The Annals of Statistics
2025, Vol. 53, No. 4, 1406–1430
https://doi.org/10.1214/25-AOS2506

SEMIPARAMETRIC MODELING AND ANALYSIS FOR LONGITUDINAL NETWORK DATA

BY YINQIU HE[1,a], JIAJIN SUN[2,b], YUANG TIAN[3,c], ZHILIANG YING[4,d] AND YANG FENG[5,e]

[1]*Department of Statistics, University of Wisconsin-Madison,* [a]*yinqiu.he@wisc.edu*
[2]*Department of Statistics, Florida State University,* [b]*jsun5@fsu.edu*
[3]*Shanghai Center for Mathematical Sciences, Fudan University,* [c]*yatian20@fudan.edu.cn*
[4]*Department of Statistics, Columbia University,* [d]*zying@stat.columbia.edu*
[5]*Department of Biostatistics, School of Global Public Health, New York University,* [e]*yang.feng@nyu.edu*

We introduce a semiparametric latent space model for analyzing longitudinal network data. The model consists of a static latent space component and a time-varying node-specific baseline component. We develop a semiparametric efficient score equation for the latent space parameter by adjusting for the baseline nuisance component. Estimation is accomplished through a one-step update estimator and an appropriately penalized maximum likelihood estimator. We derive oracle error bounds for the two estimators and address identifiability concerns from a quotient manifold perspective. Our approach is demonstrated using the New York Citi Bike Dataset.

1. Introduction. Recent years have seen an increased availability of time-varying interaction/network data (Butts (2008), Linderman and Adams (2014), Holme (2015)), with examples such as email exchange history of coworkers (Klimt and Yang (2004)) and transport records among bike-sharing stations (CitiBike (2019)). This type of data includes not only counts of pairwise interaction events but also the timestamps of these interactions.

nonnetwork count data have been studied extensively in the survival analysis literature. Andersen and Gill (1982) proposed a Poisson-type intensity-based regression model. For extensions to handle non-Poisson count data, see Pepe and Cai (1993), Cook and Lawless (2007), Lin et al. (2000), Sun and Zhao (2013).

For cross-sectional single network data, an important development is the latent space model, in which each node is represented by a latent vector and the relationship between two nodes is quantified through their inner product (Hoff, Raftery and Handcock (2002)). The latent vectors may be treated as random (Hoff (2003, 2005), Bickel et al. (2013)) or fixed (Athreya et al. (2018), Ma, Ma and Yuan (2020)).

Motivated by various scientific applications, multiple-network models and longitudinal network models have also been developed; see, for example, Zhang, Sun and Li (2020) for application in neuroscience and Butts (2008) for application in social science. Similar to the single-network models, latent space modeling in the context of multiple-network settings has been examined from different perspectives (Sewell and Chen (2015), Nielsen and Witten (2018), Matias and Robin (2014)). One line of research views the latent vectors as random (Hoff (2011), Salter-Townshend and McCormick (2017)), and another line considers fixed latent vectors (Jones and Rubin-Delanchy (2020), Levin et al. (2017)). The latter often leads to frequentist approaches, in which it is of common interest to study the estimation of

Received July 2024; revised December 2024.
MSC2020 subject classifications. Primary 62H12, 05C82; secondary 91D30, 62F12.
Key words and phrases. Network, Poisson counting process, latent space model, low rank, semiparametric, heterogeneity.

shared structure across multiple networks, and in particular, how the information accumulation across multiple networks improves the fitting quality (Arroyo et al. (2021), Zheng and Tang (2022), MacDonald, Levina and Zhu (2022), Zhang, Xue and Zhu (2020)).

In this paper, we develop a semiparametric modeling framework for longitudinal networks with interaction event counts. The model comprises a static latent space component, which accounts for inherent node interactions, and a baseline component that accommodates heterogeneity across different time points and nodes. We introduce two semiparametric estimation methods. The first method forms a generalized semiparametric one-step estimator by solving a local linear approximation of the efficient score equation. We demonstrate that the estimator automatically eliminates the identifiability issue through an interpretation in the quotient manifold induced by orthogonal transformation equivalence classes. The second method optimizes a convex surrogate loss function based on the log-likelihood, relaxing the nonconvex rank constraint of the latent space component. We establish that both methods achieve, up to a logarithmic factor, the oracle estimation error rates for the latent space component.

The rest of the paper is organized as follows. Section 2 introduces the semiparametric latent space model. Section 3 introduces the one-step estimator, provides its connection to the quotient manifold theory, and establishes the error bound for the estimator. Section 4 introduces the penalized maximum likelihood estimator and establishes its error bound. Simulation results are presented in Section 5. Section 6 applies our framework to analyze the New York Citi Bike Dataset (CitiBike (2019)). Section 7 concludes the main body of the paper with some discussions. All the technical derivations are deferred to the Supplementary Material (He et al. (2025)).

2. Semiparametric Poisson latent space model.

2.1. *Notation and model specification.* We consider longitudinal pairwise interaction counts of n subjects (nodes) over T discrete time points. Specifically, for a time point $t \in \{1, \dots, T\}$ and nodes $i, j \in \{1, \dots, n\}$, $A_{t,ij}$ denotes the number of i-j interactions at the time point t. We propose a Poisson-based latent space model

$$
\begin{aligned}
A_{t,ij} = A_{t,ji} &\sim \text{Poisson}\big\{\mathbb{E}(A_{t,ij}|z,\alpha)\big\}, \quad \text{independently with} \\
\mathbb{E}(A_{t,ij}|z,\alpha) &= \exp\big(\alpha_{it} + \alpha_{jt} + \langle z_i, z_j\rangle\big),
\end{aligned} \tag{1}
$$

which naturally adopts the exponential link function $\exp(\cdot)$ to model the event counts. For any two nodes i and j, their interaction effect is modeled through the inner product of two corresponding latent vectors $\langle z_i, z_j\rangle = z_i^\top z_j$, similarly to the inner product model of a single network (Ma, Ma and Yuan (2020)). The latent vectors z_i's do not change with respect to the time point t and represent the shared latent structures across T heterogeneous networks. For example, z_i's can encode the time-invariant geographic information in multiple transportation networks. At a given time point t, when α_{it} increases and all the other parameters are fixed, edges connecting the node i tend to have higher numbers of counts at the time point t, indicating higher baseline activity levels. Therefore, α_{it}'s model the degree heterogeneity across different nodes $i \in \{1, \dots, n\}$ and time points $t \in \{1, \dots, T\}$, and are called baseline degree heterogeneity parameters of nodes and time. In the hourly bike-sharing networks, α_{it}'s can represent distinct baseline activity levels across different stations and hours.

Model specification (1) may be expressed in vector-matrix notation as

$$
\mathbb{E}(\mathbf{A}_t|Z,\alpha) = \exp\big(\alpha_t 1_n^\top + 1_n \alpha_t^\top + ZZ^\top\big), \tag{2}
$$

where $\alpha_t = (\alpha_{1t}, \dots, \alpha_{nt})^\top \in \mathbb{R}^{n\times 1}$, $Z = (z_1, \dots, z_n)^\top \in \mathbb{R}^{n\times k}$, $1_n = (1, \dots, 1)^\top \in \mathbb{R}^{n\times 1}$, and $\exp(\cdot)$ is the elementwise exponential operation. Throughout this paper, we consider the

asymptotic regime in which the number of nodes n and the number of time periods T increase to infinity while the dimension of the latent space k is fixed. Thus, $\alpha_t 1_n^\top + 1_n \alpha_t^\top + ZZ^\top$ is a low-rank matrix. To ensure identifiability, we assume that column means of Z are zero, that is, $1_n^\top Z/n = 0$. This centering assumption is analogous to the classical two-way analysis of variance (ANOVA) modeling with interaction (Scheffé (1999)). Additionally, since $ZZ^\top = ZQQ^\top Z^\top$ for any $Q \in \mathcal{O}(k)$, where $\mathcal{O}(k) = \{Q \in \mathbb{R}^{k\times k} : QQ^\top = \mathrm{I}_k\}$, Z is identifiable up to a common orthogonal transformation of its rows.

Let $\lambda_{ij,t} = \exp(\alpha_{it} + \alpha_{jt})$. Then $\mathbb{E}(A_{t,ij}) = \lambda_{ij,t} \times \exp(\langle z_i, z_j\rangle)$. This form resembles the multiplicative counting process modeling in the analysis of recurrent event times (Cook and Lawless (2007), Sun and Zhao (2013)). In particular, the $\lambda_{ij,t}$ corresponds to the baseline rate/intensity function changing over time while the z_i corresponds to the time-invariant parameters of interest. Note that the number of baseline (nuisance) parameters is nT, and the size of the shared latent positions Z or number of parameters of interest is nk; the former increases with both n and T, whereas the latter does not increase with T. Thus, we may be able to estimate Z more accurately than the nuisance parameter.

Some notations used in this paper are summarized as follows. For a matrix $X = [x_{ij}] \in \mathbb{R}^{n\times n}$, $\mathrm{tr}(X) = \sum_{i=1}^n x_{ii}$ stands for its trace and $\sigma_{\min}(X)$ represents its minimum eigenvalue. For $X, Y \in \mathbb{R}^{n\times m}$, $\langle X, Y\rangle = \mathrm{tr}(X^\top Y)$. If $m \geq n$, for any matrix X with the singular value decomposition $X = \sum_{i=1}^n s_i u_i v_i^\top$, we let $\|X\|_* = \sum_{i=1}^n s_i$, $\|X\|_{\mathrm{F}} = \sqrt{\sum_{i=1}^n s_i^2}$, and $\|X\|_{\mathrm{op}} = \max_{i=1,\ldots,n} s_i$ stand for the nuclear norm, the Frobenius norm, and the operator norm of the matrix, respectively. For a vector $x \in \mathbb{R}^n$, $\|x\|_2 = \sqrt{x^\top x}$.

For two sequences of real numbers $\{f_n\}$ and $\{h_n\}$, $f_n \lesssim h_n$ and $f_n = O(h_n)$ mean that $|f_n| \leq c_1|h_n|$ for a constant $c_1 > 0$; $f_n \asymp h_n$ means $c_2 h_n \leq f_n \leq c_1 h_n$ for some constants $c_1, c_2 > 0$; $f_n = o(h_n)$ and $f_n \ll h_n$ mean $\lim_{n\to\infty} f_n/h_n = 0$; $f_n \gg h_n$ means $\lim_{n\to\infty} h_n/f_n = 0$. For a sequence of random variables X_n and a sequence of real numbers f_n, write $X_n = O_p(f_n)$ if for any $\epsilon > 0$, there exists finite $M > 0$ such that $\sup_n \Pr(|X_n/f_n| > M) < \epsilon$ (X_n/f_n is stochastically bounded); write $X_n = o_p(f_n)$ if for any $\epsilon > 0$, $\lim_{n\to\infty} \Pr(|X_n/f_n| > \epsilon) = 0$ (X_n/f_n converges to 0 in probability).

2.2. *Semiparametric oracle error rate and technical challenges.* To gain insights into the best possible error rate for estimating high-dimensional parameters, let's consider a simpler setting of the regular exponential family with a p_n-dimensional natural parameter vector θ_n. Suppose we have n independent observations from this family. Portnoy (1988) showed that the MLE $\hat{\theta}_n$ satisfies $\|\hat{\theta}_n - \theta_n\|_2^2 = O_p(p_n/n)$, where p_n can increase with the sample size n. For our problem of estimating Z, if the nuisance parameters α's were known, then each k-dimensional latent vector z_i would be measured by nT independent edges: $\{A_{t,ij} : j = 1, \ldots, n; t = 1, \ldots, T\}$. The result of Portnoy (1988) indicates that the oracle estimation error rate of each z_i would be $k/(nT)$. When k is fixed, the aggregated estimation error of n latent vectors $Z = [z_1, \ldots, z_n]^\top$ is expected to be of the order of $O_p(n \times k/(nT)) = O_p(1/T)$, which is referred to as the oracle estimation error rate throughout this paper. Since the α's are unknown, the asymptotic theory for the classical semiparametric models (Bickel et al. (1993)) may be modified to yield efficient score equations (projections) so that the same error rate can be achieved.

To achieve the above oracle bounds, several technical challenges remain. First, the baseline parameters α_{it} in (1) not only characterize the time-specific heterogeneity over t but also represent node-specific heterogeneity over i. Due to this two-way heterogeneity, the estimation errors of α_{it}'s are intertwined with those of node-specific parameters z_i's in a complicated way. In particular, the partial likelihood (Andersen and Gill (1982)) cannot eliminate the baseline nuisance parameters. Second, the target parameter matrix Z is only identifiable up to an orthogonal transformation. Therefore, given a true matrix $Z^\star$ and an estimate $\hat{Z}$, the ordinary

Euclidean distance between the two matrices is not a proper measure for the estimation error of $\hat{Z}$. Indeed, it is more natural to consider their distance defined up to an orthogonal group transformation $\operatorname{dist}(\hat{Z}, Z^\star) = \min_{Q\in\mathcal{O}(k)} \|\hat{Z} - Z^\star Q\|_{\mathrm{F}}$. This distance metric indicates that the intrinsic geometry of Z is non-Euclidean. Thirdly, the estimation of Z is further complicated by the high-dimensionality of the parameters and nonlinearity of the transformation in (1). Unlike classical semiparametric problems, the dimension of the parameters of interest also grows with the sample size. Furthermore, the nonlinearity of the exponential link function in (1) makes it difficult for the spectral-based analyses (Arroyo et al. (2021)) to deal with the entanglement between Z and α and achieve the semiparametric oracle rates of Z. Lastly, the different oracle rates for different parameters also render techniques under the single network setting not easily applicable: in the single network setting z_i and α_i have the same oracle rates, so it suffices to jointly control their overall error rates via the natural parameter $\Theta_{ij} = \alpha_i + \alpha_j + \langle z_i, z_j\rangle$ from a generalized linear model perspective. In contrast, in our setting, each α_{it} is measured by n independent edges: $\{A_{t,ij} : j = 1, \ldots, n\}$, and the oracle squared error rate of each α_{it} would be $O_p(1/n)$, compared to the $O_p(1/(nT))$ error rate of each z_i. Our goal is to attain the oracle estimation error rate of Z, which is a refinement of the overall error rate dominated by α's error.

To accurately estimate the latent space component Z, we next develop two methods: a generalized semiparametric one-step estimator in Section 3, and a semiparametric penalized maximum likelihood estimator in Section 4. We show that both methods achieve, up to a logarithmic factor, the oracle estimation error rates for the target parameters Z.

