Computational Statistics and Data Analysis 193 (2024) 107915

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Generalized latent space model for one-mode networks with awareness of two-mode networks

Xinyan Fan [a,1], Kuangnan Fang [b,1], Dan Pu [c,*,1], Ruixuan Qin [b,1]

[a] Center of Applied Statistics and School of Statistics, Renmin University of China, Beijing, China
[b] School of Economics, Xiamen University, Xiamen, China
[c] School of Statistics, Southwestern University of Finance and Economics, Chengdu, China

A R T I C L E I N F O

Keywords:
Latent space model
Multiplex network
One-mode network

A B S T R A C T

Latent space models have been widely studied for one-mode networks, in which the same type of nodes connect with each other. In many applications, one-mode networks are often observed along with two-mode networks, which reflect connections between different types of nodes and provide important information for understanding the one-mode network structure. However, the classical one-mode latent space models have several limitations in incorporating two-mode networks. To address this gap, a generalized latent space model is proposed to capture common structures and heterogeneous connecting patterns across one-mode and two-mode networks. Specifically, each node is embedded with a latent vector and network-specific degree parameters that determine the connection probabilities between nodes. A projected gradient descent algorithm is developed to estimate the latent vectors and degree parameters. Moreover, the theoretical properties of the estimators are established and it has been proven that the estimation accuracy of the shared latent vectors can be improved through incorporating two-mode networks. Finally, simulation studies and applications on two real-world datasets demonstrate the usefulness of the proposed model.

1. Introduction

With the increasing availability of network data, interest in the analysis of network data has increased in a wide variety of fields, including social networks (see, e.g., Lewis et al., 2012; Kossinets and Watts, 2006), biological networks (see, e.g., Bader and Hogue, 2003; Nepusz et al., 2012), and financial networks (see, e.g., Billio et al., 2012; Härdle et al., 2016). In general, network data can be represented as graphs $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where the node set $\mathcal{V}$ contains all entities and each element in the edge set $\mathcal{E}$ indicates the interaction between two entities. For example, in a brain connectivity network, nodes correspond to different brain regions, while edges reflect connectivity patterns between pairs of regions (Zhang et al., 2020a). In a social network, the node set $\mathcal{V}$ consists of a group of individuals, and the edge set $\mathcal{E}$ describes the relationship between individuals. Such a network is known as a one-mode network, since $\mathcal{V}$ contains only nodes of the same type.

* Corresponding author.
E-mail address: pudan666@163.com (D. Pu).
[1] The authors are listed in the alphabetic order.

https://doi.org/10.1016/j.csda.2023.107915
Received 8 June 2023; Received in revised form 26 December 2023; Accepted 27 December 2023

Available online 10 January 2024

To better understand complex interaction patterns in the one-mode network and fully utilize the information in the network data, network generative mechanisms have been extensively studied in recent decades. One of the most popular and important models is the latent space model, which was proposed by Hoff et al. (2002). In the latent space model, each node is embedded into a low-dimensional latent space, which flexibly characterizes several common properties of real-world networks, such as transitivity, homophily and community structures. For example, if two nodes share many common neighbors, these two nodes are more likely to connect. Such a phenomenon is called transitivity in the social network literature. This is reasonable since their latent variables are more likely to be close to each other in the latent space. For community structures, refer to Zhang et al. (2022a) and Ren et al. (2023). As a result, the latent space model and its extensions have been widely studied for one-mode networks (see, e.g., Gollini and Murphy, 2016; Salter-Townshend and McCormick, 2017; MacDonald et al., 2022; Sewell and Chen, 2015, 2017; Zhen and Wang, 2023).

Despite the developments achieved so far, the collection of additional network information provides new opportunities and challenges for the research of one-mode networks. Below are some motivating examples. As one of the most famous third-party online platforms, Yelp provides not only a friendship network among users, but also a review network among users and businesses. For the review network, two different types of nodes (i.e., users and businesses) are contained and a user-business link describes whether one user writes a review for one business. Such type of network is called a two-mode network or a bipartite network, if the node set $\mathcal{V}$ contains two different types of nodes and the edge set $\mathcal{E}$ contains connections between these different types of nodes (see, e.g., Feng et al., 2019; Huang et al., 2020a). In addition to third-party online platforms, we can easily collect the collaboration network among authors and the attendance network among authors and conferences from DBLP (Digital Bibliography & Library Project), which is a computer science bibliography website. Different authors are linked if they collaborate to write papers, and authors and conferences are linked if authors attend conferences or present papers at conferences. Specifically, such type of network data can be represented as $\mathcal{G} = (\{\mathcal{V}_1, \mathcal{V}_2\}, \{\mathcal{E}_1, \mathcal{E}_{12}\})$, where $\mathcal{V}_1$ and $\mathcal{V}_2$ are two different types of nodes, $\mathcal{E}_1$ denotes the edge set of the one-mode network between different nodes in $\mathcal{V}_1$, and $\mathcal{E}_{12}$ denotes the edge set of the two-mode network between different nodes in $\mathcal{V}_1$ and $\mathcal{V}_2$.

For the above type of network data, two types of networks are collected. In addition to the one-mode network, the two-mode network contains abundant information about the behaviors of nodes in $\mathcal{V}_1$. Therefore, incorporating the two-mode network can help us detect communities for nodes in $\mathcal{V}_1$ and predict missing links for the one-mode network. For instance, in a bibliographical network, including author-conference links helps us better identify author communities (Zhang and Chen, 2020), and in third-party online platforms, user-business links reflect individual consumption habits and may provide important information for predicting friendship links. That is, jointly modeling one-mode and two-mode networks can leverage homogeneous structures across different networks and help us better understand interested nodes or networks. When jointly analyzing multiple networks, one naive approach is to regard the two different types of nodes as nodes of the same type and treat multiple networks as an extended one-mode network. However, this approach might fail since the specific information contained in each network is ignored. For example, specific activity levels of individual nodes may vary across different networks. Fig. 1(c) shows histograms of node heterogeneity parameters for the same set of users in the two networks and demonstrates the heterogeneous connectivity patterns in different networks. More importantly, the above modeling framework directly neglects possible links among nodes in $\mathcal{V}_2$, which might lead to biased estimation and invalid inference. Therefore, it is crucial to develop a statistical model for one-mode and two-mode networks that captures homogeneous and heterogeneous structures across different networks.

The aforementioned network data are closely related to heterogeneous networks, which contain different types of nodes and different types of edges (see, e.g., Sengupta and Chen, 2015; Zhang and Chen, 2020; Huang et al., 2020b). Heterogeneous networks assume that any pairs of nodes of any type can interact with one another; that is, for two types of nodes $\mathcal{V}_1$ and $\mathcal{V}_2$, links exist within node set $\mathcal{V}_1$, within node set $\mathcal{V}_2$, and between node sets $\mathcal{V}_1$ and $\mathcal{V}_2$. In contrast to heterogeneous networks, our setting is commonplace in some practical applications. For third-party online platforms, we can easily collect the friendship network among users and the review network among users and businesses. However, it is unclear how to define the network among businesses. More importantly, the above methods focus on detecting communities for heterogeneous networks, while our method focuses on network embedding. In addition, several models have been proposed in the literature for multilevel networks, which categorize nodes into different types, with each type representing a distinct level. Accordingly, one-mode networks are defined for each type of unit, and two-mode networks are defined between adjacent types of units. Some multilevel network models include the exponential random graph model (see, e.g., Wang et al., 2013; Slaughter and Koehly, 2016) and the stochastic block model (see, e.g., Barbillon et al., 2017; Chabert-Liddell et al., 2021). However, existing methods cannot capture more complex structures and have large computational costs when considering very large networks. In addition, most existing models lack theoretical results, and the effects of modeling multiple types of networks simultaneously are rarely studied. In addition, to the best of our knowledge, latent space models have not yet been used to jointly model one-mode and two-mode networks.

In this work, we develop a latent space model for simultaneously analyzing one-mode and two-mode networks. Specifically, we assume that the latent vectors for the same type of nodes are shared in different networks, thereby providing a flexible method for modeling dependencies and sharing information between one-mode and two-mode networks. By introducing network-specific node-degree heterogeneity parameters for the common nodes, the proposed model can accommodate distinct connectivity patterns for the same nodes in different networks. Furthermore, by following Ma et al. (2020) and treating the latent vectors as fixed effects, we develop an efficient projected gradient descent algorithm for estimating the degree heterogeneity parameters and latent vectors. When compared to most existing methods that adopt Bayesian estimation for model fitting (see, e.g., Hoff, 2003; Krivitsky et al., 2009; D'Angelo et al., 2019), the proposed algorithm is more computationally efficient and more scalable for large networks. In addition, we provide the high probability upper error bound of the proposed estimators. Moreover, compared with the results of a model that utilizes the information of only the one-mode network, we theoretically prove that the estimation of the shared latent vectors can be improved by jointly modeling multiple types of networks under some mild constraints. This result provides insights into why and when simultaneously modeling multiple networks is beneficial and useful. In addition, extensive simulation studies demonstrate that incorporating information from the two-mode network improves the estimation of the shared latent vectors, and provides a better understanding of the one-mode network. This also suggests that missing links of the one-mode network can be predicted more accurately with the aid of the two-mode network.

The remainder of this paper is organized as follows. In Section 2, the generalized latent space model and its estimator are introduced. The theoretical properties are discussed in Section 3. In Sections 4 and 5, numerical studies and two real data examples are presented, respectively. Finally, the paper is concluded with a brief discussion in Section 6.

2. Models and methodology

2.1. Model and notations

Consider a network $\mathcal{G}=(\{\mathcal{V}_1,\mathcal{V}_2\},\{\mathcal{E}_1,\mathcal{E}_{12}\})$, where $\mathcal{V}_1=\{v_1^{(1)},\cdots,v_{n_1}^{(1)}\}$ and $\mathcal{V}_2=\{v_1^{(2)},\cdots,v_{n_2}^{(2)}\}$ represent two types of node sets, $\mathcal{E}_1\subset\mathcal{V}_1\times\mathcal{V}_1$ represents the edge set between nodes in $\mathcal{V}_1$, and $\mathcal{E}_{12}\subset\mathcal{V}_1\times\mathcal{V}_2$ represents the edge set between nodes in $\mathcal{V}_1$ and $\mathcal{V}_2$. To describe the network structure among nodes in $\mathcal{V}_1$, the following binary adjacency matrix $A=(a_{ij})\in\mathbb{R}^{n_1\times n_1}$ is defined: $a_{ij}=a_{ji}=1$ if $(v_i^{(1)},v_j^{(1)})\in\mathcal{E}_1$; otherwise, $a_{ij}=0$. The adjacency matrix between nodes in $\mathcal{V}_1$ and $\mathcal{V}_2$ is defined as $D=(d_{ij})\in\mathbb{R}^{n_1\times n_2}$, where $d_{ij}=1$ if $(v_i^{(1)},v_j^{(2)})\in\mathcal{E}_{12}$ and $d_{ij}=0$ otherwise.

To model network data A and D simultaneously, we propose a generalized latent space model. Specifically, we assume that nodes $v_i^{(1)}\in\mathcal{V}_1$ and $v_j^{(2)}\in\mathcal{V}_2$ can be represented by the low-dimensional vectors $Z_i\in\mathbb{R}^k$ and $W_j\in\mathbb{R}^k$, respectively. Given the node latent vectors and the degree heterogeneity parameters, different a_{ij}s and d_{ij}s are conditionally independent Bernoulli random variables, with probabilities satisfying

$$\begin{aligned}\mathrm{logit}P_{ij}^A&=\Theta_{ij}^A=\alpha_i+\alpha_j+Z_i^\top Z_j, v_i^{(1)},v_j^{(1)}\in\mathcal{V}_1,\\ \mathrm{logit}P_{ij}^D&=\Theta_{ij}^D=\beta_i+\gamma_j+Z_i^\top W_j, v_i^{(1)}\in\mathcal{V}_1,v_j^{(2)}\in\mathcal{V}_2,\end{aligned}\tag{1}$$

where $\mathrm{logit}(x)=\ln(x/(1-x))$, $\Theta^A=(\Theta_{ij}^A)\in\mathbb{R}^{n_1\times n_1}$, $\Theta^D=(\Theta_{ij}^D)\in\mathbb{R}^{n_1\times n_2}$, $P^A=(P_{ij}^A)\in\mathbb{R}^{n_1\times n_1}$ and $P^D=(P_{ij}^D)\in\mathbb{R}^{n_1\times n_2}$ are connectivity probability matrices, and α_i,β_i and γ_i are the degree heterogeneity parameters of networks A and D. Specifically, when all other parameters are fixed, if α_i increases, node $v_i^{(1)}$ is more likely to connect with other nodes, and the degree of node $v_i^{(1)}$ increases. The same interpretation can be applied to parameters β_i and γ_i. The α_i and β_i values of node $v_i^{(1)}$ are not required to be equal, and nodes in $\mathcal{V}_1$ are allowed to have different degrees of heterogeneity in various networks. In addition, to capture common structures in different networks, we assume that the latent vectors of nodes in $\mathcal{V}_1$ are shared by networks A and D. In summary, the proposed model utilizes different degree heterogeneity parameters to characterize specific connection behaviors among various networks and uses shared latent vectors to capture common information across different networks. Covariates could be introduced into this model; however, for simplicity, we consider a model with no covariates in this paper.

