of subjects into the k -th extreme latent profile. Similarly, let $\hat{\mathcal{T}} = \{\hat{\mathcal{T}}_1, \hat{\mathcal{T}}_2, \dots, \hat{\mathcal{T}}_K\}$ represent the collection of estimated partitions for all subjects, where $\hat{\mathcal{T}}_k = \{i : \hat{Z}(i, k) = 1 \text{ for } i \in [N]\}$ for $k \in [K]$. We use the measure defined in [61] to quantify the closeness of the estimated partition $\hat{\mathcal{T}}$ and the ground-truth partition $\mathcal{T}$. Denote the *Clustering error* associated with $\mathcal{T}$ and $\hat{\mathcal{T}}$ as

$$\hat{f} = \min_{\pi \in S_K} \max_{k \in [K]} \frac{|\mathcal{T}_k \cap \hat{\mathcal{T}}_{\pi(k)}^c| + |\mathcal{T}_k^c \cap \hat{\mathcal{T}}_{\pi(k)}|}{N_K}, \quad (7)$$

where S_K represents the set of all permutations of $\{1, 2, \dots, K\}$, and $\hat{\mathcal{T}}_{\pi(k)}^c$ and $\mathcal{T}_k^c$ denote the complementary sets. As stated in Reference [61], $\hat{f}$ evaluates the maximum proportion of subjects in the symmetric difference of $\mathcal{T}_k$ and $\hat{\mathcal{T}}_{\pi(k)}$. Since the observed response matrix

R is generated from adLCM with expectation R_0 , and $\hat{f}$ measures the performance of the SCK algorithm, it is expected that SCK estimates Z with a small Clustering error $\hat{f}$.

For convenience, let $\rho = \max_{j \in [J], k \in [K]} |\Theta(j, k)|$ and call it the scaling parameter. Let $B = \frac{\Theta}{\rho}$, and then we have $\max_{j \in [J], k \in [K]} |B(j, k)| = 1$ and $R_0 = \rho ZB'$. Let $\tau = \max_{i \in [N], j \in [J]} |R(i, j) - R_0(i, j)|$ and $\gamma = \max_{i \in [N], j \in [J]} \text{Var}(R(i, j))$ where $\text{Var}(R(i, j))$ means the variance of $R(i, j)$. We require the following assumption to establish theoretical guarantees of consistency for our SCK method.

Assumption 1. Assume $\gamma \geq \frac{\tau^2 \log(N+J)}{\max(N, J)}$.

The following theorem presents our main result, which provides upper bounds for the error rates of our SCK algorithm under our adLCM.

Theorem 1. Under $\text{adLCM}(Z, \Theta, \mathcal{F})$, if Assumption 1 is satisfied, with a probability of at least $1 - o((N + J)^{-3})$,

$$\hat{f} = O\left(\frac{\gamma K^2 N_{\max} \max(N, J) \log(N + J)}{\rho^2 N_{\min}^2 J}\right) \text{ and } \frac{\|\hat{\Theta} - \Theta \mathcal{P}\|_F}{\|\Theta\|_F} = O\left(\frac{\sqrt{\gamma K \max(N, J) \log(N + J)}}{\rho \sqrt{N_{\min} J}}\right),$$

where $N_{\max} = \max_{k \in [K]} \{N_k\}$, $N_{\min} = \min_{k \in [K]} \{N_k\}$, and $\mathcal{P}$ is a permutation matrix.

Because our adLCM is distribution-free, Theorem 1 provides a general theoretical guarantee of the SCK algorithm when R is generated from adLCM for any distribution $\mathcal{F}$ as long as Equation (2) is satisfied. We can simplify Theorem 1 by considering additional conditions:

Corollary 1. Under $\text{adLCM}(Z, \Theta, \mathcal{F})$, when Assumption 1 holds, if we make the additional assumption that $\frac{N_{\max}}{N_{\min}} = O(1)$ and $K = O(1)$, with a probability of at least $1 - o((N + J)^{-3})$,

$$\hat{f} = O\left(\frac{\gamma \max(N, J) \log(N + J)}{\rho^2 N J}\right) \text{ and } \frac{\|\hat{\Theta} - \Theta \mathcal{P}\|_F}{\|\Theta\|_F} = O\left(\frac{\sqrt{\gamma \max(N, J) \log(N + J)}}{\rho \sqrt{N J}}\right).$$

For the case $J = \beta N$ for any positive constant β , Corollary 1 implies that the SCK algorithm yields consistent estimation under adLCM since the error bounds in Corollary 1 decrease to zero as $N \rightarrow +\infty$ when ρ and distribution $\mathcal{F}$ are fixed.

Recall that R is an observed response matrix generated from a distribution $\mathcal{F}$ with expectation $R_0 = Z\Theta' = \rho ZB'$ under adLCM and γ is the maximum variance of $R(i, j)$ and it is closely related to the distribution $\mathcal{F}$, the ranges of R, ρ, B , and γ can vary depending on the specific distribution $\mathcal{F}$. The following examples provide the ranges of R, ρ, B , the upper bound of γ , and the explicit forms of error bounds in Theorem 1 for different distribution $\mathcal{F}$ under our adLCM. Meanwhile, based on the explicitly derived error bounds for different distribution $\mathcal{F}$, we also investigate how the scaling parameter ρ influences the performance of the SCK algorithm in these examples. For all pairs (i, j) with $i \in [N], j \in [J]$, we consider the following distributions when $\mathbb{E}(R) = R_0$ in Equation (2) holds.

Example 1. Let $\mathcal{F}$ be a Bernoulli distributionsuch that $R(i, j) \sim \text{Bernoulli}(R_0(i, j))$, where $R_0(i, j)$ is the Bernoulli probability, i.e., $\mathbb{E}(R(i, j)) = R_0(i, j)$. For this case, our adLCM degenerates to the LCM for data with binary responses. According to the properties of the Bernoulli distribution, we have the following conclusions.

- $R(i, j) \in \{0, 1\}$, i.e., $R(i, j)$ only takes two values 0 and 1.
- $B(i, j) \in [0, 1]$ and $\rho \in (0, 1]$ because $R_0(i, j)$ is a probability located in $[0, 1]$ and $\max_{i \in [N], j \in [J]} |B(i, j)|$ is assumed to be 1.
- $\tau \leq 1$ because $\tau = \max_{i \in [N], j \in [J]} |R(i, j) - R_0(i, j)| \leq 1$.

- $\gamma \leq \rho$ because $\gamma = \max_{i \in [N], j \in [J]} \text{Var}(R(i, j)) = \max_{i \in [N], j \in [J]} R_0(i, j)(1 - R_0(i, j)) \leq \max_{i \in [N], j \in [J]} R_0(i, j) = \max_{i \in [N], j \in [J]} \rho(ZB)(i, j) \leq \rho$.
- Let τ be its upper bound 1 and γ be its upper bound ρ , Assumption 1 becomes $\rho \geq \frac{\log(N+J)}{\max(N, J)}$, which means a sparsity requirement on R because ρ controls the probability of the numbers of ones in R for this case.
- Let γ be its upper bound ρ in Theorem 1, then we have

$$\hat{f} = O\left(\frac{K^2 N_{\max} \max(N, J) \log(N+J)}{\rho N_{\min}^2 J}\right) \text{ and } \frac{\|\hat{\Theta} - \Theta \mathcal{P}\|_F}{\|\Theta\|_F} = O\left(\frac{\sqrt{K \max(N, J) \log(N+J)}}{\sqrt{\rho N_{\min} J}}\right).$$

We observe that increasing ρ leads to a decrease in SCK's error rates when $\mathcal{F}$ is a Bernoulli distribution.

Example 2. Let $\mathcal{F}$ be a Binomial distribution such that $R(i, j) \sim \text{Binomial}(m, \frac{R_0(i, j)}{m})$ for any positive integer m , where $R(i, j)$ is a random variable that reflects the number of successes in a fixed number of independent trials m with the same probability of success $\frac{R_0(i, j)}{m}$, i.e., $\mathbb{E}(R(i, j)) = R_0(i, j)$. For this case, our adLCM reduces to the LCM for data with categorical responses. For a Binomial distribution, we have $\mathbb{P}(R(i, j) = r) = \binom{m}{r} (\frac{R_0(i, j)}{m})^r (1 - \frac{R_0(i, j)}{m})^{m-r}$ for $r = 0, 1, 2, \dots, m$, where $\binom{\bullet}{\bullet}$ is a binomial coefficient. By the property of the Binomial distribution, we have the following conclusions.

- $R(i, j) \in \{0, 1, 2, \dots, m\}$.
- $B(i, j) \in [0, 1]$ and $\rho \in (0, m]$ because $\frac{R_0(i, j)}{m}$ is a probability that has a range in $[0, 1]$.
- $\tau \leq m$ because $\tau = \max_{i \in [N], j \in [J]} |R(i, j) - R_0(i, j)| \leq m$.
- $\gamma \leq \rho$ because $\gamma = \max_{i \in [N], j \in [J]} \text{Var}(R(i, j)) = m \frac{R_0(i, j)}{m} (1 - \frac{R_0(i, j)}{m}) = R_0(i, j)(1 - \frac{R_0(i, j)}{m}) \leq \rho$.
- Let τ be its upper bound m and γ be its upper bound ρ , then Assumption 1 becomes $\rho \geq \frac{m^2 \log(N+J)}{\max(N, J)}$ which provides a lower bound requirement of the scaling parameter ρ .
- Let γ be its upper bound ρ in Theorem 1, then we obtain the exact forms of error bounds for SCK when $\mathcal{F}$ is a Binomial distribution, and we observe that increasing ρ reduces SCK's error rates.

Remark 5. When $\mathcal{F}$ is the Binomial distribution and Condition 1 of [28] holds ($K = O(1)$, $J = O(N)$, $N_{\max}/N_{\min} = O(1)$), the theoretical upper bounds for $\hat{f}$ and $\frac{\|\hat{\Theta} - \Theta \mathcal{P}\|_F}{\|\Theta\|_F}$ correspond to the squared error bounds established in Theorem 1 of [28]. This optimality guarantee is consistent with Assumption 1, which aligns with Assumption 1 in [28]. The relationship arises because the theoretical results in [28] apply to the overlapping case (each subject can belong to multiple latent classes). Under mild conditions, spectral clustering methods in the non-overlapping case (each subject belongs to only one latent class) typically achieve an error bound that is the square of the bound for spectral methods in the overlapping case, while requiring similar sparsity levels. This phenomenon is generally observed in the area of community detection: for instance, spectral methods for non-overlapping networks [67–70] exhibit squared error bounds relative to those for overlapping networks [64,65,71,72], despite comparable sparsity requirements. We refer readers to [73] for further discussion of this phenomenon.

Example 3. Let $\mathcal{F}$ be a Poisson distribution such that $R(i, j) \sim \text{Poisson}(R_0(i, j))$, where $R_0(i, j)$ is the Poisson parameter, i.e., $\mathbb{E}(R(i, j)) = R_0(i, j)$. By the properties of the Poisson distribution, the following conclusions can be obtained.

- $R(i, j) \in \mathbb{N}$, i.e., $R(i, j)$ is a nonnegative integer.
- $B(i, j) \in [0, 1]$ and $\rho \in (0, +\infty)$ because Poisson distribution can take any positive value for its mean.

- τ is an unknown positive value because we cannot know the exact upper bound of $R(i, j)$ when R is obtained from the Poisson distribution under the adLCM.
- $\gamma \leq \rho$ because $\gamma = \max_{i \in [N], j \in [J]} \text{Var}(R(i, j)) = \max_{i \in [N], j \in [J]} R_0(i, j) \leq \rho$.
- Let γ be its upper bound ρ , then Assumption 1 becomes $\rho \geq \frac{\tau^2 \log(N+J)}{\max(N, J)}$ which is a lower bound requirement of ρ .
- Let γ be its upper bound ρ in Theorem 1 which obtains the exact forms of error bounds for the SCK algorithm when $\mathcal{F}$ is a Poisson distribution. It is easy to observe that increasing ρ leads to a decrease in SCK's error rates.

Example 4. Let $\mathcal{F}$ be a Normal distribution such that $R(i, j) \sim \text{Normal}(R_0(i, j), \sigma^2)$, where $R_0(i, j)$ is the mean (i.e., $\mathbb{E}(R(i, j)) = R_0(i, j)$) and σ^2 is the variance parameter for Normal distribution. For this case, we have

- $R(i, j) \in \mathbb{R}$, i.e., $R(i, j)$ is a real value.
- $B(i, j) \in [-1, 1]$ and $\rho \in (0, +\infty)$ because the mean of Normal distribution can take any value. Note that, unlike the cases when $\mathcal{F}$ is Bernoulli or Poisson, B can have negative elements for the Normal distribution case.
- Similar to Example 3, τ is an unknown positive value.
- $\gamma = \sigma^2$ because $\gamma = \max_{i \in [N], j \in [J]} \text{Var}(R(i, j)) = \max_{i \in [N], j \in [J]} \sigma^2 = \sigma^2$ for Normal distribution.
- Let γ be its exact value σ^2 , then Assumption 1 becomes $\sigma^2 \max(N, J) \geq \tau^2 \log(N+J)$ which means that $\max(N, J)$ should be set larger than $\frac{\tau^2 \log(N+J)}{\sigma^2}$ for our theoretical analysis.
- Let γ be its exact value σ^2 in Theorem 1 which provides the exact forms of error bounds for SCK. We observe that increasing the scaling parameter ρ (or decreasing the variance σ^2) reduces SCK's error rates.

Example 5. Let $\mathcal{F}$ be an Exponential distribution such that $R(i, j) \sim \text{Exponential}(\frac{1}{R_0(i, j)})$, where $\frac{1}{R_0(i, j)}$ is the Exponential parameter, i.e., $\mathbb{E}(R(i, j)) = R_0(i, j)$. For this case, we have

- $R(i, j) \in \mathbb{R}_+$, i.e., $R(i, j)$ is a positive value.
- $B(i, j) \in (0, 1]$ and $\rho \in (0, +\infty)$ because the mean of Exponential distribution can be any positive value.
- Similar to Example 3, τ is an unknown positive value.
- $\gamma \leq \rho^2$ because $\gamma = \max_{i \in [N], j \in [J]} \text{Var}(R(i, j)) = \max_{i \in [N], j \in [J]} R_0^2(i, j) \leq \rho^2$ for Exponential distribution.
- Let γ be its upper bound ρ^2 , then Assumption 1 becomes $\rho^2 \geq \tau^2 \log(N+J) / \max(N, J)$, a lower bound requirement of ρ .
- Let γ be its upper bound ρ^2 in Theorem 1; the theoretical bounds demonstrate that ρ vanishes, which indicates that increasing ρ has no significant impact on the error rates of SCK.