3. Generalized semiparametric one-step estimator. In this section, we introduce our generalized semiparametric one-step estimator of Z and provide theoretical guarantees. We first introduce some notation. Model (1) leads to the following form for the log-likelihood function:

$$\begin{aligned} L(Z,\alpha) &= L(Z_v, \alpha_v) \\ &= \sum_{t=1}^{T} \sum_{1\le i\le j\le n} \{A_{t,ij}(\alpha_{it} + \alpha_{jt} + \langle z_i, z_j\rangle) - \exp(\alpha_{it} + \alpha_{jt} + \langle z_i, z_j\rangle)\}, \end{aligned} \tag{3}$$

where, for notational convenience in the differentiation of the likelihood, we use Z_v and α_v to denote vectorizations of Z and α, respectively, that is, $Z_v = (z_1^\top, \ldots, z_n^\top)^\top \in \mathbb{R}^{nk\times 1}$ and $\alpha_v = (\alpha_1^\top, \ldots, \alpha_T^\top)^\top \in \mathbb{R}^{nT\times 1}$. Then we let $\dot{L}_Z(Z,\alpha)$ and $\dot{L}_\alpha(Z,\alpha)$ denote the partial derivatives of $L(Z,\alpha)$ with respect to vectors Z_v and α_v, respectively (see the precise formulae in Section B.1 of the Supplementary Material, He et al. (2025)). Unless otherwise specified, such vectorization is applied when considering partial derivatives (for both first and higher orders) throughout the paper. Following the semiparametric literature (Tsiatis (2006)), the efficient score and the efficient Fisher information matrix for Z can be expressed as

$$\begin{aligned} S_{\mathrm{eff}}(Z,\alpha) &= \dot{L}_Z - \mathbb{E}(\dot{L}_Z \dot{L}_\alpha^\top)\{\mathbb{E}(\dot{L}_\alpha \dot{L}_\alpha^\top)\}^{-1}\dot{L}_\alpha \in \mathbb{R}^{nk\times 1}, \\ I_{\mathrm{eff}}(Z,\alpha) &= \mathbb{E}\big(S_{\mathrm{eff}}(Z,\alpha) S_{\mathrm{eff}}^\top(Z,\alpha)\big) \in \mathbb{R}^{nk\times nk}, \end{aligned} \tag{4}$$

where, for notational simplicity, (Z,α) is omitted in $\dot{L}_Z$ and $\dot{L}_\alpha$, and $\mathbb{E}(\cdot)$ refers to the expectation taken when data follows (2) with parameters (Z,α). We have derived analytic formulas for the expectation terms in (4), and express $S_{\mathrm{eff}}(Z,\alpha)$ and $I_{\mathrm{eff}}(Z,\alpha)$ as close-form functions of (Z,α) and the observed data; see Section B.1 of the Supplementary Material (He et al. (2025)).

With the above preparations, we construct our generalized semiparametric one-step estimator as

$$\hat{Z}_v = \check{Z}_v + \{I_{\mathrm{eff}}(\check{Z},\check{\alpha})\}^{+} S_{\mathrm{eff}}(\check{Z},\check{\alpha}), \tag{5}$$

where $(\check{Z}, \check{\alpha})$ denotes an initial estimate, and B^+ represents the Moore–Penrose inverse of a matrix B, which is uniquely defined and also named pseudo inverse (Ben-Israel and Greville (2003)). By the property of pseudo inverse, (5) can be equivalently written as

$$\hat{Z}_v = \check{Z}_v + \check{\mathcal{U}}\{\check{\mathcal{U}}^\top I_{\text{eff}}(\check{Z}, \check{\alpha})\check{\mathcal{U}}\}^{-1}\check{\mathcal{U}}^\top S_{\text{eff}}(\check{Z}, \check{\alpha}), \tag{6}$$

where $\check{\mathcal{U}}$ is a matrix whose columns can be any set of basis of the column space of $I_{\text{eff}}(\check{Z}, \check{\alpha})$.

The estimator originates from the one-step estimator for semiparametric problems (van der Vaart (1998), Section 25.8), which solves the efficient score equation in the presence of nuisance parameters through a linear approximation at an initial estimate. However, different from the classical one-step estimator, the efficient information matrix is singular in our setting; see Remark 1. As a result, the corresponding linear approximation equations are underdetermined. Taking pseudo inverse corresponds to solving the linear approximation equation with minimum ℓ_2 norm; the solution lies in the column space of $\check{\mathcal{U}}$, as indicated in (6).

Owing to the intricate technical challenges outlined in Section 2.2 and the singularity of $I_{\text{eff}}(\check{Z}, \check{\alpha})$, straightforward conclusions regarding the classical one-step estimator for (5) are elusive. However, we have constructed a detailed semiparametric analysis and demonstrated that our proposed one-step estimator nearly achieves the oracle error rate, given suitable initial estimators. We will present the comprehensive theoretical results in the following Section 3.1. Additionally, in Section 3.4, we will introduce an initial estimator that demonstrates both statistical validity and computational efficiency.

REMARK 1. In effect, the singularity of the efficient information matrix $I_{\text{eff}}(Z, \alpha)$ is caused by the redundancy and unidentifiability of parameters in Z (Little, Heidenreich and Li (2010)). Particularly, we find that the rank of $I_{\text{eff}}(Z, \alpha)$ equals $nk - k(k+1)/2$, which is identical to the number of free parameters in the $n \times k$ matrix Z. Below, we provide a high-level explanation, and please find a formal justification in Lemma B.3 in Section B.2.1. First, recall that to avoid the mean shift issue in (2), we have imposed k linear constraints $1_n^\top Z = 0$. Second, even with the centering constraints, Z can only be identified up to an orthogonal group transformation $\mathcal{O}(k)$. In particular, $\mathcal{O}(k)$ is the orthogonal Stiefel manifold, and its dimension is $k(k-1)/2$ (see, e.g., Absil, Mahony and Sepulchre (2008), Section 3.3.2). Intuitively, $k + k(k-1)/2 = k(k+1)/2$ degrees of freedom are removed due to the constraints and unidentifiability. This leads to $nk - k(k+1)/2$ free parameters in Z.

3.1. *Theory.* Throughout the sequel, we use $(Z^\star, \alpha^\star)$ to denote the true value of (Z, α). In other words, our observed data follow the model (1) with $(Z, \alpha) = (Z^\star, \alpha^\star)$. Besides, we denote $\Theta_{t,ij} = \alpha_{it} + \alpha_{jt} + \langle z_i, z_j \rangle$. We define the estimation error from the ith row of $\check{Z}$ as $\text{dist}_i(\check{z}_i, z_i^\star) = \|\check{z}_i - \check{Q}^\top z_i^\star\|_2$, where

$$\check{Q} = \mathop{\arg\min}_{Q \in \mathcal{O}(k)} \|\check{Z} - Z^\star Q\|_{\text{F}}, \tag{7}$$

so that $\text{dist}^2(\check{Z}, Z^\star) = \sum_{i=1}^n \text{dist}_i^2(\check{z}_i, z_i^\star)$. The goal in this subsection is to establish an error bound for the proposed one-step estimator (5) in terms of the distance defined above. To do so, we need the following regularity conditions.

CONDITION 1. *Assume the true parameters $\alpha^\star$'s and $z^\star$'s satisfy*:

(i) *There exist positive constants $M_{Z,1}$, M_α, and $M_{\Theta,1}$ such that $\|z_i^\star\|_2^2 \le M_{Z,1}$, $|\alpha_{it}^\star| \le M_\alpha$, and $\Theta_{t,ij}^\star \le -M_{\Theta,1}$ for $1 \le i, j \le n$ and $1 \le t \le T$;*

(ii) *There exists a positive constant $M_{Z,2}$ such that $\sigma_{\min}[(Z^\star)^\top Z^\star/n] \ge M_{Z,2}$;*

(iii) $1_n^\top Z^\star = 0$.

CONDITION 2. *Assume the initial estimates $\check{\alpha}$'s and $\check{z}$'s satisfy*:

(i) *There exist constants $M_b > 0$ and ϵ such that $|\check{\alpha}_{it} - \alpha^\star_{it}| + \mathrm{dist}_i(\check{z}_i, z^\star_i) \le M_b n^{-1/2} \times \log^\epsilon(nT)$ for $1 \le i \le n$ and $1 \le t \le T$;*

(ii) $1_n^\top \check{Z} = 0$.

Condition 1(i) assumes boundedness of the true parameters. It also implies $-M_{\Theta,2} \le \Theta^\star_{t,ij} \le -M_{\Theta,1}$ with $M_{\Theta,2} = M_{Z,1} + 2M_\alpha$. Condition 1(ii) excludes degenerate cases of $Z^\star$. This, together with $\|z^\star_i\|_2^2 \le M_{Z,1}$, indicates $\|Z^\star\|_{\mathrm{op}} \asymp \sqrt{n}$. Condition 1(iii) is needed in order to identify the interaction effects $Z^\star(Z^\star)^\top$ in (2). Condition 2(ii) requires $\check{Z}$ to be a feasible estimator satisfying the centering constraint. With this condition, $\hat{Z}$ is guaranteed to be feasible, namely $1_n^\top \hat{Z} = 0$. Condition 2(i) is analogous to the "$\sqrt{n}$-consistent initial estimator" used in the classical one-step estimator (van der Vaart (1998)). It assumes that the estimation errors of $\check{\alpha}_{it}$'s and $\check{z}_i$'s are uniformly bounded by $M_b \log^\epsilon(nT)/\sqrt{n}$. In Section 3.4, we will construct an initial estimator satisfying Condition 2 with the error bound of order $\log^2(nT)/\sqrt{n}$ under high probability. Recall that the classical MLE theory indicates that the optimal estimation errors for individual α_{it} and z_i would be of the orders of $n^{-1/2}$ and $(nT)^{-1/2}$, respectively. In this regard, the constructed initial $\check{\alpha}_{it}$ achieves its optimal rate up to a logarithmic factor. On the other hand, Condition 2(i) assumes that $\check{z}_i$ achieves the same error rate as $\check{\alpha}_{it}$ but not its own optimal rate $(nT)^{-1/2}$. In this connection, Condition 2 imposes a mild assumption on the initial $\check{z}_i$'s. We also would like to point out that the deterministic upper bounds in Condition 2(i) may be relaxed to probabilistic upper bounds. The current nonprobabilistic bounds are used to simplify the presentation. Theorem 1 below establishes the near optimality of the proposed one-step estimator (5).

THEOREM 1. *Assume Conditions 1–2. Let $\hat{Z}$ be the generalized one-step estimator defined as in (5). Let $\varsigma = \max\{\epsilon, 1/2\}$. For any constant $s > 0$, there exists a constant $C_s > 0$ such that when $n/\log^{2\varsigma}(T)$ is sufficiently large,*

$$\Pr\left\{\mathrm{dist}^2(\hat{Z}, Z^\star) > \frac{1}{T} \times C_s r_{n,T}\right\} = O(n^{-s}),$$

where $r_{n,T} = \max\{1, \frac{T}{n}\}\log^{4\varsigma}(nT)$.

Theorem 1 implies that with high probability, the estimation error $\mathrm{dist}^2(\hat{Z}, Z^\star)$ is $O(r_{n,T}/T)$. For ease of understanding, we first ignore the logarithmic term in $r_{n,T}$, that is, set $\varsigma = 0$. Then, the error order is reduced to

$$\frac{1}{T} \times \max\left\{1, \frac{T}{n}\right\} = \max\left\{\frac{1}{T}, \frac{1}{n}\right\}. \tag{8}$$

When $T = 1$, that is, for a single network, (8) $= O(1)$ achieves the oracle error rate. This is similar to Ma, Ma and Yuan (2020), which studied a single network; also see Section 3.2 for a detailed discussion. When $1 < T \lesssim n$, (8) $= O(1/T)$, indicating that our proposed generalized one-step estimator achieves the oracle estimation error rate in this case. When $T \gg n$, (8) $= O(1/n)$, which unfortunately, is not exactly inverse proportional to T. Intuitively, when T is very large, there is too much time heterogeneity, resulting in too many nuisance parameters α_{it}'s over time. In turn, this causes technical difficulty in analyzing the target Z. Nevertheless, we still have (8) $= O(1/n)$, which is smaller than the optimal error rate $O(1)$ for a single network. This suggests that given $T > 1$ networks, the estimation error of Z can always be improved compared to using only a single network. This can be interpreted, in a broad sense, as an "inverse proportional to T" property. We call $O(1/n)$ a sub-oracle rate throughout the paper.

Compared to (8), the established error bound $r_{n,T}/T$ contains an extra logarithmic term $\log^{4\varsigma}(nT)$, which comes from the uniform error control over high-dimensional parameters (α, Z) and the initialization error in Condition 2(i). As an example, we propose an initial estimator in Section 3.4 satisfying $\varsigma = \epsilon = 2$. For a finite ς, the error bound $O(r_{n,T}/T)$, similarly to (8), decreases as T increases. To sum up, up to logarithmic factors, $\hat{Z}$ achieves the oracle error rate $O(1/T)$, when $1 \le T \lesssim n$, and $\hat{Z}$ achieves the sub-oracle error rate $O(1/n)$, when $T \gg n$.

REMARK 2. The efficient Fisher information matrix I_{eff} in (6) can also be replaced by the "observed efficient information matrix" defined as $-H_{\text{eff}}$, where

$$H_{\text{eff}}(Z,\alpha) = \ddot{L}_{Z,Z}(Z,\alpha) - \ddot{L}_{Z,\alpha}(Z,\alpha)\ddot{L}^{-1}_{\alpha,\alpha}(Z,\alpha)\ddot{L}_{\alpha,Z}(Z,\alpha)$$

satisfies $\mathbb{E}\{-H_{\text{eff}}(Z,\alpha)\} = I_{\text{eff}}(Z,\alpha)$ (an analytic expression for $H_{\text{eff}}(Z,\alpha)$ is provided in Section B.1.4 of the Supplementary Material, He et al. (2025)). Then, the one-step estimator is given by

$$\check{Z}_v - \check{\mathcal{U}}\{\check{\mathcal{U}}^\top H_{\text{eff}}(\check{Z},\check{\alpha})\check{\mathcal{U}}\}^{-1}\check{\mathcal{U}}^\top S_{\text{eff}}(\check{Z},\check{\alpha}), \tag{9}$$

and theoretically, we could establish the same rate as that in Theorem 1. More details on this are provided in Section C.3 of the Supplementary Material (He et al. (2025)).

REMARK 3. To establish the near-oracle rate in Theorem 1, the key idea is to show that through our semiparametric efficient construction, the estimation error of α_{it}'s would not mask that of z_i's. Intuitively, this is achievable due to the fact $\mathbb{E}\{\frac{\partial S_{\text{eff}}(Z,\alpha)}{\partial \alpha_{it}}\} = 0$. This shows that a small perturbation error of α_{it}'s would not change the efficient score of z_i's in terms of the first-order expansion and therefore, its influence on estimating z_i's can be reduced (Bickel et al. (1993)). Nevertheless, establishing the near-oracle rate of Z is still challenged by the complex structures of α and Z under our model (1).