Define $Z=(Z_1,\cdots,Z_{n_1})^\top\in\mathbb{R}^{n_1\times k}$, $W=(W_1,\cdots,W_{n_2})^\top\in\mathbb{R}^{n_2\times k}$, $\alpha=(\alpha_1,\cdots,\alpha_{n_1})^\top\in\mathbb{R}^{n_1}$, $\beta=(\beta_1,\cdots,\beta_{n_1})^\top\in\mathbb{R}^{n_1}$ and $\gamma=(\gamma_1,\cdots,\gamma_{n_2})^\top\in\mathbb{R}^{n_2}$. Let $1_n\in\mathbb{R}^n$ be the all-one vector. The above model can be expressed in a matrix form as follows:

$$\begin{aligned}\Theta^A&=\alpha 1_{n_1}^\top+1_{n_1}\alpha^\top+ZZ^\top,\\ \Theta^D&=\beta 1_{n_2}^\top+1_{n_1}\gamma^\top+ZW^\top.\end{aligned}$$

In the above model, one can easily verify that α,β,γ,Z and W are not jointly identifiable. To demonstrate this claim, let c_1 and c_2 be any arbitrary nonzero constants. Define $Z'=Z+c_1 1_{n_1}1_k^\top$, $W'=W+c_2 1_{n_2}1_k^\top$, $\alpha'=\alpha-c_1 Z1_k-\frac{1}{2}kc_1^2 1_{n_1}$, $\beta'=\beta-c_2 Z1_k-\frac{1}{2}c_1c_2k1_{n_1}$ and $\gamma'=\gamma-c_1 W1_k-\frac{1}{2}c_1c_2k1_{n_2}$. Here, model (1) remains valid. To address this problem, we assume that the latent variables are centered, that is,

$$J_1Z=Z, J_2W=W,$$

where $J_1=I_{n_1}-\frac{1}{n_1}1_{n_1}1_{n_1}^\top$ and $J_2=I_{n_2}-\frac{1}{n_2}1_{n_2}1_{n_2}^\top$. Moreover, to ensure the uniqueness of β and γ, we assume $\beta^\top 1_{n_1}=0$. As a result, node degree heterogeneity parameters α, β and γ are uniquely determined, and latent vectors Z and W are uniquely identifiable up to a common orthogonal transformation of their rows. The detailed results and proofs can be found in Proposition 1 in the Appendix.

Remark 1. The proposed model can be easily extended to incorporate edges between nodes in $\mathcal{V}_2$. Denote the adjacency matrix between nodes $\mathcal{V}_2$ as $G=(g_{ij})\in\mathbb{R}^{n_2\times n_2}$. To model network data A, D and G simultaneously, the following model can be considered,

$$\begin{aligned}\mathrm{logit}P_{ij}^A&=\Theta_{ij}^A=\alpha_i+\alpha_j+Z_i^\top Z_j, v_i^{(1)},v_j^{(1)}\in\mathcal{V}_1,\\ \mathrm{logit}P_{ij}^D&=\Theta_{ij}^D=\beta_i+\gamma_j+Z_i^\top W_j, v_i^{(1)}\in\mathcal{V}_1,v_j^{(2)}\in\mathcal{V}_2,\\ \mathrm{logit}P_{ij}^G&=\Theta_{ij}^D=\delta_i+\delta_j+W_i^\top W_j, v_i^{(2)},v_j^{(2)}\in\mathcal{V}_2.\end{aligned}$$

The estimation method and optimization algorithm described below are also applicable to this general model. Motivated by practical applications, we only consider networks A and D for simplicity.

Remark 2. In practice, someone may simply regard the two different types of nodes as nodes of the same type and apply the one-mode latent space model to fit pseudo one-mode networks. Compared with the proposed model, the above approach neglects distinct connectivity patterns of node set $\mathcal{V}_1$ in different networks. In addition, the above approach directly assumes that there are no links among node set $\mathcal{V}_2$. This assumption is too naive and may lead to biased estimation and prediction. In contrast, the proposed model introduces network-specific node-degree heterogeneity parameters for node set $\mathcal{V}_1$ and makes no additional assumption about the links among node set $\mathcal{V}_2$.

Remark 3. It is worth noting that Zhang et al. (2022b) simultaneously analyzed one-mode network and node variables via shared latent variables. In addition to shared latent vectors, the proposed model assumes the two-mode network is related to node degree heterogeneity parameters β and γ. By introducing network-specific node degree heterogeneity parameters, the proposed model can capture the heterogeneous information across different networks. However, this leads to an identification problem between parameters β and γ. In addition, it is urgent to determine initial estimates for the two-mode network and investigate the properties of initial estimates. Lastly, whether the inclusion of parameters β and γ will affect the estimate of Z is questionable. This poses a great challenges to study the conditions under which the shared latent vectors can be improved.

2.2. Parameter estimation

For model (1), we employ the maximum likelihood estimator (MLE) to estimate the unknown parameters α, β, γ, Z and W. Let $\sigma(x) = 1/(1+\exp(-x))$ be the sigmoid function. For network A, the conditional negative log-likelihood function is

$$L_A = -\sum_{i=1}^{n_1}\sum_{j=1}^{n_1}\left\{A_{ij}\Theta_{ij}^A + \ln(1-\sigma(\Theta_{ij}^A))\right\}. \tag{2}$$

For network D, the negative log-likelihood function is

$$L_D = -\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\left\{D_{ij}\Theta_{ij}^D + \ln(1-\sigma(\Theta_{ij}^D))\right\}. \tag{3}$$

Define the objective function as $L(\alpha,\beta,\gamma,Z,W) = L_A + L_D$. Then, the MLE of α,β,γ,Z and W is the solution to the following constrained optimization problem:

$$\min_{\alpha,\beta,\gamma,Z,W,J_1Z=Z,J_2W=W,\beta^\top 1_{n_1}=0} L(\alpha,\beta,\gamma,Z,W). \tag{4}$$

To facilitate the above implementation and simplify the theoretical results, we include the summation of all diagonal elements in Θ^A into the objective function (2). According to Ma et al. (2020), this slight modification has limited effects in theory and in practical applications. In contrast to existing work on latent space models (see, e.g., Hoff, 2003; Chang et al., 2019), we treat the degree heterogeneity parameters and node latent vectors as fixed effects rather than random effects. Accordingly, the gradient descent algorithm can be applied to estimate the unknown parameters, and this computation is efficient and scalable (see, e.g., Zhang et al., 2020b, 2022b). Similar to Ma et al. (2020), we adopt the projected gradient descent algorithm to optimize the constrained problem (4). A detailed description of the method is presented in Algorithm 1. Details of these steps are provided in Appendix E.

As illustrated in Algorithm 1, in each iteration, we first update each parameter along the gradient direction with a prespecified step size and project the updated estimates to satisfy the identification conditions. Although Algorithm 1 is not guaranteed to find the global solution, it quickly converges to the true parameters and provides good estimation results in our simulation studies. However, Algorithm 1 requires initial inputs. We follow Ma et al. (2020) and optimize a regularized version of objective function (2) to determine the initial values Z^0 and α^0. Given Z^0, we optimize a regularized version of objective function (3) to determine initial estimates for β^0, γ^0 and W^0. The step size decays exponentially with the number of iterations, and the initial step sizes are set to $\eta_1^0 = \eta_2^0 = \eta/(2n_1)$, $\eta_3^0 = \eta/(2n_2)$, $\eta_4^0 = \eta/\|Z^0\|_2^2$ and $\eta_5^0 = \eta/\|W^0\|_2^2$ for some small constant $\eta > 0$. For the latent space dimension, we utilize a similar method to that used in Zhang et al. (2022b) and select the latent space dimension k through cross validation. In Algorithm 1, the first-order gradient information of parameters is required in each iteration. Therefore, updating parameters in each iteration requires $O\{\max(n_1^2k, n_2^2k)\}$ operations and the computational complexity of Algorithm 1 is $O\{\max(n_1^2k, n_2^2k)\}$. Our simulation studies demonstrate that this algorithm is computationally efficient and scalable for large-scale networks.

3. Theoretical results

In this section, we present the main theoretical results for the MLE of the model (1). We first provide an upper bound for the estimation error of $\Theta = \left[\Theta^A, \Theta^D\right] \in \mathbb{R}^{n_1\times(n_1+n_2)}$ and then discuss how jointly modeling network data A and D improves the estimation of the shared latent vectors Z. Before establishing the theoretical properties of the MLE, we introduce the following feasible parameter space:

Algorithm 1: Projected gradient descent algorithm for parameter estimation.

Input : adjacency matrix $A \in \mathbb{R}^{n_1 \times n_1}$; adjacency matrix $D \in \mathbb{R}^{n_1 \times n_2}$; latent space dimension k; initial estimates: $\alpha^0, \beta^0, \gamma^0, Z^0, W^0$; initial step sizes: $\eta_1^0, \eta_2^0, \eta_3^0, \eta_4^0, \eta_5^0$; number of iterations T; hyperparameter $\rho = 0.9^{1/20}$.
Output: $\hat{\alpha} = \alpha^T, \hat{\beta} = \beta^T, \hat{\gamma} = \gamma^T, \hat{Z} = Z^T, \hat{W} = W^T$.

for $t = 0, 1, \cdots, T-1$ **do**
 $\Theta^{At} = \alpha^t 1_{n_1}^\top + 1_{n_1} \alpha^{t\top} + Z^t Z^{t\top}$;
 $\Theta^{Dt} = \beta^t 1_{n_2}^\top + 1_{n_1} \gamma^{t\top} + Z^t W^{t\top}$;
 $\alpha^{t+1} = \alpha^t - \eta_1^t \nabla_\alpha L = \alpha^t + 2\eta_1^t \left(A - \sigma(\Theta^{At})\right) 1_{n_1}$;
 $\beta^{t+1} = \beta^t - \eta_2^t \nabla_\beta L = \beta^t + \eta_2^t \left(D - \sigma(\Theta^{Dt})\right) 1_{n_2}$;
 $\gamma^{t+1} = \gamma^t - \eta_3^t \nabla_\gamma L = \gamma^t + \eta_3^t \left(D - \sigma(\Theta^{Dt})\right)^\top 1_{n_1}$;
 $Z^{t+1} = Z^t - \eta_4^t \nabla_Z L = Z^t + 2\eta_4^t \left(A - \sigma(\Theta^{At})\right) Z^t + \eta_4^t \left(D - \sigma(\Theta^{Dt})\right) W^t$;
 $W^{t+1} = W^t - \eta_5^t \nabla_W L = W^t + \eta_5^t \left(D - \sigma(\Theta^{Dt})\right)^\top Z^t$;
 $\eta_1^{t+1} = \rho\eta_1^t, \eta_2^{t+1} = \rho\eta_2^t, \eta_3^{t+1} = \rho\eta_3^t, \eta_4^{t+1} = \rho\eta_4^t, \eta_5^{t+1} = \rho\eta_5^t$;
 $\omega^{t+1} = \beta^{t+1\top} 1_{n_1}, \beta^{t+1} = \beta^{t+1} - \omega^{t+1} 1_{n_1}, \gamma^{t+1} = \gamma^{t+1} + \omega^{t+1} 1_{n_2}$;
 $Z^{t+1} = J_1 Z^{t+1}, W^{t+1} = J_2 W^{t+1}$;
end

$$\begin{aligned}\mathcal{F} = \Big\{\Theta = \left[\Theta^A, \Theta^D\right] \in \mathbb{R}^{n_1 \times (n_1+n_2)} :& \\ &\Theta^A = \alpha 1_{n_1}^\top + 1_{n_1} \alpha^\top + ZZ^\top, \Theta^D = \beta 1_{n_2}^\top + 1_{n_1} \gamma^\top + ZW^\top, \\ &J_1 Z = Z, J_2 W = W, \beta^\top 1_{n_1} = 0, \|\Theta^A\|_{\max} \le M_A, \|\Theta^D\|_{\max} \le M_D \Big\},\end{aligned}$$

where $\|\cdot\|_{\max}$ represents the maximum absolute value of the entries in a matrix. Denote the true parameter as

$$\Theta^* = \left[\Theta^{*A}, \Theta^{*D}\right] = \left[\alpha^* 1_{n_1}^\top + 1_{n_1} \alpha^{*\top} + Z^* Z^{*\top}, \beta^* 1_{n_2}^\top + 1_{n_1} \gamma^{*\top} + Z^* W^{*\top}\right].$$

Denote the estimator as

$$\hat{\Theta} = \left[\hat{\Theta}^A, \hat{\Theta}^D\right] = \left[\hat{\alpha} 1_{n_1}^\top + 1_{n_1} \hat{\alpha}^\top + \hat{Z}\hat{Z}^\top, \hat{\beta} 1_{n_2}^\top + 1_{n_1} \hat{\gamma}^\top + \hat{Z}\hat{W}^\top\right],$$

where $\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{Z}$ and $\hat{W}$ are the solutions of optimization problem (4). To simplify the theoretical investigation, we assume that the true parameter and estimator lie in parameter space $\mathcal{F}$. Next, we provide the error bound for estimator $\hat{\Theta}$ in the following theorem.

Theorem 1. *Given $\Theta^*, \hat{\Theta} \in \mathcal{F}$, there exist positive constants c and C such that with a probability of at least $1 - e^{-c(2n_1+n_2)}$, we have*

$$\|\hat{\Theta} - \Theta^*\|_F \le \frac{C\sqrt{k+2}\sqrt{2n_1+n_2}}{\min_{|v|<\max\{M_A, M_D\}} b''(v)},$$

where $b(x) = \ln(1 + \exp(x))$.

The proof of Theorem 1 is provided in Appendix B. According to Theorem 1, when the information of network D is not included, i.e., when $n_2 = 0$, we have $\|\hat{\Theta}^A - \Theta^{*A}\|_F / n_1 = O_p(e^{M_A}(k/n_1)^{1/2})$. This conclusion is consistent with the results of Ma et al. (2020). Moreover, Theorem 1 indicates that $\|\hat{\Theta} - \Theta^*\|_F / [n_1(n_1+n_2)]^{1/2} = O_p(1/n_1^{1/2})$ if M_A, M_D and k are fixed. This conclusion implies that the estimation error of the joint modeling framework has the same order as the estimation error of a model that uses only the information of network A.