Example 6. Let $\mathcal{F}$ be a Uniform distribution such that $R(i, j) \sim \text{Uniform}(0, 2R_0(i, j))$, where $\mathbb{E}(R(i, j)) = \frac{0+2R_0(i, j)}{2} = R_0(i, j)$ holds immediately. For this case, we have

- $R(i, j) \in (0, 2\rho)$ because $2R_0(i, j) \leq 2\rho$.
- $B(i, j) \in (0, 1]$ and $\rho \in (0, +\infty)$ because $\text{Uniform}(0, 2R_0(i, j))$ allows $2R_0(i, j)$ to be any positive value.
- τ is an unknown positive value with an upper bound 2ρ .
- $\gamma \leq \frac{\rho^2}{3}$ because $\gamma = \max_{i \in [N], j \in [J]} \text{Var}(R(i, j)) = \max_{i \in [N], j \in [J]} \frac{(2R_0(i, j)-0)^2}{12} = \max_{i \in [N], j \in [J]} \frac{R_0^2(i, j)}{3} \leq \frac{\rho^2}{3}$ for Uniform distribution.
- Let γ be its upper bound $\frac{\rho^2}{3}$, then Assumption 1 becomes $\rho^2 \geq 3\tau^2 \log(N+J) / \max(N, J)$, a lower bound requirement of ρ .

- Since ρ disappears in the error bounds when we let $\gamma = \frac{\rho^2}{3}$ in Theorem 1, increasing ρ does not significantly influence SCK's error rates, a conclusion similar to Example 5.

Example 7. Our adLCM can also model a signed response matrix by setting $\mathbb{P}(R(i, j) = 1) = \frac{1+R_0(i, j)}{2}$ and $\mathbb{P}(R(i, j) = -1) = \frac{1-R_0(i, j)}{2}$, where $\mathbb{E}(R(i, j)) = \frac{1+R_0(i, j)}{2} - \frac{1-R_0(i, j)}{2} = R_0(i, j)$ and Equation (2) holds surely. For the signed response matrix, we have

- $R(i, j) \in \{-1, 1\}$, i.e., $R(i, j)$ only takes two values -1 and 1 .
- $B(i, j) \in [-1, 1]$ and $\rho \in (0, 1]$ because $\frac{1+R_0(i, j)}{2}$ and $\frac{1-R_0(i, j)}{2}$ are two probabilities which should be in the range $[0, 1]$. Note that, similar to Example 4, $B(i, j)$ can be negative for the signed response matrix.
- $\tau \leq 2$ because $R(i, j) \in \{-1, 1\}$ and $R_0(i, j) \in [-1, 1]$.
- $\gamma \leq 1$ because $\gamma = \max_{i \in [N], j \in [J]} \text{Var}(R(i, j)) = \max_{i \in [N], j \in [J]} (1 - R_0^2(i, j)) \leq 1$.
- When setting $\tau = 2$ and $\gamma = 1$, Assumption 1 turns to be $\max(N, J) \geq 4\log(N + J)$.
- Setting γ as its upper bound 1 in Theorem 1 gives that increasing ρ reduces SCK's error rates.

6. Simulation Studies

In this section, we conduct extensive simulation experiments to evaluate the effectiveness of the proposed method and validate our theoretical results in Examples 1–7.

6.1. Baseline Method

More than the SCK algorithm, here we briefly provide an alternative spectral method that can also be applied to fit our adLCM. Recall that $R_0 = Z\Theta'$ under adLCM; it is easy to see that $R_0(i, :) = R_0(\bar{i}, :)$ when two distinct subjects i and $\bar{i}$ belong to the same extreme latent profile for $i, \bar{i} \in [N]$. Therefore, the population response matrix R_0 features K disparate rows, and running the K-means approach on all rows of R_0 with K clusters can faithfully recover the classification matrix Z in terms of a permutation of the K extreme latent profiles. $R_0 = Z\Theta'$ also gives that $\Theta = R_0'Z(Z'Z)^{-1}$, which suggests the following ideal algorithm called Ideal RMK Algorithm 3).

Algorithm 3 Ideal RMK

Require: R_0, K .

Ensure: A permutation of Z and Θ .

- 1: Run K-means algorithm on all rows of R_0 with K clusters to obtain $Z\mathcal{P}$, a permutation of Z .
 - 2: Compute $R_0'Z\mathcal{P}((Z\mathcal{P})'Z\mathcal{P})^{-1} = \Theta\mathcal{P}$, a permutation of Θ .
-

Algorithm 4, called RMK, is a natural generalization of the Ideal RMK from the oracle case to the real case because $\mathbb{E}(R) = R_0$ under adLCM. Unlike the SCK method, the RMK method does not need to obtain the SVD of the observed response matrix R .

Algorithm 4 Response Matrix with K-means (RMK for short)

Require: R, K .

Ensure: $\hat{Z}, \hat{\Theta}$.

- 1: Run K-means algorithm on all rows of R with K clusters to obtain $\hat{Z}$.
 - 2: Obtain an estimate of Θ by setting $\hat{\Theta} = R'\hat{Z}(\hat{Z}'\hat{Z})^{-1}$.
-

The computational cost of the first step in RMK is $O(lNJK)$, where l denotes the number of iterations for the K-means algorithm. The complexity of the second step in RMK is $O(JNK)$. Therefore, the overall computational cost of RMK is $O(lNJK)$. When $J = \beta N$ for a constant value $\beta \in (0, 1]$, the complexity of RMK is $O(\beta lKN^2)$, and it is larger than

the SCK's complexity $O(KN^2)$ when $\beta l > 1$. Therefore, SCK runs faster than RMK when $\beta l > 1$, as confirmed by our numerical results in this section.

We also compare our SCK and RMK approaches with several existing methods to highlight their superior performance. The comparative methods include the probabilistic latent component analysis (PLCA) algorithm [20] (an expectation maximization method), the nonnegative matrix factorization (NMF) technique [24], and the HeteroClustering (HC) algorithm [29] (a spectral clustering approach).

6.2. Evaluation Metric

For the classification of subjects, when the true classification matrix Z is known, to evaluate how good the quality of the partition of the subjects into extreme latent profiles is, four metrics are considered including the Clustering error $\hat{f}$ computed by Equation (7). The other three popular evaluation criteria are Hamming error [74], normalized mutual information (NMI) [75–78], and adjusted rand index (ARI) [78–80].

- Hamming error is defined as

$$\text{Hamming error} = N^{-1} \min_{\mathcal{P} \in \mathcal{P}_K} \|\hat{Z} - Z\mathcal{P}\|_0,$$

where $\mathcal{P}_K$ denotes the collection of all K -by- K permutation matrices. Hamming error falls within the range $[0, 1]$, and a smaller Hamming error indicates better classification performance.

- Let C be a $K \times K$ confusion matrix such that $C(k, l)$ is the number of common subjects between $\mathcal{T}_k$ and $\hat{\mathcal{T}}_l$ for $k, l \in [K]$. NMI is defined as

$$\text{NMI} = \frac{-2 \sum_{k,l} C(k, l) \log\left(\frac{C(k, l)N}{C_{k.} C_{.l}}\right)}{\sum_k C_{k.} \log\left(\frac{C_{k.}}{N}\right) + \sum_l C_{.l} \log\left(\frac{C_{.l}}{N}\right)},$$

where $C_{k.} = \sum_{m=1}^K C(k, m)$ and $C_{.l} = \sum_{m=1}^K C(m, l)$. NMI is in the range $[0, 1]$ and the larger it is, the better it is.

- ARI is defined as

$$\text{ARI} = \frac{\sum_{k,l} \binom{C(k, l)}{2} - \frac{\sum_k \binom{C_{k.}}{2} \sum_l \binom{C_{.l}}{2}}{\binom{N}{2}}}{\frac{1}{2} [\sum_k \binom{C_{k.}}{2} + \sum_l \binom{C_{.l}}{2}] - \frac{\sum_k \binom{C_{k.}}{2} \sum_l \binom{C_{.l}}{2}}{\binom{N}{2}}},$$

where $\binom{\cdot}{2}$ is a binomial coefficient. ARI falls within the range $[-1, 1]$ and the larger it is, the better it is.

For the estimation of Θ , we use the Relative l_1 error and the Relative l_2 error to evaluate the performance. The two criteria are defined as

$$\text{Relative } l_1 \text{ error} = \min_{\mathcal{P} \in \mathcal{P}_K} \frac{\|\hat{\Theta} - \Theta\mathcal{P}\|_1}{\|\Theta\|_1} \text{ and Relative } l_2 \text{ error} = \min_{\mathcal{P} \in \mathcal{P}_K} \frac{\|\hat{\Theta} - \Theta\mathcal{P}\|_F}{\|\Theta\|_F}.$$

The smaller both values are the better.

6.3. Synthetic Data

We conduct numerical studies to examine the accuracy and the efficiency of the aforementioned approaches by changing the scaling parameter ρ and the number of subjects N . Unless specified, in all computer-generated response matrices, we set $K = 3$, $J = \frac{N}{5}$, and the $N \times K$ classification matrix Z is generated such that each subject belongs to one of the K extreme latent profiles with equal probability. For distributions that require B 's entries to be

nonnegative, we let $B(j, k) = \text{rand}(1)$ for $j \in [J], k \in [K]$, where $\text{rand}(1)$ is a random value simulated from the uniform distribution on $[0, 1]$. For a Normal distribution and signed response matrix that allow B to have negative entries, we let $B(j, k) = 2\text{rand}(1) - 1$ for $j \in [J], k \in [K]$, i.e., $B(j, k)$ is in the range $[-1, 1]$. Set $B_{\max} = \max_{j \in [J], k \in [K]} |B(j, k)|$ because the generation process of B makes $|B(j, k)| \in [0, 1]$, but it cannot guarantee that $B_{\max} = 1$, which is required in the definition of B . Therefore, we update B by $\frac{B}{B_{\max}}$. For the scaling parameter ρ and the number of subjects N , they are set independently for each distribution. After setting all model parameters (K, N, J, Z, B, ρ) , we can generate the observed response matrix R from distribution $\mathcal{F}$ with expectation $R_0 = Z\Theta' = \rho ZB'$ under our adLCM. By applying each method to R with K extreme latent profiles, we can compute the evaluation metrics of each method. In every simulation scenario, we generate 50 independent replicates and report the mean of Clustering error (as well as Hamming error, NMI, ARI, Relative l_1 error, Relative l_2 error, and running time) computed from the 50 repetitions for each method. All numerical results in this paper are reported using MATLAB R2024b on a standard personal computer (Thinkpad X1 Carbon Gen 8).

6.3.1. Bernoulli Distribution

When $R(i, j) \sim \text{Bernoulli}(R_0(i, j))$ for $i \in [N], j \in [J]$, we consider the following two simulations.

Simulation 1(a): changing ρ . Set $N = 500$. For the Bernoulli distribution, the scaling parameter ρ should be set within the range $(0, 1]$ according to Example 1. Here, for simulation studies, we let ρ be in the range $\{0.1, 0.2, 0.3, \dots, 1\}$.

Simulation 1(b): changing N . Let $\rho = 0.1$ and N be in the range $\{1000, 2000, \dots, 5000\}$.

The results are presented in Figure 1. We observe that SCK and HC outperform the other three methods, while NMF and PLCA perform the poorest in estimating (Z, Θ) . As for running time, SCK and NMF run faster than their competitors across all settings. All methods achieve better performances as ρ increases, which conforms to our analysis in Example 1. Additionally, all algorithms enjoy better performances when the number of subjects N increases, as predicted by our analysis following Corollary 1.

6.3.2. Binomial Distribution

When $R(i, j) \sim \text{Binomial}(m, \frac{R_0(i, j)}{m})$ for $i \in [N], j \in [J]$, we consider the following two simulations.

Simulation 2(a): changing ρ . Set $N = 500$ and $m = 5$. Recall that ρ 's range is $(0, m]$ when $\mathcal{F}$ is a Binomial distribution according to Example 2; here, we let ρ be in the range $\{0.2, 0.4, 0.6, \dots, 2\}$.

Simulation 2(b): changing N . Let $\rho = 0.1, m = 5$, and N be in the range $\{1000, 2000, \dots, 5000\}$.

Figure 2 presents the corresponding results. We note that SCK, RMK, and HC enjoy similar error rates and they outperform NMF and PLCA in estimating (Z, Θ) . NMF runs slightly faster than SCK, while SCK runs faster than the other three approaches for this simulation. Meanwhile, increasing ρ (and N) decreases error rates for all methods, which confirms our findings in Example 2 and Corollary 1.

6.3.3. Poisson Distribution

When $R(i, j) \sim \text{Poisson}(R_0(i, j))$ for $i \in [N], j \in [J]$, we consider the following two simulations.

Simulation 3(a): changing ρ . Set $N = 500$. Example 3 says that the theoretical range of ρ is $(0, +\infty)$ when $\mathcal{F}$ is a Poisson distribution. Here, we let ρ be in the range $\{0.2, 0.4, 0.6, \dots, 2\}$.

Simulation 3(b): changing N . Let $\rho = 0.1$ and N be in the range $\{1000, 2000, \dots, 5000\}$.

Figure 3 displays the numerical results of Simulation 3(a) and Simulation 3(b). The results are similar to those of the Binomial distribution case: SCK, RMK, and HC perform similarly, while NMF and PLCA perform poorer in estimating (Z, Θ) . SCK runs slightly slower than NMF and both methods run faster than their competitors. All methods perform better as ρ and N increase, which supports our analysis in Example 3 and Corollary 1.

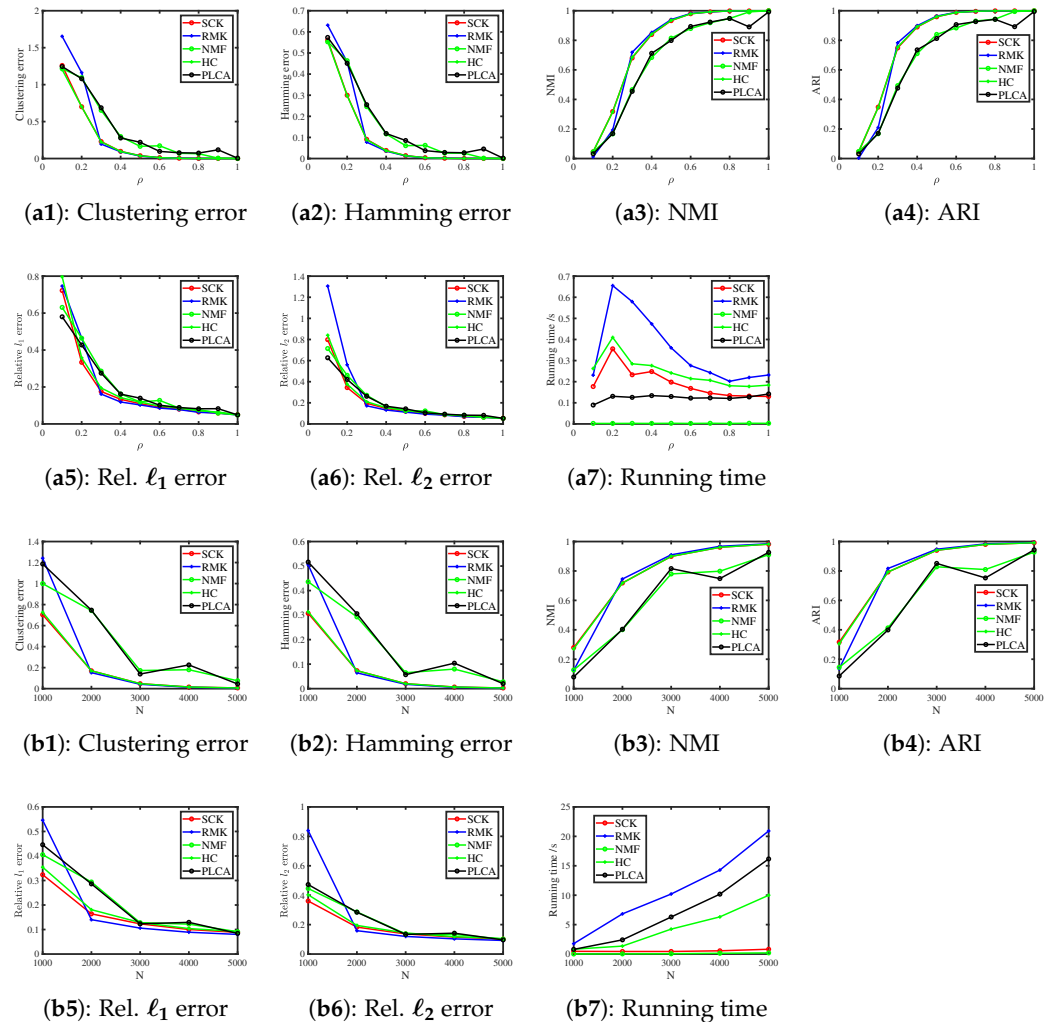


Figure 1. Numerical results of Simulation 1.