For the nuisance parameters α, we do not impose any structural assumptions, such as smoothness or sparsity. Therefore, α_{it}'s can be completely heterogeneous over all $i \in \{1,\dots,n\}$ and $t \in \{1,\dots,T\}$, and the free dimension of α is nT. Such two-way heterogeneity and high dimensionality of α make it difficult to separate the estimation error of α from that of Z as discussed in Section 2.2. Technically, we reduce the influence of estimating α by carefully investigating the semiparametric efficient score. But that influence cannot be eliminated entirely, which causes a sub-oracle rate when T is very large, as shown in (8).

For the target parameters Z, the non-Euclidean geometry and identifiability constraints of Z lead to the singularity of the efficient information matrix $I_{\text{eff}}(\check{Z},\check{\alpha})$ as mentioned in Remark 1. To address the issue, we explicitly characterize a restricted subspace of $\mathbb{R}^{nk}$ on which $I_{\text{eff}}(\check{Z},\check{\alpha})$ has positive eigenvalues. The restricted eigenspace exhibits interesting properties for characterizing the estimation error of $\hat{Z}$ and has fundamental connections with two sources of identifiability constraints of Z in Remark 1. Please see details in Section B.2.1 of the Supplementary Material (He et al. (2025)).

REMARK 4. To achieve the desired oracle rate, our constraint $T \lesssim n$ has fundamental connections to classical results in semiparametric analysis. We first elaborate on typical rates in classical semiparametric analyses and explain the connections with our result. Consider a canonical semiparametric problem with target parameters θ and nuisance parameters η with their estimates denoted by $\hat{\theta}$ and $\hat{\eta}$, respectively. Under suitable regularity conditions, the contribution from the nuisance parameters to $\|\hat{\theta} - \theta\|_2$ can often be bounded by $O(\|\hat{\eta} - \eta\|_2^2)$, a squared ℓ_2-error of nuisance parameters; see, for example, Section 25.59 in van der Vaart (1998). When θ is fixed-dimensional, its oracle estimation error rate is typically

TABLE 1
Squared estimation error rates of shared latent vectors z_i's in different models

	Models with identity link		Models with logit link	
Model	Arroyo et al. (2021)	MacDonald et al. (2022)	Ma et al. (2020) ($T = 1$)	Zhang et al. (2020)
Error Order	$O(\frac{1}{T}+\frac{1}{n})$	$O(\frac{n^\eta}{T})$ ($\eta \geq 0$, $T \ll n^{1/2}$)	$O(1)$	$O(1+\frac{1}{T})$

quadratic, that is, $\|\hat{\theta}-\theta\|_2 = O_p(n^{-1/2})$. Then for the contributed squared error $O(\|\hat{\eta}-\eta\|_2^2)$ to be no larger than $\|\hat{\theta}-\theta\|_2$, it is common to impose a quarter rate $\|\hat{\eta}-\eta\|_2 = O_p(n^{-1/4})$ on estimating the nuisance parameters. Other examples include Section 7.6 in Bickel et al. (1993), Section 3 in Murphy and van der Vaart (2000), and equation (3.8) in Chernozhukov et al. (2018), etc. We next show that our constraint $T \lesssim n$ aligns with this classical quarter rate. Specifically, as discussed in Section 2.2 of the main text, the oracle rates for estimating each target parameter $z_i^\star$ and each nuisance parameter $\alpha_{it}^\star$ are expected to be $O_p((nT)^{-1/2})$ and $O_p(n^{-1/2})$, respectively. Following the above "squared contribution" idea, when each squared oracle error rate of $\alpha_{it}^\star$ is no larger than the oracle error rate of $z_i^\star$, the constraint $T \lesssim n$ is required. In this sense, our constraint on T aligns with the common quarter rate requirement on nuisance parameters in the classical literature.

Furthermore, we would like to point out that similar error rates and constraints on T and n exist in many other related network studies. In particular, we have provided a summary of existing estimation error rates in Table 1 below, showing that all those studies require certain constraints on the growth rate of T with respect to n to achieve the oracle rate for estimating Z. In terms of the error rate and the constraints on (n, T), our result is consistent with or more relaxed than the existing results in Table 1.

3.2. *Related works*. Various models have been proposed for multiple heterogeneous networks. We briefly review related works focusing on fixed-effect continuous latent vectors.

One class of multiple-network models generalizes the random dot product graphs (Young and Scheinerman (2007), Athreya et al. (2018), Rubin-Delanchy et al. (2022)). For example, Arroyo et al. (2021) considered settings with T layers of networks, and in the tth layer network with $1 \leq t \leq T$, each edge $A_{t,ij} = A_{t,ji} \sim \text{Bernoulli}\{\mathbb{E}(A_{t,ij})\}$ independently for $1 \leq t \leq T$ and $1 \leq i < j \leq n$, where $\mathbb{E}(A_{t,ij}) = \langle z_i, z_j\rangle_{\Lambda_t}$ with $\langle z_i, z_j\rangle_{\Lambda_t} = z_i^\top \Lambda_t z_j$. Here z_i's denote the common latent positions shared across T networks, and $\Lambda_t \in \mathbb{R}^{k\times k}$ characterizes the layer-specific latent information. A similar model for directed random graphs has been studied in Zheng and Tang (2022) with nonsymmetric latent positions shared across T networks. MacDonald, Levina and Zhu (2022) proposed models based on generalized random dot product graphs and established theoretical results when $A_{t,ij} \sim \text{Normal}\{\mathbb{E}(A_{t,ij}), \sigma^2\}$ independently, where σ is a nuisance parameter, and $\mathbb{E}(A_{t,ij}) = \langle z_i, z_j\rangle_{\mathrm{I}_{p_1,q_1}} + \langle u_{i,t}, u_{j,t}\rangle_{\mathrm{I}_{p_{2,t},q_{2,t}}}$. Here z_i's denote the shared latent vectors, whereas $u_{i,t}$'s denote the layer-specific latent vectors, and $\langle\cdot,\cdot\rangle_{\mathrm{I}_{p,q}}$ defines an indefinite inner product, where $\mathrm{I}_{p,q} = \text{diag}(1,\ldots,1,-1,\ldots,-1)$ with p ones followed by q negative ones on its diagonal for $p \geq 1$ and $q \geq 0$.

From the perspective of generalized linear models, the above models can be viewed as linear responses with the identity link function. Considering a single network with binary edges, Hoff, Raftery and Handcock (2002) proposed models with the logit link function, and Ma, Ma and Yuan (2020) proposed and analyzed two estimation methods. Generalizing the idea to multilayer binary networks, Zhang, Xue and Zhu (2020) proposed the following model: for $t = 1,\ldots,T$ layers of networks, each edge $A_{t,ij} \sim \text{Bernoulli}\{\mathbb{E}(A_{t,ij})\}$ independently for $1 \leq i, j \leq n$, where $\mathbb{E}(A_{t,ij}) = \text{logit}^{-1}(\alpha_{it}+\alpha_{jt}+\langle z_i, z_j\rangle_{\Lambda_t})$ with $\text{logit}^{-1}(x) = e^x/(1+e^x)$.

Under the above models, one common interest is to estimate the latent vectors $Z = [z_1, \ldots, z_n]^\top$ shared across multiple heterogeneous networks. Table 1 summarizes the squared estimation error rates for Z in four studies with a focus on the order of the error with respect to n and T only. Our results and intuition on the oracle error rate are consistent with some existing results. But establishing similar results under our model (1) requires nontrivial technical developments due to various challenges, including two-way heterogeneity, high-dimensionality, and nonlinearity of the link function as discussed in Section 2.2. To the best of our knowledge, for multiple networks ($T > 1$) with *nonlinear* link functions in models, there is no existing result that is comparable to our derived nearly optimal rate in Theorem 1.

Besides the aforementioned works, there are other models for analyzing longitudinal or multiple networks that focus on different goals and challenges. For instance, a prevalent subclass within the latent space models is the stochastic block model (SBM) (Holland, Laskey and Leinhardt (1983), Lee and Wilkinson (2019)). For a single network, our proposed latent space model can reduce to SBM or degree-corrected block models (Karrer and Newman (2011), Zhao, Levina and Zhu (2012), Jin (2015)) under suitable parametrizations. Notably, the baseline component $\exp(\alpha_{it})$ for each vertex i is similar to degree-correction parameters in the degree-corrected block model, showing flexibility for accommodating degree heterogeneity. Typically, SBM assumes that vertices form stochastically equivalent clusters, which leads to low-rank expected adjacency matrix and discrete latent structures. The intrinsic discreteness of SBMs makes them structurally simple and helpful in identifying clusters. Alternatively, the general latent space models are advantageous when there is no apparent clustering structure, and the expected adjacency matrix violates low-rankness. As a result, the model properties and analysis techniques under those two types of models can be quite different. For multiple networks, various extensions based on block models have been proposed (Pensky and Zhang (2019), Lei, Chen and Lynch (2020), Zhang and Chen (2020), Lei and Lin (2023), Bazzi et al. (2020), Paul and Chen (2022), Agterberg, Lubberts and Arroyo (2022), Cai et al. (2022)). They often assume the expected adjacency matrix is low-rank and emphasize cluster recovery and community detection, which differs from the objectives and challenges in this work.

Another line of research proposes to impose a low-rank tensor structure for modeling the longitudinal network (Lyu, Xia and Zhang (2023), Zhang and Wang (2023)). Such models differ from our model in terms of interpretation and complexity. In particular, under fixed ranks, the total number of parameters under (1) is of the order of $O(nT)$, whereas that under a tensor structure is often reduced to the order of $O(n + T)$ or even smaller. Under (1), nT number of baseline parameters α_{it} are used to flexibly model the degree heterogeneity of different nodes and time points. This, as mentioned above, is similar to the use of degree-correction parameters in degree-corrected block models and does not have a straightforward counterpart in those related low-rank tensor models. Due to this difference in the number of parameters, the model complexities are distinct, and thus, the estimation error rates are not directly comparable.

3.3. *Geometric interpretation on quotient manifold.* The proposed (6) treats the $n \times k$ matrix Z as an nk-dimensional vector Z_v. This ignores the constraint of the space $\mathbb{R}_0^{n\times k} = \{Z \in \mathbb{R}^{n\times k} : 1_n^\top Z = 0, \det(Z^\top Z) \neq 0\}$. Moreover, viewing Z as an element of $\mathbb{R}_0^{n\times k}$ ignores that it is identified up to a rotation under our considered log-likelihood $L(Z, \alpha)$, as $L(Z, \alpha) = L(ZQ, \alpha)$ for any $Q \in \mathcal{O}(k)$. Interestingly, in our considered problem, we find that (6) implicitly takes the underlying parameter structure into account.

To demonstrate this, we first show the intrinsic parameter space of Z is a quotient set. In particular, the "rotation invariance" naturally induces an equivalence relation on $\mathbb{R}_0^{n\times k}$: $Z_1 \sim Z_2$ if and only if there exists $Q \in \mathcal{O}(k)$ such that $Z_2 = Z_1 Q$. Given the equivalence re-

lation $\sim$ on $\mathbb{R}_0^{n\times k}$ and an element $Z \in \mathbb{R}_0^{n\times k}$, elements in $\mathbb{R}_0^{n\times k}$ that are equivalent to Z form an equivalence class of Z, denoted as $[Z]$. The set of equivalence classes of all elements in $\mathbb{R}_0^{n\times k}$ is called the quotient set, denoted as $\mathbb{R}_0^{n\times k}/\sim$. Equipped with the equivalence relation $\sim$, the quotient set $\mathbb{R}_0^{n\times k}/\sim$ naturally incorporates the rotation invariance of Z and thus is the intrinsic parameter space to examine. For the simplicity of notation, we let $\mathcal{M}$ denote $\mathbb{R}_0^{n\times k}/\sim$ below.

We next show that (6) can be viewed as an approximation to the one-step Newton–Raphson update of the log profile likelihood on the quotient set $\mathcal{M}$. Specifically, the log profile likelihood of Z is $pL(Z) = L(Z, \hat{\alpha}(Z))$ with $\hat{\alpha}(Z) = \arg\max_{\alpha\in\mathbb{R}^{n\times T}} L(Z,\alpha)$. The MLE of Z, that is, the first component of $(\hat{Z}_{\mathrm{MLE}}, \hat{\alpha}_{\mathrm{MLE}})$ that maximizes $L(Z,\alpha)$, is also the maximizer of $pL(Z)$. Therefore, to investigate the MLE of the target parameter Z, it suffices to examine $pL(Z)$ (Murphy and van der Vaart (2000)). On the quotient set $\mathcal{M}$, $pL(Z)$ naturally induces a function $pL_Q : \mathcal{M} \to \mathbb{R}$ with $pL_Q([Z]) = pL(Z)$ for any $Z \in \mathbb{R}_0^{n\times k}$. To incorporate the rotation invariance relation, it is natural to search for the maximizer of pL_Q on the quotient set $\mathcal{M}$. Fortunately, a Newton–Raphson step for pL_Q can be properly constructed given the nice properties of $\mathcal{M}$ and pL_Q in Lemma 2.

LEMMA 2 (Lee, 2013, Chapter 21). *On $\mathbb{R}_0^{n\times k}$, consider the canonical atlas $\varphi : \mathbb{R}_0^{n\times k} \to \mathbb{R}^{(n-1)k}$ given by $\varphi(Z) = (z_1^\top, \ldots, z_{n-1}^\top)^\top$, and the canonical Riemannian metric $g_Z(Z_1, Z_2) = \mathrm{tr}(Z_1^\top Z_2)$ for Z_1 and Z_2 in the tangent space to $\mathbb{R}_0^{n\times k}$ at Z. (i) Endowed with the differential structure and the Riemannian metric induced by that of $\mathbb{R}_0^{n\times k}$, the quotient set $\mathcal{M}$ is a smooth Riemannian quotient manifold of dimension $nk - k(k+1)/2$. (ii) $pL_Q : \mathcal{M} \to \mathbb{R}$ is a smooth function.*

In particular, consider the Riemannian manifold $\mathcal{M}$ equipped with the Riemannian connection and the exponential retraction R (Absil, Mahony and Sepulchre (2008), Section 5). For a smooth function $f : \mathcal{M} \to \mathbb{R}$, the Newton–Raphson update at an equivalence class $[Z] \in \mathcal{M}$ is given by $R_{[Z]}(\nu)$, where ν is a tangent vector in the tangent space $T_{[Z]}\mathcal{M}$ and specified by $\mathrm{Hess}\, f([Z])[\nu] = -\mathrm{Grad}\, f([Z])$, and $\mathrm{Hess}\, f([Z])$ and $\mathrm{Grad}\, f([Z])$ denote the Hessian and gradient of f at $[Z]$, respectively (Boumal (2023)). Conceptually, $R_{[Z]}(\nu)$ defines a move in the direction of ν while staying on the manifold. Interestingly, for the function pL_Q, we prove that $R_{[Z]}(\nu)$ has an analytic (matrix) form in Proposition 3.