Compared with Ma et al. (2020), we further discuss how the estimation of the shared latent vectors Z can be improved by considering the additional information in network D. Following Zhang et al. (2022b), we analyze a one-step update to evaluate the effect of network D on the estimation of Z. Suppose we have obtained reasonable estimates $\widetilde{\alpha}, \widetilde{\beta}, \widetilde{\gamma}, \widetilde{Z}$ and $\widetilde{W}$. Based on Algorithm 1, we update $\widetilde{Z}$ in one step as follows:

$$\hat{Z} = \widetilde{Z} + 2\eta_4 \left(A - \sigma(\widetilde{\Theta}^A)\right) \widetilde{Z} + \eta_4 \left(D - \sigma(\widetilde{\Theta}^D)\right) \widetilde{W},$$

where $\widetilde{\Theta}^A = \widetilde{\alpha} 1_{n_1}^\top + 1_{n_1} \widetilde{\alpha}^\top + \widetilde{Z}\widetilde{Z}^\top$ and $\widetilde{\Theta}^D = \widetilde{\beta} 1_{n_2}^\top + 1_{n_1} \widetilde{\gamma}^\top + \widetilde{Z}\widetilde{W}^\top$. Before establishing the properties of $\hat{Z}$, we state some regular conditions.

(C1) Denote the spectral decomposition of $Z^{*\top} Z^* / n_1$ as $H_1 \Lambda_1 H_1^\top$, where $\Lambda_1 = \mathrm{diag}(\mu_1, \cdots, \mu_k)$. Assume that there exist constants τ_1 and τ_2 such that $\tau_1 \le \mu_k \le \mu_1 \le \tau_2$.
(C2) Decompose $W^{*\top} W^* / n_2$ as $H_2 \Lambda_2 H_2^\top$, where $\Lambda_2 = \mathrm{diag}(\nu_1, \cdots, \nu_k)$. Assume that there exist constants τ_3 and τ_4 such that $\tau_3 \le \nu_k \le \nu_1 \le \tau_4$.

Conditions (C1) and (C2) place requirements on the signals of the latent vectors Z^* and W^*. These conditions are similar to requirements placed on the strength of the factor loadings in factor models; see, e.g., Lam and Yao (2012) and many other studies. Based on the above conditions, the next theorem shows that the estimation accuracy of Z can be improved by including additional information from network D.

Theorem 2. *Suppose* $\widetilde{Z}$ *and* $\widetilde{\alpha}$ *satisfy* $\|\widetilde{Z} - Z^*\|_F^2 = O(1)$, $\|\widetilde{\alpha}1_{n_1}^\top - \alpha^* 1_{n_1}^\top\|_F^2 = O(n_1)$ *and the gradient of* L_A *with respect to* Z *vanishes when* $Z = \widetilde{Z}$. *Assume that* $\widetilde{\beta}, \widetilde{\gamma}$ *and* $\widetilde{W}$ *satisfy* $\|\widetilde{\beta}1_{n_2}^\top + 1_{n_1}\widetilde{\gamma}^\top - \beta^* 1_{n_2}^\top - 1_{n_1}\gamma^{*\top}\|_F^2 = O(n_1+n_2)$ *and* $\|\widetilde{W} - W^*\|_F^2 = O\{(n_1+n_2)/n_1\}$. *Under Conditions (C1) and (C2), given* $\widetilde{\alpha}, \widetilde{\beta}, \widetilde{\gamma}, \widetilde{Z}$ *and* $\widetilde{W}$, *if*

$$16 n_2 s^2 \widetilde{v}_k^2 \|\widetilde{Z} - Z^*\|_F^2 - \widetilde{v}_1 \left\|\beta^* 1_{n_2}^\top + 1_{n_1}\gamma^{*\top} + Z^* W^{*\top} - \widetilde{\beta} 1_{n_2}^\top - 1_{n_1}\widetilde{\gamma}^\top - Z^* \widetilde{W}^\top\right\|_F^2 > 0, \tag{5}$$

we have

$$\mathbb{E}\|\widehat{Z} - Z^*\|_F^2 < \|\widetilde{Z} - Z^*\|_F^2,$$

for any $0 < \eta_4 < \Delta = O\{n_1^{-1} + (n_1 n_2)^{-1/2}\}$, *where* $s = \sigma'(M_D)$, $\widetilde{v}_1$ *and* $\widetilde{v}_k$ *are the maximal and minimal eigenvalues of* $\widetilde{W}^\top \widetilde{W}/n_2$, *and* Δ *is defined as shown in (S7) in Appendix C.*

As shown in Theorem 2, the accuracy of the estimation of the common latent vectors Z can be improved by selecting an appropriate step size η_4. This result is reasonable since the two-mode network provides additional information about the shared latent vectors Z. In Theorem 2, the positivity of Δ is ensured by (5). Next, we provide some discussions to better understand (5). First, when the information in network A is limited, that is, $\tilde{Z}$ is far from Z^*, a small n_2 can lead to a positive Δ. Conversely, a large n_2 is preferred if the information in network A is rich. In addition, s in the left term of (5) is determined by M_D, which controls the sparsity of network D. If network D is denser and contains more information, the value of s is larger, and Δ is more likely to be positive. This suggests that when network D is dense and contains rich information, introducing information from network D can improve the estimation of the shared latent vectors Z, and indirectly improve the estimation of the matrix Θ^A. These interesting findings are also illustrated by our simulation studies. Finally, we note that Theorem 2 requires that the estimates of $\widetilde{\alpha}, \widetilde{\beta}, \widetilde{\gamma}, \widetilde{Z}$ and $\widetilde{W}$ are sufficiently close to the true values of the parameters. In Appendix D, we discuss the rationality of this assumption and show that these reasonable estimates are feasible.

4. Simulation studies

In this section, we conduct Monte Carlo simulations to compare the finite performance of the proposed model (gTLSM) with that of the one-mode latent space model (OLSM) and joint one-mode latent space model (jOLSM). The OLSM uses only network A to estimate Z and α. In the jOLSM, we regard the two types of nodes as being the same type and apply the one-mode latent space model proposed in Ma et al. (2020) to fit networks A and D simultaneously. To investigate the effects of the number of nodes and network density on the estimation accuracy, we consider the following four cases.

Case 1. We first study the effect of the number of nodes. We set $n_1 = 2n_2$ and vary the number of nodes of the first type (i.e., n_1) from 200 to 600. We set the latent space dimension to $k = 2$. Additional simulation results with $k = 4$ are presented in the Appendix to save space. The other model parameters are specified as follows:

(1) To control the average density of networks, generate the degree heterogeneity parameters $\alpha^*, \beta^*, \gamma^*$ with elements that are identically and independently distributed (i.i.d.) from the uniform distribution $U[-1.125, -0.375]$. Define constant $\omega^* = \beta^{*\top} 1_{n_1}$, then set $\beta^* = \beta^* - \omega^* 1_{n_1}$ and $\gamma^* = \gamma^* + \omega^* 1_{n_2}$.
(2) Generate k latent vector centers $\zeta_1, \cdots, \zeta_k \in \mathbb{R}^k$, with the elements i.i.d. from $U[-1, 1]$.
(3) Generate the latent variable $Z^* \in \mathbb{R}^{n_1 \times k}$: first generate a matrix $Z_0 \in \mathbb{R}^{n_1 \times k}$ with entries that are i.i.d. from the standard normal distribution $N(0, 1)$. Then, partition the n_1 data points into k subsets and add $\zeta_1, \cdots, \zeta_k$ to each subset. Finally, set $Z^* = J_1 Z^*$ and normalize Z^* such that $\|Z^* Z^{*\top}\|_F = n_1$.
(4) Repeat the above two steps to generate the latent variable $W^* \in \mathbb{R}^{n_2 \times k}$, except that in this case, we set $W^* = J_2 W^*$ and normalize W^* such that $\|W^* W^{*\top}\|_F = n_2$.

Given these parameters, we generate networks A and D based on model (1).

Case 2. In this case, we demonstrate the effect of the network density on the estimation accuracy. Following Zhang et al. (2022b), we consider different ranges for the degree heterogeneity parameters to achieve various network densities. Specifically, let each element of the degree heterogeneity parameters $\alpha^*, \beta^*, \gamma^*$ be independent and identically generated from $U[-2.25, -0.75]$. To keep the network density unchanged and uniquely determine parameters β and γ, then set $\omega^* = \beta^{*\top} 1_{n_1}$, $\beta^* = \beta^* - \omega^* 1_{n_1}$ and $\gamma^* = \gamma^* + \omega^* 1_{n_2}$. All other settings are the same as those in Case 1. Compared with Case 1, the network density varies from approximately 0.220 to 0.077.

Case 3. In this case, we study the relationship between the estimation error and the number of nodes of the second type (i.e., n_2). We fix $n_1 = 200$ and vary n_2 from 100 to 500. All other settings are the same as those in Case 1.

Table 1
The means and standard deviations of the relative estimation errors Δ_α, Δ_Z and Δ_Θ^A for Case 1 when $d=2$ and networks A and D are dense.

n_1	Measure	Δ_α			Δ_Z			Δ_Θ^A			Time		
		gTLSM	OLSM	jOLSM	gTLSM	OLSM	jOLSM	gTLSM	OLSM	jOLSM	gTLSM	OLSM	jOLSM
200	Mean	0.051	0.050	0.230	0.070	0.172	0.171	0.040	0.070	0.187	24.375	5.366	67.298
	SD	0.005	0.005	0.011	0.007	0.014	0.021	0.003	0.005	0.008	2.334	0.898	7.477
300	Mean	0.033	0.034	0.230	0.043	0.109	0.128	0.025	0.045	0.176	54.072	11.298	158.203
	SD	0.003	0.003	0.010	0.003	0.007	0.013	0.002	0.002	0.007	5.487	1.336	13.822
400	Mean	0.025	0.026	0.230	0.032	0.080	0.112	0.019	0.033	0.172	99.239	20.196	328.921
	SD	0.002	0.002	0.009	0.002	0.004	0.010	0.001	0.002	0.006	9.679	2.264	29.057
500	Mean	0.020	0.021	0.231	0.025	0.064	0.101	0.015	0.027	0.170	156.292	31.552	569.793
	SD	0.001	0.001	0.008	0.001	0.003	0.009	0.001	0.001	0.005	16.174	2.965	46.940
600	Mean	0.017	0.018	0.228	0.021	0.053	0.095	0.012	0.022	0.166	224.802	47.352	888.455
	SD	0.001	0.001	0.008	0.001	0.002	0.010	0.000	0.001	0.004	22.443	4.774	74.757

Table 2
The means and standard deviations of the relative estimation errors Δ_α, Δ_Z and Δ_Θ^A for Case 2 when $d=2$ and networks A and D are sparse.

n_1	Measure	Δ_α			Δ_Z			Δ_Θ^A			Time		
		gTLSM	OLSM	jOLSM	gTLSM	OLSM	jOLSM	gTLSM	OLSM	jOLSM	gTLSM	OLSM	jOLSM
200	Mean	0.039	0.045	0.105	0.146	0.339	0.165	0.037	0.059	0.082	29.698	6.387	69.308
	SD	0.003	0.003	0.006	0.013	0.045	0.014	0.002	0.005	0.004	3.524	0.941	7.392
300	Mean	0.037	0.043	0.106	0.107	0.239	0.111	0.033	0.049	0.077	66.618	13.811	180.107
	SD	0.002	0.002	0.005	0.007	0.034	0.008	0.002	0.004	0.004	4.391	1.171	15.637
400	Mean	0.036	0.042	0.106	0.090	0.192	0.088	0.031	0.044	0.076	121.476	25.901	375.150
	SD	0.002	0.002	0.005	0.006	0.024	0.007	0.001	0.002	0.003	8.810	2.442	31.608
500	Mean	0.035	0.041	0.107	0.080	0.162	0.073	0.029	0.041	0.075	194.907	41.133	649.764
	SD	0.001	0.002	0.004	0.005	0.009	0.006	0.001	0.001	0.003	12.932	2.994	49.045
600	Mean	0.034	0.041	0.106	0.073	0.146	0.065	0.029	0.040	0.074	282.370	61.835	1006.447
	SD	0.001	0.001	0.004	0.004	0.008	0.004	0.001	0.001	0.003	22.289	4.427	72.407

Table 3
The means and standard deviations of the relative estimation errors Δ_α, Δ_Z and Δ_Θ^A for Case 3 when $d=2$, $n_1=200$, and networks A and D are dense.

n_2	Measure	Δ_α			Δ_Z			Δ_Θ^A			Time		
		gTLSM	OLSM	jOLSM	gTLSM	OLSM	jOLSM	gTLSM	OLSM	jOLSM	gTLSM	OLSM	jOLSM
100	Mean	0.051	0.050	0.060	0.121	0.169	0.123	0.055	0.069	0.061	8.924	3.630	8.771
	SD	0.005	0.005	0.006	0.009	0.014	0.009	0.003	0.004	0.003	0.718	0.256	0.543
200	Mean	0.050	0.050	0.105	0.094	0.171	0.115	0.046	0.069	0.087	11.345	3.536	17.228
	SD	0.006	0.006	0.008	0.008	0.014	0.010	0.003	0.005	0.005	0.907	0.211	0.927
300	Mean	0.050	0.050	0.166	0.078	0.170	0.134	0.042	0.069	0.132	13.753	3.545	30.283
	SD	0.005	0.005	0.009	0.008	0.013	0.014	0.003	0.005	0.006	0.970	0.180	2.355
400	Mean	0.051	0.050	0.230	0.070	0.172	0.171	0.040	0.070	0.187	15.868	3.578	47.957
	SD	0.005	0.005	0.011	0.007	0.014	0.021	0.003	0.005	0.008	1.098	0.172	4.011
500	Mean	0.051	0.050	0.293	0.064	0.172	0.221	0.038	0.070	0.245	18.171	3.688	71.762
	SD	0.005	0.004	0.014	0.006	0.014	0.028	0.003	0.005	0.012	1.323	0.306	6.859

Case 4. Compared with Case 3, we assume that network D is sparser in this case. Specifically, each element of the degree heterogeneity parameters β^* and γ^* is independent and identically generated from $U[-2.25, -0.75]$. Then set $\omega^* = \beta^{*\top}1_{n_1}$, $\beta^* = \beta^* - \omega^* 1_{n_1}$ and $\gamma^* = \gamma^* + \omega^* 1_{n_2}$. Under this setting, the density of network D is approximately 0.077. The other parameters are generated in the same manner as in Case 3.