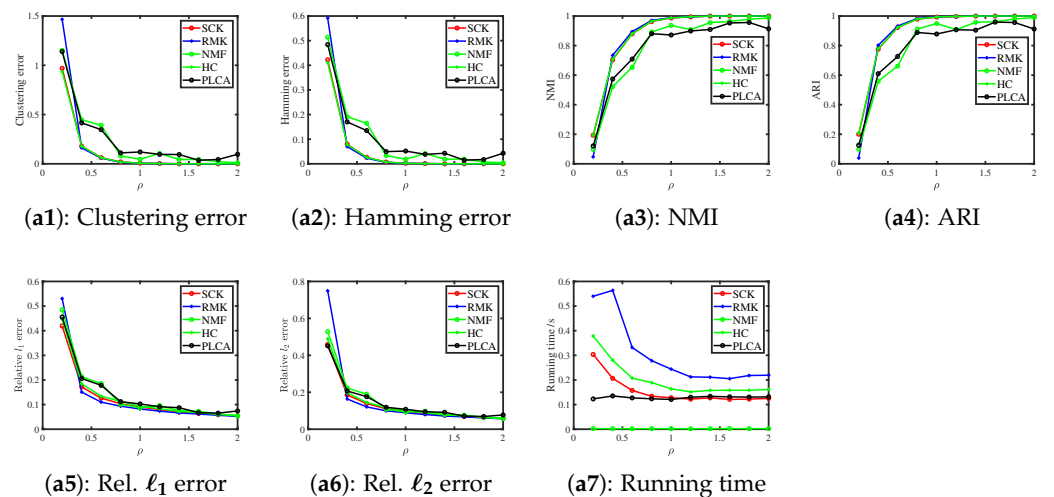


Figure 2. Cont.

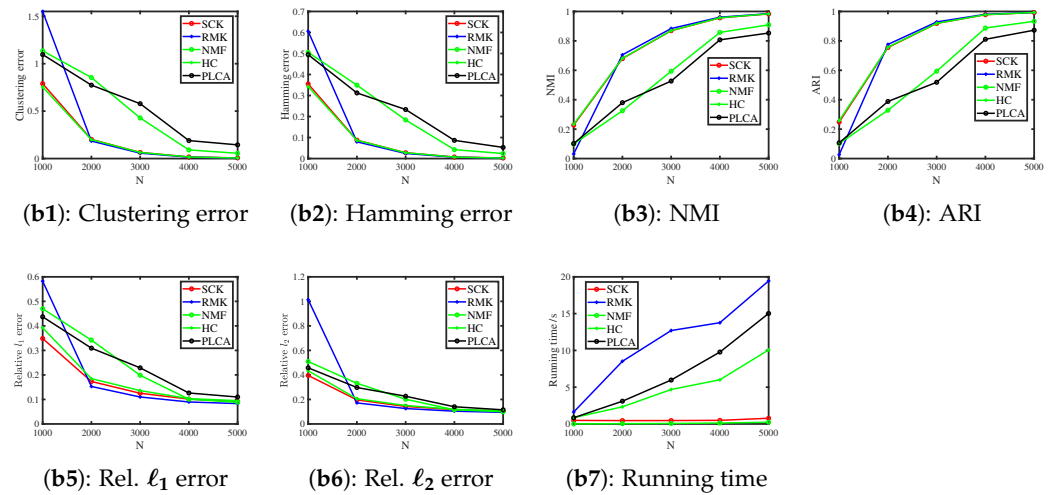


Figure 2. Numerical results of Simulation 2.

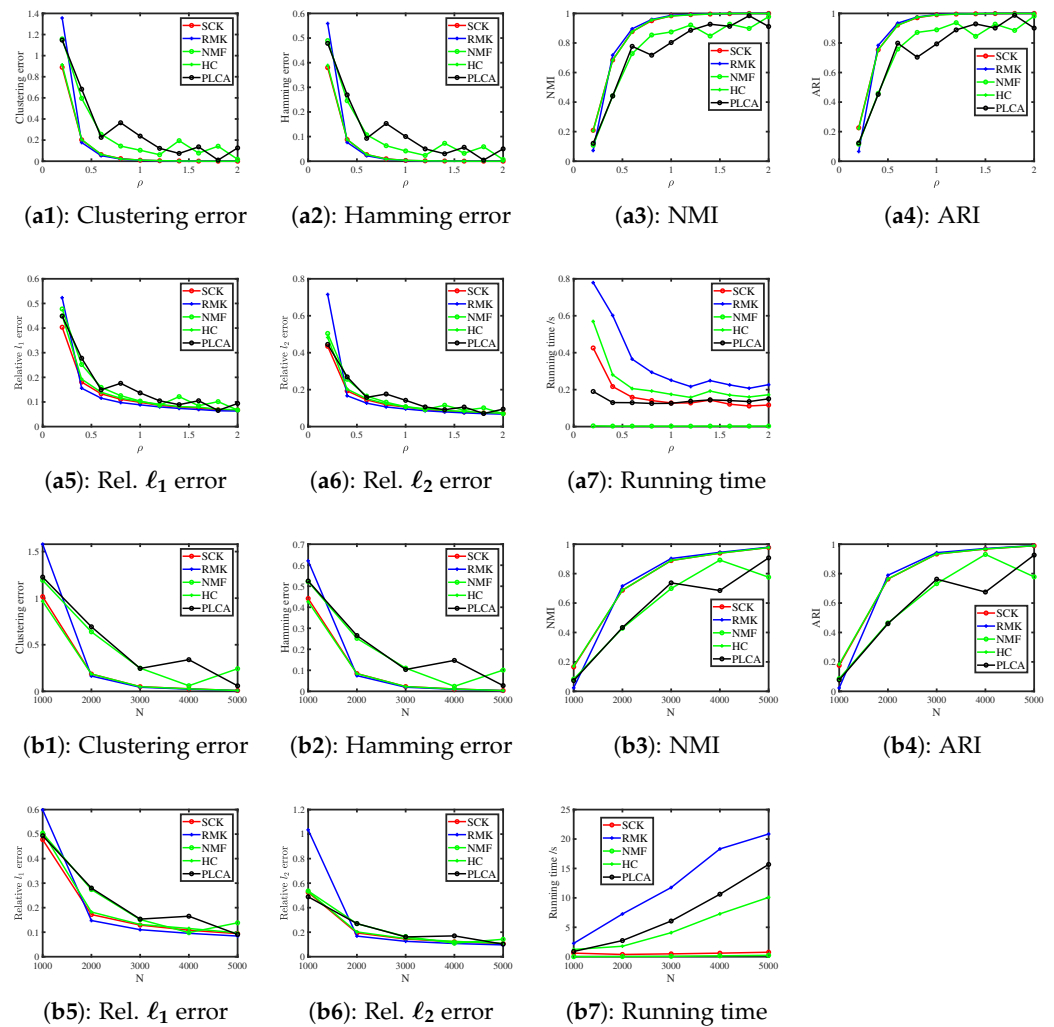


Figure 3. Numerical results of Simulation 3.

6.3.4. Normal Distribution

When $R(i, j) \sim \text{Normal}(R_0(i, j), \sigma^2)$ for $i \in [N], j \in [J]$, we consider the following two simulations.

Simulation 4(a): changing ρ . Set $N = 500$ and $\sigma^2 = 2$. According to Example 4, the scaling parameter ρ can be set as any positive value when $\mathcal{F}$ is a Normal distribution. Here, we let ρ be in the range $\{0.2, 0.4, 0.6, \dots, 2\}$.

Simulation 4(b): changing N . Let $\rho = 0.5, \sigma^2 = 2$, and N be in the range $\{1000, 2000, \dots, 5000\}$.

Figure 4 shows the results. We see that SCK, RMK, and HC have similar performances in estimating model parameters (Z, Θ) , while NMF and PLCA fail to estimate parameters in this simulation. For running time, SCK runs faster than RMK and HC. Additionally, the error rates of SCK, RMK, and HC decrease when the scaling parameter ρ and the number of subjects N increase, supporting our findings in Example 4 and Corollary 1.

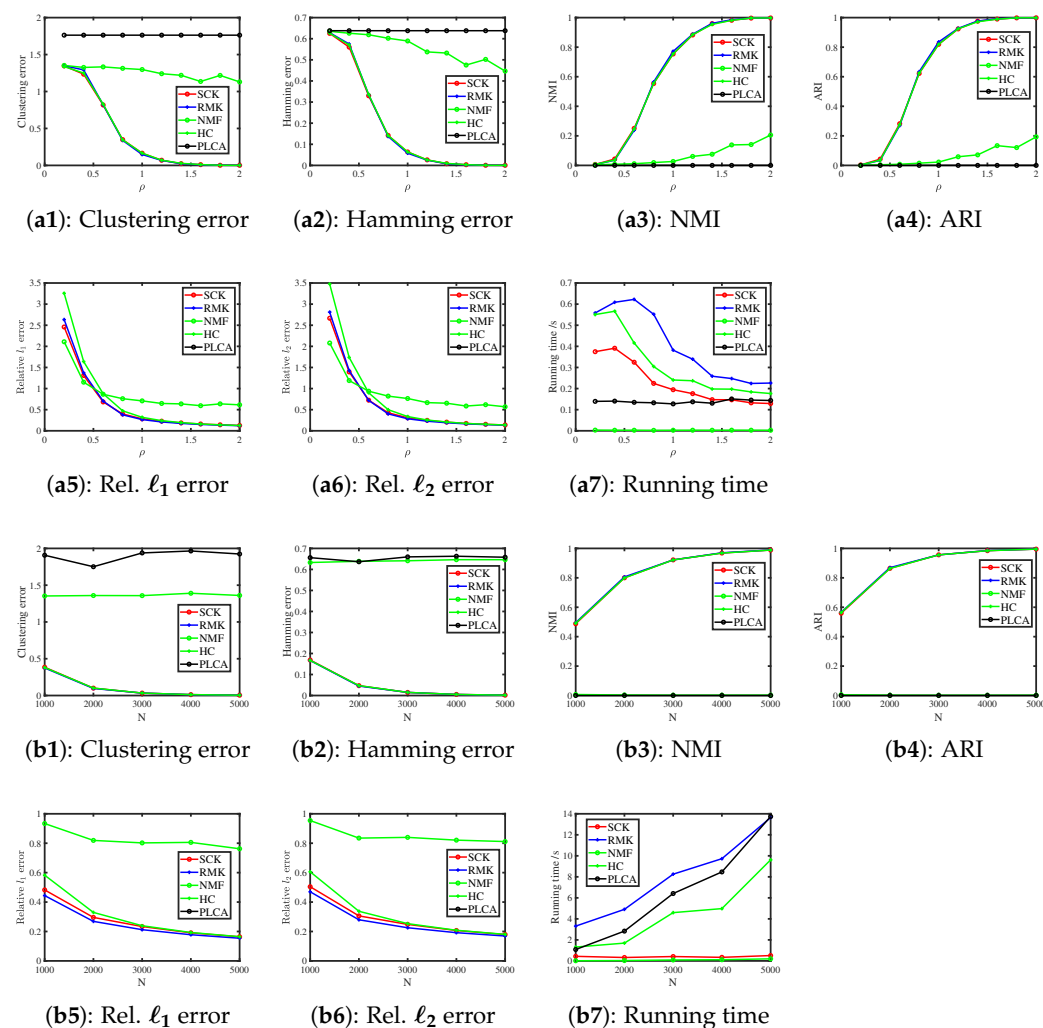


Figure 4. Numerical results of Simulation 4.

6.3.5. Exponential Distribution

When $R(i, j) \sim \text{Exponential}(\frac{1}{R_0(i, j)})$ for $i \in [N], j \in [J]$, we consider the following two simulations.

Simulation 5(a): changing ρ . Set $N = 300$. According to Example 5, the range of the scaling parameter ρ is $(0, +\infty)$ when $\mathcal{F}$ is an Exponential distribution. Here, we let ρ be in the range $\{1, 2, \dots, 20\}$ for our numerical studies.

Simulation 5(b): changing N . Let $\rho = 1$ and N be in the range $\{300, 600, \dots, 3000\}$.

Figure 5 displays the results. We see that SCK, RMK, and HC provide satisfactory estimations for Z and Θ for their small error rates, large NMI, and large ARI, while NMF and PLCA perform poorer. For running time, NMF and SCK run faster than their

competitors for large N . Meanwhile, we find that increasing ρ does not significantly influence the performances of these methods and this verifies our theoretical analysis in Example 5 that ρ disappears in the theoretical upper bounds of error rates by setting $\gamma = \rho^2$ in Theorem 1 for Exponential distribution. Furthermore, when we increase N , all methods perform better and this supports our analysis after Corollary 1.