PROPOSITION 3. *Under the same setting as Lemma 2, equip $\mathcal{M}$ with the Riemannian connection and the exponential retraction R. Consider an initial value $\check{Z}$ and its equivalence class $[\check{Z}]$. For pL_Q at $[\check{Z}]$, the Newton–Raphson update $R_{[\check{Z}]}(\nu) = [\mathrm{mat}(\check{Z}_v + \bar{\nu}_{\check{Z}})]$, where $\check{Z}_v$ is the vectorization of $\check{Z}$, $\mathrm{mat}((z_1^\top, \ldots, z_n^\top)^\top) = (z_1, \ldots, z_n)^\top \in \mathbb{R}^{n\times k}$, and*

$$\bar{\nu}_{\check{Z}} = -\check{\mathcal{U}}[\check{\mathcal{U}}^\top H_{\mathrm{eff}}\{\check{Z}, \hat{\alpha}(\check{Z})\}\check{\mathcal{U}}]^{-1}\check{\mathcal{U}}^\top S_{\mathrm{eff}}\{\check{Z}, \hat{\alpha}(\check{Z})\},$$

where $\check{\mathcal{U}}$ is the same as that in (6), and $H_{\mathrm{eff}}(Z,\alpha)$ is defined same as in Remark 2.

Note that $\check{Z}_v + \bar{\nu}_{\check{Z}}$ is similar to (9) except that $\hat{\alpha}(\check{Z})$ is replaced with $\check{\alpha}$. Given initial values $(\check{Z}, \check{\alpha})$ that are close to the true values $(Z^\star, \alpha^\star)$, we can show that $\hat{\alpha}(\check{Z})$ and $\check{\alpha}$ are close, and thus $\check{Z}_v + \bar{\nu}_{\check{Z}}$ and (9) are close; see rigorous results in Section F.4 of the Supplementary Material (He et al. (2025)). By Remark 2, (5), (6), and (9) can all be viewed as approximations to a one-step Newton–Raphson update on the quotient manifold $\mathcal{M}$, whose construction implicitly incorporates the underlying geometric structures of the quotient manifold.

REMARK 5. We provide an intuitive explanation of Proposition 3. Motivated by the fact that the log profile likelihood $pL(Z)$ depends on parameters in Z only through $ZZ^\top$, we

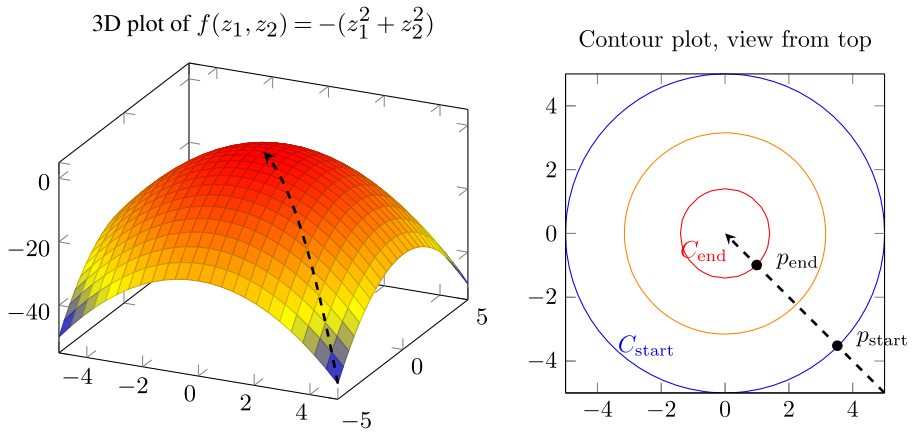


FIG. 1. *Each circle represents one equivalence class of* (z_1, z_2) *giving the same value of* $\bar{f}(z_1, z_2)$. *It suffices to search for maximization along one given direction.*

next focus on $ZZ^\top$. For ease of visualization, we first consider the parameter space $\mathbb{S}_0^{2\times 2} = \{Z \in \mathbb{R}^{2\times 2} : 1_2^\top Z = 0_{1\times 2}\}$. For $Z \in \mathbb{S}_0^{2\times 2}$, $ZZ^\top = (z_{11}^2 + z_{12}^2)(1,-1)^\top(1,-1)$ depends on Z only through the scalar $z_{11}^2 + z_{12}^2$. This suggests that the number of free parameters in $ZZ^\top$ is 1, which equals $nk - k(k+1)/2$ as analyzed in Remark 1. We visualize the function $\bar{f}(z_{11}, z_{12}) = -(z_{11}^2 + z_{12}^2)$ in Figure 1. The function has a constant value on each circle of (z_{11}, z_{12}) in the contour plot. From this perspective, each circle is an equivalence class giving the same value of $\bar{f}$. The set of all equivalence classes (circles) yields a quotient set, denoted as $\mathcal{M}_2$, and $\bar{f}$ induces a function $f : \mathcal{M}_2 \to \mathbb{R}$ given by $f([z]) = \bar{f}(z_{11}, z_{12})$, where $[z] \in \mathcal{M}_2$ represents the equivalence class specified by (z_{11}, z_{12}).

Consider the problem of maximizing the function f on $\mathcal{M}_2$. One natural idea is to search for the maximizer of f across its domain $\mathcal{M}_2$. In this case, an update step in $\mathcal{M}_2$ yields a move from one circle to another, for example, C_{start} to C_{end} in Figure 1. Such an abstract update can also be described in the Euclidean space $\mathbb{R}^2$. In particular, fix one direction that is orthogonal to a tangent line of circles, for example, the dashed line in Figure 1, move along the tangent line from the point p_{start} to the point p_{end} in Figure 1, and map p_{end} to the corresponding equivalence class C_{end}. The update from p_{start} to p_{end} in $\mathbb{R}^2$ can be analytically represented in a matrix form. Generalizing this idea, an update step on the quotient manifold $\mathbb{R}_0^{n\times k}/\sim$ can be described through a fixed search space in the Euclidean space $\mathbb{R}^{n\times k}$ (analogous to the search direction in Figure 1), and a common choice is the so-called horizontal space (Boumal (2023)). Interestingly, we prove that the horizontal space at $\check{Z}$ can be explicitly expressed through $\check{\mathcal{U}}$. This enables us to derive an analytic form $\bar{v}_{\check{Z}}$, which, after properly aligned into a matrix, defines an update in the horizontal space. The retraction R maps the post-update matrix $\text{mat}(\check{Z}_v + \bar{v}_{\check{Z}})$ to an element in $\mathbb{R}_0^{n\times k}/\sim$, which gives the targeted one-step estimator in the manifold.

3.4. *Initial estimation.* In this section, we construct an initial estimator $(\check{Z}, \check{\alpha})$ and prove that it satisfies Condition 2 with $\epsilon = 2$ and high probability. The proposed initialization algorithm consists of two stages outlined below, while the implementation details are provided in Section D.1 of the Supplementary Material (He et al. (2025)). In the following, we define $E_t^\star = \mathbb{E}(\mathbf{A}_t)$, $\Theta_t^\star = \log(E_t^\star)$ for $t = 1, \ldots, T$, and $G^\star = Z^\star(Z^\star)^\top$. We note that

$$\alpha_t^\star = H_n^{-1}\Theta_t^\star 1_n, \quad \text{and} \quad G^\star = \sum_{t=1}^{T}(\Theta_t^\star - \alpha_t^\star 1_n^\top - 1_n(\alpha_t^\star)^\top)/T, \tag{10}$$

where $H_n = n\mathrm{I}_n + 1_n 1_n^\top$.

Outline of the initialization algorithm.

(1) The first stage obtains an "initial of initial" estimator $(\mathring{Z}, \mathring{\alpha})$ as follows. A similar "double-SVD" idea has been introduced in Zhang, Chen and Li (2020).

(a) Construct an estimator $\mathring{E}_t$ for $E_t^\star$ by applying the universal singular value thresholding (Chatterjee (2015)) to each matrix $\mathbf{A}_t$. Take $\mathring{\Theta}_t = \log(\mathring{E}_t)$ as an estimator for $\Theta_t^\star$.
(b) Let $\mathring{\alpha}_t = H_n^{-1}\mathring{\Theta}_t 1_n$, motivated by (10). Take $\mathring{\alpha} = (\mathring{\alpha}_1, \ldots, \mathring{\alpha}_T)$.
(c) Let $\mathring{G} = \sum_{t=1}^T (\mathring{\Theta}_t - \mathring{\alpha}_t 1_n^\top - 1_n \mathring{\alpha}_t^\top)/T$, motivated by (10).
Let $\mathring{Z} = \mathring{U}_k \mathring{D}_k^{1/2}$, where $\mathring{U}_k \mathring{D}_k \mathring{U}_k^\top$ denote the top-k eigen-decomposition of $\mathring{G}$.

(2) The second stage utilizes the projected gradient descent method (Chen and Wainwright (2015)) with $L(Z, \alpha)$ as an objective function to maximize, and $(\mathring{Z}, \mathring{\alpha})$ from the first stage as initial values. In particular:

(a) Update Z and α along their corresponding gradient directions with pre-specified step sizes.
(b) Project the updated estimates to the constraint set $\mathcal{S}_C$ induced by the conditions of identifiability and boundedness of parameters in Condition 1. Specifically,

$$\mathcal{S}_C = \Big\{(Z, \alpha) : Z \in \mathbb{R}^{n\times k}, 1_n^\top Z = 0, \max_i \|z_i\|_2^2 \le M_{Z,1}, \alpha \in \mathbb{R}^{n\times T}, \max_{i,t} |\alpha_{it}| \le M_\alpha\Big\}.$$

(c) Repeat (a) and (b) until convergence. The resulting values, denoted as $(\check{Z}, \check{\alpha})$, are used as the initial estimator in the proposed one-step estimator.

In Theorem 4, we establish an elementwise error bound of $(\check{Z}, \check{\alpha})$ obtained through the proposed initialization algorithm above.

THEOREM 4. *Assume Conditions 1 and 3. For any constant $s > 0$, there exists a constant $C_s > 0$ such that when $n/\log^{k+3}(T)$ is sufficiently large,*

$$\Pr\Big[\max_{1\le i\le n, 1\le t\le T}\big\{|\check{\alpha}_{it} - \alpha_{it}^\star| + \mathrm{dist}_i(\check{z}_i, z_i^\star)\big\} > \frac{C_s \log^2(nT)}{\sqrt{n}}\Big] = O\big((nT)^{-s}\big).$$

Theorem 4 implies that with high probability, the initial estimator $(\check{Z}, \check{\alpha})$ satisfy Condition 2 with $\epsilon = 2$.

4. Semiparametric penalized maximum likelihood estimator. The log-likelihood $L(Z, \alpha)$ is nonconvex in Z, and also requires specification of the dimension of each z_i. On the other hand, when the log-likelihood (3) is reparametrized through $G = (G_{ij})_{n\times n}$ with $G_{ij} = \langle z_i, z_j\rangle$, the log-likelihood function

$$l(G, \alpha) = \sum_{t=1}^T \sum_{1\le i\le j\le n} \big[A_{t,ij}(\alpha_{it} + \alpha_{jt} + G_{ij}) - \exp(\alpha_{it} + \alpha_{jt} + G_{ij})\big]$$

is convex in α and G. Nevertheless, $G = ZZ^\top$ itself overparameterizes the model, and the dimension of z_i corresponds to the rank of G in this model specification. Subject to the rank constraint, solving MLE of G is a nonconvex optimization problem, even though $l(G, \alpha)$ is convex. To overcome this issue, we construct a penalized maximum likelihood estimator of G with a nuclear norm penalization that relaxes the exact rank constraint on G. We further show that the proposed estimator can achieve the corresponding almost-oracle error rate.

We first characterize the parameter space of G and its relationship with Z. From the perspective of likelihood functions, matrices G and Z such that $L(Z, \alpha) = l(G, \alpha)$ can be viewed

as equivalent parameters. Accordingly, any $Z \in \mathbb{R}_0^{n\times k}$ uniquely defines $G = ZZ^\top \in \mathbb{S}_{0,+}^{n,k}$, where $\mathbb{S}_{0,+}^{n,k}$ represents the class of all $n \times n$ positive semidefinite symmetric matrices with rank k and $G1_n = 0_n$. Moreover, any matrix $G \in \mathbb{S}_{0,+}^{n,k}$ induces $G^{1/2} = U_k D_k^{1/2} \in \mathbb{R}_0^{n\times k}$, where $U_k D_k U_k^\top$ is the top-k eigenvalue components of G. Then G uniquely identifies an equivalence class $\{G^{1/2}Q : Q \in \mathcal{O}(k)\} = \{Z \in \mathbb{R}_0^{n\times k} : ZZ^\top = G\}$.

The penalized maximum likelihood estimator is defined as the solution $(\hat{G}, \hat{\alpha})$ to the following convex optimization problem:

$$\begin{aligned} &\max_{G,\alpha} l(G,\alpha) - \lambda_{n,T}\|G\|_* \\ &\quad \text{subject to} \quad G \in \mathbb{S}_+^n, \quad G1_n = 0_n, \quad |G_{ij}| \le M_{Z,1}, \end{aligned} \tag{11}$$

where $\mathbb{S}_+^n$ represents the class of $n \times n$ positive semidefinite matrices, $M_{Z,1}$ is a constant same as in Condition 1, and $\lambda_{n,T}$ is a prespecified tuning parameter.

To scale the error of G to the error of Z, we present the estimation error of G in terms of $\|\hat{G} - G^\star\|_{\mathrm{F}}^2/n$; see Remark 6. The following Theorem 5 shows that the penalized MLE $\hat{G}$ solved from (11) almost achieves the oracle rate, similarly to Theorem 1.

THEOREM 5. *Assume $G^\star = Z^\star Z^{\star\top}$ with $Z^\star \in \mathbb{R}_0^{n\times k}$ satisfying Condition 1. Suppose we choose $\lambda_{n,T} \asymp \sqrt{nT}\log(nT)$. Then for any constant $s > 0$, there exists a constant $C_s > 0$ such that when $n/\log(T)$ is sufficiently large,*

$$\Pr\left\{\|\hat{G} - G^\star\|_{\mathrm{F}}^2/n > \frac{1}{T} \times C_s r'_{n,T}\right\} = O(n^{-s}),$$

where $r'_{n,T} = \max\{1, \frac{T}{n}\}\log^2(nT)$.

Theorem 5 implies that with high probability, the estimation error $\|\hat{G} - G^\star\|_{\mathrm{F}}^2/n$ is $O(r'_{n,T}/T)$. Similarly to Theorem 1, when ignoring the logarithmic term in $r'_{n,T}/T$, the error order reduces to (8). Therefore, up to the logarithmic term, the penalized MLE achieves the oracle estimation error rate $O(1/T)$ when $1 \le T \lesssim n$, and it achieves the sub-oracle rate $O(1/n)$ when $T \gg n$.