For each setting, we assume that the latent space dimension k is known and replicate the experiment 100 times. Then, we calculate the means and standard deviations of the relative estimation errors $\Delta_\alpha = \|\hat{\alpha} - \alpha^*\|_2^2 / \|\alpha^*\|_2^2$, $\Delta_Z = \|\hat{Z}\hat{Z}^\top - Z^* Z^{*\top}\|_F^2 / \|Z^* Z^{*\top}\|_F^2$ and $\Delta_\Theta^A = \|\hat{\Theta}^A - \Theta^{A*}\|_F^2 / \|\Theta^{A*}\|_F^2$. The average CPU time (in seconds) is also recorded using a PC with 2.10 GHz and 96 GB RAM. The simulation results for Cases 1–4 are shown in Tables 1–4. We can draw the following conclusions based on the estimation results. First, in Case 1, as the number of nodes n_1 and n_2 increases, the performance of the MLE improves for all three models, and the proposed model outperforms the other models. This result is not surprising based on Theorem 1. Second, regardless of whether the network is dense or sparse, the estimation errors of Z and Θ^A in the proposed model are smaller than those in the one-mode latent space model. This result implies that incorporating information from network D improves the estimation accuracies of Z and Θ^A. Third, as shown in Tables 3 and 4, the estimation errors of Z and Θ^A in the proposed model decrease as n_2 increases. However, the

Table 4
The means and standard deviations of the relative estimation errors Δ_α, Δ_Z and Δ_Θ^A for Case 4 when $d = 2$, $n_1 = 200$, network A is dense and network D is sparse.

n_2	Measure	Δ_α			Δ_Z			Δ_Θ^A			Time		
		gTLSM	OLSM	jOLSM	gTLSM	OLSM	jOLSM	gTLSM	OLSM	jOLSM	gTLSM	OLSM	jOLSM
100	Mean	0.052	0.050	0.057	0.148	0.169	0.162	0.064	0.069	0.071	12.602	5.024	13.968
	SD	0.005	0.005	0.006	0.012	0.014	0.014	0.004	0.004	0.005	1.761	0.840	1.984
200	Mean	0.051	0.050	0.077	0.134	0.171	0.179	0.059	0.069	0.087	16.602	5.175	26.311
	SD	0.006	0.006	0.006	0.011	0.014	0.016	0.004	0.005	0.007	2.043	1.004	3.421
300	Mean	0.051	0.050	0.100	0.120	0.170	0.211	0.055	0.069	0.109	20.794	4.980	43.015
	SD	0.005	0.005	0.008	0.011	0.013	0.021	0.004	0.005	0.008	3.499	0.823	5.471
400	Mean	0.052	0.050	0.120	0.111	0.172	0.264	0.052	0.070	0.136	24.976	5.089	68.174
	SD	0.005	0.005	0.010	0.009	0.014	0.026	0.003	0.005	0.011	3.510	0.879	8.622
500	Mean	0.052	0.050	0.138	0.103	0.171	0.321	0.050	0.070	0.163	29.241	5.110	100.659
	SD	0.005	0.004	0.013	0.010	0.014	0.034	0.004	0.005	0.014	3.401	0.928	12.985

estimation errors in the joint one-mode latent space model increase since the model is not specified correctly. Fourth, a comparison of Tables 3 and 4 shows that when network D is sparse, the performance of the proposed model is worse than the performance when network D is dense. The above two findings suggest that if network D is denser or if the number of nodes of the second type is larger, incorporating information from network D is more beneficial for improving the estimation of the shared latent vectors. These findings are consistent with our theoretical results. Fifth, our simulation studies demonstrate that this algorithm is computationally efficient and scalable for large-scale networks. For example, to fit a generalized network with $n_1 = 600$ and $n_2 = 1200$ in Case 1, the average CPU time is 282.370 seconds.

5. Real data analysis

To demonstrate the practicality of the proposed model, we apply the proposed generalized latent space model to analyze two real datasets, namely, the Last.fm dataset and the Yelp dataset.

5.1. Last.fm dataset analysis

The Last.fm dataset was collected from the Last.fm online music system (http://www.last.fm). This dataset includes two types of networks: a user-user friendship network and a user-listened-artist relationship network. For ease of illustration, we remove users and artists with extremely small node degrees, with 134 users and 411 artists included in the subsequent analysis. The density of the user-user network is 0.07, and the density of the user-artist network is 0.1.

Next, we apply the proposed model to fit the user-user and user-artist networks. Following Zhang et al. (2020b), we set the dimension of the latent space to $k = 3$ for ease of visualization. After the latent positions of the users are determined, we apply the k-means algorithm to cluster the users. The latent positions of the users are shown in Fig. 1(a). The clusters are well separated in the 3-dimensional embedding space. Then, we count the top 10 artists that were most listened to by the users in each cluster, yielding 24 artists. For each artist, we calculate the proportion of users in each cluster who listen to that artist. As shown in Fig. 1(b), the proportions differ considerably in various clusters. Next, we visualize the node heterogeneity parameters of users in the two networks, finding that users have distinct connectivity patterns in the two networks, which demonstrates the rationality of choosing network-specific node heterogeneity parameters. Considering the practical meaning of the node heterogeneity parameters, we follow Wu et al. (2022) and apply four-quadrant clustering to segment users. Specifically, we set the 80% quantile of parameters $\hat{\alpha}$ and $\hat{\beta}$ as thresholds and divide the users into four classes. Fig. 1(d) shows that users in class 4 have many friends and listen to a wide range of artists. Conversely, users in class 3 have a large number of friends but listen to only a few artists. Thus, the music platform should focus on these users and recommend more niche singers, thereby improving the evaluations of these users and their friends on the platform and different artists.

Finally, it is of interest to investigate whether including additional information from other networks could assist in predicting missing links of the target network due to the correlation between different networks. To this end, we compare the prediction performance of the proposed model with that of the one-mode latent space model and joint one-mode latent space model. Specifically, we randomly remove some links in the user-user friendship network, fit the models based on the remaining links and attempt to predict the missing links. The missing ratio was varied from 0.05 to 0.2. The above procedure was repeated 20 times, and the means and standard deviations of three indices, namely, the AUC value, log-likelihood value and mean square error between the predicted probabilities and observed values, were calculated. The prediction results are shown in Table 5. It is not surprising that the joint one-mode latent space model has the largest mean square error and smallest log-likelihood value, since it does not take the node type information and heterogeneous node connection patterns of different networks into account. In addition, the proposed model has superior performance compared with the one-mode latent space model, which indicates that incorporating information from the user-artist network is beneficial for improving the link prediction performance.

Fig. 1. The results for Last.fm dataset: (a) visualization of the latent vectors; (b) proportion of users in each cluster who listen to each artist; (c) histograms of the degree heterogeneity parameters; (d) locations of the degree heterogeneity parameters.

Table 5
Detailed prediction results for the Last.fm dataset. AUC: the AUC value; Logf: the log-likelihood value; Error: the mean square error between the predicted probabilities and observed values.

Ratio	Measure	gTLSM			OLSM			jOLSM		
		AUC	Logf	Error	AUC	Logf	Error	AUC	Logf	Error
0.05	Mean	0.761	-0.229	0.247	0.689	-0.250	0.257	0.714	-0.314	0.295
	SD	0.044	0.030	0.022	0.051	0.037	0.022	0.039	0.019	0.013
0.10	Mean	0.767	-0.228	0.247	0.682	-0.251	0.258	0.714	-0.320	0.300
	SD	0.026	0.013	0.009	0.033	0.018	0.010	0.028	0.008	0.006
0.15	Mean	0.761	-0.229	0.248	0.680	-0.251	0.258	0.707	-0.328	0.305
	SD	0.017	0.015	0.011	0.026	0.017	0.011	0.018	0.010	0.008
0.20	Mean	0.759	-0.233	0.250	0.674	-0.256	0.260	0.699	-0.337	0.311
	SD	0.013	0.011	0.008	0.029	0.015	0.009	0.015	0.008	0.006

5.2. *Yelp dataset analysis*

The Yelp dataset was downloaded from https://www.yelp.com/dataset. This dataset contains rich information about different businesses, reviews and user data. The network includes two types of nodes (users and businesses) and two types of links (user-user and user-business). The user-user network represents the friend relationships among different users. The user-business network is constructed as follows: one user and one business are linked if the user writes a review for this business. Since the Yelp dataset

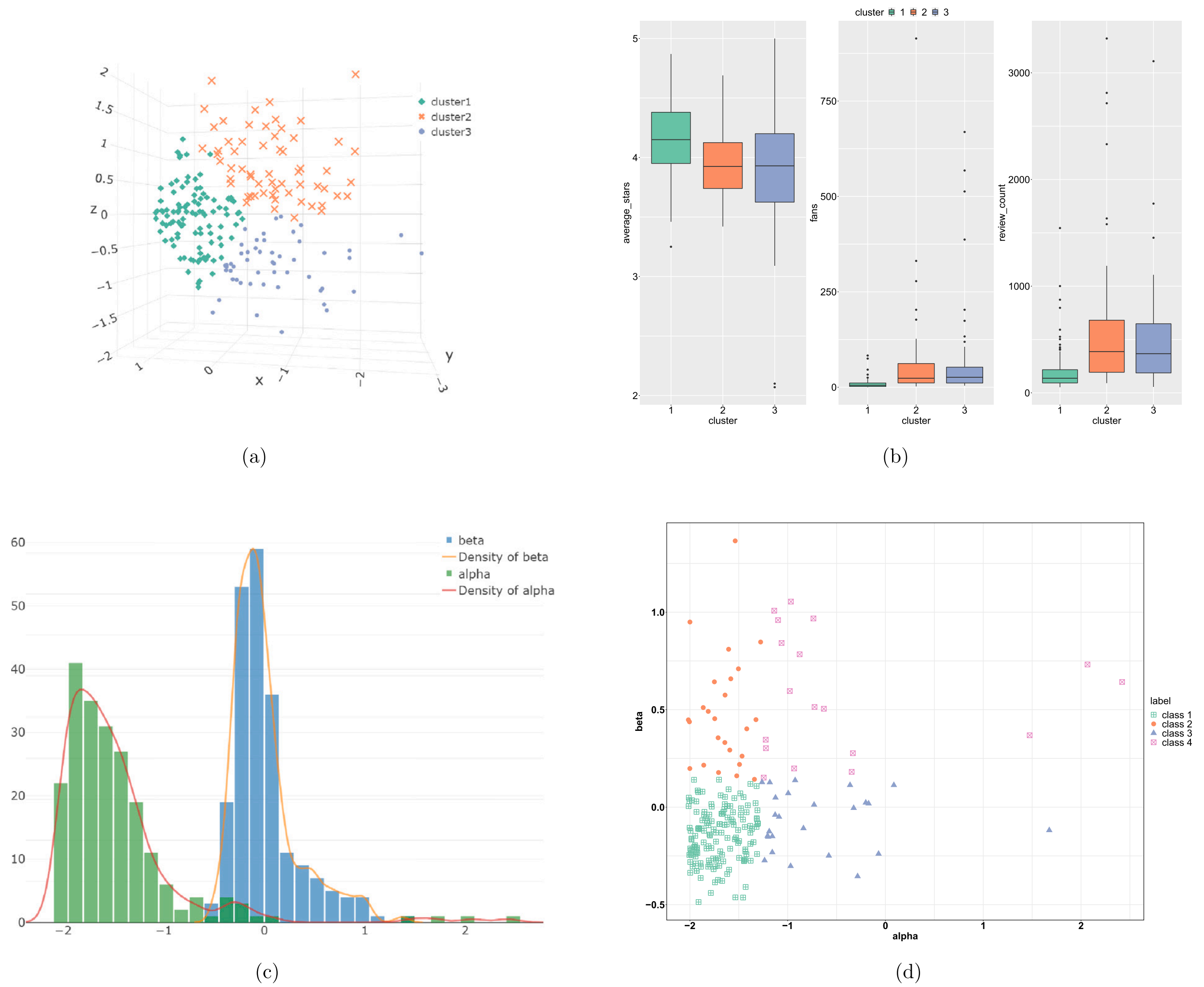

Fig. 2. The results for the Yelp dataset: (a) visualization of the latent vectors; (b) boxplots of the average star rating, number of fans and number of reviews for each cluster; (c) histograms of the degree heterogeneity parameters; (d) locations of the degree heterogeneity parameters.

includes 1,987,897 users and 150,346 businesses, we extract a subset of the Yelp dataset, focusing on businesses located in Saint Petersburg. Because the user-business network is extremely sparse, we remove some users and businesses so that the node degree of the user is greater than 20 and the node degree of the business is greater than 10. Finally, 212 users and 208 businesses are investigated in the subsequent analysis. The density of the user-user network is 0.068, and the density of the user-business network is 0.115.

We apply the proposed model with $k = 3$ to jointly model the user-user and user-business networks. The estimated latent vectors for the users are shown in Fig. 2(a). For users in each cluster, we analyze the average star rating given by the users to different businesses, the total number of reviews provided by the users and the number of fans of users. Fig. 2(b) shows that users in cluster 2 have more fans and have written more reviews with higher star ratings, which can be regarded as superstars. These characteristics are very different from those of users in cluster 1, who have few fans and have provided few reviews. Moreover, we can see that cluster 1 contains more than 50% of the total users, which implies that most users in the platform are inactive. Furthermore, we visualize the connectivity patterns between users in the two networks. In the user-user network, most users have few connections; however, some users are closely connected to other users. In the user-business network, most users have similar connectivity patterns. Thus, we classify the users into four classes based on their node degree heterogeneity parameters. Businesses should focus on users with more friends and adopt different strategies for users in classes 3 and 4 to improve their star rating.