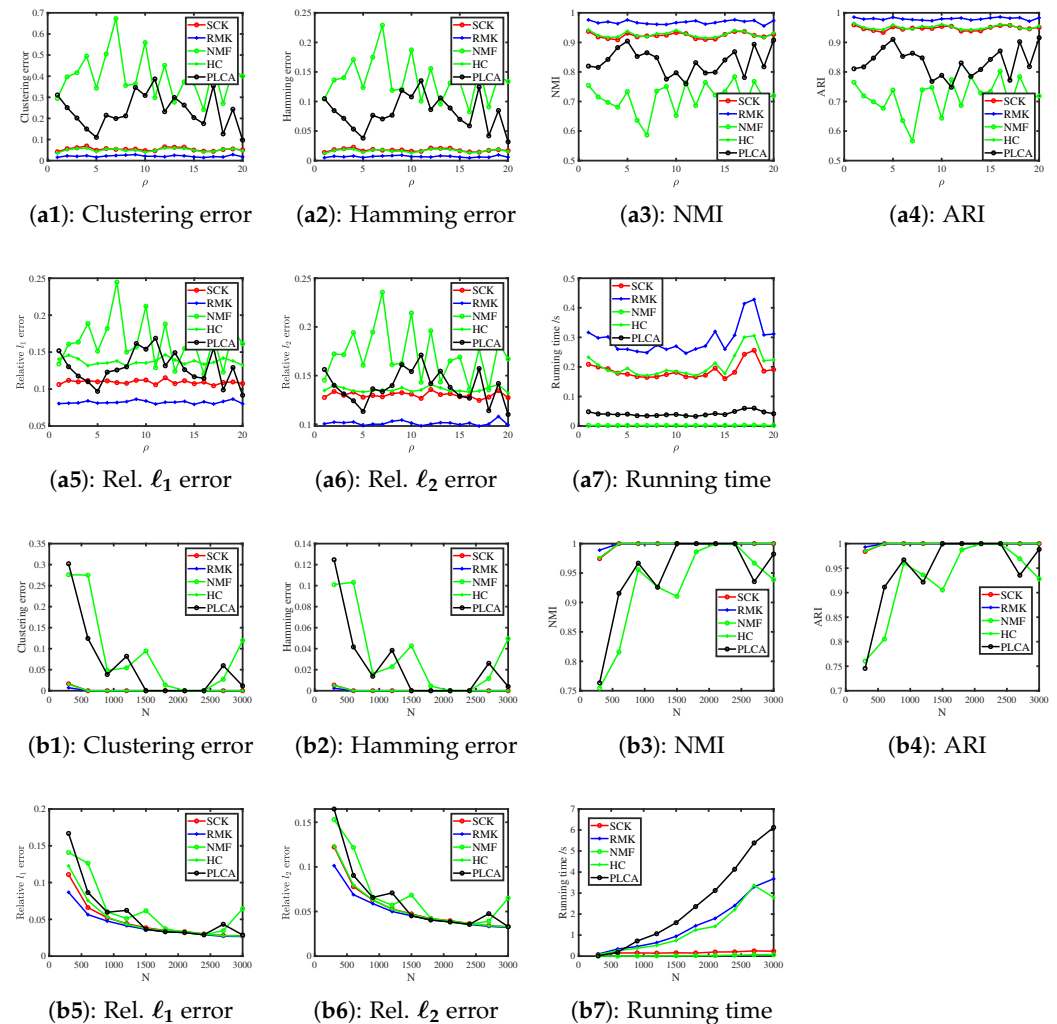


Figure 5. Numerical results of Simulation 5.

6.3.6. Uniform Distribution

When $R(i, j) \sim \text{Uniform}(0, 2R_0(i, j))$ for $i \in [N], j \in [J]$, we consider the following two simulations.

Simulation 6(a): changing ρ . Set $N = 120$. According to Example 6, the scaling parameter ρ can be set as any positive value when $\mathcal{F}$ is a Uniform distribution. Here, we let ρ be in the range $\{1, 2, \dots, 20\}$.

Simulation 6(b): changing N . Let $\rho = 1$ and N be in the range $\{300, 600, \dots, 3000\}$.

Figure 6 displays the numerical results. We see that increasing ρ does not significantly decrease or increase the estimation accuracies of these methods, which verifies our theoretical analysis in Example 6. For all settings, SCK runs faster than RMK and HC. When increasing N , the Clustering error and Hamming error (NMI and ARI) for SCK, RMK, and HC are 0 (1), and this suggests that they return the exact estimation of the classification matrix $\mathbf{Z}$. This phenomenon occurs because N is set quite large for Uniform distribution in Simulation 6(b). For the estimation of Θ , the error rates for SCK, RMK, and HC decrease when we increase N and this is consistent with our findings following Corollary 1. The

numerical results for running time are similar to previous simulations, and we omit the detailed analysis here for brevity.

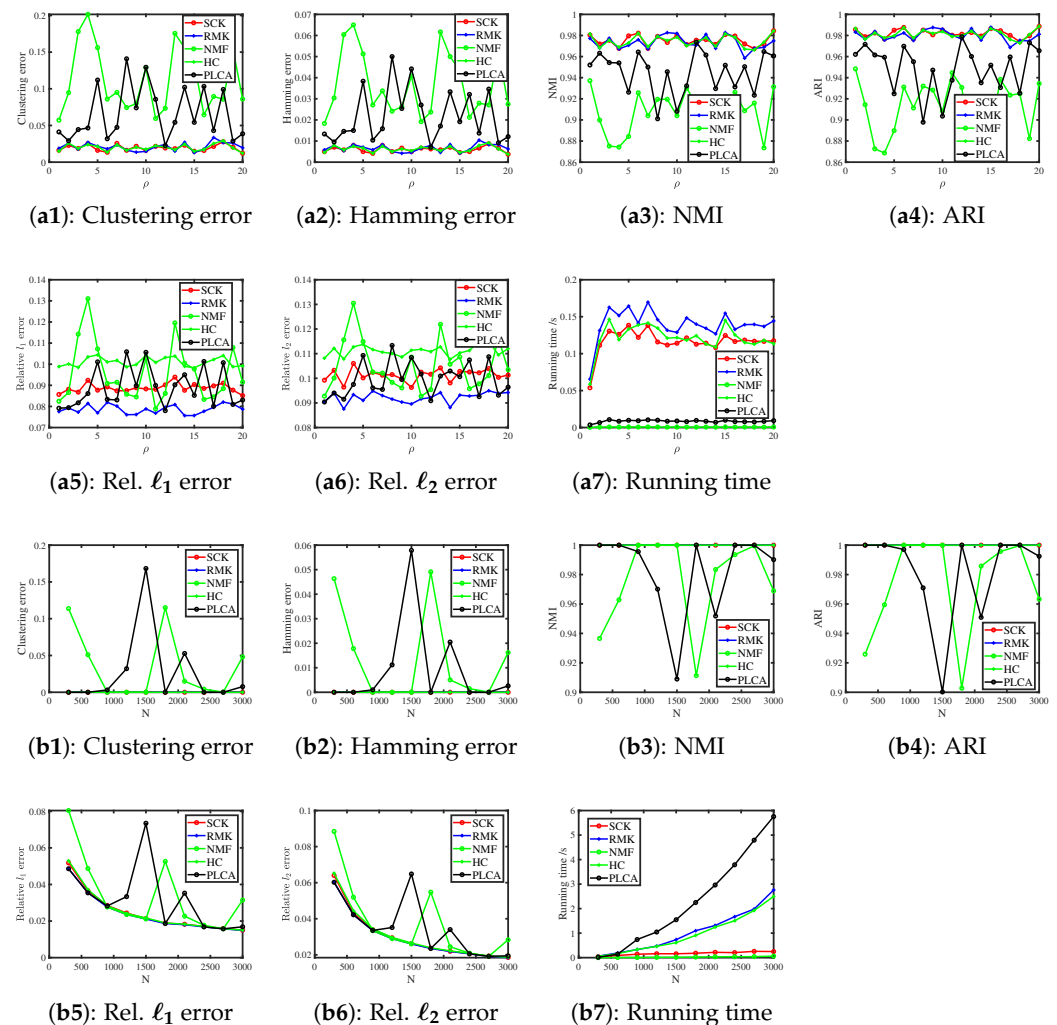


Figure 6. Numerical results of Simulation 6.

6.3.7. Signed Response Matrix

For signed response matrices when $\mathbb{P}(R(i, j) = 1) = \frac{1+R_0(i, j)}{2}$ and $\mathbb{P}(R(i, j) = -1) = \frac{1-R_0(i, j)}{2}$ for $i \in [N], j \in [J]$, we consider the following two simulations.

Simulation 7(a): changing ρ . Set $N = 500$. Recall that the theoretical range of the scaling parameter ρ is $(0, 1]$ for signed response matrices according to our analysis in Example 7; here, we let ρ be in the range $\{0.1, 0.2, \dots, 1\}$.

Simulation 7(b): changing N . Let $\rho = 0.2$ and N be in the range $\{1000, 2000, \dots, 5000\}$.

Figure 7 shows the results. We see that increasing ρ and N improves the estimation accuracies of SCK, HC, and RMK, which confirms our analysis in Example 7 and Corollary 1. Meanwhile, PLCA and NMF almost fail to estimate $\mathbf{Z}$ in this simulation. Additionally, it is easy to see that SCK, RMK, and HC enjoy similar performances in estimating $\mathbf{Z}$ and Θ , and SCK requires less computation time compared to RMK and HC.

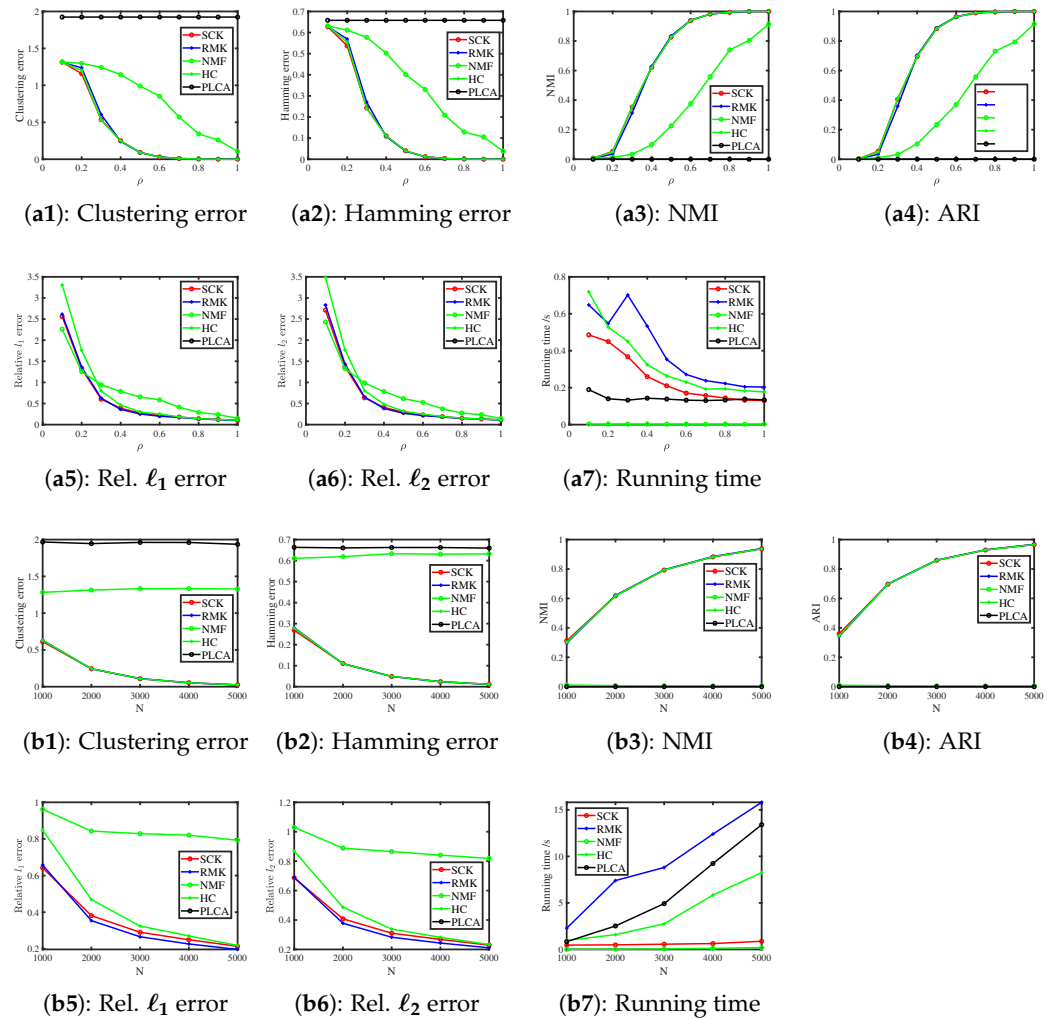


Figure 7. Numerical results of Simulation 7.

6.3.8. Simulated Arbitrary-Distribution Response Matrices

For visuality, we plot two response matrices R generated from the Normal distribution and the Poisson distribution under adLCM. Let $K = 2$, $N = 16$, $J = 10$, $\sigma^2 = 1$, $\ell(i) = 1$, $\ell(i + 8) = 2$ for $i \in [8]$, and $\Theta(j, 1) = 100$, $\Theta(j, 2) = 110 - 10j$ for $j \in [10]$. Because $R_0 = Z\Theta'$ has been set, we can generate R under different distributions with expectation R_0 under the proposed adLCM. Here, we consider the following two settings.

Simulation 8 (a): When $R(i, j) \sim \text{Normal}(R_0(i, j), \sigma^2)$ for $i \in [N]$, $j \in [J]$.

Simulation 8 (b): When $R(i, j) \sim \text{Poisson}(R_0(i, j))$ for $i \in [N]$, $j \in [J]$.

Figure 8 displays the response matrices R generated for Simulation 8 (a) and 8 (b), respectively. Error rates of these methods for the corresponding observed response matrices are displayed in Table 1. We also plot the estimated item matrix $\hat{\Theta}$ for our SCK and RMK in Figure 9. We see that all approaches exactly recover Z from R , while they estimate Θ with slight perturbations. Meanwhile, since Z , Θ , and K are known for this simulation, the R provided in Figure 8 can be regarded as benchmark response matrices, and readers can apply SCK and RMK (and other methods) to R to check their effectiveness in estimating Z and Θ .

In summary, we conduct extensive numerical experiments across a wide spectrum of data-generating distributions to evaluate the SCK and RMK methods compared against various benchmarks. The proposed SCK algorithm consistently demonstrates superior performance and efficiency for latent class analysis of various types of response. Across all simulation settings, SCK achieves highly accurate estimates of both the latent class memberships Z and

item parameters Θ , often matching or exceeding the performance of alternative methods like RMK and HC while significantly outperforming NMF and PLCA. Notably, SCK demonstrates exceptional computational efficiency, consistently running faster than RMK and HC, and substantially faster than the EM-based PLCA method, which is the slowest across all scenarios. Furthermore, the experiments reveal a critical limitation of NMF and PLCA: the two methods consistently fail to provide reliable parameter estimates when the response matrix contains negative values, as evidenced by their poor performance in simulations involving Normal distributions and signed response matrices. In contrast, SCK robustly handles all distributions, including those with negative or continuous responses, confirming its versatility as a powerful and efficient tool for the analysis of data under the proposed adLCM framework.