REMARK 6. Intuitively, n^2 parameters in $G^\star$ are redundant and essentially determined by only nk parameters in $Z^\star$. To align with the error rate in Theorem 1, we consider the estimation error of $\hat{G}$ with a scaling factor $1/n$ in Theorem 5. When the rank k is known, the one-step estimator $\hat{Z}$ induces the estimator $\hat{Z}\hat{Z}^\top$ for $G^\star$ satisfying $\|\hat{Z}\hat{Z}^\top - G^\star\|_{\mathrm{F}}^2/n \asymp \mathrm{dist}^2(\hat{Z}, Z^\star) \asymp 1/T$; see Lemmas G.3 and G.4 in the Supplementary Material (He et al. (2025)). When k is unknown, Theorem 5 suggests the penalized MLE $\hat{G}$ achieves the same error rate as that of $\hat{Z}\hat{Z}^\top$.

REMARK 7. The idea of using the nuclear norm penalization as a convex relaxation of the exact rank constraint originates from the low-rank matrix recovery literature (Candès and Tao (2010), Candès and Recht (2012), Davenport et al. (2014)). It was also utilized by Ma, Ma and Yuan (2020) to fit the latent space model for a single network. But different from these studies, we have two components G and α in our model with different but entangling oracle error rates. To establish a sharp error rate for $\hat{G}$, we develop a novel semiparametric analysis for the penalized MLE through the profile likelihood (Murphy and van der Vaart (2000)).

In particular, we note that $\hat{G}$ is also the penalized MLE of the log profile likelihood $pl(G) = l(G, \hat{\alpha}(G))$, where $\hat{\alpha}(G) = \arg\max_{\alpha \in \mathbb{R}^{n \times T}} l(G, \alpha)$. Let $\dot{l}_\alpha$ and $\ddot{l}_{\alpha\alpha}$ represent the first-order derivatives and the second-order derivatives of $l(G, \alpha)$ with respect to α_v, respectively. We can obtain $\hat{\alpha}(G) - \alpha = (\ddot{l}_{\alpha\alpha})^{-1}\dot{l}_\alpha + \text{high-order terms}$, and $\mathbb{E}(\dot{l}_\alpha) = 0$. Intuitively, this suggests that the perturbation error from $\hat{\alpha}(G)$ can be small in terms of the first-order expansions, so its impact on the profile likelihood of G can be reduced.

Despite the wide use of the profile likelihood method (Murphy and van der Vaart (2000), Severini and Wong (1992)), our analysis is faced with unique technical challenges. First, not only are the nuisance parameters α high-dimensional but also they encode two-way heterogeneity over both nodes and time. As a result, to separate the estimation error of G from that of α requires delicate analysis that differs significantly from the existing studies. Second, the target parameter G is high-dimensional and intrinsically redundant with a low-rank structure. The redundancy of parameters in G, as explained in Remark 1, leads to singularity issues when analyzing the likelihood function and calls for new technical developments. Meanwhile, the low-rank structure of G motivates the use of the nuclear norm penalization, which needs to be properly taken care of in our semiparametric analysis.

REMARK 8 (Relationship between the two estimators). The two estimators are fundamentally connected as they are both motivated by maximizing the semiparametric profile likelihood. In particular, the penalized MLE is a convex relaxation of maximizing the log-likelihood function $l(ZZ^\top, \alpha)$, which is equivalent to maximizing the log profile likelihood $pL(Z)$. The one-step estimator solves the maximization of the quadratic function $q(x) = pL(\check{Z}) + (x - \check{Z}_v)^\top S_{\text{eff}}(\check{Z}, \check{\alpha}) - \frac{1}{2}(x - \check{Z}_v)^\top I_{\text{eff}}(\check{Z}, \check{\alpha})(x - \check{Z}_v)$. In settings with fixed-dimensional target parameters, the log profile likelihood is approximated by its local quadratic expansion similar to $q(x)$ (Murphy and van der Vaart (2000)). From an alternative point of view, the one-step estimator gives an approximate solution to the efficient score equation, which, in general, is equivalent to maximizing the profile likelihood. Therefore, we expect that the two proposed estimators are similar up to the convex relaxation, and they can achieve the oracle error rate as the nuisance α is eliminated through the profile likelihood.

In practice, the penalized MLE differs from the nonconvex one-step estimator since it adopts a convex relaxation and does not require a prespecification of the true latent dimension k. Nevertheless, the computation of the one-step estimator is more efficient than solving penalized MLE, and it can achieve a better performance when oracle k is specified. Both convex and nonconvex estimators are widely examined in the related literature (Ma, Ma and Yuan (2020)), and their theoretical properties are of interest to study.

REMARK 9. Given $\hat{G}$ solved from (11), we can further construct an estimator for $Z^\star$. If the true k is given, we let $\hat{Z}_k = \hat{U}_k \hat{D}_k^{1/2}$, where $\hat{U}_k \hat{D}_k \hat{U}_k^\top$ is the top-k eigenvalue components of the penalized MLE $\hat{G}$. When k is unknown, we can consistently estimate k. Intuitively, since k equals the rank of $G^\star$, we can estimate k as the number of significant eigenvalues of an estimator $\hat{G}_0$ that approximates $G^\star$ sufficiently well. For instance, we can set $\hat{G}_0$ to be the penalized MLE $\hat{G}$ in (11) or the "double-SVD" estimator $\mathring{G}$ in Section 3.4, both of which can be obtained without prespecifying k. We have developed rigorous theoretical guarantees and conducted numerical experiments in Section I of the Supplementary Material (He et al. (2025)). Notably for the one-step estimator in Section 3, we can also consistently estimate k when it is unknown.

5. Simulation studies. We conduct simulation studies to examine how the estimation errors vary with respect to n and T. To this end, we consider the following two scenarios of (n, T):

(a) fixing $n = 200$ and varying $T \in \{5, 10, 20, 40, 80\}$;
(b) fixing $T = 20$ and varying $n \in \{100, 200, 400, 800\}$.

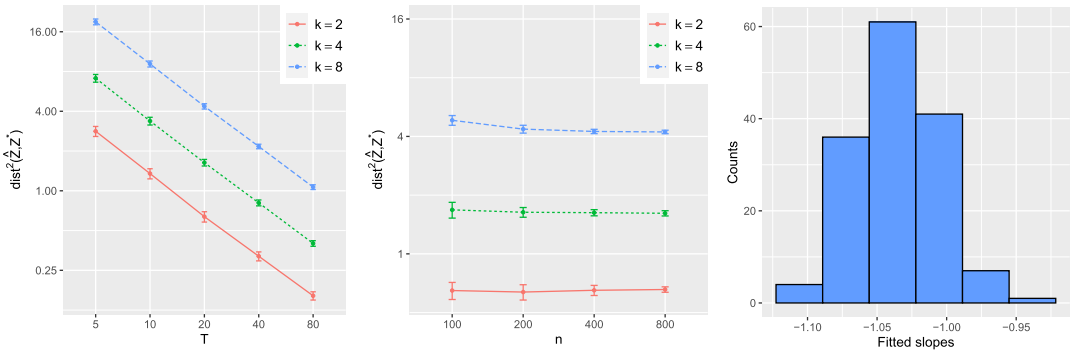
In each scenario, we allow varying $k \in \{2, 4, 8\}$. Given (n, T, k), we generate the shared latent vectors $Z^\star$ as follows: independently sample w_i following a uniform distribution on $\mathbb{B}_2^k = \{x \in \mathbb{R}^k : \|x\|_2 \le 1\}$, let $W = [\tilde{w}_1, \ldots, \tilde{w}_n]^\top$ where $\tilde{w}_j = w_j - \sum_{l=1}^n w_l/n$, and set $Z^\star = \sqrt{n}W/\|WW^\top\|_{\mathrm{F}}^{1/2}$. In this way, $Z^\star$ has zero column means, and $G^\star = Z^\star Z^{\star\top}$ satisfies $\|G^\star\|_{\mathrm{F}}/n = 1$. For the baseline heterogeneity parameters $\alpha^\star$, we consider two cases:

(I) $\alpha_{it}^\star$ are independently sampled from $U(-2, 0)$;
(II) $\alpha_{it}^\star = t/T + \beta_{it}$ with β_{it} independently sampled from $U(-3, -1)$ for $1 \le i \le \lfloor n/2 \rfloor$, and $\alpha_{it}^\star = -2t/T + \beta_{it}$ with β_{it} independently sampled from $U(-2, 0)$ for $\lfloor n/2 \rfloor < i \le n$.

Intuitively, $\alpha^\star$'s in Cases (I) and (II) represent different types of heterogeneity; the former is uniform over i, whereas the latter has a two-block structure over i.

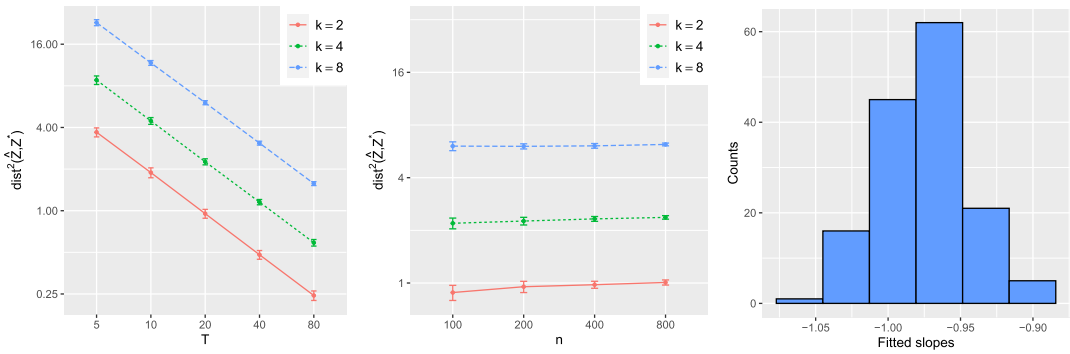
Under each model configuration and for each estimator, we estimate the error $\mathrm{dist}^2(\hat{Z}, Z^\star)$ over 50 Monte Carlo simulations. We compute $\mathrm{dist}^2(\hat{Z}, Z^\star) = \|\hat{Z} - Z^\star V U^\top\|_{\mathrm{F}}^2$ where $U\Sigma V^\top$ is the singular value decomposition of $\hat{Z}^\top Z^\star$; see Schönemann (1966). For the penalized MLE, we obtain $\hat{Z}$ following Remark 9 given true k; the corresponding errors of $\hat{G}$ follow similar patterns and are provided in Section H.1 of the Supplementary Material (He et al. (2025)).

Under Case (I), we present the empirical estimation errors of the one-step estimator and the penalized MLE in Figures 2 and 3, respectively. In each figure, panel (a) suggests that $\mathrm{dist}^2(\hat{Z}, Z^\star)$ is inverse proportional to T when n is fixed. Furthermore, panel (c) shows that all the fitted slopes are close to -1. In addition, panel (b) shows that $\mathrm{dist}^2(\hat{Z}, Z^\star)$ does not change too much as n increases. In conclusion, the numerical results are consistent with the oracle theoretical rate $O(1/T)$ for the estimation error of $Z^\star$. Comparing Figures 2 and 3, we find that under the same (n, T, k), the one-step estimator can achieve a slightly smaller estimation error than the penalized MLE. This might be because biases are introduced with the convex relaxation. However, the penalized MLE could be more flexible and robust when the true k is unknown in applications. Under Case (II), we present the empirical estimation errors $\mathrm{dist}^2(\hat{Z}, Z^\star)$ of the one-step estimator and the penalized MLE in Figures 4 and 5,



(a) $\mathrm{dist}^2(\hat{Z}, Z^\star)$ versus T (b) $\mathrm{dist}^2(\hat{Z}, Z^\star)$ versus n (c) Histogram of fitted slopes

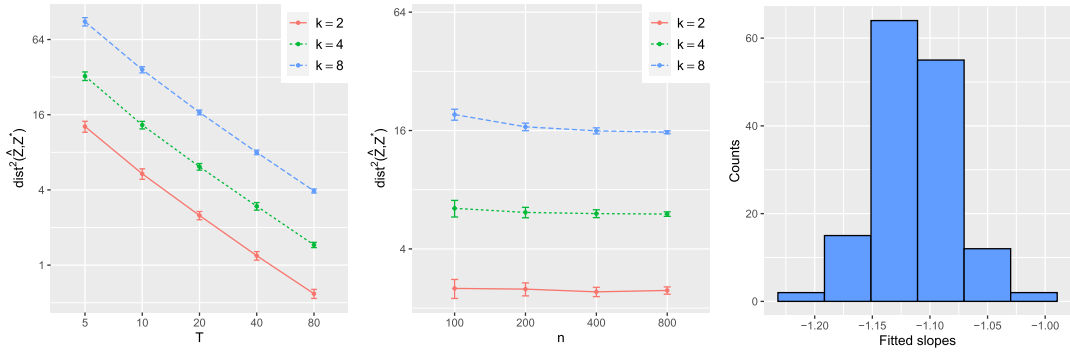
FIG. 2. *Case (I): Empirical estimation errors of the one-step estimator. Panel (a) presents* $\mathrm{dist}^2(\hat{Z}, Z^\star)$ *(averaged over* 50 *repetitions) versus* T *in the scenario (a). Panel (b) presents* $\mathrm{dist}^2(\hat{Z}, Z^\star)$ *(averaged over* 50 *repetitions) versus* n *in the scenario (b). In (a) and (b), axes are in the log scale, three lines correspond to results under* $k \in \{2, 4, 8\}$*, respectively, and error bars are obtained by* $\pm$ *the standard deviation from* 50 *repetitions. Panel (c) presents the slopes from regressing* $\log \mathrm{dist}^2(\hat{Z}, Z^\star)$ *on* $\log T$ *with fixed* $(n, k) \in \{200\} \times \{2, 4, 8\}$ *in the* 50 *repetitions under the scenario (a).*



(a) $\mathrm{dist}^2(\hat{Z}, Z^\star)$ versus T (b) $\mathrm{dist}^2(\hat{Z}, Z^\star)$ versus n (c) Histogram of fitted slopes

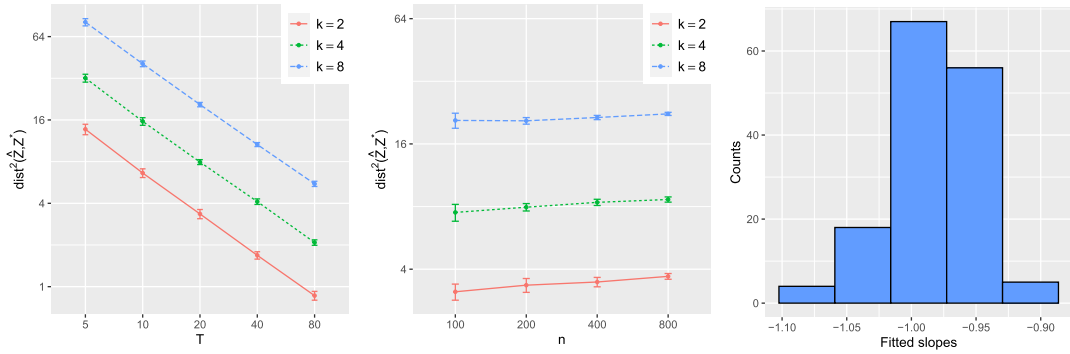
FIG. 3. *Case (I): Empirical estimation errors of the penalized MLE. Panels (a)–(c) are presented similarly to Figure 2.*

respectively. We can see that the patterns are similar to those in Figures 2–3. The results suggest the oracle theoretical rate $O(1/T)$ can hold under a variety of types of heterogeneity of $\alpha^\star$.