Finally, we compare the missing link prediction performance of the proposed model, the one-mode latent space model and the joint one-mode latent space model. The prediction results are shown in Table 6. The proposed model has higher AUC values, higher log-likelihood values and smaller errors than the one-mode latent space model and joint one-mode latent space model. These results demonstrate that leveraging information from the user-business network is beneficial for predicting the missing links of the user-user network.

Table 6
Detailed prediction results for the Yelp dataset. AUC: the AUC value; Logf: the log-likelihood value; Error: the mean square error between the predicted probabilities and observed values.

Ratio	Measure	gTLSM			OLSM			jOLSM		
		AUC	Logf	Error	AUC	Logf	Error	AUC	Logf	Error
0.05	Mean	0.925	-0.146	0.193	0.913	-0.149	0.195	0.890	-0.182	0.213
	SD	0.021	0.012	0.010	0.026	0.013	0.011	0.026	0.011	0.009
0.10	Mean	0.925	-0.150	0.196	0.912	-0.154	0.198	0.889	-0.187	0.216
	SD	0.014	0.006	0.005	0.017	0.007	0.006	0.019	0.007	0.005
0.15	Mean	0.921	-0.152	0.197	0.905	-0.156	0.199	0.881	-0.191	0.218
	SD	0.013	0.005	0.004	0.014	0.005	0.004	0.019	0.005	0.003
0.20	Mean	0.920	-0.154	0.198	0.903	-0.159	0.201	0.878	-0.194	0.220
	SD	0.009	0.004	0.004	0.012	0.005	0.004	0.013	0.004	0.003

6. Conclusions

In this article, we propose a novel method for one-mode networks with awareness of two-mode networks. The proposed model provides a flexible method for modeling common structures and heterogeneous connectivity patterns across different networks. In addition, a feasible and efficient algorithm for estimating the degree parameters and latent vectors is proposed. Moreover, we provide an upper bound for the estimators and theoretically demonstrate that incorporating two-mode networks improves the estimation accuracy of the shared latent positions under some typical conditions. These findings are supported by the results of numerical experiments on various simulated and real examples.

To conclude the article, we discuss several interesting topics that could be investigated in future works. First, our model can be extended to take node features into account. How to leverage nodal information and additional network information to assist in modeling one-mode networks is an interesting topic. Second, the proposed model assumes that nodes of the first type in different networks are represented by the same latent vectors, and it is of interest to assume that some, but not all, latent positions are shared in different networks. Finally, the model can be extended to impose nonlinear structures on the latent vectors to capture more complex network structures.

Appendix

This appendix is organized as follows. Appendix A presents the result of Proposition 1 as well as its proof. Appendices B–C present the proofs of Theorems 1 and 2, respectively. Appendix D discusses the properties of initial estimates. Appendix E shows the details of Algorithm 1. Additional simulation results are shown in Appendix F.

Appendix A. Proof of Proposition 1

Lemma 1. *For any $t=(t_1,\cdots,t_{n_1})^\top\in\mathbb{R}^{n_1}$ and $s=(s_1,\cdots,s_{n_2})^\top\in\mathbb{R}^{n_2}$, if $t1_{n_2}^\top 1_{n_2}+1_{n_1}s^\top 1_{n_2}=0$, then $t_1=\cdots=t_{n_1}=-\frac{1}{n_2}\sum_{i=1}^{n_2}s_i$.*

Proof. This lemma is modified from Lemma 3 of Zhang et al. (2020b).

Proposition 1. *Suppose that two sets of parameters $(\alpha,\beta,\gamma,Z,W)$ and $\left(\alpha',\beta',\gamma',Z',W'\right)$ satisfy the following conditions:*
(1) $J_1Z=Z,\ J_1Z'=Z'$;
(2) $J_2W=W,\ J_2W'=W'$;
(3) $\beta^\top 1_{n_1}=0,\ \beta'^\top 1_{n_1}=0$.
Then

$$\alpha 1_{n_1}^\top+1_{n_1}\alpha^\top+ZZ^\top=\alpha' 1_{n_1}^\top+1_{n_1}\alpha'^\top+Z'Z'^\top, \tag{S1}$$

$$\beta 1_{n_2}^\top+1_{n_1}\gamma^\top+ZW^\top=\beta' 1_{n_2}^\top+1_{n_1}\gamma'^\top+Z'W'^\top, \tag{S2}$$

implies that there exists an orthonormal matrix $O\in\mathbb{R}^{k\times k}$ where $OO^\top=O^\top O=I_k$, such that

$$\alpha=\alpha',\beta=\beta',\gamma=\gamma',Z=Z'O,W=W'O.$$

Proof. Since $J_1Z=Z,\ J_1Z'=Z'$, we have $ZZ^\top 1_{n_1}=Z'Z'^\top 1_{n_1}=0$. Right multiplying 1_{n_1} to both sides in (S1) gives

$$\alpha 1_{n_1}^\top 1_{n_1}+1_{n_1}\alpha^\top 1_{n_1}=\alpha' 1_{n_1}^\top 1_{n_1}+1_{n_1}\alpha'^\top 1_{n_1}.$$

That is,

$$(\alpha-\alpha')1_{n_1}^\top 1_{n_1}+1_{n_1}(\alpha-\alpha')^\top 1_{n_1}=0.$$

By Lemma 1, we have $\alpha_i-\alpha_i'=-\frac{1}{n_1}\sum_{i=1}^{n_1}(\alpha_i-\alpha_i')$ and $\alpha=\alpha'$. Under conditions (1)–(2), we have $ZW^\top 1_{n_2}=Z'W'^\top 1_{n_2}=0$ and $1_{n_1}^\top ZW^\top=1_{n_1}^\top Z'W'^\top=0$. This leads to

$$\beta 1_{n_2}^\top 1_{n_2}+1_{n_1}\gamma^\top 1_{n_2}=\beta' 1_{n_2}^\top 1_{n_2}+1_{n_1}\gamma'^\top 1_{n_2}, 1_{n_1}^\top\beta 1_{n_2}^\top+1_{n_1}^\top 1_{n_1}\gamma^\top=1_{n_1}^\top\beta' 1_{n_2}^\top+1_{n_1}^\top 1_{n_1}\gamma'^\top.$$

Hence,

$$(\beta-\beta')1_{n_2}^\top 1_{n_2}+1_{n_1}(\gamma-\gamma')^\top 1_{n_2}=0, 1_{n_1}^\top(\beta-\beta')1_{n_2}^\top+1_{n_1}^\top 1_{n_1}(\gamma-\gamma')^\top=0.$$

By Lemma 1, we have $\beta_i-\beta_i'=-\frac{1}{n_2}\sum_{i=1}^{n_2}(\gamma_i-\gamma_i')$ and $\gamma_i-\gamma_i'=-\frac{1}{n_1}\sum_{i=1}^{n_1}(\beta_i-\beta_i')$. Since $\beta^\top 1_{n_1}=0$ and $\beta'^\top 1_{n_1}=0$, we have $\beta=\beta'$ and $\gamma=\gamma'$. Combining (S1)–(S2), we have

$$ZZ^\top=Z'Z'^\top, ZW^\top=Z'W'^\top.$$

Therefore, there exists an orthonormal matrix $O\in\mathbb{R}^{k\times k}$ where $OO^\top=O^\top O=I_k$, such that $Z=Z'O$ and $W=W'O$.

Appendix B. Proof of Theorem 1

Lemma 2. *Let $\mathcal{X}\in\mathbb{R}^{n_1\times\cdots\times n_K}$ be a K-way tensor. Assume each element $\mathcal{X}_{i_1\cdots i_k}$ is independent, zero mean and satisfies $\mathbb{E}(e^{t\mathcal{X}_{i_1\cdots i_k}})\le e^{\sigma^2t^2/2}$. Then there exist positive constants c and C which only depend on σ^2 and K such that with probability at least $1-e^{-c\sum_{k=1}^K n_k}$, $\|\mathcal{X}\|_2\le C\sqrt{\sum_{k=1}^K n_k}$.*

Proof. This lemma is from theorem 1 of Tomioka and Suzuki (2014) and we cite it as a lemma.

Proof of Theorem 1. Define $b(x)=\ln(1+\exp(x))$. Accordingly,

$$L(\alpha,\beta,\gamma,Z,W)=-\sum_{i=1}^{n_1}\sum_{j=1}^{n_1}\left\{A_{ij}\Theta_{ij}^A-b(\Theta_{ij}^A)\right\}-\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\left\{D_{ij}\Theta_{ij}^D-b(\Theta_{ij}^D)\right\}.$$

Since $(\hat\alpha,\hat\beta,\hat\gamma,\hat Z,\hat W)=\operatorname{argmin}_{\Theta\in\mathcal{F}}L(\alpha,\beta,\gamma,Z,W)$, then

$$L(\hat\alpha,\hat\beta,\hat\gamma,\hat Z,\hat W)-L(\alpha^*,\beta^*,\gamma^*,Z^*,W^*)\le 0. \tag{S3}$$

Further, we have

$$\begin{aligned}
&L(\hat\alpha,\hat\beta,\hat\gamma,\hat Z,\hat W)-L(\alpha^*,\beta^*,\gamma^*,Z^*,W^*)\\
&=\sum_{i=1}^{n_1}\sum_{j=1}^{n_1}\left\{A_{ij}\Theta_{ij}^{*A}-A_{ij}\hat\Theta_{ij}^A+b(\hat\Theta_{ij}^A)-b(\Theta_{ij}^{*A})\right\}\\
&+\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\left\{D_{ij}\Theta_{ij}^{*D}-D_{ij}\hat\Theta_{ij}^D+b(\hat\Theta_{ij}^D)-b(\Theta_{ij}^{*D})\right\}\\
&=\sum_{i=1}^{n_1}\sum_{j=1}^{n_1}\left\{\left(\Theta_{ij}^{*A}-\hat\Theta_{ij}^A\right)\left(A_{ij}-b'(\Theta_{ij}^{*A})\right)+b(\hat\Theta_{ij}^A)-b(\Theta_{ij}^{*A})+b'(\Theta_{ij}^{*A})\left(\Theta_{ij}^{*A}-\hat\Theta_{ij}^A\right)\right\}\\
&+\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\left\{\left(\Theta_{ij}^{*D}-\hat\Theta_{ij}^D\right)\left(D_{ij}-b'(\Theta_{ij}^{*D})\right)+b(\hat\Theta_{ij}^D)-b(\Theta_{ij}^{*D})+b'(\Theta_{ij}^{*D})\left(\Theta_{ij}^{*D}-\hat\Theta_{ij}^D\right)\right\}\\
&=\sum_{i=1}^{n_1}\sum_{j=1}^{n_1}\left\{\left(\Theta_{ij}^{*A}-\hat\Theta_{ij}^A\right)\left(A_{ij}-b'(\Theta_{ij}^{*A})\right)+\frac{1}{2}b''(\tilde\Theta_{ij}^A)\left(\Theta_{ij}^{*A}-\hat\Theta_{ij}^A\right)^2\right\}\\
&+\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\left\{\left(\Theta_{ij}^{*D}-\hat\Theta_{ij}^D\right)\left(D_{ij}-b'(\Theta_{ij}^{*D})\right)+\frac{1}{2}b''(\tilde\Theta_{ij}^D)\left(\Theta_{ij}^{*D}-\hat\Theta_{ij}^D\right)^2\right\},
\end{aligned}$$

where $\tilde\Theta_{ij}^A=\lambda_{1,ij}\hat\Theta_{ij}^A+(1-\lambda_{1,ij})\Theta_{ij}^{*A}$ for some $\lambda_{1,ij}\in(0,1)$ and $\tilde\Theta_{ij}^D=\lambda_{2,ij}\hat\Theta_{ij}^D+(1-\lambda_{2,ij})\Theta_{ij}^{*D}$ for some $\lambda_{2,ij}\in(0,1)$. Since we assume $\|\hat\Theta^A\|_{\max}\le M_A$ and $\|\Theta^{*A}\|_{\max}\le M_A$, this implies that $\|\tilde\Theta^A\|_{\max}<M_A$. Similarly, $\|\tilde\Theta^D\|_{\max}<M_D$. Together with (S3), we have

$$
\begin{aligned}
&\frac{1}{2} \min_{|v|<\max\{M_A,M_D\}} b''(v)\|\Theta^* - \hat{\Theta}\|_F^2 \\
\le & \frac{1}{2} \min_{|v|<M_A} b''(v)\|\Theta^{*A} - \hat{\Theta}^A\|_F^2 + \frac{1}{2} \min_{|v|<M_D} b''(v)\|\Theta^{*D} - \hat{\Theta}^D\|_F^2 \\
\le & \sum_{i=1}^{n_1}\sum_{j=1}^{n_1} \left\{ \left(\hat{\Theta}_{ij}^A - \Theta_{ij}^{*A}\right)\left(A_{ij} - b'(\Theta_{ij}^{*A})\right)\right\} + \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} \left\{ \left(\hat{\Theta}_{ij}^D - \Theta_{ij}^{*D}\right)\left(D_{ij} - b'(\Theta_{ij}^{*D})\right)\right\} \\
= & \langle E^A, \hat{\Theta}^A - \Theta^{*A}\rangle + \langle E^D, \hat{\Theta}^D - \Theta^{*D}\rangle \\
= & \langle E, \hat{\Theta} - \Theta^*\rangle,
\end{aligned} \tag{S4}
$$

where $E^A = A - b'(\Theta^{*A}) \in \mathbb{R}^{n_1\times n_1}$, $E^D = D - b'(\Theta^{*D}) \in \mathbb{R}^{n_1\times n_2}$ and $E = [E^A, E^D] \in \mathbb{R}^{n_1\times(n_1+n_2)}$. Furthermore, we have

$$
\left|\langle E, \hat{\Theta} - \Theta^*\rangle\right| \le \|E\|_2 \sqrt{\operatorname{rank}(\hat{\Theta} - \Theta^*)}\|\hat{\Theta} - \Theta^*\|_F \le 2\sqrt{k+2}\cdot\|E\|_2\|\hat{\Theta} - \Theta^*\|_F, \tag{S5}
$$

where the second inequality is by the fact that $\operatorname{rank}(\hat{\Theta} - \Theta^*) \le 4(k+2)$. By Hoeffding's lemma, entries in E are independent, mean-zero sub-gaussian random variables with $\mathbb{E}(e^{tE_{ij}}) \le e^{t^2/8}$. Using Lemma 2, with probability at least $1 - e^{-c(2n_1+n_2)}$, we have

$$
\|E\|_2 \le C\sqrt{2n_1 + n_2} \tag{S6}
$$

Based on (S4), (S5) and (S6), we can obtain that

$$
\|\hat{\Theta} - \Theta^*\|_F \le \frac{C\sqrt{k+2}\sqrt{2n_1+n_2}}{\min_{|v|<\max\{M_A,M_D\}} b''(v)}.
$$