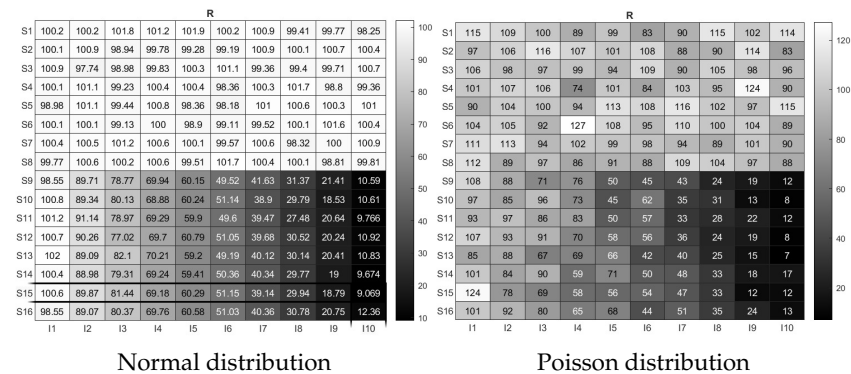


Figure 8. Illustration for response matrices R generated from adLCM. In both panels, S_i denotes subject i and I_j denotes item j for $i \in [16], j \in [10]$.

Table 1. Error rates of different methods for R in Figure 8. Values outside the parentheses correspond to the first panel of Figure 8, and those inside correspond to the second panel of Figure 8.

Method	Clustering Error	Hamming Error	NMI	ARI	Rel. ℓ_1 Error	Rel. ℓ_2 Error
SCK	0 (0)	0 (0)	1.000 (1.000)	1.000 (1.000)	0.0024 (0.0254)	0.0032 (0.0295)
RMK	0 (0)	0 (0)	1.000 (1.000)	1.000 (1.000)	0.0024 (0.0245)	0.0032 (0.0295)
NMF	0 (0)	0 (0)	1.000 (1.000)	1.000 (1.000)	0.0024 (0.0245)	0.0032 (0.0295)
HC	0 (0)	0 (0)	1.000 (1.000)	1.000 (1.000)	0.0024 (0.0254)	0.0032 (0.0307)
PLCA	0 (0)	0 (0)	1.000 (1.000)	1.000 (1.000)	0.0024 (0.0245)	0.0032 (0.0295)

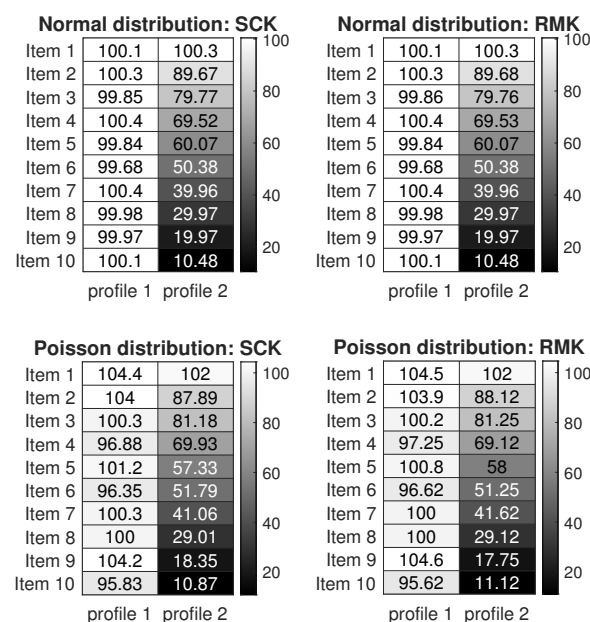


Figure 9. Heatmap of the estimated item parameter matrix $\hat{\Theta}$ of SCK and RMK for R in Figure 8.

7. Real Data Applications

As the main goal of this paper is to introduce the proposed adLCM and the SCK algorithm for arbitrary-distribution response matrices, this section reports empirical results on two datasets. Because the true classification matrix and the true item parameter matrix are unknown for real data, and SCK runs much faster than RMK and HC though they may perform similarly in simulations, we only report the outcomes of the SCK approach. For real-world datasets, the number of extreme latent profiles K is often unknown. Here, we infer K for real-world data using the following strategy:

$$K = \arg \min_{k \in [\text{rank}(R)]} \|R - \hat{Z}\hat{\Theta}'\|, \quad (8)$$

where $\hat{Z}$ and $\hat{\Theta}$ are outputs in Algorithm 2 with inputs R and k . The method specified in Equation (8) selects K by picking the one that minimizes the spectral norm difference between R and $\hat{Z}\hat{\Theta}'$. The determination of the number of extreme latent profiles K in our adLCM in a rigorous manner with theoretical guarantees remains a future direction.

7.1. International Personality Item Pool (IPIP) Personality Test Data

Background. We apply SCK to an experiment personality test dataset called the International Personality Item Pool (IPIP) personality test, which is obtainable for download at https://openpsychometrics.org/_rawdata/ (accessed on 11 August 2025). This data consists of 1005 subjects and 40 items. The IPIP data also records the age and gender of each subject. After dropping subjects with missing entries in their responses, age, or gender, and dropping two subjects that are neither male nor female, there are 896 subjects left, i.e., $N = 896, J = 40$. All items are rated on a 5-point scale, where 1 = Strongly disagree, 2 = Disagree, 3 = Neither agree nor disagree, 4 = Agree, 5 = Strongly agree, i.e., $R \in \{1, 2, 3, 4, 5\}^{896 \times 40}$, a response matrix. Items 1–10 measure the personality factor Assertiveness (short as “AS”); Items 11–20 measure the personality factor Social confidence (short as “SC”); Items 21–30 measure the personality factor Adventurousness (short as “AD”); and Items 31–40 measure the personality factor Dominance (short as “DO”). The details of each item are depicted in Figure 10.

Analysis. We apply Equation (8) to infer K for the IPIP dataset and find that the estimated value of K is 3. We then apply the SCK algorithm to the response matrix R with $K = 3$ to obtain the 896×3 matrix $\hat{Z}$ and the 40×3 matrix $\hat{\Theta}$. The running time for SCK on this dataset is around 0.2 s.

Results. For convenience, we denote the estimated three extreme latent profiles as profile 1, profile 2, and profile 3. Based on $\hat{Z}$ and the information of age and gender, we can obtain some basic information (shown in Table 2) such as the size of each profile, number of males (females) in each profile, and the average age of males (and females) in each profile. From Table 2, we see that the number of females is larger than that of males for profile 1, while profiles 2 and 3 have more males. The average age of males (and females) in profile 2 is smaller than that of profiles 1 and 3, while the average age of females in profile 3 is the largest. We can also obtain the average point on each item for males (and females) in each estimated extreme latent profile and the results are shown in Figure 10. We observe that males in profile 3 tend to be more confident, more creative, more social, and more open to changes than males in profiles 1 and 2; males in profile 3 are more (less) dominant than males in profile 1 (profile 2). Males in profile 2 are more confident, creative, social, open to changes, and dominant than males in profile 1. Meanwhile, in the three estimated extreme latent profiles, females enjoy similar personalities to males. We also find that males in profile 3 (profile 2) are more (less) confident, creative, social, open to changes, and dominant than females in profile 3 (profile 2). Furthermore, it is interesting to see that, though males

in profile 1 are less confident, creative, social, and open to changes than females in profile 1, they are more dominant than females in profile 1. We also plot the average point on each item in each estimated extreme latent profile regardless of gender in Figure 10 where we can draw similar conclusions as those for males. In Figure 10, we also plot the heatmap of the estimated item parameter matrix $\hat{\Theta}$. By comparing these results shown in Figure, we see that the (j, k) -th element in the matrix shown in the third panel of Figure 10 is close to $\hat{\Theta}(j, k)$ for $j \in [40], k \in [3]$. Such a result implies that the behavior differences on each item for every extreme latent profile are governed by the item parameter matrix Θ .

Remark 6. Recall that $\mathbb{E}(R) = R_0 = Z\Theta'$ under the adLCM; then we have $R_0(i, j) = \Theta(j, \ell(i))$ for $i \in [N], j \in [J]$. Then we have $\sum_{\ell(i) \equiv k} R_0(i, j) = \sum_{\ell(i) \equiv k} \Theta(j, \ell(i)) = \sum_{\ell(i) \equiv k} \Theta(j, k) = N_k \Theta(j, k)$, which gives that $\Theta(j, k) = \frac{\sum_{\ell(i) \equiv k} R_0(i, j)}{N_k}$ for $k \in [K]$. This explains why the average value on the j -th item in the k -th estimated extreme latent profile approximates $\hat{\Theta}(j, k)$ for $j \in [J], k \in [K]$.

Table 2. Descriptive statistics for each estimated extreme latent profile from $\hat{Z}$ on the IPIP dataset.

Statistic	Profile 1	Profile 2	Profile 3
Size	276	226	394
# Male	123	129	241
# Female	153	97	153
Average age (male)	35.98	32.82	35.90
Average age (female)	35.54	31.38	38.71

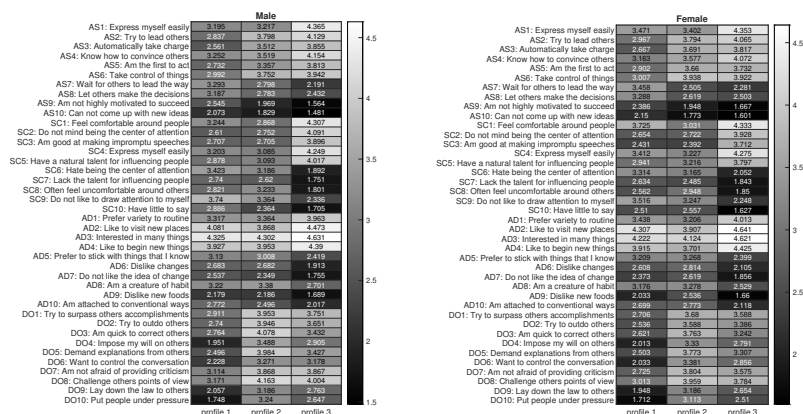
7.2. Big Five Personality Test with Random Number (BFPTRN) Data

Background. Our SCK method is also applied to personality test data: the Big Five Personality Test with Random Number (BFPTRN) data. This dataset can be downloaded from the same URL as the IPIP data. This data asks respondents to generate random numbers in certain ranges attached to 50 personality items. The Big Five personality traits are extraversion (items E1–E10), neuroticism (items N1–N10), agreeableness (items A1–A10), conscientiousness (items C1–C10), and openness (items O1–O10). The original BFPTRN data contains 1369 subjects. After excluding subjects with missing responses or missing random numbers and removing those with random numbers exceeding the specified range, there remain 1155 subjects, i.e., $N = 1155, J = 50$. All items are rated using the same 5-point scale as the IPIP data, which results in $R \in \{1, 2, 3, 4, 5\}^{1155 \times 50}$. The detail of each item and each range for random numbers can be found in Figure 11.

Analysis. The estimated number of extreme latent profiles for the BFPTRN dataset is 3. Applying the SCK approach to R with $K = 3$ produces the 1155×3 matrix $\hat{Z}$ and the 50×3 matrix $\hat{\Theta}$. SCK takes around 1.6 s to process this data.

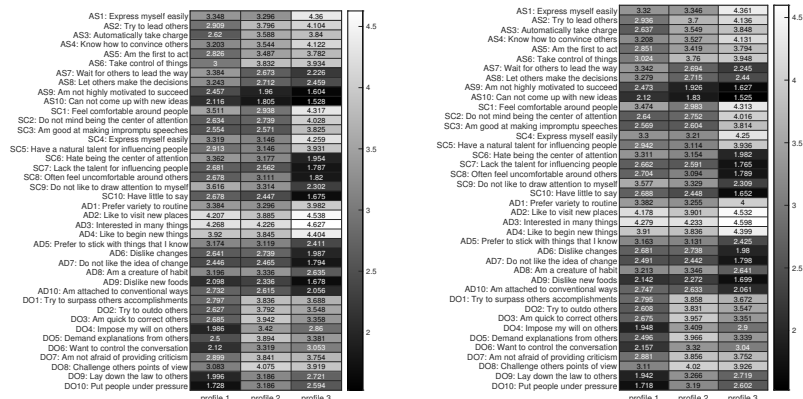
Results. Without confusion, we also let profile 1, profile 2, and profile 3 represent the three estimated extreme latent profiles. Profiles 1, 2, and 3 have 409, 320, and 426 subjects, respectively. Similar to the IPIP data, based on $\hat{Z}$ and $\hat{\Theta}$, we can also obtain the heatmap of the average point on each subject for every profile, the heatmap of the average random number on each range for every profile, and the heatmap of $\hat{\Theta}$ as shown in Figure 11. We observe that there is no significant connection between the average point and the average random number on each item in each estimated extreme latent profile. From the first panel of Figure 11, we find that: for extroversion, subjects in profile 1 are the most extroverted, while subjects in profile 2 are the most introverted; for neuroticism, subjects in profile 3 are emotionally stable, while subjects in profiles 1 and 2 are emotionally unstable; for agreeableness, subjects in profiles 1 and 3 are easier to get along with than subjects in profile 2; for conscientiousness, subjects in profile 3 are more responsible than those in profiles 1

and 2; and for openness, subjects in profiles 1 and 3 are more open than those in profile 2. Meanwhile, the matrix shown in the first panel of Figure 11 approximates $\hat{\Theta}$ well, which has been explained in Remark 6.



Average point on each item for male in each estimated extreme latent profile

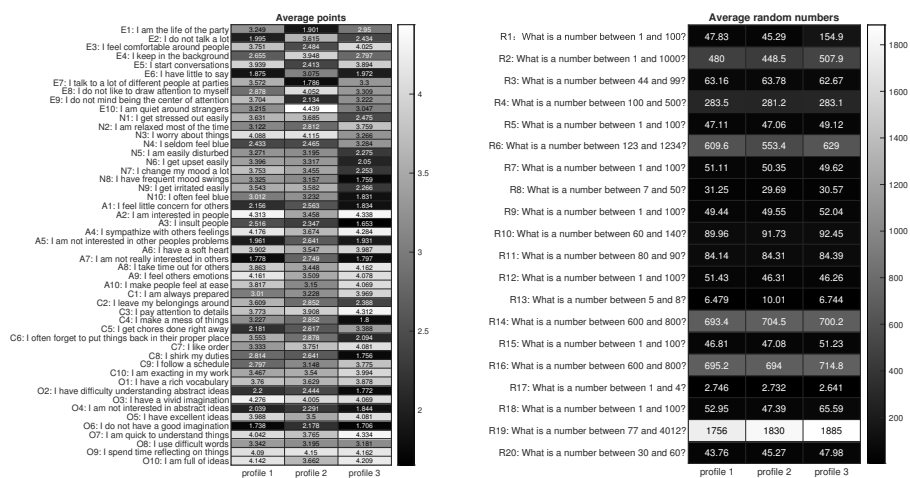
Average point on each item for female in each estimated extreme latent profile



Average point on each item in each estimated extreme latent profile

Heatmap of $\hat{\Theta}$

Figure 10. Numerical results for the IPIP data.