(a) $\mathrm{dist}^2(\hat{Z}, Z^\star)$ versus T (b) $\mathrm{dist}^2(\hat{Z}, Z^\star)$ versus n (c) Histogram of fitted slopes

FIG. 4. *Case (II): Empirical estimation errors of the one-step estimator. Panels (a)–(c) are presented similarly to Figure 2.*



(a) $\mathrm{dist}^2(\hat{Z}, Z^\star)$ versus T (b) $\mathrm{dist}^2(\hat{Z}, Z^\star)$ versus n (c) Histogram of fitted slopes

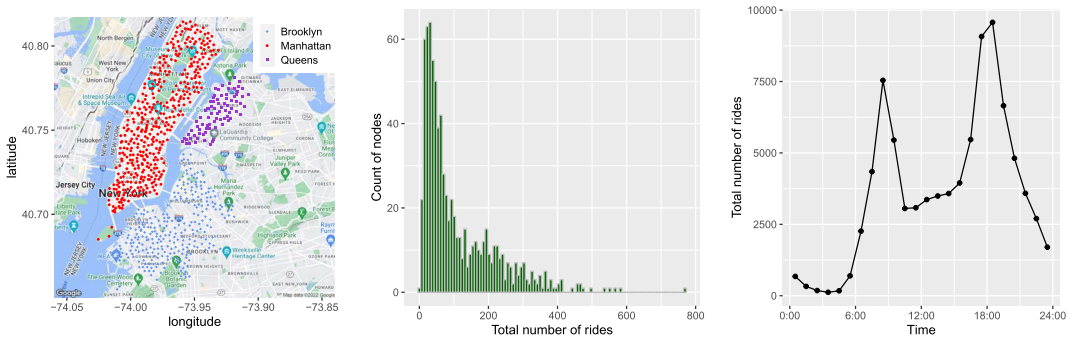
FIG. 5. *Case (II): Empirical estimation errors of the penalized MLE. Panels (a)–(c) are presented similarly to Figure 2.*

REMARK 10. When estimating latent vectors with the penalized MLE, we set the tuning parameter $\lambda_{n,T} = c_\lambda \sqrt{nT\hat{\mu}}$ with a fixed constant $c_\lambda = 0.5$ and $\hat{\mu} = \sum_{t,i,j} A_{t,ij}/(n^2 T)$ being the average over all entries. In general applications, practitioners may also choose c_λ using the network cross-validation (Chen and Lei (2018), Li, Levina and Zhu (2020)) or based on their data interpretation. Notably, when focusing on the aggregation effect over T networks, that is, how the estimation error changes with respect to T, we find that this phenomenon is not sensitive to the choice of c_λ. See more discussions in Section H.4 of the Supplementary Material (He et al. (2025)).

6. Analysis of New York Citi Bike dataset. We illustrate the use of the proposed methods by analyzing the New York Citi Bike data (CitiBike (2019)). The data set contains over 2.3 million rides between bike stations in New York in August 2019. Each ride is identified by two stations and a time stamp (the hire starting time). We focus on a weekday, August 1st, 2019, and keep the rides that last between one minute and 3 hours. The processed data contains 85,854 rides between 782 stations over 24 hours. The 782 bike stations form a common set of nodes across hours. A pair of stations defines a network edge, and the number of events between them in each hour gives the hourly edge weight. We provide exploratory data visualization in Figure 6. Panels (b) and (c) of Figure 6 show that the number of ride events is very heterogeneous across different bike stations and hours, respectively, motivating the application of the model (1).

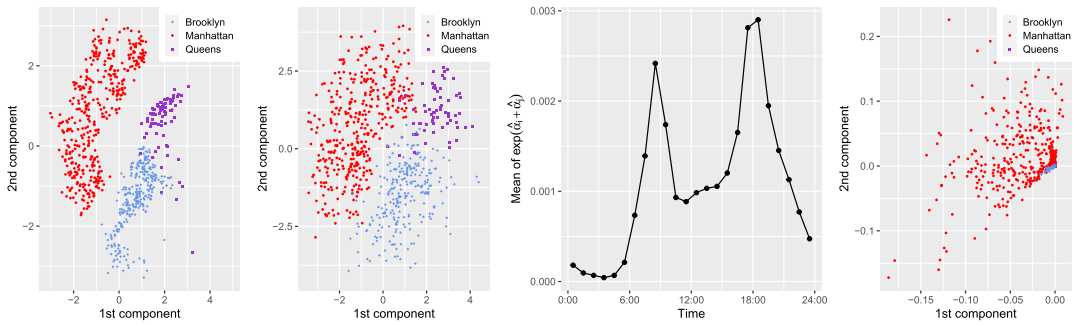
We next fit the model (1) with 2-dimensional latent vectors z_i's. The choice of $k = 2$ is for the ease of interpretation below and is consistent with the strategy in Remark 9. Figure 7(a) visualizes the estimated positions $\hat{z}_i$'s obtained by the one-step estimator, while the results of the penalized MLE are similar and deferred to the Supplementary Material (He et al. (2025)).

To interpret the results, we compare the estimated latent space vectors $\hat{z}_i$'s, visualized in Figure 7(a), with true geographic locations (in latitudes and longitudes) of stations, visualized in Figure 6(a). We can see that, overall, the estimated latent positions can match the true geographic positions. Particularly, the stations located in Manhattan are well separated from those in Brooklyn and Queens. This could be because the borough of Manhattan is separated from the other two boroughs by the East River, and it is less common to use shared bikes to travel between Manhattan and the other two boroughs. On the other hand, the estimated station locations in Brooklyn and Queens overlap. This could be because Queens and Brooklyn are not physically divided by a river, and it is easier to travel between these two boroughs by biking. In the meantime, some bike stations were estimated to be in the Central Park area.



(a) True Station Locations (b) Node-Level Histogram (c) Ride Counts Over Time

FIG. 6. *Illustration of Citi Bike Data*: *Panel* (*a*) *presents true locations of bike stations on Google Maps, colored by three boroughs of New York City. Panel* (*b*) *presents the histogram of the total number of rides by nodes* (*bike stations*). *Panel* (*c*) *presents the total number of rides over the* 24*-hour period.*



(a) 24-Hour Estimates (b) 1-Hour Estimates (c) Baseline Estimates (d) COSIE Estimates

FIG. 7. *Estimation Results*: *Panel* (*a*) *shows the estimated latent positions* $\hat{z}_i \in \mathbb{R}^2$ *using all the data across the* 24*-hour period. Each point is colored based on which borough the corresponding bike station is located, and we present the results after a rotation for better visualization. Panel* (*b*) *shows the estimated latent positions* $\hat{z}_i \in \mathbb{R}^2$ *using one-hour data during* 21:00-22:00 *on the selected day, where points are similarly colored and rotated to those in the panel* (*a*). *Panel* (*c*) *presents the mean of estimated baseline levels* $\sum_{i,j=1}^{n} \exp(\hat{\alpha}_{it} + \hat{\alpha}_{jt})/n^2$ *for* $t = 1, \ldots, 24$, *that is, across the* 24*-hour period under the model* (1). *Panel* (*d*) *presents the estimated two-dimensional shared latent vectors under COSIE with* 24*-hour data.*

This could be because people would ride through Central Park frequently, so two stations that are far away geographically might be viewed as close to each other based on the interactions. Besides the latent vectors, we visualize the estimated baseline levels across the 24-hour period in Figure 7(c). The results show a time-varying pattern that is consistent with the true changing pattern of the total number of rides across the 24-hour periods, visualized in Figure 6(c). In summary, the fitted parameters $\hat{z}_i$'s and $\hat{\alpha}_{it}$'s under the model (1) are reasonable and interpretable.

In addition, comparing panels (a) and (b) in Figure 7, we can see that the alignment between the estimated latent space vectors and the true geographic locations improves as we utilize observations over more hours. The results suggest that the proposed methods can effectively extract static latent space information and accommodate time heterogeneity simultaneously.

As a comparison, we fit the data by the common subspace independent edge (COSIE) model in Arroyo et al. (2021) with two-dimensional shared latent vectors. We estimate the model by the multiple adjacency spectral embedding (MASE) method in Arroyo et al. (2021). We present the estimated latent vectors in Figure 7(d). Compared to Figure 7(a), Figure 7(d) does not appear to show a direct connection between the estimated latent positions and the geographic positions of bike stations. This may be because Arroyo et al. (2021) targeted at $\mathbf{A}_t$'s that are adjacency matrices with binary edges. We present the results by COSIE only to illustrate the differences between different mean models of $\mathbf{A}_t$'s.

REMARK 11. The analyzed networks are undirected and mainly encode the total usage of bike rides between stations. The learned embeddings z_i's can be easy to interpret and useful for understanding the usage of shared bikes and for guiding the design of stations. Alternatively, one may examine bike rides with directional information, giving directed networks. Then the model (1) may be generalized as $\exp(\alpha_{it} + \beta_{jt} + \langle z_i, w_j \rangle)$, where for each vertex i, α_{it} and β_{it} characterize the "leave" and "arrival" baseline activity levels at the time point t, and z_i and w_i are time-invariant latent vectors characterizing the out-degree and in-degree properties, respectively. It would be an interesting future research direction to extend this study to directed networks.

7. Discussion. In this work, we propose a longitudinal latent space model tailored for recurrent interaction events. We develop two novel semiparametric estimation techniques, that is, the generalized semiparametric one-step updating and the penalized maximum likelihood estimation, and show that the resulting estimators attain the oracle estimation error rate for the shared latent structure. The first approach utilizes the semiparametric efficient score equation to construct a second-order updating estimator. We show that the estimator possesses a geometric interpretation on the quotient manifold, which automatically overcomes the nonuniqueness issue due to overparametrization. The second approach corresponds to a convex relaxation of the low-rank static latent space component.

By separating the (primary) parameters of interest associated with the static latent space from the dynamic nuisance parameters, a strategy commonly found in semiparametrically efficient parameter estimation, we are able to delineate the oracle rates of convergence for the primary and the nuisance parameters according to their dimensions. This strategy also helps us to untangle the static and time-heterogeneous components inherent in the network model and construct the oracle estimators.

There are a few other interesting future works. First, the ability to accurately estimate latent structures could enable important downstream analysis such as prediction, hypothesis testing, and change-point detection. For instance, it may be useful to ascertain a change point in the structure of the latent space (Bhattacharjee, Banerjee and Michailidis (2020), Enikeeva and Klopp (2021), Madrid Padilla, Yu and Priebe (2022)). The achievement of oracle estimation error rates could facilitate the quantification of uncertainty in estimators, which in turn lays a strong foundation for conducting reliable statistical inference.

Second, this paper focuses on the variance in estimation error rates as a function of n and T, while treating the latent dimension k and network sparsity level as fixed. Extending the current methodology and theory to the cases when k grows (Choi, Wolfe and Airoldi (2012)) as well as sparse networks (Qin and Rohe (2013), Le, Levina and Vershynin (2017)) are important topics.

Third, this work aims to unveil the fundamental relationship between the estimation errors and the degree of baseline heterogeneity. We focus on the most challenging scenario where the degree of baseline heterogeneity increases linearly with respect to n and T. It is possible to impose additional structures to reduce baseline heterogeneity, such as assuming $\{\alpha_{i1}, \ldots, \alpha_{iT}\}$ to be piecewise constants. Nevertheless, as T increases, the intrinsic number of parameters would eventually become large to keep up with the increasing data complexity. The developed results would also provide us with techniques for investigating such scenarios. Which structural assumptions are appropriate may vary across different applications and require case-by-case analyses in future research.

Fourth, the proposed model has the potential for further extensions to capture more complex network structures. Currently, heterogeneity across networks is only characterized at the first-order baseline levels α_{it}'s, while the second-order interaction terms are modeled by the time-invariant components z_i's. We find that this model adequately describes the analyzed dataset. But more generally, it may also be of interest to incorporate time-varying interaction terms, which would further increase the model complexity and pose new theoretical challenges. Moreover, the proposed model adopts the Euclidean inner product to describe the interactions between nodes, which can be limited to capturing homophilic network structures. Recently, researchers proposed to use indefinite inner products to capture heterophilic structures (Rubin-Delanchy et al. (2022), Lei (2021), MacDonald, Levina and Zhu (2022)). We discuss the feasibility of generalizing the proposed analysis to heterophilic products in Section J of the Supplementary Material (He et al. (2025)). While developing comprehensive results would require careful consideration of specific model properties, the foundational results and techniques developed under (1) pave the way for studying more general models.

Finally, it would be worthwhile to generalize our results to other models, including various distributions for weighted edges, continuous time stamps, or additional covariates influencing the network structure (Hoff, Raftery and Handcock (2002), Vu et al. (2011), Perry and Wolfe (2013), Hoff (2015), Kim et al. (2023), Sit, Ying and Yu (2021), Weng and Feng (2021), Huang, Sun and Feng (2023)). We believe that the proposed semiparametric analysis framework can function as a valuable building block for establishing sharp estimation error rates under those models.

APPENDIX: DETAILS FOR THE INITIALIZATION ALGORITHM

We present the details of the initialization Algorithm 1 below. More explanations on the details can be found in Section D of the Supplementary Material (He et al. (2025)).