Appendix C. Proof of Theorem 2

Suppose the gradient of L_A with respect to Z will vanish, then the update of $\widetilde{Z}$ is

$$
\widehat{Z} = \widetilde{Z} + \eta_4\left(D - \sigma(\widetilde{\Theta}^D)\right)\widetilde{W}.
$$

Define $\widetilde{\delta}_i = (\widetilde{\beta}_i + \widetilde{\gamma}_1, \cdots, \widetilde{\beta}_i + \widetilde{\gamma}_{n_2})^\top \in \mathbb{R}^{n_2}$. For a particular node i, we have

$$
\widehat{Z}_i = \widetilde{Z}_i + \eta_4\widetilde{W}^\top(D_{i.}^\top - \sigma(\widetilde{\delta}_i + \widetilde{W}\widetilde{Z}_i)).
$$

Hence,

$$
\begin{aligned}
&\widehat{Z}_i - Z_i^* = \widetilde{Z}_i - Z_i^* + \eta_4\widetilde{W}^\top(D_{i.}^\top - \sigma(\widetilde{\delta}_i + \widetilde{W}\widetilde{Z}_i)) \\
=&\widetilde{Z}_i - Z_i^* + \eta_4\widetilde{W}^\top\left[D_{i.}^\top - \sigma(\delta_i^* + W^*Z_i^*) + \sigma(\delta_i^* + W^*Z_i^*) - \sigma(\widetilde{\delta}_i + \widetilde{W}Z_i^*)\right. \\
&\left.+\sigma(\widetilde{\delta}_i + \widetilde{W}Z_i^*) - \sigma(\widetilde{\delta}_i + \widetilde{W}\widetilde{Z}_i)\right] \\
=&\left[I_k - \eta_4\widetilde{W}^\top \operatorname{diag}(\sigma'(\xi_i))\widetilde{W}\right](\widetilde{Z}_i - Z_i^*) + \eta_4\widetilde{W}^\top\left[\sigma(\delta_i^* + W^*Z_i^*) - \sigma(\widetilde{\delta}_i + \widetilde{W}Z_i^*)\right] \\
&+\eta_4\widetilde{W}^\top\left[D_{i.}^\top - \sigma(\delta_i^* + W^*Z_i^*)\right] \\
:=&S_{i1} + S_{i2},
\end{aligned}
$$

where $\xi_i \in \mathbb{R}^{n_2}$ is between $\widetilde{\delta}_i + \widetilde{W}Z_i^*$ and $\widetilde{\delta}_i + \widetilde{W}\widetilde{Z}_i$. Given $\widetilde{\alpha}, \widetilde{\beta}, \widetilde{\gamma}, \widetilde{Z}$ and $\widetilde{W}$, we have

$$
\mathbb{E}\|\widehat{Z} - Z^*\|_F^2 = \sum_{i=1}^{n_1}\mathbb{E}\|\widehat{Z}_i - Z_i^*\|^2 = \sum_{i=1}^{n_1}\|S_{i1}\|_2^2 + \sum_{i=1}^{n_1}\mathbb{E}\|S_{i2}\|_2^2 + 2\sum_{i=1}^{n_1}\mathbb{E}S_{i1}^\top S_{i2}.
$$

This third term equals to 0 since S_{i2} is a mean zero vector. For the first term, we denote $T_1 = \operatorname{diag}(\sigma'(\xi_1), \cdots, \sigma'(\xi_{n_1})) \in \mathbb{R}^{n_1n_2\times n_1n_2}$, $T_2 = I_{n_1}\otimes I_k - \eta_4(I_{n_1}\otimes\widetilde{W}^\top)T_1(I_{n_1}\otimes\widetilde{W}) \in \mathbb{R}^{n_1k\times n_1k}$ and $\delta = \beta 1_{n_2}^\top + 1_{n_1}\gamma^\top = (\delta_1, \cdots, \delta_{n1})^\top \in \mathbb{R}^{n_1\times n_2}$. Then we have

$$
\begin{aligned}
\sum_{i=1}^{n_1}\|S_{i1}\|_2^2 =& \sum_{i=1}^{n_1}\left\|\left[I_k - \eta_4\widetilde{W}^\top \operatorname{diag}(\sigma'(\xi_i))\widetilde{W}\right](\widetilde{Z}_i - Z_i^*)\right. \\
&\left.+\eta_4\widetilde{W}^\top\left[\sigma(\delta_i^* + W^*Z_i^*) - \sigma(\widetilde{\delta}_i + \widetilde{W}Z_i^*)\right]\right\|_2^2 \\
=&\left\|T_2 \operatorname{vec}(\widetilde{Z}^\top - Z^{*\top})\right. \\
&\left.+\eta_4 \operatorname{vec}\left\{\widetilde{W}^\top\left[\sigma(\delta^* + Z^*W^{*\top}) - \sigma(\widetilde{\delta} + Z^*\widetilde{W}^\top)\right]^\top\right\}\right\|_F^2.
\end{aligned}
$$

Denote $\widetilde{\nu}_i$ $(i = 1, \cdots, k)$ as the i-th largest eigenvalue of $\widetilde{W}^\top\widetilde{W}/n_2$. Then we have

$$\left\|T_2\,\mathrm{vec}(\widetilde{Z}^\top - Z^{*\top})\right\|_F^2 \le \lambda_{\max}^2(T_2)\|\widetilde{Z} - Z^*\|_F^2 \le (1 - s\eta_4 n_2\widetilde{\nu}_k)^2\|\widetilde{Z} - Z^*\|_F^2,$$

where $s = \sigma'(M_D) > 0$. In addition,

$$\begin{aligned}
&\left\|\mathrm{vec}\left\{\widetilde{W}^\top\left[\sigma(\delta^* + Z^*W^{*\top}) - \sigma(\widetilde{\delta} + Z^*\widetilde{W}^\top)\right]^\top\right\}\right\|_F^2\\
\le&\left\|\widetilde{W}\right\|_F^2\left\|\sigma(\delta^* + Z^*W^{*\top}) - \sigma(\widetilde{\delta} + Z^*\widetilde{W}^\top)\right\|_F^2\\
\le&\frac{1}{16}\left\|\widetilde{W}\right\|_F^2\left\|\delta^* + Z^*W^{*\top} - \widetilde{\delta} - Z^*\widetilde{W}^\top\right\|_F^2\\
\le&\frac{1}{16}n_2\widetilde{\nu}_1\left\|\delta^* + Z^*W^{*\top} - \widetilde{\delta} - Z^*\widetilde{W}^\top\right\|_F^2.
\end{aligned}$$

This leads to

$$\sum_{i=1}^{n_1}\|S_{i1}\|_2^2 \le \left((1 - s\eta_4 n_2\widetilde{\nu}_k)\|\widetilde{Z} - Z^*\|_F + \frac{1}{4}\eta_4\sqrt{n_2\widetilde{\nu}_1}\left\|\delta^* + Z^*W^{*\top} - \widetilde{\delta} - Z^*\widetilde{W}^\top\right\|_F\right)^2.$$

For the second term, denote $V_i = \mathrm{diag}(\sigma'(\delta_i^* + W^*Z_i^*))$ and then we have

$$\begin{aligned}
\sum_{i=1}^{n_1}\mathbb{E}\|S_{i2}\|_2^2 &= \eta_4^2\sum_{i=1}^{n_1}\mathrm{tr}(\widetilde{W}^\top V_i\widetilde{W})\\
&= \eta_4^2\,\mathrm{tr}\left[(I_{n_1}\otimes\widetilde{W}^\top)\,\mathrm{diag}\left\{\mathrm{vec}\left[\sigma'\left((\Theta^{*D})^\top\right)\right]\right\}(I_{n_1}\otimes\widetilde{W})\right]\\
&\le \frac{1}{4}\eta_4^2 n_1 n_2 k\widetilde{\nu}_1.
\end{aligned}$$

Combining above results, we have

$$\begin{aligned}
\mathbb{E}\|\widehat{Z} - Z^*\|_F^2 \le \Bigg(&(1 - s\eta_4 n_2\widetilde{\nu}_k)\|\widetilde{Z} - Z^*\|_F + \frac{1}{4}\eta_4\sqrt{n_2\widetilde{\nu}_1}\Big\|\delta^* + Z^*W^{*\top}\\
&-\widetilde{\delta} - Z^*\widetilde{W}^\top\Big\|_F\Bigg)^2 + \frac{1}{4}\eta_4^2 n_1 n_2 k\widetilde{\nu}_1.
\end{aligned}$$

If $\eta_4 = 0$, then $\mathbb{E}\|\widehat{Z} - Z^*\|_F^2 = \|\widetilde{Z} - Z^*\|_F^2$. To guarantee $\mathbb{E}\|\widehat{Z} - Z^*\|_F^2 < \|\widetilde{Z} - Z^*\|_F^2$, a sufficient condition is

$$\begin{aligned}
\|\widetilde{Z} - Z^*\|_F^2 \ge \Bigg(&(1 - s\eta_4 n_2\widetilde{\nu}_k)\|\widetilde{Z} - Z^*\|_F + \frac{1}{4}\eta_4\sqrt{n_2\widetilde{\nu}_1}\Big\|\delta^* + Z^*W^{*\top}\\
&-\widetilde{\delta} - Z^*\widetilde{W}^\top\Big\|_F\Bigg)^2 + \frac{1}{4}\eta_4^2 n_1 n_2 k\widetilde{\nu}_1.
\end{aligned}$$

That is,

$$\begin{aligned}
&-\eta_4^2\left[\left(sn_2\widetilde{\nu}_k\|\widetilde{Z} - Z^*\|_F - \frac{1}{4}\sqrt{n_2\widetilde{\nu}_1}\left\|\delta^* + Z^*W^{*\top} - \widetilde{\delta} - Z^*\widetilde{W}^\top\right\|_F\right)^2 + \frac{1}{4}n_1 n_2 k\widetilde{\nu}_1\right]\\
&-2\eta_4\left[\frac{1}{4}\sqrt{n_2\widetilde{\nu}_1}\|\widetilde{Z} - Z^*\|_F\left\|\delta^* + Z^*W^{*\top} - \widetilde{\delta} - Z^*\widetilde{W}^\top\right\|_F - sn_2\widetilde{\nu}_k\|\widetilde{Z} - Z^*\|_F^2\right] \ge 0.
\end{aligned}$$

Define

$$\Delta = \frac{2\left(sn_2\widetilde{\nu}_k\|\widetilde{Z} - Z^*\|_F^2 - \frac{1}{4}\sqrt{n_2\widetilde{\nu}_1}\|\widetilde{Z} - Z^*\|_F\left\|\delta^* + Z^*W^{*\top} - \widetilde{\delta} - Z^*\widetilde{W}^\top\right\|_F\right)}{\left(sn_2\widetilde{\nu}_k\|\widetilde{Z} - Z^*\|_F - \frac{1}{4}\sqrt{n_2\widetilde{\nu}_1}\left\|\delta^* + Z^*W^{*\top} - \widetilde{\delta} - Z^*\widetilde{W}^\top\right\|_F\right)^2 + \frac{1}{4}n_1 n_2 k\widetilde{\nu}_1}. \tag{S7}$$

Under condition (5) in Theorem 2, we have that $\Delta > 0$. Then we have, if $0 < \eta_4 < \Delta$, $\mathbb{E}\|\widehat{Z} - Z^*\|_F^2 < \|\widetilde{Z} - Z^*\|_F^2$. This implies that, if the step size η_4 is appropriately chosen, the estimation of Z will be improved by incorporating the two-mode network D.

Next, we investigate the order of Δ. Note that

$$\begin{aligned}
&\left\|\delta^* + Z^*W^{*\top} - \widetilde{\delta} - Z^*\widetilde{W}^\top\right\|_F^2\\
=&\left\|\beta^*1_{n_2}^\top + 1_{n_1}\gamma^{*\top} - \widetilde{\beta}1_{n_2}^\top - 1_{n_1}\widetilde{\gamma}^\top\right\|_F^2 + \left\|Z^*W^{*\top} - Z^*\widetilde{W}^\top\right\|_F^2\\
\le&\left\|\beta^*1_{n_2}^\top + 1_{n_1}\gamma^{*\top} - \widetilde{\beta}1_{n_2}^\top - 1_{n_1}\widetilde{\gamma}^\top\right\|_F^2 + \|Z^*\|_F^2\left\|W^* - \widetilde{W}\right\|_F^2,
\end{aligned}$$

whose order is $O(n_1+n_2)$ under Condition (C1). For any $i=1,\cdots,k$, using Weyl's inequality, we have

$$|\widetilde{v}_i - v_i| \le \left\|\widetilde{W}^\top\widetilde{W}/n_2 - W^\top W/n_2\right\|_2 \le \left\|\widetilde{W}^\top\widetilde{W}/n_2 - W^\top W/n_2\right\|_F$$
$$\le n_2^{-1}\left\|\widetilde{W}-W\right\|_F\left\|\widetilde{W}+W\right\|_F = O\left(\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right).$$

By Condition (C2), $\widetilde{v}_i$ $(i=1,\cdots,k)$ is of constant order $O(1)$. Accordingly, the order of $\Delta = O\left\{\left[\frac{1}{n_1}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\right]^{1/2}\right\}$.