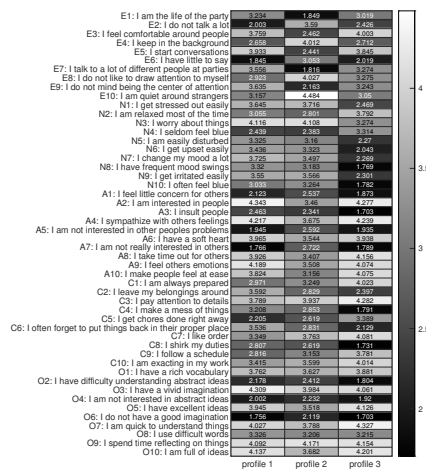
Heatmap of $\hat{\Theta}$

Figure 11. Numerical results for the BFPTRN data.

8. Conclusions and Future Work

In this paper, we introduced the arbitrary-distribution latent class model (adLCM), a novel class of latent class analysis models for data with arbitrary-distribution responses. We studied its model identifiability, developed an efficient inference method SCK to fit adLCM, and built a theoretical guarantee of estimation consistency for the proposed method under adLCM. On the methodology side, the new model adLCM provides exploratory and useful tools for latent class analysis in applications where the data may have arbitrary-distribution responses. adLCM allows the observed response matrix to be generated from any distribution as long as its expectation follows a latent class structure modeled by adLCM. In particular, the popular latent class model is a sub-model of our adLCM, and data with signed responses can also be modeled by adLCM. Ground-truth latent classes of data with responses generated from adLCM serve as benchmarks for evaluating latent class analysis approaches. On the algorithmic side, the SVD-based spectral method SCK is efficient and easy to implement. SCK requires no tuning parameters and it is applicable for data with arbitrary-distribution responses. This means that researchers in fields such as social, psychological, behavioral, and biological sciences, and beyond can design their tests/evaluations/surveys/interviews without worrying that the response should be binary or positive, as our method SCK is applicable for any kind of response matrices in latent class analysis. On the theoretic side, we established the rate of convergence for our method SCK under the proposed adLCM. We found that SCK exhibits different behaviors when the response matrices are generated from different distributions, and we conducted extensive experiments to verify our theoretical findings. Empirically, we applied our method to two real personality test datasets with meaningful results. We expect that our adLCM and SCK method will have broad applications for latent class analysis in understanding human behaviors for diverse fields, similar to the widespread use of latent class models in recent years.

There are several future directions worth exploring. First, methods with theoretical guarantees should be designed to determine the number of extreme latent profiles K for observed response matrices generated from any distribution $\mathcal{F}$ under adLCM. Following [81], a possible approach to estimate K is to count the number of significant singular values (i.e., those above a noise threshold) of R . Second, the grade of membership (GoM) model [52,82] provides a richer modeling capacity than the latent class model since GoM allows a subject to belong to multiple extreme latent profiles. Therefore, following the distribution-free idea developed in this work, it is meaningful to extend the model GoM

to data with arbitrary-distribution responses. Third, like the LCM being equipped with individual covariates [50,83–87], it is worth considering adding individual covariates into the adLCM analysis. Fourth, our adLCM only considers static latent class analysis and it is meaningful to extend adLCM to the dynamic case [88]. Fifth, our SCK is a spectral clustering method, and it is possible to speed it up by application of the random-projection techniques [63] or the distributed spectral clustering idea [89] to deal with large-scale data for latent class analysis. Finally, while our proposed methods can estimate latent class memberships and item parameters, a meaningful extension would be to develop methodology and theoretical guarantees for explicitly estimating the latent distribution parameters in future work.

Author Contributions: H.Q.: conceptualization, data curation, formal analysis, funding acquisition, methodology, project administration, resources, software, validation, visualization, writing—original draft, and writing—review and editing. X.X.: data curation, funding acquisition, project administration, resources, software, validation, visualization, writing—original draft, and writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: Qing’s work was sponsored by the Scientific Research Foundation of Chongqing University of Technology (Grant No. 2024ZDR003), the Science and Technology Research Program of Chongqing Municipal Education Commission (Grant No. KJQN202401168), and the Natural Science Foundation of Chongqing, China (Grant No: CSTB2023NSCQ-LZX0048). Xu’s research was supported the National Natural Science Foundation of China (Grant No. 12301358 and 42450275) and the Fundamental Research Funds for the Central Universities (No. 2042025kf0051), Wuhan University.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Data and code will be made available on request.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Proofs Under adLCM

Appendix A.1. Proof of Proposition 1

Proof. According to Lemma 1, we know that $\mathbf{U} = \mathbf{Z}\mathbf{X}$, where $\mathbf{X} = \mathbf{\Theta}'\mathbf{V}\mathbf{\Sigma}^{-1}$. Similarly, $\mathbf{U}$ can be rewritten as $\mathbf{U} = \tilde{\mathbf{Z}}\tilde{\mathbf{X}}$, where $\tilde{\mathbf{X}} = \tilde{\mathbf{\Theta}}'\mathbf{V}\mathbf{\Sigma}^{-1}$. Then, for $i \in [N]$, we have

$$\mathbf{U}(i,:) = \mathbf{Z}(i,:) \mathbf{X} = \mathbf{X}(\ell(i),:) = \tilde{\mathbf{X}}(\tilde{\ell}(i),:) = \tilde{\mathbf{X}}(\tilde{\ell}(i),:), \quad (\text{A1})$$

where $\tilde{\ell}(i)$ denotes the extreme latent profile that the i -th subject belongs in the alternative classification matrix $\tilde{\mathbf{Z}}$. For $\bar{i} \in [N]$ and $\bar{i} \neq i$, we have

$$\mathbf{U}(\bar{i},:) = \mathbf{Z}(\bar{i},:) \mathbf{X} = \mathbf{X}(\ell(\bar{i}),:) = \tilde{\mathbf{X}}(\tilde{\ell}(\bar{i}),:) = \tilde{\mathbf{X}}(\tilde{\ell}(\bar{i}),:). \quad (\text{A2})$$

When $\ell(i) = \ell(\bar{i})$, by the second statement of Lemma 1, we get $\mathbf{U}(i,:) = \mathbf{U}(\bar{i},:)$. Combining this fact (i.e., $\mathbf{U}(i,:) = \mathbf{U}(\bar{i},:)$) with Equations (A1) and (A2) leads to

$$\mathbf{X}(\ell(i),:) = \mathbf{X}(\ell(\bar{i}),:) = \tilde{\mathbf{X}}(\tilde{\ell}(i),:) = \tilde{\mathbf{X}}(\tilde{\ell}(\bar{i}),:) \text{ when } \ell(i) = \ell(\bar{i}). \quad (\text{A3})$$

Equation (A3) implies that $\tilde{\ell}(i) = \tilde{\ell}(\bar{i})$ when $\ell(i) = \ell(\bar{i})$, i.e., for any two distinct subjects i and $\bar{i}$, they are in the same extreme latent profile under $\tilde{\mathbf{Z}}$ when they are in the same extreme latent profile under $\mathbf{Z}$. Therefore, we have $\tilde{\mathbf{Z}} = \mathbf{Z}\mathbf{P}$, where $\mathbf{P}$ is a permutation matrix. Combining $\tilde{\mathbf{Z}} = \mathbf{Z}\mathbf{P}$ with $\mathbf{Z}\mathbf{\Theta}' = \tilde{\mathbf{Z}}\tilde{\mathbf{\Theta}}'$ leads to $\mathbf{Z}\mathbf{\Theta}' = \tilde{\mathbf{Z}}\tilde{\mathbf{\Theta}}' = \mathbf{Z}\mathbf{P}\tilde{\mathbf{\Theta}}'$, which gives that

$$\mathbf{Z}(\mathbf{\Theta}' - \mathbf{P}\tilde{\mathbf{\Theta}}') = \mathbf{0}. \quad (\text{A4})$$

Taking the transposition of Equation (A4) gives

$$(\Theta - \tilde{\Theta}\mathcal{P}')Z' = 0. \quad (\text{A5})$$

Right multiplying Z at both sides of Equation (A6) gives

$$(\Theta - \tilde{\Theta}\mathcal{P}')Z'Z = 0. \quad (\text{A6})$$

Since each extreme latent profile is not an empty set, the $N \times K$ classification matrix Z has a rank K , which gives that the $K \times K$ matrix $Z'Z$ is nonsingular. Therefore, right multiplying $(Z'Z)^{-1}$ at both sides of Equation (A6) gives $\Theta = \tilde{\Theta}\mathcal{P}'$, i.e., $\tilde{\Theta} = \Theta\mathcal{P}$ since $\mathcal{P}$ is a permutation matrix. $\square$

Appendix A.2. Proof of Lemma 1

Proof. For the first statement: Since $R_0 = Z\Theta' = U\Sigma V'$, $V'V = I_{K_0 \times K_0}$, and the $K_0 \times K_0$ diagonal matrix Σ is nonsingular, we have $U = Z\Theta'V\Sigma^{-1} \equiv ZX$, where $X = \Theta'V\Sigma^{-1}$. Hence, the first statement holds.

For the second statement: For $i \in [N]$, $U = ZX$ gives $U(i,:) = Z(i,:)X = X(\ell(i),:)$. Then, if $\ell(\bar{i}) = \ell(i)$, we have $U(\bar{i},:) = X(\ell(\bar{i}),:) = X(\ell(i),:) = U(i,)$, i.e., U has K distinct rows. Thus, the second statement holds.

For the third statement: Since $R_0 = Z\Theta' = U\Sigma V'$, we have $\Theta Z' = V\Sigma U' \Rightarrow \Theta Z'Z = V\Sigma U'Z \Rightarrow \Theta = V\Sigma U'Z(Z'Z)^{-1}$ where the $K \times K$ matrix $Z'Z$ is nonsingular because each extreme latent profile has at least one subject, i.e., $\text{rank}(Z'Z) = \text{rank}(Z) = K$. Thus, the third statement holds.

For the fourth statement: Recall that when $K_0 = K$, we have $U \in \mathbb{R}^{N \times K}$, $V \in \mathbb{R}^{J \times K}$, Σ is a $K \times K$ full-rank diagonal matrix, and X is a $K \times K$ matrix, where $U'U = I_{K \times K}$, $V'V = I_{K \times K}$. Let $\Delta = \text{diag}(\sqrt{N_1}, \sqrt{N_2}, \dots, \sqrt{N_K})$, then

$$R_0 = Z\Theta' = Z\Delta^{-1}\Delta\Theta'. \quad (\text{A7})$$

It is straightforward to verify that $Z\Delta^{-1}$ is a column orthogonal matrix, i.e., $(Z\Delta^{-1})'Z\Delta^{-1} = I_{K \times K}$.

Since $K_0 = K$, we have $\text{rank}(\Delta\Theta') = K$. Let $\tilde{U}\tilde{\Sigma}\tilde{V}' = \Delta\Theta'$ be the compact SVD of $\Delta\Theta'$, where $\tilde{\Sigma}$ is a $K \times K$ diagonal matrix, $\tilde{U} \in \mathbb{R}^{K \times K}$, $\tilde{V} \in \mathbb{R}^{J \times K}$, $\tilde{U}'\tilde{U} = I_{K \times K}$, and $\tilde{V}'\tilde{V} = I_{K \times K}$. Note that $\tilde{U} \in \mathbb{R}^{K \times K}$ and $\tilde{U}'\tilde{U} = I_{K \times K}$ imply $\text{rank}(\tilde{U}) = K$. Equation (A7) implies

$$R_0 = Z\Theta' = U\Sigma V' = Z\Delta^{-1}\Delta\Theta' = Z\Delta^{-1}\tilde{U}\tilde{\Sigma}\tilde{V}'. \quad (\text{A8})$$

Note that $U, V, Z\Delta^{-1}\tilde{U}$, and $\tilde{V}$ are all orthonormal matrices. Moreover, note that Σ and $\tilde{\Sigma}$ are $K \times K$ diagonal matrices. Then we have

$$U = Z\Delta^{-1}\tilde{U}, \Sigma = \tilde{\Sigma}, \text{ and } V = \tilde{V}. \quad (\text{A9})$$

Recall that $U = ZX$; Equation (A9) gives that $X = \Delta^{-1}\tilde{U} \in \mathbb{R}^{K \times K}$ and $\text{rank}(X) = K$ because $\text{rank}(\Delta) = K$ and $\text{rank}(\tilde{U}) = K$. We can easily verify that the rows of $\Delta^{-1}\tilde{U}$ are perpendicular to each other and the k -th row has length $\sqrt{1/N_k}$ for $k \in [K]$, i.e., $\sqrt{XX'} = \sqrt{\Delta^{-1}\tilde{U}\tilde{U}'\Delta^{-1}} = \sqrt{\Delta^{-2}} = \Delta^{-1}$. Thus, the fourth statement holds.

Remark A1. In this remark, we provide the reason why the fourth statement does not hold when $K_0 < K$. For this case, the rank of $\Delta\Theta'$ is K_0 , and thus $\tilde{U} \in \mathbb{R}^{K \times K_0}$ and $\text{rank}(\tilde{U}) = K_0$.

Then we have $X = \Delta^{-1}\tilde{U} \in \mathbb{R}^{K \times K_0}$ and $\text{rank}(X) = K_0$. Thus, $\text{rank}(XX') = K_0 < K = \text{rank}(\Delta^{-2})$, which implies $\sqrt{XX'} \neq \Delta^{-1}$ when $K_0 < K$ and the fourth statement does not hold. $\square$

Appendix A.3. Proof of Theorem 1

First, the following two lemmas are provided for our further proof.

Lemma A1. Under $\text{adLCM}(Z, \Theta, \mathcal{F})$, we have

$$\max(\|\hat{U}\hat{O} - U\|_F, \|\hat{V}\hat{O} - V\|_F) \leq \frac{2\sqrt{2K}\|R - R_0\|}{\rho\sigma_K(B)\sqrt{N_{\min}}},$$

where $\hat{O}$ is a K -by- K orthogonal matrix.