Stage 1. Stage 1 utilizes the universal singular value thresholding (USVT) method in Chatterjee (2015) to obtain an "initial of initial" estimator. In particular, for $t = 1, \dots, T$, lines 3–4 of Algorithm 1 finds the singular value decomposition of $\mathbf{A}_t$ and retains the top singular components with the singular values $s_{t,l} > \tau_t$. Line 5 applies $f_{M_{\Theta,2}} : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ to $\widetilde{E}_t$ to ensure that all elements in $\mathring{E}_t$ are in $[e^{-M_{\Theta,2}}, 1]$. Specifically, given a matrix $X = (X_{ij})_{1\le i,j\le n} \in \mathbb{R}^{n\times n}$, the (i, j)th element of $f_{M_{\Theta,2}}(X)$ is defined as

$$\big(f_{M_{\Theta,2}}(X)\big)_{ij} = \begin{cases} X_{ij} & \text{if } e^{-M_{\Theta,2}} \le X_{ij} \le 1; \\ e^{-M_{\Theta,2}} & \text{if } X_{ij} < e^{-M_{\Theta,2}}; \\ 1 & \text{if } X_{ij} > 1. \end{cases} \tag{12}$$

Line 10 projects $\widetilde{G}$ into $\mathbb{S}^n_+$ and obtain $\mathring{G}$. In particular, $\mathcal{P}_{\mathbb{S}^n_+}(\cdot)$ represents the projection mapping onto $\mathbb{S}^n_+$, that is, for $X \in \mathbb{R}^{n\times n}$, we define

$$\mathcal{P}_{\mathbb{S}^n_+}(X) = \operatorname*{arg\,min}_{Y\in\mathbb{S}^n_+} \|X - Y\|, \tag{13}$$

where $\|\cdot\|$ represents the Euclidean norm. In line 11, we define $U_k = (u_1, \dots, u_k) \in \mathbb{R}^{n\times k}$ and $D_k = \operatorname{diag}(d_1, \dots, d_k) \in \mathbb{R}^{k\times k}$, where $d_1 \ge \cdots \ge d_k$ are the largest k eigenvalues of $\mathring{G}$, and $u_j \in \mathbb{R}^n$ is the eigenvector of $\mathring{G}$ corresponding to d_j for $j = 1, \dots, k$.

Stage 2. Stage 2 sets $(\mathring{Z}, \mathring{\alpha})$ obtained in Stage 1 as the initial estimator (line 13) and consists of two iterative sub-stages.

Stage 2-1 (*lines* 14–18). For $r = 1, \dots, R_1 - 1$, let (Z^r, α^r) denote the updated parameters along gradients after the rth iteration. Lines 15–16 of Algorithm 1 updates the parameters (Z^r, α^r) along the directions

$$g_Z(Z^r, \alpha^r) = \sum_{t=1}^{T} \big(\mathbf{A}_t - \exp(\Theta_t^r)\big) Z^r \in \mathbb{R}^{n\times k}, \tag{14}$$

$$g_{\alpha_t}(Z^r, \alpha^r) = \big(\mathbf{A}_t - \exp(\Theta_t^r)\big) 1_n \in \mathbb{R}^{n\times 1} \tag{15}$$

with $\Theta_t^r = \alpha_t^r 1_n^\top + 1_n(\alpha_t^r)^\top + Z^r(Z^r)^\top$ for $t = 1, \dots, T$. Lines 17–18 projects the updated estimator $(\widetilde{Z}^{r+1}, \widetilde{\alpha}^{r+1})$ onto the following sets:

$$\mathcal{C}_Z = \Big\{ Z \in \mathbb{R}^{n\times k} : 1_n^\top Z = 0, \max_{1\le i\le n} \|z_i\|_2^2 \le M_{Z,1} \Big\}, \tag{16}$$

$$\mathcal{C}_\alpha = \Big\{ \alpha \in \mathbb{R}^{n\times T} : \max_{1\le i\le n, 1\le t\le T} |\alpha_{it}| \le M_\alpha \Big\}, \tag{17}$$

$$\mathcal{C}'_Z = \Big\{ Z \in \mathbb{R}^{n\times k} : \max_{1\le i\le n} \|z_i\|_2^2 \le M_{Z,1} \Big\}. \tag{18}$$

Algorithm 1: Initialization algorithm for the semiparametric one-step estimator

Input: Data: $\mathbf{A}_1, \ldots, \mathbf{A}_T \in \mathbb{R}^{n\times n}$; thresholds: $\tau_1, \ldots, \tau_T$; latent space dimension: k; step sizes: η_Z, η_α; number of iterations: R_1, R_2.

Output: $(\check{\alpha}, \check{Z})$.

1 **Stage 1:**

2 **for** $t = 1, \ldots, T$ **do**

3 Find the singular value decomposition of $\mathbf{A}_t = \sum_{l=1}^{n} s_{t,l} u_{t,l} v_{t,l}^\top$, where for $l = 1, \ldots, n$, $s_{t,l} \in \mathbb{R}$, $u_{t,l} \in \mathbb{R}^n$, and $v_{t,l} \in \mathbb{R}^n$ represent the singular values, the left singular vectors, and the right singular vectors of $\mathbf{A}_t$, respectively.

4 Let $\widetilde{E}_t = \sum_{\{l: s_{t,l} > \tau_t\}} s_{t,l} u_{t,l} v_{t,l}^\top$.

5 Let $\mathring{E}_t = f_{M_{\Theta,2}}(\widetilde{E}_t)$, where $f_{M_{\Theta,2}}(x)$ is defined in (12).

6 Let $\mathring{\Theta}_t = \log(\mathring{E}_t)$.

7 Let $\mathring{\alpha}_t = \arg\min_\alpha \|\mathring{\Theta}_t - \alpha 1_n^\top - 1_n \alpha^\top\|_{\mathrm{F}}^2 = (n\mathrm{I}_n + 1_n 1_n^\top)^{-1} \mathring{\Theta}_t 1_n$.

8 **end**

9 Let $\widetilde{G} = \sum_{t=1}^{T} (\mathring{\Theta}_t - \mathring{\alpha}_t 1_n^\top - 1_n \mathring{\alpha}_t^\top)/T$.

10 Let $\mathring{G} = \mathcal{P}_{\mathbb{S}_+^n}(\widetilde{G})$, where $\mathcal{P}_{\mathbb{S}_+^n}(\cdot)$ is defined in (13).

11 Let $\mathring{Z} = \mathring{U}_k \mathring{D}_k^{1/2}$ where $\mathring{U}_k \mathring{D}_k \mathring{U}_k^\top$ is the top-k eigen components of $\mathring{G}$.

12 **Stage 2:**

13 Let $Z^0 = \mathring{Z}$ and $\alpha^0 = \mathring{\alpha} = (\mathring{\alpha}_1, \ldots, \mathring{\alpha}_T)$.

14 **for** $r = 0, 1, \ldots, R_1 - 1$ **do**

15 $\widetilde{Z}^{r+1} = Z^r + \eta_Z g_Z(Z^r, \alpha^r)$ with $g_Z(Z^r, \alpha^r)$ in (14).

16 $\widetilde{\alpha}_t^{r+1} = \alpha_t^r + \eta_\alpha g_{\alpha_t}(Z^r, \alpha^r)$ with $g_{\alpha_t}(Z^r, \alpha^r)$ in (15) for $t = 1, \ldots, T$.

17 $Z^{r+1} = \mathcal{P}_{\mathcal{C}_Z}(\widetilde{Z}^{r+1})$ with $\mathcal{C}_Z$ defined in (16).

18 $\alpha^{r+1} = \mathcal{P}_{\mathcal{C}_\alpha}(\widetilde{\alpha}^{r+1})$ with $\mathcal{C}_\alpha$ in (17).

19 **end**

20 **for** $r = R_1, \ldots, R_2 - 1$ **do**

21 $\widetilde{Z}^{r+1} = Z^r + \eta_Z g_{Z,R_1}(Z^r, \alpha^r)$ with $g_{Z,R_1}(Z^r, \alpha^r)$ in (19).

22 $\widetilde{\alpha}_t^{r+1} = \alpha_t^r + \eta_\alpha g_{\alpha_t,R_1}(Z^r, \alpha^r)$ with $g_{\alpha_t,R_1}(Z^r, \alpha^r)$ in (20) for $t = 1, \ldots, T$.

23 $Z^{r+1} = \mathcal{P}_{\mathcal{C}'_Z}(\widetilde{Z}^{r+1})$ with $\mathcal{C}'_Z$ defined in (18).

24 $\alpha^{r+1} = \mathcal{P}_{\mathcal{C}_\alpha}(\widetilde{\alpha}^{r+1})$ with $\mathcal{C}_\alpha$ defined in (17).

25 **end**

26 Let $\check{\alpha} = \alpha^{R_2}$ and $\check{Z} = J Z^{R_2}$, where J is defined in (21).

Similarly to (13), for $\mathcal{C} \subseteq \mathbb{R}^{m\times q}$ and $X \in \mathbb{R}^{m\times q}$, let $\mathcal{P}_{\mathcal{C}}(X) = \arg\min_{Y\in\mathcal{C}} \|Y - X\|$. The projection mapping is uniquely defined as $\mathcal{C}_Z$, $\mathcal{C}_\alpha$, and $\mathcal{C}'_Z$ are closed and convex sets (Section 8.1, Boyd and Vandenberghe (2004)).

Stage 2-2 (*lines* 20–25). The second part of Stage 2 utilizes an alternating version of the projected gradient descent to obtain the error bound of individual rows of Z and individual elements in α. In particular, in lines 21–22, we define

$$g_{Z,R_1}(Z^r, \alpha^r) = \sum_{t=1}^{T} (\mathbf{A}_t - \exp(\Theta_t^{r,R_1})) Z^{R_1} \in \mathbb{R}^{n\times k}, \tag{19}$$

$$g_{\alpha_t,R_1}(Z^r,\alpha^r)=(\mathbf{A}_t-\exp(\Theta_t^{r,R_1}))1_n\in\mathbb{R}^{n\times 1}, \tag{20}$$

where $\Theta_t^{r,R_1}=\alpha_t^r 1_n^\top+1_n(\alpha_t^{R_1})^\top+Z^r(Z^{R_1})^\top$ with $\alpha_t^{R_1}$ and $Z^{R_1}=(z_1^{R_1},\ldots,z_n^{R_1})^\top$ being estimates from *Stage 2-1*. Compared to Θ_t^r in (14)–(15), Θ_t^{r,R_1} is defined with part of the parameters fixed at $\alpha_t^{R_1}$ and Z^{R_1}.

The final output of Algorithm 1 is $\check{\alpha}=\alpha^{R_2}$ and $\check{Z}=JZ^{R_2}$ with

$$J=\mathrm{I}_n-\frac{1}{n}1_n1_n^\top. \tag{21}$$

We call J the centering matrix as $1_n^\top J=0$, which gives $1_n^\top\check{Z}=0$.

CONDITION 3. *Assume the tuning parameters in Algorithm 1 satisfy the following*:

(i) *Thresholds* $\tau_t=\tau n^{1/2}$ *for some large enough constant* $\tau>0$.
(ii) *Step sizes* $\eta_Z=\eta/(T\|Z^0\|_{\mathrm{op}}^2)$ *and* $\eta_\alpha=\eta/(2n)$ *for some small constant* $\eta>0$.
(iii) *Numbers of iterations* R_1 *and* R_2 *are sufficiently large*.

REMARK 12. In simulations, we choose parameters following Remark D.2 in the Supplementary Material (He et al. (2025)). Condition 3 on parameters are imposed for the convenience of proofs. It essentially requires that the tuning parameters satisfy certain orders with respect to n and T and can be approximated by the practical choices in Remark D.2.

Acknowledgments. We are grateful to the Editor, Dr. Enno Mammen, an Associate Editor, and two anonymous referees for their helpful comments and suggestions.

Y. Feng is the corresponding author.

Funding. Z. Ying was partially supported by NSF Grant DMS-2015417.

Y. Feng was partially supported by NIH Grant 1R21AG074205-01, NSF Grant DMS-2324489, NYU University Research Challenge Fund, and a grant from NYU School of Global Public Health.

Y. He, Sun, and Tian contributed equally to this work.

SUPPLEMENTARY MATERIAL

Supplement to "Semiparametric modeling and analysis for longitudinal network data." (DOI: 10.1214/25-AOS2506SUPP; .pdf). Due to space limitation, additional results and proofs are deferred to the supplementary material. The codes are published at He et al. (2025).

REFERENCES

ABSIL, P.-A., MAHONY, R. and SEPULCHRE, R. (2008). *Optimization Algorithms on Matrix Manifolds*. Princeton Univ. Press, Princeton, NJ. MR2364186 https://doi.org/10.1515/9781400830244

AGTERBERG, J., LUBBERTS, Z. and ARROYO, J. (2022). Joint spectral clustering in multilayer degree-corrected stochastic blockmodels. arXiv preprint. Available at arXiv:2212.05053.

ANDERSEN, P. K. and GILL, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10** 1100–1120. MR0673646

ARROYO, J., ATHREYA, A., CAPE, J., CHEN, G., PRIEBE, C. E. and VOGELSTEIN, J. T. (2021). Inference for multiple heterogeneous networks with a common invariant subspace. *J. Mach. Learn. Res.* **22** Paper No. 142, 49. MR4318498

ATHREYA, A., FISHKIND, D. E., TANG, M., PRIEBE, C. E., PARK, Y., VOGELSTEIN, J. T., LEVIN, K., LYZINSKI, V., QIN, Y. et al. (2018). Statistical inference on random dot product graphs: A survey. *J. Mach. Learn. Res.* **18** Paper No. 226, 92. MR3827114

BAZZI, M., JEUB, L. G., ARENAS, A., HOWISON, S. D. and PORTER, M. A. (2020). A framework for the construction of generative models for mesoscale structure in multilayer networks. *Phys. Rev. Res.* **2** 023100.

BEN-ISRAEL, A. and GREVILLE, T. N. E. (2003). *Generalized Inverses: Theory and Applications*, 2nd ed. *CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC* **15**. Springer, New York. MR1987382

BHATTACHARJEE, M., BANERJEE, M. and MICHAILIDIS, G. (2020). Change point estimation in a dynamic stochastic block model. *J. Mach. Learn. Res.* **21** Paper No. 107, 59. MR4119175

BICKEL, P., CHOI, D., CHANG, X. and ZHANG, H. (2013). Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels. *Ann. Statist.* **41** 1922–1943. MR3127853 https://doi.org/10.1214/13-AOS1124

BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. *Johns Hopkins Series in the Mathematical Sciences*. Johns Hopkins Univ. Press, Baltimore, MD. MR1245941

BOUMAL, N. (2023). *An Introduction to Optimization on Smooth Manifolds*. Cambridge Univ. Press, Cambridge. MR4533407

BOYD, S. and VANDENBERGHE, L. (2004). *Convex Optimization*. Cambridge Univ. Press, Cambridge. MR2061575 https://doi.org/10.1017/CBO9780511804441

BUTTS, C. T. (2008). A relational event framework for social action. *Sociol. Method.* **38** 155–200.

CAI, T., XIA, D., ZHANG, L. and ZHOU, D. (2022). Consensus knowledge graph learning via multi-view sparse low rank block model. arXiv preprint. Available at arXiv:2209.13762.

CANDÈS, E. J. and RECHT, B. (2012). Exact matrix completion via convex optimization. *Commun. ACM* **55** 111–119.