Appendix D. The properties of reasonable estimates

Lemma 3. *For any $\Theta,\widetilde{\Theta}\in\mathcal{F}$, the following equalities hold:*

$$\|\Theta^A - \widetilde{\Theta}^A\|_F^2 = \|\alpha 1_{n_1}^\top + 1_{n_1}\alpha^\top - \widetilde{\alpha}1_{n_1}^\top - 1_{n_1}\widetilde{\alpha}^\top\|_F^2 + \|ZZ^\top - \widetilde{Z}\widetilde{Z}^\top\|_F^2,$$
$$\|\Theta^D - \widetilde{\Theta}^D\|_F^2 = \|\beta 1_{n_2}^\top + 1_{n_1}\gamma^\top - \widetilde{\beta}1_{n_2}^\top - 1_{n_1}\widetilde{\gamma}^\top\|_F^2 + \|ZW^\top - \widetilde{Z}\widetilde{W}^\top\|_F^2.$$

Proof. Since the proof is similar, we only show the proof of the second equality.

$$\begin{aligned}\|\Theta^D - \widetilde{\Theta}^D\|_F^2 =& \operatorname{tr}\left[\left(\beta 1_{n_2}^\top + 1_{n_1}\gamma^\top - \widetilde{\beta}1_{n_2}^\top - 1_{n_1}\widetilde{\gamma}^\top + ZW^\top - \widetilde{Z}\widetilde{W}^\top\right)^\top\right.\\
&\left.\left(\beta 1_{n_2}^\top + 1_{n_1}\gamma^\top - \widetilde{\beta}1_{n_2}^\top - 1_{n_1}\widetilde{\gamma}^\top + ZW^\top - \widetilde{Z}\widetilde{W}^\top\right)\right]\\
=&\|\beta 1_{n_2}^\top + 1_{n_1}\gamma^\top - \widetilde{\beta}1_{n_2}^\top - 1_{n_1}\widetilde{\gamma}^\top\|_F^2 + \|ZW^\top - \widetilde{Z}\widetilde{W}^\top\|_F^2\\
&+2\operatorname{tr}\left[\left(\beta 1_{n_2}^\top + 1_{n_1}\gamma^\top - \widetilde{\beta}1_{n_2}^\top - 1_{n_1}\widetilde{\gamma}^\top\right)^\top\left(ZW^\top - \widetilde{Z}\widetilde{W}^\top\right)\right].\end{aligned}$$

Since $ZW^\top 1_{n_2} = \widetilde{Z}\widetilde{W}^\top 1_{n_2} = 0$ and $1_{n_1}^\top ZW^\top = 1_{n_1}^\top\widetilde{Z}\widetilde{W}^\top = 0$, the third term equals to 0. This completes the proof.

Lemma 4. *The estimated $\widetilde{Z}$ and $\widetilde{\alpha}$ from Algorithm 1 in Ma et al. (2020) satisfy the conditions $\|\widetilde{\alpha}1_{n_1}^\top - \alpha^* 1_{n_1}^\top\|_F^2 = O_p(n_1)$, $\|\widetilde{Z}-Z^*\|_F^2 = O_p(1)$ and the gradient of the negative log-likelihood function L_A with respect to Z when $Z=\widetilde{Z}$ will vanish.*

Proof. This lemma is from lemma 3 and 4 of Zhang et al. (2022b).

Lemma 5. *Suppose Conditions (C1)–(C2) hold. Assume D is generated from model (1) and we have $\widetilde{Z}$ satisfying $\|\widetilde{Z}-Z^*\|_F^2 = O_p(1)$ and $J_1\widetilde{Z} = \widetilde{Z}$. If we minimize the negative log-likelihood function L_D given $\widetilde{Z}$, then we have $\|\widetilde{\beta}1_{n_2}^\top + 1_{n_1}\widetilde{\gamma}^\top - \beta^* 1_{n_2}^\top - 1_{n_1}\gamma^{*\top}\|_F^2 = O_p(n_1+n_2)$ and $\|\widetilde{W}-W^*\|_F^2 = O_p\{(n_1+n_2)/n_1\}$.*

Proof. Define $\bar{\Theta}^D = \beta^* 1_{n_2}^\top + 1_{n_1}\gamma^{*\top} + \widetilde{Z}W^{*\top}$ and $\widetilde{\Theta}^D = \widetilde{\beta}1_{n_2}^\top + 1_{n_1}\widetilde{\gamma}^\top + \widetilde{Z}\widetilde{W}^\top$. Note that $(\widetilde{\beta},\widetilde{\gamma},\widetilde{W}) = \operatorname{argmin}_{J_2 W = W} L_D(\beta,\gamma,W,\widetilde{Z})$. Therefore, we have

$$\begin{aligned}0 \le& L_D(\beta^*,\gamma^*,W^*,\widetilde{Z}) - L_D(\widetilde{\beta},\widetilde{\gamma},\widetilde{W},\widetilde{Z})\\
=&\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\left\{D_{ij}\widetilde{\Theta}_{ij}^D - D_{ij}\bar{\Theta}_{ij}^D + b\left(\bar{\Theta}_{ij}^D\right) - b\left(\widetilde{\Theta}_{ij}^D\right)\right\}\\
=&\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\left\{\left(\widetilde{\Theta}_{ij}^D - \bar{\Theta}_{ij}^D\right)\left(D_{ij} - b'\left(\bar{\Theta}_{ij}^D\right)\right) + b\left(\bar{\Theta}_{ij}^D\right) - b\left(\widetilde{\Theta}_{ij}^D\right) + b'\left(\bar{\Theta}_{ij}^D\right)\left(\widetilde{\Theta}_{ij}^D - \bar{\Theta}_{ij}^D\right)\right\}\\
=&\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\left\{\left(\widetilde{\Theta}_{ij}^D - \bar{\Theta}_{ij}^D\right)\left(D_{ij} - b'\left(\bar{\Theta}_{ij}^D\right)\right) - \frac{1}{2}b''\left(\bar{\bar{\Theta}}_{ij}^D\right)\left(\widetilde{\Theta}_{ij}^D - \bar{\Theta}_{ij}^D\right)^2\right\},\end{aligned}$$

where $\bar{\bar{\Theta}}_{ij}^D = \lambda_{3,ij}\bar{\Theta}_{ij}^D + (1-\lambda_{3,ij})\widetilde{\Theta}_{ij}^D$ for some $\lambda_{3,ij}\in(0,1)$. This implies that

$$\frac{1}{2}\min_{|v|<M_D} b''(v)\|\widetilde{\Theta}^D - \bar{\Theta}^D\|_F^2 \le \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\left\{\left(\widetilde{\Theta}_{ij}^D - \bar{\Theta}_{ij}^D\right)\left(D_{ij} - b'\left(\bar{\Theta}_{ij}^D\right)\right)\right\}$$
$$=\langle\bar{E},\widetilde{\Theta}^D - \bar{\Theta}^D\rangle \le \|\bar{E}\|_2\sqrt{\operatorname{rank}(\widetilde{\Theta}^D - \bar{\Theta}^D)}\|\widetilde{\Theta}^D - \bar{\Theta}^D\|_F,$$

where $\bar{E} = D - b'\left(\bar{\Theta}^D\right)$. Therefore, we have

$$\|\widetilde{\Theta}^D - \bar{\Theta}^D\|_F \le \frac{2\|\bar{E}\|_2\sqrt{\text{rank}(\widetilde{\Theta}^D - \bar{\Theta}^D)}}{\min_{|v|<M_D} b''(v)}.$$

Note that $\text{rank}(\widetilde{\Theta}^D - \bar{\Theta}^D) \le 2(k+2)$. Define $\Theta^{*D} = \beta^* 1_{n_2}^\top + 1_{n_1}\gamma^{*\top} + Z^* W^{*\top}$. Then

$$\begin{aligned}\|\bar{E}\|_2 = \left\|D - b'\left(\bar{\Theta}^D\right)\right\|_2 &\le \left\|D - b'\left(\Theta^{*D}\right)\right\|_2 + \left\|b'\left(\Theta^{*D}\right) - b'\left(\bar{\Theta}^D\right)\right\|_2 \\ &\le \left\|D - b'\left(\Theta^{*D}\right)\right\|_2 + \frac{1}{4}\left\|Z^* W^{*\top} - \widetilde{Z} W^{*\top}\right\|_2.\end{aligned}$$

Using Lemma 2, with probability at least $1 - e^{-c(n_1+n_2)}$, we have

$$\left\|D - b'\left(\Theta^{*D}\right)\right\|_2 \le C\sqrt{n_1 + n_2},$$

which implies that $\left\|D - b'\left(\Theta^{*D}\right)\right\|_2 = O_p(\sqrt{n_1+n_2})$. For the second term, it is bounded by $\frac{1}{4}\|W^*\|_F\|Z^* - \widetilde{Z}\|_F = O_p(\sqrt{n_2})$. If k and M_D are fixed, then we have $\|\widetilde{\Theta}^D - \bar{\Theta}^D\|_F^2 = O_p(n_1 + n_2)$. By Lemma 5, we have $\|\widetilde{\Theta}^D - \bar{\Theta}^D\|_F^2 \ge \|\beta^* 1_{n_2}^\top + 1_{n_1}\gamma^{*\top} - \widetilde{\beta} 1_{n_2}^\top - 1_{n_1}\widetilde{\gamma}^\top\|_F^2$ and

$$\|\widetilde{\Theta}^D - \bar{\Theta}^D\|_F^2 \ge \|\widetilde{Z}\widetilde{W}^\top - \widetilde{Z} W^{*\top}\|_F^2 \ge n_1 \widetilde{\mu}_k \|\widetilde{W} - W^*\|_F^2,$$

where $\widetilde{\mu}_i$ $(i = 1, \cdots, k)$ is the i-th largest eigenvalue of $\widetilde{Z}^\top \widetilde{Z}/n_1$. By Weyl's inequality, we can easily show that $\widetilde{\mu}_i$ $(i = 1, \cdots, k)$ is of constant order $O(1)$. To guarantee above inequality hold, we have $\|\widetilde{W} - W^*\|_F^2 = O\{(n_1+n_2)/n_1\}$. This completes the proof.

Appendix E. Details of Algorithm 1

To optimize the constrained problem (4), the projected gradient descent algorithm is applied. In each iteration, we first update each parameter along the gradient direction and then project the updated estimates to satisfy the identification conditions. Next we will discuss these two steps in detail.

Step 1 (Descent). To update each parameter along the gradient direction, the first-order derivatives of each parameter are required. After some tedious calculations, we have

$$\begin{aligned}\nabla_\alpha L &= -2\left(A - \sigma(\Theta^A)\right) 1_{n_1}, \\ \nabla_\beta L &= -\left(D - \sigma(\Theta^D)\right) 1_{n_2}, \\ \nabla_\gamma L &= -\left(D - \sigma(\Theta^D)\right)^\top 1_{n_1}, \\ \nabla_Z L &= -2\left(A - \sigma(\Theta^A)\right) Z - \left(D - \sigma(\Theta^D)\right) W, \\ \nabla_W L &= -\left(D - \sigma(\Theta^D)\right)^\top Z,\end{aligned}$$

where $\Theta^A = \alpha 1_{n_1}^\top + 1_{n_1}\alpha^\top + ZZ^\top$ and $\Theta^D = \beta 1_{n_2}^\top + 1_{n_1}\gamma^\top + ZW^\top$. Accordingly, each parameter is updated along the gradient direction with a small step size.

Step 2 (Projection). Define the constraint set

$$\mathcal{C}_Z = \{Z \in \mathbb{R}^{n_1 \times k}, J_1 Z = Z\},$$

$$\mathcal{C}_W = \{W \in \mathbb{R}^{n_2 \times k}, J_2 W = W\},$$

$$\mathcal{C}_\beta = \{\beta \in \mathbb{R}^{n_1}, \beta^\top 1_{n_1} = 0\}.$$

The projection steps for parameters Z, W and β are straightforward and have the following closed form expressions,

$$\mathcal{P}_{\mathcal{C}_Z}(Z) = J_1 Z, \mathcal{P}_{\mathcal{C}_W}(W) = J_2 W, \mathcal{P}_{\mathcal{C}_\beta}(\beta) = \beta - \beta^\top 1_{n_1}.$$

By applying the above formula to update Z, W and β, the renewed estimates satisfy the identification conditions.

Appendix F. Additional simulation studies

This section presents two subsections. In Section F1, we first present the simulation results for Cases 1–4 when the latent space dimension is $k = 4$. Section F2 introduces additional simulation studies to investigate whether cross validation can select the latent space dimension correctly.