Proof. According to the proof of Lemma 3 in [62], there is a $K \times K$ orthogonal matrix $\hat{O}$ such that

$$\max(\|\hat{U}\hat{O} - U\|_F, \|\hat{V}\hat{O} - V\|_F) \leq \frac{\sqrt{2K}\|\hat{R} - R_0\|}{\sqrt{\lambda_K(R_0R_0')}}.$$

Because $\hat{R}$ is the top K SVD of R and $\text{rank}(R_0) = K$, we have $\|R - \hat{R}\| \leq \|R - R_0\|$. Then we have $\|\hat{R} - R_0\| = \|\hat{R} - R + R - R_0\| \leq 2\|R - R_0\|$, which gives

$$\max(\|\hat{U}\hat{O} - U\|_F, \|\hat{V}\hat{O} - V\|_F) \leq \frac{2\sqrt{2K}\|R - R_0\|}{\sqrt{\lambda_K(R_0R_0')}}. \quad (\text{A10})$$

For $\lambda_K(R_0R_0')$, because $R_0 = Z\Theta' = \rho ZB'$ and $\lambda_K(Z'Z) = N_{\min}$, we have

$$\begin{aligned} \lambda_K(R_0R_0') &= \lambda_K(Z\Theta'\Theta Z') = \lambda_K(\rho^2 ZB'BZ') = \rho^2 \lambda_K(B'BZ'Z) \\ &\geq \rho^2 \lambda_K(Z'Z) \lambda_K(B'B) = \rho^2 N_{\min} \lambda_K(B'B). \end{aligned}$$

Combining Equation (A10) with $\lambda_K(R_0R_0') \geq \rho^2 N_{\min} \lambda_K(B'B)$ gives

$$\max(\|\hat{U}\hat{O} - U\|_F, \|\hat{V}\hat{O} - V\|_F) \leq \frac{2\sqrt{2K}\|R - R_0\|}{\rho\sigma_K(B)\sqrt{N_{\min}}}.$$

$\square$

Lemma A2. Under $\text{adLCM}(Z, \Theta, \mathcal{F})$, if Assumption 1 is satisfied, then with a probability of at least $1 - o((N + J)^{-3})$,

$$\|R - R_0\| \leq C\sqrt{\gamma \max(N, J) \log(N + J)},$$

where C is a positive constant.

Proof. This lemma holds by setting α in Lemma 2 [90] as 3, where Lemma 2 of [90] is obtained from the rectangular version of Bernstein inequality in [91]. $\square$

Proof. Now, we prove the first statement of Theorem 1. Set $\varsigma > 0$ as a small value, by Lemma 2 of [61] and the fourth statement of Lemma 1, if

$$\frac{\sqrt{K}}{\varsigma} \|U - \hat{U}\hat{O}\|_F \left(\frac{1}{\sqrt{N_k}} + \frac{1}{\sqrt{N_l}} \right) \leq \sqrt{\frac{1}{N_k} + \frac{1}{N_l}}, \text{ for each } 1 \leq k \neq l \leq K, \quad (\text{A11})$$

and then the Clustering error $\hat{f} = O(\varsigma^2)$ using the K-means algorithm. By setting $\varsigma = \sqrt{\frac{2KN_{\max}}{N_{\min}}} \|U - \hat{U}\hat{O}\|_F$, we see that Equation (A11) always holds for all $1 \leq k \neq l \leq K$. Thus we get $\hat{f} = O(\varsigma^2) = O(\frac{KN_{\max}\|U - \hat{U}\hat{O}\|_F^2}{N_{\min}})$. According to Lemma A1, we have

$$\hat{f} = O\left(\frac{K^2 N_{\max} \|R - R_0\|^2}{\rho^2 \sigma_K^2(B) N_{\min}^2}\right).$$

By Lemma A2, we have

$$\hat{f} = O\left(\frac{\gamma K^2 N_{\max} \max(N, J) \log(N + J)}{\rho^2 \sigma_K^2(B) N_{\min}^2}\right).$$

Next, we prove the second statement of Theorem 1. Since $U = ZX$ by Equation (3) in Lemma 1 and $U'U = I_{K \times K}$, we have $X'Z'ZX = I_{K \times K}$ which gives that $(Z'Z)^{-1} = XX'$ and $\lambda_1(XX') = \sigma_1^2(XX') = \frac{1}{\lambda_K(Z'Z)} = \frac{1}{N_{\min}}$. We also have $Z(Z'Z)^{-1} = ZXX' = UX'$. Similarly, we have $\hat{Z}(\hat{Z}'\hat{Z})^{-1} \approx \hat{U}\hat{X}'$, where $\hat{X}$ is the $K \times K$ centroid matrix returned by the K-means method for $\hat{U}$. Recall that $\hat{R} = \hat{U}\hat{\Sigma}\hat{V}'$, then combine it with Equation (4) and Lemma A2, and we have

$$\begin{aligned} \|\hat{\Theta} - \Theta\mathcal{P}\| &= \|\hat{V}\hat{\Sigma}\hat{U}'\hat{Z}(\hat{Z}'\hat{Z})^{-1} - V\Sigma U'Z(Z'Z)^{-1}\mathcal{P}\| = \|\hat{R}'\hat{Z}(\hat{Z}'\hat{Z})^{-1} - R_0'Z(Z'Z)^{-1}\mathcal{P}\| \\ &= \|(\hat{R}' - R_0')\hat{Z}(\hat{Z}'\hat{Z})^{-1} + R_0'(\hat{Z}(\hat{Z}'\hat{Z})^{-1} - Z(Z'Z)^{-1})\mathcal{P}\| \\ &\leq \|(\hat{R}' - R_0')\hat{Z}(\hat{Z}'\hat{Z})^{-1}\| + \|R_0'(\hat{Z}(\hat{Z}'\hat{Z})^{-1} - Z(Z'Z)^{-1})\mathcal{P}\| \\ &\leq \|\hat{R} - R_0\| \|\hat{Z}(\hat{Z}'\hat{Z})^{-1}\| + \|R_0\| \|\hat{Z}(\hat{Z}'\hat{Z})^{-1} - Z(Z'Z)^{-1}\mathcal{P}\| \\ &\leq 2\|R - R_0\| \|\hat{Z}(\hat{Z}'\hat{Z})^{-1}\| + \|R_0\| \|\hat{Z}(\hat{Z}'\hat{Z})^{-1} - Z(Z'Z)^{-1}\mathcal{P}\| \\ &= 2\|R - R_0\| \|\hat{Z}(\hat{Z}'\hat{Z})^{-1}\| + \|\rho ZB'\| \|\hat{Z}(\hat{Z}'\hat{Z})^{-1} - Z(Z'Z)^{-1}\mathcal{P}\| \\ &\leq 2\|R - R_0\| \|\hat{Z}(\hat{Z}'\hat{Z})^{-1}\| + \rho\|Z\| \|B\| \|\hat{Z}(\hat{Z}'\hat{Z})^{-1} - Z(Z'Z)^{-1}\mathcal{P}\| \\ &= 2\|R - R_0\| \|\hat{Z}(\hat{Z}'\hat{Z})^{-1}\| + \\ &\quad \rho\sigma_1(B)\sqrt{N_{\max}} \|\hat{Z}(\hat{Z}'\hat{Z})^{-1} - Z(Z'Z)^{-1}\mathcal{P}\| \\ &= 2\|R - R_0\| \|\hat{Z}(\hat{Z}'\hat{Z})^{-1} - Z(Z'Z)^{-1}\mathcal{P} + Z(Z'Z)^{-1}\mathcal{P}\| + \\ &\quad \rho\sigma_1(B)\sqrt{N_{\max}} \|\hat{Z}(\hat{Z}'\hat{Z})^{-1} - Z(Z'Z)^{-1}\mathcal{P}\| \\ &\leq 2\|R - R_0\| (\|\hat{Z}(\hat{Z}'\hat{Z})^{-1} - Z(Z'Z)^{-1}\mathcal{P}\| + \|Z(Z'Z)^{-1}\mathcal{P}\|) + \\ &\quad \rho\sigma_1(B)\sqrt{N_{\max}} \|\hat{Z}(\hat{Z}'\hat{Z})^{-1} - Z(Z'Z)^{-1}\mathcal{P}\| \\ &\leq 2\|R - R_0\| (\|\hat{Z}(\hat{Z}'\hat{Z})^{-1} - Z(Z'Z)^{-1}\mathcal{P}\| + \|Z(Z'Z)^{-1}\| \|\mathcal{P}\|) + \\ &\quad \rho\sigma_1(B)\sqrt{N_{\max}} \|\hat{Z}(\hat{Z}'\hat{Z})^{-1} - Z(Z'Z)^{-1}\mathcal{P}\| \\ &= 2\|R - R_0\| (\|\hat{Z}(\hat{Z}'\hat{Z})^{-1} - Z(Z'Z)^{-1}\mathcal{P}\| + \frac{1}{\sqrt{N_{\min}}}) + \\ &\quad \rho\sigma_1(B)\sqrt{N_{\max}} \|\hat{Z}(\hat{Z}'\hat{Z})^{-1} - Z(Z'Z)^{-1}\mathcal{P}\| \\ &= O(\|R - R_0\| (\|\hat{U}\hat{X}' - UX'\mathcal{P}\| + \frac{1}{\sqrt{N_{\min}}})) \\ &\leq O(\|R - R_0\| (\|\hat{U}\hat{X}'\| + \|UX'\mathcal{P}\| + \frac{1}{\sqrt{N_{\min}}})) \\ &\leq O(\|R - R_0\| (\|\hat{U}\| \|\hat{X}\| + \|U\| \|X\| \|\mathcal{P}\| + \frac{1}{\sqrt{N_{\min}}})) \\ &= O(\|R - R_0\| (\|\hat{X}\| + \|X\| + \frac{1}{\sqrt{N_{\min}}})) \\ &= O(\frac{\|R - R_0\|}{\sqrt{N_{\min}}}) \\ &= O(\sqrt{\frac{\gamma \max(N, J) \log(N + J)}{N_{\min}}}) \end{aligned}$$

Since $(\hat{\Theta} - \Theta\mathcal{P})$ is a $J \times K$ matrix and $K \ll J$ in this paper, we have $\text{rank}(\hat{\Theta} - \Theta\mathcal{P}) = K$. Since $\|M\|_F \leq \sqrt{\text{rank}(M)}\|M\|$ holds for any matrix M , we have $\|\hat{\Theta} - \Theta\mathcal{P}\|_F \leq \sqrt{K}\|\hat{\Theta} - \Theta\mathcal{P}\|$. Thus, we have

$$\|\hat{\Theta} - \Theta\mathcal{P}\|_F = O\left(\sqrt{\frac{\gamma K \max(N, J) \log(N + J)}{N_{\min}}}\right). \quad (\text{A12})$$

Combining Equation (A12) with the fact that $\|\Theta\|_F \geq \|\Theta\| = \|\rho B\| = \rho\|B\| = \rho\sigma_1(B) \geq \rho\sigma_K(B)$ gives

$$\frac{\|\hat{\Theta} - \Theta\mathcal{P}\|_F}{\|\Theta\|_F} \leq \frac{\|\hat{\Theta} - \Theta\mathcal{P}\|_F}{\rho\sigma_K(B)} = O\left(\frac{\sqrt{\gamma K \max(N, J) \log(N + J)}}{\rho\sigma_K(B)\sqrt{N_{\min}}}\right).$$

Recall that the J -by- K matrix B satisfies $\max_{j \in [J], k \in [K]} |B(j, k)| = 1$; we have $\sigma_K(B)$ is at least of the order $\sqrt{J} - \sqrt{K - 1}$ with high probability by applying the lower bound of the smallest singular value of a random rectangular matrix in [92]. Since $K \ll J$ in this paper, we have

$$\hat{f} = O\left(\frac{\gamma K^2 N_{\max} \max(N, J) \log(N + J)}{\rho^2 N_{\min}^2 J}\right) \text{ and } \frac{\|\hat{\Theta} - \Theta\mathcal{P}\|_F}{\|\Theta\|_F} = O\left(\frac{\sqrt{\gamma K \max(N, J) \log(N + J)}}{\rho \sqrt{N_{\min} J}}\right).$$

□

References

- Dayton, C.M.; Macready, G.B. Concomitant-variable latent-class models. *J. Am. Stat. Assoc.* **1988**, *83*, 173–178. [\[CrossRef\]](#)
- Hagenaars, J.A.; McCutcheon, A.L. *Applied Latent Class Analysis*; Cambridge University Press: Cambridge, UK, 2002.
- Magidson, J.; Vermunt, J.K. Latent class models. In *The Sage Handbook of Quantitative Methodology for the Social Sciences*; SAGE: Newcastle upon Tyne, UK, 2004; pp. 175–198.
- Guo, G.; Zhang, J.; Thalmann, D.; Yorke-Smith, N. Etaf: An extended trust antecedents framework for trust prediction. In Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014), Beijing, China, 17–20 August 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 540–547.
- Harper, F.M.; Konstan, J.A. The movielens datasets: History and context. *ACM Trans. Interact. Intell. Syst. (TIIS)* **2015**, *5*, 1–19. [\[CrossRef\]](#)
- Meyer, G.J.; Finn, S.E.; Eyde, L.D.; Kay, G.G.; Moreland, K.L.; Dies, R.R.; Eisman, E.J.; Kubiszyn, T.W.; Reed, G.M. Psychological testing and psychological assessment: A review of evidence and issues. *Am. Psychol.* **2001**, *56*, 128. [\[CrossRef\]](#)
- Silverman, J.J.; Galanter, M.; Jackson-Triche, M.; Jacobs, D.G.; Lomax, J.W.; Riba, M.B.; Tong, L.D.; Watkins, K.E.; Fochtmann, L.J.; Rhoads, R.S.; et al. The American Psychiatric Association practice guidelines for the psychiatric evaluation of adults. *Am. J. Psychiatry* **2015**, *172*, 798–802. [\[CrossRef\]](#) [\[PubMed\]](#)
- De La Torre, J.; van der Ark, L.A.; Rossi, G. Analysis of clinical data from a cognitive diagnosis modeling framework. *Meas. Eval. Couns. Dev.* **2018**, *51*, 281–296. [\[CrossRef\]](#)
- Chen, Y.; Li, X.; Zhang, S. Joint maximum likelihood estimation for high-dimensional exploratory item factor analysis. *Psychometrika* **2019**, *84*, 124–146. [\[CrossRef\]](#)
- Shang, Z.; Erosheva, E.A.; Xu, G. Partial-mastery cognitive diagnosis models. *Ann. Appl. Stat.* **2021**, *15*, 1529–1555. [\[CrossRef\]](#)
- Poole, K.T. Nonparametric unfolding of binary choice data. *Political Anal.* **2000**, *8*, 211–237. [\[CrossRef\]](#)
- Clinton, J.; Jackman, S.; Rivers, D. The statistical analysis of roll call data. *Am. Political Sci. Rev.* **2004**, *98*, 355–370. [\[CrossRef\]](#)
- Bakker, R.; Poole, K.T. Bayesian metric multidimensional scaling. *Political Anal.* **2013**, *21*, 125–140. [\[CrossRef\]](#)
- Chen, Y.; Ying, Z.; Zhang, H. Unfolding-model-based visualization: Theory, method and applications. *J. Mach. Learn. Res.* **2021**, *22*, 548–598.
- Martinez-Moya, J.; Feo-Valero, M. Do shippers' characteristics influence port choice criteria? Capturing heterogeneity by using latent class models. *Transp. Policy* **2022**, *116*, 96–105. [\[CrossRef\]](#)
- Formann, A.K.; Kohlmann, T. Latent class analysis in medical research. *Stat. Methods Med. Res.* **1996**, *5*, 179–211. [\[CrossRef\]](#)
- Kongsted, A.; Nielsen, A.M. Latent class analysis in health research. *J. Physiother.* **2017**, *63*, 55–58. [\[CrossRef\]](#)
- Wu, Z.; Deloria-Knoll, M.; Zeger, S.L. Nested partially latent class models for dependent binary data; estimating disease etiology. *Biostatistics* **2017**, *18*, 200–213. [\[CrossRef\]](#)