CANDÈS, E. J. and TAO, T. (2010). The power of convex relaxation: Near-optimal matrix completion. *IEEE Trans. Inf. Theory* **56** 2053–2080. MR2723472 https://doi.org/10.1109/TIT.2010.2044061

CHATTERJEE, S. (2015). Matrix estimation by universal singular value thresholding. *Ann. Statist.* **43** 177–214. MR3285604 https://doi.org/10.1214/14-AOS1272

CHEN, K. and LEI, J. (2018). Network cross-validation for determining the number of communities in network data. *J. Amer. Statist. Assoc.* **113** 241–251. MR3803461 https://doi.org/10.1080/01621459.2016.1246365

CHEN, S., LIU, S. and MA, Z. (2022). Global and individualized community detection in inhomogeneous multilayer networks. *Ann. Statist.* **50** 2664–2693. MR4500621 https://doi.org/10.1214/22-aos2202

CHEN, Y. and WAINWRIGHT, M. J. (2015). Fast low-rank estimation by projected gradient descent: General statistical and algorithmic guarantees. arXiv preprint. Available at arXiv:1509.03025.

CHERNOZHUKOV, V., CHETVERIKOV, D., DEMIRER, M., DUFLO, E., HANSEN, C., NEWEY, W. and ROBINS, J. (2018). Double/debiased machine learning for treatment and structural parameters. *Econom. J.* **21** C1–C68. MR3769544 https://doi.org/10.1111/ectj.12097

CHOI, D. S., WOLFE, P. J. and AIROLDI, E. M. (2012). Stochastic blockmodels with a growing number of classes. *Biometrika* **99** 273–284. MR2931253 https://doi.org/10.1093/biomet/asr053

CITIBIKE (2019). New York Citi Bike Data in August 2019. https://github.com/cedoula/bikesharing. [Original source: https://citibikenyc.com/system-data].

COOK, R. J. and LAWLESS, J. F. (2007). *The Statistical Analysis of Recurrent Events*. *Statistics for Biology and Health*. Springer, New York. MR3822124

DAVENPORT, M. A., PLAN, Y., VAN DEN BERG, E. and WOOTTERS, M. (2014). 1-bit matrix completion. *Inf. Inference* **3** 189–223. MR3311452 https://doi.org/10.1093/imaiai/iau006

ENIKEEVA, F. and KLOPP, O. (2021). Change-point detection in dynamic networks with missing links. arXiv preprint. Available at arXiv:2106.14470.

HE, Y., SUN, J., TIAN, Y., YING, Z. and FENG, Y. (2025). Supplement to "Semiparametric modeling and analysis for longitudinal network data." https://doi.org/10.1214/25-AOS2506SUPP

HOFF, P. D. (2003). Random effects models for network data. *Citeseer*.

HOFF, P. D. (2005). Bilinear mixed-effects models for dyadic data. *J. Amer. Statist. Assoc.* **100** 286–295. MR2156838 https://doi.org/10.1198/016214504000001015

HOFF, P. D. (2011). Hierarchical multilinear models for multiway data. *Comput. Statist. Data Anal.* **55** 530–543. MR2736574 https://doi.org/10.1016/j.csda.2010.05.020

HOFF, P. D. (2015). Multilinear tensor regression for longitudinal relational data. *Ann. Appl. Stat.* **9** 1169–1193. MR3418719 https://doi.org/10.1214/15-AOAS839

HOFF, P. D., RAFTERY, A. E. and HANDCOCK, M. S. (2002). Latent space approaches to social network analysis. *J. Amer. Statist. Assoc.* **97** 1090–1098. MR1951262 https://doi.org/10.1198/016214502388618906

HOLLAND, P. W., LASKEY, K. B. and LEINHARDT, S. (1983). Stochastic blockmodels: First steps. *Soc. Netw.* **5** 109–137. MR0718088 https://doi.org/10.1016/0378-8733(83)90021-7

HOLME, P. (2015). Modern temporal network theory: A colloquium. *Eur. Phys. J. B* **88** 1–30.

HUANG, S., SUN, J. and FENG, Y. (2023). PCABM: Pairwise covariates-adjusted block model for community detection. *J. Amer. Statist. Assoc.* **119** 2092–2104. MR4797925 https://doi.org/10.1080/01621459.2023.2244731

JIN, J. (2015). Fast community detection by SCORE. *Ann. Statist.* **43** 57–89. MR3285600 https://doi.org/10.1214/14-AOS1265

JONES, A. and RUBIN-DELANCHY, P. (2020). The multilayer random dot product graph. arXiv preprint. Available at arXiv:2007.10455.

KARRER, B. and NEWMAN, M. E. J. (2011). Stochastic blockmodels and community structure in networks. *Phys. Rev. E* **83** 016107, 10. MR2788206 https://doi.org/10.1103/PhysRevE.83.016107

KIM, B., NIU, X., HUNTER, D. and CAO, X. (2023). A dynamic additive and multiplicative effects network model with application to the United Nations voting behaviors. *Ann. Appl. Stat.* **17** 3283–3299. MR4661698 https://doi.org/10.1214/23-aoas1762

KLIMT, B. and YANG, Y. (2004). The Enron corpus: A new dataset for email classification research. In *European Conference on Machine Learning* 217–226. Springer, Berlin.

LE, C. M., LEVINA, E. and VERSHYNIN, R. (2017). Concentration and regularization of random graphs. *Random Structures Algorithms* **51** 538–561. MR3689343 https://doi.org/10.1002/rsa.20713

LEE, C. and WILKINSON, D. J. (2019). A review of stochastic block models and extensions for graph clustering. *Appl. Netw. Sci.* **4** 1–50.

LEE, J. M. (2013). *Introduction to Smooth Manifolds*, 2nd ed. *Graduate Texts in Mathematics* **218**. Springer, New York. MR2954043

LEI, J. (2021). Network representation using graph root distributions. *Ann. Statist.* **49** 745–768. MR4255106 https://doi.org/10.1214/20-aos1976

LEI, J., CHEN, K. and LYNCH, B. (2020). Consistent community detection in multi-layer network data. *Biometrika* **107** 61–73. MR4064140 https://doi.org/10.1093/biomet/asz068

LEI, J. and LIN, K. Z. (2023). Bias-adjusted spectral clustering in multi-layer stochastic block models. *J. Amer. Statist. Assoc.* **118** 2433–2445. MR4681594 https://doi.org/10.1080/01621459.2022.2054817

LEVIN, K., ATHREYA, A., TANG, M., LYZINSKI, V. and PRIEBE, C. E. (2017). A central limit theorem for an omnibus embedding of multiple random dot product graphs. In 2017 *IEEE International Conference on Data Mining Workshops* (*icdmw*) 964–967. IEEE Comput. Soc., Los Alamitos.

LI, T., LEVINA, E. and ZHU, J. (2020). Network cross-validation by edge sampling. *Biometrika* **107** 257–276. MR4108931 https://doi.org/10.1093/biomet/asaa006

LIN, D. Y., WEI, L. J., YANG, I. and YING, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **62** 711–730. MR1796287 https://doi.org/10.1111/1467-9868.00259

LINDERMAN, S. and ADAMS, R. (2014). Discovering latent network structure in point process data. In *International Conference on Machine Learning* 1413–1421.

LITTLE, M. P., HEIDENREICH, W. F. and LI, G. (2010). Parameter identifiability and redundancy: Theoretical considerations. *PLoS ONE* **5** e8915.

LYU, Z., XIA, D. and ZHANG, Y. (2023). Latent space model for higher-order networks and generalized tensor decomposition. *J. Comput. Graph. Statist.* **32** 1320–1336. MR4669249 https://doi.org/10.1080/10618600.2022.2164289

MA, Z., MA, Z. and YUAN, H. (2020). Universal latent space model fitting for large networks with edge covariates. *J. Mach. Learn. Res.* **21** Paper No. 4, 67. MR4071187

MACDONALD, P. W., LEVINA, E. and ZHU, J. (2022). Latent space models for multiplex networks with shared structure. *Biometrika* **109** 683–706. MR4472842 https://doi.org/10.1093/biomet/asab058

MADRID PADILLA, O. H., YU, Y. and PRIEBE, C. E. (2022). Change point localization in dependent dynamic nonparametric random dot product graphs. *J. Mach. Learn. Res.* **23** Paper No. [234], 59. MR4577673

MATIAS, C. and MIELE, V. (2017). Statistical clustering of temporal networks through a dynamic stochastic block model. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 1119–1141. MR3689311 https://doi.org/10.1111/rssb.12200

MATIAS, C., REBAFKA, T. and VILLERS, F. (2018). A semiparametric extension of the stochastic block model for longitudinal networks. *Biometrika* **105** 665–680. MR3842891 https://doi.org/10.1093/biomet/asy016

MATIAS, C. and ROBIN, S. (2014). Modeling heterogeneity in random graphs through latent space models: A selective review. In *MMCS*, *Mathematical Modelling of Complex Systems*. *ESAIM Proc. Surveys* **47** 55–74. EDP Sci., Les Ulis. MR3419385 https://doi.org/10.1051/proc/201447004

MURPHY, S. A. and VAN DER VAART, A. W. (2000). On profile likelihood. *J. Amer. Statist. Assoc.* **95** 449–465. MR1803168 https://doi.org/10.2307/2669386

NIELSEN, A. M. and WITTEN, D. (2018). The multiple random dot product graph model. arXiv preprint. Available at arXiv:1811.12172.

PAUL, S. and CHEN, Y. (2022). Null models and community detection in multi-layer networks. *Sankhya A* **84** 163–217. MR4402750 https://doi.org/10.1007/s13171-021-00257-0

PENSKY, M. and ZHANG, T. (2019). Spectral clustering in the dynamic stochastic block model. *Electron. J. Stat.* **13** 678–709. MR3914178 https://doi.org/10.1214/19-ejs1533

PEPE, M. S. and CAI, J. (1993). Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. *J. Amer. Statist. Assoc.* **88** 811–820.

PERRY, P. O. and WOLFE, P. J. (2013). Point process modelling for directed interaction networks. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 821–849. MR3124793 https://doi.org/10.1111/rssb.12013

PORTNOY, S. (1988). Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity. *Ann. Statist.* **16** 356–366. MR0924876 https://doi.org/10.1214/aos/1176350710

QIN, T. and ROHE, K. (2013). Regularized spectral clustering under the degree-corrected stochastic blockmodel. *Adv. Neural Inf. Process. Syst.* **26**.

RUBIN-DELANCHY, P., CAPE, J., TANG, M. and PRIEBE, C. E. (2022). A statistical interpretation of spectral embedding: The generalised random dot product graph. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 1446–1473. MR4494166 https://doi.org/10.1111/rssb.12509

SALTER-TOWNSHEND, M. and MCCORMICK, T. H. (2017). Latent space models for multiview network data. *Ann. Appl. Stat.* **11** 1217–1244. MR3709558 https://doi.org/10.1214/16-AOAS955

SCHEFFÉ, H. (1999). *The Analysis of Variance*. *Wiley Classics Library*. Wiley, New York. Reprint of the 1959 original, a Wiley Publication in Mathematical Statistics. MR1673563

SCHÖNEMANN, P. H. (1966). A generalized solution of the orthogonal Procrustes problem. *Psychometrika* **31** 1–10. MR0215870 https://doi.org/10.1007/BF02289451

SEVERINI, T. A. and WONG, W. H. (1992). Profile likelihood and conditionally parametric models. *Ann. Statist.* **20** 1768–1802. MR1193312 https://doi.org/10.1214/aos/1176348889

SEWELL, D. K. and CHEN, Y. (2015). Latent space models for dynamic networks. *J. Amer. Statist. Assoc.* **110** 1646–1657. MR3449061 https://doi.org/10.1080/01621459.2014.988214

SIT, T., YING, Z. and YU, Y. (2021). Event history analysis of dynamic networks. *Biometrika* **108** 223–230. MR4226200 https://doi.org/10.1093/biomet/asaa045

SUN, J. and ZHAO, X. (2013). *Statistical Analysis of Panel Count Data*. *Statistics for Biology and Health*. Springer, New York. MR3136574 https://doi.org/10.1007/978-1-4614-8715-9

TSIATIS, A. A. (2006). *Semiparametric Theory and Missing Data*. *Springer Series in Statistics*. Springer, New York. MR2233926

VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247 https://doi.org/10.1017/CBO9780511802256

VU, D., HUNTER, D., SMYTH, P. and ASUNCION, A. (2011). Continuous-time regression models for longitudinal networks. In *Advances in Neural Information Processing Systems* 2492–2500.

WENG, H. and FENG, Y. (2021). Community detection with nodal information: Likelihood and its variational approximation. *Stat* **11** Paper No. e428, 17. MR4394988 https://doi.org/10.1002/sta4.428

XU, K. S. and HERO, A. O. (2014). Dynamic stochastic blockmodels for time-evolving social networks. *IEEE J. Sel. Top. Signal Process.* **8** 552–562.

YANG, T., CHI, Y., ZHU, S., GONG, Y. and JIN, R. (2011). Detecting communities and their evolutions in dynamic social networks—a Bayesian approach. *Mach. Learn.* **82** 157–189. MR3108191 https://doi.org/10.1007/s10994-010-5214-7

YOUNG, S. J. and SCHEINERMAN, E. R. (2007). Random dot product graph models for social networks. In *Algorithms and Models for the Web-Graph*. *Lecture Notes in Computer Science* **4863** 138–149. Springer, Berlin. MR2504912 https://doi.org/10.1007/978-3-540-77004-6_11

ZHANG, H., CHEN, Y. and LI, X. (2020). A note on exploratory item factor analysis by singular value decomposition. *Psychometrika* **85** 358–372. MR4128065 https://doi.org/10.1007/s11336-020-09704-7

ZHANG, H. and WANG, J. (2023). Efficient estimation for longitudinal network via adaptive merging. arXiv preprint. Available at arXiv:2211.07866.

ZHANG, J. and CHEN, Y. (2020). Modularity based community detection in heterogeneous networks. *Statist. Sinica* **30** 601–629. MR4213981

ZHANG, J., SUN, W. W. and LI, L. (2020). Mixed-effect time-varying network model and application in brain connectivity analysis. *J. Amer. Statist. Assoc.* **115** 2022–2036. MR4189774 https://doi.org/10.1080/01621459.2019.1677242

ZHANG, X., XUE, S. and ZHU, J. (2020). A flexible latent space model for multilayer networks. In *International Conference on Machine Learning* 11288–11297.

ZHAO, Y., LEVINA, E. and ZHU, J. (2012). Consistency of community detection in networks under degree-corrected stochastic block models. *Ann. Statist.* **40** 2266–2292. MR3059083 https://doi.org/10.1214/12-AOS1036

ZHENG, R. and TANG, M. (2022). Limit results for distributed estimation of invariant subspaces in multiple networks inference and PCA. arXiv preprint. Available at arXiv:2206.04306.