Table S1
The means and standard deviations of the relative estimation errors Δ_α, Δ_Z and Δ_Θ^A for Case 1 when $d=4$ and networks A and D are dense.

n_1	Measure	Δ_α			Δ_Z			Δ_Θ^A			Time		
		gTLSM	OLSM	jOLSM	gTLSM	OLSM	jOLSM	gTLSM	OLSM	jOLSM	gTLSM	OLSM	jOLSM
200	Mean	0.050	0.053	0.233	0.133	0.375	0.269	0.058	0.132	0.217	33.942	10.013	96.525
	SD	0.005	0.005	0.014	0.009	0.038	0.023	0.004	0.012	0.011	7.633	3.235	26.700
300	Mean	0.034	0.035	0.236	0.084	0.224	0.181	0.038	0.080	0.196	68.993	17.991	223.422
	SD	0.003	0.003	0.011	0.005	0.020	0.017	0.002	0.006	0.008	16.453	6.498	50.438
400	Mean	0.025	0.026	0.244	0.061	0.161	0.144	0.027	0.058	0.191	109.615	27.247	465.358
	SD	0.002	0.002	0.010	0.003	0.012	0.025	0.001	0.004	0.009	9.673	5.876	77.200
500	Mean	0.020	0.021	0.247	0.048	0.125	0.130	0.022	0.045	0.189	171.077	39.101	832.483
	SD	0.001	0.002	0.009	0.002	0.005	0.039	0.001	0.002	0.014	16.848	4.830	118.496
600	Mean	0.017	0.018	0.248	0.039	0.103	0.116	0.018	0.037	0.187	257.306	56.772	1360.265
	SD	0.001	0.001	0.008	0.001	0.003	0.036	0.001	0.001	0.011	19.260	6.145	155.262

Table S2
The means and standard deviations of the relative estimation errors Δ_α, Δ_Z and Δ_Θ^A for Case 2 when $d=4$ and networks A and D are sparse.

n_1	Measure	Δ_α			Δ_Z			Δ_Θ^A			Time		
		gTLSM	OLSM	jOLSM	gTLSM	OLSM	jOLSM	gTLSM	OLSM	jOLSM	gTLSM	OLSM	jOLSM
200	Mean	0.038	0.042	0.105	0.297	0.766	0.401	0.050	0.097	0.103	34.244	10.740	89.912
	SD	0.003	0.003	0.006	0.022	0.072	0.059	0.003	0.006	0.007	7.271	3.854	17.497
300	Mean	0.036	0.041	0.106	0.193	0.484	0.244	0.040	0.071	0.090	73.582	21.738	228.943
	SD	0.002	0.002	0.006	0.011	0.052	0.039	0.002	0.005	0.005	11.132	6.435	42.111
400	Mean	0.035	0.040	0.107	0.150	0.369	0.180	0.036	0.060	0.085	134.718	37.373	471.169
	SD	0.002	0.002	0.005	0.007	0.042	0.027	0.001	0.004	0.004	27.351	9.483	79.127
500	Mean	0.034	0.040	0.108	0.126	0.294	0.147	0.034	0.053	0.083	211.619	58.159	811.782
	SD	0.001	0.002	0.004	0.005	0.030	0.024	0.001	0.003	0.004	30.342	14.199	123.078
600	Mean	0.034	0.040	0.107	0.110	0.248	0.137	0.032	0.049	0.082	306.241	81.523	1276.058
	SD	0.001	0.002	0.004	0.004	0.025	0.035	0.001	0.002	0.005	38.771	15.431	197.858

Table S3
The means and standard deviations of the relative estimation errors Δ_α, Δ_Z and Δ_Θ^A for Case 3 when $d=4$, $n_1=200$, and networks A and D are dense.

n_2	Measure	Δ_α			Δ_Z			Δ_Θ^A			Time		
		gTLSM	OLSM	jOLSM	gTLSM	OLSM	jOLSM	gTLSM	OLSM	jOLSM	gTLSM	OLSM	jOLSM
100	Mean	0.054	0.053	0.060	0.257	0.364	0.249	0.097	0.129	0.098	9.295	4.736	10.222
	SD	0.006	0.006	0.005	0.014	0.030	0.014	0.005	0.010	0.005	0.685	0.960	1.517
200	Mean	0.052	0.053	0.103	0.198	0.374	0.230	0.079	0.132	0.119	11.907	4.712	19.267
	SD	0.006	0.006	0.008	0.013	0.041	0.017	0.005	0.013	0.007	0.886	0.883	2.487
300	Mean	0.051	0.053	0.164	0.159	0.371	0.239	0.067	0.131	0.162	14.438	4.810	33.986
	SD	0.006	0.006	0.011	0.011	0.037	0.018	0.005	0.012	0.008	0.805	0.853	4.357
400	Mean	0.050	0.053	0.233	0.133	0.375	0.269	0.058	0.132	0.217	16.897	4.898	54.064
	SD	0.005	0.005	0.014	0.009	0.038	0.023	0.004	0.012	0.011	1.212	0.910	6.889
500	Mean	0.050	0.053	0.303	0.116	0.372	0.316	0.053	0.131	0.280	19.262	4.854	82.433
	SD	0.005	0.006	0.019	0.009	0.040	0.026	0.004	0.012	0.014	1.303	0.867	14.117

F1. Simulation results with $k=4$

We first present the simulation results for Cases 1–4 when the latent space dimension is $k=4$. The conclusions based on Tables S1–S4 are similar to those of the main paper.

F2. Selection results for latent space dimension

Following Zhang et al. (2022b), we select the latent space dimension k through cross validation. Specifically, we randomly remove some elements of the two adjacency matrix A and D, then use the remaining elements to fit the proposed model and predict those removed elements. Repeat above procedure several times and choose k which yields the minimum prediction error. To examine the robustness of cross validation on the selection of latent space dimensions, we consider different settings. We fix $n_1=200, 400$ and vary n_2 from 200 to 600. Three different densities of the network are considered. Set the true latent space dimension $k=2$ or 4. Other settings are the same as those in Case 1. Replicate the experiment 100 times and calculate the percentage of correct selection.

Table S4
The means and standard deviations of the relative estimation errors Δ_α, Δ_Z and Δ^A_Θ for Case 4 when $d=4$, $n_1=200$, network A is dense and network D is sparse.

n_2	Measure	Δ_α			Δ_Z			Δ^A_Θ			Time		
		gTLSM	OLSM	jOLSM	gTLSM	OLSM	jOLSM	gTLSM	OLSM	jOLSM	gTLSM	OLSM	jOLSM
100	Mean	0.055	0.053	0.060	0.323	0.364	0.352	0.117	0.129	0.129	12.732	6.370	14.977
	SD	0.007	0.006	0.005	0.022	0.030	0.038	0.007	0.010	0.011	1.706	1.615	2.929
200	Mean	0.055	0.053	0.085	0.292	0.374	0.388	0.108	0.132	0.148	17.230	6.576	30.581
	SD	0.006	0.006	0.009	0.017	0.041	0.043	0.007	0.013	0.011	2.207	1.552	7.387
300	Mean	0.054	0.053	0.113	0.264	0.371	0.444	0.099	0.131	0.169	21.797	6.821	49.301
	SD	0.006	0.006	0.012	0.023	0.037	0.059	0.008	0.012	0.011	3.123	1.631	8.026
400	Mean	0.053	0.053	0.137	0.237	0.375	0.523	0.091	0.132	0.191	26.358	6.724	78.441
	SD	0.005	0.005	0.014	0.016	0.038	0.062	0.006	0.012	0.014	3.595	1.742	13.018
500	Mean	0.053	0.053	0.159	0.218	0.371	0.600	0.085	0.131	0.214	31.541	6.521	111.294
	SD	0.005	0.005	0.018	0.019	0.040	0.060	0.006	0.012	0.022	4.113	1.597	17.506

Table S5
Detailed selection results for latent space dimension.

n_1	n_2	density	$d=2$			$d=4$		
			Under	Correct	Over	Under	Correct	Over
200	200	0.053	0.03	0.97	0	1	0	0
		0.110	0	1	0	0.42	0.58	0
		0.220	0	1	0	0.01	0.99	0
	400	0.053	0.02	0.98	0	0.94	0.06	0
		0.110	0	1	0	0.09	0.91	0
		0.220	0	1	0	0	1	0
	600	0.053	0.02	0.98	0	0.81	0.19	0
		0.110	0	1	0	0.04	0.96	0
		0.220	0	1	0	0	1	0
400	200	0.053	0	1	0	0.33	0.67	0
		0.110	0	1	0	0	1	0
		0.220	0	1	0	0	1	0
	400	0.053	0	1	0	0.19	0.81	0
		0.110	0	1	0	0	1	0
		0.220	0	1	0	0	1	0
	600	0.053	0	1	0	0.02	0.98	0
		0.110	0	1	0	0	1	0
		0.220	0	1	0	0	1	0

The selection results are shown in Table S5. From the table, we conclude that cross validation performs well and the percentage of correct selection almost reaches 1 as the number of nodes increases and the density of the network increases.

References

Bader, G.D., Hogue, C.W., 2003. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinform. 4, 1–27.
Barbillon, P., Donnet, S., Lazega, E., Bar-Hen, A., 2017. Stochastic block models for multiplex networks: an application to a multilevel network of researchers. J. R. Stat. Soc., Ser. A, Stat. Soc. 180, 295–314.
Billio, M., Getmansky, M., Lo, A.W., Pelizzon, L., 2012. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. J. Financ. Econ. 104, 535–559.
Chabert-Liddell, S.C., Barbillon, P., Donnet, S., Lazega, E., 2021. A stochastic block model approach for the analysis of multilevel networks: an application to the sociology of organizations. Comput. Stat. Data Anal. 158, 107179.
Chang, X., Huang, D., Wang, H., 2019. A popularity-scaled latent space model for large-scale directed social network. Stat. Sin. 29, 1277–1299.
D'Angelo, S., Murphy, T.B., Alfò, M., 2019. Latent space modelling of multidimensional networks with application to the exchange of votes in eurovision song contest. Ann. Appl. Stat. 13, 900–930.
Feng, L., Zhou, C., Zhao, Q., 2019. A spectral method to find communities in bipartite networks. Physica A, Stat. Mech. Appl. 513, 424–437.
Gollini, I., Murphy, T.B., 2016. Joint modeling of multiple network views. J. Comput. Graph. Stat. 25, 246–265.
Härdle, W.K., Wang, W., Yu, L., 2016. Tenet: tail-event driven network risk. J. Econom. 192, 499–513.
Hoff, P.D., 2003. Random effects models for network data. In: Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers. National Academy Press, pp. 303–322.
Hoff, P.D., Raftery, A.E., Handcock, M.S., 2002. Latent space approaches to social network analysis. J. Am. Stat. Assoc. 97, 1090–1098.
Huang, D., Wang, F., Zhu, X., Wang, H., 2020a. Two-mode network autoregressive model for large-scale networks. J. Econom. 216, 203–219.
Huang, W., Liu, Y., Chen, Y., 2020b. Mixed membership stochastic blockmodels for heterogeneous networks. Bayesian Anal. 15, 711–736.
Kossinets, G., Watts, D.J., 2006. Empirical analysis of an evolving social network. Science 311, 88–90.
Krivitsky, P.N., Handcock, M.S., Raftery, A.E., Hoff, P.D., 2009. Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. Soc. Netw. 31, 204–213.

Lam, C., Yao, Q., 2012. Factor modeling for high-dimensional time series: inference for the number of factors. Ann. Stat. 40, 694–726.
Lewis, K., Gonzalez, M., Kaufman, J., 2012. Social selection and peer influence in an online social network. Proc. Natl. Acad. Sci. 109, 68–72.
Ma, Z., Ma, Z., Yuan, H., 2020. Universal latent space model fitting for large networks with edge covariates. J. Mach. Learn. Res. 21, 1–67.
MacDonald, P.W., Levina, E., Zhu, J., 2022. Latent space models for multiplex networks with shared structure. Biometrika 109, 683–706.
Nepusz, T., Yu, H., Paccanaro, A., 2012. Detecting overlapping protein complexes in protein-protein interaction networks. Nat. Methods 9, 471–472.
Ren, M., Zhang, S., Wang, J., 2023. Consistent estimation of the number of communities via regularized network embedding. Biometrics 79, 2404–2416.
Salter-Townshend, M., McCormick, T.H., 2017. Latent space models for multiview network data. Ann. Appl. Stat. 11, 1217–1244.
Sengupta, S., Chen, Y., 2015. Spectral clustering in heterogeneous networks. Stat. Sin. 25, 1081–1106.
Sewell, D.K., Chen, Y., 2015. Latent space models for dynamic networks. J. Am. Stat. Assoc. 110, 1646–1657.
Sewell, D.K., Chen, Y., 2017. Latent space approaches to community detection in dynamic networks. Bayesian Anal. 12, 351–377.
Slaughter, A.J., Koehly, L.M., 2016. Multilevel models for social networks: hierarchical Bayesian approaches to exponential random graph modeling. Soc. Netw. 44, 334–345.
Tomioka, R., Suzuki, T., 2014. Spectral norm of random tensors. ArXiv preprint. arXiv:1407.1870.
Wang, P., Robins, G., Pattison, P., Lazega, E., 2013. Exponential random graph models for multilevel networks. Soc. Netw. 35, 96–115.
Wu, Y., Lan, W., Zou, T., Tsai, C.L., 2022. Inward and outward network influence analysis. J. Bus. Econ. Stat. 40, 1617–1628.
Zhang, J., Chen, Y., 2020. Modularity based community detection in heterogeneous networks. Stat. Sin. 30, 601–629.
Zhang, J., He, X., Wang, J., 2022a. Directed community detection with network embedding. J. Am. Stat. Assoc. 117, 1809–1819.
Zhang, J., Sun, W.W., Li, L., 2020a. Mixed-effect time-varying network model and application in brain connectivity analysis. J. Am. Stat. Assoc. 115, 2022–2036.
Zhang, X., Xu, G., Zhu, J., 2022b. Joint latent space models for network data with high-dimensional node variables. Biometrika 109, 707–720.
Zhang, X., Xue, S., Zhu, J., 2020b. A flexible latent space model for multilayer networks. In: International Conference on Machine Learning. PMLR, pp. 11288–11297.
Zhen, Y., Wang, J., 2023. Community detection in general hypergraph via graph embedding. J. Am. Stat. Assoc. 118, 1620–1629.