19. Van der Heijden, P.G.; Dessens, J.; Bockenholt, U. Estimating the concomitant-variable latent-class model with the EM algorithm. *J. Educ. Behav. Stat.* **1996**, *21*, 215–229. [\[CrossRef\]](#)
20. Smaragdis, P.; Raj, B.; Shashanka, M. Shift-invariant probabilistic latent component analysis. *J. Mach. Learn. Res.* **2007**, *5*.
21. Bakk, Z.; Vermunt, J.K. Robustness of stepwise latent class modeling with continuous distal outcomes. *Struct. Equ. Model. Multidiscip. J.* **2016**, *23*, 20–31. [\[CrossRef\]](#)
22. Chen, H.; Han, L.; Lim, A. Beyond the EM algorithm: Constrained optimization methods for latent class model. *Commun. Stat.-Simul. Comput.* **2022**, *51*, 5222–5244. [\[CrossRef\]](#)
23. Gu, Y.; Xu, G. A joint MLE approach to large-scale structured latent attribute analysis. *J. Am. Stat. Assoc.* **2023**, *118*, 746–760. [\[CrossRef\]](#)
24. Shashanka, M.; Raj, B.; Smaragdis, P. Probabilistic latent variable models as nonnegative factorizations. *Comput. Intell. Neurosci.* **2008**, *2008*, 947438. [\[CrossRef\]](#)
25. Anandkumar, A.; Ge, R.; Hsu, D.; Kakade, S.M.; Telgarsky, M. Tensor decompositions for learning latent variable models. *J. Mach. Learn. Res.* **2014**, *15*, 2773–2832.
26. Zeng, Z.; Gu, Y.; Xu, G. A Tensor-EM Method for Large-Scale Latent Class Analysis with Binary Responses. *Psychometrika* **2023**, *88*, 580–612. [\[CrossRef\]](#)
27. Chen, L.; Gu, Y. A spectral method for identifiable grade of membership analysis with binary responses. *Psychometrika* **2024**, *89*, 626–657. [\[CrossRef\]](#)
28. Qing, H. Finding mixed memberships in categorical data. *Inf. Sci.* **2024**, *676*, 120785. [\[CrossRef\]](#)
29. Lyu, Z.; Chen, L.; Gu, Y. Degree-heterogeneous Latent Class Analysis for high-dimensional discrete data. *J. Am. Stat. Assoc.* **2025**, *1–14*. [\[CrossRef\]](#)
30. Formann, A.K. Constrained latent class models: Theory and applications. *Br. J. Math. Stat. Psychol.* **1985**, *38*, 87–111. [\[CrossRef\]](#)
31. Lindsay, B.; Clogg, C.C.; Grego, J. Semiparametric estimation in the Rasch model and related exponential response models, including a simple latent class model for item analysis. *J. Am. Stat. Assoc.* **1991**, *86*, 96–107. [\[CrossRef\]](#)
32. Zhang, N.L. Hierarchical latent class models for cluster analysis. *J. Mach. Learn. Res.* **2004**, *5*, 697–723.
33. Yang, C.C. Evaluating latent class analysis models in qualitative phenotype identification. *Comput. Stat. Data Anal.* **2006**, *50*, 1090–1104. [\[CrossRef\]](#)
34. Xu, G. Identifiability of restricted latent class models with binary responses. *Ann. Stat.* **2017**, *45*, 675–707. [\[CrossRef\]](#)
35. Xu, G.; Shang, Z. Identifying latent structures in restricted latent class models. *J. Am. Stat. Assoc.* **2018**, *113*, 1284–1295. [\[CrossRef\]](#)
36. Ma, W.; Guo, W. Cognitive diagnosis models for multiple strategies. *Br. J. Math. Stat. Psychol.* **2019**, *72*, 370–392. [\[CrossRef\]](#)
37. Gu, Y.; Xu, G. Partial identifiability of restricted latent class models. *Ann. Stat.* **2020**, *48*, 2082–2107. [\[CrossRef\]](#)
38. Newman, M.E. Analysis of weighted networks. *Phys. Rev. E* **2004**, *70*, 056131. [\[CrossRef\]](#)
39. Derr, T.; Johnson, C.; Chang, Y.; Tang, J. Balance in signed bipartite networks. In Proceedings of the 28th ACM International Conference on Information and Knowledge Management, Beijing, China, 3–7 November 2019; pp. 1221–1230.
40. Goldberg, K.; Roeder, T.; Gupta, D.; Perkins, C. Eigentaste: A constant time collaborative filtering algorithm. *Inf. Retr.* **2001**, *4*, 133–151. [\[CrossRef\]](#)
41. Gibson, W.A. Three multivariate models: Factor analysis, latent structure analysis, and latent profile analysis. *Psychometrika* **1959**, *24*, 229–252. [\[CrossRef\]](#)
42. Mislevy, R.J.; Verhelst, N. Modeling item responses when different subjects employ different solution strategies. *Psychometrika* **1990**, *55*, 195–215. [\[CrossRef\]](#)
43. Lubke, G.H.; Muthén, B. Investigating population heterogeneity with factor mixture models. *Psychol. Methods* **2005**, *10*, 21. [\[CrossRef\]](#)
44. McLachlan, G.J.; Do, K.A.; Ambroise, C. *Analyzing Microarray Gene Expression Data*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2005.
45. Lubke, G.; Muthén, B.O. Performance of factor mixture models as a function of model size, covariate effects, and class-specific parameters. *Struct. Equ. Model. Multidiscip. J.* **2007**, *14*, 26–47. [\[CrossRef\]](#)
46. Kim, Y.; Muthén, B.O. Two-part factor mixture modeling: Application to an aggressive behavior measurement instrument. *Struct. Equ. Model.* **2009**, *16*, 602–624. [\[CrossRef\]](#)
47. Tein, J.Y.; Cox, S.; Cham, H. Statistical power to detect the correct number of classes in latent profile analysis. *Struct. Equ. Model. Multidiscip. J.* **2013**, *20*, 640–657. [\[CrossRef\]](#)
48. Goodman, L.A. Exploratory Latent Structure Analysis Using Both Identifiable and Unidentifiable Models. *Biometrika* **1974**, *61*, 215–231. [\[CrossRef\]](#)
49. Agresti, A. *Categorical Data Analysis*; Wiley: New York, NY, USA, 2013.
50. Forcina, A. Identifiability of extended latent class models with individual covariates. *Comput. Stat. Data Anal.* **2008**, *52*, 5263–5268. [\[CrossRef\]](#)

51. Gyllenberg, M.; Koski, T.; Reilink, E.; Verlaan, M. Non-uniqueness in probabilistic numerical identification of bacteria. *J. Appl. Probab.* **1994**, *31*, 542–548. [\[CrossRef\]](#)
52. Woodbury, M.A.; Clive, J.; Garson Jr, A. Mathematical typology: A grade of membership technique for obtaining disease definition. *Comput. Biomed. Res.* **1978**, *11*, 277–298. [\[CrossRef\]](#)
53. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
54. Dey, K.K.; Hsiao, C.J.; Stephens, M. Visualizing the structure of RNA-seq expression data using grade of membership models. *PLoS Genet.* **2017**, *13*, e1006599. [\[CrossRef\]](#)
55. Ke, Z.T.; Wang, M. Using SVD for topic modeling. *J. Am. Stat. Assoc.* **2024**, *119*, 434–449. [\[CrossRef\]](#)
56. Tolley, H.D.; Manton, K.G. Large sample properties of estimates of a discrete grade of membership model. *Ann. Inst. Stat. Math.* **1992**, *44*, 85–95. [\[CrossRef\]](#)
57. Erosheva, E.A.; Fienberg, S.E.; Joutard, C. Describing disability through individual-level mixture models for multivariate binary data. *Ann. Appl. Stat.* **2007**, *1*, 346. [\[CrossRef\]](#)
58. Gormley, I.; Murphy, T. A grade of membership model for rank data. *Bayesian Anal.* **2009**, *4*, 265–295. [\[CrossRef\]](#)
59. Gu, Y.; Erosheva, E.E.; Xu, G.; Dunson, D.B. Dimension-grouped mixed membership models for multivariate categorical data. *J. Mach. Learn. Res.* **2023**, *24*, 1–49.
60. Massa, P.; Salvetti, M.; Tomasoni, D. Bowling alone and trust decline in social network sites. In Proceedings of the 2009 Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing, Chengdu, China, 12–14 December 2009; IEEE: Piscataway, NJ, USA, 2009; pp. 658–663.
61. Joseph, A.; Yu, B. Impact of regularization on spectral clustering. *Ann. Stat.* **2016**, *44*, 1765–1791. [\[CrossRef\]](#)
62. Zhou, Z.; Amini, A.A. Analysis of spectral clustering algorithms for community detection: the general bipartite setting. *J. Mach. Learn. Res.* **2019**, *20*, 1–47.
63. Zhang, H.; Guo, X.; Chang, X. Randomized spectral clustering in large-scale stochastic block models. *J. Comput. Graph. Stat.* **2022**, *31*, 887–906. [\[CrossRef\]](#)
64. Mao, X.; Sarkar, P.; Chakrabarti, D. Estimating mixed memberships with sharp eigenvector deviations. *J. Am. Stat. Assoc.* **2021**, *116*, 1928–1940. [\[CrossRef\]](#)
65. Qing, H.; Wang, J. Regularized spectral clustering under the mixed membership stochastic block model. *Neurocomputing* **2023**, *550*, 126490. [\[CrossRef\]](#)
66. Chen, Y.; Chi, Y.; Fan, J.; Ma, C. Spectral methods for data science: A statistical perspective. *Found. Trends Mach. Learn.* **2021**, *14*, 566–806. [\[CrossRef\]](#)
67. Lei, J.; Rinaldo, A. Consistency of spectral clustering in stochastic block models. *Ann. Stat.* **2015**, *43*, 215–237. [\[CrossRef\]](#)
68. Paul, S.; Chen, Y. Spectral and matrix factorization methods for consistent community detection in multi-layer networks. *Ann. Stat.* **2020**, *48*, 230–250. [\[CrossRef\]](#)
69. Lei, J.; Lin, K.Z. Bias-adjusted spectral clustering in multi-layer stochastic block models. *J. Am. Stat. Assoc.* **2023**, *118*, 2433–2445. [\[CrossRef\]](#)
70. Qing, H. Community detection in multi-layer networks by regularized debiased spectral clustering. *Eng. Appl. Artif. Intell.* **2025**, *152*, 110627. [\[CrossRef\]](#)
71. Jin, J.; Ke, Z.T.; Luo, S. Mixed membership estimation for social networks. *J. Econom.* **2024**, *239*, 105369. [\[CrossRef\]](#)
72. Qing, H. Discovering overlapping communities in multi-layer directed networks. *Chaos Solitons Fractals* **2025**, *194*, 116175. [\[CrossRef\]](#)
73. Qing, H. A useful criterion on studying consistent estimation in community detection. *Entropy* **2022**, *24*, 1098. [\[CrossRef\]](#)
74. Jin, J. Fast community detection by SCORE. *Ann. Stat.* **2015**, *43*, 57–89. [\[CrossRef\]](#)
75. Strehl, A.; Ghosh, J. Cluster ensembles—A knowledge reuse framework for combining multiple partitions. *J. Mach. Learn. Res.* **2002**, *3*, 583–617.
76. Danon, L.; Diaz-Guilera, A.; Duch, J.; Arenas, A. Comparing community structure identification. *J. Stat. Mech. Theory Exp.* **2005**, *2005*, P09008. [\[CrossRef\]](#)
77. Bagrow, J.P. Evaluating local community methods in networks. *J. Stat. Mech. Theory Exp.* **2008**, *2008*, P05001. [\[CrossRef\]](#)
78. Luo, W.; Yan, Z.; Bu, C.; Zhang, D. Community detection by fuzzy relations. *IEEE Trans. Emerg. Top. Comput.* **2017**, *8*, 478–492. [\[CrossRef\]](#)
79. Hubert, L.; Arabie, P. Comparing partitions. *J. Classif.* **1985**, *2*, 193–218. [\[CrossRef\]](#)
80. Vinh, N.X.; Epps, J.; Bailey, J. Information theoretic measures for clusterings comparison: Is a correction for chance necessary? In Proceedings of the 26th Annual International Conference on Machine Learning, Montreal, QC, Canada, 14–18 June 2009; pp. 1073–1080.
81. Rohe, K.; Qin, T.; Yu, B. Co-clustering directed graphs to discover asymmetries and directional communities. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 12679–12684. [\[CrossRef\]](#)

82. Erosheva, E.A. Comparing latent structures of the grade of membership, Rasch, and latent class models. *Psychometrika* **2005**, *70*, 619–628. [[CrossRef](#)]
83. Huang, G.H.; Bandeen-Roche, K. Building an identifiable latent class model with covariate effects on underlying and measured variables. *Psychometrika* **2004**, *69*, 5–32. [[CrossRef](#)]
84. Reboussin, B.A.; Ip, E.H.; Wolfson, M. Locally dependent latent class models with covariates: An application to under-age drinking in the USA. *J. R. Stat. Soc. Ser. A Stat. Soc.* **2008**, *171*, 877–897. [[CrossRef](#)]
85. Vermunt, J.K. Latent class modeling with covariates: Two improved three-step approaches. *Political Anal.* **2010**, *18*, 450–469. [[CrossRef](#)]
86. Di Mari, R.; Bakk, Z.; Punzo, A. A random-covariate approach for distal outcome prediction with latent class analysis. *Struct. Equ. Model. Multidiscip. J.* **2020**, *27*, 351–368. [[CrossRef](#)]
87. Bakk, Z.; Di Mari, R.; Oser, J.; Kuha, J. Two-stage multilevel latent class analysis with covariates in the presence of direct effects. *Struct. Equ. Model. Multidiscip. J.* **2022**, *29*, 267–277. [[CrossRef](#)]
88. Asparouhov, T.; Hamaker, E.L.; Muthén, B. Dynamic latent class analysis. *Struct. Equ. Model. Multidiscip. J.* **2017**, *24*, 257–269. [[CrossRef](#)]
89. Wu, S.; Li, Z.; Zhu, X. A Distributed Community Detection Algorithm for Large Scale Networks Under Stochastic Block Models. *Comput. Stat. Data Anal.* **2023**, *187*, 107794. [[CrossRef](#)]
90. Qing, H.; Wang, J. Community detection for weighted bipartite networks. *Knowl.-Based Syst.* **2023**, *274*, 110643. [[CrossRef](#)]
91. Tropp, J.A. User-Friendly Tail Bounds for Sums of Random Matrices. *Found. Comput. Math.* **2012**, *12*, 389–434. [[CrossRef](#)]
92. Rudelson, M.; Vershynin, R. Smallest singular value of a random rectangular matrix. *Commun. Pure Appl. Math. J. Issued Courant Inst. Math. Sci.* **2009**, *62*, 1707–1739. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.